Improved detection of pancreatic neuroendocrine tumours through the development of a biomarker panel

Ponni Balasundaram

University College London (UCL)

Division of Medicine
Institute for Liver and Digestive Health
Royal Free Campus
London

Thesis submitted for the degree of Doctor of Philosophy

Declaration

I Ponni Balasundaram confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Pancreatic neuroendocrine tumour (pNET) patients present with non-specific symptoms leading to delayed diagnosis which in turn impacts survival. The current gold standard marker Chromogranin A (CgA) has limitations, including associated confounding factors such as proton pump inhibitor use and limited utility for insulinomas. In clinical practice, pNETs need to be discriminated from other pancreatic conditions such as pancreatic ductal adenocarcinoma (PDAC), acute pancreatitis (AP) and chronic pancreatitis (CP), but currently no test is available for this distinction to be made.

A multianalyte biomarker approach for pNET detection, using seven previously explored GEP-NET markers was explored for their suitability for pNET detection. A training cohort of pNET patient and healthy control samples was used to develop machine learning (ML) algorithms across seven different marker combinations. Based on this work, a three-marker combination of CgA, VGF-nerve growth factor inducible peptide (VGF-NGF) and Angiopoietin-2 (ANG2) was identified as suitable across the algorithms assessed, and that the seven markers could be reduced to three markers without a large impact on performance, meaning a more cost-effective test using fewer markers. The suitability of this three-marker combination was also confirmed in internal validation. However, at external validation using an independent sample of pNETs, VGF-NGF was deemed not to be suitable and the performance of ANG2 and CgA algorithms was less than that seen using the training cohort. Models created based on the training pNET case and healthy control evaluation were also not found to discriminate between AP, CP, PDAC and pNETs.

In summary, reduction to a two or three-marker pNET detection panel performed well in the ML model training stage but did not show strong performance in external validation. Further markers, perhaps identified in early pNET biological models are likely to be needed in combination for better identification in a point of care test.

Impact Statement

The work from my PhD explored the role that machine learning (ML) algorithms can play in the diagnosis of pNET patients using biomarkers. The methods of ML applied to this work such as C5.0 decision trees, logistic regression, support vector machines and random forest's allowed for ML algorithms which have a widespread utility in healthcare to be used in an application to diagnose pNET patients; this therefore had the impact to improve diagnosis in these patients. The thesis explored these ML algorithms in detail using different approaches for validation including a train test approach, k-fold cross validation and external validation using independent cohorts. The work from my thesis illustrated some of the limitations that these ML algorithms and validation approaches can have. My thesis also highlighted the challenges in biomarker work, in that certain biomarkers that hold promise in preliminary stages can fall through at different points as illustrated in this thesis. The thesis began with the analysis of seven GEP-NET markers which lead to three markers to be taken forward to external validation and from this only two markers evaluated in the ML models. Moreover, the work from this thesis has shown that different approaches are needed when trying to diagnose pNET patients from healthy controls, compared to when pNET patients need to be differentially diagnosed from AP, CP and PDAC patients and that one a strategy fits all is not the answer. The use of independent pNET cohorts was a particular strength in this work. The work also brought to light the limitations that exist for the established NET marker CgA. As there is no standardised FDA approved test, with many ways to assay CgA available, this thesis illustrated this challenge as well as highlighting this for the wider field of NET research. As if ML algorithms using CgA were to be adopted into clinical practice, having standardisation as part of this process would be key. My thesis work also illustrated the need for identifying newer pNET markers, particularly for early stage whereby the impacts of a biomarker test would be most beneficial for pNET patients in terms of survival. This would mean developing newer experimental models that better recapitulate the early pNET tumour.

Acknowledgements

Abbreviations

A&E   Accident and emergency
ASCL1   Achaete-scute family BHLH transcription factor 1
ACTHoma   Adrenocorticotropic hormone-producing pancreatic neuroendocrine
             neoplasm
AI   Artificial intelligence
AJCC   American Joint Committee on Cancer
ANG   Angiopoietin
ANN   Artificial neural network
ANOVA   Analysis of variance
AORs   Adjusted odds ratio
AP   Acute pancreatitis
ARVCF   Armadillo repeat protein deleted in velo-cardio-facial-syndrome
ASPEN   Asymptomatic sporadic non-functioning pancreatic neuroendocrine
ATRX   ATRX chromatin remodeler
AUC   Area under the curve
Bcl-2   B-cell lymphoma 2
BDNF   Brain-derived neurotrophic factor
BET   Bromodomain and extra terminal domain
BMI   Body mass index
BRCA   Breast cancer gene
CA19-9   Carbohydrate antigen 19-9
cAMP   Cyclic adenosine monophosphate
CAP   Capecitabine
CART   Classification and regression tree
CART   Cocaine and amphetamine regulated transcript
CDKN1A   Cyclin dependent kinase inhibitor 1
CEA   Carcinoembryonic antigen
CEACAM5   Carcinoembryonic antigen-related cell adhesion molecule 5
cfDNA   Cell-free DNA
CgA   Chromogranin A
CgB   Chromogranin B
CHAID   Chi-squared automatic interaction detection
CI   Confidence interval
COPD   Chronic obstructive pulmonary disease
CP   Chronic pancreatitis
CREB   cAMP response element-binding protein
csv   Comma separated value
CT   Computed tomography
CTC   Circulating tumour cells
CV   Cross validation
CYFRA-21   Cytokeratin 19 fragment
DAXX   Death domain associated protein
DCG   Dense core granules
DCIS   Ductal carcinoma in situ
DNA   Deoxyribonucleic acid
ECM   Extracellular matrix
ENETS   European Neuroendocrine Tumour Society
ERK   Extracellular regulated kinase
EUROPAC   European Registry of Hereditary Pancreatitis and Familial
             Pancreatic Cancer

F-actin	Fibrous actin
FBG	Fasting blood glucose
FCV	Fold cross validation
FDA	Food and Drug Administration
FDG	Fluorodeoxyglucose
FIT	Faecal immunochemical test
FN	False negative
FP	False positive
F-pNET	Functional pNET
FPR	False positive rate
GAL-1	Galectin-1
G-actin	Globular actin
GEP-NET	Gastro-entero-pancreatic neuroendocrine tumour
GI	Gastrointestinal
GP	General practitioner
GPNMB	Transmembrane glycoprotein NMB
GRFoma	Growth hormone releasing factor producing pancreatic neuroendocrine neoplasm
hBA1c	Haemoglobin A1c
HDL-C	High density lipoprotein-C
H&E	Haematoxylin and eosin
HE-4	Human epididymis protein 4
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HSL	Health services laboratories
IBD	Inflammatory bowel disease
ID3	Iterative dichotomiser 3
ILK	Integrin-linked kinase
iNOS	Inducible nitric oxide synthase
IRMA	Immunoradiometric assay
ITGB5	Integrin beta-5
k-FCV	k-fold cross validation
KLK8	Kallikrein-8
KS	Kolmogorov-Smirnov
LCN2	Lipocalin-2
LDA	Linear discriminant analysis
LDH	Lactate dehydrogenase
lncRNA	Long non-coding RNA
LOOCV	Leave-one-out cross validation
LR	Logistic regression
LRG1	Leucine rich alpha-2-glycoprotein 1
Lu	Lutetium
LYN	Tyrosine-protein kinase Lyn
MAC2BP	Mac-2-binding protein
MAD	Mothers against decapentaplegic
MAPK	Mitogen-activated protein kinase
MCP1	Monocyte chemoattractant protein-1
MCLA	Matlab classification learner app
MDC	Multi-Disciplinary Diagnostic Centre
MEK	Mitogen-activated protein kinase kinase
MEN1	Multiple endocrine neoplasia 1
MEN4	Multiple endocrine neoplasia 4

MetAP2	Methionine aminopeptidase 2
MiNEN	Mixed neuroendocrine-non neuroendocrine
miRNA	MicroRNA
ML	Machine learning
MMP	Matrix metalloproteinase
MRI	Magnetic Resonance Imaging
mTOR	Mammalian Target of Rapamycin
MWA	Microwave Ablation
NAFLD	Non-alcoholic Fatty Liver Disease
NASH	Non-alcoholic steatohepatitis
NCAM	Neural Cell Adhesion Molecule
NEC	Neuroendocrine carcinoma
NEN	Neuroendocrine neoplasm
NET	Neuroendocrine tumour
NF1	Neurofibromatosis Type 1
NFKB	Nuclear Factor Kappa B
NF-pNET	Non-functional pNET
NHS	National Health Service
NICE	National Institute For Health and Care Excellence
NO	Nitric oxide
NSE	Neuron specific enolase
OCP	Oral contraceptive pill
OR	Odds ratio
OS	Overall survival
PCR	Polymerase chain reaction
PDAC	Pancreatic ductal adenocarcinoma
PDGF-AA	Platelet-derived growth factor-AA
PET	Positron emission tomography
PFS	Progression free survival
PI3K	Phosphoinositide 3-kinase
PINCH	Cystine-histidine-rich protein
pNET	Pancreatic neuroendocrine tumour
proSAAS	Proprotein convertase subtilisin/kexin type 1 inhibitor
proGRP	Progastrin-releasing peptide
PRRT	Peptide receptor radionuclide therapy
PTHrPoma	Parathyroid hormone related peptide hormone-producing pancreatic neuroendocrine neoplasm
qPCR	quantitative polymerase chain reaction
QUEST	Quick, unbiased, efficient statistical tree
RB1	RB transcriptional co-repressor 1
RBF	Radial basis function
RF	Random forest
RFA	Radiofrequency ablation
RIA	Radioimmunoassay
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
SI-NET	Small intestinal NET
SIRT	Selective interval radiation therapy
SLE	Systemic lupus erythematosus
SPARC	Secreted protein acidic and cysteine rich
SPSS	Statistical package for the social sciences
SRS	Somatostatin receptor scintigraphy

SRS-SPECT	SRS single photon emission computerised tomography
SSA	Somatostatin analogue
SSTR	Somatostatin receptor
SV2	Synaptic vesicle 2
SVM	Support vector machine
TACE	Trans-arterial chemoembolization
TAE	Trans-arterial embolization
TB4	Thymosin beta 4
TB4SO	Thymosin beta 4 sulphoxide
TIE2	Tyrosine-protein kinase receptor TIE2
TIMP1	Tissue inhibitor metalloproteinase
TN	True negative
TP	True positive
TP53	Tumour protein p53
TPA	Tissue plasminogen activator
TRACE	Time-resolved amplified cryptate emission
TRAIL	TNF-related apoptosis-inducing ligand
TTR	Transthyretin
UCH-L1	Ubiquitin C-terminal hydrolase L1
UOL	University of Liverpool
US	Ultrasound
USA	United States of America
uTIMP1	Urinary TIMP1
UK	United Kingdom
VCAM1	Vascular cell adhesion protein 1
VEGF	Vascular endothelial growth factor
VGF-NGF	VGF-nerve growth factor inducible peptide
VHL	Von-Hippel-Lindau
VIPoma	Vasoactive intestinal peptide hormone-producing pancreatic neuroendocrine neoplasm
VTE	Venous thromboembolism
WHDA	Watery diarrhoea, hypokalaemia, achlorhydria
WHO	World Health Organisation
XGBoost	eXtreme gradience boosting
5-FU	5-Flurouracil

List of Tables

List of Figures

List of Appendices

Publications
(Poster presentation for UKINETS 2019, Nov 2019 Endocrine Abstracts)
The role of Thymosin Beta 4 as an autocrine mitogen in neuroendocrine
tumours

Table of Contents

Chapter 1 – Introduction

1.1 Neuroendocrine Tumours

Neuroendocrine tumours (NETs) are a wide spectrum of malignancies which arise from neuroendocrine cells throughout the body[1]. The discovery of NETs within the gastrointestinal tract and pancreas began in 1870 when Rudolf Heidenhain discovered neuroendocrine cells[2]. However, Siegfried Oberndorfer was the first to describe these tumours and coined the term carcinoid ("karzinoide") in 1907[3]. The term "neuroendocrine" is applied to dispersed cells with "neuro" and "endocrine" properties. The "neuro" property is based on the identification of dense core granules (DCGs), similar to the DCGs which are found in serotonergic neurons that store biologically-active monoamines[4]. The "endocrine" property of these cells refers to the synthesis and secretion of these monoamines[5]. The neuroendocrine system itself is extensive and includes endocrine glands, such as the parathyroid, pituitary and the adrenal medulla, and endocrine islet tissue embedded within glandular tissues, such as the thyroid and pancreas, as well as scattered cells in the exocrine parenchyma, including the cells of the digestive and respiratory tracts[4]. Consequently, NETs are heterogenous in nature. The wider group of NETs include various types of gastro-entero-pancreatic NETs (GEP-NETs) which include various types of pancreatic neuroendocrine tumours (pNETs).

1.1.1 Epidemiology of Neuroendocrine Tumours
A recent 2019 UK NET registry analysis included NET patients diagnosed between 2013 and 2015[6]. Among the 15,222 cases, 7724 (50.7%) were GEP-NENs. The patient cohort included both NETs and NECs. The age adjusted incidence rate of all GEP-NENs was 4.8 cases per 100,000 however this included NEC patients thus the true incidence rate of GEP-NETs was not specified[7]. 17.7% of all cases were pancreatic[6]. Findings from the Surveillance Epidemiology End Result 18 in the for pNETs diagnosed between 2000 and 2016 showed that annual incidence rates increased of pNETs from 0.27 to 1.00 per 100,000[8]. This was largely explained by an increase in the number of patients diagnosed with localised disease in more recent years (2012-2016).

1.1.2 Functional and non-functional pNETs

pNETs consist of both functional and non-functional types. Functional pNETs (F-pNETs) are associated with specific functional syndromes including gastrinomas (Zollinger-Ellison syndrome), insulinomas, glucagonomas, VIPomas (Verner-Morrison syndrome, pancreatic cholera, watery diarrhoea, hypokalemia, achlorhydia (WDHA) syndrome), GRFomas (growth hormone releasing factor secreting), adrenocorticotropin (ACTHomas), somatostatinomas, pNETs causing carcinoid syndrome, and finally pNETs causing hypercalcaemia parathyroid hormone-related peptide (PTHrPomas). Table 1.1 illustrates the different types of pNETs, associated secretory products and clinical features. Non-functional pNETs (NF-pNETs) are not associated with a functional syndrome, however similar to other pNETs 60-100% secrete various peptides including chromogranin A (CgA), neuron-specific Enolase (NSE) pancreatic polypeptide (50-70%), ghrelin and neurotensin[9]. However, NF-pNETs tend to be asymptomatic and incidentally diagnosed. Sporadic, small NF-pNETs are associated with aggressive behaviour and poorer prognosis, including extra-pancreatic extension, lymph nodal metastasis, distant metastasis, recurrence, and mortality[10].

1.1.3 Sporadic pNETs

pNETs can arise either sporadically or from familial syndromes. The genetic background of sporadic pNETs and the mutational landscape of sporadic pNETs with and without liver metastasis was explored recently[11]. The study involved the collection of tumour tissue from 14 patients who had sporadic pNETs[11]. Next generation sequencing of 612 cancer-associated genes carried out on tumour tissue samples[11] identified 63 somatic mutations in 53 genes. Amongst these mutations, multiple endocrine neoplasia 1 (MEN1) was identified as the most recurrently mutated gene[11]. Other novel genes which were recurrently mutated included adrenoceptor alpha 2B, ARVGF delta catenin family member, carbamoyl-phosphate synthase 2, aspartate transcarbamylase, dihydroorotase and neuregulin 1[11]. Out of the 53 genes with mutations, 11 of these were enriched in the PI3K/AKT signalling pathway[11]. Additionally, the analysis found that sporadic pNETs with liver metastasis had distinctly different mutational profiles compared to pNETs without liver metastasis[11]. 13 genes were mutated in pNETs with liver metastasis alone[11]. These genes included ATRX chromatin remodeller, thioredoxin reductase 2, janus kinase 3, armadillo

repeat protein deleted in velo-cardio-facial syndrome (ARVCF) delta catenin family member, integrin subunit alpha V, RAD50 double strand break repair protein, protein kinase C delta, tubulin alpha 3c, DNA methyltransferase 3 alpha, integrin subunit alpha L, partner and localiser of BRCA2 and pyruvate dehydrogenase kinase 2[11]. Kyoto Encyclopaedia of Genes and genomes signalling pathway enrichment analysis found that the 13 genes were involved in homologous recombination, chemokine signalling pathways and type II diabetes mellitus[11]. The relationship between pNETs and type II diabetes was examined in a study by Fan et al., with this study finding that pNET patients with type II diabetes had an increased risk for tumour metastasis (OR of 2.81 P=0.001). The Fan et al., study had cohort of 299 pNET patients. However, in this study the two pNET patients with pNET and type II diabetes did not demonstrate liver metastasis[11].

1.1.4 Familial Syndromes Associated with pNETs

The majority of pNETs are sporadic in nature, however they can also be associated with a number of genetic syndromes including MEN1 and Von Hippel-Lindau (VHL) disease[12]. pNETs are thought to occur in 30-80% of MEN1 patients, >15% of VHL patients, <10% of neurofibromatosis type 1 (NF1) patients, and <1% of tuberous sclerosis patients[13,14]. While MEN1 remains the most common hereditary syndrome associated with pNETs, over the past decade the spectrum of inherited pNETs has been expanded via the inclusion of multiple endocrine neoplasia 4 (MEN4) and glucagon cell adenomatosis (Mahvash disease)[15].

MEN1 is characterised by the combined occurrence of two or more tumours which usually involve the pancreatic islets and the parathyroid, anterior pituitary and adrenal glands[16]. It is an autosomal dominant disorder and patients carry germline mutations of the MEN1 tumour suppressor gene located on chromosome 11q13 and consequent loss of a 610 amino acid nuclear protein known as menin, which has a role in suppressing proliferation[17,18]. The majority of MEN1 patients develop tumours by the age of 45[19]. VHL-driven pNETs occur in young patients and within the pancreas. They are typically non-functional and multifocal[15]. The NF-1 gene functions as a tumour suppressor and the NF-1 protein plays a role in MAPK signalling as it contains a 360-residue GAP domain that is involved in the MAP-kinase pathway[15]. From a morphological

point of view, pNETs seen in NF-1 patients are indistinguishable from their sporadic counterparts, hence their diagnosis is difficult in the absence of genetic or more specific clinical characteristics[15].

Type of F-pNET	Secretory Product	Clinical features
Gastrinoma	Gastrin	Zollinger Ellison syndrome Mostly malignant NET of the duodenum and pancreas
Insulinoma	Insulin	Usually benign Associated with hypoglycaemia
Glucagonoma	Glucagon	Rare pancreatic tumours
VIPoma	Vasoactive intestinal peptide	VIPoma syndrome is also known as WHDA syndrome

Table 1.1: Summary of the different types of F-pNET, including the secretory product and clinical features.

1.1.5 pNET Grading and Staging

1.1.5.1 pNET Grading

World Health Organisation (WHO) classifications have evolved significantly in the last two decades in particular for digestive neuroendocrine neoplasms (NENs)[20]. The latest WHO classification published in 2022[21] groups all NENs regardless of their primary location and is a progress for clinicians. In this classification NENs are separated into two main groups, the well differentiated neoplasms (called neuroendocrine tumours, NETs) and the poorly differentiated neoplasms (called neuroendocrine carcinomas, NECs). NECs are further divided into small cell -NEC and large cell NEC. The classification was an important step towards a common classification blueprint for all NENs in the body that share many characteristics[20]. Proliferations and grading refer to the biological aggressivity of NEN and proliferation is a key point of the digestive WHO classification and NETs are graded into G1, G2 and G3 grades[20].

In addition to their central role in grade classification[22], Ki-67 index and mitotic rate influence patient prognosis[23] and biomarker levels[24]. A study in which 202 patients with pancreatic NECs were enrolled, including 172 well differentiated and 30 poorly differentiated carcinomas, showed that the major risk factor for progression is Ki-67 index[23]. This therefore is an important measure to help

guide clinicians to better manage patients and plan appropriate follow up programmes and therapeutic strategies. 5-year survival rates for G1 NETs, G2 NETs and NECs were 91%, 69% and 10%, respectively[24]. Moreover, patients who had G1 or G2 NETs and who were treated with surgical resection had a good prognosis, while NEC patients exhibited distant metastasis and had a poor prognosis[24]. Therefore NET grading can provide not only prognostic information but also information on the most suitable treatment strategy for patients[24]. The relationship between the grade of pNET and biomarker levels has revealed that pNET patients with different tumour grades have different levels of biomarkers[24]. Elevated tumour markers (carbohydrate antigen 19-9 (CA19-9)), carcinoembryonic antigen (CEA), NSE, progastrin-releasing peptide (ProGRP) and lactate dehydrogenase (LDH) levels increased with pathological grade of pNET[24]. Serum levels of NSE and LDH were elevated in pNECs compared to G1 and G2 pNETs[24]. They also found a positive correlation between the Ki-67 index and NSE levels (R=0.3030, p=0.021) and LDH levels (r=0.578, P<0.001)[24].

1.1.5.2 pNET Staging

The two most widely used staging systems for pNETs have been produced by the European Neuroendocrine Tumour Society (ENETs) and the American Joint Committee on Cancer (AJCC)[25]. Both staging systems use a TNM approach. The T refers to the size of the primary tumour, N to metastasis to regional lymph nodes, and M to distant metastases[26]. In the AJCC 7th staging system[26] T1 tumours are limited to the pancreas and are ≤2 cm in greatest dimension[25]. T2 tumours are limited to the pancreas and >2 cm in greatest dimension, T3 tumours are spread beyond the pancreas but with no involvement of the superior mesenteric artery[25]. T4 tumours have involvement of the celiac axis or superior mesenteric artery and are unresectable[25]. In the AJCC 8th and ENETs staging systems, T1 refers to a ≤2 cm tumour limited to the pancreas, T2 refers to a 2-4 cm tumour limited to the pancreas, T3 refers to a tumour limited to the pancreas and >4cm in size or one that has invaded the duodenum or common bile duct, and T4 refers to a tumour that has invaded adjacent structures[25]. N0 refers to no regional lymph node metastasis, while N1 refers to regional node metastasis. This classification of nodal involvement is the same in the AJCC 7th stage, AJCC 8th stage and ENETs systems[25]. In the AJCC 7th staging system,

the M0 category refers to no distant metastasis and M1 refers to distant metastasis[25]. The AJCC 8th and the ENETs staging have a detailed M1 classification with M1 being further subdivided into M1a referring to metastasis confined to liver, M1b referring to at least one extrahepatic stie and M1c referring to both hepatic and extrahepatic metastases[25].

1.1.6 Therapeutics and Treatments for pNETs

Understanding the different treatment options and therapeutics available for pNET patients is important to illustrate the clinical complexity of this group of patients. The following Sections 1.1.6.1 to 1.1.6.9 briefly explain the different treatment options available to patients with non-metastatic and metastatic pNETs.

1.1.6.1 Active Surveillance for non-metastatic pNETs

Active surveillance is an option for patients whereby treatment is not initiated but instead a watch and wait approach is taken. The Asymptomatic Sporadic non-functioning Pancreatic Neuroendocrine Neoplasms (ASPEN) study aims to prospectively explore whether active surveillance as an approach for asymptomatic NF-PNENs of <2cm in size is safe when compared to surgery[27]. However an important issue with active surveillance is the absence of prognostic markers that can predict tumour progression during follow-up[28].

1.1.6.2 Surgery for non-metastatic pNETs

Surgical resection offers the only potentially curative treatment for patients with F-pNETs and NF-pNETs. Patients are selected for surgical resections based on tumour functionality, grade, stage, and association with MEN1. Gastrinomas and non-MEN1 genotypes are positive indicators for surgery, as >98% survival was obtained for these groups after surgical resection[29].

1.1.6.3 RFA for non-metastatic pNETs

Interventional radiology also encompasses techniques such as radiofrequency ablation (RFA), which provides a relatively new treatment option. A recent RFA study included a cohort of 18 adults, comprising seven insulinoma patients and 11 patients with NF-pNETs, who underwent RFA between March 2017 and October 2018. Technical success was achieved in 26 out of 27 lesions and

there were no major complications 48 hours post procedure. There were also no clinically-significant recurrences during a mean follow-up period of 8.7±4.6 months[30].

1.1.6.4 Surgery if feasible for metastatic pNETs

Surgery is the first course of treatment for liver metastasis[31]. The risk of recurrence is significant and full excision of metastasis is rarely achievable[32]. It may be performed in the palliative setting for symptomatic control or when hepatic involvement is smaller than 50% and at least 90% of the tumour can be resected[33,34].

1.1.6.5 Somatostatin analogues for metastatic pNETs

Somatostatin analogues (SSAs) act on the somatostatin receptor and work by controlling hormonal secretion and tumour growth as functional and NF-pNETs express at least one out of the five subtypes of somatostatin receptor[35]. SSAs exert antiproliferative activity against pNETs with ki-67 > 10% and can be used instead of other more toxic treatments[36]. There is an association between lower tumour burden (25%) and potential for benefit in this population. Moreover, patients with hepatic tumour burden <25% showed longer PFS compared with those with high liver tumour load[36]. In this study a substantial proportion of patients displayed a liver tumour load greater than 25% which reflects the current 'real life' practice among NET referral centres[36].

1.1.6.6 PRRT for metastatic pNETs

Another therapy for pNETs is Peptide Receptor Radionuclide Therapy (PRRT). The presence of somatostatin receptors (SSTRs) on GI-NETs and pNETs has been exploited in order to target radiotherapy by using radiolabelled SSAs (indium-111, yttrium-90 (Y90) or lutetium-177 (Lu-177))[37]. After SSTR binding and internalisation of the receptor complex, the ionising radiation is released which causes damage to tumour DNA and cell death[18]. Treatment of metastatic pNETs with 177Lu-DOTATATE was shown to be a safe and effective therapy resulting in radiological symptomatic and biochemical response in a high percentage of patients with metastatic functioning pNETs[38].

1.1.6.7 Molecular targeted therapies for metastatic pNETs

Molecular therapies for NETs include mammalian target of rapamycin (mTOR) inhibitors and anti-angiogenic compounds that target vascular endothelium growth factor (VEGF). Other therapies include anti-angiogenic compounds that block the actions of VEGF in promoting the survival and growth of blood vessels[18]. Sunitinib maleate is a tyrosine kinase inhibitor that can irreversibly inhibit several kinases including the VEGFR family. Sunitinib has been approved for the treatment of locally advanced or metastatic pNETs[39]. Continuous daily administration of Sunitinib at a dose of 37.5mg improved progression free survival, overall survival and the objective response rate as compared with placebo among patients with advanced pNETs[39]. Treatment of patients with progressive metastatic well-differentiated pancreatic tumours with the biologic Bevacizumab, which is an anti-VEGF monoclonal antibody, in combination with 5-FU/Streptozocin was carried out through the multicentre, non-randomised BETTER Phase II trial and this achieved a progression-free survival (PFS) of 23.7 months which was much longer than the initial 9 months that was assumed in the protocol and also better than the reported PFS with either chemotherapy alone or new targeted therapies in the same population[40]. The mTOR inhibitor Everolimus has been shown to have efficacy in Phase II and Phase III studies in patients with pNETs[41] in the RADIANT-3 study, where patients treated with Everolimus had a significantly longer PFS (11 versus 4.6 months) than patients receiving placebo[42].

1.1.6.8 Chemotherapy for metastatic pNETs

When pNETs progress under somatostatin analogue (SSA) therapy, or exhibit symptoms or aggressive features and/or Ki67 values >10%, cytotoxic chemotherapy should be considered as the first-line treatment of choice[43]. The two classes of drugs which are used to treat well-differentiated pNETs are alkylating agents (e.g. streptozocin, temzolomide, and dacarbazine) and antimetabolites (e.g. 5-Flurouracil (5-FU) and capecitabine (CAP))[44]. In western countries streptozocin has been used as a first-line anti-cancer drug in patients with unresectable NETs[45].

1.1.6.9 Locoregional treatments of hepatic metastases

Intra-arterial therapies for NETs are based on the knowledge that most liver metastases from NETs are hypervascular and take their blood supply from the hepatic artery, while the normal liver blood supply is from the portal vein[46]. Trans-arterial embolization (TAE) causes ischemia and necrosis of the lesions by injecting various particles (gelfoam, polyvinyl alcohol, microspheres)[44].Trans-arterial chemoembolization (TACE) was developed in 1990s based on the rationale of embolising the blood vessels after delivering chemotherapy directly to the tumour cells by systemic injection[44]. Both the high drug concentration and ischaemia of the cells can enhance their response to the treatment. The most commonly used agents are doxorubicin and streptozotocin alone or combined with other agents[44].

Selective interval radiation therapy (SIRT) is radioemobilisation using resin-based (Sirspheres) and glass-bead (Therapspheres) micron sized particles, loaded with Yttrium-90 radioisotype, is increasingly being used, delivering high irradiation directly to the tumour[44]. One series of 148 patients has shown positive response in 62.9% of patients, and stable disease in 22.7%[47]. Thermal ablation is performed by using radiofrequency ablation (RFA) or microwave ablation (MWA) which delivers high frequency current to the lesion inducing heat which destroys the proteins leading to cellular death[44]. RFA is more frequently used than microwave ablation, showing overall good clinical response but as with surgery intra-hepatic recurrence remains a problem[44]. The classical indications for thermal ablation in liver metastases are less than 5 lesions, and less than 5cm in size, however in NETs a more extensive metastatic spared is frequent and the technique is sometimes used beyond these indications[44].

1.2 pNET Detection and Diagnosis

The earlier pNETs are detected and diagnosed the better the patients' outlook. Methods for diagnosing pNETs including histology and imaging which are discussed in subsections 1.21 and 1.22 respectively. The clinical pathways and routes for patient diagnosis are reviewed in 1.23. pNET symptoms, comorbidities and risk factors are considered in 1.24 and 1.25. Blood based pNET detection is discussed in a separate section (1.3).

1.2.1 Histology

NETs present as solid or small trabecular clusters or dispersed among other cells which then make them difficult to identify in sections that are stained with haematoxylin and eosin (H&E). However, immunostaining enables exact identification of NETs. NETs at the ultrastructural level are defined by their cytoplasmic membrane-bound dense-cored secretory granules (diameter >80nm) and small clear vesicles (40-80nm) which are related to the synaptic vesicles of neurons[48].

pNETs are characterised as well-demarcated, usually solitary, round tumours with a diameter of 1-4 cm. These tumours can occur in all parts of the pancreas[48]. If associated with a hormonal syndrome the hormone causing the syndrome can be detected immunocytochemically. Histologically, in approximately 5% of insulinomas amyloid deposits are found in the tumour. The colour of the cut surface varies but is mostly white to tan, depending on the degree of fibrosis and amyloid deposition. Fibrous capsules are not a characteristic finding although they can be present. The tumours are mostly located in the pancreatic parenchyma. The following histopathological features of pNETs are characteristic in the majority of cases[49].

- Tumour cells are arranged in solid nests or in ribbon-like trabecular and/or gland like formations.
- Tumour nests are surrounded by thin vascular stroma. Perivascular pseudorossette arrangements are considered highly specific in pNETs.
- The cytoplasm of tumour cells contains neuroendocrine granules with CgA presence demonstrated via immunohistochemistry.
- Nuclear atypia is mild to moderate and round to oval in shape. The characteristic nuclear morphology of the nuclei with coarsely granular chromatin and subtle nucleoli is referred to as the "salt and pepper" pattern.

Immunohistochemical examinations are vital for the diagnosis of pNETs[49]. The neuroendocrine differentiation of the tumour is confirmed via the expression of CgA and synaptophysin[49]. CgA is a large secretory granule associated universal marker of neuroendocrine tissues and tumours[48]. However, its expression level depends on the cell type and the number of secretory granules present in the cells[48]. Synaptophysin is a small vesicle associated marker and

an integral membrane glycoprotein that occurs in the presynaptic vesicles of neurons. Other small vesicle associated markers include synaptic vesicle protein 2 (SV2) and synaptobrevin[48]. Somatostatin receptors (SSTRs) are widely expressed in different NETs and comprise five subtypes termed SSTR1, SSTR2, SSTR3 SSTR4 and SSTR5. These subtypes can be identified via immunohistochemistry and also through autoradiography[50] or otreoscan[51]. Serotonin-producing NETs and gastrinomas are more commonly positive for SSTR2 and SSTR5 compared to insulinomas[52].

Cytosolic markers of NETs include NSE[48]. However, commercially available antibodies have a limited sensitivity due to unspecific staining of certain non-endocrine tissues such as striated muscles[48]. Additionally, NSE has also been recognised in some non-neuroendocrine tissues including pseudopapillary neoplasm of the pancreas and serous cystic neoplasms of the pancreas[48]. Thus, these limitations suggest that NSE alone as a marker is not suitable for histological diagnosis[48]. Cell membrane markers of NETs include the neural cell-adhesion molecule (NCAM, CD56) which belongs to a group of cell surface glycoproteins involved in direct cell to cell adhesion[48]. However, limitations of NCAM include the presence not only in neuroendocrine cells but also expression in non-endocrine normal tissues such as the renal tubules and thyroid follicle and neoplastic tissues including non-small lung carcinomas[48].

1.2.2 Imaging

The detection of a F-pNET can be a diagnostic challenge. While these tumours affect very prominent hormonal systems, the resulting effect may be small[53]. In contrast, NF-pNET symptoms are usually secondary to tumour mass effects, such as abdominal pain, jaundice and weight loss, or to the effects of metastatic disease[53]. Patients with NF-pNETs more often present with larger tumours which are more easily detected compared to functioning pNETs[53]. This is due to the lack of symptoms associated with NF-pNETs compared to F-pNETs, thus the size of the tumour being larger at diagnosis due to a delayed diagnosis. Moreover, imaging of PDACs and pNETs can produce similar results leading to misdiagnosis of pNETs as PDACs[54]. Additionally, pNETs need to be distinguished from other hypervascular lesions such as primary exocrine tumours (e.g. serous cystic adenoma, solid pseudopapillary neoplasia, and acinar cell carcinoma), hypervascular metastases (e.g. renal cell carcinoma,

carcinoid, and medullar thyroid carcinoma), neurogenic tumours (e.g.
Schwannoma), vascular lesions (e.g. aneurysm, pseudoaneurysm, and
arteriovenous malformation) and developmental lesions (e.g. intrapancreatic
splenule)[55]. Imaging options for pNETs include anatomical (computed
tomography (CT), magnetic resonance imaging (MRI), ultrasound (US)) and
functional (scintigraphy and positron emission tomography (PET))[53]. CT is the
first line imaging modality of choice for suspected pNET patients and allows for
disease identification in the pancreas and assessment of disease extent[51]. On
CT and MRI images, NF-pNETs appear as a large pancreatic mass that has
heterogenous enhancement due to necrotic and haemorrhagic changes.
Moreover, the development of multidetector CT technology has allowed high
resolution images to be obtained during multiple phases of enhancement[56].
However, meticulous technique, awareness of atypical appearances such as
isoattenuating, hypoattenuating and cystic tumours, and careful scrutiny of
tumours near vessels are critical to obtaining the highest possible sensitivities[56].
On CT images typical insulinomas are well defined, hypervascular, and show
intense enhancement during the arterial/pancreatic phase[51]. Moreover,
amongst all pNETs, gastrinomas are most often associated with MEN1
mutation[51]. After contrast injection, gastrinomas have a delayed enhancement
persistent on the delayed phase of CT imaging due to the presence of fibrosis[51].
The superior soft tissue contrast resolution of MRI, which is the ability to
distinguish various structures based on their different signal intensity
characteristics, makes it advantageous for the investigation of pNETs compared
to CT[53]. However, limitations to MRI include its lower availability and prolonged
image acquisition time compared to CT[53]. CT sensitivity for the diagnosis of
GEP-NETs is 57-94% with sensitivities for MRI approaching 94%. Thus, the
sensitivities are similar for both imaging modalities. With both CT and MRI,
pNET detection increases with tumour size[57].

Another method for imaging pNETs is ultrasound (US). The various US
approaches include conventional transabdominal US, endoscopic US (EUS),
(US by inserting a transducer via the endoscope) and intra-operative US (direct
contact of the transducer with the organ surface during surgery)[53]. Similar to
MRI, US is advantageous as it does not expose the patient to potentially
ionising radiation[53]. Moreover, US provides guidance of the biopsy needle for
both fine needle aspiration for cytology and core biopsy for histopathological

examination[53]. EUS has a sensitivity of 87% in the diagnosis of GEP-NETs[57], but it is more suited for the detection of small (< 2cm) pNETs, such as gastrinomas and insulinomas[57]. Moreover, the sensitivity for EUS is higher for head than tail lesions due to differences in the proximity of the endoscope[51]. The sensitivity of EUS is higher for pancreatic gastrinomas compared to extra-pancreatic gastrinomas, which is thought to be due to their generally smaller size[51]. EUS is also thought to be helpful for the detection of adjacent lymph node metastasis within the gastrinoma triangle[51]. The gastrinoma triangle itself is defined as the confluence of the cystic and common bile duct superiorly, the 2nd and 3rd parts of the duodenum inferiorly and the neck and body of the pancreas medially, both dorsally and ventrally[58]. Moreover early screening using EUS can identify pNETs in asymptomatic patients with MEN1[59]. Thus identifying these tumours before the development of symptoms and metastases could facilitate prompt surgical intervention and improve prognosis[59]. EUS can also be used to guide RFA to the tumour and thus be an alternative to surgery[30].

Isotope-imaging modalities have become increasingly relevant for the management of pNET patients[51]. Functional imaging with SSAs is useful to detect the primary tumour, assess disease extent and recurrence, and to finally select patient candidates for PRRT with Y90 or Lu177-labelled SSAs[51]. Poorly differentiated NECs have lower expression levels of somatostatin receptors and functional imaging with SSAs is limited for these tomours[51]. SRS (somatostatin receptor scintigraphy) may be able to detect the pancreatic primary tumour when morphological imaging and EUS show no lesions[51], because SRS is more sensitive in detecting well-differentiated gastrinoma, glucagonoma, VIPoma and non-functioning pNETs[51]. However, it has a low sensitivity for detecting insulinomas due to their lower SSTR2 subtype expression[51]. FDG (fluorodeoxyglucose) PET (positron emission tomography)/CT is widely used in oncology but FDG is not considered a good tracer for NET tumours. Moreover, gallium-68 SSA PET has been demonstrated to be more sensitive than SRS-SPECT (SRS single photon emission computerised tomography) and is expected to become the preferred mode for functional imaging of pNETs[51]. Imaging predictors of aggressive behaviour in pNETs include large size, low vascularisation, vascular encasement, ill-defined margins, pancreatic and/or bile duct dilatation, complex cystic morphology, liver involvement, a large number of

metastatic sites and rapid spontaneous tumour progression[51]. Liver involvement and a higher number of metastatic sites represent an advanced disease state.

1.2.3 Clinical pathways to diagnosis

Patients with NETs face delays in the time taken to be diagnosed. A web-based survey found that the median time from first symptoms to diagnosis is 53.8 months[60]. Moreover 58% of respondents were referred to secondary care where they were seen a median of 3 times[60]. Furthermore, 30% of patients presented acutely to Accident and emergency (A&E) services[60]. There is therefore a clear clinical need to improve diagnostic pathways for NET patients. When considering patients with pNETs, most occur sporadically and a minority are associated with familial conditions such as MEN1. Patients who have been diagnosed as MEN1 or are at an increased risk of MEN1 affects due to first degree relative diagnosis should be offered a programme of combined clinical, biochemical and radiological screening[61]. As a minimum, screening for GEP-NETs in this group should include an annual plasma biochemical evaluation comprising measurements of gastrin, glucagon, VIP, pancreatic polypeptide, CgA and insulin with associated fasting glucose level[61]. A consensus for optimum radiological screening in these patients has not been established and this depends instead on clinical judgement, resources and patient preferences. However, a suggested minimum imaging protocol includes annual pancreatic and duodenal visualisation with MRI, CT or EUS[61]. As most pNETs are sporadic identification of patients is typically through presentation with symptoms in general practice. The National Institute For Health and Care Excellence (NICE) guidelines pathway (Dec 2021) for pancreatic cancer indicates patients suspected of pancreatic cancer should be referred to secondary care via a suspected cancer pathway, which should provide an appointment within 2 weeks for suspected pancreatic cancer patients if they are aged 40 years and over and have jaundice[62]. The NICE guidelines also suggest an urgent direct access CT scan should be performed within 2 weeks, or an urgent ultrasound scan if CT is not available, to assess people aged 60 years and over with weight loss and any of the following symptoms: diarrhoea, back pain, abdominal pain, nausea, vomiting, constipation, or new-onset diabetes[62].

There are also patients who present with non-specific symptoms, who consequently don't qualify for referral via the two-week wait pathway. A non-

specific symptom can be caused by a range of conditions in addition to cancer. Thus, identifying appropriate diagnostic tests and referral routes for these patients can be challenging. This leads to patients often seeing their GP multiple times before referral, which contributes to longer intervals from presentation to diagnosis. Improved clinical pathways are needed to diagnose these patients. The MDC (multi-disciplinary diagnostic centre) provides another route whereby patients with non-specific but concerning symptoms which could be indicative of cancer can be seen. The MDC aims to improve outcomes for such patients by providing rapid access to a variety of diagnostic tests within a single diagnostic centre, with the aim of speeding up diagnosis[63]. The MDC concept was first trialled in Denmark as part of their three-legged cancer stategy[64] and subsequently in the United Kingdom (UK) (5 projects encompassing 10 centres in total)[63]. The UK criteria for referral to MDCs followed two key points. Firstly, that the patient must be considered as being of clinical concern with non-specific symptoms potentially indicative of cancer (or other serious disease) and that their presenting symptoms are not sufficiently clear to indicate an appropriate tumour-specific urgent referral pathway. Patients with a previous cancer were included and painless jaundice despite being a site-specific symptom was included as a referral criterion in the London MDC. Table 1.2 summarises the approaches taken at the different UK MDCs. Eligible referral criteria were included on MDC referral templates to assist with data consistency, data quality and to clearly indicate patient symptoms as appropriate. However, there was still some missing data associated with the evaluation arrangements and despite measures implemented to promote consistency with data collection and reporting, a certain degree of variation was present due to localised data arrangements[63].

MDC	Referral Criteria	Referrer
London (5 centres)	Abdominal symptoms with no clear referral pathway, new unexplained abdominal pain, unexplained weight loss, persistent nausea/appetite loss, GP clinical suspicion and painless jaundice	GP
Oxford (1 centre)	Severe unexplained fatigue, unexplained weight loss, persistent nausea or appetite loss, now atypical pain, unexplained laboratory findings, no organ specific symptoms, no symptoms fulfilling referral via the standard 2 week wait	GP

	pathway, over 40 years old and GP clinical suspicion ('gut feeling')	
Airedale (1 centre)	Persistent unexplained weight loss, persistent unexplained abdominal pain, non-specific but concerning symptoms (high risk of cancer), GP suspicion. Too unwell for 2-week wait referral	GP A&E Secondary care clinic
Leeds (1 centre)	Appetite loss and nausea (unexplained, 40 years and over), weight loss (unexplained, 40 years and over), abdominal pain without rectal bleeding or weight loss (<3-month duration or recent change in the character/severity, 50 years and over), anaemia (non-iron deficient, without bleeding, 50 years and over), hypercalcaemia (unexplained and persisting <12 months), thrombocythemia (unexplained and persisting <12 months), GP suspicion and general condition ('poor' general condition)	GP Acute Medicine
Greater Manchester (2 centres)	Non-specific abdominal pain, unexplained weight loss, severe unexplained fatigue, nausea/appetite loss, splenomegaly, hepatomegaly, lymphadenopathy, GP clinical suspicion and non-iron deficiency anaemia	GP

Table 1.2: Referral criteria and referral route for the different MDC pilot centres in England[63].

The MDC pilot found that 8% of referrals resulted in a cancer diagnosis with 241 cancers diagnosed form a referral cohort of 2961. Moreover, a relatively large proportions of the diagnosed patients had stage IV cancers compared to Stage 1 to 3. Several of the non-specific symptoms were systemic, consistent with metastatic cancer[63]. Despite this a substantial portion of early-stage cancers of the lung, upper gastrointestinal (GI) tract and haematological cancers were diagnosed, which indicated that MDCs provide a route for timely diagnosis for these cancers. However, it does remain a possibility that some of the cancers, including those diagnosed at an early stage, were diagnosed 'incidentally' via symptoms not caused by cancer. Also the number of NETs diagnosed via this route was not mentioned[63]. The MDC pilots found a strong association between cancer diagnosis and patient age and identified GP 'clinical suspicion' as a strong predictor of cancer amongst the non-specific symptoms. Hence the pilot demonstrated that a cancer referral pathway for patients with non-specific but concerning symptoms could be a valuable referral option for patients suspected of cancer across a broad range of anatomical sites[63].

1.2.4 pNET symptoms

There is a delay in the diagnosis of patients with pNETs with the median duration from the time of first symptoms to diagnosis being 24 months[65]. Respondents saw their GPs about their symptoms 5 times over a median 18-month period, with 31% of patients being diagnosed following unplanned emergency admission[65]. Due to the absence of a UK screening programme for the detection of pancreatic cancer, identification of patients suspected of this disease through symptomatic presentation in general practice remains the key avenue for early diagnosis[66]. In a real-life general practice scenario, GPs have the considerable challenge of distinguishing patients with suspected malignancy versus other benign conditions[66]. Often these patients present with non-specific symptoms that can be easily missed by the clinician and thus delay a cancer diagnosis. A recent and first-of-its-kind large population-based study examining the symptomology of PDAC and pNENs was carried out by Liao and co-workers in 2021[66]. This open cohort study included 15,194,279 patients aged 25 years and over. From this group 23,640 PDAC cases and 596 pNEN cases were identified. Most pNENs were neuroendocrine carcinomas (45.97%) and most of the pNETs were well differentiated (38.59%).

This study systematically examined the symptomology of PDACs and pNETs and quantified the association of 42 potential symptoms in different time windows relative to the date of diagnosis. The study identified 2 symptoms for pNETs within 3 months of diagnosis (alarm symptoms) - GI bleeding and jaundice. The study also identified symptoms within 1 year of diagnosis of a pNET and these were diarrhoea, bowel change, vomiting, indigestion, abdominal mass, abdominal pain and weight loss (longer than 2 years). However, the identified symptoms overlapped with those of PDAC as illustrated in Figure 1.1 and no pNET-specific symptoms were identified in this study. In contrast, several PDAC-specific symptoms were identified. Consequently, a patient presenting to their GP with a PDAC-specific symptom might be referred for PDAC as opposed to pNET assessment, while a patient presenting with a pNET symptom might be referred for pNET and PDAC assessment. In both cases other diagnoses could not be excluded based on non-specific symptomatic presentation.

A limitation of this study was the low number of pNET cases as this may have resulted in the full burden of symptoms associated with pNETs not being

statistically significant. Hence it is not clear whether a lower number of significant symptoms for pNETs was due to a lack of statistical power, or if pNETs had a less prominent symptomatology, or if it was a combination of both factors. Additionally, symptomatology was not explored with functionality, which might have revealed different symptoms between NF-pNETs and F-pNETs. The study also did not explore symptoms associated with early and late stage pNETs. There was also information bias in the electronic health records and the authors were unable to evaluate the accuracy of information recording across practices. The implications of this study were that the identified symptoms should be communicated to the public to encourage individuals to seek medical attention if they have any of these symptoms. Public and patient engagement events could assist in raising public awareness of these symptoms as signs of pancreatic cancer which could result in patients seeing their GPs more promptly when they notice these changes.
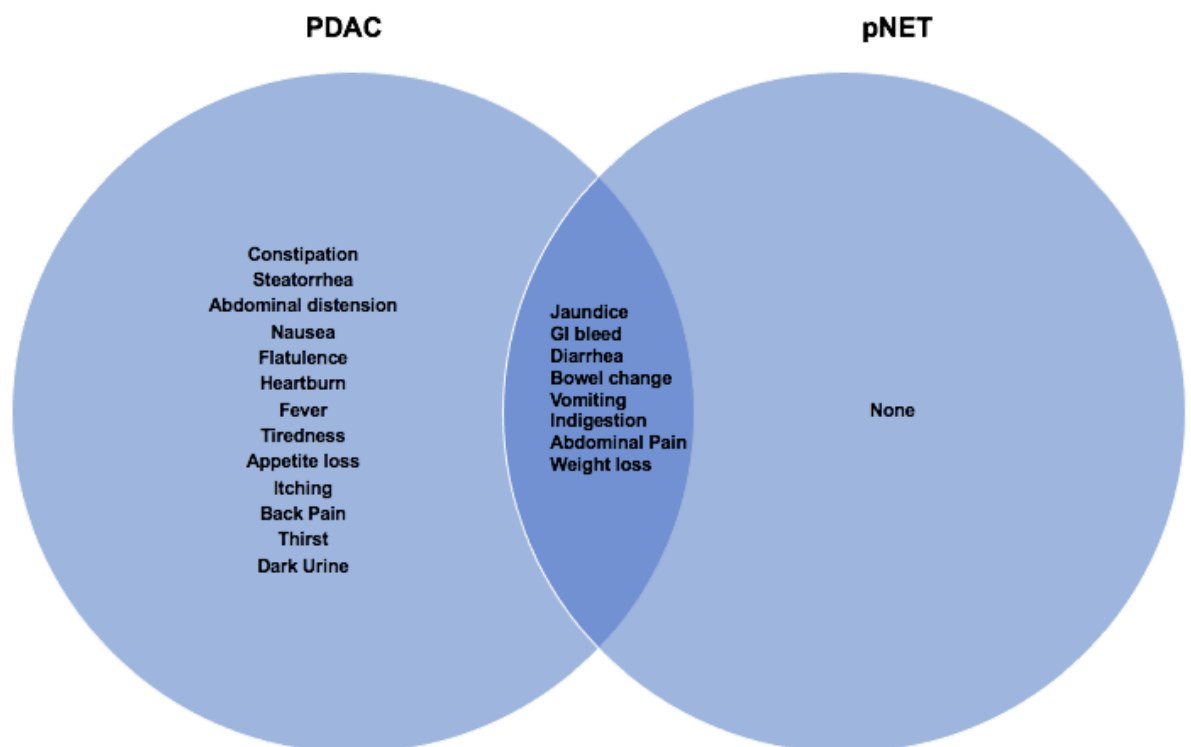
Figure 1.1: Symptomology Venn Diagram for PDAC and pNET.
The Venn diagram illustrates PDAC-specific symptoms, overlapping symptoms for pNETs and PDAC, and the absence of any pNET-specific symptoms as identified by Liao and co-workers.

1.2.5 pNET comorbidities and risk factors

A large scale population level based study identified pNET comorbidities as type 2 diabetes, venous thromboembolism (VTE), Cushing's syndrome and pancreatic cyst[66]. Type 2 diabetes, especially recent-onset of type 2 diabetes, has been shown to be associated with pNET risk in various case control studies [67,68,69,70]. In the Liao et al. (2021) population level study comorbidity analysis identified an overlap between those seen for pNENs and for PDACs[66]. The identified pNET comorbidities of type 2 diabetes, VTE, Cushing's syndrome and pancreatic cyst are all statistically associated with PDAC as well and no pNEN-specific comorbidities were identified. However, additional comorbidities were identified for PDACs but not for pNENs. These included AP, cholangitis, family history of GI cancer and type 1 diabetes. Figure 1.2 illustrates the specific and shared comorbidities of PDACs and pNENs.

The relationship between pNETs and diabetes is of particular interest. Other case control studies have identified diabetes as a risk factor for pNETs[70,67,68,69,71]. This includes the 2008 Hassan et al. study which reported an odds ratio (OR) of 2.80 (confidence interval (CI) 1.50-5.20) for pNETs and diabetes[70]. CIs are at 95% unless specified otherwise. This study compared 740 patients with histologically confirmed, well-differentiated, low/intermediate grade NETs and 924 healthy controls. Within the NET cohort there were 60 patients with pNETs. The study found an association between recent onset diabetes and pNETs in 55% of the patients, but no such association with long-term diabetes. Moreover, the significant relationship between diabetes and pNETs in men and women was observed only for diabetes that had been diagnosed within less than a year of NET diagnosis.

A study by Halfdanarson et al. in 2014 reported an OR of 4.80 (CI 2.30-9.90) for pNET and diabetes[67]. In this study, patients were asked to complete a risk factor questionnaire that included detailed information on family history of malignancies, personal history of malignant and non-malignant disorders and habits including tobacco and alcohol use. Only histologically or clinically confirmed low to intermediate grade pNETs were included in the study. Patients having MEN1 or other inherited syndromes such as VHL were excluded. They found that cases were more likely than controls to report being diagnosed with diabetes prior to being diagnosed with pNETs (19% vs 11% p=0.003). Moreover, 6 patients were diagnosed with diabetes at the same time or after the

pNET diagnosis and 16 were diagnosed with diabetes within three years prior to pNET diagnosis. Additionally, this study excluded patients with insulinomas. However due to a lack of information on gene mutations resulting in MEN1 there is a likelihood that patients with clinically unrecognised MEN1 may have been included. The study also found that pNET patients were more likely than controls to report a family member with sarcoma (p=002), pNET (p=0.024), gall bladder cancer (p=0.024), ovarian cancer (p=0.04) and gastric cancer (p=0.01). They also found that more cases had a body mass index (BMI) of 30 or higher compared to controls (30% vs 24% p=0.013)[67].
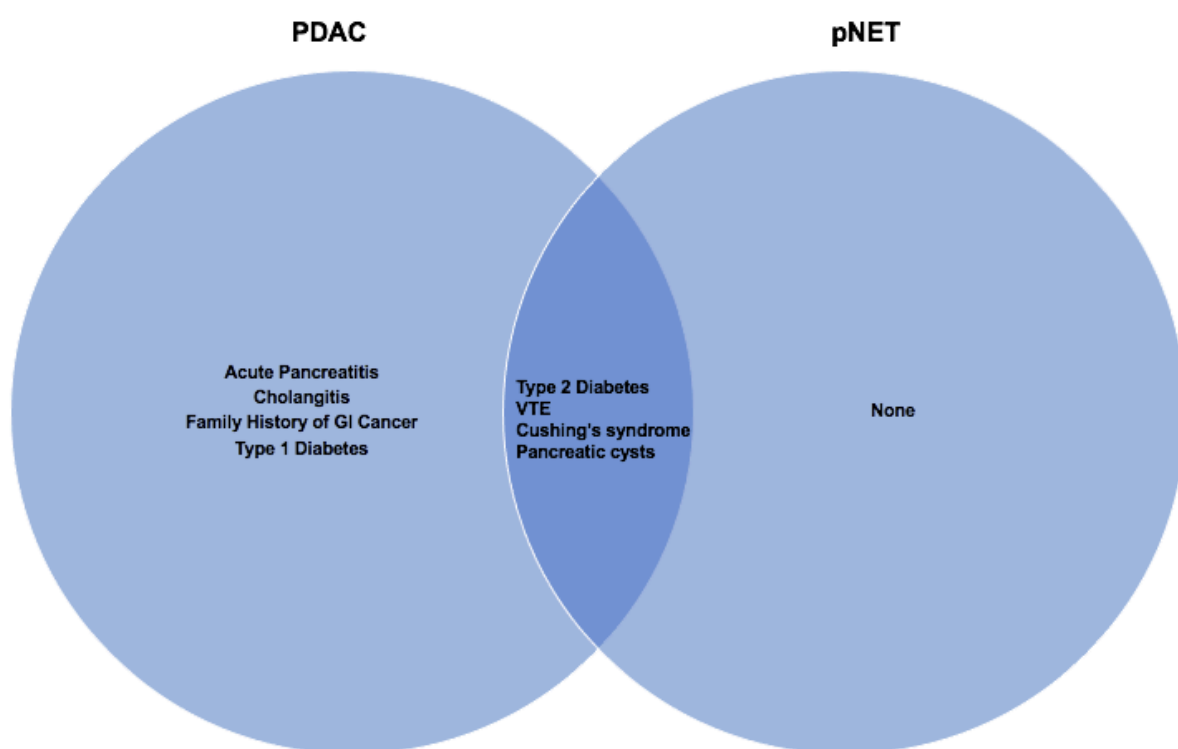
Figure 1.2: Comorbidities Venn Diagram for PDAC and pNET.

The Venn Diagram illustrates the specific comorbidities seen in PDAC alone, pNET alone and in both conditions identified from the 2021 Liao et al. study.

A report by Capurso et al., describes another case control study which comprised of 162 pNETs and 648 controls[68]. In this study participants were interviewed using a specific questionnaire on demographics and potential risk factors, which included smoking, alcohol, height, weight, medical history and family history of cancer[68]. The study found that a first-degree family history of any cancer was a significant risk factor with an OR of 2.2 (CI (1.5-3.2)[68]. Moreover, they found that a family history of PDAC was more frequent in pNETs than in controls (4.2 vs 1.25, p=0.01)[68]. They also found that high

alcohol intake OR 4.8, CI(2.4-9.5), history of CP (OR of 8.6, (CI 1.4-51) and recent onset of diabetes OR 40.1, (4.8-328.9) were all independent risk factors[68]. Interestingly, they found that a history of diabetes was also associated with metastatic disease at the time of diagnosis[68].

Another study that explored risk factors was performed by Girladi et al, who investigated a total of 184 cases and 248 controls, with 100 of the cases being pNENs[69]. This study also reported that type 2 diabetes was a pNEN risk factor (OR 3.01, (CI 1.15-7.89)[69]. They also found increased risk associated with a history of pancreatic cancer (OR 5.88, (CI 2.43-1422), increased risk with a family history of pancreatic tumours (OR 1.94, (CI 1.19-3.7) and an increased risk of pNETs for patients who had a family history of lung tumours (OR 3.01, (CI 1.15-7.89). The case group for this study included both pNENs and lung NENs, but not MEN1 cases, but tumour functionality details were not collected. The mean age of controls was significantly younger compared to both lung (p<0.001) and pancreatic NENs (p<0.001). Moreover, patients with both lung NENs and pNENs were heavier than controls (p=0.033) and among the pNET group only the prevalence of diabetes was significantly higher compared to controls (p=0.002). pNENs patients and lung NEN patients both had a significantly higher prevalence of a family history of tumours compared to controls (p<0.001 and p=0.001, respectively). Additionally, patients with lung NENs and pNENs reported a significantly higher prevalence of family history of lung cancer compared to controls. Increased age was shown to be a risk factor for developing pNENs and lung NENs (OR=1.03, CI 1.01-1.05). Based on these data, the authors suggested a link between neuroendocrine neoplasia and cancer in general through a cancer predisposing landscape. Genes involved in familial susceptibility are well known for both lung NENs and pNENs[69]. Heritable syndromes associated with NEN development including MEN1, VHL, NF-1 Cowden, tuberous sclerosis and Li-Fraumeni syndromes. As genome-wide association data are lacking for patients with neuroendocrine cancer, more subtle genetic associations have not been elucidated. MEN1 cases were excluded from the study and additional hidden genetic traits may exist within this cohort. Another limitation of this study is the small sample size, most likely as a result of the low pNEN prevalence.

A study by Ben et al. also explored pNET risk factors. This study included 385 pNET patients and 614 age and sex-matched controls[71]. Patients with inherited

syndromes including MEN1 and VHL were excluded, but pNET functionality and grade data were included. Most F-pNETs were insulinomas and most pNETs were well-differentiated G1 tumours. The study also investigated risk factors for the total pNET group and for the F-pNET and NF-pNET subgroups. Multivariable analysis showed that only heavy alcohol use was independently associated with F-pNET risk with multivariate adjusted odds ratio (AORs) (AOR 1.87, CI 1.01-3.51). AORs are ORs which have been adjusted to account for other predictors. Multivariable analysis, which is an analysis with one outcome (dependent) and multiple independent (a.k.a predictor or explanatory variables) was used in this study with multivariable analysis with adjustment for risk factors for NF-pNETs showing that ever/heavy smoking, first degree family history of cancer and history of diabetes were independent risk factors for NF-pNETs. Additionally, multivariable analysis with adjustments for ever smoking and first degree family history of cancer showed that new-onset diabetes (<1 year) was significantly associated with development of NF-pNETs with an AOR of 2.67 (CI 1.37-5.20). As seen with the other studies, the increased risk of NF-pNETs in subjects who have a family history of cancer could be due to unknown genetic factors or shared environmental factors.

The limitations of this study, as with the other previously mentioned studies, include not considering medications such as statins and aspirin which are both associated with the risk of PDAC[72]. Additionally, in this study data on the usage of hypoglycaemia treatments such as metformin, thiazolidines and insulin were not available for most of the diabetic individuals[71]. This may be important as metformin has been shown to have anti-proliferative effects in pNET cell lines[73] and insulin therapy is associated with an increased risk of colorectal cancer[74]. Thus, the association between certain medications and pNET risk deserves more exploration.

The Ben at el. study also did not distinguish between type 1 and type 2 diabetes. However, it is most likely that the majority of the diabetic individuals had type 2 diabetes because of the late onset and because they were treated only with oral hypoglycaemic agents[71]. Importantly, this study shows the different risk factors between F-pNETs and NF-pNETs suggesting different biological behaviour and clinical characteristics exist between the two disease subtypes.

1.3 pNET diagnosis utilising blood-based assays

Diagnostic markers for NETs were first evaluated at the tissue level, whereby several proteins involved in hormone secretion, such as synaptophysin[75] and chromogranin A (CgA)[76], were found to be specific to the neuroendocrine cell phenotype[77]. These initial findings in NET tissues were significant in the further development and consequent implementation of clinical tests for circulating biomarkers, including CgA and pancreatic polypeptide[77]. These biomarkers now provide clinicians with a useful and less invasive way of diagnosing, monitoring and predicting treatment responses in patients[77]. The five year survival rate of pNET patients is 37.6% and is lowest in comparison to the other assessed NETs[78]. This highlights the importance of having more sensitive and specific biomarkers to improve pNET detection, ideally at an early stage. Multiplex tests for pancreatic cancer diagnosis are discussed in section 1.3.1. microRNA (miRNAs) for pNET diagnosis are discussed in section 1.3.2 and metabolomics approaches are discussed in section 1.3.3.

1.3.1 Multiplex tests for Cancer Diagnosis

Multiplex blood tests, which use a number of analytes, that have been used for the detection diagnosis of pancreatic cancers include the Galleri, CancerSEEK and NETest tests.

1.3.1.1 Galleri (GRAIL) test

The Galleri (GRAIL) blood test was developed by the healthcare company GRAIL and works by a targeted methylation analysis of circulating cell-free DNA (cfDNA) to detect and localise multiple cancer types across all stages[79]. A study exploring the ability for this blood test to detect cancer was carried out. The study consisted of 6689 participants which in turn consisted of 2482 cancer patients across 50 different cancer types and 4207 non-cancer patients. cfDNA samples were prepared and divided into training and validation sets. They found a consistent level of performance in the training and validation sets with a specificity of 99.3% in the latter. Moreover, they found a stage 1-3 sensitivity of 67.3% in a pre-specified set of 23 cancer types (anus, bladder, colon, oesophagus, head and neck, liver/bile duct, lung, lymphoma, ovary, pancreas, plasma cell neoplasm and stomach) and a sensitivity of 43.9% in all cancer types. Detection increased with increasing stage, for example sensitivity was

39% at stage 1 compared to 93% at stage 4. There were some limitations to this study, including the small sample size compared to a larger scale prospective study and the smaller sample numbers for individual type of cancer. This means that the representation of heterogeneity within some cancer types was less. Additionally, it is unclear as to how many NETs were identified using the Galleri test. Larger population-based studies are in progress to assess the suitability of the Galleri test for the detection of cancer. A UK GRAIL pilot study with 165,000 participants in total began in 2021[80]. The participants included a group of 140,000 participants aged 50 to 79 with no symptoms who will have blood tests every three years. Anyone within this group who has a positive test will be referred for further investigation by the NHS[80]. The other participant group of 25,000 participants with possible cancer symptoms will also be offered testing after referral to hospital to see if the Galleri test speeds up the diagnosis time. The results from this study are expected in 2023 and, dependent on the findings of the study, this will be expanded to around 1 million participants in 2024 and 2025. The SYMPLIFY study is the first large scale evaluation of a multi cancer early detection test in individuals presenting in primary care and referred for diagnostic follow up for suspicion of cancer. The study enrolled 6238 patients aged 18 and over in England and Wales who were referred for urgent imaging, endoscopy or other diagnostic modalities to investigate symptoms suspicious for possible gynaecological, lung, lower GI or upper GI cancer or who have presented with non-specific symptoms. Within the study 368 (6.7%) of the 5461 evaluable patients were diagnosed with cancer through standard care, with the most common cancer diagnosed being colorectal (37,2%). GRAIL'S multi detection cancer test detected a cancer signal in 323 people, 244 in whom cancer was diagnosed, resulting in a PPV of 75.5%, NPV of 97.6% and specificity of 98.4%. The overall sensitivity of the test was 66.3% with this ranging from 24.2% in stage 1 cancers to 95.3% in stage 4 cancer[81].

1.3.1.2 CancerSEEK

The CancerSEEK test was developed in the USA by scientists at John Hopkins and aims to detect 8 types of cancer (liver, stomach, ovary, pancreas, oesophagus, colon, lung and breast) through assessment of the levels of circulating proteins and mutations in cfDNA[82]. The test was applied to 1005 patients with these eight types of non-metastatic, clinically detected cancer, and

resulted in test sensitivities ranging from 69% to 98% for 5 cancer types (ovary, liver, stomach, pancreas and oesophagus) at > 99% specificity. CancerSEEK localised the site of the cancer with a median of 83% of patients, but there were some important limitations to this study. First, the cohort consisted of patients with known cancers diagnosed on the basis of symptoms of disease; however, in practice in a true screening setting most individuals would have less advanced disease. Moreover, the controls in this study were limited to healthy individuals but in a true cancer screening setting, individuals may have other diseases or inflammatory conditions which could result in a greater proportion of false positives than observed in the study. The estimate for the cost of the test is estimated to be less than $500, which is comparable to or lower than other screening tests for single cancers such as colonoscopy. As with the Galleri test, to truly evaluate CancerSEEK, larger prospective studies in a genuine clinical setting are required.

1.3.1.3 NETest

The NETest utilises 51 different circulating ribonucleic acids (RNAs) for NET detection with an AUC of 0.98[83]. Despite the strong performance which is superior to the widely used CgA test there are important issues to consider with this test. These includes cost, practicality and reproducibility in early clinical use associated with the amplification of the 51 different RNA targets using polymerase chain reaction (PCR) thus better diagnostic tests are needed for early pNET detection are needed, ideally a point of care test for use in primary care.

1.3.2 MicroRNAs

miRNAs are a series of small noncoding RNAs which are able to regulate gene expression[84]. Circulating miRNAs derived from a tumour can be detected via quantitative polymerase chain reaction (qPCR)[84]. Certain miRNAs have been shown to be elevated in pNETs. A study by Thorns et al., aimed to identify miRNA signatures of pNEN in tissue and serum, with the study including tissue samples from 37 pNEN patients, 9 patients with non-neoplastic pancreatic pathology, 7 samples of micro-dissected pancreatic islets and serum samples from 27 pNEN patients and 15 healthy controls[85]. In this study miRNA expression profiles were established through utility of qPCR for 754 miRNAs.

They found that miR-193b was more highly expressed in pNEN tissue than the islet cells (p=0.039). When comparing the miRNA levels in the serum from patients with pNEN and healthy controls, 13 miRNAs were more abundant in the serum of patients. miRNA-193b was significantly more abundant in the serum of patients (P=2.02×10^{-5}). The mechanism of miRNA access to the blood is unclear. They may be released into the blood during tumour necrosis or apoptosis. However, in G1 and G2 tumours, necrosis is not common and so this is most likely not the source of circulating miRNA in pNEN. miRNA may also be secreted in an ATP-dependent manner as exosome enclosed or agonaute-2 (AGO2/NPM1)-associated miRNAs[86,]. The tumour itself may not be the source of miRNA but the tumour may directly or indirectly induce the secretion of miR-193b from other cells. miR193B has been shown to be overexpressed in high grade prostate carcinomas, however sera levels of this miRNA were not looked at in this study[87,]. Another miRNA, miRNA-1920 was shown to discriminate between PDAC and pNET with an AUC of 0.80 in sera[88]. However, the clinical utility of circulating miRNAs for pNET diagnosis could be hampered by the lack of standardised measurements[84].

1.3.3 Metabolomics

Metabolomic approaches have shown to be suitable for detection of NETs from healthy controls. A study by Kinross et al.,2013 explored metabolomic approaches in urine for NET detection, with this study consisting of 28 patients with NENs who were prospectively recruited and 17 healthy controls. Within the group of NETs there were 10 pNETs. Urine samples were subjected to 1H nuclear magnetic resonance spectroscopy profiling with data obtained from this used for supervised for supervised and unsupervised analysis. They found an AUC of 0.90 to differentiate heathy vs NET. Analysis also showed the ability to distinguish small bowel NETs and pNETs. A subclass analysis also demonstrated class separation between functional and non-functional NETs and those with metastases. Thus metabolomic profiling could provide novel biomarkers for NETs[89].

1.4 GEP-NET serum markers

CgA, tissue inhibitor of metalloproteinases 1 (TIMP1), VGF-nerve growth factor (VGF-NGF), thymosin beta 4 (TB4), NSE, Mac-2-binding protein (MAC2BP) and angiopoietin-2 (ANG2) were used for the development of a GEP-NET biomarker panel by our group (unpublished). These 7 markers are described further in sections 1.4.1-1.4.8. They were also the selected markers for my research project in which a pNET biomarker panel was developed.

1.4.1 Chromogranin A

CgA is a 46kDa glycoprotein and a member of the granin family[90], which has 9 members[91], namely CgA, chromogranin B (CgB), secretogranins SG2 to SG7, VGF and Proprotein convertase subtilisin/kexin type 1 inhibitor (proSAAS)[91]. The shared properties of this family include their acidic pI, heat stability due to their hydrophilic nature, multiple dibasic cleavage sites, and ability to form aggregates that bind calcium[92]. Granins are also major components of large dense core secretory vesicles, and are co-secreted with peptide hormones and amines[93]. CgA was the first to be discovered in 1965[94] and was subsequently purified from the bovine adrenal medulla in 1967[95]. CgB was subsequently identified in bovine adrenal medulla in 1985[96].

A general characteristic of GEP-NETs is the expression of CgA which is released from many non-transformed neuroendocrine cells[97]. In pNETs increased CgA expression is associated with advanced disease progression and is considered a good marker for treatment response to Everolimus for advanced pNETs[98]. Moreover CgA has been suggested as a predictor of radiological disease with predictive value 6 months before radiological progression of pNETs[99]. CgA is a widely used serum diagnostic biomarker for many pNETs but not for insulinomas[90]. This highlights the crucial point that insulinomas are biologically different to other NENs, a property that needs further exploration. Moreover, it strengthens the reasoning for a multianalyte approach for pNET detection as opposed to a single analyte approach.

Nobles et al. found that 2 of 21 (10%) insulinoma patients had an elevated CgA level and 38% had an elevated NSE level[100]. Additionally, 9 of 13 (69%) NF pancreatic islet cell tumour patients had an elevated CgA level and 4 (31%) had an elevation in NSE[100]. Despite their small cohort sizes, these results suggest a higher suitability of NSE for insulinomas compared to CgA and of CgA

compared to NSE as a marker for NF-pNETs. This highlights biological differences between NF-pNETs vs insulinomas, and potentially other pNET subtypes.

A further study supports the notion that CgA is not a reliable marker for insulinomas. When comparing healthy controls, insulinomas and non-insulinomas, Qiao et al. reported serum CgA levels being slightly elevated in patients with insulinomas and significantly increased in patients with non-insulinoma pNETs[90]. They also reported that the median CgA level in insulinoma patients was 64.8ng/ml, which was significantly lower than the median of 192.5ng/ml seen in non-insulinoma pNETs[90], once again highlighting differences between insulinomas and non-insulinoma pNETs. This could be attributed to the fact that >90% of the insulinomas were benign, similar to other studies[101] and reflecting the characteristic absence of metastasis in most insulinoma patients[90].

The presence of metastasis does not fully account for increased CgA levels in non-insulinoma patients. For example, a specific study of gastrinomas showed that CgA levels were raised in this group despite the absence of metastasis[102]. Moreover, in addition to the absence of metastasis, high levels of insulin in insulinoma patients might inhibit secretion of CgA by these tumour cells[90]. Therefore, both metastasis and tumour subtype are potentially important determinants of CgA levels[90].

1.4.2 Tissue Inhibitor of Metalloproteinase 1

Tissue Inhibitor of Metalloproteinase 1 (TIMP1) belongs to the TIMP family, which includes TIMP1, 2, 3 and 4[103]. The TIMP1 protein has a main function of inhibiting matrix metalloproteinases (MMPs) by forming non-covalent 1:1 stoichiometric complexes, thereby regulating the balance of matrix remodelling during the process of degradation of the extracellular matrix (ECM)[104].This inhibitory effect on the MMPs is significant for the tumour invasion property and development of metastatic disease. TIMP1 is known to inhibit apoptosis and in lung adenocarcinoma cells TIMP1 has an anti-apoptotic effect via an interaction with B-cell lymphoma 2 (Bcl-2)[105]. Bcl-2 is also an anti-apoptotic molecule, with it's overexpression known to increase TIMP1 expression[106].

TIMP1 has been identified as a secreted protein in 2 NET cell lines, BON-1 and SHP-77[107]. However, its function in pNET pathology is unclear. It is thought that

TIMPs can regulate cellular functions via MMP-dependent and independent mechanisms. In the MMP-dependent mechanism, TIMPs interact with soluble or membrane-bound MMPs or other metalloproteinases in the cellular periphery modulating cleavage of ECM proteins, which include collagens, laminin and fibronectin as well as signalling proteins such as cytokines, chemokines and growth factors that can activate or supress receptor mediated signalling in the cell[108]. The MMP-independent mechanism involves a direct interaction of TIMPs with a specific cell surface receptor to initiate intracellular signalling cascades, leading to altered gene expression and changes in cellular behaviour[108]. Elevated stromal expression of TIMP1 promoted liver metastasis in two independent tumour models[109]. This induced the hepatocyte growth factor signalling pathway and expression of several metastasis associated genes[109]. Furthermore, TIMP1, but not TIMP2, prevented cytokine-induced apoptosis and cytokine-mediated inhibition of glucose-stimulated insulin secretion in rat islets and β-cells[110]. Moreover, TIMP1 did not impact nitric oxide (NO) production or inducible nitric oxide synthase (iNOS) gene expression[110]. As overcoming cytokine-induced β-cell damage is an important goal for protecting pancreatic β-cells from immune mediated destruction in type 1 diabetes, this work infers TIMP1 may be an ideal gene to prevent cytokine-mediated β-cell destruction and dysfunction in models of type 1 diabetes as well as islet transplantation rejection[110]. This thus, may also be of relevance to patients with type 1 diabetes.

TIMP1 has been suggested to be a suitable biomarker for malignancies such as PDAC[111] and colorectal cancer[112]. Specifically, urinary TIMP1 (uTIMP1) could differentiate between PDAC and control samples with a sensitivity of 90% and specificity of 70%[111]. Importantly uTIMP-1 could also distinguish between PDAC and pNET samples, which is potentially of clinically importance because discriminatory biomarker tests for pancreatic diseases are needed to help clinicians make accurate diagnosis when patients could present with overlapping symptoms. Moreover, the results from this study suggest that TIMP1 has a more specific role in the progression of exocrine PDACs than pNETs, thereby highlighting differences in the biology of both types of pancreatic cancers and the selective utility of this biomarker in being able to differentiate between endocrine and exocrine cancers of the pancreas. This study also found that MMP2 was an independent predictor of the presence of

pNET[111]. Moreover, urinary MMP2 levels were 1.7ng/ml (1.1-3.3) for pNETs and 1.2ng/ml (1.0-1.6) for PDAC compared to 0.8ng/ml (0.2-1.0) for controls. It has been suggested that lipocalin-2 (LCN2) and TIMP1 are potential serum markers for the early detection of familial pancreatic cancer (PDAC)[113]. This group of patients has a higher risk of developing pancreatic cancers, which have a prevalence of 3.5% (range 1.1-3.5%)[114].

1.4.3 Mac-2-binding protein

Mac-2-binding protein (MAC2BP), also known as galectin 3 binding protein, is a secreted glycoprotein which is found in various bodily fluids including serum, saliva, semen, and urine[115]. MAC2BP and galectins are located in the ECM, where they may play a role in cell attachment by interactions with $\beta1$ integrins, collagens and fibronectin[116]. It is implicated in various cancers including breast[117], oesophageal squamous cell carcinoma[118], nasopharyngeal cancers[119] as well as NETs[107]. MAC2BP interaction to integrins on tumour cells leads to activation of the Protein Kinase B (AKT) and extracellular signal-regulated kinase (ERK) pathways that are associated with increased survival, proliferation, motility an migration of cancer cell lines[120]. High expression of MAC2BP is associated with poor survival in node-negative breast cancer patients who were not receiving systemic adjuvant therapies[121]. MAC2BP is a novel E-selectin ligand expressed by breast cancer and patients with breast cancer who have MAC2BP overexpressing tumours are more likely to develop distant metastasis than those with low MAC2BP expression[121]. Mechanistically, Shirure et al., proposed three possible scenarios for MAC2BP expressing breast cancer cells with these being either metastasis, no adhesion or deletion. The first scenario is that breast cancer cells expressing high levels of MAC2BP bind to galectin-1 (Gal-1) but also possess enough free epitopes for E-selectin binding, thus, these cells are more likely to form metastatic lesions. The second scenario is that breast cancer cells that express low levels of MAC2BP may not bind to the endothelium due to blockade of E-selectin ligand function by Gal-1. The third scenario is that breast cancer cell with an absence of Gal-1 may lead to deletion of cancer cells by immune cells. However, in pNETs a Gal-1 based mechanism such a seen in breast cancer is not relevant because despite pNETs expressing E-selectin[122] they do not express Gal-1, or other galectins such as Gal-3 or Gal-4[123] cells[117]. A study in which the quantitative proteomic

profiling of the extracellular matrix of pancreatic islets during the angiogenic switch and insulinoma progression revealed that galectin-1 was present in lower abundance in the ECM of both angiogenic islets and insulinomas[124].

Cell secretome analysis has revealed MAC2BP secretion by 3 NET cell lines, including the pNET cell line BON-1[107]. MAC2BP was expressed by the majority of NET tissue samples and serum MAC2BP levels were significantly elevated in NET patients compared to healthy controls[107]. In terms of the primary site of the NET, serum MAC2BP levels were significantly elevated in midgut NETs (3.34µg/ml; range of 0.82-10.66µg/ml) and pNETs (2.67µg/ml; range of 1.37-10.50µg/ml) compared to controls[107]. A logistic regression (LR) receiver operator characteristic (ROC) curve for all NETs vs controls based on serum MAC2BP level revealed an Area under the curve (AUC) of 0.77.

1.4.4 Neuron-Specific Enolase

Neuron-specific enolase (NSE) belongs to a group of intracellular enzymes known as enolases, which catalyse the conversion of 2-phospho-D-glycerate to phosphoenolpyruvate in the glycolytic pathway[125]. This pathway is involved in the conversion of glucose into pyruvate[126]. They are cytoplasmic enzymes required for aerobic glycolysis, which provides the main source of energy in cancer cells aiding cell proliferation[125]. Enolases are functionally active as dimers composed of non-covalently linked alpha (α), beta (β), and gamma (γ) subunits[125]. The $\gamma\gamma$ homodimer is known as NSE[127]. The biological half-life of NSE in bodily fluids is around 24 hours, and elevated levels of NSE in serum and or cerebrospinal fluid may activate different pathways leading to neuroinflammation[128]. Cell surface enolase is found on microglia, activated macrophages, neurons and astroctyes[129]. The expression of NSE at the cell surface promotes cell migration, survival and growth as well as initiating inflammatory processes after injury[129]. Cell surface enolase can act via different pathways including the mitogen-activated protein kinase kinase (MEK)/ ERK pathway, phosphoinositide 3-kinase (PI3K)/AKT pathway. The MEK/ERK pathway induces cellular proliferation, differentiation and migration and the PI3K-AKT pathway promotes glycolysis, cellular proliferation and migration[129]. NSE is important in diabetes and cancer. Increased levels of NSE are seen in diabetic patients and higher levels of NSE have been associated with diabetic

neuropathy[128]. NSE is also overexpressed in tumours where its primary role is to accelerate glycolysis[125]. Glycolysis is mediated via the enzymatic part of the peptide whereas other functions for NSE are mediated by interactions through the C-terminal part of the peptide[130]. The C-terminal part of NSE promotes cell survival and neurite outgrowth via activation of the PI3K/AKT and MAPK/ERK pathways[130]. NSE is also greatly up-regulated in glioblastoma cells which were exposed to hypoxia and serum starvation[131].

NSE is localised both in the cytoplasm and also on membranes of primary and metastatic cerebral tumours and on glial filaments of glioma cells[127]. In neuroblastoma cells, NSE co-localised with actin filaments, and the co-localisation depended on the presence of gamma-1-syntrophin[132]. This suggests that NSE might be involved in migration of tumour cells. In lymphoma cells, NSE regulates the polarisation of macrophages to the M2 phenotype, and contributes to the proliferation and progression of lymphoma[133].

NSE has been used as a tumour marker for monitoring the response to chemotherapy in advanced adenocarcinoma, squamous cell carcinoma and small cell lung cancer[134]. Moreover, serum NSE level has been associated with CTC levels in lung cancer patients[135]. NSE has also been associated with the development of intracranial metastasis in patients with lung or breast cancer[136]. However, 28.3% of patients in this study showed no increase in 4 serum tumour markers including NSE. This high percentage suggested better monitoring of these patients and better markers are needed for this subgroup of patients[136]. NSE was not significant for monitoring the efficacy of nivolumab in patients with advanced non-small cell lung cancer whereas the reduction in serum level of two other markers, Cytokeratin 19 fragment (CYFRA21-1) or carcinoembryonic antigen (CEA), might be reliable biomarkers[134]. In small cell carcinoma of the pancreas, NSE has been shown to be a good marker, both for diagnostic purposes but also for an evaluation to treatment[137]. NSE is also secreted by three NET cell lines BON1, NCI-H727 and SHP-77[107].

1.4.5 Angiopoietin-2

Angiogenesis is a vital process in cancer growth, maintenance and metastasis[138], and the angiopoietin (ANG)/TIE (tyrosine kinase with immunoglobulin (Ig) and epidermal growth factor (EGF) homology domains) signalling pathway is one of the main pathways linked to angiogenesis[139].

Angiopoietin-2 (ANG2) is a growth factor that activates the ANG/TIE signalling pathway[139] by binding to the TIE2 receptor tyrosine kinase. Unlike angiopoietin-1 (ANG1), ANG2 acts in an autocrine manner and its expression is tightly regulated in healthy tissue[139]. ANG2 signals synergistically with VEGF as well as with ANG1 through TIE-dependent and independent pathways. ANG2 may be involved in the facilitation of endothelial cell migration and proliferation in pro-angiogenic signalling[140].

ANG2 mRNA was found to be significantly overexpressed in hypervascular hepatocellular carcinoma, and overexpression of ANG2 led to rapid tumour growth in an animal model[141]. NETs are highly vascularised in nature[142] and an enhanced peripheral blood plasma TIE2 concentration is a diagnostic marker for NET patients with a similar level of performance to CgA[140]. Higher plasma levels of ANG2 in conjunction with a positive correlation with TIE2 levels in NET patients with metastasis indicated that both ANG2 and TIE2 are predictive of metastases[140]. In NET patients' sera ANG2 levels were significantly elevated compared to healthy controls, however ANG1 levels displayed no differences. Additionally, serum ANG2 levels were more elevated in patients with distant metastases than in patients without metastasis. Moreover, time to disease progression was worse in patients with elevated serum ANG2. Hence ANG2 would be a useful serum marker for monitoring and assessing the prognosis of NET patients[138].

1.4.6 Thymosin Beta 4

Thymosin beta 4 (Tβ4) is a 43 amino acid peptide with a molecular weight of 4.9kD[143]. It is a part of the thymosin family and was first isolated from bovine thymus tissue[144]. It is present in all mammalian cells investigated except erythrocytes[145]. The family consists of α, β and γ thymosins[146]. Within the thymosin beta family, three subtypes exist including Tβ4, 10 and 15[147]. Tβ4 has a well-studied intracellular role, involving the regulation of actin polymerisation[148]. Actin polymerisation is a process which is vital for many aspects of cellular function including cytoskeleton maintenance, cell morphology and cell migration[149]. Within this process, Tβ4 binds to monomeric actin, known as G-actin. This interaction prevents its polymerisation into F-actin microfilaments.

There is also growing evidence for extracellular Tβ4 signalling, for which two potential mechanisms have been suggested. In the first it is internalised and then acts intracellularly to promote cell migration, while in the second it signals via a surface receptor[150]. Extracellular Tβ4 can modulate platelet activity and has a biphasic effect on thrombus formation. Work by Kaur et al., showed that low concentrations of TB4 (0.2μM to 0.5μM) increased both the rate constant of platelet deposition by around 1.5 fold and the total number of deposited platelets by around 3 fold[151]. However, at higher concentrations (>1μM) the Tβ4-potentiating effect was diminished to near control levels[151]. Thus, at low concentrations, Tβ4 could increase the affinity of platelet receptors for their ligands and hence promote platelet deposition. However at higher concentrations Tβ4 can bind to fibrinogen and prevent fibrinogen–platelet interactions and thereby attenuate platelet deposition[151,145]. This dual action of Tβ4 has potentially important implications for several disease conditions. In conditions such as thrombocytopaenia, gray platelet syndrome, in which platelet adhesion and aggregation are halted, the introduction of submicromolar amounts of Tβ4 to the blood would be expected to increase platelet activity[145]. Whereas in conditions such as diabetes, atherosclerosis and coronary artery disease supramicromolar levels of Tβ4 could be used to prevent platelet hyperactivity[145].

Tβ4 has also been shown to have various important extracellular functions encompassing many aspects of tumour biology. This includes apoptosis of Hela cells[152], tumour invasion by A549 and H157 cells,[153] and metastasis as seen for mouse melanoma B16 cell lines[154]. Tβ4 has also been shown to play a role in promoting cardiomyocyte migration and cardiac repair, which are important in the treatment of conditions such as myocardial infarction[155]. The mechanism by which Tβ4 is thought to act on cardiomyocytes is via the formation of a functional complex with PINCH and integrin-linked kinase (ILK), resulting in activation of the AKT protein kinase, which promotes cardiomyocyte migration and survival during cardiac repair[155]. Tβ4 is also known to upregulate transcription of laminin 5, which induces cell differentiation, adhesion and motility, and it protects cardiomyocytes from oxidative stress by targeting anti-oxidative enzymes and anti-apoptotic genes[156]

Tβ4 is also thought to have anti-inflammatory properties by inhibiting TNFα-induced Nuclear factor kappa B (NFKB) activation[157]. Tβ4 is involved in the pathology of diabetic peripheral neuropathy, but mechanisms for this have not been fully established[158]. It has been suggested that activation of ANG/TIE2 signalling contributes to the anti-inflammatory nature of Tβ4[158] as inhibition of TIE2 activity with an anti-TIE2 antibody abolished Tβ4-suppressed p- NFKB and vascular cell adhesion protein 1 (VCAM1) expression, but not monocyte chemotactic protein-1 (MCP1) expression. Thus, blockage of the ANG1/TIE2 pathway with the antibody against TIE2 may not affect MCP1. This study provided in vivo evidence that the ANG1/TIE2 pathway contributes to the therapeutic effect of Tβ4 on diabetic peripheral neuropathy[158].

Tβ4 also has an oxidised form known as thymosin beta sulphoxide (Tβ4SO). Tβ4SO has an important role in neutrophils. It is an anti-inflammatory peptide which down regulates neutrophil mediated inflammation. Mechanistically this is via the action of neutrophil apoptosis and their clearance via phagocytic macrophages[159].

1.4.7 VGF-nerve growth factor

The VGF-nerve growth factor inducible peptide (VGF) gene encodes a 68kDa peptide that comprises 615 amino acids (human) and 617 amino acids (mouse and rat)[160]. VGF is a member of the chromogranin/secretogranin family[161]. VGF is stored in large dense core vesicles which are located in neuroendocrine, endocrine and neuronal cells[161]. The VGF gene was originally found on the basis of its regulation via NGF in rat phaeochromocytoma cells[162]. VGF plays a role in various neurological diseases, including Huntington's disease[163], amyotrophic lateral sclerosis[161] and bipolar disorder[164]. VGF is regulated by both brain-derived neurotrophic factor (BDNF) and serotonin[165]. Localisation studies have indicated that VGF is distributed widely across the nervous system in neurons and in a number of neuroendocrine tissues, which include the pituitary and adrenal medulla, and also in various gastrointestinal and pancreatic endocrine cell types[160]. The receptor for VGF has not been identified[165]. The relationship between VGF and pNENs has been explored[166]. In vitro and in vivo data demonstrated that long non-coding RNA (lncRNA) H19 overexpression promoted metastasis and tumour growth with H19 knockdown leading to the opposite effects[166]. This study demonstrated an H19 interaction

with VGF as well as VGF upregulation in pNENs[166]. Mechanistically H19 activated phosphoinositide 3-kinase(PI3K)/AKT/ cyclic adenosine monophosphate (cAMP) response element-binding protein (CREB) signalling and promoted pNEN progression by interacting with VGF[166]. Thus, VGF has a potentially important role in pNEN progression in man.

1.5 GEP-NET serum biomarker confounding factors

During the process of developing a biomarker test a key consideration is the effect of confounding factors. Confounding factors can be internal or external to patients[167]. Internal confounding factors include age, gender, diet and other metabolic factors[167]. In practice however, it is difficult to identify and control for all potential confounding factors. All 7 biomarkers within the GEP-NET panel discussed in Section 1.4 have confounding factors, which are discussed in subsections 1.5.1 to 1.5.7.

1.5.1 Confounding factors for CgA in serum and plasma

CgA levels are elevated in the plasma or sera of various malignancies including colon cancer[168], hepatocellular carcinoma[169], lung cancer[170], bladder cancer[171],melanoma[172], multiple myeloma[173], ovarian cancer[174], head and neck cancer[175], prostate cancer[176], PDAC[177], small cell lung cancer[170], neuroblastoma[178] and medullary thyroid carcinoma[179]. CgA levels are also elevated or reduced in either the plasma or sera of non-malignant conditions including chronic obstructive pulmonary disease (COPD)[180], chronic pancreatitis[181], cirrhosis[182], diabetes[183], heart failure[183], hypertension[184], idiopathic pulmonary fibrosis[180], inflammatory bowel disease[185], kidney dysfunction[186], major depressive disorder[187], myocardial infarction[188], non-alcoholic fatty liver disease (NAFLD)[189], rheumatoid arthritis[190], sepsis[191], systemic lupus erythaematosus[190], type 1 diabetes[192], type 2 diabetes[193], vertigo[194], diabetic neuropathy[195], epilepsy[196], hyperthyroidism[197], schizophrenia[198], hypothyroidism[183], chronic diarrhoea[183], peptic ulcer[183], chronic periodontitis[199], birth[200], Meniere's disease (an inner ear disorder)[194] and gastritis[201]. Treatments associated with influencing CgA levels include, proton pump inhibitors[202], endocrine treatment for prostate cancer[203], abiraterone acetate therapy[204] and the nGR-hTNF vaccine for metastatic

melanoma[172]. Other factors influencing CgA levels in plasma and sera include age[188,190], creatinine clearance[188], marathon running[205], and alpha-fetoprotein[206].

1.5.2 Confounding factors for TIMP1 in serum and plasma

Serum levels of TIMP1 are also increased in patients with various cancers. These include melanoma[207], metastatic breast cancer[208], hepatocellular carcinoma[209], ovarian cancer[210], PDAC[211], multiple myeloma[212], and lung adenocarcinoma[213]. TIMP1 serum or plasma levels are also elevated or decreased in range of non-malignant conditions. These include COPD[214], rheumatoid arthritis[215], psoriasis[216], ulcerative colitis[217], hypertension[218], rotator cuff tear[219], metabolic syndrome[220], systemic sclerosis[221], migraine aura[222], burns[223], polycystic kidney disease[224], autoimmune ear disease[225], cystic fibrosis[226], pneumonia[227], juvenile idiopathic arthritis[228], jaundice[229], cachexia[229] and osteoarthritis[230]. Moreover, TIMP1 predicts adiposity in humans with plasma levels of TIMP1 being significantly different between lean, obese and visceral obese patients[231]. Certain medications and treatments are known to have an impact on TIMP1 levels including methotrexate[215]. Oral contraceptive treatment for 6 months significantly reduced TIMP1 concentration levels in patients who had this treatment for polycystic ovary syndrome[232]. Statin treatment is also known to impact TIMP1 levels[233]. TIMP1 is also a powerful predictor of long-term mortality in heart failure patients treated with cardiac resynchronisation therapy[234].

1.5.3 Confounding factors for MAC2BP in serum and plasma

MAC2BP has been explored in the serum or plasma of patients with various malignancies. These include breast cancer[235], gastric cancer[236], lung cancer[237], lymphoma[238], hepatocellular carcinoma[239] and oesophageal squamous cell carcinoma[118]. MAC2BP has also been shown to be a plasma or serum biomarker of in conditions including chronic pancreatitis (CP)[240], non-alcoholic steatohepatitis (NASH)[241], angina[242], atherosclerosis[243], cirrhosis[239], coronary artery disease[244], diabetic retinopathy[245] and juvenile idiopathic arthritis[246]. For non-alcoholic fatty liver disease (NAFLD) patients, MAC2BP is associated with diabetes[247], metabolic syndrome [247], obesity[247] and hypertension[247]. Moreover in NAFLD patients MAC2BP is also associated with levels of albumin[247], uric

acid[247], alpha-fetoprotein[247], BMI[247], fasting blood glucose (FBG)[247], haemoglobin A1C (hBa1C)[247], alanine aminotransferase[247], aspartate aminotransferase[247], total bilirubin[247], and high density lipoprotein cholesterol (HDL-C)[247].

1.5.4 Confounding factors for NSE in serum and plasma

NSE has also been shown to be a blood-based biomarker for malignancies including, metastatic breast cancer[248], small cell cancers of the urinary bladder[249] and uterine cervix[250], colon cancer[251], gastric cancer[252], lung cancer[253], lymphoma[254], multiple myeloma[255], glioma[256], non-small cell lung cancer[134], head and neck cancer[257], prostate cancer[258], metastatic prostate cancer[259], squamous carcinoma of the lung[253], oral squamous cell carcinoma, neuroblastoma[260], adult T cell luekemia[254], ovarian dysgerminoma[261], ovarian choriocarcinoma [262], small cell carcinoma of the pancreas[137], medullary thyroid carcinoma[263], meningioma[100] and bone cancer[264]. NSE is also a blood-based biomarker for non-malignant conditions including; bipolar disorder[265], diabetes[266], liver failure[267], diabetic neuropathy[268], Human immunodeficiency virus (HIV)[269], idiopathic pulmonary fibrosis[270], major depressive disorder[265], metabolic syndrome[271], migraine[272], delirium in polytrauma[273], obesity[274], preeclampsia[275], pregnancy[275], pulmonary tuberculosis[276], spinal-cord injury[277], systemic sclerosis[278] , type 1 diabetes[268], type 2 diabetes[268], vascular dementia[279], vertigo[280], stroke[281], silicosis[282], interstitial lung disease[283], epilepsy[284], chronic severe traumatic brain injury[285], multiple sclerosis[286] and lean mass[205]. NSE levels are also influenced by various treatments, procedures and chemicals including magnesium sulphate[287], octreotide[288], dexmedetomidine[273], fentanyl[289], tuberculosis treatment[276], carbamazepine[284], oxcarbazepine[284], mercury[290] and cardiopulmonary by-pass[291].

1.5.5 Confounding factors for ANG2 in serum and plasma

ANG2 is known to be a blood-based biomarker for patients with malignancies including bladder cancer[292], breast cancer[293], chronic myeloid leukaemia[294], chronic lymphocytic leukaemia[295] multiple myeloma[296], type 1 endometrial cancer[297], ovarian cancer , colon cancer[299], gastric cancer[300], cervical cancer[301], glioma[302], meningioma[302], hepatocellular carcinoma[303], acute myeloid leukaemia[303], angiosarcoma[305], lymphoma[306], lung cancer[307], melanoma[308],

renal cancer[309], head and neck cancer[310], oesophageal cancer[311] and medullary thyroid cancer[179].

However, similar to other marker proteins, ANG2 levels in serum and plasma are also affected by various cardiac and circulatory conditions such as angina[312], aortic stenosis[313], coronary artery disease[314], hypertension[315], bronchopulmonary dysplasia-associated pulmonary hypertension[316], hereditary haemorrhagic telangiectasia[317], venous thromboembolism[318], heart failure[319], stroke[320], myocardial infarction[321], peripheral artery disease[322] and Crimean-Congo haemorrhagic fever[323].

ANG2 levels in serum or plasma have also been found to be of importance in respiratory conditions such as COPD[324], idiopathic pulmonary fibrosis[325], pneumonia[326], interstitial lung disease[327], tuberculosis lymphadenitis [328] and lung transplantation[329], and in gastrointestinal conditions such as AP[330], chronic hepatitis-associated cirrhosis[331], small bowel angiodysplasia[332], NASH[333], and ulcerative colitis[334]. ANG2 levels are also influenced in other conditions including atopic dermatitis[335], rheumatoid arthritis[336], sepsis[191], systemic sclerosis[337], multiple sclerosis[338], systemic lupus erythematosus [339], spinal cord injury[340], malaria[341], hyperthyroidism[342], Kawasaki disease [343], kidney dysfunction[344], juvenile idiopathic arthritis[228], human immunodeficiency virus (HIV)[345], psychological stress[346], pregnancy[347], sleep apnoea[348], brain arteriovenous malformation[349], dengue and dengue haemorrhagic fever[350], multiple trauma[351], polycystic kidney disease[352], severe malnutrition[345], pre-eclampsia[353], Kaposiform lymphangiomatosis[354], type 2 diabetes, obesity[348], metabolic syndrome[355], lean mass[356], C1 inhibitor-deficient hereditary angioedema[357] and diabetic retinopathy[358].

Certain medications are also known to have an influence on ANG2 levels. Treprostinil or Sunitinib treatment have been associated with decreased levels of circulating ANG2 levels[359,360]. Likewise, itraconazole and propranolol treatment caused a decrease in serum ANG2 levels in patients treated for infantile haemangioma[361]. Laser treatment for retinopathy also caused an increase in serum ANG2 levels from prior to post treatment in infants[362].

1.5.6 Confounding factors for Tβ4 in serum and plasma

Tβ4 has been shown in various studies to be a blood-based biomarker for obstructive sleep apnoea syndrome[363], acute on chronic liver failure[364],

rheumatoid arthritis[365], sepsis[366], psoriasis[367], pre-eclampsia[368], congenital heart disease[369], cirrhosis[370], heart failure[371], bipolar disorder[372], sleep apnoea syndrome[363], Crohn's disease[373], ulcerative colitis[373], non-Inflammatory Bowel Disease (IBD) GI disease[373], knee osteoarthritis[374], rheumatoid arthritis[365], coronary angioplasty[375], and major depressive disorder and bipolar[372]. Tβ4 levels were also decreased in several liver-related conditions. Patients with NAFLD[376], a chronic hepatitis B infection, compensated liver cirrhosis, or chronic or acute on chronic liver failure[370] had lower levels of Tβ4 compared to healthy controls. Moreover, in patients that had chronic liver disease there was a positive relationship between Tβ4 level and albumin, choline esterase level, and platelet count[370]. There was a negative relationship between Tβ4 and aminotransferase aspartate aminotransferase and total bilirubin levels, as well as prothrombin time and Child-Pugh score[370].

1.5.7 Confounding factors for VGF in plasma and serum

VGF levels are elevated or decreased in certain conditions including bipolar disorder[164], major depressive disorder[164] and obesity[377]. Patients with major depressive disorder presented with significantly lower serum VGF levels compared to healthy controls[164]. In contrast, VGF levels were significantly higher in bipolar patient serum compared to controls[164]. VGF levels were highest in the plasma of obese individuals with type 2 diabetes, followed by obese (BMI >30) individuals without type 2 diabetes, and lowest in lean (BMI <25) individuals, suggesting that obesity and type 2 diabetes each contribute to an elevation of VGF level[377].

1.6 Artificial Intelligence and Healthcare

Artificial intelligence (AI) is a term used for a collection of established technologies called decision support technologies, knowledge-based systems, or expert systems. Early AI applications were focused on diagnosis and therapy recommendations in medical settings. Moreover, early types of AI used symbolic approaches which were based on rules and knowledge, while present-day AI implements statistical methods alongside symbolic approaches. Currently, there is increasing and renewed interest in AI utility whereby a new generation of clinical decision support is facilitated by the availability of powerful

computing tools to help manage and analyse big data to generate new knowledge[378]. Modern medicine is faced with the major challenge of acquiring, analysing and applying large amounts of information to solve complex clinical problems[379].

AI systems include artificial neural networks (ANNs), fuzzy expert systems, evolutionary computation and hybrid intelligent systems. ANNs are mainly concerned with learning, whereas fuzzy logic focuses on imprecision, and evolutionary computation is based on search and optimisation[379]. ANNs are computational analytical tools which are based on the biological nervous system. They consist of networks of highly interconnected computer processors (neurons) which can perform parallel computations for data processing and knowledge representation. Fuzzy logic was popularised by Lofti Zadeh in 1965[380]. It is a science of reasoning, thinking and inference which bases itself on everything being a matter of degree. Thus, instead of assuming everything is either black or white (conventional logic) fuzzy logic recognises that in reality most things would fall somewhere in between (i.e. in differing shades of grey). This method uses continuous set membership from 0 to 1 in contrast to using sharp distinctions i.e. 0 for false and 1 for true[379]. Evolutionary computation is a broad term for several computational techniques that are based on natural evolution, which incorporates the mechanisms of natural selection and survival of the fittest into solving real world problems. John Holland[381] developed the "Genetic algorithms", which are a class of stochastic search and optimisation algorithms based on natural biological evolution. The advantages of each of these technologies can be combined to produce hybrid intelligent systems, which work in a complementary synergistic manner[379]. Hybrid systems include ANNs for designing fuzzy systems, fuzzy systems that design ANNs, and Genetic algorithms for automatically training and generating ANN architectures. AI can be classified in different ways, but in its simplest form it can be divided into two main categories termed strong AI and weak AI[277]. Strong AI refers to a programmed machine that can take on human-level cognition with the machine having the ability to learn on its own, simultaneously conducting a number of complex tasks based on what is already known[382]. Currently, strong AI does not exist. However, weak AI does exist, and it is the process by which a machine is trained to complete a specific designated task. The machine works by simply acting upon and is bound by the rules and algorithms set for it. Machine

learning (ML) comes under the weak AI umbrella. Subdivisions of ML include supervised learning, unsupervised learning, semi-supervised learning and deep learning. Figure 1.3 illustrates the different types of AI.
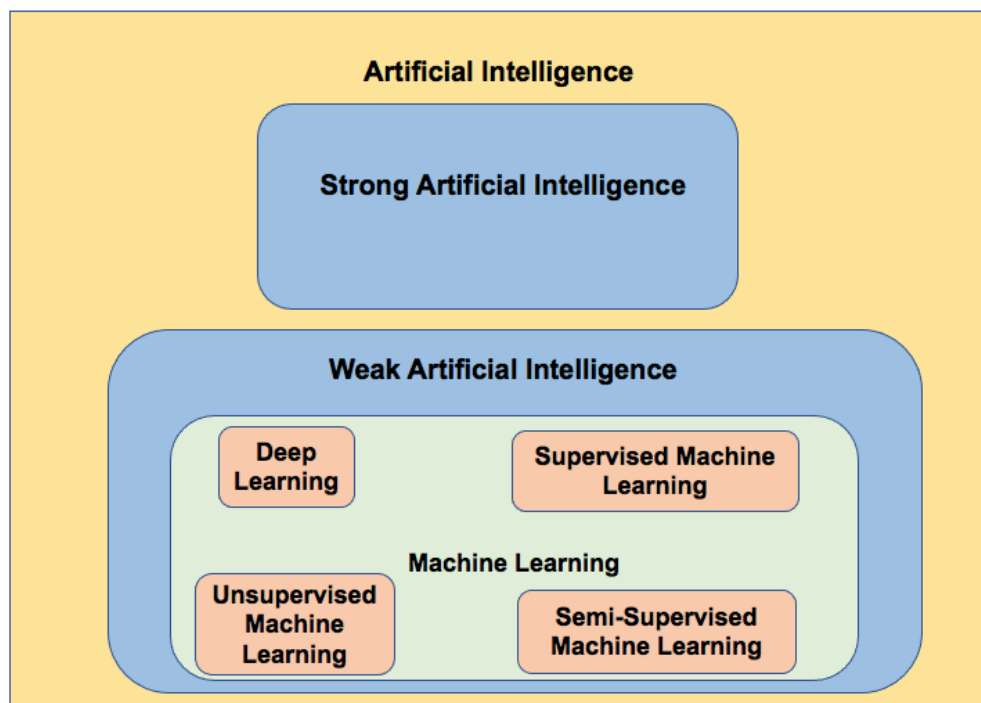
Figure 1.3: Different types of AI.

AI can be divided into strong and weak AI[382]. Strong AI is currently non-existent however weak AI encompasses ML, which in turn encompasses deep learning, as well as supervised, unsupervised and semi-supervised ML. Figure adapted from Du et al.,2020.

AI has the potential to transform healthcare. For example, AI is a tool that can aid clinicians in the decision-making process. With the generation of vast amounts of genomic, biomarker and phenotypic data across the healthcare system, AI is likely to play an increasingly central role in their utility[378]. The challenges of real-world implementation mean that little change to clinical practice from AI is likely expected in the next 10 years. However, change in 10 years' time is certain with massive changes in 20 years' time. Medical students starting today are likely by mid-career to be substantially impacted by this technology. Clinical specialities like radiology may not have disappeared by then, but are expected to be heavily transformed[383]. Healthcare AI techniques are emerging in imaging-based specialities such as radiology and gastroenterology[384]. Moreover, modern imaging modalities, which include endoscopic and cross-sectional imaging, generate a greater amount of visual

information than the human eye can distinguish. Thus, harnessing the technology that AI provides offers clear benefits to these specialities. Historically, the evaluation of AI was limited to the design and development phase, due to the rare implementation and use of AI systems within routine clinical practice. Moreover, during the design and development phase, evaluation focuses upon the performance of algorithms in terms of discrimination, accuracy and precision[378]. Depending on the use, one performance measure could be considered more important compared to another. For example, for a triage-based algorithm, a high level of discrimination is needed. However, an algorithm that predicts mortality or complication risks in shared decision making needs to be highly accurate and precise for all types of patients. Another key consideration is the decision-making performance of humans with and without AI assistance. Once an algorithm is developed, clinical validation of its utility is also needed. Wider considerations include how meaningful the algorithm is operationally, how it fits the clinical workflow, if it represents up to date clinical knowledge, if it will change clinical decisions, and the level of confidence that can be given. Another important consideration is that humans may become over reliant on decision support systems and delegate full responsibility to a decision support system rather than continue to be vigilant. This could have dangerous consequences if a decision support system is wrong or fails[378].

1.7 Machine Learning

Machine learning (ML) is the name given to both an academic discipline as well as a set of techniques that allow computers to undertake complex AI tasks. ML encompasses elements of mathematics, statistics and computer science and is an engine helping to drive advances within the development of AI. Two areas in particular that may benefit from the application of ML techniques in the medical field are diagnosis and outcome prediction. ML techniques themselves are based on algorithms, which are sets of mathematical procedures that describe the relationship between variables. Algorithms working in different ways depending on their type, and they can be combined synergistically in ML[385]. Statistical techniques and ML techniques have similarities and the boundary between the two is ill-defined. However, to understand the difference between

the two, the goal of each technique should be understood. ML is an extension of statistical approaches, as it has a learning component, with this learning component allowing for iterative learning. In addition to supervised, unsupervised, and semi-supervised methods, ML includes the area of deep learning, which originated from ANN research[386,382].

1.7.1 Unsupervised Machine Learning

Unsupervised learning does not involve a pre-defined outcome. Unsupervised techniques are exploratory and are used to find undefined patterns or clusters which occur within datasets[385]. The techniques used for unsupervised ML are often referred to as dimension reduction techniques and include principal component analysis[387] and latent Dirichlet analysis[388].

1.7.2 Supervised Machine Learning

In supervised learning a set of input variables such as blood metabolite or gene expression levels are utilised to predict a quantitative response variable for example hormone level or a qualitative response variable such as healthy versus diseased individual[389]. For supervised ML algorithms, a labelled training dataset is used to train the underlying algorithm[390]. The trained algorithm is then fed onto the unlabelled test dataset to categorise each sample[390]. Supervised learning algorithms suit classification and regression problems[390]. In classification problems the output variables are discrete, such as 'red' or 'black' or whether someone was 'diabetic' or 'non-diabetic'[390]. However, for regression problems, the corresponding output is a continuous value such as the risk of developing cardiovascular disease for an individual. The following supervised types of ML which will be used within this project are described - decision trees (Section 1.7.2.1), LR (Section 1.7.2.2), support vector machines (SVM, Section 1.7.2.3) and random forests (Section 1.7.2.4). Table 1.7.2.5 described the advantages and disadvantages of supervised ML models discussed in the subsections of Section 1.7.2.

1.7.2.1 Decision Trees

Decision trees are highly versatile and can be used for a wide range of classification problems. A key application of decision trees is to predict outcomes. Decision tree models include the classification and regression tree

(CART) [391], C5.0 tree[392], quick, unbiased, efficient, statistical tree (QUEST)[393] and chi-squared automatic interaction detection (CHAID)[394] tree. Binary decision trees were first developed at the Bell Laboratories in 1958[395]. The C5.0 decision tree originates from the C4.5 tree, which in turn originates from the iterative dichotomiser 3 (ID3) decision tree[396]. The C4.5 tree improved on the ID3 tree by taking information gain as the selecting criterion, which was a limitation of the latter. C4.5 trees also introduced new functions including pruning technology. C5.0 decision trees have the additional feature of boosting.
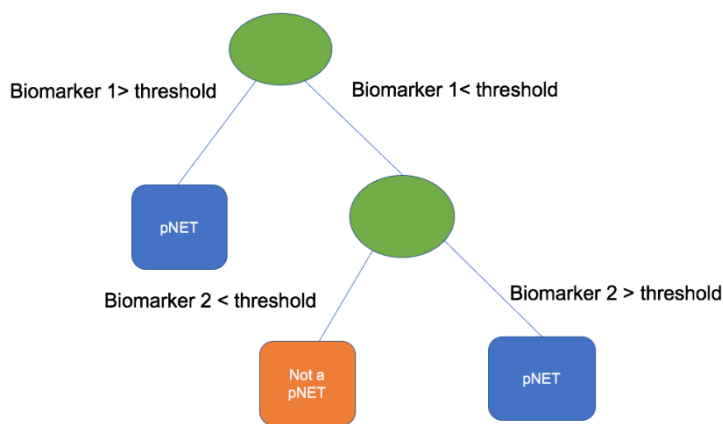
Figure 1.4: Illustration of how a decision tree would work for pNET diagnosis.

Green ovals denote splitting nodes and child nodes are shown in blue (pNET) and orange (control).

A decision tree has different components including nodes and branches. Figure 1.4 illustrates how a decision tree model would work for pNET diagnosis. There are 3 types of nodes, namely decision, internal and child nodes[397]. A root node is also a decision node and represents a choice that divides all samples into two (binary classification) mutually exclusive subsets. Internal nodes are also known as chance nodes and represent the possible choices at that point in a tree. The top edge of each chance node is connected to its parent node and the bottom edge is connected to its child or leaf nodes. Leaf nodes are also known as end nodes and represent the final result of the combined decisions or events. A tree also consists of branches, which connect nodes. A decision tree is formed using a hierarchy of such branches. Each path from the root node through the internal nodes to a leaf node represents a classification decision rule. Table 1.3 summarises the different decision tree approaches in Statistical Package for the Social Sciences (SPSS) Modeler software.

Decision trees differ in the approach taken for splitting. When building a decision tree, the most important input values classify the records at each root and early internal node into two categories based on the value of each corresponding nodal variable. Splitting options include using entropy, Gini index, classification error, information gain, gain ratio and twoing criteria[397,398]. This classifying or splitting procedure continues until stopping criteria or homogeneity are met[397]. In most cases not all of the potential input values are used to build a decision tree and, in some instances, a specific input variable may be used multiple times at different levels of a decision tree[397]. Stopping decision tree development is also important to consider, because complexity and robustness are competing characteristics that need to be considered[397]. A more complex model might fit the decision tree training data well but is likely to have a poorer ability to predict future samples due to overfitting[397]. Overfitting of decision trees needs to be controlled using stopping rules. Common parameters used in such stopping rules include the minimum number of records in a leaf, the minimum number of records in a node prior to splitting, and the distance or depth of any leaf from the root node[397]. Decision trees can also be pruned[398], which involves reducing the size of the tree in order to avoid unnecessary complexity and overfitting[398].

Advantages of decision trees include simplifying complex relationships between input and target variables, being relatively easy to understand and interpret, being non-parametric and without distributional assumptions, the ability to handle missing values, the ability to handle heavy skews without the need to transform data, and its robustness to outliers[397]. However, they also have disadvantages, which include being subject to overfitting and underfitting particularly with small datasets[397]. This can limit the generalisability and robustness of the resulting models[397].

	C&R Tree	QUEST	Random Trees	CHAID	C5.0	Trees-AS
Input fields	Continuous Categorical	Continuous	Continuous Categorical	Continuous Categorical	Any type	Continuous Categorical
Target fields **(Only one field can be specified**)	Continuous Categorical Flag Nominal Ordinal	Categorical Flag Nominal	Continuous Categorical Flag Nominal Ordinal	Continuous Categorical Flag Nominal Ordinal	Flag Nominal Ordinal	Continuous Categorical Flag Nominal Ordinal

Type of split	Binary split	Binary split	Binary split	2 or more	2 or more	2 or more
Method used for splitting	For a Categorical output a dispersion measure is used e.g. (Gini coefficient)	Chi-square for categorical inputs ANOVA for continuous inputs	Same approach as C&R tree	Chi-square test	Information theory measure is used, the information gain ratio	Chi-square test
Missing value handling	Substitute prediction fields where needed	Substitute prediction fields where needed	Can handle missing values	Missing values assigned a new category	Fractioning method	Can handle missing values
Pruning	Option to grow the tree fully and then prune back	Option to grow the tree fully and then prune back	Allows control of the minimum subgroup size. However, no pruning	Allows control of the minimum subgroup size	Option to grow the tree fully and then prune back	Allows control of the minimum subgroup size
Interactive tree building	Available	Available	Not Available	Available	Not Available	Not Available
Prior probabilities	Available	Available	Not Available	Not Available	Not Available	Not Available
Rule sets	Available	Available	Not Available	Available	Available	Not available

Table 1.3: Decision tree options available in SPSS Modeler.
The main parameters available for each decision tree type are shown[399,400]

1.7.2.2 Logistic Regression

When considering a problem in which a dependent variable is categorical, a common approach is to use logistic regression (LR), which takes its name from the type of curve used for data fitting[401]. Categorical variables are used in biomedicine for discrete parameters, such as whether a drug was administered, or if a patient survived[401]. Categorical variables may have more than two values. LR with two dependent categorical variables is known as binary LR, while LR involving more than 2 dependent categorical variables is known as multivariate LR. Two advantages of using LR is that it is relatively simple to perform and easy to interpret[401].

In LR the probability of event Y (P(Y)) occurring is predicted, given known values of different factors (x_n, vector-containing predictors). The general form of the functional dependence given by this regression is through this formula[402].

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + .. b_N X_N)}} = \frac{1}{1 + e^{-(\underline{b}^T . \underline{X})}}$$

P(Y) can also be defined as the probability of Y belonging to a certain class and b_n are coefficients that are determined within a LR analysis[402]. A more compact version of the formula is shown on the right hand side of the second equals sign and is used where $\underline{b}$ is the coefficients vector and $\underline{X}$ is the predictor vector having 1 as the first value[402].

Different approaches can be adopted within an LR analysis, including the enter, forward stepwise and backward stepwise methods. In the enter method all the input variables are entered simultaneously[1]. In a "forward stepwise" regression analysis various factors are introduced one by one, beginning with the strongest and stopping when addition of another factor does not significantly improve the prediction. In a backward stepwise approach all the factors are initially introduced and then individual factors are withdrawn until the overall prediction does not deteriorate[403].

LR has been used in cancer biomarker algorithm development for various cancers including breast[404], gastric[405], ovarian[406] and pancreatic cancers[407]. The approach taken to optimising a serum biomarker algorithm for ovarian cancer was to compare different methodologies for algorithm development including LR, k-nearest neighbour and linear discriminant analysis (LDA) [406]. In this study the LR algorithm produced the best performance based on a combination of 4 biomarkers (Human epididymis protein 4(HE-4), Platelet-derived growth factor AA (PDGF-AA), prolactin and Transthyretin (TTR)). LR produced an AUC of 0.95 compared to 0.94 using a k-nearest algorithm and 0.92 using linear discriminant analysis[406]. Moreover, a LR model has been incorporated in to a PancRisk score for the improved earlier detection of PDAC utilising urine biomarkers[407].

1.7.2.3 Support Vector Machines

The SVM algorithm was developed by Vapnik and Cortes in 1995[408]. Its characteristics include the algorithm's ability for good generality under the principle of structural risk minimisation, being able to deal with non-linear problems though kernel methods, the ability to deal with noise through the introduction of slack variables, and producing limited solutions as the optimal hyperplane depends only on the support vectors and guarantees convergence[409].

SVMs identify the hyperplane that separates a set of data points into two classes. This is achieved by maximising the marginal distances (the distance between the decision hyperplane and the data points closest to the boundary)[410] to produce the maximum margin hyperplane[411]. The samples closest to hyperplane boundaries are known as the support vectors[412]. The aim of a SVM is for the resulting classifier to achieve considerable generalisability and to be utilisable for the reliable classification of new samples[410].

In reality it is not straightforward to separate data linearly, as noise is a phenomenon of real-world data which blurs linear boundaries within classification problems[411]. A solution is to relax the constraint of maximising the margins of a linear separator[411], resulting in a soft margin classifier[413].

For a soft margin classifier, additional coefficients such as epsilon (ε) are utilised known and is known as a slack variable. The soft margin approach can work for data that are close to being linearly separable, however it returns poor results when inputs are strongly influenced by errors and other sources of variation[411]. The C parameter, known as the penalty parameter is important as it controls the extent of misclassification[414]. If C is low, this means that that the boundary which is chosen has a larger margin, meaning a potentially greater number of misclassifications, resulting in a model which is underfitted[414]. However, a higher C means that the boundary chosen has a smaller margin, with a smaller number of misclassifications and thus potentially resulting in overfitted models[414]. Thus, finding a balance between an overfitted SVM and an underfitted SVM model is key.

SVMs map input vectors into a high dimensional feature space Z via non-linear mapping[408]. Within the feature space, a linear decision is constructed to ensure a high generalisability of the network[408]. The kernel feature of SVMs allow for linear and non-linear data to be separated linearly. Different kernel functions

include linear, polynomial, sigmoid and the Radial Basis Function (RBF) which is also known as a Gaussian kernel[414]. The performance of a SVM is dependent on the choice of the kernel[415]. Table 1.4 summarises the different types of kernel with their associated formulas and parameters.

The most common type is the linear kernel, which can be used directly when samples are separable in a low dimensional space. Unlike other kernels, there are no parameters for the linear kernel function.

The polynomial kernel is a global learner with a poor local performance. The parameters in this kernel include 'd', which represents the dimensionality of the kernel function. The dimensionality of the mapping function grows with the value of d, and the higher the dimensionality the easier the classification, but computational complexity increases at the same time. Complex classifiers have a good performance with a high classification accuracy at the training data level but are prone to overfitting[414]. There are 3 parameters, gamma, coef and d for the polynomial kernel. The default values are gamma set to 1/k (k is the number of classes), coef set to 0, and the polynomial degree set to 3[414]. The polynomial kernel is a non-stationary kernel and it is particularly suited to problems where all the training samples are normalised[412]. This kernel non-linearly maps samples into a higher dimensional space and so, unlike the linear kernel, this can deal with attributes that are non-linear.

The ability for the algorithm to generalise decreases with the increase of the adjustable gamma parameter[414], which specifies the radius of the RBF[416]. As it plays an important role in the performance of the RBF kernel it should be carefully tuned. The gamma parameter defines how far the influence of a single training data set reaches, meaning that low values of gamma have a far reach[417]. Low gamma value also imply that points need not be close together to be considered the same class. Conversely, larger gamma values imply that points should be closer together to be co-classified. Moreover, when the gamma is very small, the model is too constrained and cannot capture the complexity and the shape of the data[417]. When the gamma value is small it has less influence, and the resulting model behaves similarly to a linear model. Larger gamma value models tend to overfit whereas smaller gamma value models tend to underfit.

Applications of the sigmoid kernel are rare[414]. The sigmoid kernel satisfies Mercer's theorem[412] and the underlying kernel function is the equivalent of a

two-layer neural network[414]. Thus, the sigmoid kernel is widely used in the context of neural networks [414]. The gamma and r values must be carefully chosen to avoid poor and potentially worse than random performance of the resulting SVM model[412]. Gamma is a scaling parameter for the input samples and r is a shifting parameter that controls the mapping threshold.

Type of SVM	Formula	Parameters
Linear	$K(x,x_i)= x.x_i$	None
Polynomial	$K(x,x_i)=[y*(x.x_i)+ coef]^d$	y, coef, d
RBF	$K(x,x_i)=exp(-y*\|\|x-x_i\|\|^2)$	y
Sigmoid	$K(x,x_i)=tanh(y(x.x_i)+coef)$	y, coef

Table 1.4: Different SVM kernels with their associated formulas and kernel parameters.
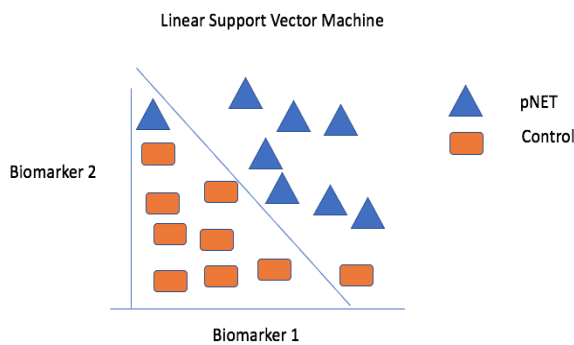
Figure 1.5: Illustration of how a Linear SVM would work for pNET diagnosis.
The level of Biomarkers 1 and 2 are shown on the x and y axis, respectively. Orange rectangles and blue triangles represent control and case samples, respectively. A straight line defined by shows the separation using Linear SVM for the cases and controls.

SVMs are supervised ML algorithms which have been widely employed for their discriminative ability in many different disciplines, including areas of oncology[418,419], cardiology[420,421], respiratory medicine[422,423] and neurology[424,425]. The utility of these algorithms has been particularly evident in improving cancer detection[426,427,428]. For example, a study by Tong et al. found that an SVM algorithm could differentiate the samples in their validation set with an accuracy of 86.1%, which was higher than the accuracy obtained using RF (82.5%) or LR (78.7%) algorithms[426]. Figure 1.5 illustrates how a Linear SVM model could be used to discriminate between pNETs and controls based on two biomarker assays.

1.7.2.4 Random Forests

In essence a random forest (RF) is a collection of a type of decision tree known as CART that have been constructed using bootstrap datasets with a random subset of predictor variables for modelling outcomes. Thus, RF is in effect an

ensemble of multiple decision trees. The CART is user friendly and produces a visual representation that can be read in a similar way to a flow chart to predict outcomes[429]. CARTs are constructed using binary splits of a dataset[429]. CARTs can be simple or detailed in nature, depending on the number of predictor variables included in a CART model, outcome categories and stopping criteria. More detailed trees are commonly employed in RF models[429]. RFs work by iteratively developing decision trees to be used in the context of classification or regression problems[429]. The RF ML method was developed in 2001 by Leo Breiman et al.[430]. A RF is built via multiple steps based on a dataset divided into two groups termed the in-bag (training) set and the out-of-bag (validation) set[429]. The in-bag dataset is used to grow the CART forest by randomly selecting a subset of predictor variables to make binary splits at each tree using the best features[429]. Each tree is grown until the stopping criterion has been met. This process is repeated until the specified number of trees has been made[429]. The out-of-bag dataset is then run down all the trees in the forest and the results from each tree are compiled to give a prediction for each sample[429]. Decision trees have drawbacks, particularly in providing poor accuracy for complex datasets (large datasets as well as datasets with complex variable interactions)[431]. RFs can however address the pitfalls of decision trees. RFs share some of the beneficial properties of decision trees, such as the ability to interpret relationships between predictors and an outcome[429] but they can provide higher accuracy than a single decision tree model. RF algorithms are focused on result prediction as opposed to production of an explainable model. Despite the absence of a well-defined model, they have been increasingly used as a standard analytical tool[432]. The price for accurate classification via a RF is reduced user understanding and interpretability of results[429]. This latter is a key disadvantage of the RF model, which is often described as 'black box' because it does not provide users with a describable model due to the complexity of the algorithm[429] and because gaining an insight into the RF prediction rule is difficult due to the typically large number of trees[432]. Initially their usage was limited to innovation-friendly scientists or ML experts[432], but RFs have since become increasingly well known by various non-computational communities[432]. Bagging is employed by RFs to enhance accuracy when random features are used to give estimates of the generalisation error of the combined ensemble trees[430]. Moreover, the out-of-bag error estimate removes the need for an

additional test set[430]. Additionally, cross validation (CV) entails bias of an unknown extent, while out-of-bag estimates are unbiased[430].

A major advantage of RFs is their ability to deal with datasets that consist of large numbers of predictor variables. However, in practice the number of predictors required to obtain the outcome predictions are limited to improve efficiency[431]. RFs differ from traditional statistical techniques whereby classical assumptions are not necessary hence RF approach often has much lower prediction error rates compared to traditional models, which makes it ideal for classification problems[429]. A final advantage of RFs is that they can be employed situations in which the number of predictor variables is much bigger than the sample size unlike other models[429].

Supervised ML Approach	Advantages	Disadvantages
LR	• Robustness • High precision • Works well with small sizes • Interpretable model • Simplicity	• Problems with high co-linearity • Gaussian distributed residuals • Not the most powerful algorithm. Can be outperformed.
SVM	• Robust • High precision	• Slow in training • Blackbox approach • Sensitive with respect to kernel choice
Decision Tree	• Easy to interpret • Deals well with outliers • Works with many datatypes • Fast predictive speed • Variable distribution not important • Boosting option available	• Overfitting • Slow in training
Random Forest	• Feature of importance • Works well with large datasets • Strong performance	• Blackbox approach • Slow in training • Overfitting • An ensemble approach is less

		interpretable than a single decision tree

Table 1.5: Advantages and disadvantages of different supervised ML approaches.

1.7.2.6 Validation of supervised machine learning models

A reliable way to validate the performance of a supervised ML model is by assessing it's classification performance using a separate dataset to that used to train the model. However, this approach requires a large amount of data which is often not available, for example due to the high associated costs or limited data availability. Another approach is known as the Train and Test approach in which a portion of the dataset is allocated to training and another portion to validation. Using unseen data to test a ML model provides an unbiased estimate of what the performance would be when the model is deployed for actual predictions in real-world situations. Cross validation (CV) approaches address situations in which the availability of datasets is limited. Instead of training a fixed model only once via a Train/Test split, in CV several models are created iteratively on different portions of the data. K-fold is a common CV approach in which a portion of the data is separated for validation, leaving the rest to train a model and predict class outcomes. This is then repeated several times by leaving out a different portion of the data for the validation until all of the data is used. A model's performance is then calculated as a mean of classification performances in each of the validation folds. The advantage of a k fold approach is that is economical as it allows utilisation of all the data for training and reuse for validation. If the validation were to be performed with a separate dataset, double the amount of data would be needed to have the same amount for training and validation. Another approach known as nested CV avoids pooling the training and validation data but is economical as it allows all the data to be used for training and reused for validation. With nested CV a portion of data is split and in each fold a model is developed on the reduced training set from scratch including feature selection and parameter tuning. This is repeated after splitting a different portion of the data for validation and each time a new model is developed for training until all the data is used[433].

1.7.2.7 Evaluating supervised machine learning models

1.7.2.7.1 Receiver operating characteristic and Area under the curve

A Receiver operating characteristic (ROC) curve is used for the visualisation, organisation and selection of classifiers based on their performance. ROC curves are 2D graphs which have the true positive rate plotted on the Y axis and the false positive rate plotted on the x axis. Thus, a ROC curve depicts the relative trade-off between benefit (true positives) and cost (false positives). ROC curves are compared via calculation of corresponding AUC values. The area under the curve (AUC) is the area under a curve as a proportion of a unit square defined by the both axes ranging from 0 and 1. However because random guessing produces a diagonal line between (0,0) and (1,1) with an area of 0.5, no effective classifier should have an AUC of less than 0.5[434].

1.7.2.7.2 Confusion matrix

In the field of ML a confusion matrix is also known as an error or contingency matrix. A confusion matrix has 4 different components, namely true positives (TP), which are the positive cases that the classifier has correctly identified; true negatives (TN), which are negative cases that the classifier has correctly identified; false positives (FP), which are negative cases that the classifier has incorrectly identified; and false negatives (FN), which are the positive cases that the classifier has incorrectly identified.

The following parameters are based on the confusion matrix and can be used to evaluate different classification models. These include accuracy, F1 score, precision, sensitivity (TP rate), specificity and FP rate[390].

1.7.2.7.3 McNemar test

The McNemar test can be employed to compare the performance of ML models. This test is a 2x2 cross classification of paired (or matched) responses to a dichotomous item. A McNemar test is a type of chi-squared test that uses dependent (i.e. correlated or paired) data as opposed to independent (unrelated) samples. It is non-parametric and thus distribution free, and can therefore be used with sample data that are not normally distributed[435].

1.8 Cancer Screening Programmes in the UK

Once a biomarker test and algorithm are validated using well characterised patient and healthy cohorts, they may be applied to undiagnosed populations. Cancer screening at a population level is seen for a few types of cancer where there is a suitable test and justification based on a cost-benefit assessment that typically requires a high prevalence such as that found in the UK for breast, cervical and bowel cancer. The screening approaches for these cancers and the associated issues of FN and FP results are discussed in subsection 1.8.1. However, as pNETs have a much lower population prevalence, use of a pNET test could only be justified for a smaller, high-risk group. Targeted screening programmes for cancers with lower prevalence such as pancreatic cancer are discussed in subsection 1.8.2.

1.8.1 Population cancer screening programmes

There were 367,167 new cases of cancer between 2015 and 2017 in the UK[436]. However, the National Health Service (NHS) has only three established population screening programs - for breast, cervical and bowel cancer[437]. Breast cancer is the most common type of cancer in the UK, accounting for 15% of all new cancer cases (2016-2018)[438]. Moreover, between 2016 and 2018 there were 11,547 deaths from breast cancer in the UK, highlighting why screening for breast cancer is important in order to improve outcomes by detecting disease earlier[438]. The UK NHS breast cancer screening program is available to women aged between 50 and 70 as well as some trans and non-binary people, and these individuals are invited to partake in screening via mammogram analysis every three years[439]. However, women aged over 70 can still have screening mammograms even though they don't fall within the recommended screening programme age range. Cancer is found in 9 out of every 1000 women that undergo screening[439]. Screening for breast cancer is also adapted to reflect the level of risk in different age groups. Women with an increased risk of breast cancer due to a family history of the condition who are younger than 40 years are offered yearly MRI scans from the age of 30 or 40 dependent on level of risk[439]. For patients who have been shown to have a gene mutation through testing that increases the risk of breast cancer the recommendations are slightly different with UK guidelines recommend yearly MRI scans from the age of 20 for women who have a TP53 mutation and age

30 for breast cancer gene 1 (BRCA1) or breast cancer gene 2 (BRCA2) mutations[439].

The UK NHS bowel cancer screening programme is organised separately by each nation[440]. The programme in England invites people aged between 60 and 74 years to be screened every 2 years[440] using a faecal immunochemical test (FIT), which is an immunoassay that measures human haemoglobin in faeces in terms of micrograms of blood per gram of faeces (μg/g)[441]. People over the age of 74 can request a screening kit from the bowel cancer screening program[440]. Dependent on the results of the FIT, patients may be referred for further investigation and diagnosis[440].

The NHS screening program for cervical cancer targets women aged 25 to 64[442]. Women aged between 25 and 49 are invited every 3 years, and those aged over 49 are invited every 5 years until the age of 64[442]. Cervical screening is also available for trans men and non-binary people within this age range who have a cervix[442]. GPs can arrange cervical screening for patients aged over 64[442]. Cervical screening in England, Wales and Scotland involves testing cervical cells for human papilloma virus (HPV)[442].

When participants consider breast, bowel or cervical cancer screening, it is important for them to understand that while screening is widely beneficial, there are also potential harms of screening, such as FN or FP results[439]. A FN result means a screening test has failed to identify a cancer in a patient and hence the opportunity to promptly treat that cancer has been missed[439]. In contrast, a FP result means a screening test has incorrectly detected the presence of cancer[439]. A FP result can lead to patient anxiety and further testing such as a biopsy assessment[439].

When considering breast screening there are also issues of overdiagnosis and overtreatment[439]. Overdiagnosis could lead to identifying breast cancers that won't ever cause any problems[439]. For example, breast screening can pick up ductal carcinoma in situ (DCIS)[439] but it's not possible to know whether DCIS will develop into cancer. Hence many DCIS patients have surgery, radiotherapy and hormone therapy, potentially unnecessarily[439].

The risks in bowel cancer screening also include overdiagnosis and unnecessary treatment, in this case for bowel growths that may never cause any harm[440]. Some people with an increased risk of bowel cancer may undergo screening earlier than the normal schedule of the NHS bowel screening

program[440]. These patients include patients with conditions such as familial adenomatous polyposis, lynch syndrome, a family history of bowel cancer, ulcerative colitis, Crohn's, polyps in the bowel, and previous history of bowel cancer[440].

Despite cervical cancer screening saving thousands of lives every year in the UK, similar potential harms include FN and FP results, overdiagnosis and overtreatment[442]. Treatment is offered to everyone with above a threshold count of abnormal cells as it is impossible to know whether they will go on to develop cancer or not[442]. Treatment can lead to problems such as bleeding or infection[442]. Moreover, removal of cervical tissue can elevate the risk of premature birth in subsequent pregnancies[442].

1.8.2 Targeted cancer screening for pancreatic cancer

There is no population screening programme for pancreatic cancer. The European Registry of Hereditary Pancreatitis and Familial Pancreatic Cancer (EUROPAC) programme exists to identify and monitor high-risk patients[443]. Inclusion in this study is for patients who have 2 or more first degree relatives with pancreatic cancer, patients who have 2 or more relatives with pancreatic cancer on the same side of the family, or if the patient's family has a known familial syndrome and at least one family member having pancreatic cancer and a faulty gene[443]. Participation in the EUROPAC study involves a CT scan, EUS, and a blood test to check CA-19-9 and sugar levels,[443] and screening usually starts from the age of 40[443].

There is no similar screening programme for pNETs due to their far lower prevalence. Patients with MEN1 are offered yearly screening to check for early signs of tumour development,[61] but only a small portion of pNETs are familial. Thus, the prompt identification of patients with sporadic pNETs is an unmet clinical need. The characterisation of any particular combinations of symptoms and comorbidities associated with pNETs would allow primary care physicians and MDC specialists to identify high-risk patients with individually non-specific symptoms and establish points of care in which a biomarker test for pNETs could justifiably be used.

1.9 Summary

Despite substantial advances being made for several more prevalent cancers, the development of novel biomarkers for NETs has not been as successful[444]. Pharmacoeconomics, and the heterogenous nature of NETs with respect to site, secretory product, proliferative control and molecular genetic changes, make it challenging to identify a single panel of pNET markers[444]. The delay in pNET diagnosis is partly due to the symptomatic presentation of these tumours being non-specific in nature, but it is also due to the absence of a suitable test. One reason for the latter is a deficit in the identification of early diagnostic biomarkers[445]. Moreover, the development of robust circulating pNET biomarkers for the diagnosis and prediction of disease progression would greatly facilitate early tumour detection and targeted management[446,447]. The current gold standard marker used clinically is CgA, which has important drawbacks to consider, including multiple confounding factors such as the use of proton pump inhibitors[448] and renal failure[449]. There is also the important drawback that CgA is not a reliable diagnostic biomarker for insulinomas[90]. Recent efforts have shifted toward a multianalyte approach with the rationale that targeting a greater number of markers as opposed to a single analyte approach would improve the performance of test sensitivity and specificity. Pancreatic cancers present challenges to accurately diagnose due to the presence of fibrosis and cystic disease that can confound imaging and clinical interpretation[450]. Thus, developing a biomarker panel for the identification of pNETs which is both cost-effective and has high accuracy is an unmet clinical need.

1.10 Aims and Hypotheses

1.10.1 Hypotheses

- Earlier detection of pNETs via the usage of a biomarker panel and a corresponding algorithm will lead to a higher proportion of patients being diagnosed at a stage at which surgery is possible and, consequently, an improved 5-year survival rate.
- The biomarker panel will be used a part of a triage pathway in primary care for patients suspected of having a pNET based on their presentation with a characteristic combination of symptoms .

1.10.2 Aims

This study aims to

- Develop and validate a diagnostic blood biomarker panel that can be used to detect patients with pNETs and to differentiate pNETs from benign, non-cancer pNETs and lower-risk pancreatic lesions.

- Detect pNETs at a greater sensitivity and specificity than chromogranin A (CgA) the current gold standard test used in clinical practice.

Chapter 2 – Methods

2.1 General summary of ML model creation using SPSS Modeler

An SPSS Statistics data file, containing the biomarker data and whether a patient had a pNET or not was used for the creation of ML models. In SPSS Statistics (IBM), the NET column contained "0" and "1" values and was set to be a numeric data. The "0" and "1" values were used to represent control and case, respectively. The columns of biomarker data were also set to be numeric data. Creation of models in SPSS Modeler (IBM) employed the following general method. First, the SPSS Statistics file containing the biomarker data was dragged into the area in SPSS Modeler known as the stream, where models were created. The SPSS Statistics file was then connected to the Type node, located under the Field Ops or Favourites tabs. Within the Type node, the NET field and the different biomarker fields could then be seen. The measurement option for the NET field was changed to "flag" and the role option to "target". The Type node was then connected to the ML model algorithm of interest located in the Modelling tab. The algorithms employed were SVM (SVM node), LR (Logistic Node), Random Forest (RF node) and C5.0 decision tree (C5.0 node). Within the Modelling nodes, parameters for the models were defined and the models were run. Once a model had been run, a gold diamond nugget was produced that contained the model information. Each nugget could be connected to other nodes to obtain further information for the corresponding created model. Additional nodes included the Table node, which provided prediction information, the Analysis node, which provided the AUC and confusion matrix, and the Evaluation Node for obtaining ROC curves. Section 2.2 describes the methodology for the creation of LR models for the whole training dataset in SPSS Modeler. Section 2.3 provides the methodology for the creation of C5.0 decision trees in SPSS Modeler for the whole training dataset. Section 2.4.1 provides the methodology for the creation of SVM models in SPSS Modeler. Finally, Section 2.5 provides the methodology for the creation of RF models in SPSS Modeler for the whole training dataset.

2.2 Creation of Logistic Regression models from the UOL training dataset in SPSS Modeler

2.2.1 Generating the Logistic Regression model

SPSS Modeler 18.2.1 software was used to create LR models for the University of Liverpool (UOL) pNET and control data. The record containing the missing ANG2 value was first removed from the analysis using SPSS Statistics Version 27, resulting in a dataset that contained 49 controls and 57 cases of the starting 50 controls and 57 cases. The SPSS Statistics file containing the levels of seven biomarkers was used to generate the results for all 127 possible biomarker combinations. For model generation this file was dragged into the blank canvas of SPSS Modeler to create a node. The Type node was then dragged into the same canvas and these two nodes were linked using the Connect option. Clicking on the Type node revealed the NET field and the different biomarker fields. The Measurement option for the NET field was changed to "flag" and the Role option to "target". The Preview option allowed a snapshot of the data to be viewed.

From the node palette, the Logistic option was selected to carry out LR modelling in SPSS Modeler. The Logistic node was dragged into the canvas and a connection formed between the Type node and the Logistic node. Within the Logistic node, under the Fields option, "Custom Field Assignments" was selected with target and inputs added. The Target was NET and inputs were the different combinations of biomarker data. Below the model options, the "Use partitioned data" and the "Build model for each split" options were unselected. The Procedure option chosen was "Binomial". Beneath the Binomial Procedure the method used was the "Enter" method. Underneath the Expert setting, the "Simple mode" (default) was kept and in the Analyse option the "Calculate predictor importance" option was selected, before each LR model was generated. Model generation was indicated on the canvas by a gold diamond-shaped icon known as the "model nugget for the LR model". Clicking on the Logistic model gold nugget provided further information on the created model. This included the equations for the LR analysis and predictor importance information for the different markers. A summary table was also shown which provided the characteristics of the LR model and the settings used to build the model.

2.2.2 Performance Evaluation

For each LR model generated as described in the previous section the Logistic node was connected to the Analysis node located in the Output tab. Within the Analysis node the coincidence matrices (for symbolic targets) and evaluation metric (AUC and Gini, binary classifiers only) were then selected. Run was then clicked to provide the AUC value and confusion matrix for each model.

2.2.3 ROC curves

ROC curves were also obtained for each LR model by selecting the Graphs option within the Evaluation node. The gold diamond Logistic model node containing the LR model was then connected to the Evaluation node. Clicking on the Evaluation node provided further options and in the Plot setting the Chart Type was changed from the default to ROC, with all other settings kept at default. Run was then clicked to generate the interactive ROC curve.

2.3 Creation of non-boosted C5.0 decision trees from the UOL training dataset in SPSS Modeler

C5.0 non-boosted trees were created in SPSS Modeler. Section 2.3.1 explains the process for generating a C5.0 decision tree model, Section 2.3.2 describes the methodology for performance evaluation, and Section 2.3.3 explains the methodology for C5.0 decision tree ROC curve analysis.

2.3.1 Generating a C5.0 decision tree model

A SPSS file containing the UOL biomarker data was saved in SPSS Statistics Version 25 and used by SPSS Modeler 18.2.1 to create C5.0 decision tree models on the stream canvas. This file used had a pre-defined missing ANG2 value already specified in the SPSS Statistics file.

Specifically, a blank canvas was opened and streams were created by linking nodes as previously explained. The SPSS Modeler stream represented the data flow through each operation and the tools required to build the model were located in the Nodes palette.

The SPSS Statistics file containing the biomarker data for the 50 controls and 57 cases was dragged into the canvas to create a File node, which was then connected to a Type node that had been dragged from the Node palette in the

Field ops section as previously described. The Type node fields included the NET and biomarker data. The measurement for NET was changed to "flag" and the role changed to "target". The asterix (*) next to ANG2 denoted a missing value, which had been defined previously in the SPSS Statistics file and did not need to be redefined.

From the node palette the modelling option was selected to reveal the different modelling options available. The C5.0 node was selected and dragged into the canvas and connected to the Type node as described previously. Clicking on the C5.0 node provided further information and parameters required to build the C5.0 models.

Selecting the C5.0 node revealed five tabs, including the Fields and Models tabs. In the Fields tab, "Use custom field assignments" was selected with target and inputs added. The target was NET and inputs were the biomarker combination used to generate each model. In the Model tab, further parameters for the decision tree were selected according to on the type of C5.0 decision tree to be built. For all trees, "Use partitioned data" and "Build model for each split" were deselected, as was "Group symbolics", and the Output Type was set to "Decision tree". The default "Expected noise" setting of 0 was used for all trees. The model options containing simple and expert options was set to simple and contained the Favor accuracy (default) or Favor generality. Accuracy or generality were selected for different models. Relevant option boxes were ticked or unticked according to the type of decision tree being created (cross validated, non-cross validated, boosted or non-boosted). "Calculate predictor importance" was selected in the Analyse tab.

Table 2.1 summarises the settings employed to create different types of C5.0 decision trees.

	C5.0 Decision tree Non boosted Non cross validated Generalise setting	C5.0 Decision tree Non boosted Non cross validated Accuracy setting
Output type	Decision Tree	Decision Tree
Partitioning data	None	None
Boosting	No boosting	No boosting
Boosting settings	N/A	N/A
Cross Validating	None	None
Cross Validating settings	N/A	N/A
Mode	Simple	Simple
Favour	Generalise	Accuracy
Expected noise (%)	0	0

Table 2.1: Summary table for the settings used to model different types of C5.0 decision trees in SPSS Modeler 18.2.1.
C5.0 decision trees include non-boosted, non-cross-validated general trees; non-boosted, non-cross-validated accuracy trees; and boosted, cross-validated general trees (described in Section 2.8).

Once the model was successfully generated, a gold diamond-shaped "model nugget" appeared, which contained information on the model. On the Model tab, thresholds and predictor of importance information were shown, and the Viewer button provided a visual representation of the decision tree.

The Summary tab provided further details of each decision tree, including the tree depth, fields used, and inputs used in that model. Summaries of the build settings and model training were also available. The Settings tab provided additional information, including "calculate confidences" and "generate SQL of this model".

2.3.2 Performance Evaluation

The AUC value and Gini score (for non-boosted trees) for C5.0 decision tree models were obtained to evaluate performance. For each model, the Analysis node in the Output tab was selected, dragged onto the canvas, and connected to the C5.0 model nugget. The default settings for the Analysis node included "Evaluation metric" (AUC and Gini binary classifiers only) selection. Run was then selected to provide the AUC value and Gini score for that model.

2.3.3 ROC curves

To generate a ROC curve, from the Nodes pallet, the Graphs tab was selected and the Evaluation node dragged onto the canvas and then connected to the gold diamond C5.0 model node. In the Plot settings the "Chart type" was changed from the default to ROC with other settings kept at default. Run was then clicked to generate a ROC curve.

2.4 Creation of Support Vector Machine (SVM) models from the UOL training dataset using SPSS Modeler and the MATLAB Classification Learner App

2.4.1 SVM model generation from the UOL training set using SPSS Modeler 18.2.1

Linear and RBF SVM models were created from the whole UOL training dataset using SPSS Modeler. Section 2.4.1.1 explains the pre-processing and scaling of the UOL data that was carried out prior to Linear and RBF SVM model generation, while Sections 2.4.1.2 and 2.4.1.3 describe the process of Linear and RBF SVM generation, respectively, using SPSS Modeler.

2.4.1.1 Data pre-processing and scaling prior to SPSS Modeler entry

Prior to data analysis using SPSS Modeler, the UOL biomarker data (n=107) was pre-processed by removing a missing value record (n=1) and scaled in Microsoft Excel. The resulting dataset contained 49 controls and 57 cases (n=106) and scaling used the min-max process, which scaled the data using the equation:

$$x_{normalised} = (x - minimum(x)) / (maximum(x) - minimum(x))$$

Data scaling was required for SVM modelling as this is a distance-based algorithm.

2.4.1.2 Linear and RBF SVM model creation using SPSS Modeler

SPSS Statistics Version 27 was used to create the data input file for SPSS Modeler 18.2.1 from the min-max scaled data.

To create each SVM model the SPSS Statistics data file containing the scaled data was dragged into the stream area of SPSS Modeler, and the resulting File node was connected to the Type node. Double clicking the Type node allowed the fields in the model to be seen. The NET field in the "measurement" option was changed to "flag" and the "role" for the NET field changed to "target". Clicking the preview button allowed the dataset to be previewed. The Type node was then connected to the SVM node, which was located in the Modelling tab. Double clicking on the SVM node showed the type of SVM to create (linear, RBF, sigmoid or polynomial), to adjust parameters, and to run the model. Within the SVM node were further subtabs including Field, Model, Expert and Analyse. The Field subtab was kept on the default setting "use pre-defined roles". In

Model subtab, "use partitioned data" and "build model for each split" were both deselected. The Expert subtab allowed the parameters of the model to be adjusted, including the C parameter (also known as the regularisation parameter) for Linear SVM models, and the kernel type (gamma) for RBF SVM models. For Liner SVM models, C parameter values from 1 to 10 were employed, as recommended by the manufacturer. For RBF SVM models, C and gamma parameters were explored to compare the performance of the resulting models using a grid approach. C parameter values were chosen between 1 and 10 and the gamma value ranged from 0.4 to 0.95. In the Analyse tab, "calculate predictor importance" was selected to assess the contribution of each marker to a model.

Models were then created, resulting in the generation of a SVM model nugget, which contained model parameter information and predictor of importance data. Further information for a model was obtained by connecting the model nugget to other nodes. These included the Analysis tab to obtain the AUC value and confusion matrix information; the Evaluation tab (plot) to get a ROC curve; and Table to see the predictions for individual case and control samples.

2.4.2 SVM model generation for the UOL training set using MATLAB

Optimised Linear and Optimised RBF SVM models were generated for the whole training dataset using the MATLAB R2020a Classification Learner App (MCLA). Section 2.4.2.1 describes the process of uploading the biomarker data and Sections 2.4.2.2 and 2.4.2.3 describe the Optimised Linear and RBF SVM parameters, respectively. Sections 2.4.2.4 to 2.4.2.6 them describe the methods to obtain minimum classification error plots, confusion matrix for the models, and ROC curves, respectively.

2.4.2.1 Uploading of Biomarker data into MATLAB CLA

Although RBF is known as Gaussian in MATLAB, for consistency it will be referred to in this thesis as RBF. In order to create an Optimised Linear or RBF SVM model in MATLAB, the "Classification Learner App" was selected under the Apps tab. "New session" was selected and within this the "file" option was selected to allow the Microsoft Excel spreadsheet containing the pre-processed min-maxed scaled data for the seven biomarkers (n=106) to be imported and the response status to be indicated as "categorical 2 unique". Specific markers were then selected via tick boxes according to the input required for each model

to be generated. No cross validation (CV) options were selected. "Start session" was then selected to open a new session. "Optimised SVM" was selected, followed by "advanced options" to fine tune each model.

2.4.2.2 Optimised Linear SVM parameters

In the Select SVM Hyperparameters to Optimise tab the Box constraint level was "optimised" and the Kernel function was "linear". "Not optimise" and "standardise data" options were not selected as the data had previously been standardised. Within the Optimiser options the "optimiser of grid search" was selected with a training time limit of 300 seconds. The number of grid divisions was set at 10 and the box constraint level range was 0.001-1000.

2.4.2.3 Optimised RBF (Gaussian) SVM model parameters

In the Select SVM Hyperparameters to Optimise tab the selected Kernel function was "Gaussian" with the optimise option not selected. The selected Box constraint level was "optimised" and the "kernel scale" was set in the optimised settings. The "standardise data" option was not ticked in both the "values" and the "optimise" options as the data had already been scaled. The optimiser options were the same as those used for the Optimised Linear SVM parameters (Section 2.4.2.2). The box constraint level was set to 0.001-1000 and the kernel scale to 0.001-1000.

2.4.2.4 Confusion Matrix

A confusion matrix was obtained for the Optimised Linear and Optimised RBF models in the MCLA. The matrix was available in different formats and used to calculate sensitivity and specificity values based on the number of observations plot.

2.4.2.5 ROC curve

ROC curves were obtained for the models within the MCLA. ROC curves in MCLA were obtained once the model was generated by clicking on the ROC curves tab.

2.5 Creation of Random Forest models from the UOL training dataset using SPSS Modeler

Section 2.5.1 summaries the process involved in the generation of RF models using SPSS Modeler, followed by Section 2.5.2 which explains performance evaluation of the RF models and Section 2.5.3 which describes how ROC curves were generated for these models.

2.5.1 RF model generation

SPSS Version 26 was used to remove one control record from the UOL dataset which had a missing ANG2 value, thereby reducing the total dataset to n=106. Individual input files were generated for the different combinations. A stream was created in SPSS Modeler Version 18.2.1 and an RF model was generated from the input dataset by dragging the File node into the stream. The File node was connected to the Type node using the connect command, as previously described (Section 2.1). Double clicking on the Type node allowed the fields of the dataset to be seen. The NET field was changed to "flag" and the role changed to "target". The "preview" command allowed a small preview of the dataset to be seen. The Type node was then connected to the RF node, which was located in the Modelling tab. Default settings with "predefined" data were used to generate the model. The gold diamond nugget represented the resulting RF model, its features of importance, and the parameters used.

2.5.2 RF model performance evaluation

To generate the AUC value for each RF model, the Analysis node was selected from the Output tab and dragged onto the canvas. The gold diamond nugget was connected to the Analysis node, which the employed default settings with the evaluation metric (AUC and Gini binary classifiers only) selected. "Run" was selected to generate the AUC result for the model. This procedure was repeated 10 times for each combination to generate 10 different models, each with an associated AUC value. The AUC average, standard deviation (SD) and range were obtained for each set of 10 runs.

2.5.3 RF model ROC curve generation

ROC curves were generated for each RF model as described for C5.0 decision trees in Section 2.3.3.

2.6 Train and Test assessment of models using SPSS Modeler

A train and test approach was used to assess the different types of model created using the UOL data. For this analysis the whole UOL dataset was divided into 2 separate datasets with one defined as the training data and another as the validation data. The training portion (n=53) consisted of 25 controls and 28 cases while the validation portion (n=53) consisted of 24 controls and 29 cases. The methods for this approach in SPSS Modeler are shown in Sections 2.6.1 to 2.6.4 for non-boosted C5.0 decision trees, SVM, LR and RF, respectively.

2.6.1 C5.0 decision tree Train and Test assessment

The four input biomarker combinations present in the top five resulting C5.0 non-boosted decision tree models created from the whole UOL training data (with the ANG2 missing value entry removed, n=106) were assessed. Training (n=53) and test (n=53) datasets were created in SPSS Statistics for each of these combinations (VAMP, AMNTP, TAPM and CM).

The training dataset for each combination was dragged into the stream area and connected to the Type node. Within the Type node the NET field measurement was changed to "flag" and role changed to "target", before the node was connected to the C5.0 node. The settings for the decision trees were the same as those used previously (Section 2.2.1). The model nugget was connected to the Analyse node to generate the AUC value and ROC curve, as described previously (Sections 2.2.2 and 2.3.3, respectively).

The validation dataset was then dragged into the stream and connected to the Type node, and the NET field measurement and the role were changed as for the training dataset. The gold diamond representing a model was copied and pasted next to the Type node for the validation dataset, to which it was then connected. The model nugget was connected to the Analyse node to obtain the AUC value and the ROC curve for the validation portion, as previously described (Sections 2.2.2 and 2.2.3, respectively).

2.6.2 SVM Train and Test assessment

The raw biomarker data (n=106) was edited in Excel to create a min-max normalised training and validation portions that contained the same cases used for other algorithms. For both portions, min-max normalisation was carried out

using the equation in Section 2.4.1.1 and only the normalised data were used. These include a 7-marker train and test split dataset; a 4-marker CVAM train and test split dataset; a 3-marker CVA train and test split dataset; and 2-marker train and test split dataset for the CV, AV and AC combinations.

To create SVM models, the SPSS file containing the normalised training portion of the data was dragged into the stream area of SPSS Modeler and connected to the Type node (Favourite tab). Within the Type node the NET field measurement and role were changed as described (Section 2.6.1) before this node was connected to the SVM node.

Train and test Linear and RBF SVMs were created using the method described in Section 2.4.1.2 (Linear) and Section 2.4.1.3 (RBF). Once the SVM settings were selected and a model nugget for the training portion was generated, this was connected to the Analyse node to obtain the AUC value and ROC curve as described (Sections 2.2.2 and 2.2.3, respectively).

The normalised validation portion was dragged into the stream and connected to the Type node. The NET field measurement and role were changed as for the training portion before the gold diamond model nugget was copied and pasted next to the Type node. A connection between the Type node for the normalised validation portion and the SVM model nugget was then established. The model nugget was connected to the Analyse node where the AUC and ROC curve for this portion were obtained as described (Sections 2.2.2 and 2.2.3, respectively).

2.6.3 LR Train and Test assessment

Train and Test assessment of LR algorithms were carried out for LR models created from the CVAM, CVA, CV, CA, VA, C, V and A data.

The training dataset was dragged into the stream area and connected to the Type node, in which the NET field measurement and role were changed as described (Section 2.6.1). The Type node was then connected to the Logistic Node. The settings used to create each model were as previously descried (Section 2.2.1). The training LR model nugget was connected to the Analyse node where the AUC value and ROC curve were obtained as described (Sections 2.2.2 and 2.2.3, respectively).

The validation portion was dragged into the stream, connected to the Type node, and the NET field measurement and the role were changed as for the

training portion. The LR model nugget obtained for the training portion was copied and pasted next to the Type node before the model nugget and Type node for the validation portion were connected. The LR model nugget was then connected to the Analyse node where the AUC value and ROC curve for the validation portion were obtained as previously described (Sections 2.2.2 and 2.2.3, respectively).

2.6.4 Random Forest Train and Test assessment

Train and Test assessment was carried out for RF models created from the TCVAPM, CVAPM, VAPM, CVAM, CVA, TPM, CA, CM, VA, and A data.

The training dataset was dragged into the stream area and connected to the Type node, within which the NET field measurement and role were changed as described (Section 2.6.1) before connection to the Logistic Node. The settings used to create each model were as descried (Section 2.2.1). The RF model nugget was connected to the Analyse node, where the AUC value and ROC curve were obtained as described (Section 2.2.2).

The validation portion was dragged into the stream, connected to the Type node, and the NET field measurement and the role were changed as described (Section 2.6.1). The model nugget obtained for the training portion was copied and pasted next to the Type node before connection to the Type node. The RF model nugget was connected to the Analyse node, where the AUC value and the ROC curve for the validation portion were obtained as described (Sections 2.2.2 and 2.2.3, respectively).

2.7 SVM K-Fold Cross Validation in MATLAB Classification Learner App

2.7.1 Creation of Optimised Linear and RBF SVM cross-validated models

The methodology used to create the optimised SVM k-FCV models was the same as that described for the non-validated optimised SVM models (Section 2.4.2) apart from the inclusion of 5FCV, 10FCV or 20FCV. The k value was selected using the slider control (as opposed to the "no validation" option used in Section 2.4.2) before the session was started.

The k-FCV analysis in MCLA partitions the data into k folds, and for each fold it trains a model using the out-of-fold observations and assesses model performance using in-fold data. The results are reported as the average test

error over all folds. The advantage of this method is that it gives a good estimate of the predictive accuracy of the final model which has been trained using all the data. It requires multiple fits but makes efficient use of all data and is therefore ideal for smaller datasets.

This optimised SVM model was selected in the new session and the advanced options were used for further fine tuning of each model. Subsequent steps for 5FCV, 10FCV and 20FCV were as described previously for Optimised Linear and RBF SVM models (Sections 2.4.2.2 and 2.4.2.3, respectively). ROC curves for 5FCV, 10FCV and 20FCV were based on the validation portion.

2.8 10-fold Cross Validation of C5.0 Boosted Decision Trees

Boosted C5.0 decision trees were created in SPSS Modeler using the complete UOL training dataset (n=107). The build settings for 10FCV are shown in Table 2.2. The number of trials setting, which enables control of how many models are used for the boosted model, was kept at the default (10). The number of folds specifies the number of divisions of the data and this was also kept at the default value of 10. The whole process was repeated 10 times and the AUCs obtained for the runs were compared. Additional results included the Predictor Importance for each run, the percentage of boost, and the markers that had been used in each run.

	Setting
Output type	Decision Tree
Partitioning data	None
Boosting	Boosting
Boosting	Number of trials=10
Cross Validating	Number of folds =10
Mode	Simple
Favour	Generalise
Expected Noise (%)	0

Table 2.2: C5.0 Boosted Decision Tree cross validation settings.

2.9 RFH Cohort Validation of ML algorithms based on the UOL cohort

ML algorithms created using the CA data in the whole UOL dataset (Sections 2.2 – 2.5) were externally validated in MCLA using the RFH pNET biomarker data for CgA and ANG2 (n=51)..

A dataset containing only RFH pNET cases was entered into the stream and connected to the Type node. The NET field measurement and the role were changed as described (Section 2.6.1). The model nugget for each type of CA model (C5.0 non-boosted decision tree, LR, RBF SVM, Linear SVM and RF) obtained in the UOL analysis was connected to the Type node. The various model nuggets were also connected to the Analyse node to allow the confusion matrix comprising the correctly identified cases and unidentified cases to be obtained. For the RF algorithm, 10 runs were carried out to compare performance across the runs.

A similar approach was taken for the single C and A marker algorithms, excluding the Linear and RBF SVM models.

2.10 McNemar test for model selection

A McNemar test was carried out to compare the predictions obtained for selected models created using SPSS Modeler. These comparisons are shown in Table 2.3.

Model 1	Model 2	Model type
C, A, V	CA, CV, AV	LR
CA, AV, CV	CVA	LR RBF SVM Linear SVM
CVA	CVAM	LR RBF SVM Linear SVM

Table 2.3: Models selected for comparison using a McNemar test.

For these analyses to be carried out model streams were created for the LR (selected 1, 2, 3 and 4-marker combinations) and SVM (selected 2, 3 and 4-marker combinations) models (Table 2.3). The Table node located in the Output tab was used to calculate the 106 different predictions for the models that are needed for the McNemar test. The results file containing the predictions for each of the models was exported in comma separated value (csv) format and then imported into SPSS Version 27 for the McNemar test, which compares the predictions of two models. The test's confusion matrix has four components A, B, C and D. A is the number of correct predictions for Model 1 and Model 2. B is

the number of correct predictions for Model 1 and incorrect predictions for Model 2. C is the number of incorrect predictions for Model 1 and correct predictions for Model 2. D is the number of incorrect predictions for Model 1 and Model 2. Thus, the B and C values indicate the amount of difference between the two models.

A binomial McNemar test, a variation of the McNemar test is favoured from the standard McNemar test when the sample size is small or when B+C is less than 25[451], the McNemar test using a binomial distribution was carried out in Chapter 3. To carry out the McNemar test, the Analyse tab was selected to allow "descriptive statistics" and then the "crosstabs" option to be selected. The Model 1 predictions were added to the rows and the Model 2 predictions were added to the columns. The McNemar test was selected in the statistics box and an output file containing the McNemar test p value was generated. A statistically significant difference between Model 1 and Model 2 was indicted by p<0.05.

2.11 Sample size calculation for the Pancreatic Diseases cohort

Sample size calculations to obtain the number of participants required in each of the four groups (AP, pNETs, PDAC and CP) drawn from the ADEPTs cohort were carried out using the following methodology, based on Hajian-Tilaki et al., 2014[452]. The number of samples required for a defined sensitivity level ($n_{sensitivity}$) was obtained using the following equation:

$$n_{sensitivity} = \frac{Z^2_{\frac{\alpha}{2}} \times sensitivity(1 - sensitivity)}{d^2 \times prevalence}$$

The $n_{sensitivity}$ value was then divided by 4 to obtain the required number of samples per group. A range of sensitivities was selected to cover the potential sensitivity of the diagnostics test (0.6, 0.7, 0.8 and 0.9), across a range of d (maximum marginal error) values (0.1 ,0.15 and 0.2). The prevalence of pNETs used in the equation was 0.25 as there were four groups, and the $Z^2_{\alpha/2}$ value for a 5% type-1 error is 1.96.

2.12 ELISA of the External Validation cohorts

Analyses of the circulating CgA, VGF-NGF and ANG2 marker levels were carried out for the Royal Free Hospital (RFH) cohort of pNET cases, healthy control samples from the UKCTOCs biobank, and the ADEPTs cohort. Different ELISA protocols were used for the analysis of serum ANG2 (Section 2.12.2), serum CgA (Section 2.12.3) and serum VGF (Section 2.12,4).

2.12.1 Serum sample preparation
Serum samples for the RFH pNET cohort and the ADEPTs cohort were processed according to the ADEPTs protocol. In summary, for sera the blood samples were centrifuged at 2000g for 10 mins at room temperature and aliquots of the sera supernatant were stored for future ELISA experiments

2.12.2 ANG2 ELISA
ANG2 ELISAs were carried using Quantikine^{TM} ANG2 ELISA kits (R&D Systems) with reagents, working standards and samples prepared according to the manufacturer's protocol. 100 µl of Assay diluent RD1-76 was added to each well, followed by 50µl of standard, control or sample. The plate was covered with an adhesive stripped and incubated for 2 h at room temperature on a horizontal orbital microplate shaker set at 500rpm. Each well was then aspirated and washed four times with wash buffer. 200µl of Human ANG2-conjugate was added to each well and the plate was then incubated for 2 h on the shaker as before. Aspiration and washing were repeated as previously described before 200µl of substrate solution were added to each well. The plate was incubated for 30 min at room temperature and 50µl of stop solution was then added to each well. Absorbance was read using the DS2 plate reader (Dynex) at 450nm with wavelength correction set at 540nm. After adjusting for dilution, sample concentrations were determined from standard curves created using the OriginPro 2022b Standard Curve App (OriginLab).

2.12.3 CgA ELISA
Serum CgA ELISA assays were carried out using CgA ELISA kits (CisBio). 500µl of diluent were mixed with 10µl of calibrator, control or sample in an Eppendorf tube and 200µl of each was added to each ELISA plate well. The plate was covered with adhesive film and agitated for 1 h at 700 rpm at room

temperature. Wells were then aspirated and washed three times with 200µl of wash buffer. 200µl of HRP conjugate was added to each well and the plate was covered with adhesive film and incubated for 2 h at room temperature with agitation at 700 rpm. The wells were then aspirated and washed three times as before. 100µl of TMB was dispensed into each well and the plate was covered with adhesive film. The colorimetric reaction was developed for 10 min at room temperature with agitation at 700rpm. 50µl of stop solution was dispensed to each well and plates were read using a DS2 plate reader (Dynex) at 450nm. Sample concentrations were determined using Origin Pro (Section 2.12.2).

2.12.4 VGF ELISA
Serum VGF ELISA assays employed the VGF ELISA kit (Cloud-Clone Corp). 100µl of diluted sample, blank or standard were added to each well of the plate, which was then sealed and then incubated at 37°C for 1 h. The liquid was removed from each well without washing before 100µl of Detection Reagent A working solution was added to each well. The plate was sealed and incubated for 1 h at 37°C. Wells were then aspirated and washed three times with wash solution. 100µl of Detection Reagent B working solution was added to each well and the plate was sealed and incubated for 30 min at 37°C. The wells were aspirated and washed 5 times before 90µl of Substrate solution was added to each well. The plate was then incubated for 10 to 20 min at 37°C. 50µl of stop solution were then added to each well and absorbance was read at 450nm using a DS2 plate reader (Dynex). Sample concentrations were determined from standard curves using the OriginPro Standard Curve App (Section 2.12.2).

2.13 Clinical characteristics

2.13.1 Clinical characteristics of the RFH cohort and the QCancer® protocol
Samples were obtained for patients at RFH with ethical approval for this present. Patients provided consent and blood samples were provided by the patients. Sera samples were processed as explained in Section 2.12.1 and aliquots of sera were stored in the -80 freezer for experiments.
The QCancer questionnaire algorithm was pioneered by Professor Julia Hipplsey-Cox and was created from the QResearch database. The QCancer questionnaire is an algorithm that combines symptomatic information, alongside

family history questions and lifestyle questions to generate a predicted risk score of having cancer. Female and male versions of the QCancer questionnaire are available at (www.qcancer.org). Clinical information for the RFH cohort was obtained from the RFH computer system, and I invited patients who were part of the RFH pNET cohort to complete the QCancer questionnaire. The questionnaire data were analysed using the QCancer algorithm.

2.13.2 Plasma CgA levels of RFH pNET patients

RFH pNET patients have a CgA assay carried out as part of their clinical work up and I included these data in the external validation of detection algorithms. These blood test results were recorded on the ADEPTs database, but with no units. As the assays had been carried out by HSL Analytics using a Diasource assay that reports blood test results in U/L, these units were assigned to the CgA data. The healthy reference range for controls was obtained from the Diasource ELISA protocol. As my machine learning algorithms used CgA measurements in nmol/L, the U/L values were converted to nmol/L.

2.13.3 Inclusion and Exclusion criteria for the ADEPTs Pancreatic Diseases cohort

An inclusion and exclusion criteria for the ADEPTs pancreatic diseases cohort was created. The inclusion criteria were patients diagnosed with pNETs, PDAC, CP and AP. An exclusion criterion that was originally proposed was pregnancy, patients diagnosed with another cancer, patients on PPIs, patients with any mental health conditions (diagnosed depression, schizophrenia and bipolar disorder), however in practice an exclusion criterion including depression and PPI usage was not possible with the available samples.

2.13.4 Assessing relationships between biomarkers and clinical parameters and assessing confounding factors

Statistical association of individual biomarkers with clinical parameters and confounding factors were both assessed in Chapter 5 and Chapter 6. Tests used in these analyses included the Independent-Samples Mann Whitney U Test and the Independent-Samples Kruskal-Wallis Test, which were both performed using SPSS Version 28.

In SPSS, the Analyse tab was selected, and within this, "Non-parametric tests" was selected. The default option of "automatically compare distributions across

groups" was retained within the Objective tab. Within the fields tab, specific biomarker and classification data were selected as the test and the group variables, respectively. The software automatically selects the Mann Whitney U Test if assessing a group variable which has two groups, and the Independent-Samples Kruskal-Wallis Test if the group variable has more than two groups.

2.14 External validation of ML models in SPSS Modeler

LR, C5.0 decision tree, RF, RBF and Linear SVM models created using the UOL training dataset were validated using the external validation dataset consisting of ANG2 and CgA levels for the RFH pNET cases (n=60) and the UKCTOCs controls (n=51). External validation was applied to the LR, C5.0 non-boosted decision tree (general and accuracy) and RF models created using the single ANG2 and CgA markers; and the LR, C5.0 non boosted decision tree (general and accuracy), Linear SVM and RBF SVM models created using the combination of ANG2 and CgA markers.

The model to be validated was re-opened in SPSS, and the corresponding SPSS dataset was dragged into the stream. The Type node which forms the bridge between the dataset and the model nugget was then dragged into the canvas with these two nodes connected. Clicking on the Type node revealed the NET field and the different biomarker fields. The measurement option for the NET field was changed to "flag" and the role option to "target". The preview option was selected to view a snapshot of the data.The model nugget created using the training dataset was copied and pasted to the stream. The Type node was then connected to the copied model nugget, which was connected to other nodes including the Analyse node in which the confusion matrix that provided the correctly identified cases and unidentified cases was obtained. The model nugget was also connected to the Evaluation node to obtain the ROC curve and AUC value (Section x.x). The same procedure was carried out for each of the RF algorithms (A, C and AC), however as there were ten runs, each of the ten resulting models was opened and the same procedure carried out for each of these.

For the Linear and RBF SVM AC dataset (n=217), the data were minmax scaled in Microsoft Excel using the min-max approach (Section 2.4.1.1).

2.15 Validation of ML models derived from the Pancreatic Diseases cohort

LR, C5.0 decision tree, RF, RBF and Linear SVM models created using the UOL training dataset in SPSS Modeler were evaluated by assessing diagnostic performance of the individual CgA and ANG2 marker data in the pancreatic diseases cohort (n=59) . LR, C5.0 decision tree and RF models created using the combined UOL and external validation dataset (n=217) were also evaluated using the combined CgA and ANG2 marker data in the pancreatic diseases cohort (n=59). ML model evaluation in SPSS Modeler was carried out as previously described (Section 2.14).

2.16 Further algorithm development from the Combined UOL and external validation dataset

SPSS Modeler was used to create non-boosted general and accuracy C5.0 decision tree, RF and LR models as described (Sections 2.3, 2.5 and 2.2, respectively), but based on the combined UOL and external validation dataset (n=217). Similarly, the MCLA was used to create Optimised Linear and RBF SVM models as described (Section 2.4.2) using the combined UOL and external validation dataset.

2.17 Further algorithm development from the ADEPTs Pancreatic Diseases dataset

Further LR, RF, and non-boosted C5.0 decision tree algorithms were developed from the Pancreatic diseases cohort for the ANG2 and CgA makers using SPSS Modeler. Non-boosted C5.0 decision trees were created using a newer version of SPSS Modeler 18.3 (trial version). A general overview of the creation of algorithms in SPSS Modeler is described in Section 2.1. The creation of Optimised Linear and RBF SVM models (CgA and ANG2 combination) was carried out using MATLAB Classification Learner App as described (Section 2.4.2).

Chapter 3 - Development of a biomarker panel and algorithm for pNET detection using the University of Liverpool (UOL) training dataset

3.1 Introduction to the machine learning models used for analyses.

As discussed in Section 1.6, AI could revolutionize healthcare via its widespread range of potential applications, such as assisting physicians to make better clinical decisions[453]. As this project was centred around the development of models to detect patients with pNETs based on a biomarker assay and corresponding algorithm, exploring ML methods was pivotal to this research aim. However, the field of ML is vast, with many methodologies for supervised and unsupervised ML, each with its associated strengths and weaknesses. Thus, it was important that the most suitable methodologies were identified and tested to address the research aim.

The research aim was addressed using labelled data, meaning that it was known whether or not samples were from a pNET patient. Accordingly, supervised ML approaches were assessed. Key to these algorithms was to use a biomarker (independent or input variables) dataset obtained for each of the labelled cases to produce a prediction of case or control (dependent or output variable) for each case, and then to learn iteratively from the accuracy of the results[385].

Supervised ML algorithms are typically used on two types of problem, namely a classification problem or a regression problem. In a classification problem, the output variable is discrete. For example, in this project the output value for a healthy control could be defined as 0 and a pNET patient as 1 (binary classification). In contrast for a regression problem the output variable is continuous. A classification approach was used in this project.

The four supervised ML approaches used to develop the "pNET" or "not pNET" binary classification algorithms in this project were LR, SVM, C5.0 decision tree and RF. For LR to be used as a binary classifier, a threshold level of probability is assigned to differentiate the two classes[402]. The main reasons for selecting LR in my research were its relative simplicity and the ability to easily review the quantitative basis for individual classifications[401]. The second selected

approach was SVM, which has the advantage of being useful for linearly and non-linearly separable data via the use of different kernels. These include RBF (Gaussian), Polynomial and Sigmoid kernels, as well as Linear SVM[414]. Moreover, the RBF kernel has two parameters (C and gamma) that need to be carefully tuned using a grid-based approach, which can be time consuming. One disadvantage to the SVM approach is that it is a distance-based algorithm for which normalisation is an important pre-processing step[454]. Normalisation is not needed for the other ML algorithms such as RFs and decision trees as these are not distance-based algorithms.

Within previous biomarker applications, SVM had been shown to have superior performance to LR and RF in the development of a gastric cancer biomarker panel[455]. Although, care is needed in comparing the properties of different markers and panels, the properties and proven application of this algorithm made it important to assess in the development of a pNET biomarker panel.

The C5.0 decision tree approach was the third method that I selected for the development of a pNET detection algorithm due to its ability to deal with missing values and its superior performance in comparison to other algorithms including LR and artificial neural networks[456]. Advantages of a decision tree approach include its easy interpretability, non-parametric approach without distributional assumptions, handling of missing values, easy handling of skewed data and robustness to outliers[397]. Similar to LR, the main advantage of a decision tree is its simplicity and easy interpretability[397]. There is also a clear way to visually represent these algorithms, which can help guide a clinician or patient through the decision-making process. As there are many different types of decision trees, it was important to choose the most suitable type for the data. The different types of decision tree have differences in the types of input field, target (output field), splitting method, ability to handle missing values and pruning options. The C&R, QUEST, CHAID and C5.0 decision tree models can all have the target variable as the flag type in SPSS Modeler. The target variable in this case is whether the patient had a pNET or not and the flag aspect in SPSS Modeler denotes that this target variable is binomial. Moreover, as the input variables for these four tree types can be continuous, they were all compatible with the biomarker assay data. The C5.0 decision tree was selected as the decision tree to take forward for further analysis as it produced the best decision tree model based on area under the receiver-operated characteristics curve

(AUC) when compared to other decision tree models including the QUEST decision tree and CHAID decision tree in SPSS Modeler when using the Auto Classifier node (Appendix D). The Auto Classifier node in SPSS Modeler allows different algorithms to be compared based on different properties including AUC.

The final approach used for algorithm development was the RF. The RF is an ensemble (combination of algorithms) approach developed by Breimann et al[430]. It is a collection of decision trees which are combined to produce a final output. Single decision trees have easy interpretability, however they tend to perform poorly on their own and hence an ensemble approach such as RFs can produce more accurate results, as the best features for splitting at each node are selected. Random selection in this process causes the individual decision trees that make up the ensemble to emphasise different features, with the resulting diversity of the trees being able to capture more complex feature patterns than a single decision tree. RFs also reduce the chance of the training data to be overfitted, which improves predictive accuracy[457]. RFs have been shown to have good performance compared to LR[432]. However, RFs have disadvantages including their black box approach, which makes interpretation of the model difficult.

Other methodologies for supervised ML exist, including neural networks and LDA but a focus for this project was given to the four ML approaches previously mentioned. Neural networks have the disadvantages of being black box in nature, overtraining and having chance effects[458]. Thus, the neural network approach is similar to the SVM and RF in its black box nature. LDA and LR are both used widely for analysing categorical outcome variable problems, however the disadvantage of LDA compared to LR is that LDA makes more assumptions about the underlying data. Thus, LR is more flexible in this respect[459]. Hence a LR approach was more favourable for this analysis. It was therefore decided that the four ML approaches used for this project included two basic, more interpretable algorithms (C5.0 decision trees and LR) and two computationally advanced, black box model algorithms (SVMs and RFs). Hence, in the development of the pNET model a range of complexity in approaches were used.

In this chapter, I describe how the four above mentioned algorithms were explored using the UOL case-control data as a training dataset to allow marker

and algorithm performances to be determined. Comparison of the performance of different biomarker combinations and algorithms then allowed the best performing candidate tests to be identified. Validation of these marker combinations is explored in Chapters 4 and 5. Validated tests were then used to analyse other pNET and pancreatic disease cohorts using samples obtained from a bona fide clinical pathway (Chapter 6).

3.2 C5.0 non-boosted decision trees models

General and accuracy C5.0 decision trees were created from the UOL training dataset (n=107) using the methodology described in Section 2.3. C5.0 decision trees can be created in SPSS Modeler as non-boosted or boosted types. Boosting for C5.0 decision trees works by building multiple models in a sequence. The first model is initially built with the second model subsequently built in a way that it focuses on the records that were misclassified by the first model. The third model is then built focussing on the second model's misclassifications, and this process is then repeated. Finally, the cases are then classified via applying the whole set of models to them and a weighted voting procedure is used to combine the separate predictions into one overall prediction. Thus the C5.0 method uses gradient boosting which enables the algorithm to learn from classification errors of prior trees[460]. Boosted C5.0 decision tree cross-validated models will be explored in Chapter 4. In contrast the non-boosted C5.0 model does not do this, and a single model is developed. In addition, C5.0 decision trees can be created using the "favor accuracy" or "generality" setting (referred to as general C5.0 decision trees in this thesis). The generality setting aims to produce C5.0 decision tree models that are less susceptible to overfitting, however models built using the generality setting are not guaranteed to generalise better than other models. Thus to truly evaluate the generality of a model a validation approach should be used. The accuracy setting aims to create the most accurate decision tree possible. General non-boosted C5.0 decision tree results will be discussed in Section 3.2.1 and accuracy C5.0 non-boosted decision tree results will be discussed in Section 3.2.2, followed by a comparison of these models in Section 3.2.3.

3.2.1 General non-boosted C5.0 decision trees

General non-boosted C5.0 decision trees were created for all 127 marker combinations of up to all seven candidate serum markers in the training dataset (Tables 3.1 to 3.7). Models were evaluated and compared using AUC values, for which 0.5 indicated performance no better than random (i.e. no usable model).

Single marker analysis revealed that CgA (C) was the best single marker with an AUC of 0.813 (Table 3.1). The best two-marker combination was CM with an AUC of 0.850 (Table 3.2). The best three-marker combinations were CMP, CMT and MNC although each of these models utilised only CgA and MAC2BP (M) (Table 3.3). This result indicated that TIMP1 (P), TB4 (T) and NSE (N) made no significant contribution to the resulting three-marker models.

Marker entered	Marker used	Predictor Importance	AUC	Gini
C	C	C=1.00	0.813	0.627
A	A	A=1.00	0.767	0.534
V	V	V=1.00	0.681	0.361
M	None	Not Available	0.500	0.000
N	None	Not Available	0.500	0.000
T	None	Not Available	0.500	0.000
P	None	Not Available	0.500	0.000

Table 3.1: C5.0 general non-boosted decision tree results for single markers.
Results obtained when one marker was entered for general non-boosted C5.0 model creation for the training dataset (n=107). Markers entered were C (CgA), A (ANG2), V (VGF-NGF), M (MAC2BP), N (NSE), T (TB4) and P (TIMP1). The biomarkers entered and the biomarkers utilised in the individual models are shown. Three usable models were created. The marker key is used in all subsequent tables.

Markers entered	Markers used	Predictor Importance	AUC	Gini
CM	CM	C=0.90 M=0.10	0.850	0.700
CA	AC	A=0.68 C=0.32	0.835	0.670
CN	C	C=1.00	0.813	0.627
CP	C	C=1.00	0.813	0.627
CT	C	C=1.00	0.813	0.627
AM	A	A=1.00	0.767	0.534
AN	A	A=1.00	0.767	0.534
AP	A	A=1.00	0.767	0.534
AT	A	A=1.00	0.767	0.534
AV	A	A=1.00	0.767	0.534
CV	C	C=1.00	0.751	0.502
MV	V	V=1.00	0.681	0.361
NV	V	V=1.00	0.681	0.361
TV	V	V=1.00	0.681	0.361
VP	V	V=1.00	0.681	0.361
MN	None	Not Available	0.500	0.000
MP	None	Not Available	0.500	0.000
MT	None	Not Available	0.500	0.000
PN	None	Not Available	0.500	0.000

TN	None	Not Available	0.500	0.000
TP	None	Not Available	0.500	0.000

Table 3.2: C5.0 general non-boosted decision tree results for two-marker combinations.
Results obtained when two marker combinations were entered for C5.0 general
non-boosted decision tree creation. 21 combinations were assessed with usable models produced for 15
combinations.

Markers entered	Markers used	Predictor Importance	AUC	Gini
CMP	CM	C=0.90 M=0.10	0.850	0.700
CMT	CM	C=0.90 M=0.10	0.850	0.700
MNC	CN	C=0.90 M=0.10	0.850	0.700
PAV	AVP	A=0.66 V=0.20 P=0.15	0.844	0.688
AMP	AMP	A=0.75 M=0.21 P=0.04	0.844	0.689
CAV	AC	A=0.68 C=0.32	0.835	0.670
CTA	AC	A=0.68 C=0.32	0.835	0.670
CAM	AC	A=0.68 C=0.32	0.835	0.670
CAP	AC	A=0.68 C=0.32	0.835	0.670
CAN	AC	A=0.68 C=0.32	0.835	0.670
PVC	CV	C=0.80 V=0.20	0.815	0.629
MVC	CV	C=0.80 V=0.20	0.815	0.629
CNT	C	C=1.00	0.813	0.627
CNP	C	C=1.00	0.813	0.627
CTP	C	C=1.00	0.813	0.627
VAM	A	A=1.00	0.767	0.534
ATP	A	A=1.00	0.767	0.534
VAT	A	A=1.00	0.767	0.534
TNA	A	A=1.00	0.767	0.534
TAM	A	A=1.00	0.767	0.534
PAN	A	A=1.00	0.767	0.534
MNA	A	A=1.00	0.767	0.534
VAN	A	A=1.00	0.767	0.534
CVT	C	C=1.00	0.751	0.502
CNV	C	C=1.00	0.751	0.502
VMP	V	V=1.00	0.681	0.361
MVT	V	V=1.00	0.681	0.361
TNV	V	V=1.00	0.681	0.361
VTP	V	V=1.00	0.681	0.361
MNV	V	V=1.00	0.681	0.361
PNV	V	V=1.00	0.681	0.361
NMT	None	Not Available	0.500	0.000
MTP	None	Not Available	0.500	0.000
TNP	None	Not Available	0.500	0.000
NMP	None	Not Available	0.500	0.000

Table 3.3: C5.0 general non-boosted decision tree results for three-marker combinations
Results obtained when three-marker combinations were entered for general non-boosted C5.0
decision tree models. Usable models were produced for 31 out of the 35 models.

The best 4-marker model employed the APMV combination with all four
markers being utilised (Table 3.4). This revealed that C was not universally
required for the best performing models, despite it being the best individual
marker (Table 3.1). The best five-marker model employed the VTAMP and

VNAMP combinations. Notably, both used the optimal four-marker combination of APMV and omitted C (Table 3.5). In addition to identifying the optimal markers combinations, these results further strengthened the argument that NSE (N) and TB4 (T) were not important markers for pNET detection as they were absent from all optimal combinations of up to five markers. The best six-marker model employed the TVAMNP combination, but, like the optimal five-marker model, this model used only the same four markers APMV and achieved the same AUC of 0.888 (Table 3.6) as the optimal four-marker model. Increasing data input to a seven-marker combination did not lead to a better performing model as the seven-marker combination entered produced a model which just utilised two of the markers, ANG2 and CgA with a poorer AUC of 0.835 (Table 3.7). Interestingly the top performing models at the four, five and six-marker combinations entered used the APMV markers and produced an AUC of 0.888 (Tables 3.4, 3.5 and 3.6) and thus follows this pattern, however when the seven markers are entered only two markers (ANG2 and CgA). This is not consistent with the pattern seen at the four, five and six-marker level which consistently identified the best performing model. Thus, entering seven markers into the model did not produce the optimal model.

Consequently, the best C5.0 non-boosted general model utilised four of the seven markers, ANG2, VGF, MAC2BP and TIMP1, to produce an AUC of 0.888 (Tables 3.4, 3.5 and 3,6), which indicates that the markers TB4, NSE and CgA can be omitted from the starting panel to produce the best performing non-boosted general decision tree.

Markers entered	Markers used	Predictor Importance	AUC	Gini
AMVP	APMV	A=0.81 P=0.20 M=0.09 V=1×10^{-22}	0.888	0.776
CTMN	CM	C=0.90 M=0.10	0.850	0.700
CMTP	CM	C=0.90 M=0.10	0.850	0.700
CMPN	CM	C=0.90 M=0.10	0.850	0.700
TAPV	AVP	A=0.66 V=0.20 P=0.15	0.844	0.688
TAMP	AMP	A=0.75 M=0.21 P=0.04	0.844	0.689
AMPN	AMP	A=0.75 M=0.21 P=0.04	0.844	0.689
AVPN	AVP	A=0.66 V=0.20 P=0.15	0.844	0.688
AMVC	AC	A=0.68 C=0.32	0.835	0.670
CAMP	AC	A=0.68 C=0.32	0.835	0.670
AMCN	AC	A=0.68 C=0.32	0.835	0.670
ACPN	AC	A=0.68 C=0.32	0.835	0.670
CTAM	AC	A=0.68 C=0.32	0.835	0.670
CATP	AC	A=0.68 C=0.32	0.835	0.670
CTAV	AC	A=0.68 C=0.32	0.835	0.670
AVCN	AC	A=0.68 C=0.32	0.835	0.670

CTAN	AC	A=0.68 C=0.32	0.835	0.670
CAPV	AC	A=0.68 C=0.32	0.835	0.670
CMVP	CV	C=0.80 V=0.20	0.815	0.629
CNVP	CV	C=0.80 V=0.20	0.815	0.629
CMNV	CV	C=0.80 V=0.20	0.815	0.629
TMVC	CV	C=0.80 V=0.20	0.815	0.629
CVTP	CV	C=0.80 V=0.20	0.815	0.629
CNTP	C	C=1.00	0.813	0.627
TNMA	A	A=1.00	0.767	0.534
AMVN	A	A=1.00	0.767	0.534
ATPN	A	A=1.00	0.767	0.534
AMVT	A	A=1.00	0.767	0.534
TAVN	A	A=1.00	0.767	0.534
CTVN	C	C=1.00	0.751	0.502
TMPV	V	V=1.00	0.681	0.361
TMNV	V	V=1.00	0.681	0.361
NMVP	V	V=1.00	0.681	0.361
TNVP	V	V=1.00	0.681	0.361
TNMP	Not Available	None	0.500	0.000

Table 3.4: C5.0 general non-boosted decision tree results for four-marker combinations.

Results obtained when four-marker combinations were entered for C5.0 general non-boosted model creation. Models were created from 34 of the 35 possible combinations.

Markers entered	Markers used	Predictor Importance	AUC	Gini
VTAMP	APMV	A=0.81 P=0.10 M=0.09 V=1x10^{-22}	0.888	0.776
VNAMP	APMV	A=0.81 P=0.10 M=0.09 V=1x10^{-22}	0.888	0.776
AMNTP	AMPT	A=0.69 M=0.13 P=0.10 T=0.08	0.871	0.742
TNPMC	CM	C=0.90 M=0.10	0.850	0.700
TAPVN	AVP	A=0.66 V=0.20 P=0.15	0.844	0.688
CTAMN	AC	A=0.68 C=0.32	0.835	0.670
CTAMP	AC	A=0.68 C=0.32	0.835	0.670
VAPMC	AC	A=0.68 C=0.32	0.835	0.670
AVMNC	AC	A=0.68 C=0.32	0.835	0.670
AVTNC	AC	A=0.68 C=0.32	0.835	0.670
CTAMV	AC	A=0.68 C=0.32	0.835	0.670
ANPMC	AC	A=0.68 C=0.32	0.835	0.670
CTAPV	AC	A=0.68 C=0.32	0.835	0.670
AVPNC	AC	A=0.68 C=0.32	0.835	0.670
CTAPN	AC	A=0.68 C=0.32	0.835	0.670
CTMPV	CV	C=0.80 V=0.20	0.815	0.629
TVMNC	CV	C=0.80 V=0.20	0.815	0.629
VNPMC	CV	C=0.80 V=0.20	0.815	0.629
CTPVN	CV	C=0.80 V=0.20	0.815	0.629
AMNTV	A	A=1.00	0.767	0.534
TMPVN	V	V=1.00	0.681	0.361

Table 3.5: C5.0 general non-boosted decision tree results for five-marker combinations.

Results obtained from all 21 five-marker combinations.

Markers entered	Markers used	Predictor Importance	AUC	Gini
TVAMNP	APMV	A=0.81 P=0.10 M=0.09 V=1x10^{-22}	0.888	0.776
CVAMNP	AC	A=0.68 C=0.32	0.835	0.670
CATMNP	AC	A=0.68 C=0.32	0.835	0.670
CTAMNV	AC	A=0.68 C=0.32	0.835	0.670

CVMPTA	AC	A=0.68 C=0.32	0.835	0.670
CVNPTA	AC	A=0.68 C=0.32	0.835	0.670
CVMNPT	CV	C=0.80 V=0.20	0.815	0.629

Table 3.6: C5.0 general non-boosted decision tree results for six-marker combinations.
Results obtained from all six-marker combinations.

Markers entered	Markers used	Predictor Importance	AUC	Gini
CTAMNVP	AC	A=0.68 C=0.32	0.835	0.670

Table 3.7: Performance of the C5.0 general non-boosted decision tree created from the seven-marker combination.

3.2.2 Accuracy non-boosted C5.0 decision trees

Accuracy non-boosted C5.0 decision trees were also created for the 127 biomarker combinations to compare whether the accuracy setting produced models that performed better than the general setting for C5.0 non-boosted decision trees. Results for the accuracy setting were mostly found to be identical to the general setting across most combinations (Appendix A), although there were certain exceptions (Table 3.9). Single marker analysis revealed that CgA was the best single marker with an AUC of 0.813 (Table 3.8). The best two-marker model was produced from the CM combination with an AUC of 0.850 (Table 3.8). The best three-marker models arose from CMP, CMT and MNC input combinations, although each of these models utilised only the two markers C and M, and produced as expected an AUC unchanged from the two-marker CM model (Table 3.8). This result indicated that, as for the general models, TIMP1, TB4 and NSE were not significant contributors to accuracy model building.

Number of markers entered	Best marker combination entered	Markers used	Predictor Importance	AUC (Accuracy)
1	C	C	C=1.00	0.813
2	CM	C M	C=0.90 M=0.10	0.850
3	CMP, CMT, MNC	C M	C=0.90 M=0.10	0.850
4	TAMP	AMTP	A=0.79 M=0.14 T=0.06 P=0.01	0.867
5	AMNTP	AMPT	A=0.69 M=0.13 P=0.10 T=0.08	0.871
6	TVAMNP	APVM	A=0.59 P=0.25 V=0.14 M=0.02	0.864
7	CTAMNVP	AC	A=0.68 C=0.32	0.835

Table 3.8: Best performing C5.0 non-boosted accuracy decision tree models for all marker combinations.

The best model for four-marker model was obtained using AMTP, with all these markers being utilised in the resulting model that had an AUC of 0.867 (Table 3.8). The best model for 5-marker inputs was obtained using AMNTP, with only the AMTP markers utilised in the resulting model that had an AUC of 0.871 (Table 3.8). Both models used the AMTP markers however, the resulting decision trees were different which is reflected in the performance of the models with this being 0.867 and 0.871. The thresholds used for the markers were the same, as was the arrangement of the decision tree itself (Figure 3.1), however the predictor importance values differed, as seen in (Table 3.8).
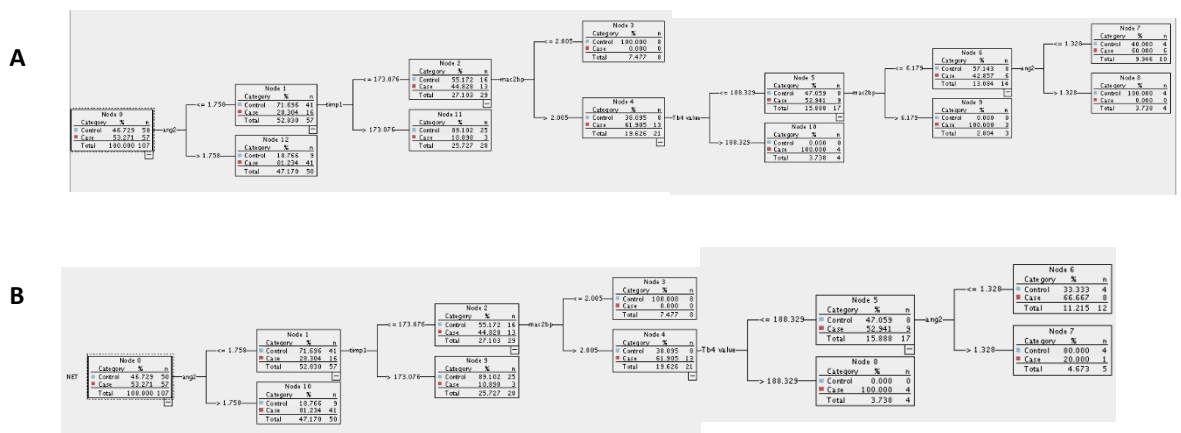
Figure 3.1: C5.0 non-boosted accuracy decision tree AMPT models with differing AUCs.
Decision tree schematics illustrating that the same four markers can result in different models with different AUCs. (A) C5.0 non boosted accuracy setting decision tree obtained when AMNTP was entered and AMPT was used. The AUC obtained for this decision tree was 0.871. (B) C5.0 non-boosted accuracy setting decision tree obtained when TAMP was entered and AMTP was used. The AUC obtained for this decision tree model was 0.867.

The best model for six-marker inputs was obtained using TVAMNP with only the AMVP markers being used in the resulting model, which had an AUC of 0.864 (Table 3.8). The best model for seven-marker inputs was the same as seen with the general non-boosted decision tree and used just two of the markers (AC) and had an AUC of 0.835 (Table 3.8).

3.2.3 General and Accuracy Non-Boosted C5.0 Decision tree comparison
General and accuracy non-boosted C5.0 decision trees produced identical results for most of the combinations, suggesting that this specific setting did not have an impact on the decision tree models that were created. However, there were certain exceptions as shown in (Table 3.9). For these marker combinations the setting of generality or accuracy mattered. For five of the combinations entered the generality setting within SPSS Modeler produced

models with higher AUC values compared to those produced using the accuracy setting. While the accuracy setting was expected to produce better performing trees, in these results the accuracy setting bettered the generality setting in only one of the trees, the TAMP entered combination. This therefore suggests that the optimal model is not consistently identified in both the generality and accuracy setting.

Marker combination entered	Marker combination used	Predictor Importance (Accuracy)	AUC (Accuracy)	Maker combination used	Predictor Importance (General)	AUC (General)
AMVP	AMVP	A=0.59 P=0.25 V=0.14 M=0.02	0.864	AMVP	A=0.81 P=0.10 M=0.09 V=1x10^{-22}	0.888
TAMP	AMTP	A=0.79 M=0.14 T=0.06 P=0.01	0.867	AMP	A=0.75 M=0.21 P=0.04	0.844
VNAMP	AMVP	A=0.59 P=0.25 V=0.14 M=0.02	0.864	AMVP	A=0.81 P=0.10 M=0.09 V=1x10^{-22}	0.888
VTAMP	AMVP	A=0.59 P=0.25 V=0.14 M=0.02	0.864	AMVP	A=0.81 P=0.10 M=0.09 V=1x10^{-22}	0.888
TVAMNP	AVMP	A=0.59 P=0.25 V=0.14 M=0.02	0.864	AMVP	A=0.81 P=0.10 M=0.09 V=1x10^{-22}	0.888

Table 3.9: C5.0 non-boosted decision trees for which the accuracy or general setting impacted the AUC.

3.3 Random Forest models

RF models were created for all 127 marker combinations of the seven biomarkers using the UOL (n=106) dataset. The missing ANG2 was removed from the dataset due to the algorithm not being able to process this. Methodology was carried out using the methodology described in Section 2.5. Single marker analysis revealed that ANG2 produced the best RF models with an average AUC of 0.993 (Table 3.10). The single marker AUC results were high with the average AUCs above ranging from 0.988 to 0.993. This therefore suggests that this algorithm can produce a high performance without the need for additional markers. When increasing the markers combination to two markers again, high AUCs were obtained with the range of average AUCs ranging from 0.994-0.999 hence the addition of a marker betters the results

comparatively to a one-marker approach. The two-marker combination analysis revealed that the best combination was CT with an average AUC of 0.999 (Table 3.11). For the three-marker analysis the average AUC range was even better with this being between 0.996-1.00. Thus, having a better average AUC range compared to the single marker and two-marker combinations. The three-marker combination analysis revealed that the best combination was CAP which had an average AUC of 1.000 (Table 3.12). However, the performance of the markers did not improve considerably with four to seven-marker combinations (Appendix B). Moreover, the performance was constantly high across all marker combinations in general, and thus suggests that the RF algorithm may be overfitting. The RF algorithm should be less susceptible compared to a single decision tree, thus, the impact of overfitting for this algorithm should be further evaluated by carrying out a validation.

Marker	Average AUC	Standard Deviation	Range
A	0.993	0.002	0.989-0.996
T	0.991	0.006	0.979-0.999
P	0.989	0.004	0.984-0.996
V	0.989	0.005	0.981-0.997
C	0.989	0.005	0.983-0.997
M	0.988	0.004	0.981-0.994
N	0.988	0.005	0.981-0.996

Table 3.10: RF model results for single marker inputs.
The result shown are the average AUC across 10 runs are shown along with the associated standard deviation. The range for the 10 runs for each marker is also shown. ANG2 as a single maker had the best average AUC of 0.993.

Markers	Average AUC	Standard Deviation	Range
CT	0.999	0.001	0.996-1.000
CA	0.998	0.002	0.996-1.000
CN	0.998	0.003	0.992-1.000
MP	0.998	0.002	0.994-1.000
AM	0.997	0.001	0.996-0.999
AN	0.997	0.002	0.994-1.000
CP	0.997	0.001	0.994-0.998
AP	0.997	0.003	0.991-1.000
CV	0.997	0.003	0.992-1.000
PN	0.997	0.004	0.990-1.000
TA	0.997	0.003	0.991-1.000
VA	0.997	0.002	0.994-1.000
TP	0.996	0.002	0.991-0.999
CM	0.996	0.003	0.990-1.000
VP	0.996	0.003	0.989-0.999
MN	0.995	0.005	0.982-0.999
TN	0.995	0.004	0.987-1.000
TV	0.995	0.004	0.987-1.000

TM	0.994	0.004	0.988-0.999
VN	0.994	0.006	0.981-1.000

Table 3.11: RF model results for two-marker inputs.

SPSS modeler was used for the creation of the models. 21 combinations for the 2 markers were carried out with 10 runs for each of the combinations. The standard deviations and the range for this are also shown. Average AUC results across all two-marker combinations were 0.994 or above.

Markers	Average AUC	Standard deviation	Range
CAP	1.000	0.001	0.998-1.000
CVA	0.999	0.002	0.994-1.000
APN	0.999	0.002	0.993-1.000
CAN	0.999	0.001	0.997-1.000
CNM	0.999	0.001	0.997-1.000
CPM	0.999	0.001	0.998-1.000
CVM	0.999	0.002	0.996-1.000
CVP	0.999	0.002	0.994-1.000
PNM	0.999	0.002	0.995-1.000
TCA	0.999	0.001	0.997-1.000
TCP	0.999	0.001	0.997-1.000
TCV	0.999	0.001	0.998-1.000
TVA	0.999	0.001	0.996-1.000
VNM	0.999	0.001	0.998-1.000
VPM	0.999	0.002	0.995-1.000
ANM	0.998	0.002	0.994-1.000
CAM	0.998	0.002	0.995-1.000
CPN	0.998	0.002	0.994-1.000
TAM	0.998	0.003	0.991-1.000
TAP	0.998	0.001	0.996-1.000
TCN	0.998	0.002	0.994-1.000
TPM	0.998	0.002	0.993-1.000
VAM	0.998	0.002	0.995-1.000
VAN	0.998	0.002	0.994-1.000
APM	0.997	0.003	0.991-1.000
CVN	0.997	0.004	0.985-1.000
TCM	0.997	0.004	0.996-1.000
TAN	0.997	0.002	0.993-1.000
TNM	0.997	0.004	0.989-1.000
TPN	0.997	0.002	0.992-1.000
TVN	0.997	0.003	0.990-1.000
TVP	0.997	0.003	0.990-1.000
VAP	0.997	0.002	0.993-0.999
VPN	0.997	0.004	0.989-1.000
TVM	0.996	0.003	0.992-1.000

Table 3.12: RF model results for the three-marker combinations.

35 combinations were analysed and for each combination 10 runs were carried out. Average AUCs were above 0.996 or above across all three-marker combinations.

3.4 SVM models for the seven-marker combination

The nomenclature used for certain SVM-related terms differed between SPSS Modeler and MCLA. The SPSS Modeler terms RBF and "C parameter" are synonymous with the MCLA terms Gaussian and "box constraint", respectively. Finally, in MCLA the RBF SVM kernel scale is synonymous with the RBF SVM

gamma setting in SPSS Modeler (referred to simply as gamma within this thesis). These terms are used interchangeably in this thesis.

Linear and RBF SVM models were created for the 7-marker combination using the UOL training dataset. Due to the SVM being a distance-based algorithm, the data were scaled using a min-max approach as described in Section 2.4.1.1. The methodologies for Linear and RBF SVM model generation using SPSS Modeler are described in Sections 2.4.1.2 and 2.4.1.3, respectively, while Optimised Linear and Optimised RBF SVM seven-marker models were created using the MCLA as described in Section 2.4.2.

Linear SVMs were constructed in SPSS Modeler across a range of C parameter settings (1 to 10). The C parameter is a tuning parameter for the Linear SVM which influences the performance of a model. Increasing the C parameter means that the cost of misclassification increases and thus, higher C parameter models are more likely to be overfitted to the data due to a harder margin. Increasing the C parameter was found to increase the AUC, with a peak value of 0.871 at C=6 and 7 (Figure 3.2) with this being the best result for the seven Linear SVMs across the tested scale. ANG2 was the most important of the seven markers in all these models (Table 3.13).

For RBF SVM generation, the C and RBF gamma tuning parameters were both optimised by varying C (1-10) and gamma values (0.4-0.95). The gamma parameter focuses on how closely fitted the model is to the training data, with lower gamma values potentially being too constrained and higher gamma values potentially causing overfitting. Higher C and gamma values were found to improve performance in terms of AUC (Figure 3.2). The best RBF SVM model for the seven-maker combination was seen at C=10 and gamma=0.95 with an AUC of 0.906.

For the Optimised Linear and Optimised RBF SVMs constructed in the MCLA, selection of the best box constraint and of kernel scale values were carried out automatically. The box constraint range used for the Optimised Linear SVM and Optimised RBF SVM was between 0.001-1000. The kernel scale for the Optimised RBF SVM was between 0.001-1000. The C parameter and box constraint do not have units associated with them. The Optimised RBF SVM had an AUC of 1.00 compared to the AUC of 0.89 for the Optimised Linear SVM (Table 3.14), and the corresponding ROC curves are shown in (Figure 3.3). Similarly, accuracy which is the percentage of correctly identified instances and

misclassification cost of the Optimised Linear SVM were lower than the Optimised RBF SVM at 82.1% and 100% accuracy, respectively (Table 3.14), and at 0 and 19 misclassifications, respectively (Table 3.14). However, the extremely high performance of the Optimised RBF SVM indicates overfitting was likely. Moreover, the box constraint (equivalent to C parameter) level for the Optimised RBF model was 10, which was at the highest end of the scale used for the SPSS Modeler C parameter range of 1 to 10 (Figure 3.2), again indicating overfitting because the higher the box constraint or C parameter value, the greater the likelihood of overfitting due to a harder margin. The kernel scale for the RBF optimised model was 0.0046 and was not within the corresponding scale in SPSS Modeler (0.4 to 0.95) hence suggesting that the scale used in the SPSS Modeler analysis may not be the most suitable for obtaining the best performing model. The box constraint level used for the Optimised Linear SVM model in MCLA was 215.4435, which was much higher than the C parameter range used for the SPSS Modeler analysis (1 to 10) (Figure 3.2). This means that as the box constraint value is higher the margin is smaller, meaning that the penalty is higher for misclassification leading to overfitting. Hence the implications are that the Optimised Linear SVM model developed in MCLA is likely to be overfitted and only by carrying out validation, can this be explored further.

A

B
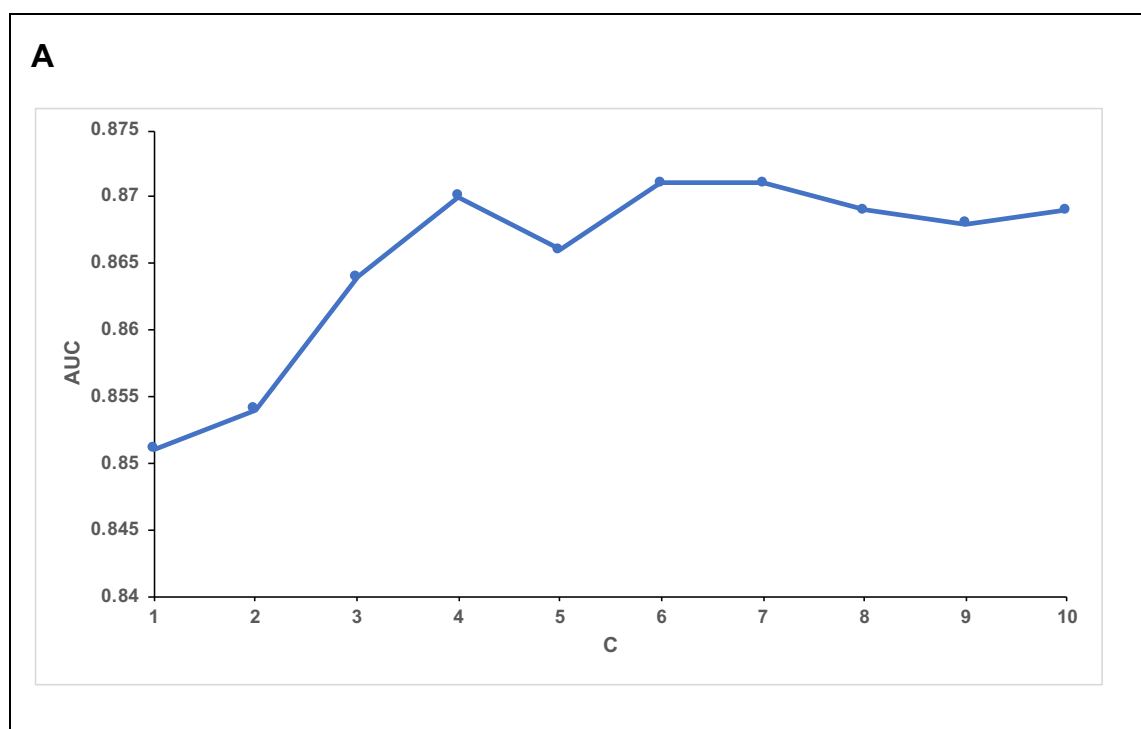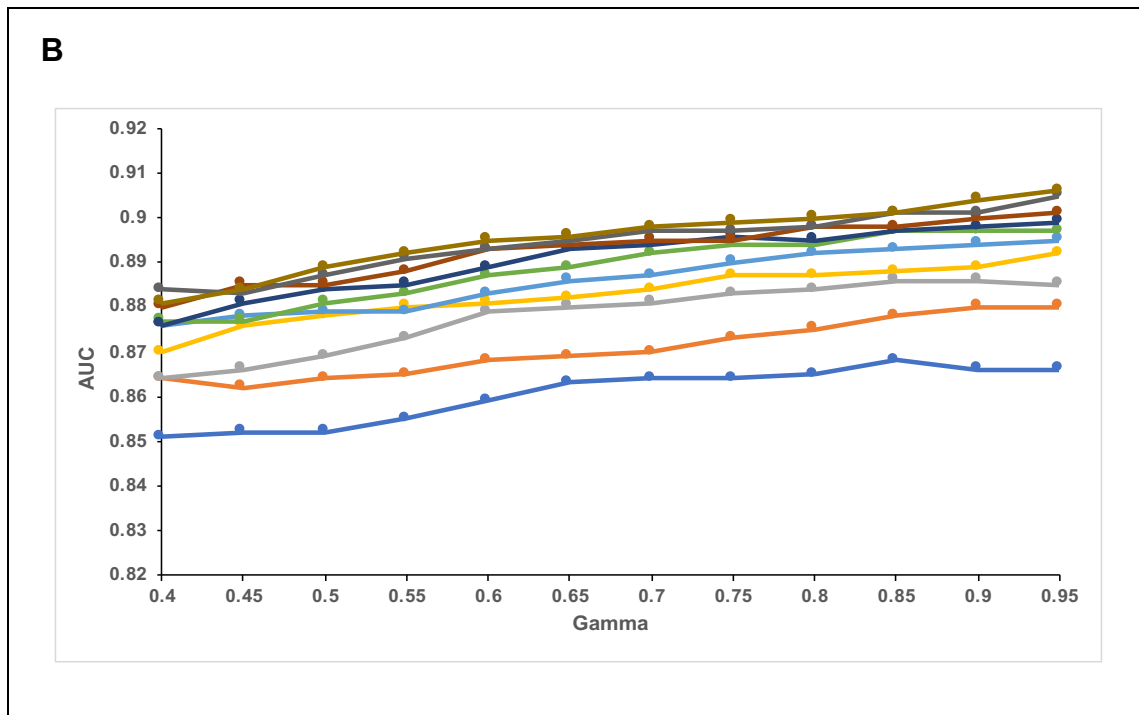
Figure 3.2: Linear and RBF SVM models for the seven-marker combination.

(A) Linear SVM for the seven-marker combination. Increasing the C value caused an increase in performance as reflected in the increase in AUC value. (B) RBF SVM for the six-marker combination excluding TB4. TB4 was excluded due to its poor performance in the seven marker SVM analysis. AUCs for the range of C values between 1 and 10 are shown as C=1 (blue), C=2 (orange), C=3 (grey), C=4 (yellow), C=5 (light blue), C=6 (green), C=7 (dark blue), C=8 (brown), C=9 (dark grey), C=10 (light brown). The general trend is that increasing C and gamma values leads to better performance. The same colour key for the C parameter is used in all subsequent RBF results.

C	1st rank	2nd rank	3rd rank	4th rank	5th rank	6th rank	7th rank
1	A=0.33	V=0.25	M=0.13	C=0.12	P=0.07	N=0.05	T=0.04
2	A=0.36	V=0.21	M=0.14	C=0.14	N=0.06	P=0.05	T=0.03
3	A=0.38	V=0.19	M=0.15	C=0.15	N=0.06	P=0.04	T=0.03
4	A=0.40	V=0.31	M=0.12	C=0.15	N=0.07	T=0.03	P=0.02
5	A=0.39	V=0.22	M=0.12	C=0.15	N=0.07	T=0.03	P=0.02
6	A=0.40	V=0.21	C=0.14	M=0.12	N=0.08	*P=0.03	*T=0.03
7	A=0.40	V=0.20	M=0.12	C=0.14	N=0.08	*P=0.03	*T=0.03
8	A=0.41	V=0.20	C=0.14	M=0.11	N=0.08	*P=0.03	*T=0.03
9	A=0.42	V=0.20	C=0.14	*M=0.08	*N=0.08	*P=0.03	*T=0.03
10	A=0.43	V=0.20	C=0.14	*M=0.08	*N=0.08	*P=0.03	*T=0.03

Table 3.13: Predictor importance for the ten Linear SVM models for the seven-marker combination.

*In all ten models A was the most important marker, while T and P were the least important. * denotes same ranking.*

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.89	82.1%	Box constraint=215.4435	19
Optimised RBF SVM	1.00	100%	Kernel scale= 0.0046 Box constraint level=10	0

Table 3.14: Seven-marker Linear and RBF Optimised SVM Models.

Results for the Optimised Linear SVM and RBF SVM show that the Optimised RBF SVM model performed better than the Linear SVM based on AUC, accuracy and misclassification cost.
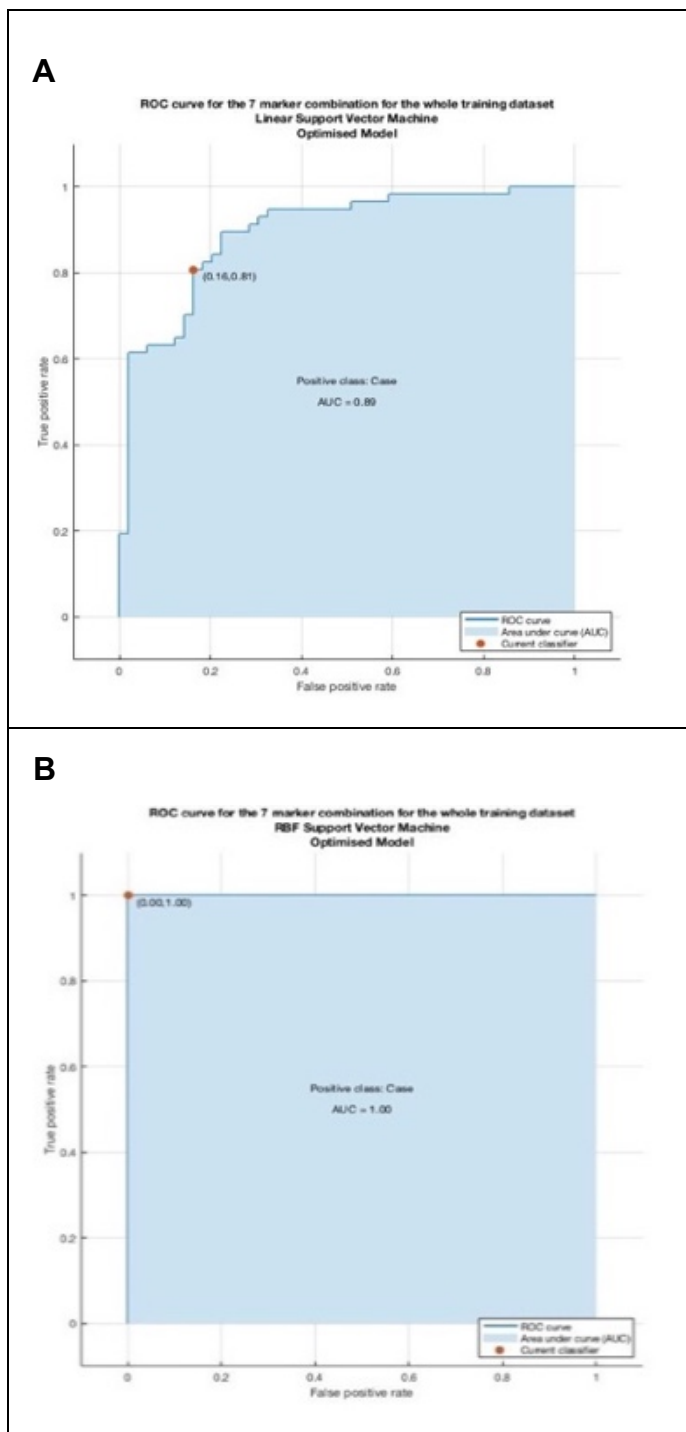
Figure 3.3: ROC curves for the seven-marker Optimised SVM models.
(*A*) *ROC curve for the Optimised Linear SVM.* (*B*) *ROC curve for the Optimised RBF SVM.*

3.5 SVM models for six-marker combinations excluding NSE, TB4 or TIMP1

Predictor Importance analysis from the seven-marker Linear SVM analysis in SPSS Modeler revealed that the lowest performing markers in the generated models were TIMP1, NSE and TB4 (Table 3.13). Thus, six-marker Linear and RBF SVM models excluding these markers were created in order to assess the impact of the loss of these markers on the performance of the models. Optimised Linear and RBF SVM models were carried out due to the limitations of the SPSS Modeler approach for SVM which include the grid approach not necessarily including the optimum Box constrain and gamma values. Optimised Linear and RBF six-marker models excluding either NSE, TIMP1 or TB4 were generated out using the MCLA. Results for the SVM analysis using SPSS Modeler showed as expected that increasing the C parameter value increased the performance of the Linear SVM models for these models (Figures 3.4, 3.6 and 3.8). Likewise, for the RBF SVM increasing values of gamma and C parameter led to better performances of these six-marker models (Figures 3.4, 3.6 and 3.8). For the six-marker models excluding either NSE, TIMP1 or TB4 the Optimised RBF model performed better than the Optimised Linear model as reflected by the AUCs values, accuracy values and the misclassification cost values (Tables 3.7, 3.9 and 3.11 and Figures 3.5, 3.7 and 3.9).

B
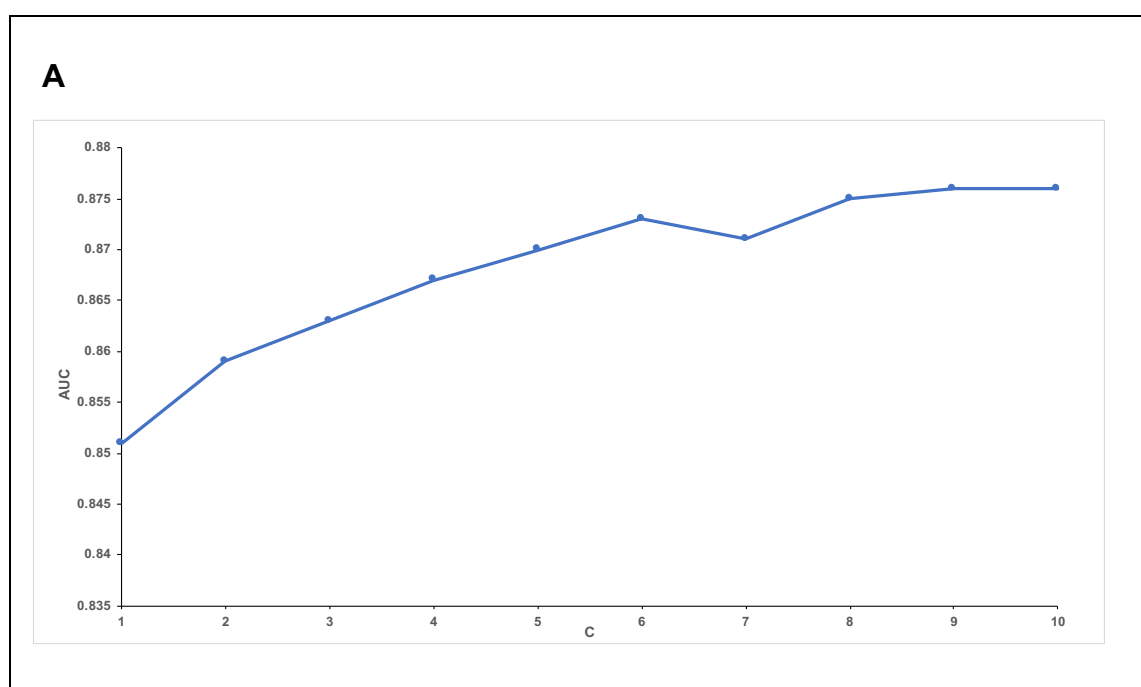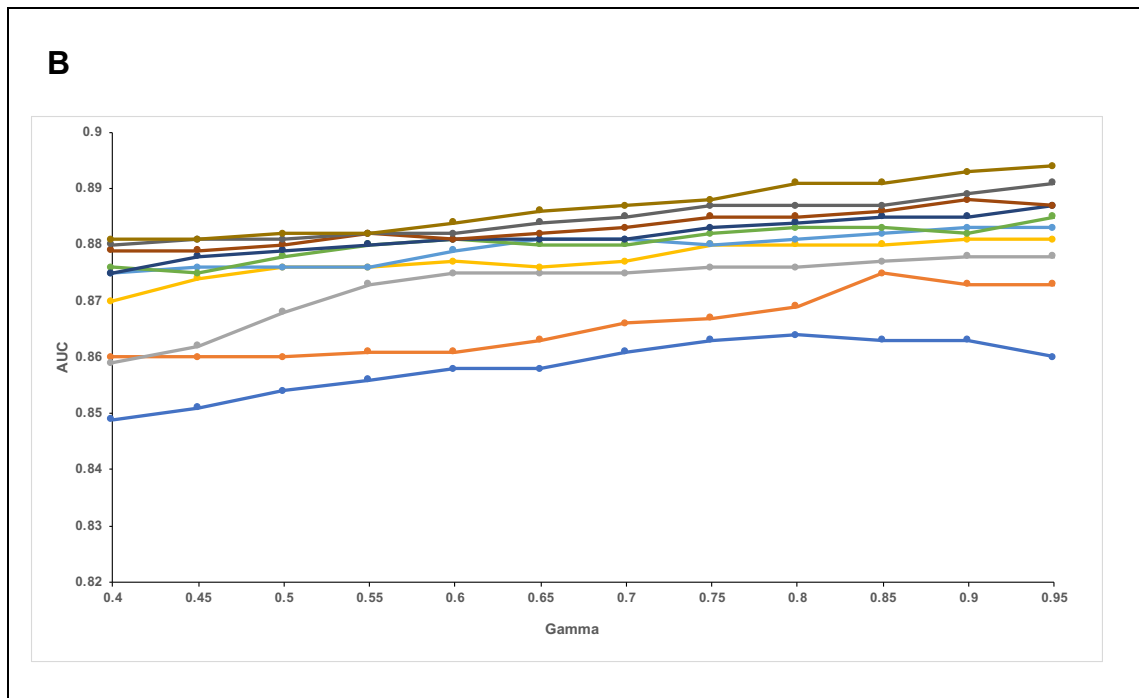
Figure 3.4: Linear and RBF SVM models generated for the six-marker combination excluding TB4.
*(**A**) Effect of the C parameter on the performance of Linear SVM models derived from the 6-marker combination excluding TB4 (UOL training dataset, n=106). The data was min-max scaled. (**B**) Effect of the C and gamma parameter on the performance of RBF SVM models derived from the 6-marker combination excluding TB4 (UOL training dataset, n=106).*
.

C	1st rank	2nd rank	3rd rank	4th rank	5th rank	6th rank
1	A=0.32	V=0.31	C=0.13	M=0.13	P=0.06	N=0.05
2	A=0.35	V=0.28	C=0.14	M=0.13	*N=0.05	*P=0.05
3	A=0.35	V=0.24	C=0.17	M=0.14	N=0.06	P=0.04
4	A=0.37	V=0.25	C=0.17	M=0.13	N=0.06	P=0.03
5	A=0.38	V=0.24	C=0.17	M=0.12	N=0.07	P=0.03
6	A=0.38	V=0.24	C=0.17	M=0.12	N=0.07	P=0.03
7	A=0.38	V=0.24	C=0.17	M=0.12	N=0.07	P=0.03
8	A=0.39	V=0.23	C=0.16	M=0.12	N=0.07	P=0.02
9	A=0.40	V=0.23	C=0.16	M=0.11	N=0.07	P=0.03
10	A=0.40	V=0.23	C=0.16	M=0.11	N=0.07	P=0.03

Table 3.15: Predictor Importance for the six-marker Optimised Linear SVM excluding TB4 and with min-max scaling.
*ANG2 was the most important marker for Optimised Linear SVM models across (C=1 to C=10) using the six-marker combination excluding TB4. * denotes same ranking.*

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.89	79.2%	Box constraint=2.1544	22
Optimised RBF SVM	1.00	100%	Kernel scale=0.0046416 Box constraint=10	0

Table 3.16: Optimised Linear and RBF SVM results for the six-marker combination excluding TB4.

Optimised RBF SVM performed better than the Optimised Linear SVM for the six-marker combination excluding TB4 as reflected by the higher AUC and accuracy and lower misclassification cost.
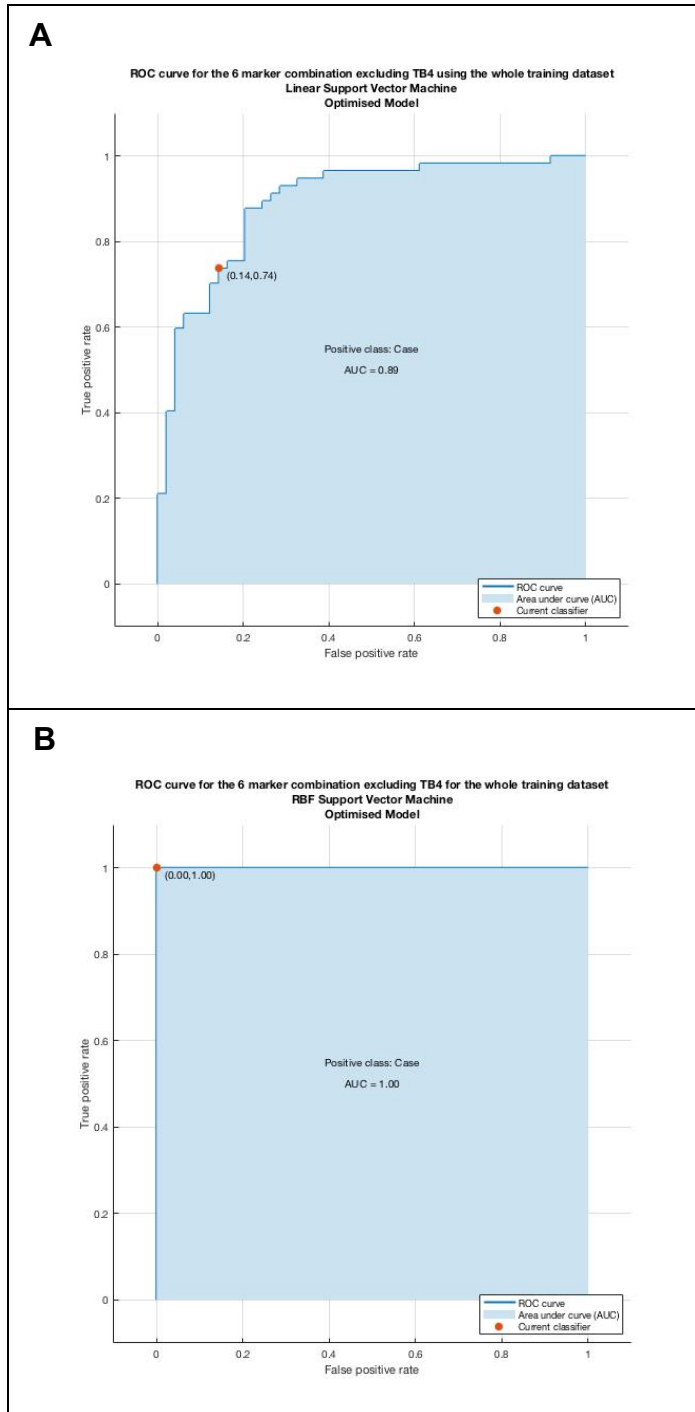
Figure 3.5: ROC curves obtained for the six-marker combination excluding TB4 Optimised SVM models min-max scaled using the MCLA. *(A) ROC curve for the Optimised Linear Support Vector Machine. (B) ROC curve for the Optimised RBF Support Vector Machine.*
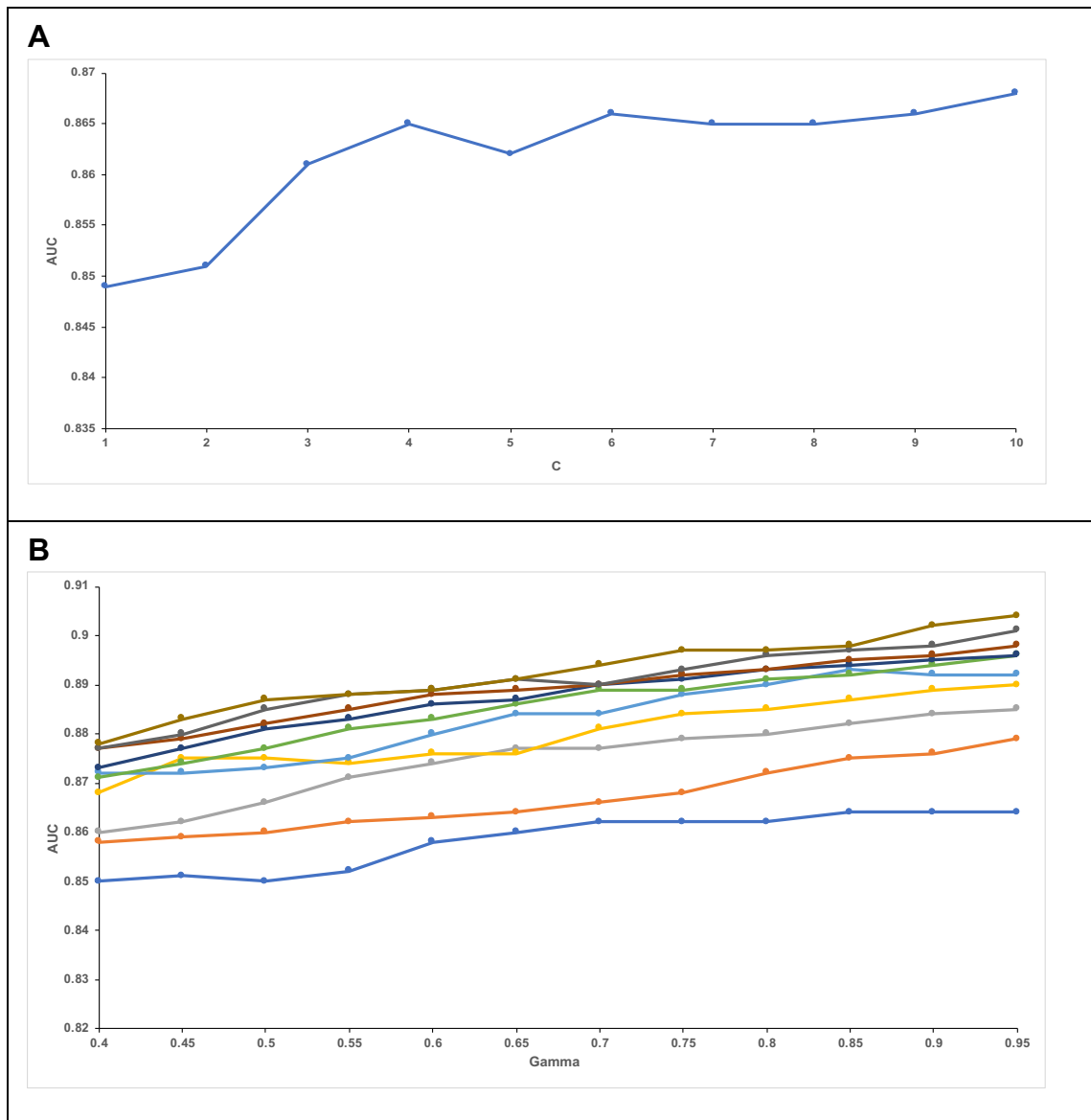
Figure 3.6: Linear and RBF SVM models for the six-marker combination excluding NSE.
(A) Effect of the C parameter on the performance of Linear SVM models derived from the 6-marker combination excluding NSE (UOL training dataset, n=106). The data was min-max scaled. (B) Effect of the C and gamma parameter on the performance of RBF SVM models derived from the 6-marker combination excluding NSE (UOL training dataset, n=106).

C	1st rank	2nd rank	3rd rank	4th rank	5th rank	6th rank
1	A=0.33	V=0.31	C=0.12	M=0.12	P=0.08	T=0.04
2	A=0.34	V=0.31	C=0.14	M=0.13	P=0.05	T=0.03
3	A=0.35	V=0.29	C=0.16	M=0.14	P=0.04	T=0.03
4	A=0.37	V=0.32	C=0.15	M=0.11	T=0.03	P=0.02
5	A=0.36	V=0.33	C=0.15	M=0.11	T=0.03	P=0.02
6	A=0.36	V=0.31	C=0.15	M=0.11	*P=0.03	*T=0.03
7	A=0.37	V=0.31	C=0.15	M=0.11	*P=0.03	*T=0.03
8	A=0.39	V=0.30	C=0.15	M=0.11	*P=0.03	*T=0.03
9	A=0.40	V=0.30	C=0.15	M=0.09	*P=0.03	*T=0.03
10	A=0.41	V=0.30	C=0.15	M=0.09	*P=0.03	*T=0.03

Table 3.17: Predictor Importance for the six-marker Optimised Linear SVM excluding NSE and with min-max scaling.
*ANG2 (A) was the most important marker across C=1 to C=10 values. * denotes same ranking*

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.88	80.2%	Box constraint level=215.4435	21
Optimised RBF SVM	1.00	100%	Kernel scale=0.0046 Box constraint=10	0

Table 3.18: Optimised SVM models generated from the six-marker combination excluding NSE and min-max scaled.

Optimised RBF SVM performed better compared to Optimised Linear SVM as reflected by the higher AUC and accuracy and lower misclassification cost.
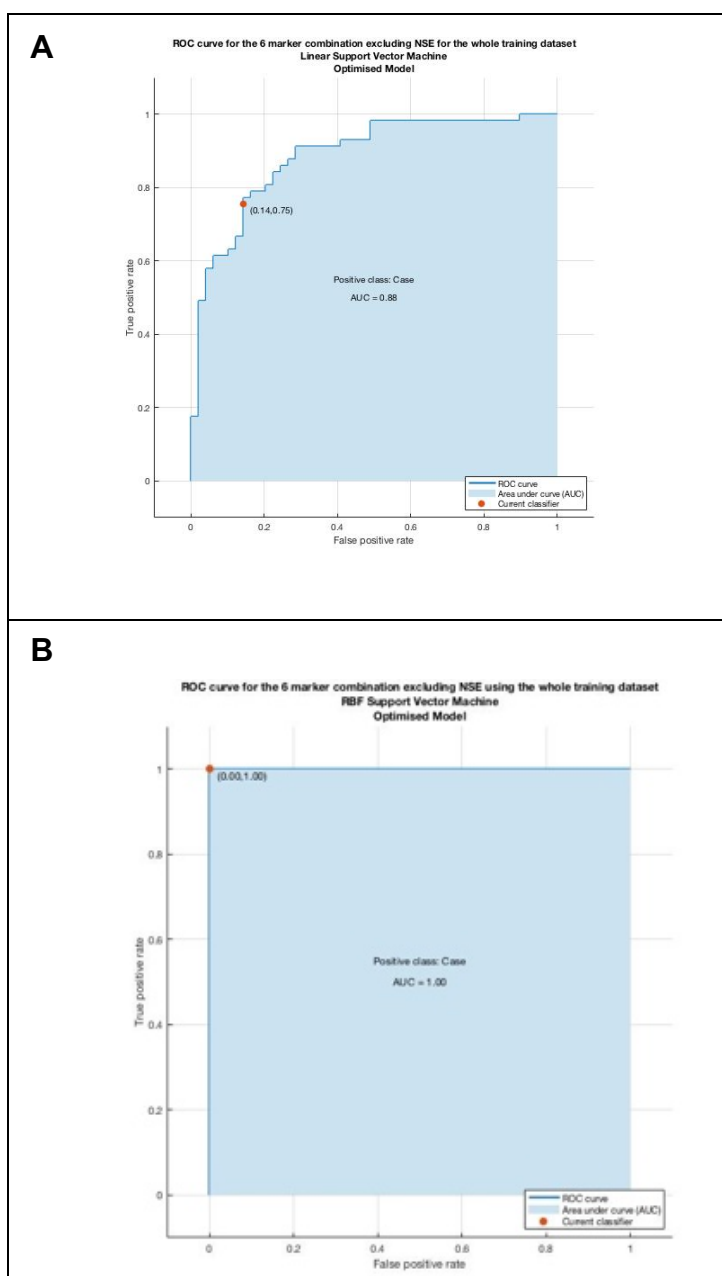
Figure 3.7: ROC curves for the Optimised SVM models generated from the six-marker combination excluding NSE and min-max scaled.

(A) ROC curve for the Optimised Linear Support Vector Machine. (B) ROC curve for the Optimised RBF Support Vector Machine.
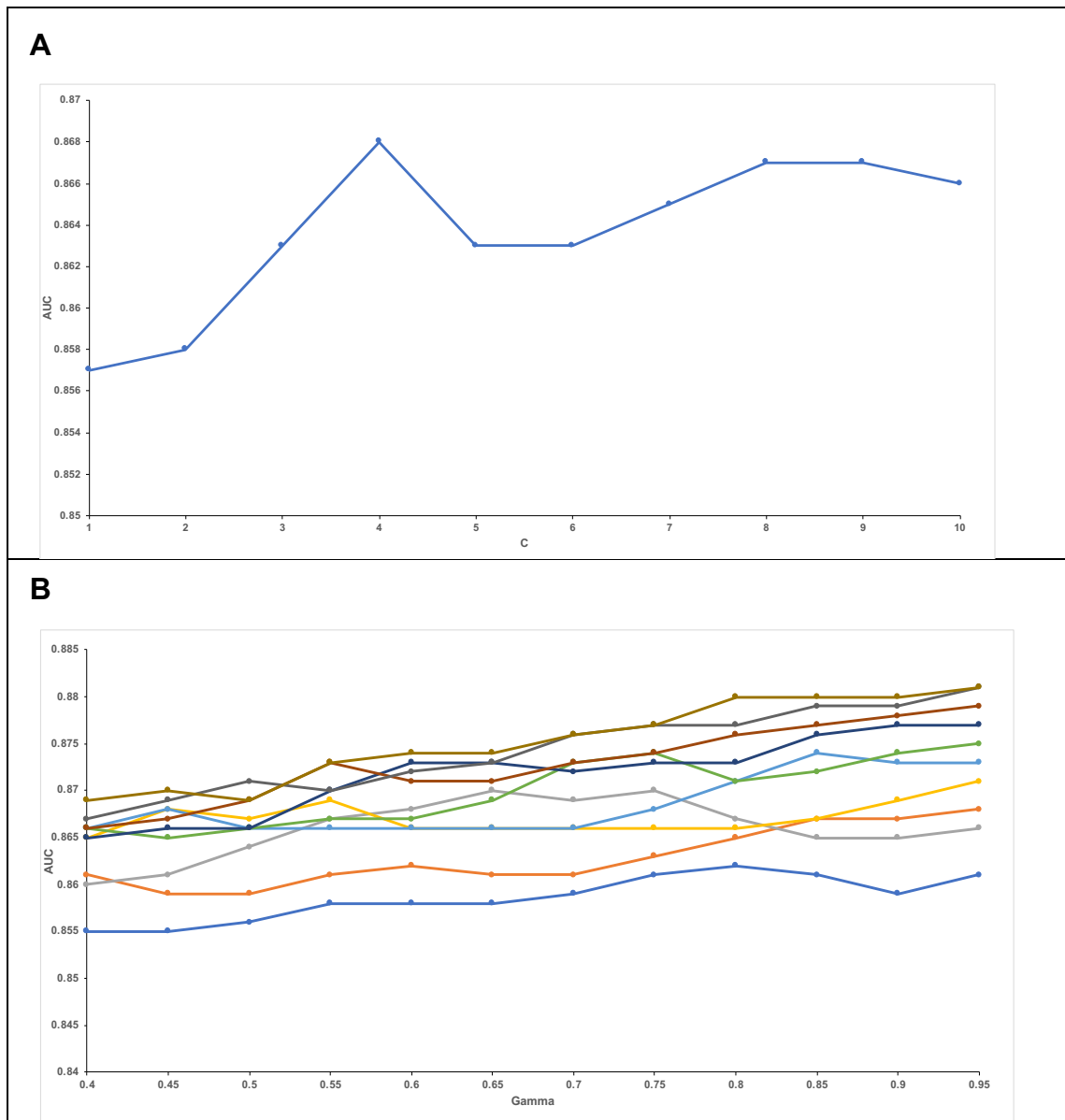
Figure 3.8: Linear and RBF SVM models for the six-marker combination excluding TIMP1.
(A)Effect of the C parameter on the performance of Linear SVM models derived from the 6 marker combination excluding TIMP1 (UOL training dataset, n=106). The data was min-max scaled. (B) Effect of the C and gamma parameter on the performance of RBF SVM models derived from the 6 marker combination excluding TIMP1 (UOL training dataset, n=106).

C regularisation Parameter	1st rank	2nd rank	3rd rank	4th rank	5th rank	6th rank
1	V=0.33	A=0.29	M=0.16	C=0.12	N=0.05	T=0.04
2	V=0.33	A=0.31	M=0.16	C=0.13	N=0.06	T=0.02
3	A=0.33	V=0.28	M=0.16	C=0.14	N=0.06	T=0.02
4	A=0.35	V=0.29	C=0.14	M=0.14	N=0.07	T=0.02
5	A=0.33	V=0.31	C=0.14	M=0.13	N=0.07	T=0.02
6	A=0.32	V=0.31	C=0.14	M=0.13	N=0.07	T=0.02
7	A=0.33	V=0.31	M=0.13	C=0.13	N=0.08	T=0.02
8	A=0.34	V=0.30	C=0.13	M=0.13	N=0.08	T=0.02
9	A=0.35	V=0.30	C=0.13	M=0.11	N=0.09	T=0.02
10	A=0.36	V=0.30	C=0.13	M=0.11	N=0.09	T=0.02

Table 3.19: Predictor Importance for the six-marker Linear SVM (excluding TIMP1) models produced with min-max scaling.

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.89	82.1%	Box constraint=46.4159	19
Optimised RBF SVM	1.00	100%	Kernel Scale=0.0046416 Box constraint=10	0

Table 3.20: Six-marker (excluding TIMP1) combination Optimised SVM model created with min-max scaled.
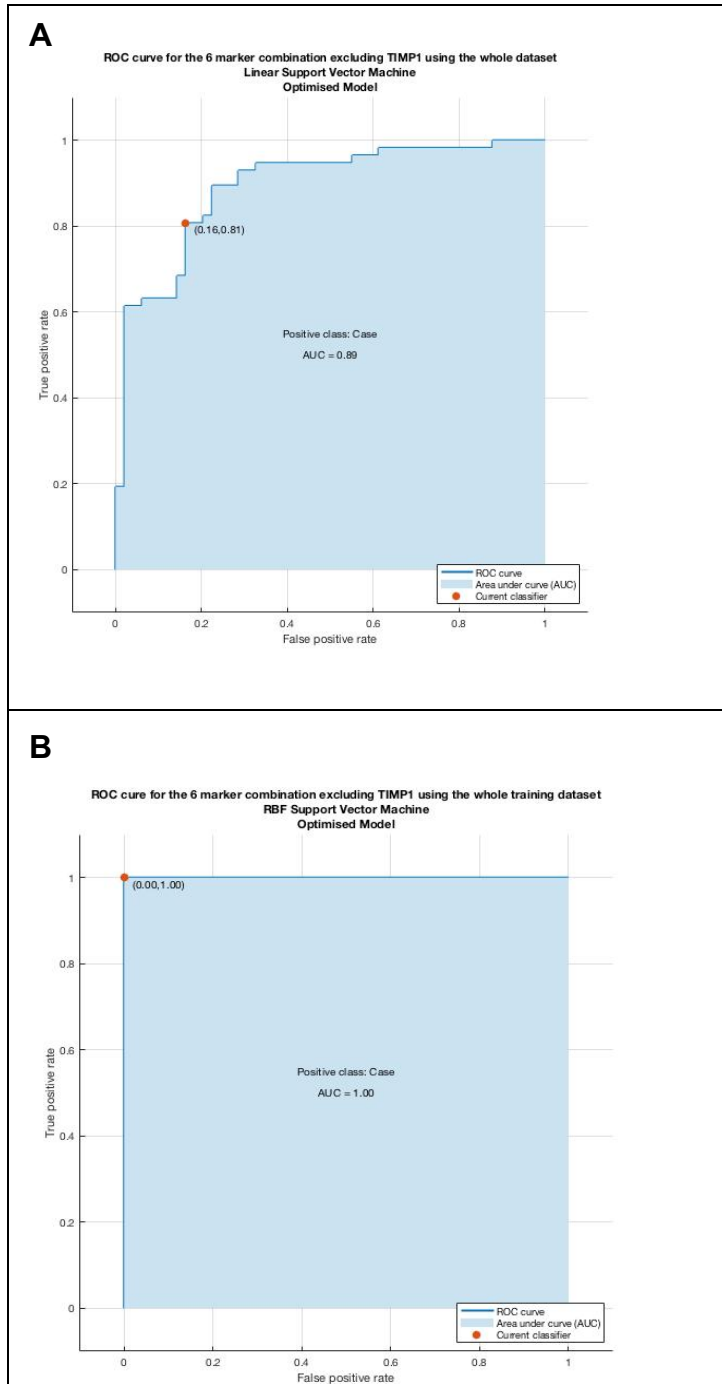
Figure 3.9: ROC curves for the six-marker (excluding TIMP1) Optimised SVM model generated using min-max scaling.
(A) ROC curve for the Optimised Linear SVM. (B) ROC curve for the Optimised RBF SVM.

3.6 SVM models for the CVAM combination

The four best markers identified from the Predictor Importance for the Linear SVM for the six-marker combinations excluding either NSE, TB4 or TIMP1 were identified as CGA, ANG2, VGF, and MAC2BP (Tables 3.15, 3.17 and 3.19). Linear SVM, RBF SVM and Optimised Linear and Optimised RBF models were therefore created for the four-marker CVAM combination to assess whether the input can be reduced further without loss of performance by the simultaneous removal of all three markers (NSE, TB4 and TIMP1). From this analysis, the best performing Linear SVM AUC was obtained for the C=9 model (Figure 3.10). The best performing RBF SVM models were seen for the C=10 and gamma = 0.8-0.95 models (Figure 3.10). The particularly high C and gamma values for the best performing RBF SVM models suggested that these models were overfitted to the training data. This required further exploration through validation. The Optimised Linear SVM and Optimised RBF SVM model created in the MCLA revealed AUC values of 0.88 and 1.00 respectively (Table 3.22). Thus, RBF SVM had superior performance to Linear SVM for the Optimised CVAM models, although the very high performance of the former indicated overfitting.

B
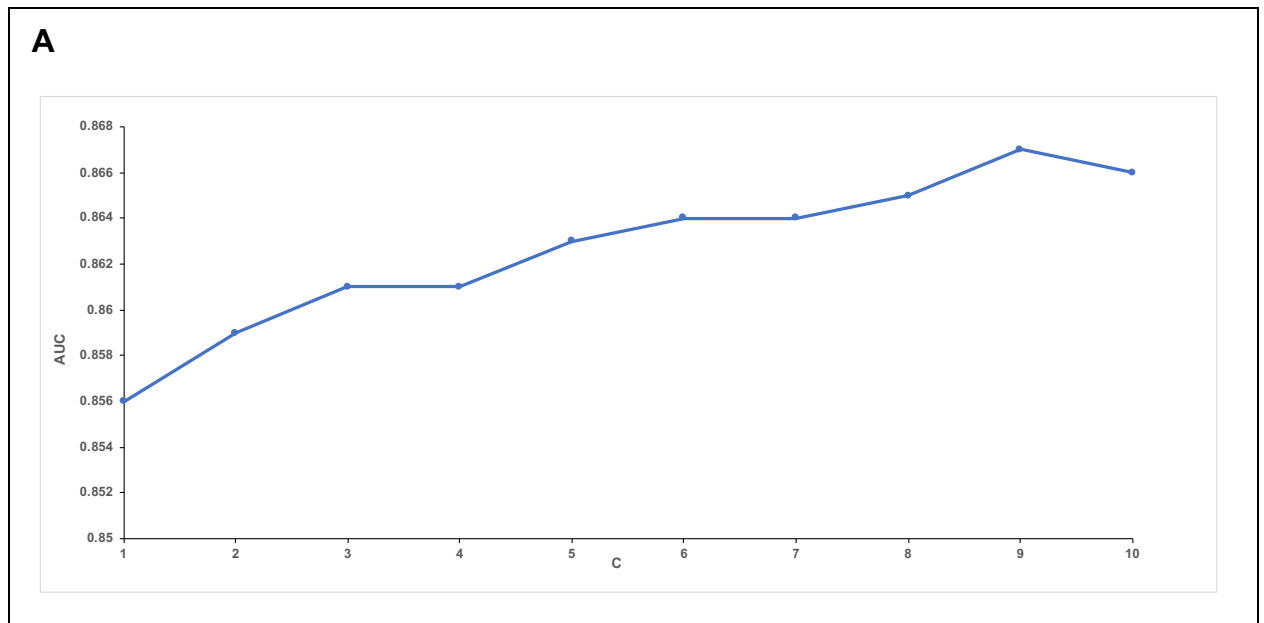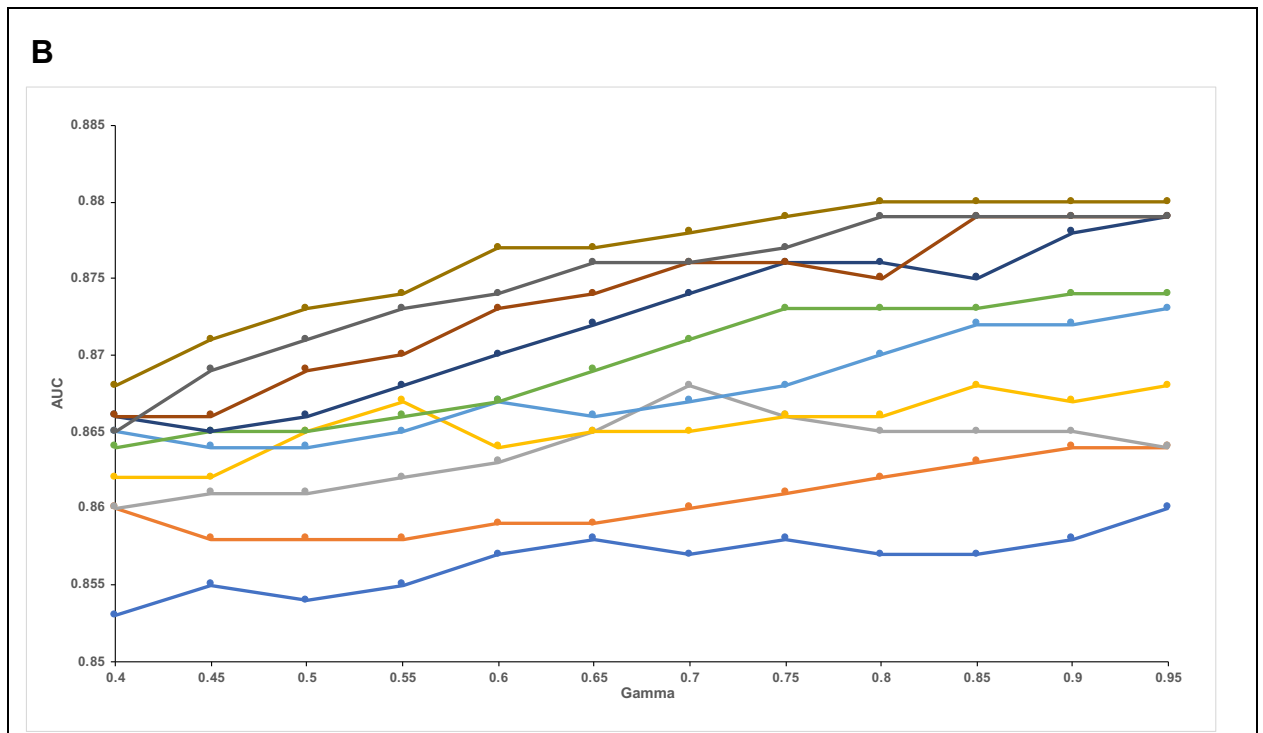
Figure 3.10: Linear and RBF SVM models for the CVAM combination.
(A) Effect of the C parameter on the performance of Linear SVM models derived from the CVAM combination (UOL training dataset, (n=106)). (B) Effect of the C and gamma parameter on the performance of RBF SVM models derived from the CVAM combination (UOL training dataset, n=106).

C	1st rank	2nd rank	3rd rank	4th rank
1	A=0.36	V=0.33	C=0.22	M=0.19
2	A=0.36	C=0.24	V=0.20	M=0.20
3	A=0.35	C=0.26	M=0.21	V=0.19
4	A=0.35	C=0.26	V=0.20	M=0.18
5	A=0.36	C=0.26	V=0.21	M=0.17
6	A=0.35	C=0.26	V=0.22	M=0.17
7	A=0.35	C=0.26	V=0.22	M=0.17
8	A=0.36	C=0.26	V=0.21	M=0.17
9	A=0.39	C=0.26	V=0.19	M=0.16
10	A=0.41	C=0.25	V=0.18	M=0.16

Table 3.21: Predictor Importance results for the CVAM Linear SVM models.
Models for the C parameter range (1-10) were generated for the min-max scaled data. Predictor Importance analysis revealed that A (ANG2) was the most important marker in all 10 models.

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.88	81.1%	Box constraint level=46.4159	20
Optimised RBF SVM	1.00	100%	Box constraint level=10 Kernel Scale=0.0046416	0

Table 3.22: Performance of the CVAM Optimised SVM models produced using min-max scaling.

The Optimised RBF SVM performed better than the Optimised Linear SVM model as reflected by higher AUC and Accuracy, and lower misclassification cost.
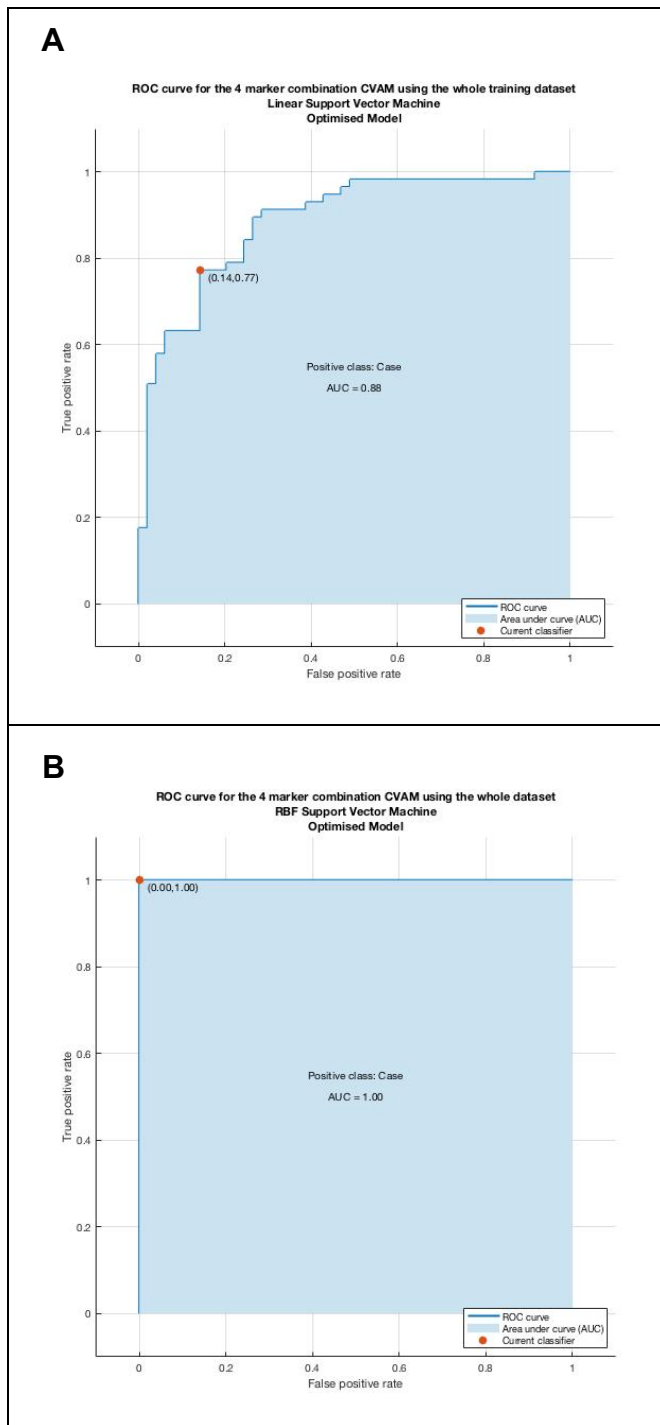
Figure 3.11: ROC curves for CVAM Optimised SVM models produced using min-max scaling.
(A) ROC curve for the Optimised Linear SVM. (B) ROC curve for the Optimised RBF SVM.

3.7 SVM models for the CVA combination

The Linear SVM for the CVAM combination revealed that the top three performing models used the markers CGA, VGF and ANG2 (Table 3.21). This three-marker combination was therefore explored using Linear, RBF, Optimised Linear, and Optimised RBF SVM approaches. Linear and RBF SVM were carried out in SPSS Modeler with the methodology described in Section 2.4.1 All optimised models were created utilising the MCLA methodology described in Section 2.4.2.

The results showed that increasing the C parameter value up to C=4 increased the AUC obtained for Linear SVMs as expected, however there was an unexpected decrease in the AUC for C>4, as increasing C value should increase performance although the C=10 AUC value was still higher than the value seen at C=1 (Figure 3.12). This inflection indicated the point at which the C parameter started to limit the performance of this combination. The highest AUC of 0.863 was seen at C=4 (Figure 3.12). The predictor importance for the Linear SVM revealed that ANG2 was the most important marker within the three-marker combination, followed by VGF and then CGA (Table 3.23). This pattern was seen across all ten SVM models.

B
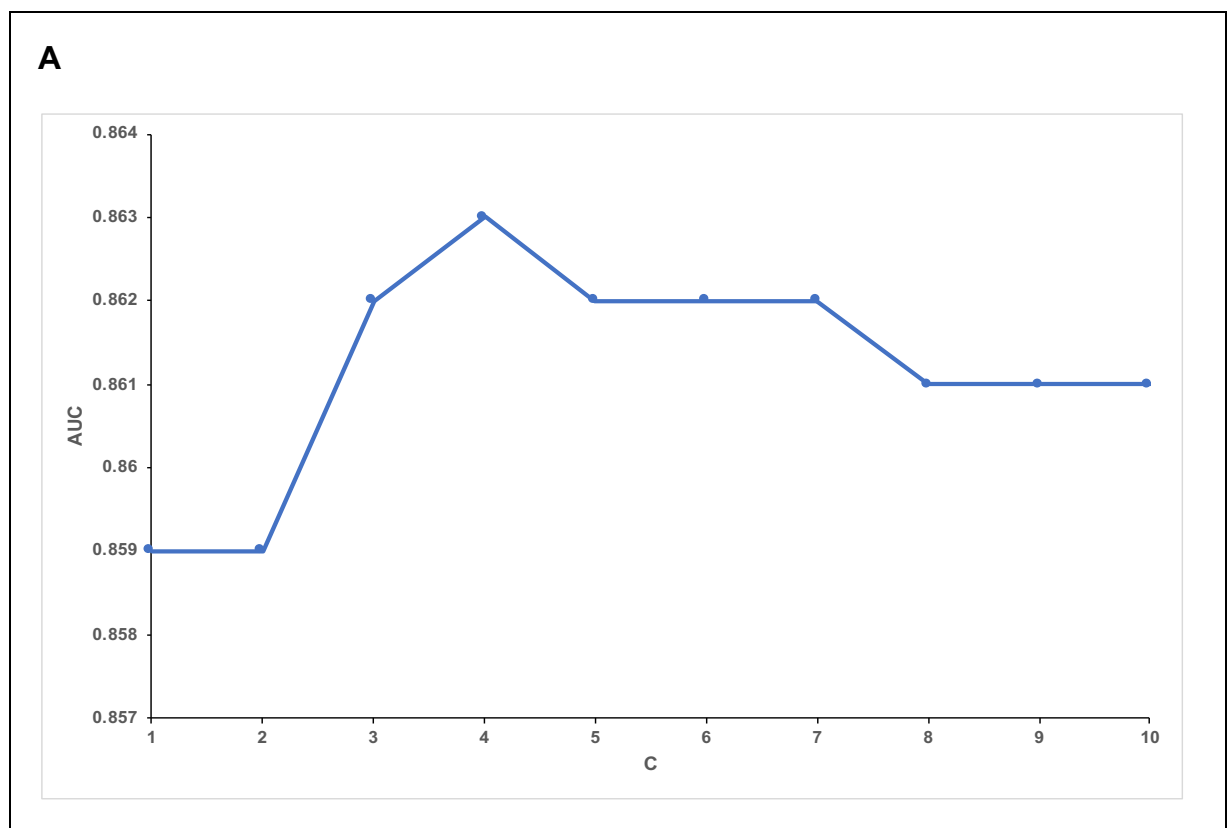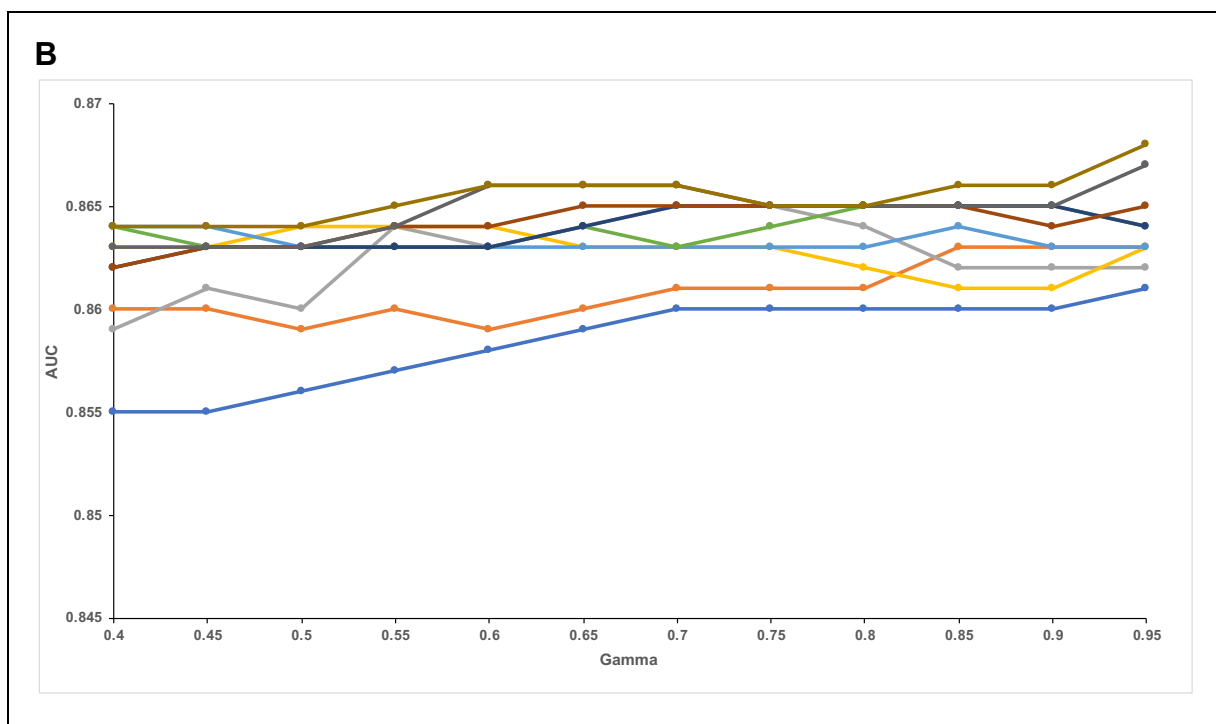
Figure 3.12: Linear and RBF SVM models for the CVA combination.
(A) Effect of the C parameter on the performance of Linear SVM models derived from the CVA combination (UOL training dataset, n=106). The data were min-max scaled. (B) Effect of the C and gamma parameter on the performance of RBF SVM models derived from the CVA combination (UOL training dataset, n=106).

C	1st rank	2nd rank	3rd rank
1	A=0.43	V=0.37	C=0.20
2	A=0.40	V=0.38	C=0.22
3	A=0.39	V=0.36	C=0.25
4	A=0.42	V=0.33	C=0.25
5	A=0.42	V=0.34	C=0.24
6	A=0.42	V=0.34	C=0.24
7	A=0.42	V=0.34	C=0.24
8	A=0.42	V=0.34	C=0.24
9	A=0.43	V=0.33	C=0.24
10	A=0.43	V=0.33	C=0.24

Table 3.23: Predictor Importance for the CAV Linear SVM models.
A range of C values from 1 to 10 were explored. Across the models, ANG2 was consistently the 1st ranking marker, followed by VGF and then CGA.

In the RBF SVM for CVA, the general trend was for increasing C and increasing gamma to lead to an increase in AUC, as expected (Figure 3.13). The highest AUC of 0.868 was seen at C=10 and gamma=0.95 (Figure 3.13). However, the fact that the best AUC was seen at the top end of the C and gamma ranges suggested that this model was overfitted.

The Optimised SVM results revealed that the Optimised RBF SVM performed better than the Optimised Linear SVM as their AUC values were 1.00 and 0.88, respectively (Table 3.24). Once again, the very high performance of the former suggested overfitting.

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.88	78.3%	Box constraint level=215.4435	23
Optimised RBF SVM	1.00	100%	Box constraint level=10 Kernel scale= 0.0046416	0

Table 3.24: Summary of results for the CVA Optimised SVM models produced using min-max scaling.

The Optimised RBF SVM model performed better than the Optimised Linear SVM model as reflected by the AUC, accuracy and misclassification cost.
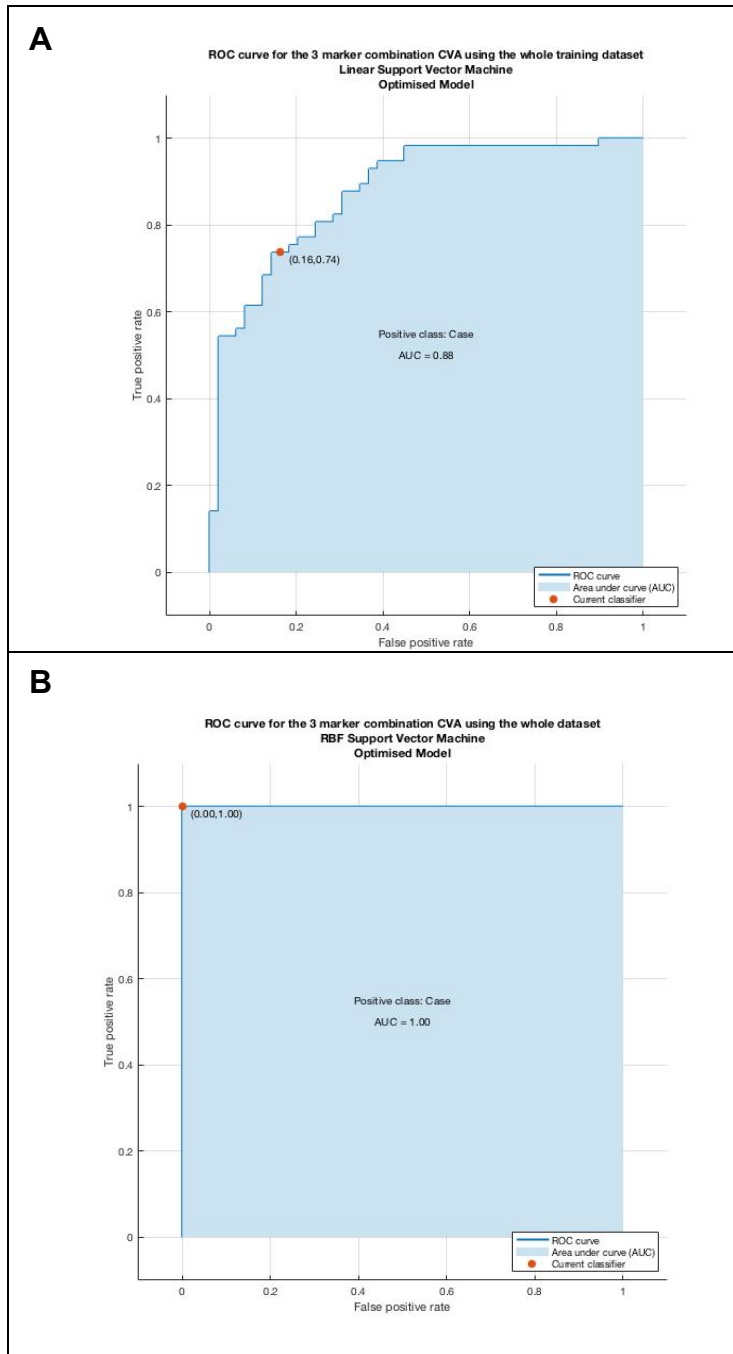
Figure 3.13: ROC curves for the CVA Optimised SVM models produced using min-max scaling.
(A) ROC curve for the Optimised Linear SVM. *(B)* ROC curve for the Optimised RBF SVM.

3.8 SVM for the AV, CA and CV combinations

Two-marker SVMs were generated using the UOL training dataset (n=106) for the different combinations of the three-marker CVA combination (VA, CA and CV) to observe whether decreasing the biomarker panel from the three-marker CVA combination to the two-marker combinations impacted performance. Linear and RBF SVM were carried out using SPSS Modeler whilst Optimised Linear and Optimised RBF SVM models were created using MCLA. Methodology for the Linear and RBF SVMs are in Section 2.4.1 and methodology for the Optimised SVM models is in Section 2.4.2.

3.8.1 Linear and RBF SVM using SPSS Modeler

For the CA Linear SVM model, unlike with previous combinations, the AUC initially decreased with increasing C parameter, followed by an increase above C=5 (Figure 3.15). For the CA RBF SVM model, smaller C parameter values had lower AUCs compared to the larger C parameter values with increasing gamma values leading to higher performing models (Figure 3.15).

For the VA combination, increasing the C parameter for the Linear SVM models led to an increase in the AUC to a peak of 0.820 at C=3 (Figure 3.14). However, the AUC dropped to 0.818 at C=4 and was then unchanged up to C=10 (Figure 3.14). This therefore showed that increasing the C parameter beyond 4 did not improve classification and that this was optimal. For the VA RBF SVM, increasing the C value and gamma value lead to increased performance (Figure 3.14).

For the CV combination with the Linear SVM, the general trend of increasing the C parameter caused an increase in AUC as expected (Figure 3.16). The RBF SVM for the CV combination also increased with increasing C and gamma values (Figure 3.16). However, the trend for the RBF SVM models lacked the consistency seen for models produced using higher marker combinations (Figure 3.14). Generally, the AUCs seen for the CA, VA, and CV combinations dropped from that seen at the 3 marker CVA level for Hence dropping to a two-marker level would result in poorer performance for both the Linear and RBF SVMs for the two-marker CA combination.
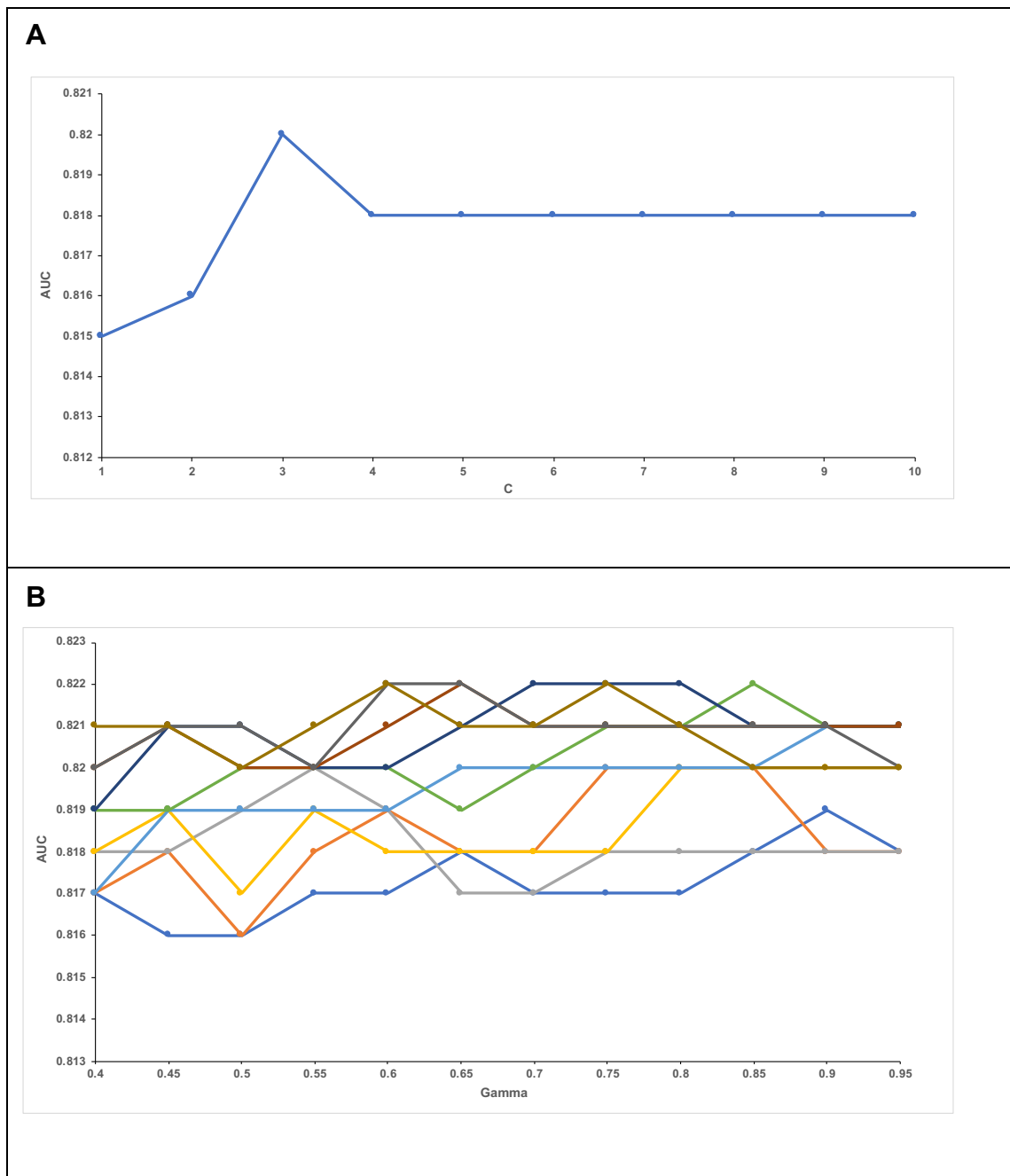
Figure 3.14: Linear and RBF SVM models for the VA combination.
(A) Effect of the C parameter on the performance of Linear SVM models derived from the VA combination UOL training dataset, (n=106). The data were min-max scaled. (B) Effect of the C and gamma parameter on the performance of RBF SVM models for the VA combination (UOL training dataset, n=106).
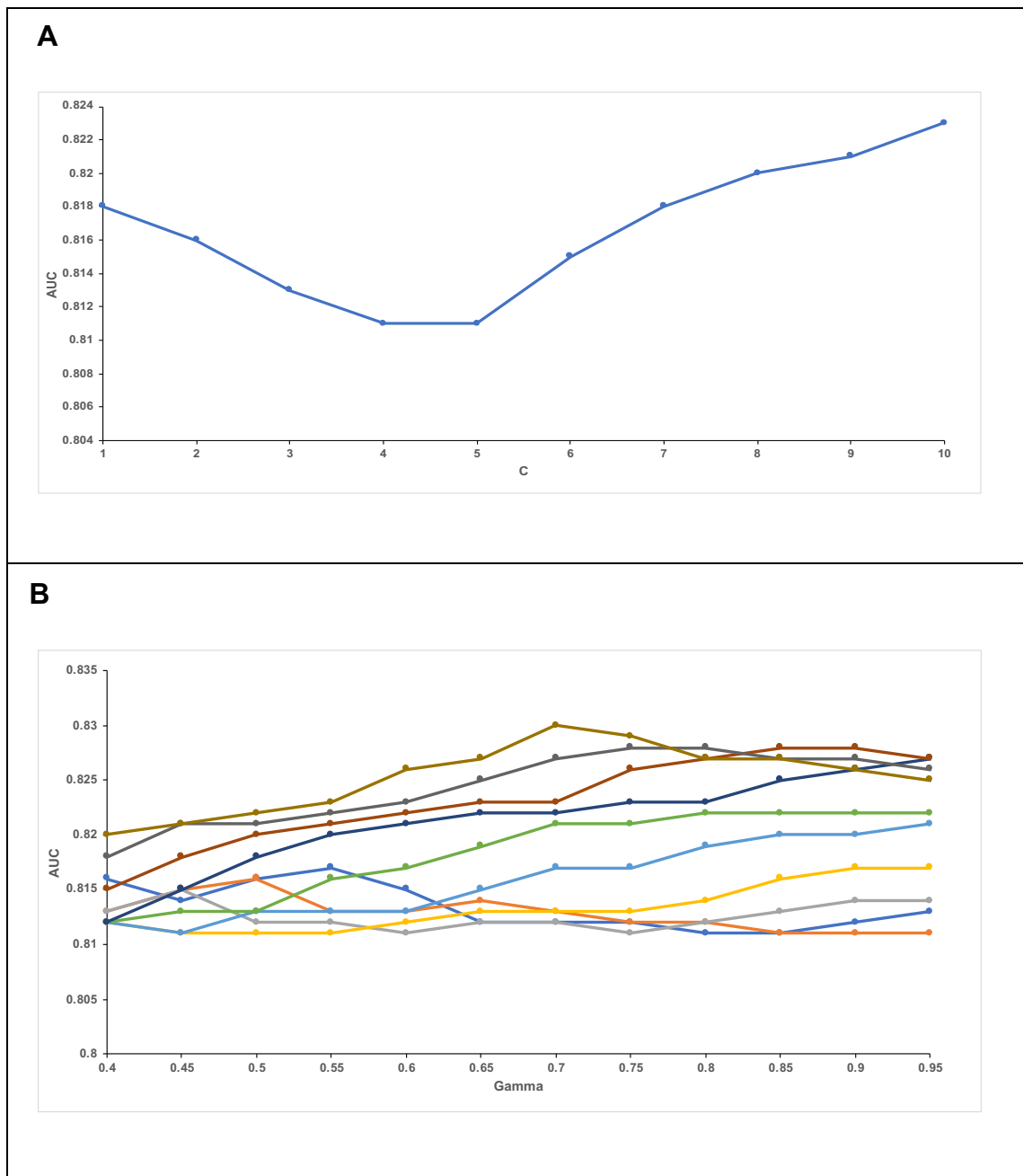
Figure 3.15: Linear and RBF SVM models for the CA combination.
(A) Effect of the C parameter on the performance of Linear SVM models derived from the CA combination (UOL training dataset, (n=106). The data was min-max scaled. (B) The effect of the C and gamma parameter on the performance of RBF SVM models for the CA combination (UOL training dataset, n=106).
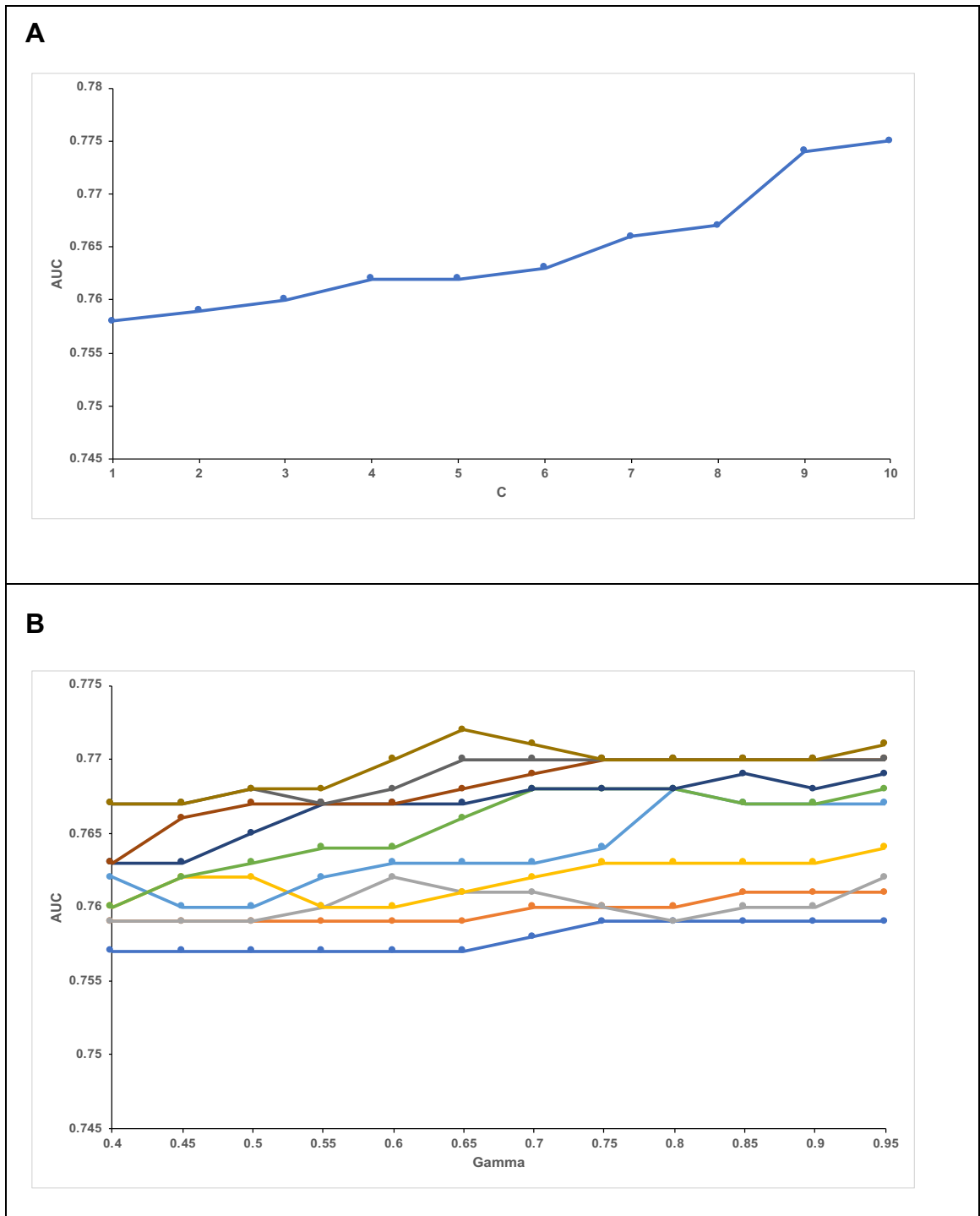
Figure 3.16: Linear and RBF SVM models for the CV combination.
(A) Effect of the C parameter on the performance of Linear SVM models derived from the CV combination (UOL training dataset, (n=106). The data was min-max scaled. The AUC obtained increases as the C parameter increases. (B) The effect of the C and gamma parameter on the performance of RBF SVM models for the CV combination (UOL training dataset, n=106).

3.8.2 Optimised Linear and SVM using MCLA

The Optimised Linear SVM for the VA combination produced an AUC of 0.82 compared to an AUC of 1.00 obtained for the Optimised RBF SVM (Table 3.25). The Optimised Linear SVM results for the CA combination revealed an AUC of 0.82 compared to an AUC of 1.00 obtained for the Optimised RBF SVM (Table 3.26). Finally, the Optimised Linear SVM results for the CV combination had an AUC of 0.79 compared to the Optimised RBF SVM which had an AUC of 1.00 (Table 3.27). Thus, the VA and CA combinations performed the best in Optimised Linear SVM compared to the CV combination which had a lower AUC of 0.79. This shows that ANG2 was important in the best performing Optimised Linear SVM model. For the RBF Optimised SVM the VA, CA and CV combinations had an AUC of 1.00. Thus, producing perfectly performing models. This therefore suggests that these RBF Optimised SVM models are overfitted and would need to be evaluated using cross validation.

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.82	75.5%	Box constraint level=0.1	26
Optimised RBF SVM	1.00	100%	Box constraint level=10 Kernel scale=0.0046416	0

Table 3.25: Summary of results for the VA Optimised SVM models produced using min-max scaled data.
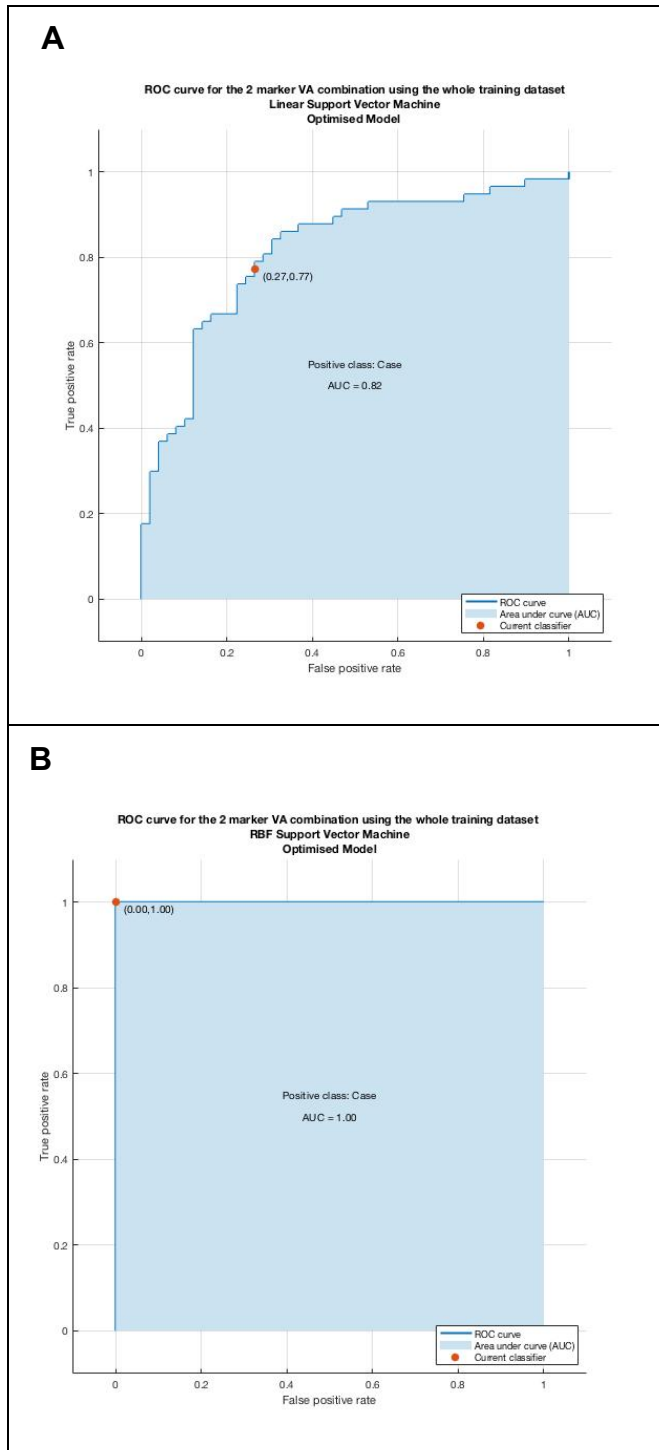
Figure 3.17: ROC curves obtained from Optimised SVM models for the VA combination produced using min-max scaled data.
 (A) ROC curve for the Optimised Linear SVM. (B) ROC curve for the Optimised RBF SVM.

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.82	76.4%	Box constraint level=0.001	25
Optimised RBF SVM	1.00	100%	Box constraint level=10 Kernel scale= 0.0046416	0

Table 3.26: Summary of results for the CA Optimised SVM models produced using min-max scaled data.
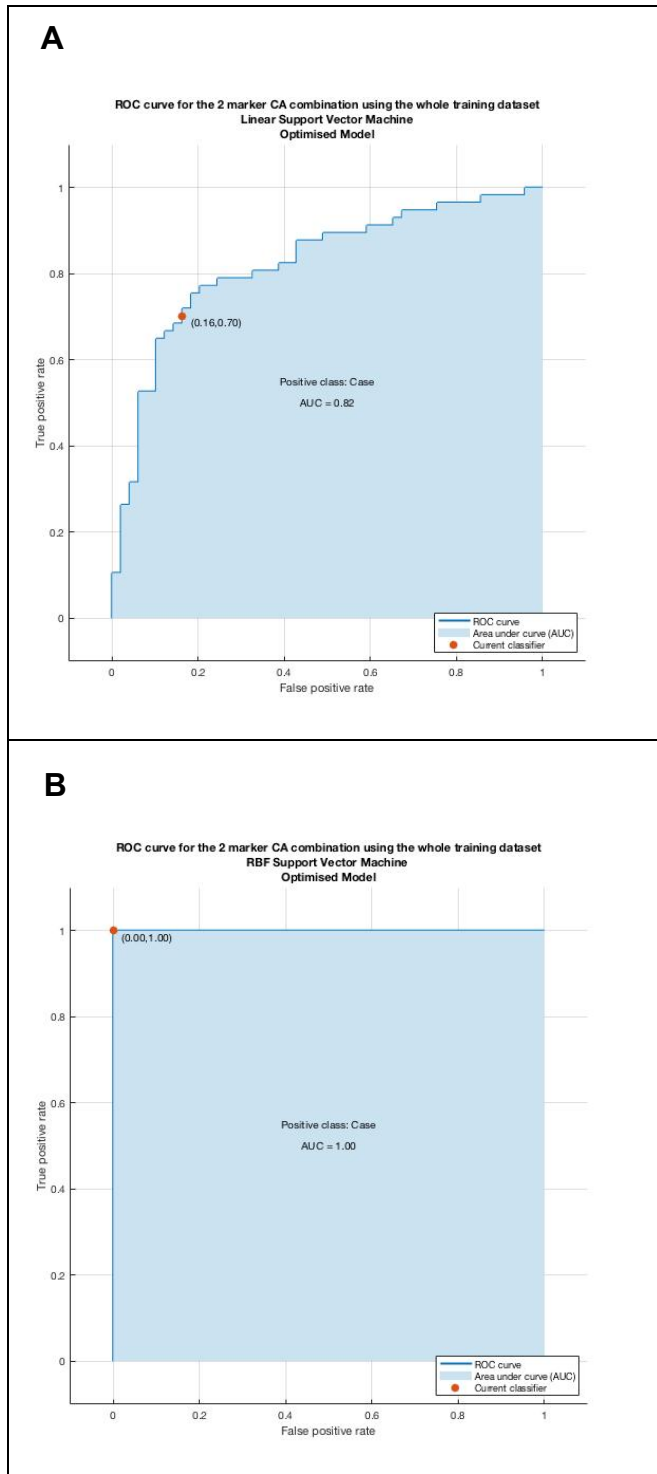
Figure 3.18: ROC curves for the CA combination Optimised SVM models produced using min-max scaled data.
(A) ROC curve for the Optimised Linear SVM. (B) ROC curve for the Optimised RBF SVM

Model	AUC	Accuracy	Parameters	Misclassification cost
Optimised Linear SVM	0.79	73.6%	Box constraint level=10	28
Optimised RBF SVM	1.00	100%	Box constraint level=10 Kernel scale= 0.001	0

Table 3.27: Summary of results for the 2 marker CV Optimised SVM models produced using min-max scaled data.
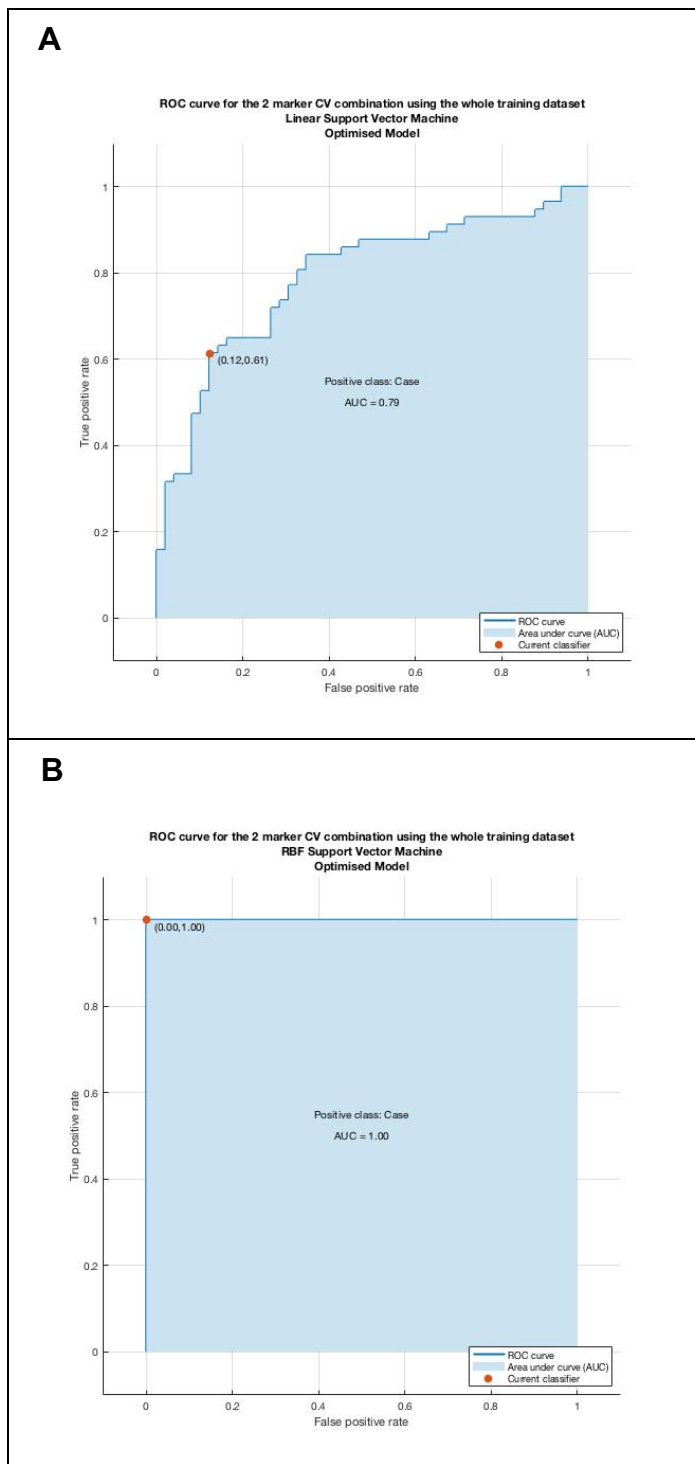
Figure 3.19: ROC curves for the 2 marker CV combination Optimised SVM models for the whole training dataset (n=106) min-max scaled using the MATLAB Classification Learner App.
(A) ROC curve for the Optimised Linear SVM.
(B) ROC curve for the Optimised RBF SVM.

3.9 Logistic Regression for the development of a pNET algorithm using SPSS Modeler (n=106)

The LR approach was used to analyse data in the whole training set (n=106), which had one control sample missing a data value removed, in order to compare the performance of this algorithm with other supervised ML methods. The methodology for this approach is described in Section 2.2. 127 combinations of the markers were assessed in order to obtain the best LR models. Single-marker LR analysis revealed that ANG2 was the best single marker with an AUC of 0.758 (Table 3.28). Two-marker LR analysis revealed that the CA combination produced the highest AUC of 0.849 (Table 3.29), which was higher than the best single marker as expected, the three-marker analysis revealed that the CVA combination produced the best AUC of 0.873, (Table 3.30). The four-marker analysis revealed that the CVAN combination produced an AUC of 0.880. The five-marker analysis revealed that the AVMNC combination produced an AUC of 0.883. The six-maker LR analysis showed that the best combination was the CVAMNP combination which produced an AUC of 0.885. The best performing LR model was the seven-marker model which produced an AUC of 0.886. For LR, increasing the number of markers in the model, improved model performance (Figure 3.20). LR also showed that at the single-marker level TB4, NSE and TIMP1 performed the poorest with AUCs of 0.559, 0.564 and 0.570 respectively. The four-marker CVAN (Figure 3.21) and CVAM (Figure 3.22) combinations were the top performing four-marker combinations. Within both these models NSE and MAC2BP were the least important markers as assessed by the predictor of importance. Removal of NSE or MAC2BP from the four-marker combination reduced performance as the AUC for the three marker CVA combination (Figure 3.23) was 0.873. For the two-marker CA (Figure 3.24) CV (Figure 3.25) and VA (Figure 3.26) combinations, the CA combination had the best AUC of 0.849, however all three two-marker combinations were the top three performing two-marker combinations. Finally, at the one-marker level, comprising C (Figure 3.27), A (Figure 3.28), and V (Figure 3.29), there was a further drop in AUC which illustrated how the AUC dropped with a reduction of markers from the original four-marker combination.

Marker	Predictor Importance	AUC
A	A=1.00	0.758
C	C=1.00	0.750
V	V=1.00	0.691
M	M=1.00	0.667
P	P=1.00	0.570
N	N=1.00	0.564
T	T=1.00	0.559

Table 3.28 LR model results for single-marker inputs

Combination	Predictor Importance	AUC
CA	A=0.53 C=0.47	0.849
VA	A=0.62 V=0.38	0.819
CV	C=0.58 V=0.42	0.795
CM	C=0.57 M=0.43	0.778
CT	C=0.82 T=0.18	0.774
AN	A=0.76 N=0.24	0.772
AM	A=0.69 M=0.31	0.771
CN	C=0.74 N=0.26	0.767
CP	C=0.82 P=0.18	0.766
TA	A=0.85 T=0.15	0.763
AP	A=0.92 P=0.08	0.755
MV	M=0.51 V=0.49	0.738
VN	V=0.64 N=0.36	0.701
TV	V=0.80 T=0.20	0.697
VP	V=0.70 P=0.30	0.696
TM	M=0.78 T=0.22	0.685
MN	M=0.67 N=0.33	0.671
MP	M=0.54 P=0.46	0.671
TP	P=0.52 T=0.48	0.627
PN	N=0.63 P=0.37	0.616
TN	N=0.61 T=0.39	0.586

Table 3.29 LR model results for two-marker inputs

Combination	Predictor Importance	AUC
CVA	A=0.44 C=0.28 V=0.28	0.873
CAM	A=0.44 C=0.35 M=0.21	0.856
CAN	A=0.46 C=0.35 N=0.19	0.849
CAP	A=0.49 C=0.47 P=0.04	0.849

TCA	C=0.48 A=0.46 T=0.06	0.843
VAN	A=0.48 V=0.32 N=0.20	0.829
VAM	A=0.46 V=0.33 M=0.21	0.823
TVA	A=0.51 V=0.41 T=0.08	0.820
VAP	A=0.64 V=0.31 P=0.04	0.819
CVP	C=0.56 V=0.31 P=0.12	0.810
CVN	C=0.42 V=0.34 N=0.24	0.803
CVM	C=0.39 M=0.31 V=0.29	0.802
TCV	C=0.57 V=0.33 T=0.10	0.798
ANM	A=0.54 M=0.23 N=0.22	0.791
TCP	C=0.63 T=0.22 P=0.15	0.786
TCM	C=0.61 M=0.29 T=0.10	0.784
CNM	C=0.56 M=0.29 N=0.15	0.781
TCN	C=0.71 N=0.15 T=0.14	0.780
CPM	C=0.56 M=0.33 P=0.10	0.776
CPN	C=0.56 N=0.27 P=0.17	0.774
TAN	A=0.68 N=0.24 T=0.08	0.773
APM	A=0.63 M=0.33 P=0.05	0.773
APN	A=0.70 N=0.24 P=0.06	0.772
TAM	A=0.64 M=0.27 T=0.10	0.771
TAP	A=0.81 T=0.11 P=0.08	0.759
VNM	V=0.42 M=0.30 N=0.28	0.750
TVM	M=0.49 V=0.44 T=0.07	0.741
VPM	V=0.43 M=0.40 P=0.16	0.739
VPN	V=0.48 N=0.31 P=0.21	0.716
TVP	V=0.67 P=0.18 T=0.15	0.705
TVN	V=0.63 N=0.27 T=0.09	0.702
TPM	M=0.62 T=0.19 P=0.18	0.687
PNM	M=0.60 N=0.26 P=0.14	0.679
TNM	M=0.57 N=0.33 T=0.10	0.678
TPN	N=0.49 P=0.28 T=0.23	0.634

Table 3.30 LR model results for three-marker inputs

Combination	Predictor Importance	AUC
CVAN	A=0.39 C=0.27 V=0.18 N=0.16	0.880
CVAM	C=0.37 A=0.31 M=0.16 V=0.16	0.879
TCVA	A=0.36 C=0.35 V=0.25 T=0.04	0.874
CVAP	A=0.41 C=0.36 V=0.20 P=0.03	0.873
CANM	A=0.45 C=0.30 M=0.14 N=0.11	0.859
CAPM	A=0.44 M=0.27 C=0.26 P=0.02	0.856

TCAM	C=0.40 A=0.33 M=0.21 T=0.06	0.853
CAPN	A=0.48 C=0.37 N=0.12 P=0.03	0.849
TCAN	A=0.50 C=0.24 N=0.18 T=0.07	0.848
TCAP	C=0.43 A=0.42 T=0.10 P=0.04	0.846
VANM	A=0.34 V=0.31 N=0.22 M=0.13	0.834
VAPN	A=0.52 V=0.28 N=0.17 P=0.04	0.829
TVAN	A=0.49 V=0.29 N=0.17 T=0.44	0.828
VAPM	V=0.38 A=0.36 M=0.23 P=0.04	0.825
TVAM	A=0.44 V=0.30 M=0.21 T=0.05	0.823
TVAP	A=0.58 V=0.31 T=0.07 P=0.05	0.820
CVPN	C=0.47 V=0.23 N=0.20 P=0.11	0.818
CVPM	C=0.44 V=0.27 M=0.23 P=0.06	0.810
TCVP	C=0.46 V=0.34 P=0.11 T=0.09	0.808
CVNM	C=0.31 M=0.28 V=0.25 N=0.17	0.808
TCVN	C=0.44 V=0.26 N=0.23 T=0.07	0.805
TCVM	C=0.38 V=0.30 M=0.25 T=0.06	0.805
APNM	A=0.50 M=0.25 N=0.21 P=0.04	0.791
TANM	A=0.47 M=0.28 N=0.17 T=0.08	0.789
TCNM	C=0.46 M=0.27 N=0.16 T=0.11	0.788
TCPN	C=0.57 N=0.22 P=0.11 T=0.11	0.785
TCPM	C=0.54 M=0.28 T=0.10 P=0.09	0.784
CPNM	C=0.48 M=0.27 N=0.18 P=0.07	0.783
TAPM	A=0.66 M=0.20 T=0.10 P=0.04	0.772
TAPN	A=0.64 N=0.20 T=0.09 P=0.07	0.771
VPNM	V=0.44 M=0.31 N=0.18 P=0.07	0.752
TVNM	V=0.42 M=0.30 N=0.24 T=0.04	0.747
TVPM	V=0.42 M=0.40 P=0.09 T=0.09	0.741
TVPN	V=0.50 N=0.28 P=0.14 T=0.08	0.719
TPNM	M=0.48 N=0.32 T=0.11 P=0.10	0.687

Table 3.31 LR model results for four-marker inputs.

Combination	Predictor Importance	AUC
AVMNC	A=0.32 C=0.24 V=0.19 N=0.13 M=0.12	0.883
VAPMC	A=0.38 C=0.27 V=0.21 M=0.12 P=0.01	0.882
AVPNC	A=0.35 C=0.32 V=0.21 N=0.10 P=0.02	0.881
AVTNC	V=0.38 A=0.30 C=0.20 N=0.09 T=0.03	0.880
CTAMV	A=0.35 C=0.31 M=0.16 V=0.16 T=0.03	0.878
CTAPV	A=0.38 V=0.31 C=0.23 T=0.04 P=0.03	0.874
ANPMC	A=0.40 C=0.28 N=0.15 M=0.15 P=0.02	0.859
CTAMN	A=0.41 C=0.28 N=0.13 M=0.13 T=0.05	0.859
CTAMP	A=0.35 C=0.34 M=0.25 T=0.05 P=0.01	0.854

CTAPN	A=0.41 C=0.37 N=0.14 T=0.05 P=0.03	0.848
AMNTV	V=0.35 A=0.34 N=0.16 M=0.13 T=0.02	0.835
VNAMP	A=0.39 V=0.24 M=0.19 N=0.15 P=0.03	0.833
TAPVN	A=0.42 V=0.30 N=0.20 P=0.05 T=0.05	0.828
VTAMP	A=0.48 V=0.26 M=0.18 T=0.04 P=0.03	0.824
CTPVN	C=0.44 V=0.25 N=0.19 P=0.07 T=0.06	0.815
VNPMC	C=0.31 V=0.29 M=0.22 N=0.14 P=0.04	0.812
CTMPV	C=0.44 M=0.23 V=0.21 P=0.08 T=0.05	0.811
TVMNC	C=0.35 V=0.24 M=0.20 N=0.18 T=0.03	0.809
AMNTP	A=0.51 N=0.21 M=0.20 T=0.05 P=0.03	0.789
TNPMC	C=0.42 M=0.26 N=0.20 P=0.06 T=0.05	0.788
TMPVN	V=0.44 M=0.24 N=0.19 P=0.08 T=0.05	0.748

Table 3.32 LR models for the five-marker inputs

Combination	Predictor Importance	AUC
CVAMNP	A=0.40 C=0.24 V=0.14 N=0.11 M=0.09 P=0.02	0.885
CTAMNV	C=0.28 V=0.22 A=0.22 M=0.14 N=0.13 T=0.01	0.884
CVMPTA	A=0.30 C=0.29 V=0.21 M=0.15 T=0.03	0.881
CVNPTA	A=0.32 V=0.26 C=0.24 N=0.11 T=0.04 P=0.03	0.880
CATMNP	A=0.35 C=0.34 M=0.15 N=0.12 T=0.03 P=0.01	0.858
TVAMNP	A=0.43 V=0.19 M=0.17 N=0.17 P=0.03 T=0.02	0.834
CVMNPT	C=0.37 V=0.26 M=0.18 N=0.13 P=0.03 T=0.02	0.811

Table 3.33: LR models for the six-marker inputs.

Combination	Predictor Importance	AUC
CTAMNVP	A=0.35 V=0.21 C=0.18 N=0.12 M=0.10 P=0.02 T=0.02	0.886

Table 3.34: LR models for the seven-marker input.
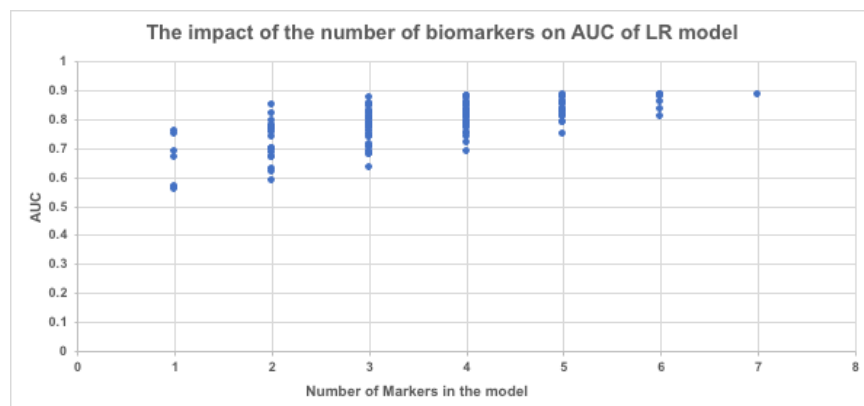
Figure 3.20: The effect of the number of biomarkers on AUC of LR model.

LR model AUC results for the 127 biomarker combinations (one to seven marker) for the UOL training dataset (n=106).
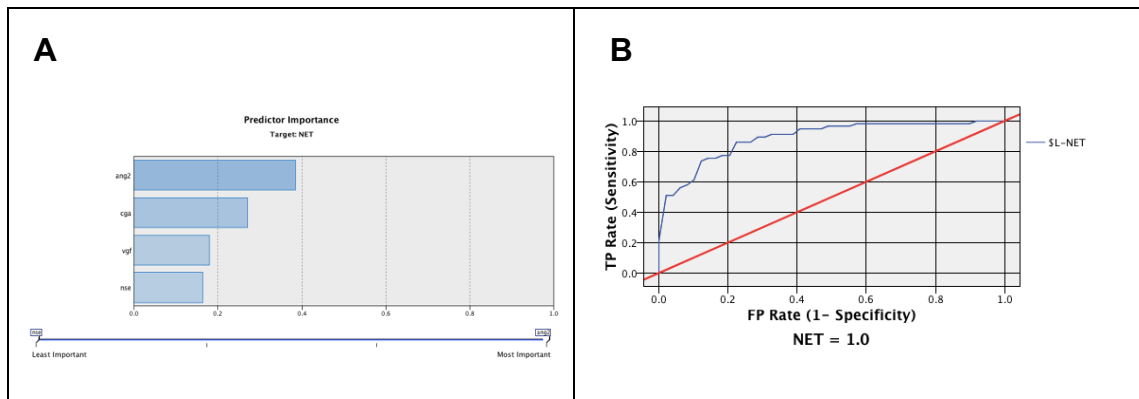
Figure 3.21: Predictor Importance and ROC curve analysis for the CVAN input.
(A) Predictor Importance results reveal that the most important markers for the model were A=0.39 C=0.27 V=0.18 N=0.16. (B) The ROC curve for the CVAN LR mode with an AUC of 0.880.
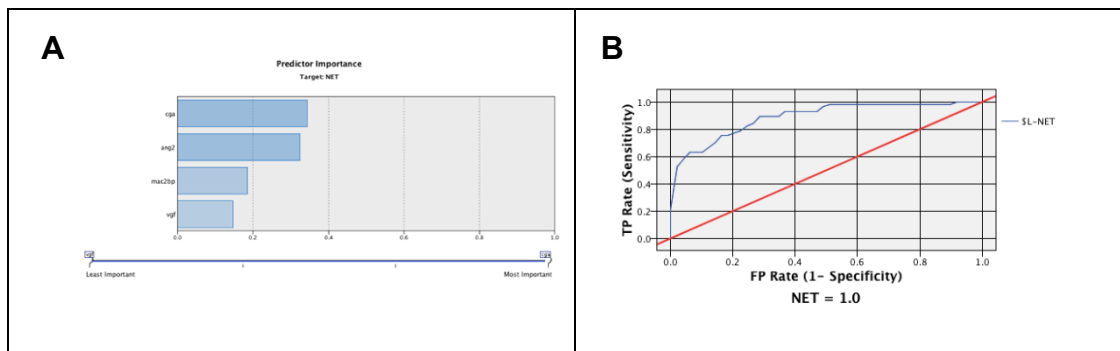
Figure 3.22: Predictor Importance and ROC curve analysis for the CVAM input.
(A) Predictor Importance results reveal that the most important markers for the model were C (0,34) and A (0.32). (B) The ROC curve for the CVAM LR model with an AUC of 0.879.
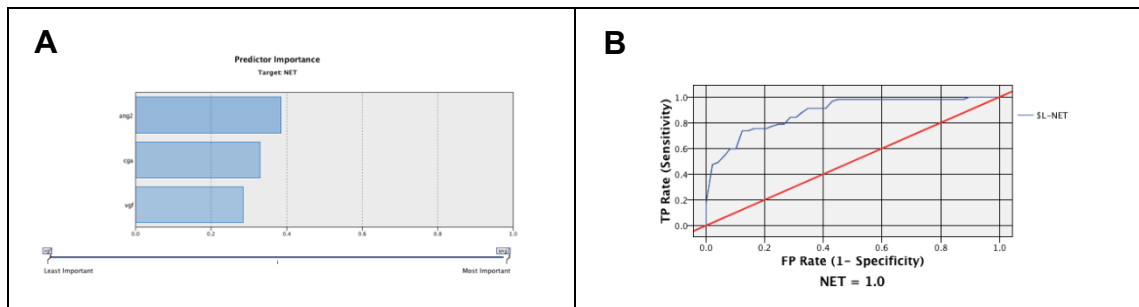
Figure 3.23: Predictor Importance and ROC curve analysis for the CVA input.
(A) Predictor Importance results reveal that the most important markers for the model were A=0.39. C=0.33 and V=0.29. (B) The ROC curve for the CVA LR model with an AUC of 0.873.
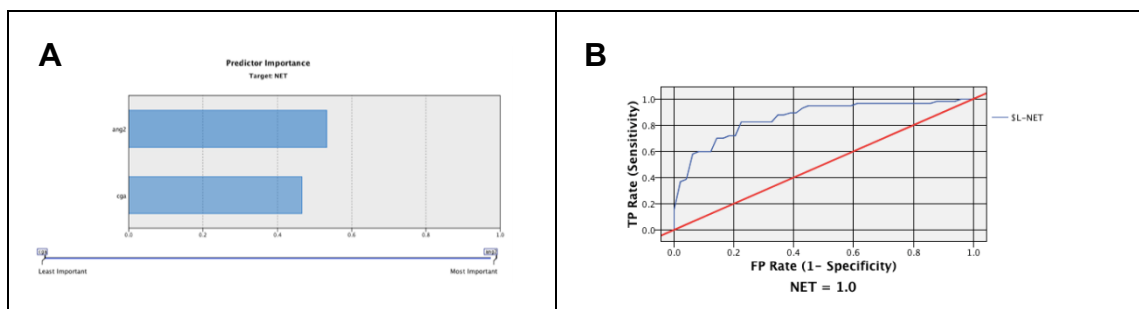
Figure 3.24: Predictor Importance and ROC curve analysis for the CA for input.
(A) Predictor Importance results reveal that the most important markers for the model were A=0.53 and C=0.47. (B) ROC curve for the CA LR model with an AUC of 0.849.
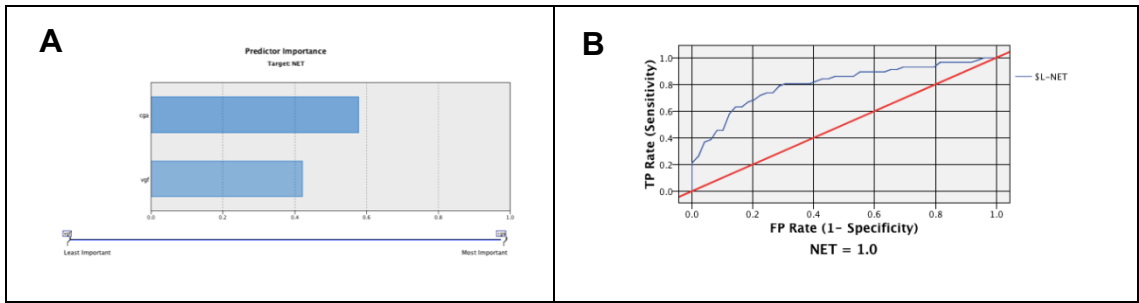
Figure 3.25: Predictor Importance and ROC curve analysis for the CV input.
(A) *Predictor Importance results reveal that the most important markers for the model were C=0.58 V=0.42.* *(B)* *ROC curve analysis reveals an AUC of 0.795 for the CV combination using LR.*
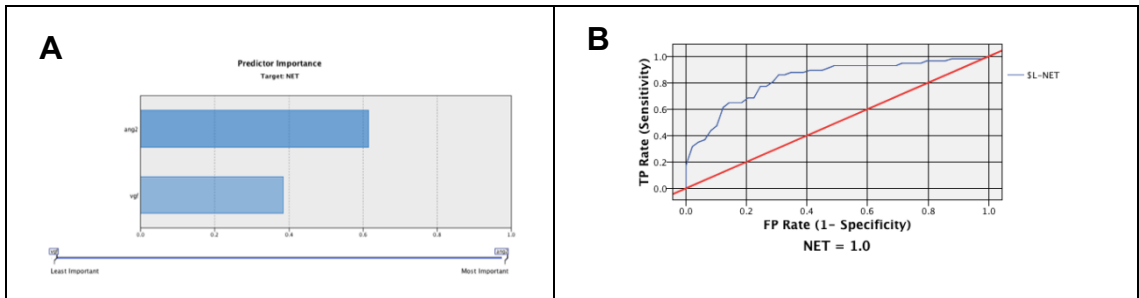
Figure 3.26: Predictor importance and ROC curve analysis for the VA input.
(A) *Predictor Importance results reveal that the most important markers for the model were A=0.62 V=0.38.* *(B)* *ROC curve analysis reveals an AUC of 0.819.*
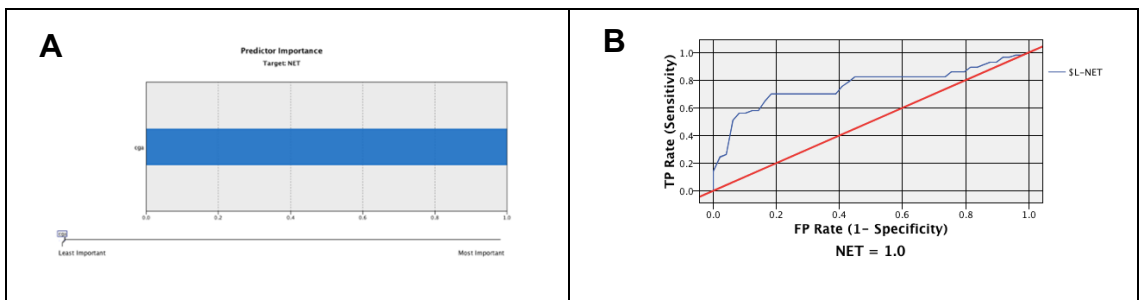
Figure 3.27: Predictor importance and ROC curve analysis for the C input.
(A) *Predictor Importance results reveal that C was 1.00.* *(B)* *ROC curve analysis reveals an AUC of 0.750.*
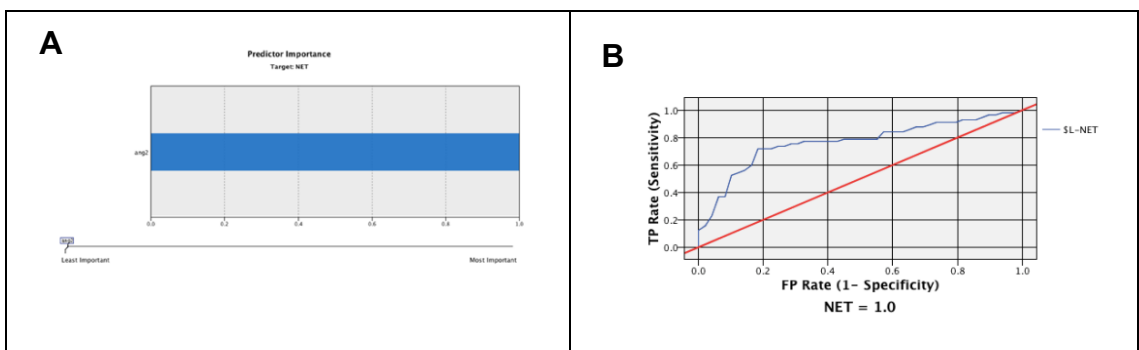
Figure 3.28: Predictor Importance and ROC curve analysis for the A input.
(A) *Predictor Importance results reveal that A was 1.00.* *(B)* *ROC curve analysis reveals an AUC of 0.758.*
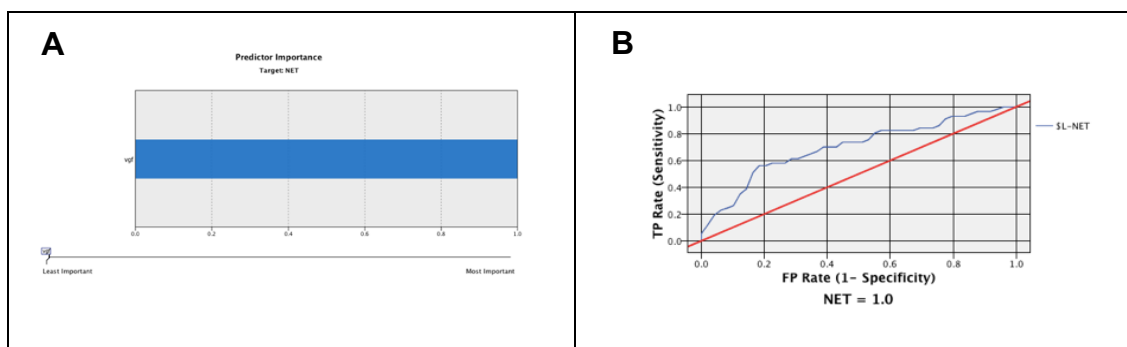
Figure 3.29: Predictor Importance and ROC curve analysis for the V input.
(*A*) *Predictor Importance results reveal that V was 1.00. (**B**) ROC curve analysis reveals an AUC of 0.691.*

Combination	Predictor Information	AUC
CTAMNVP	A=0.35 V=0.21 C=0.18 N=0.12 M=0.10 P=0.02 T=0.02	0.886
CVAMNP	A=0.40 C=0.24 V=0.14 N=0.11 M=0.09 P=0.02	0.885
AVMNC	A=0.32 C=0.24 V=0.19 N=0.13 M=0.12	0.883
CVAN	A=0.39 C=0.27 V=0.18 N=0.16	0.880
CVAM	C=0.34 A=0.32 M=0.18 V=0.15	0.879
CVA	A=0.39 C=0.33 V=0.29	0.873
CAM	A=0.44 C=0.35 M=0.21	0.856
VAM	A=0.46 V=0.33 M=0.21	0.823
CVM	C=0.39 M=0.31 V=0.29	0.802
CA	A=0.53 C=0.47	0.849
CV	C=0.58 V=0.42	0.795
VA	A=0.62 V=0.38	0.819
C	C=1.00	0.750
A	A=1.00	0.758
V	V=1.00	0.691

Table 3.35: Summary of LR model performance.
The CTAMNVP combination performed best with an AUC of 0.886 while the single marker V had the poorest performance with an AUC of 0.691.

3.10 McNemar test for model selection

To decide which ML models to take forward, the McNemar test was used for binary comparisons of model performance (Table 3.37). Detailed methodology for this is described in Section 2.9. When comparing two ML classifiers (classifier one and classifier two respectively), four different outcomes are present, firstly that classifier one and classifier two correctly classify, the second outcome is that classifier one correctly classifies but classifier two incorrectly classifies, the third outcome is that classifier one incorrectly classifies and classifier two correctly classifies. The final outcome is that classifier one and

classifier two both incorrectly classify. Crucial for the McNemar test is the sum of the B and C values. The B value refers to the number of correctly classified instances in the 1st model and the number of incorrectly classified instances in the second model, while the C value refers to the number of incorrectly classified instances in the 1st model and correctly classified instances in the 2nd model. The lower the B+C total the more similar are both models. Conversely, the larger the B+C value the greater the differences between the two models. The standard McNemar test can be used on larger B+C values but for instances where B+C < 25 the binomial corrected McNemar test is recommended[451]. For Linear SVM and RBF SVM it was seen that TB4, NSE and TIMP1 were not particularly important markers. Work from C5.0 decision trees also showed that NSE was not a significantly important marker. However, RF models consistently showed high performance across all marker combinations, even at the single-marker level. Thus, based on this rationale, NSE, TB4 and TIMP1 were excluded. This then led to a decision of an optimal four-marker combination for SVM would be the CVAM combination based on NSE, TB4 and TIMP1 being the least important markers at the seven and six-marker level. For LR however, the CVAN combination performed the best as a four-marker combination with an AUC of 0.880, however the CVAM combination performed similarly to the CVAN combination, thus the CVAM four-marker combination was used as a starting point to examine whether reduction to a three-marker CVA combination, two-marker (CV, CA or VA), or single-marker (A, C or V) input; caused a significant difference in model prediction performance by McNemar test. This would provide further rationale and support to reduce the marker combination from a four-marker combination to a three or two-marker panel.

	Linear SVM	RBF SVM	LR
CVAM	Linear SVM CVAM C=9 AUC= 0.867	RBF SVM CVAM C=10 gamma=0.80 AUC=0.880	Binomial LR model AUC=0.879
CVA	Linear SVM CVA C=4 AUC=0.863	RBF SVM CVA C=10 gamma=0.95 AUC=0.868	Binomial LR model AUC=0.873
CA	Linear SVM CA C=10 AUC=0.823	RBF SVM CA C=10 gamma=0.70 AUC=0.830	Binomial LR model AUC=0.849
VA	Linear SVM VA C=3 AUC=0.820	RBF SVM VA C=9 gamma=0.60 AUC=0.822	Binomial LR model AUC=0.819

CV	Linear SVM CV C=10 AUC=0.775	RBF SVM CV C=10 gamma=0.65 AUC=0.772	Binomial LR model AUC=0.795
C	Not Applicable	Not Applicable	Binomial LR model AUC=0.750
A	Not Applicable	Not Applicable	Binomial LR model AUC=0.758
V	Not Applicable	Not Applicable	Binomial LR model AUC=0.691

Table 3.36: Models used for McNemar testing.
Top performing Linear and RBF SVM models for the four-marker CVAM, three-marker CVA, two-marker CA, VA and CV combinations. LR models for these marker combinations are also shown.

The McNemar test produced B+C values of 3, 5 and 5 for the CVA vs CVAM Linear SVM, RBF SVM and LR models, respectively (Table 3.37). These low B+C values meant that only 3 instances, 5 instances and 5 instances, respectively, were dissimilar between the models, which suggested that these models were similar. This was supported by the P values of 1.00 (Table 3.37) for all three. Moreover, the results across these three models indicated that MAC2BP was a redundant marker in LR, Linear SVM and RBF SVM models and supported the rationale to decrease the marker panel from CVAM to CVA. When comparing the single-marker LR results vs the LR results obtained for the two-marker combinations, the B+C values were much higher than those of the three to four-marker comparisons, which meant that there were more dissimilarities in agreement of classification between the compared models, but the only comparison that was statistically significant was V vs VA (Table 3.37). Hence, the addition of ANG2 is significantly important for model performance.

Combination	Models	B+C	Result
CVA vs CVAM	Linear SVM CVAM C=9 Linear SVM CVA C=4	3	P=1.00 Not significant
	RBF SVM CVAM C=10 gamma=0.80 RBF SVM CVA C=10 gamma=0.95	5	P=1.00 Not significant
	LR	5	P=1.00 Not significant
CA vs CVA	Linear SVM CA C=10 Linear SVM CVA C=4	19	P=1.00 Not significant
	RBF SVM CA C=10 gamma=0.70 RBF SVM CVA C=10 gamma=0.95	19	P=1.00 Not significant
	LR	17	P=0.629 Not significant
VA vs CVA	Linear SVM VA C=3 Linear SVM CVA C=4	5	P=0.375 Not significant
	RBF SVM VA C=9 gamma=0.60 RBF SVM CVA C=10 gamma=0.95	5	P=0.375 Not significant

	LR	14	P=0.791 Not significant
CV vs CVA	Linear SVM CV C=10 Linear SVM CVA C=4	26	P=0.327 Not significant
	RBF SVM CV C=10 gamma=0.65 RBF SVM CVA C=10 gamma=0.95	26	P=0.327 Not significant
	LR	20	P=0.167 Not significant
C vs CV	LR	20	P=0.824 Not significant
A vs CV	LR	27	P=1.000 Not significant
V vs CV	LR	27	P=0.122 Not significant
C vs CA	LR	31	P=0.473 Not significant
A vs CA	LR	15	P=1.00 Not significant
V vs CA	LR	54	P=0.076 Not significant
C vs VA	LR	38	Not significant P=0.256
A vs VA	LR	17	P=0.629 Not significant
V vs VA	LR	35	P=0.041 Significant

Table 3.37: Binomial McNemar test results used to determine models to take forward for external validation.

158
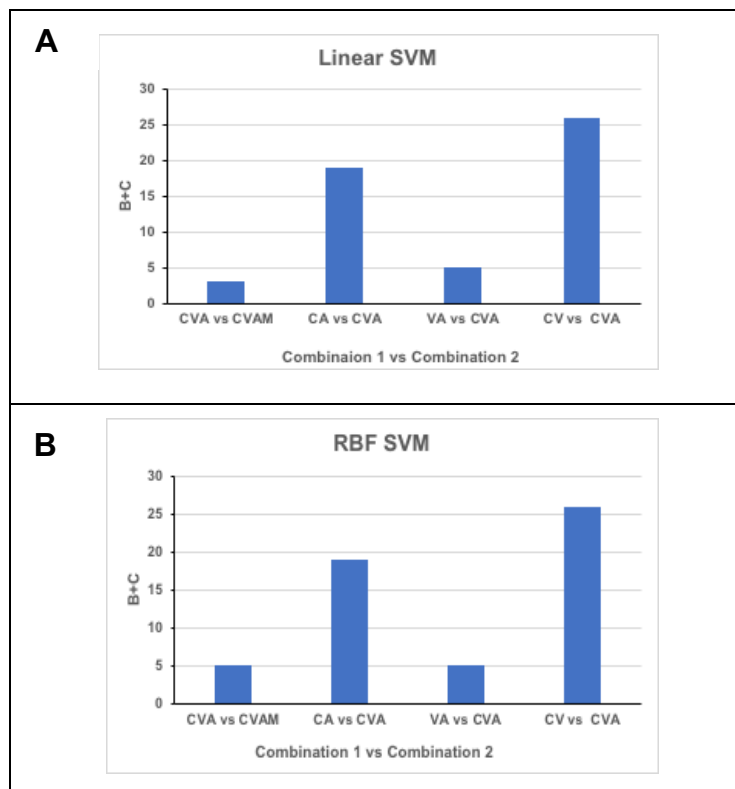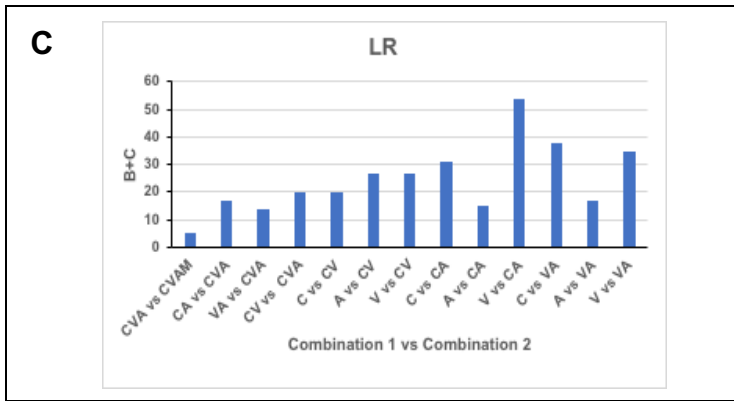
Figure 3.30:McNemar B+C results

Graphical illustration of B+C results from the McNemar test for the 4 marker CVAM, 3 marker CVA, 2 marker CV, CA and VA and 1 marker C,A and V. (A) Linear SVM, (B) RBF SVM and (C) LR.
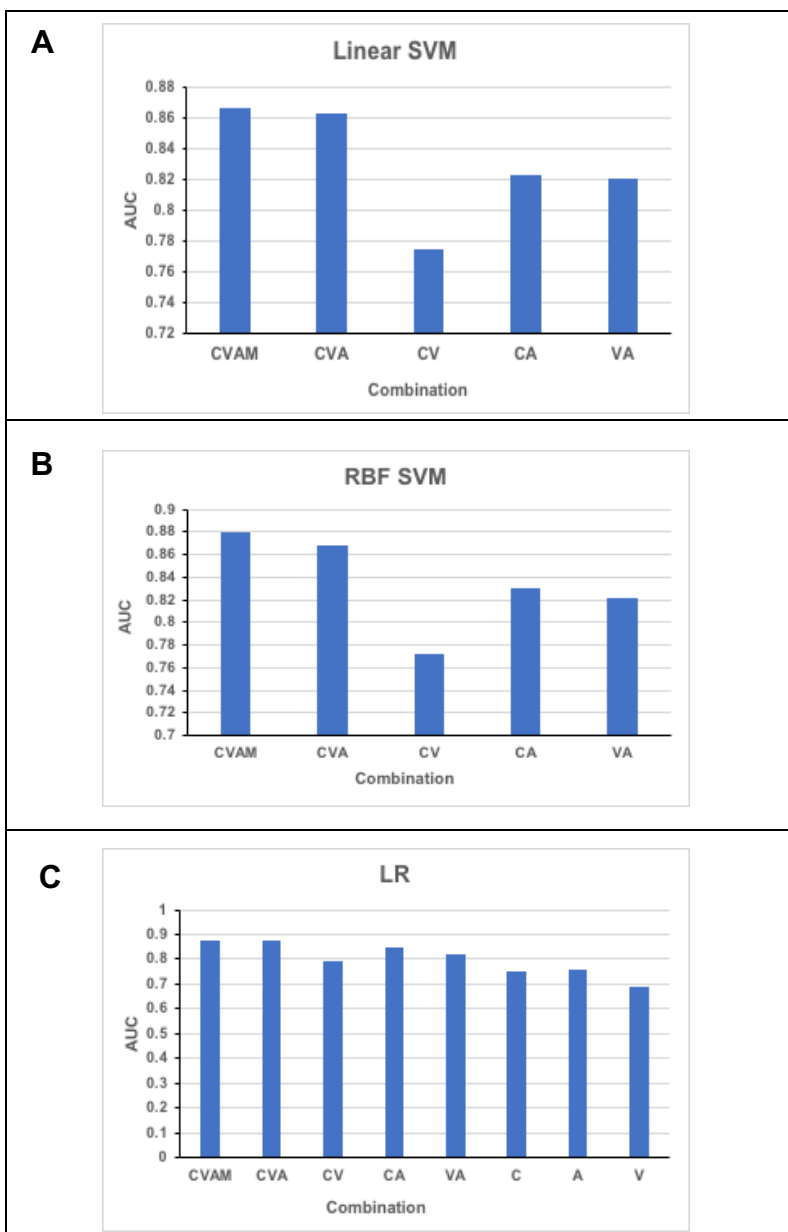
Figure 3.31: Impact of omitting markers from the four-marker CVAM combination on AUC

Graphical illustration of AUC results for the top performing four-marker CVAM, three-marker CVA, two-marker CV, CA and VA, and one-marker C,A and V models (A) Linear SVM, (B) RBF SVM and (C) LR.

3.11 CVA model comparisons

The confusion matrices obtained for the CVA ML models (Appendix C) were utilised to calculate the sensitivities and specificities of each model. A summary of these values is shown in (Table 3.38). For all the above-mentioned models utilising the CVA combination, apart from the RF and the Optimised RBF SVM model, some controls and cases were incorrectly predicted. The LR, Linear SVM and the Optimised Linear SVM models had the lowest sensitivities (74%), while the RF and Optimised RBF SVM had the best sensitivities (100%). The LR model had the lowest specificity (78%), while the RF and Optimised RBF SVM models had the best specificities (100%).

CVA model	Sensitivity	Specificity
LR	74%	78%
Linear SVM (C=4)	74%	80%
RBF SVM (C=10, gamma=0.95)	77%	84%
Optimised Linear SVM	74%	84%
Optimised RBF SVM	100%	100%
Random Forest	100%	100%

Table 3:38: Sensitivities and specificities of different CVA ML models.
The RF model (Run 1) and the Optimised RBF SVM model had the highest sensitivity and specificity of 100%.

3.12 Discussion

Four different ML approaches were applied to the UOL pNET case and control training dataset to obtain the best performing and most suitable pNET diagnostic test.

The best non-boosted C5.0 decision tree generated using either the general or accuracy settings produced an AUC of 0.888 based on the four markers ANG2, VGF, MAC2BP and TIMP1 (AVMP). This was followed in performance by a tree produced from five input markers (AMNTP) that used only four of these markers (AMPT) and had an AUC of 0.871 (Table 3.9). Accordingly, these results indicated the AVMP decision tree model should be taken forward for validation. The RF approach using one to seven-marker inputs (Tables 3.10 to 3.12) generated models that performed consistently well across all marker combinations with little discrimination in terms of AUC value between three, four, five, six and seven-marker combinations (Table 3.12 and Appendix B). This made it difficult to decide which marker combination to take forward. The

RF algorithm is a black box algorithm, which is a disadvantageous in practice as interpretability of a model that would be used as a clinical decision support tool is an important consideration in addition to performance. The very high performance even at the single marker level, with AUCs even at the single marker level reaching >0.99 is indicative of overfitting and would need to be tested at validation. Moreover, the very high performance of all RF models indicated this approach had suffered from widespread overfitting. Thus, it was important to validate a variety of RF models across the one to seven-marker range to assess the impacts of overfitting and thus these selected combinations would be explored using internal validation approaches in Chapter 4. Moreover, unlike with the other approaches, specific maker combinations were not selected to assess due to a similar performance across the combinations for a given number of markers.

SVM approaches were explored in SPSS Modeler. For Linear SVMs a C parameter range of between 1 and 10 was used and for RBF SVMs a C parameter range of 1 to 10 and a RBF gamma range of 0.4 to 0.95 were used. Grid optimisation was carried out, meaning that RBF SVM models were generated for all combinations of each C and gamma value within their respective ranges. A limitation of this approach is that C and gamma values outside these ranges were not considered. As a result, there was a possibility that the optimal SVM model was not identified. To address this limitation MCLA software was also used for SVM model development. MCLA allowed optimisation of both the Linear and RBF SVM generation via a wider range of gamma (0.001-1000) and C (0.001-1000) values. The MCLA software also automated the process of grid optimisation, which was a strength. Both Linear and RBF SVM were therefore explored using SPSS Modeler and MCLA. My results demonstrated that a reduction of markers used to generate SVM models did not initially have a large impact on performance. However, at the two-marker level, a reduction in performance was seen in the SPSS Modeler analysis and Optimised Linear SVM results (MCLA). However, the Optimised RBF SVM results had an AUC of 1.00 even at the two-marker level. As this very high level of performance could have been caused by overfitting to the UOL dataset, it was important to determine how well the Optimised RBF algorithm would perform using k-fold cross-validation (k-FCV) thus, k-FCV models for the seven-marker combination, six-marker combinations excluding NSE, TB4 and TIMP1,

the four-marker CVAM, the three-marker CVA combination and the two-marker CV, CA and VA combination will be assessed in Chapter 4.

LR models were created for all 127 combinations of markers. LR results revealed that the seven-marker combination CTAMNVP had the best AUC of 0.886 (Table 3.34). At the six-marker level the CVAMNP combination had an AUC of 0.885 and was the best performing six-marker combination (Table 3.33), thus illustrating that removal of TB4 did not influence model performance as reflected by AUC. The best five-marker combination was the AVMNC combination which had an AUC of 0.883 (Table 3.32). This suggested that TB4 and TIMP1 were not important for model performance. The best four-marker combination was the CVAN combination with an AUC of 0.880 followed by the CVAM with an AUC of 0.879.The LR performance decreased as any one of the four markers was removed (Table 3.31). These LR results clearly support my initial hypothesis that a marker-panel approach is more accurate for pNET detection compared to a single-marker approach.

The McNemar test was utilised for LR and SVM models to assess statistically whether the number of markers in the best performing combination (CVAM) could be reduced to a three-marker (CVA) or even a two-marker combination (CV, CA, or AV) with no significant loss of performance despite the decrease in AUC. The McNemar test results revealed that a reduction of markers did not lead to a statistically significant difference between SVM models as determined by p value (Table 3.37). However generally, as the marker number decreased model B+C values increased. Similar test results were obtained for comparisons of LR models with the exception that a statistically significant difference was obtained between the V and VA models and the B+C value for the VA vs CVA SVM (Linear and RBF) comparisons was low (Table 3.37). Based on the changes in B+C and AUC values (Figure 3.30 and Figure 3.31), reducing the CVAM panel to CVA was justified, but further reduction to two markers or a single marker was not justified due to the marginal decrease in AUC and increase in B+C combined with the cost saving and technical simplification resulting from the removal of Mac2BP assays.

Clinical use of the single CgA marker for pNET detection has been explored in several studies and found to be limited. Qiao et al. (2014) found that the diagnostic accuracy of CgA for insulinoma pNETs was 0.724 compared to 0.805 for non-insulinomas[90]. Indeed, a study by Hong et al. (2020) found that CgA

discriminated non-insulinoma pNETs from non-pNET patients with a remarkably high AUC of 0.897[461]. Their study found that a non-insulinoma subtype and the presence of liver metastases were associated with elevated serum CgA. Pulvirenti et al. (2019), who compared CgA levels in patients with well differentiated pNETs, other pancreatic conditions and healthy controls, concluded that CgA functions as a poor diagnostic marker for pNETs due to low sensitivity[462]. Their CgA test to distinguish pNET patients from healthy controls had an AUC of 0.77, and they also found that treatment with PPIs increased CgA levels in pNET patients, indicating compromised specificity. Moreover, serum CgA levels were increased in just 6/27 pNET patients with localised disease (Stage 1 to 3) compared to 33/38 patients with stage 4 disease. Hence clinical characteristics of study cohorts are also important considerations. For example, if the cohort had an increased number of late stage pNETs or a lack of insulinomas CgA would be expected to perform relatively well as a marker. However, CgA seems to have predictive value 6 months prior to radiological progression prediction in pNET patients, highlighting that CgA as a marker has utility in certain circumstances[99].

Many studies have suggested that CgA alone is inadequate as a clinical NET or pNET marker and that a wider panel of markers alongside CgA is needed to improve pNET patient identification, particularly for pNETs of lower grade, earlier stage and subtypes such as insulinomas. Despite CgA being used in practice for diagnosis, prognosis, response to treatment and surveillance of recurrence, the evidence is inconsistent from mainly retrospective studies which are small and based on heterogenous populations undergoing many treatments[463]. Moreover, the drawbacks associated with CgA such as PPI usage and other confounders may be offset via a multiple-biomarker approach. The models generated here using combinations of markers had improved performance across all four ML algorithms compared to the single CgA marker currently in clinical use. Results from my work show that CgA as a single marker yielded an AUC of 0.75 in an LR algorithm, 0.813 in a C5.0 non-boosted decision tree algorithm, and an average of 0.989 across 10 runs of an RF algorithm, which was likely to have been overfitted. Superior levels of performance resulted when CgA was included with additional markers, indicating non-redundancy. Taking the CVA input as an example, the LR model had an AUC of 0.873, the best Linear SVM model had an AUC of 0.863, and

the best RBF SVM model had an AUC of 0.868. The Optimised Linear SVM for CVA had an AUC 0.88 and the Optimised RBF SVM for CVA had an AUC of 1.00, although both were likely to have been overfitted. Finally, the C5.0 non-boosted decision tree produced when CVA was entered had an AUC of 0.835. These results illustrate the strength of a multi-marker approach as opposed to a single marker approach across supervised ML algorithms. However, for the RF, C5.0 decision tree and SVM increasing the markers in the panel above a certain point ceased to increase the AUC value, most likely because of functional redundancy between markers above that point. However, this was not seen with LR, as with LR increasing the number of markers in the panel led to an increase in performance with the more markers added to the model bettering performance with the seven marker LR model having the best performance with an AUC of 0.886 (Table 3.34).

Regarding the clinical characteristics of the UOL cohort, it contained a total of 57 pNET cases and 50 controls. Within the pNET cases 29 there were females, 27 were males and 1 was unknown. Within the UOL controls there were 27 females and 23 males. Within the pNET cases there were 4 confirmed insulinomas, 3 confirmed gastrinomas. 4 patients with MEN1 disease, and 22 idiopathic pNETs. Therefore, the number of confirmed insulinomas, gastrinomas and MEN1 related pNETs were similar and insulinomas do not make up a large portion of the group. Data on the epidemiology of insulinomas in the modern era is rare[464], however a recent study from Finland reported an increasing incidence of insulinomas from 0.5 cases per million/year in the 1980s to 0.9 cases per million/year in the 2000s[465]. pNETs arise in 30-80% of patients with MEN1[466]. The worldwide incidence of gastrinomas is about 0.5 to 3 cases/million per year[467]. As for CgA assays it has been shown that their performance is poor for insulinomas[90] it is likely that only a few cases would be impacted based on the UOL cohort numbers of insulinomas. UOL TNM staging was available for only 19 out of the 57 pNETs and thus was a limitation of the cohort. Lymph node involvement was specified for only 24 out of the 57 pNETs within the cohort and 11 of these specifying no lymph node involvement and 13 specifying some lymph node involvement. Vascular Invasion was specified for 17 out of the 57 pNETs. Within this 7 were yes, an additional 2 were IV, 7 no and 1 unknown. Perineural invasion was specified for 18 out of the 57 pNETs. 6 were yes, 10 no, 1 unknown and 1 n/a.

In terms of the grade of pNET, 23 were well differentiated, 1 was poorly differentiated, and grade information was unavailable for 33 patients. As high grade is known to be a predictor of increased CgA levels[468], within this cohort there was only one poorly differentiated tumour, however as grade information was unavailable for 33 patients it is unclear as to the numbers of these patients that may also have had a CgA level due to a higher grade, and thus this is a limitation. 47 patients did not have diabetes, 7 patients had diabetes (type undefined) and one patient had type-1 diabetes. Several pNET patients in the cohort had multiple tumours. A main limitation for this UOL cohort was that the clinical profile was incomplete for many of the patients. However, it was clear that the UOL pNET cohort was diverse in terms of tumour grade and type of pNET. Thus, the cohort mirrored the heterogeneity of pNETs that are most encountered in clinical practice.

Combining CgA with other markers has been reported in previous studies. A study by Kovesdi et al. (2020) showed that a combination of CgA and miRNAs in a regression model resulted in a higher AUC value compared to those obtained using individual markers[469]. The comparative and combined utility of cocaine and amphetamine regulated transcript (CART), CgA and chromogranin B (CgB) for NET diagnosis and distinguishment between stable and progressive disease in different NET substyles were looked at using multiple logistic regression analysis. Least significant variables were omitted form the model using a backwards selection procedure. Measuring CgA in combination with CgB significantly improved diagnostic accuracy in gut NETs and NETs with an unknown primary origin, with CgB being the best diagnostic marker for pNETs with no diagnostic advantage in also measuring CgA or CART. For distinguishing between controls (n=50) and pNETs (n=117), the LR model had an AUC of 0.78 using CART, CgA and CgB which reduced to 0.77 when using just CgB and CART which were identified as the most significant markers. Moreover, a combination approach for distinguishing stable pNETs (n=116) and progressive pNETs (n=36) revealed an AUC of 0.86 for the 3 markers (CART, CgA and CgB) however the reduced model using the most significant markers (CgA and CART) revealed that CgB did not contribute anything further to the model as the AUC remained at 0.86[445]. This suggests that CgA may be a later stage marker compared to CgB due to CgB being in the optimal model for distinguishing pNET cases from controls but not for distinguishing stable and

progressive disease. Within my optimal CVA marker combination, VGF-NGF has not been explored as a serum pNET marker previously, but ANG2 has been explored[140,470]. A study by Melen Mucha found that ANG2 was levels were elevated in metastatic disease compared to local disease[140]. Moreover, an additional study found that serum ANG2 levels were elevated in patients with metastatic disease compared to those with localised disease without metastasis which in turn was higher than healthy controls[470]. Time to disease progression was worse for patients with serum ANG2 levels >4756pg/ml. The study also found no significant difference in ANG2 levels when comparing patients with functional tumours and non-functional tumours. Additionally, ANG2 serum levels were no different between the 29 patients who were on SST analogues and the 18 patients who were not. However, ANG2 has not been explored previously in a purely pNET cohort which made the analysis of the UOL pNET cohort for this marker unique.

Based on the findings from the UOL training dataset the CVA combination was optimal relative to other marker combinations in terms of performance and marker number. However, it was important to validate how well this marker combination, as well as other well-performing combinations, using internal and external methods. This was because overfitting and underfitting are both important phenomena. It was likely that the best RBF SVM models which tended to have higher C and gamma values produced using the training dataset were overfitted to the UOL dataset. Likewise, RF models that performed extremely well could also have been overfitted, which required assessment by validation. Validation approaches such as k-fold, split set and external validation are explored in Chapters 4 and 5.

Chapter 4 - Validation of pNET biomarker panel and algorithms

4.1 Introduction to validation

When ML models are created using a whole training dataset as described in Chapter 3, a limitation to this approach is not knowing the extent to which these models are overfitted to the dataset from which they are derived. Overfitting occurs when a model is highly specific to the dataset used to create it and are unable to generalise well to other datasets[471]. The converse of overfitting is underfitting. Underfitting is when a model lacks some of the relevant assumptions or complexity in order to accurately reflect the data. Thus, an underfitted model will make inaccurate predictions from new data. An underfitted model will also display high bias and low variance, whereas an overfitted model will display low bias and high variance. Bias measures the error of the model compared to the true underlying model whereas variance measures how much the model would change if it were given different training data. In ML models there is typically a trade-off between bias and variance. Ultimately a ML model should be detailed enough to capture the pattern but simple enough not to be excessively specific to the training data[471]. Overfitting to a certain extent is expected and a modest drop in performance due to a marginal level of overfitting and differences between datasets due to their small size. It is known that smaller datasets with a are more likely to overfit, and are less likely to represent the population in question[472]. However larger drops in performance would indicate too much overfitting for real world applications and thus such models would not be taken forward. Thus, future work would need to explore refining the best models based on larger datasets.

Thus, validation approaches for ML algorithms are important in order to assess the effect of overfitting and the generalisability of ML models. Different validation options have been used in ML biomarker studies, including the train and test approach[473], k-FCV[474], a type of k-FCV known as Leave-one-out cross validation (LOOCV)[475] and external validation[476]. These approaches are explained in Chapter 1, Section 1.7.2.6.

In this chapter, different validation approaches were taken to explore the pNET UOL case and control data. First, a train and test approach entailed splitting the UOL case-control cohort into two separate datasets, one for training and one for

validation. The training dataset was employed to produce models that were tested using the validation dataset. Using separate data to test a ML model provides an unbiased estimate of the expected performance of the model when deployed in real world situations[433]. Train and test analysis was carried out for LR (Section 4.2), C5.0 decision trees (Section 4.3), RF (Section 4.4) and SVM (Section 4.5), using SPSS Modeler.

The second type of validation explored in this chapter was k-FCV. For 5FCV k=5, for 10FCV k=10 and for 20FCV k=20. For a 10FCV, the data are divided into 10 folds with one-fold used for validation whilst the model is trained using the remaining subset. This process is repeated 10 times until each subset is used exactly once for validation. In a 5FCV and a 20FCV the data are divided into 5 or 10 folds, respectively. 5FCV, 10FCV and 20FCV were carried out for the Optimised Linear and Optimised RBF SVM utilising MCLA (Section 4.7). Optimised Linear and Optimised RBF SVM approaches were carried out in Chapter 3 using MCLA, and thus k-FCV for the same type of ML algorithm. k-FCV is very economical as it allows the use of all the data for training to be re-used for validation[433].The default of 10FCV was carried out for the C5.0 decision trees produced from the seven-marker combination (Section 4.6). If validation were to be carried out with a separate dataset, twice the amount of data would be needed to provide the same amount of data for both training and validation[433]. However, validating a model using data that was also used to develop the model can inflate performance estimates[433]. Hence, cross validation is used when data are limited. Cross validation typically gives more accurate out of sample error estimation compared to the other validation approaches[433]. Out of sample error is a measure of how well the model would perform on new data or a test set. By carrying out validation using the two different approaches of train and test as well as k-FCV allowed all four types of ML approaches to be further evaluated.

External validation approaches use a separate dataset[433]. The advantage of using unseen data to test a ML model is that is can give a more unbiased estimate of performance in real-world situations[433]. External validation using an independent pNET RFH cohort is explored in Chapter 5.

4.2 Train and Test approach for Logistic Regression

A train and test approach was taken to assess the extent of overfitting when utilising the LR algorithm for the UOL cohort. The UOL dataset (n=106) was divided into two separate datasets with one containing the training data and another containing the validation data. The training dataset was used to create the model, with the validation dataset used to evaluate the model. Detailed methodology for this approach in SPSS Modeler can be found in Section 2.6.3. The aim of this analysis was to assess the extent of overfitting for the models using these markers and to compare their performance in the training and validation portions. The marker combinations assessed in this analysis were CVAM, CVA, CA, CV and AV, as well as the single markers C, V and A.

4.2.1 Train and Test analysis of the CVAM LR model

An LR model was created using the training portion (n=53) and tested using the validation portion (n=53) of the CGA, VGF-NGF, ANG2 and MAC2BP data. The resulting LR model showed that VGF-NGF was the most important marker from the predictor importance, followed by CGA, then ANG2 and finally MAC2BP (Figure 4.1A). The performance of the model as reflected by AUC values was 0.891 in the training portion compared to 0.839 in the validation portion (Figure 4.1B and C). This modest drop in performance (0.052 units) indicated the model had not been overfitted to the training dataset.
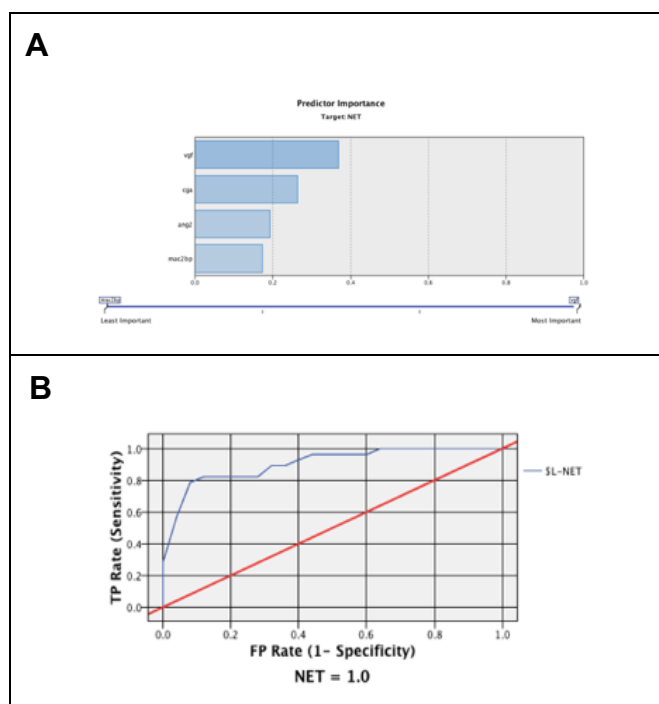
Figure 4.1: **Results obtained for the CVAM combination LR model using the Train and Test approach in SPSS modeler**. **(A)** Predictor Importance results for the model with V=0.37, C=0.26, A=0.19 and M=0.17. **(B)** ROC curve obtained for the Training portion (n=53) with an AUC of 0.891. **(C)** ROC curve obtained for the Validation portion (n=53) with an AUC of 0.839.

C

NET = 1.0
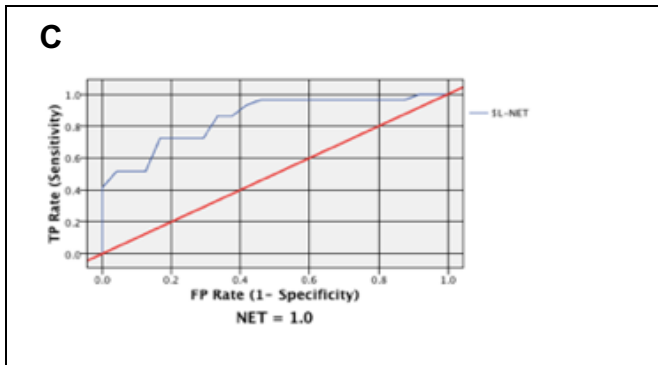
4.2.2 Train and Test analysis of the CVA LR model

Results for this model's predictor importance analysis based on the training dataset showed that VGF-NGF was the most important marker followed by ANG2 and then CGA (Figure 4.2A). The AUC obtained for the training portion was 0.894 compared to 0.849 for the validation portion (Figure 4.2B and C). As for the CVAM model, a modest drop in performance (0.045 units) indicated the CVA model had not been overfitted to the training dataset.
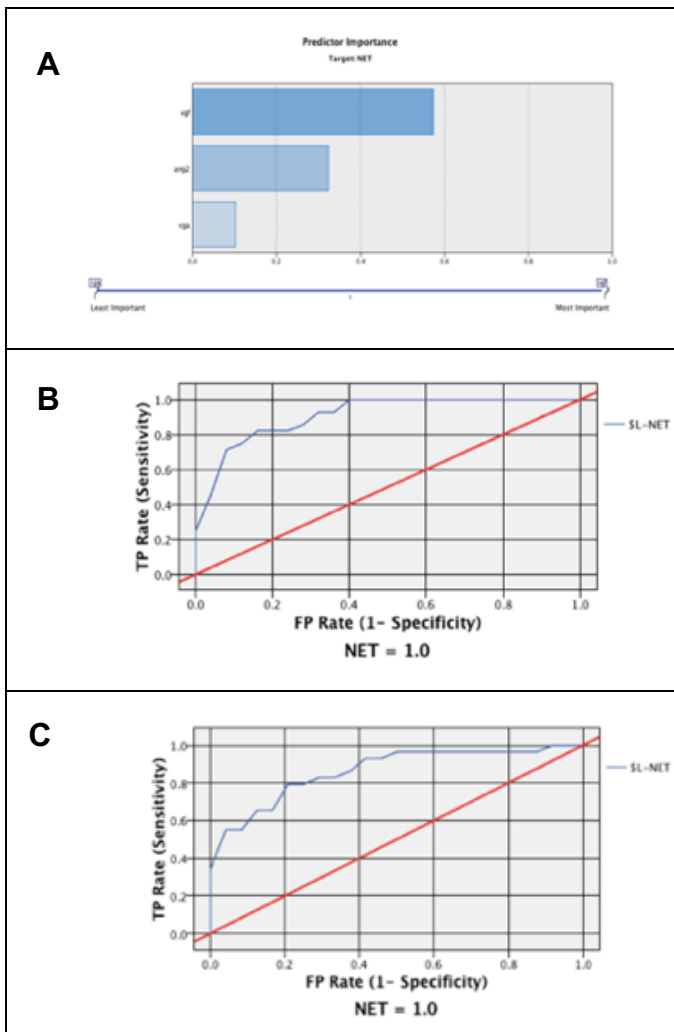
Figure 4.2: Results obtained for the 3 marker CVA combination LR model using the Train and Test approach in SPSS modeler. Training portion (n=53) and Validation portion (n=53). **(A)** Predictor Importance results for the model with V=0.57, A=0.32 and C=0.10. **(B)** ROC curve obtained for the Training portion (n=53) with an AUC of 0.894. **(C)** ROC curve obtained for the Validation portion (n=53) with an AUC of 0.849 .

4.2.3 Train and Test analysis of the CV, CA and AV LR models

The results of the Train and Test analysis of the three different two-marker combinations of CGA, VGF and ANG2 are shown in Figures 4.3 to 4.5. As expected, the performance of each model was higher using the validation data compared to using the training data. The CA model produced the best AUC of 0.870 for the training portion, and this dropped 0.048 units to 0.822 for the validation portion (Table 4.1). The VA model had the poorest performance in the training stage with an AUC of 0.822, which dropped only 0.005 units to 0.817 for the validation stage. The CV combination had an intermediate AUC of 0.850 for the training stage, however this dropped to the lowest AUC of 0.724 at the validation stage. This large drop (0.126 units) suggested that the CV model had been overfitted to the training data more than the other two-marker models. In addition, the combinations of AC and CV performed better in training than the VA combination suggesting the presence of CgA improves LR model performance.
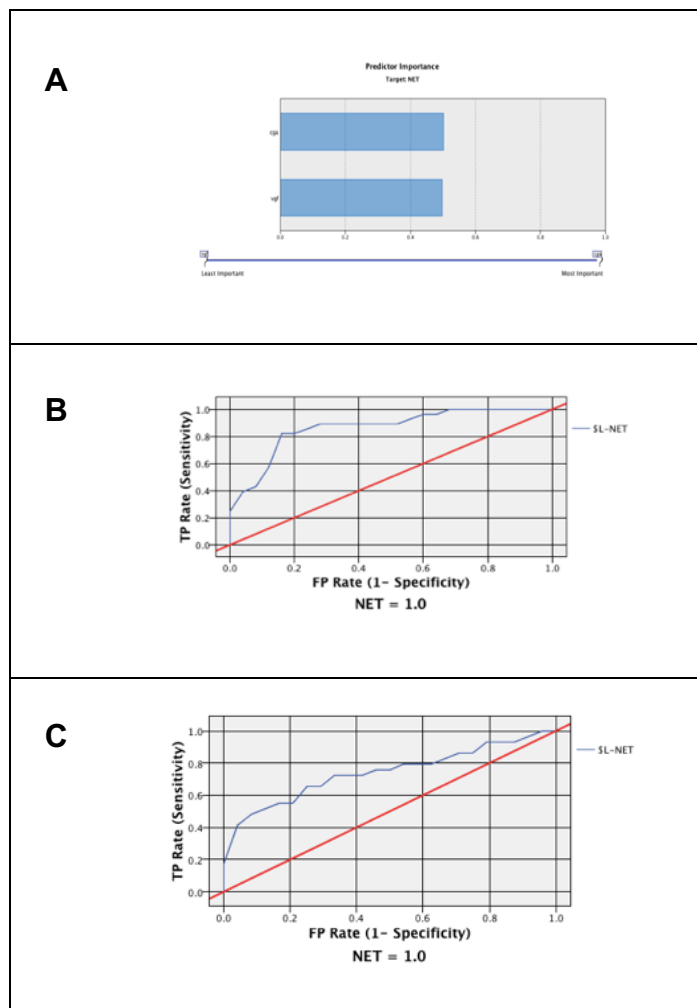
Figure 4.3: Results obtained for the 2 marker CV combination LR model using the train and test approach in SPSS modeler. Training portion (n=53) and Validation portion (n=53). **(A)** Predictor Importance results for the model with C=0.50 and V=0.50. **(B)** ROC curve obtained for the Training portion (n=53) with an AUC of 0.85. **(C)** ROC curve obtained for the Validation portion (n=53) with an AUC of 0.724.
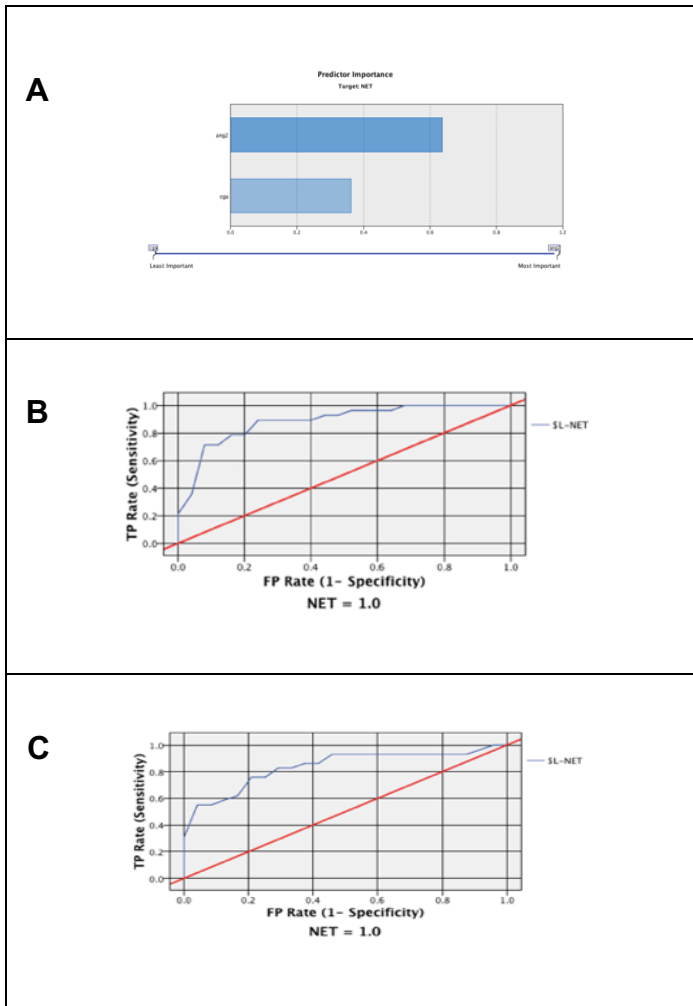
Figure 4.4: Results obtained for the 2 marker CA combination LR model using the train and test approach in SPSS modeler. Training portion (n=53) and Validation portion (n=53). **(A)** Predictor Importance results for the model with A=0.76 and C=0.24. **(B)** ROC curve obtained for the Training portion (n=53) with an AUC of 0.87. **(C)** ROC curve obtained for the Validation portion (n=53) with an AUC of 0.822
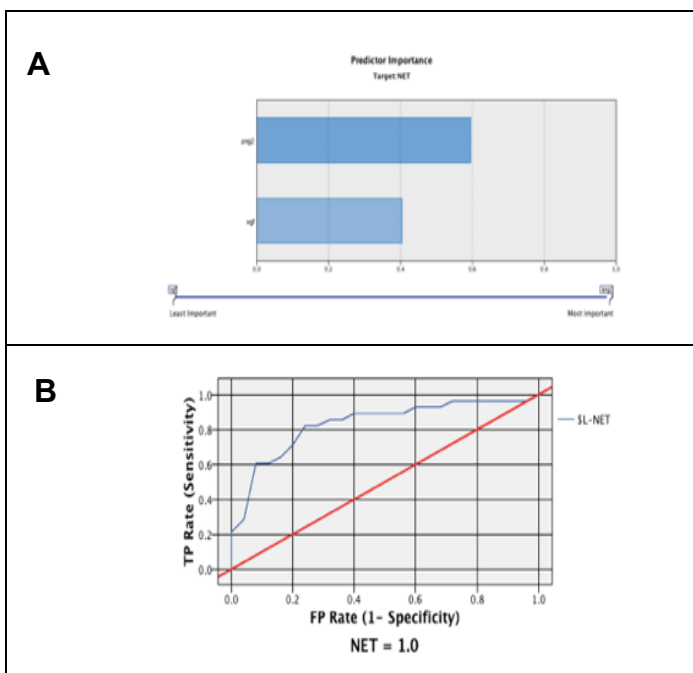
Figure 4.5: Results obtained for the 2 marker VA combination LR model using the train and test approach in SPSS modeler. Training portion (n=53) and Validation portion (n=53). **(A)** Predictor Importance results for the model with A=0.59 and V=0.41. **(B)** ROC curve obtained for the Training portion (n=53) with an AUC of 0.822. **(C)** ROC curve obtained for the Validation portion (n=53) with an AUC of 0.817.

NET = 1.0
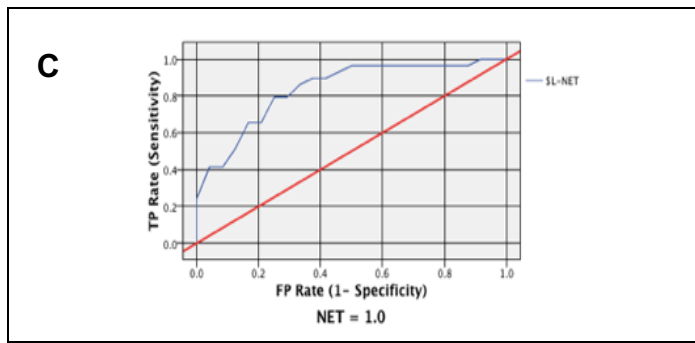
4.2.4 Train and Test analyses of the C, V and A LR models

The results from the training set analysis revealed that the best performing single marker was CGA with an AUC of 0.797 (Table 4.1), while the poorest performing single marker was VGF with an AUC of 0.770. In the validation stage ANG2 performed the best with an AUC of 0.761, while VGF performed the poorest with an AUC of 0.644. ROC curves for the single-marker analyses confirmed that the training stage performance was higher than the validation stage (Figures 4.6 to 4.8). Hence highlighting the overfitting of these models.
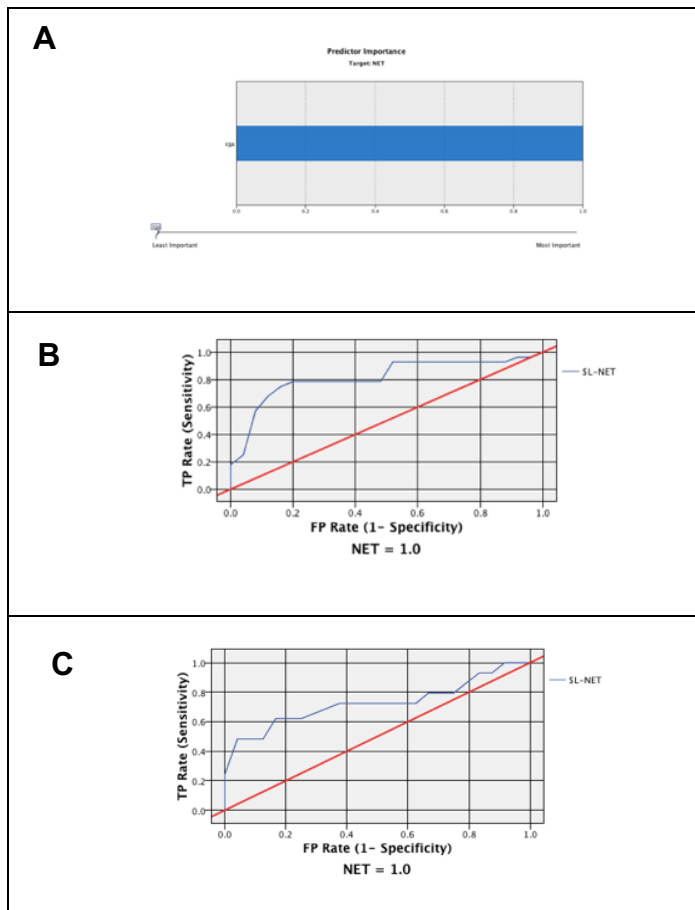
Figure 4.6: Results obtained for the C single marker LR model using the train and test approach in SPSS modeler. Training portion (n=53) and Validation portion (n=53). **(A)** Predictor Importance results for the model with C=1.00. **(B)** ROC curve obtained for the Training portion (n=53) with an AUC of 0.797. **(C)** ROC curve obtained for the Validation portion n=53 with an AUC of 0.711.
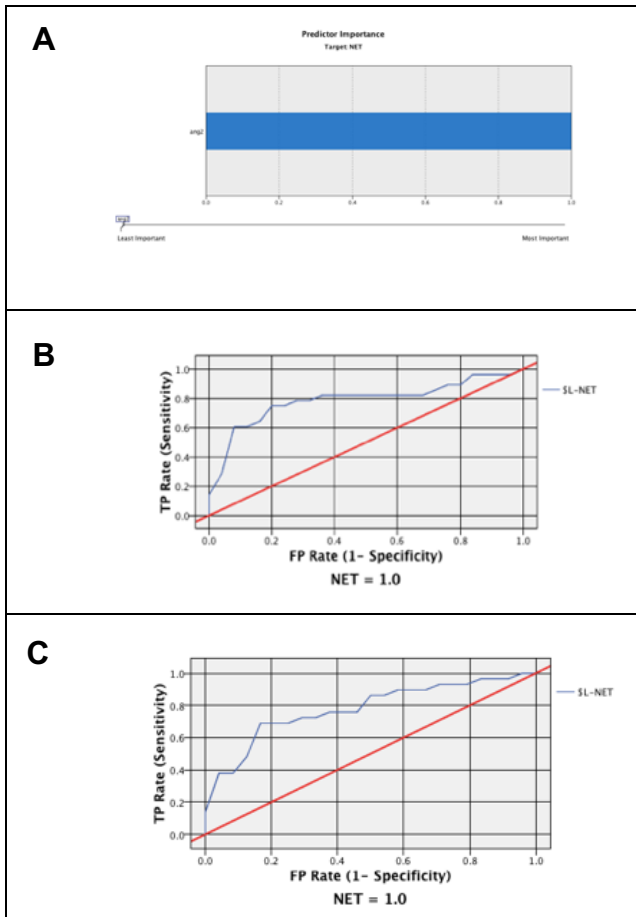
Figure 4.7: Results obtained for the A single marker LR model using the train and test approach in SPSS modeler. Training portion (n=53) and Validation portion (n=53). (A) Predictor Importance results for the model with A=1.00. **(B)** ROC curve obtained for the Training portion (n=53) with an AUC of 0.770. **(C)** ROC curve obtained for the Validation portion (n=53) with an AUC of 0.761.
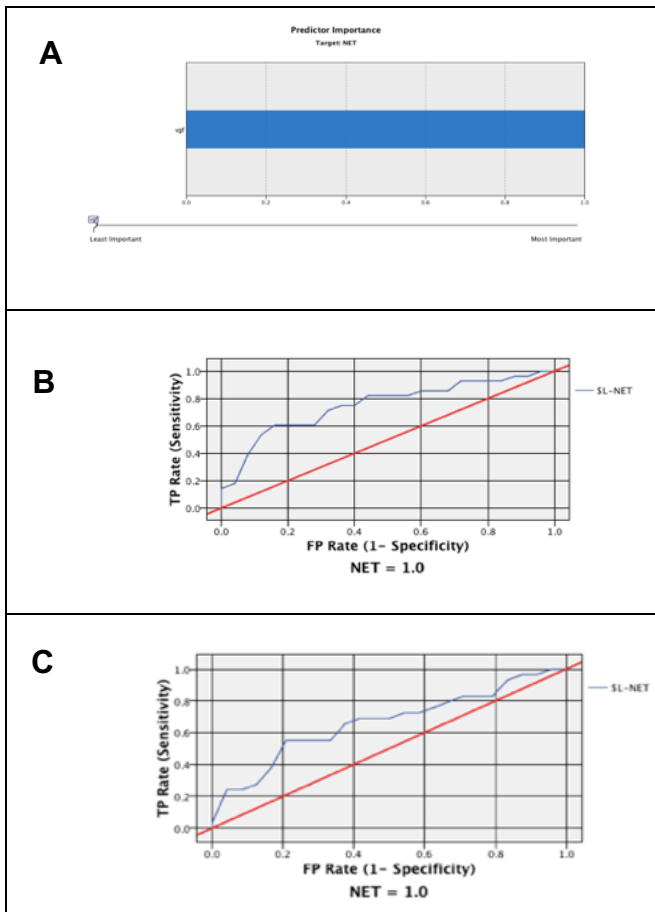
Figure 4.8: Results obtained for the V single marker LR model using the train and test approach in SPSS modeler. Training portion (n=53) and Validation portion (n=53**). (A)** Predictor Importance results for the model with V=1.00. **(B)** ROC curve obtained for the Training portion (n=53) with an AUC of 0.736. **(C)** ROC curve obtained for the Validation portion (n=53) with an AUC of 0.644.

	Training portion AUC	Validation portion AUC	Difference in AUC (%)
CVAM	0.891	0.839	0.052 (6)
CVA	0.894	0.849	0.045 (5)
CV	0.85	0.724	0.126 (15)
VA	0.822	0.817	0.005 (1)
CA	0.870	0.822	0.048 (6)
C	0.797	0.711	0.086 (11)
A	0.770	0.761	0.009 (1)
V	0.736	0.644	0.092 (13)

Table 4.1: LR Train and Test results.
The training portion (n=53) and validation portion (n=53) results are shown for the selected marker combinations.

4.3 Train and Test approach for C5.0 decision trees

Non-boosted general and accuracy decision tree models were obtained from the previous dataset (n=107) analysis and the top five models were identified (Chapter 3). A Train and Test approach was then taken to assess the degree of overfitting of these five non-boosted C5.0 decision trees. To allow comparison with the Train and Test analysis of other supervised ML algorithms, the entry with the single missing ANG2 value was removed and the resulting dataset (n=106) used to produce the training and validation datasets. The training portion consisted of 25 controls and 28 cases, while the validation portion consisted of 24 controls and 29 cases. A detailed methodology for this approach can be found in Section 2.6.1. The VAMP C5.0 non-boosted general decision tree model is described in Section 4.3.1, the VAMP C5.0 accuracy non-boosted model is described in Section 4.3.2, the general AMNTP C5.0 non-boosted model is described in Section 4.3.3, TAPM C5.0 accuracy non-boosted model described in Section 4.3.4 and general CM entered model described in Section 4.3.5.

4.3.1 Train and Test analysis of the VAMP general non-boosted C5.0 decision tree

The VAMP general non-boosted C5.0 decision tree was previously identified as one of the top five models in the whole dataset analysis. Thus, this marker combination was reviewed using train (n=53) and test (n=53) subsets to assess overfitting and performance. The C5.0 non-boosted decision tree obtained from the training subset revealed that all four markers (VGF, ANG2, MAC2BP and TIMP1) were used in the model (Figure 4.9). The ROC curves
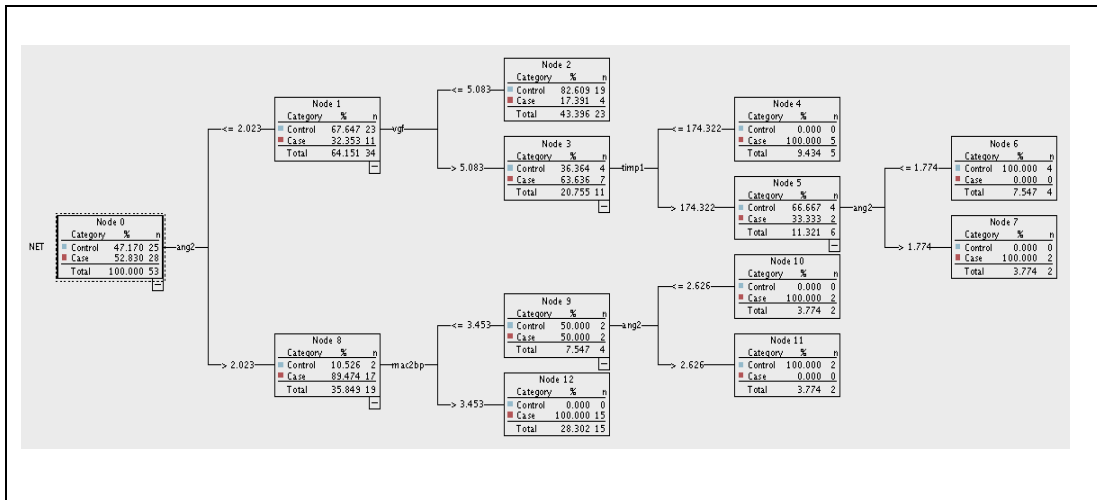
Figure 4.9: VAMP C5.0 non-boosted decision tree.
The schematic illustrates the tree obtained from the training subset (n=53) when the VAMP markers were entered and AVPM were used for model creation in the Train and Test analysis**.**

obtained from the training and validation subsets had AUC values of 0.934 and 0.691, respectively (Figure 4.10). The large difference (0.243 units) indicated strong overfitting of this decision tree model.
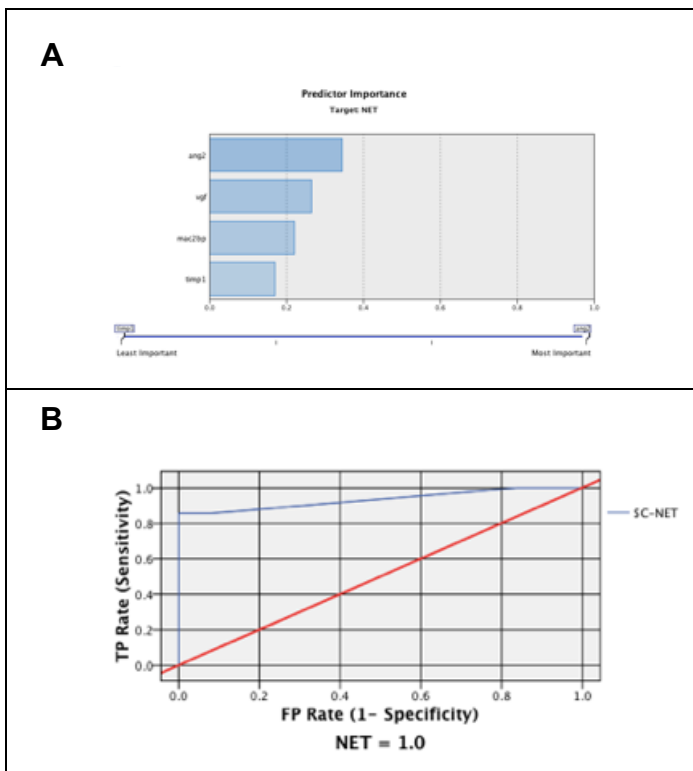
Figure 4.10: Train and Test approach for the VAMP combination using the C5.0 non-boosted general decision tree.
(A) Predictor Importance A=0.34 V=0.26 M=0.22 and P=0.17**. (B)** ROC curve obtained for the training portion with an AUC of 0.934 (n=53). **(C)** ROC curve obtained for the validation portion with an AUC of 0.691 (n=53).

C
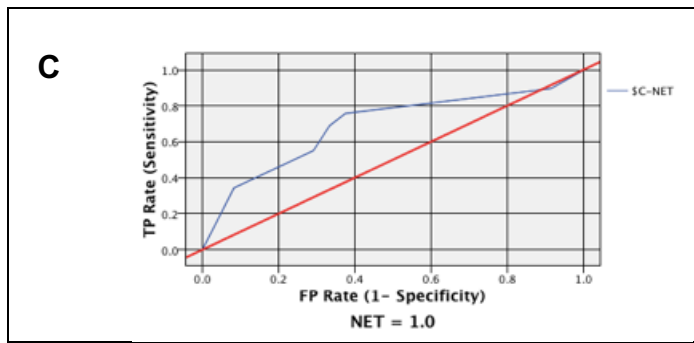
4.3.2 Train and Test analysis of the VAMP accuracy non-boosted C5.0 decision tree

The VAMP combination of markers also produced a C5.0 non-boosted accuracy decision tree model as one of the top five models produced using the whole dataset, and with results differing from those of the general setting model. A train (n=53) and test (n=53) assessment of the accuracy model revealed that all four markers were used (Figure 4.11). Moreover, the C5.0 non-boosted decision tree produced from the training set was identical to that obtained using the general setting (Figure 4.9). Therefore, the general or accuracy setting did not have any effect on the train and test approach, as seen for the whole dataset analysis in Chapter 3. Consequently, the ROC curves and AUCs obtained for the training and validation stages were the same as the corresponding values obtained for the general C5.0 non-boosted decision tree (Figure 4.10). The large difference in AUC of 0.243 units suggests that the model was overfitting. This also suggests that this combination of markers may not be suitable to take forward.
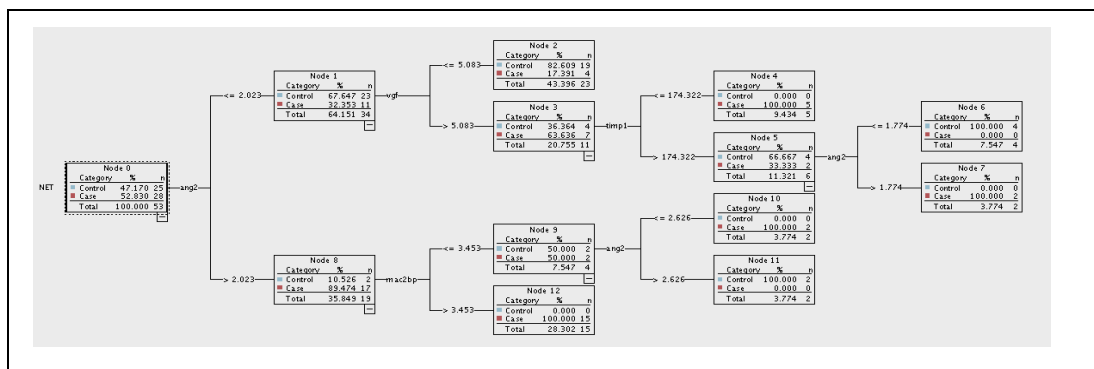
Figure 4.11: Train and Test approach for the VAMP combination using the C5.0 non-boosted accuracy decision tree C5.0 created using the training portion of the split dataset n=53 for when VAPM markers were entered and AVMP were used.
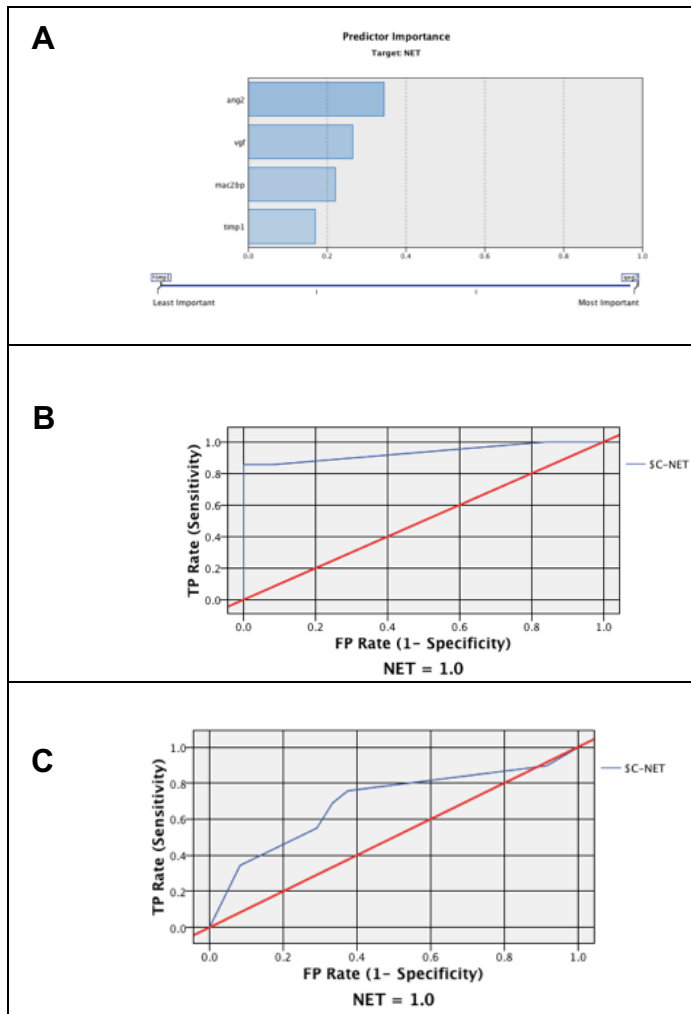
Figure 4.12: Train and Test approach for the VAMP combination using the C5.0 non-boosted accuracy decision tree. (A) Predictor Importance A=0.34 V=0.26 M=0.22 and P=0.17. **(B)** ROC curve obtained for the training portion with an AUC of 0.934 (n=53). **(C)** ROC curve for the validation portion with an AUC of 0.691 (n=53).

4.3.3 Train and Test analysis of the AMNTP general C5.0 non-boosted decision tree

The AMNTP C5.0 non-boosted decision tree was also identified as a top decision tree model (Chapter 3) and a train and test analysis of this marker combination was therefore explored. The decision tree obtained from the training subset (Figure 4.13) employed only three of the five entered markers. Predictor importance results showed that ANG2 was the most important marker in the model (Figure 4.14). The difference in AUC was 20%. This large difference illustrated that the decision tree model was overfitting and again questioned the suitability of taking forward this marker combination and type of algorithm forward based on these results.
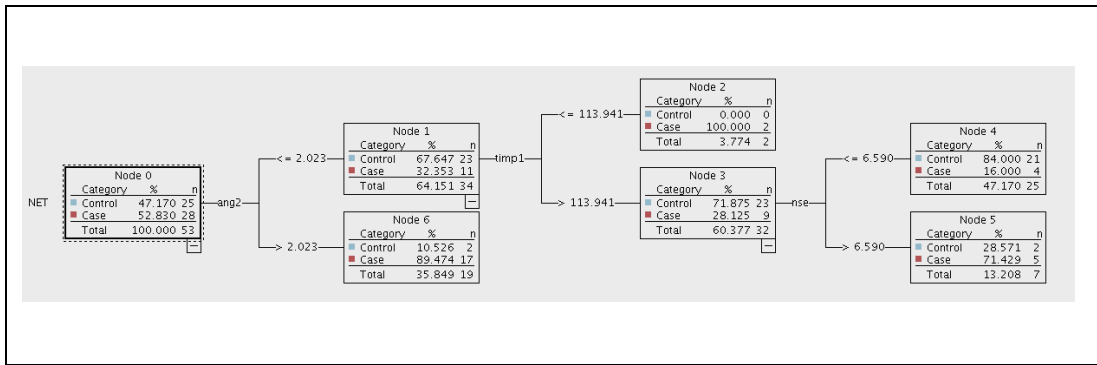
Figure 4.13: Train and Test approach for the C5.0 non-boosted general decision tree model created for the training portion (n=53) of the dataset when the AMNTP combination was entered, and APN were used.
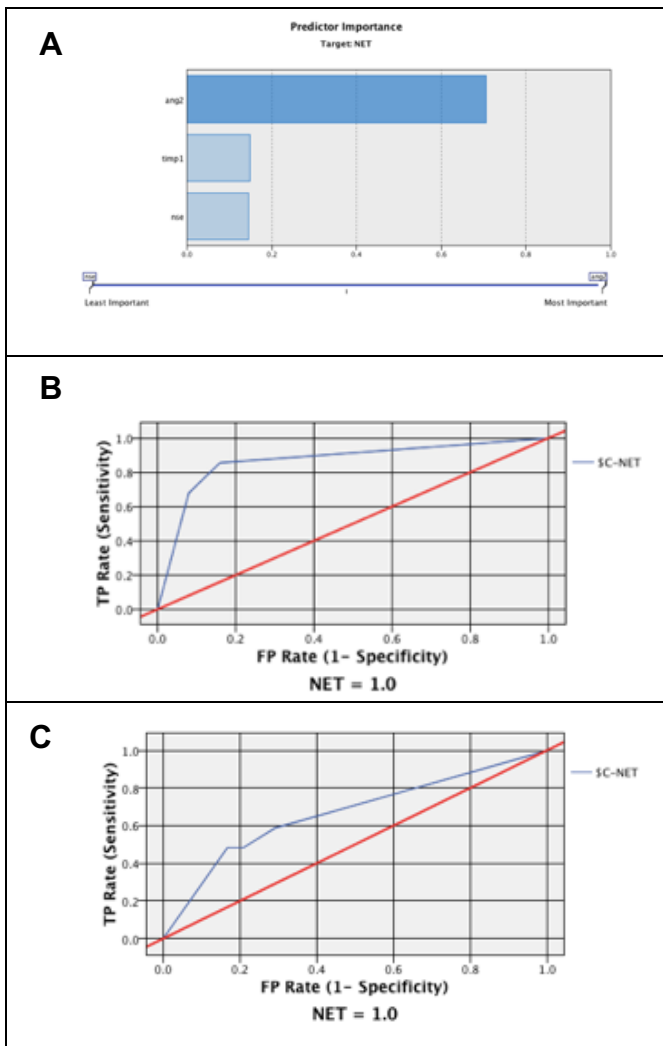
Figure 4.14: AMNTP entered APN used C5.0 non-boosted general decision tree using the train and test approach. **(A)** Predictor Importance A=0.71 P=0.15 N=0.15 **(B)** ROC curve for the training portion (n=53) with an AUC of 0.866. **(C)** ROC curve for the validation portion with an AUC of 0.667 (n=53).

4.3.4 Train and Test analysis of the TAPM accuracy C5.0 non-boosted decision tree

The C5.0 non-boosted accuracy decision tree produced from TAPM data was also identified as one of the top five models in Chapter 3, and a train and test analysis of these markers was therefore performed. The C5.0 decision tree obtained using the training portion (n=53) shows that only ANG2 and TIMP1

179

were used in this model despite four markers being entered (Figure 4.15). Predictor importance analysis showed that ANG2 had greater importance than TIMP1 in the model. The ROC curves for this model for the training and validation portions are shown in Figure 4.16. The difference in AUC was 15% and was thus lower than seen with other combinations looked at.
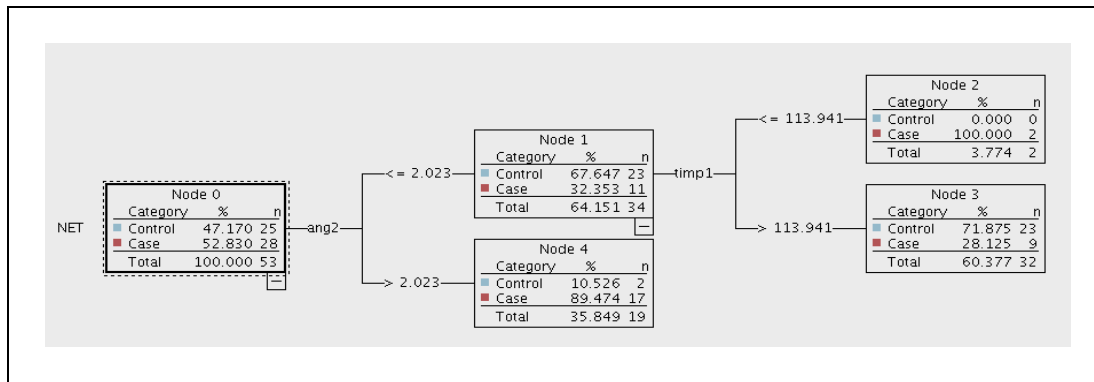
Figure 4.15: TAPM Train and Test C5.0 non-boosted accuracy decision tree.
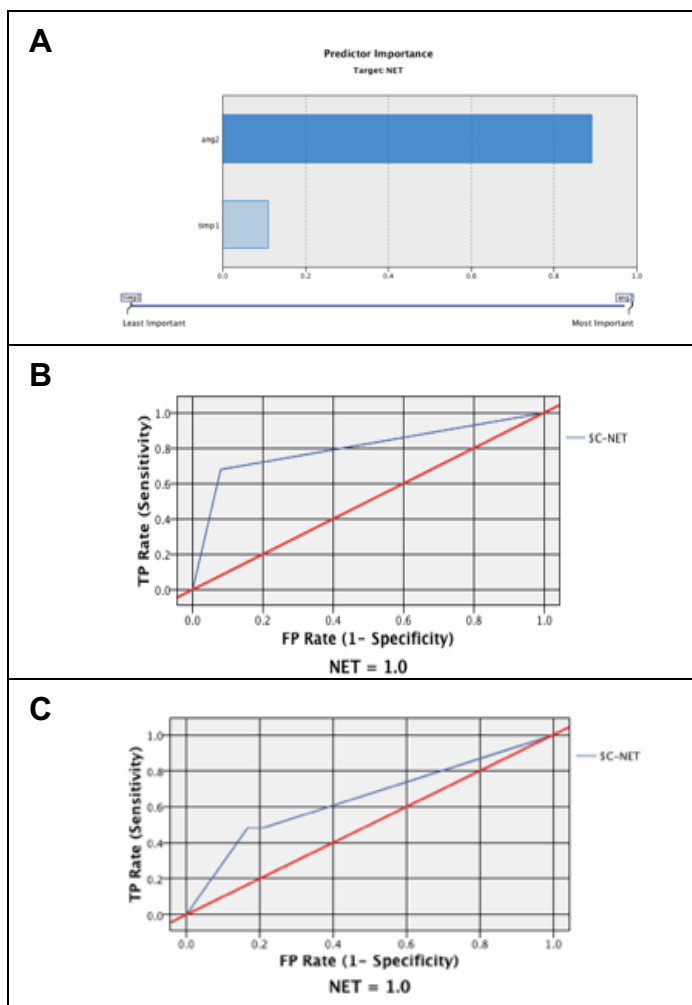
Figure 4.16: C5.0 non-boosted accuracy decision tree using a train and test approach when TAPM were entered and AP were used. **(A)** Predictor Importance analysis A=0.89 and P=0.11 **(B)** ROC curve for the training portion (n=53) with an AUC of 0.796. **(C)** ROC curve for the validation portion (n=53) with an AUC of 0.647.

4.3.5 Train and Test analysis of the CM general C5.0 non-boosted decision tree

Finally, a C5.0 non-boosted decision tree model was created from the CM combination. The model revealed that only CgA was used. Moreover, this model performed better in the training portion compared to the validation portion with AUCs of 0.795 and 0.693 respectively, a difference of 0.102 units (Table 4.2). The difference in AUC of 10% was modest and not as high as seen with other marker combinations.
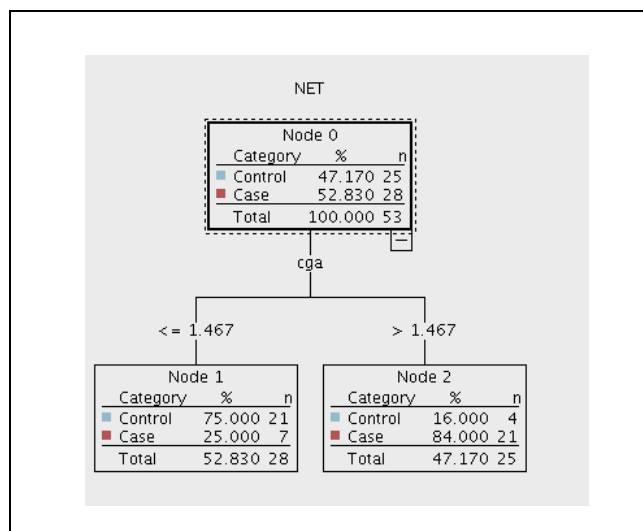
Figure 4.17: Train and Test C5.0 non-boosted general decision tree produced using CM training data entry, and only C was used (n=53).
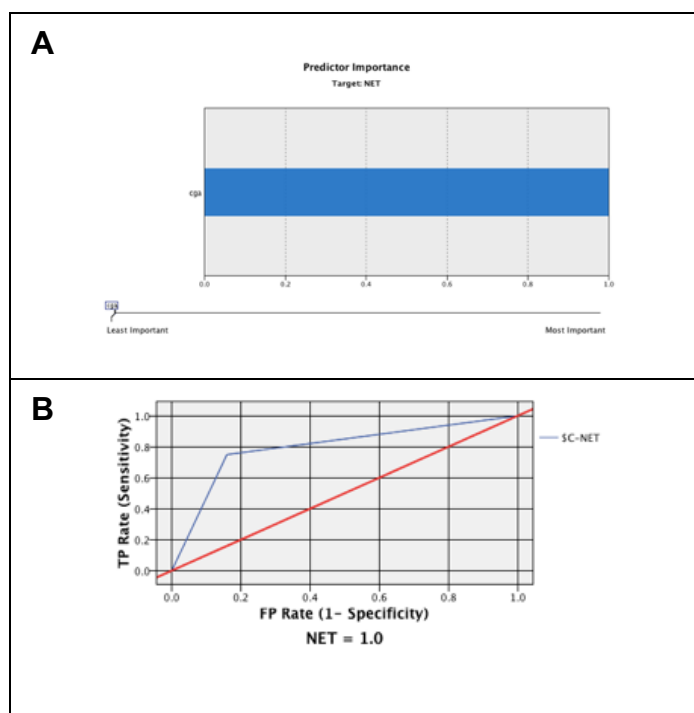
Figure 4.18: Train and Test analysis of the CM C5.0 non-boosted general decision tree.
(A) Predictor Importance C=1.00 **(B)** ROC curve with an AUC of 0.795 for the Training portion (n=53). **(C)** ROC curve with an AUC of 0.693 for the Validation portion (n=53).

C

NET = 1.0
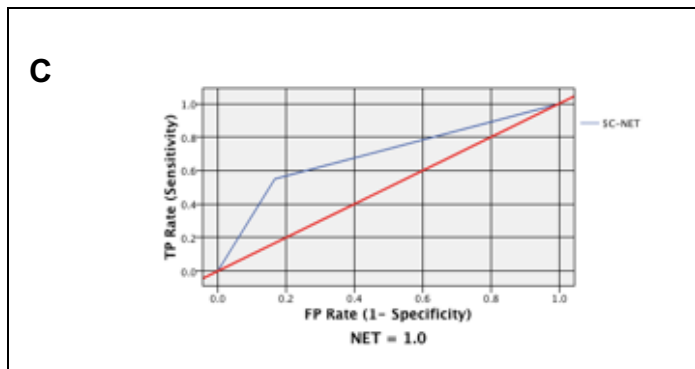
4.3.6 Summary of C5.0 non-boosted decision trees using the train and test approach

The train and test analysis of the top five C5.0 non-boosted decision trees identified in Chapter 3 revealed that the worst performing model at the training stage was produced from CM data and utilised just the C marker. The best performing model in the training stage was VAPM which used all four markers and had an AUC of 0.934 (Table 4.2). However, in the validation stage this dropped to 0.691, a large decrease of 0.243 (26%), indicating substantial overfitting.

Moreover, differences between the whole training dataset approach as described in Chapter 3 and the train and test approach for the top non-boosted C5.0 decision trees (Table 4.2) are seen for the different combinations entered. For the VAPM entered combination the same markers are used in both the whole training dataset analysis and the train and test approach. However, the predictor of importance for the markers vary, with the model created in the whole training dataset analysis the order of importance is APMV whereas for the train and test approach analysis is it AVMP and so VGF is of greater importance in the train and test compared to TIMP1. Interestingly, the AUC obtained for the training portion for the train test approach was higher than that seen when using the whole training dataset, with this being 0.864 (accuracy), 0.888 (general) and 0.934 (general and accuracy train and test). This could be attributed to the smaller size of the dataset used to build the model (n=106) vs (n=53) (for the training portion).For the AMNTP combination, for the whole training dataset the markers used were AMPT with an AUC of 0.871, this was higher than that seen for the test and train analysis for when this combination was entered. The model created when the AMNTP combination was entered

also differed in the markers that were used in the model, with just APN being used in the model. The AUC was also lower, with this being 0.866 compared to 0.871. This could perhaps be attributed to fewer markers being used to build the model in the train and test approach compared to the whole training dataset (3 vs 4). For the TAPM combination for the whole training dataset analysis, all four markers were used with an AUC of 0.867. However, for the train and test approach only two out of the four markers were used A and P and the AUC obtained for the training portion was 0.796 which was lower. Again, this could be attributed to fewer markers being used to build the model for the train and test approach. Finally, for the CM combination for the whole training dataset analysis, both C and M were used for model building and an AUC of 0.850 was obtained. However, for the train and test approach when both these markers were entered, only C was used for model building with an AUC of 0.795 in the training portion. The lower AUC seen when the same markers are entered in these two different approaches could be attributed to two. markers being used in the model when built on the whole training dataset vs when the model was built using just the training portion.

Markers entered	Markers used	Predictor Importance	Training portion AUC	Validation portion AUC	Difference in AUC (%)
VAPM	AVPM	A=0.34 V=0.26 M=0.22 P=0.17	0.934	0.691	0.243 (24)
VAPM	AVMP	A=0.34 V=0.26 M=0.22 P=0.17	0.934	0.691	0.243 (24)
AMNTP	APN	A=0.71 P=0.15 N=0.15	0.866	0.667	0.199 (20)
TAPM	AP	A=0.89 P=0.11	0.796	0.647	0.149 (15)
CM	C	C=1.00	0.795	0.693	0.102 (10)

Table 4.2: Summary of the 5 C5.0 non-boosted decision tree models examined using the train and test approach.

4.4 Train and Test assessment of Random Forest models

The RF models across all combinations performed very well and thus overfitting was a possible explanation, as described in Chapter 3. A detailed methodology for the train and test assessment of RF models is described in Section 2.6.4. Briefly, a single run was carried out for one to six-marker combinations to assess the likelihood of overfitting, this would provide an overview across several marker combinations, it was also unlikely that a seven-marker model

would be taken forward as performance at the six-marker level illustrates this (Table 4.3). As expected, the training portion AUC was higher than the validation portion AUC across all marker combinations. Moreover, the difference in AUC across the training portion and validation portion was >0.167 for all the explored combinations, which indicated that all the RF models had been strongly overfitted to the training data across all combinations and highlights the low ability of RF models to generalise. Although many of the validation AUCs were higher compared to the C5.0 non-boosted decision trees. This could be due to the training portion AUCs being much higher than that being seen for the C5.0 non-boosted decision trees thus stronger performing models were made to begin with and thus, despite the drop due to overfitting in RF, the validation portion AUCs for certain combinations are higher than that seen for decision trees. Also, the combination itself could be influencing this to a certain extent. As only 5 entered combinations were explored with the C5.0 non boosted compared to 10 combinations for RF.

Combination	Training portion AUC	Validation portion AUC	Difference in AUC (%)
A	0.999	0.681	0.318 (32)
VA	0.999	0.749	0.250 (33)
CA	1.00	0.751	0.249 (25)
CM	0.993	0.627	0.366 (37)
CVA	1.00	0.828	0.172 (17)
CVAM	0.996	0.825	0.171 (17)
TAPM	0.999	0.668	0.331 (33)
VAPM	0.999	0.687	0.312 (31)
CVAPM	0.997	0.809	0.188 (19)
TCVAPM	1.00	0.833	0.167 (17)

Table 4.3: Summary of the RF model performance using the Train and Test approach.

4.5 Train and Test approach for SVM model validation

SVM models produced by SPSS Modeler from the whole UOL dataset were previously explored in Chapter 3. The train and test approach was also used to assess overfitting of these models. As SVMs are distance-based algorithms, normalisation was needed. The detailed methodology used for this is described in Section 2.6.2. Briefly, the whole dataset was divided as before into training and validation datasets. The training portion contained 25 controls and 28 cases, while the validation portion contained 24 controls and 29 cases. Both datasets were normalised using the min-max methodology as previously

described (Section 2.4.1.1). Ten different C parameter values were assessed for the Linear SVM kernel, and 120 different combinations of gamma and C parameter were assessed for the RBF SVM kernel. A train and test analysis was then conducted for the best performing SVM models identified in Chapter 3, namely, the seven-marker combination (Sections 4.5.1 and 4.52), and the CVAM (Sections 4.5.3 and 4.5.4), CVA (Sections 4.5.5 and 4.5.6) and CV, CA and AV combinations (Sections 4.5.7 and 4.5.8).

4.5.1 Train and Test analysis of seven-marker Linear SVM models

Increasing the C parameter value from one to ten for the training and validation portions caused an increase in the performance of the seven-marker SVM model (Figure 4.19). Specifically, the training and validation AUCs ranged from 0.844-0.893 and 0.825-0.845, respectively. The small difference between corresponding training and validation AUC values indicated low levels of overfitting for the seven-marker Linear SVM models. Moreover, the difference in AUC between the train and test portions was lowest at C=1, which supports the rationale that smaller C value models tend to overfit less than higher C value models (Figure 4.19).

Moreover, at each C parameter value the training AUC was higher than that seen at validation (Figure 4.19), indicating the presence of overfitting. The highest AUC was obtained at C=6 for the training data, however this model also had the largest difference between training and validation AUCs, indicating the highest level of overfitting. The difference in AUC between the validation and training portions tended to increase, indicating increasing levels of overfitting, with increasing C parameter value. As the difference in AUC was lowest at C=1 this model was most likely to have been underfitted to the data.
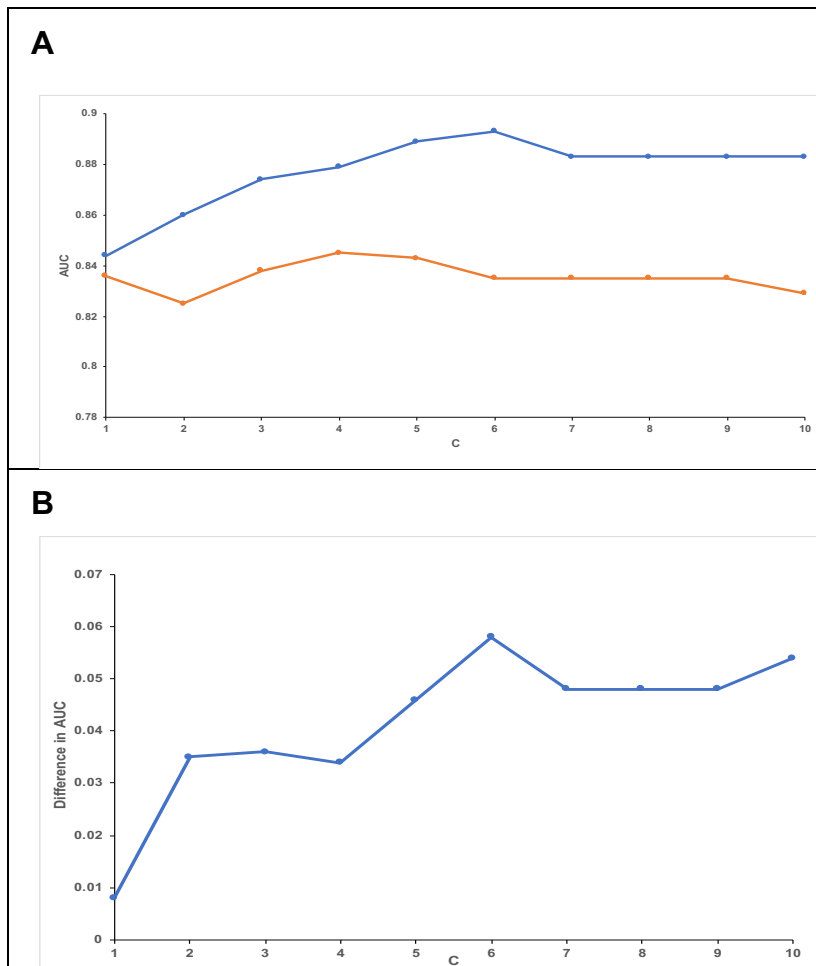
Figure 4.19: Train and Test analysis of the Linear SVM seven-marker model.
Data were min-max scaled, and a C range of 1 to 10 was used with the training (n=53) and validation (n=53) portions.
(A) Comparison between the training (n=53, blue line) and validation (n=53, red line) portion.
(B) Difference in AUC between training and validation portions at each C parameter value.

The predictor importance results (Table 4.4) indicated that VGF-NGF was the most important marker in nine out of then ten models and that thymosin beta 4 was the least important marker in nine out of the ten models

C	1st Marker	2nd Marker	3rd Marker	4th Marker	5th Marker	6th Marker	7th Marker
1	M=0.32	V=0.28	A=0.11	C=0.09	N=0.08	T=0.06	P=0.06
2	V=0.29	M=0.29	C=0.13	A=0.12	N=0.08	P=0.05	T=0.04
3	V=0.28	M=0.29	C=0.16	A=0.14	N=0.08	P=0.05	T=0.04
4	V=0.27	M=0.22	A=0.17	C=0.16	N=0.08	P=0.05	T=0.04
5	V=0.27	A=0.20	M=0.18	C=0.17	N=0.08	P=0.05	T=0.04
6	V=0.26	A=0.22	C=0.17	M=0.17	N=0.08	P=0.05	T=0.04
7	V=0.26	A=0.23	C=0.17	M=0.16	N=0.08	P=0.06	T=0.04
8	V=0.27	A=0.23	C=0.17	M=0.16	N=0.08	P=0.06	T=0.04
9	V=0.27	A=0.23	C=0.17	M=0.16	N=0.08	P=0.06	T=0.04
10	V=0.27	A=0.22	C=0.17	M=0.16	N=0.08	P=0.06	T=0.04

Table 4.4: Train and Test predictor importance analysis for the seven-marker Linear SVM models created using the training portion (n=53).

4.5.2 Train and Test analysis of seven-marker RBF SVM models

A train and test dataset approach was carried out in SPSS Modeler as previously described (Section 2.6.2). 120 combinations of C parameter and gamma were looked at to assess the impact of the C parameter and gamma values on the performance of the models in the training and validation portions and, consequently, the extent of overfitting. As the C parameter value increased at each gamma value, the performance of each model also increased, as expected (Figure 4.20). Additionally, at higher gamma values performance was better than at lower values, also as expected. The optimal training result was seen at C=7 and gamma=0.95, C=8 and gamma=0.85, C=9 and gamma=0.95, C=10 and gamma=0.9, and C=10 and gamma=0.95. This indicated that higher C parameter and higher gamma values produced better performing models in the training stage.
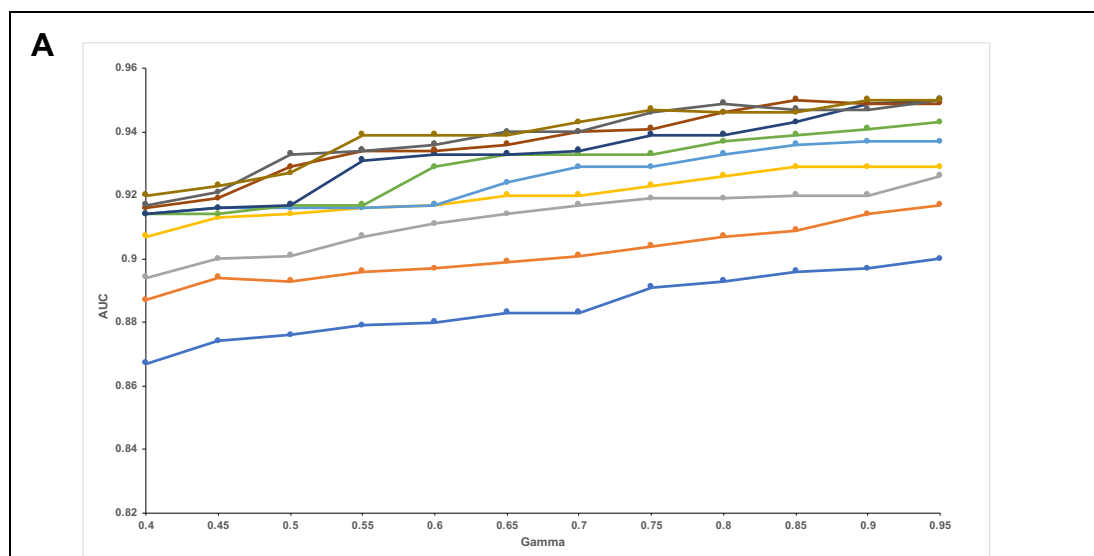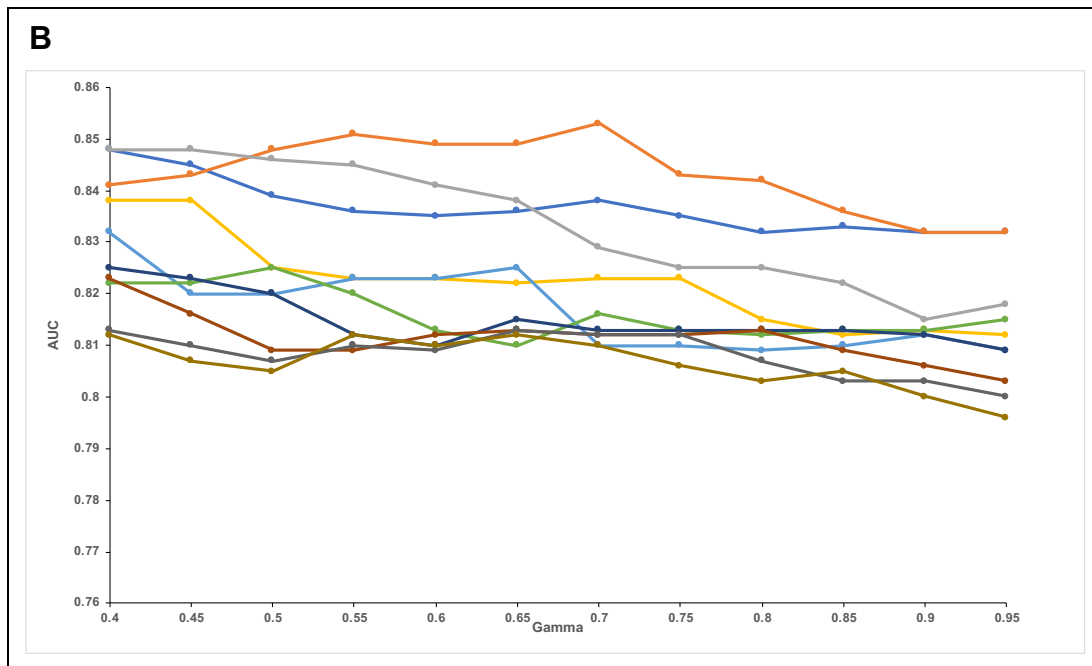
Figure 4.20: Train and Test analysis of RBF SVM seven-marker models.
Data were min-max scaled and a C range of 1 to 10 and a gamma range of 0.4 to 0.95 were used. In total 120 combinations of C and gamma values were examined. **(A)** Training portion dataset (n=53) results. AUCs for the range of C values between 1 and 10 is seen with C=1 (blue), C=2 (orange), C=3 (grey), C=4 (yellow), C=5 (light blue), C=6 (green), C=7 (navy blue), C=8 (brown), C=9 (dark grey), C=10 (light brown) are shown. The colour key for C parameter is used in all subsequent RBF results. **(B)** Validation portion dataset (n=53) results.

The validation results were lower than those of the training portion indicative of overfitting (Figure 4.20). Moreover, as the C parameter value increased the performance of the models decreased, thereby highlighting that these models had been overfitted to their cognate data and had a poor ability to generalise despite the high C parameter and high gamma value models performing well at training. Thus, the best models at the validation level had low C parameter and gamma values.

4.5.3 Train and Test analysis of CVAM Linear SVM models

Using the train and test approach described in Section 2.6.2, C parameter values from 1 to 10 were used. As expected, the results showed that the training portion performed better than the validation portion, with the best training set model in terms of AUC produced at C=6 (Figure 4.21). The difference in AUC was the lowest at C=1, suggesting that this model was least likely to be overfitted. Predictor importance results revealed that VGF-NGF was the most important marker in six out of the ten models, while CGA was the least important marker in all ten models (Table 4.5).
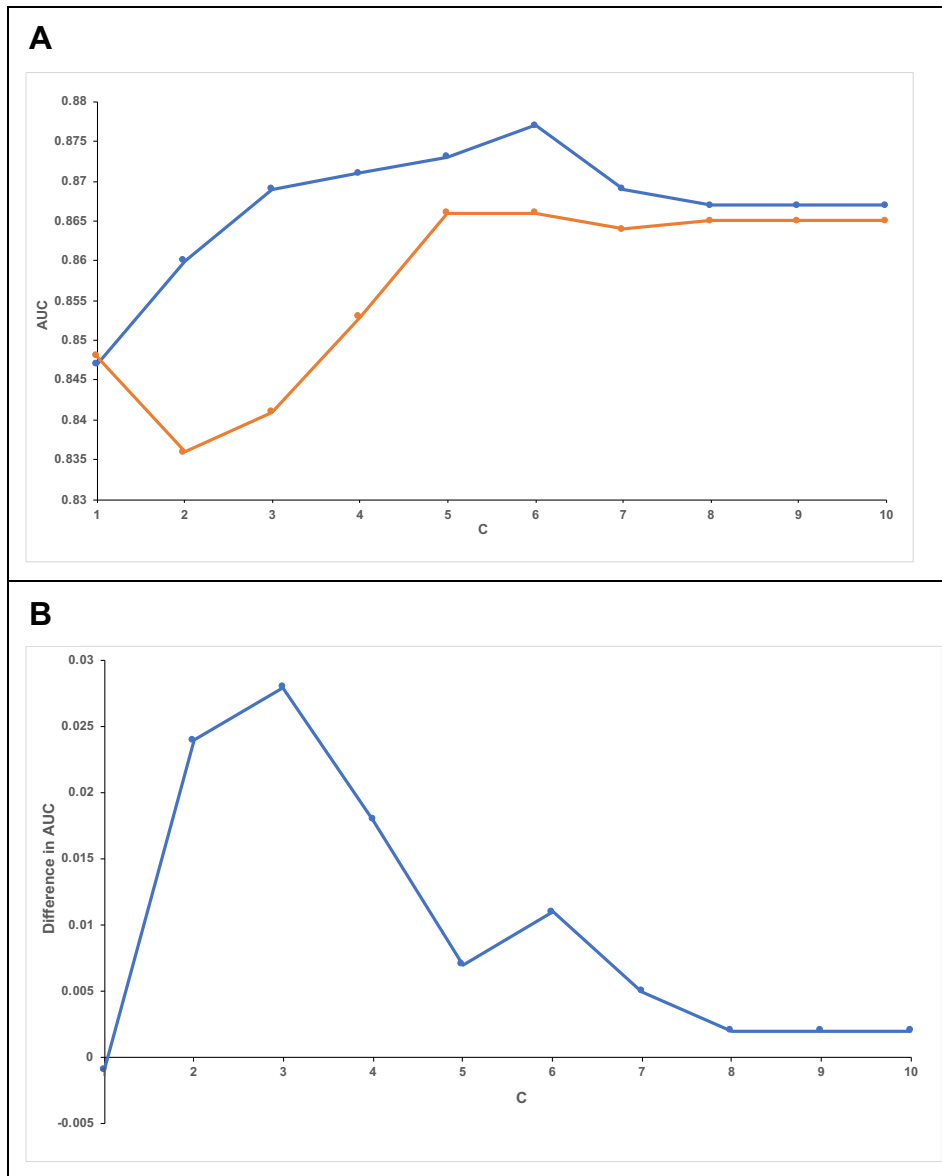
Figure 4.21: Train and Test analysis of the CVAM Linear SVM model. Data were min-max scaled, and a C range of 1 to 10 was used with the training (n=53) and validation (n=53) portions. **(A)** Comparison of Linear SVM results for the training and validation portions. **(B)** Difference in AUC between training and validation portions for the ten SVM models.

C	1st Marker	2nd Marker	3rd Marker	4th Marker
1	M=0.35	V=0.29	A=0.22	C=0.14
2	V=0.34	M=0.29	A=0.21	C=0.16
3	V=0.34	M=0.27	A=0.20	C=0.19
4	V=0.35	A=0.23	M=0.22	C=0.19
5	V=0.34	A=0.26	M=0.21	C=0.19
6	V=0.33	A=0.29	M=0.20	C=0.18
7	V=0.32	A=0.31	M=0.19	C=0.18
8	A=0.33	V=0.31	M=0.18	C=0.18
9	A=0.33	V=0.31	M=0.18	C=0.18
10	A=0.33	V=0.31	M=0.18	C=0.18

Table 4.5: Train and Test predictor importance for the ten CVAM Linear SVM models created using the training portion.
VGF was the most important marker in lower C parameter value models (C=2 to 7), while ANG2 was more important in higher C value models (C=8 to 10). The least important marker was CGA across all ten models.

4.5.4 Train and Test analysis of CVAM RBF SVM models

A train and test analysis examined 120 combinations of C parameter and gamma values to assess the change in CVAM model performance between training and validation stages as an indicator of the extent of overfitting (Figure 4.22). As expected, increasing the C parameter value resulted in increased performance at the training and validation levels, although only for C parameter values up to 4 or 5. For these C parameter values, increasing the gamma value tended to improved performance at both levels. This therefore supported the notion that both an optimal C and gamma level were important for the RBF SVM.
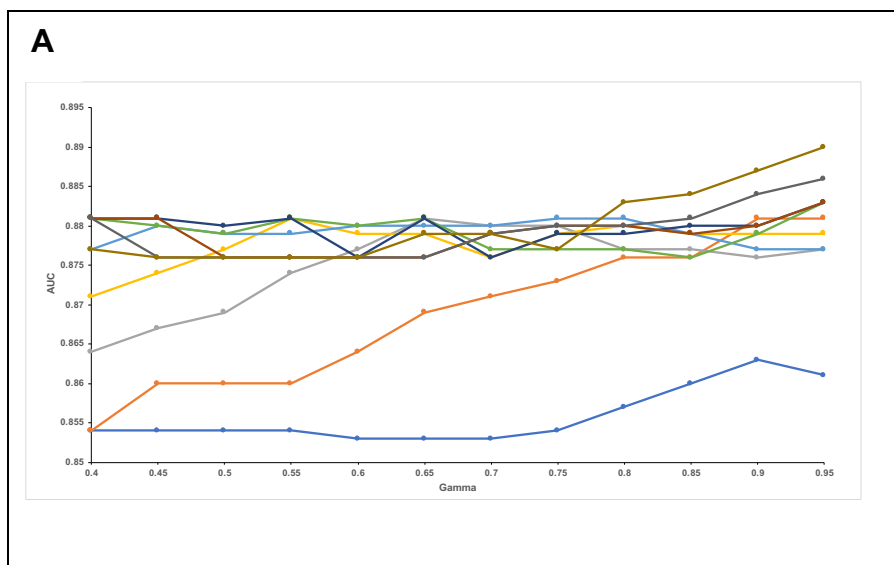
Figure 4.22: RBF SVM 4 marker CVAM combination Train and Test approach using SPSS Modeler. Data was min-max scaled. A C range of 1 to 10 was used and a gamma range of 0.4 to 0.95 was used. In total 120 combinations of C and gamma were examined. **(A)** Training

190

B
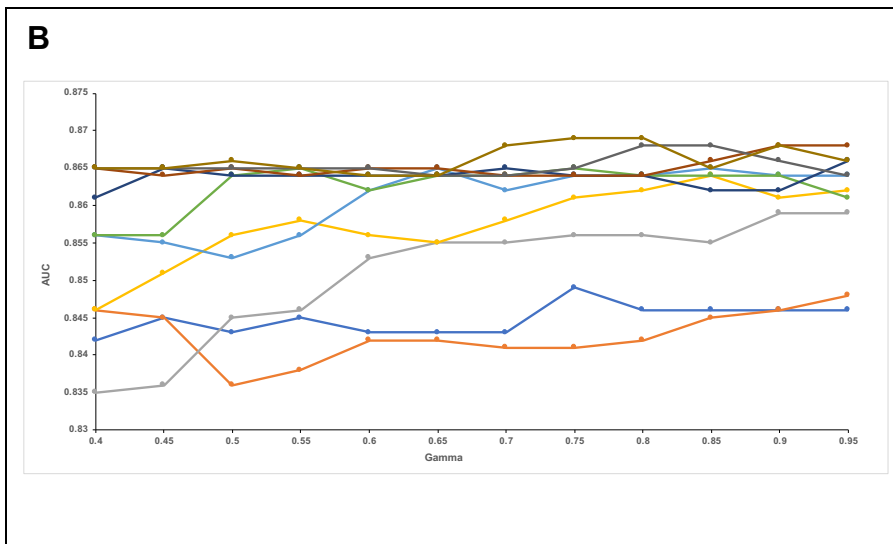
4.5.5 3 Train and Test analysis of CVA Linear SVM models

The extent of overfitting was explored using C parameter values from 1 to 10. The results showed that the training portion performed better than the validation portion across all C parameter values, as expected (Figure 4.23). However, the performance within either portion changed little with increasing C parameter value, suggesting that the C parameter was not a significant determinant of these models. The small difference (approx. 0.03 units) in performance across all C parameter values indicated that the training model was not overfitted at any value. This therefore suggested that overfitting was not of concern with these models and a wider suitability of models across the C parameters (1-10).

A
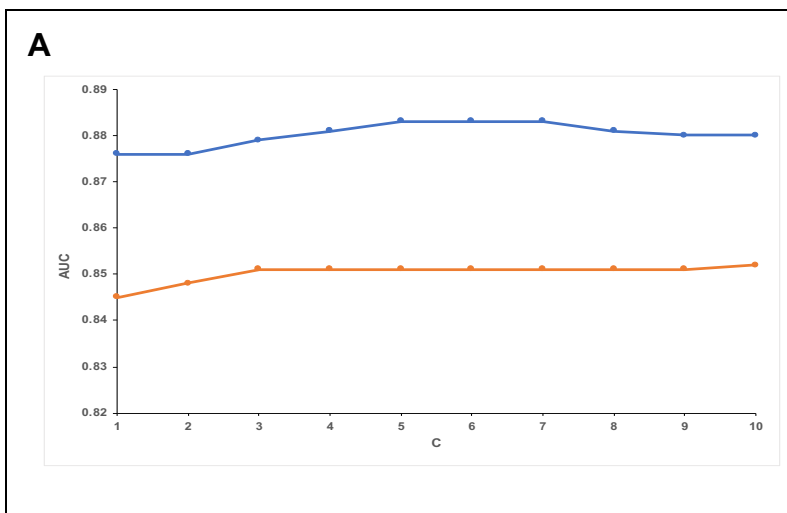
Figure 4.23: Linear SVM 3 marker CVA combination Train and Test approach using SPSS Modeler. Data were min-max scaled. A C range of 1 to 10 was used. The training portion contained (n=53) and validation portion contained (n=53). **(A)** Linear SVM results comparison between the training and validation portion results, using the 3-marker combination.
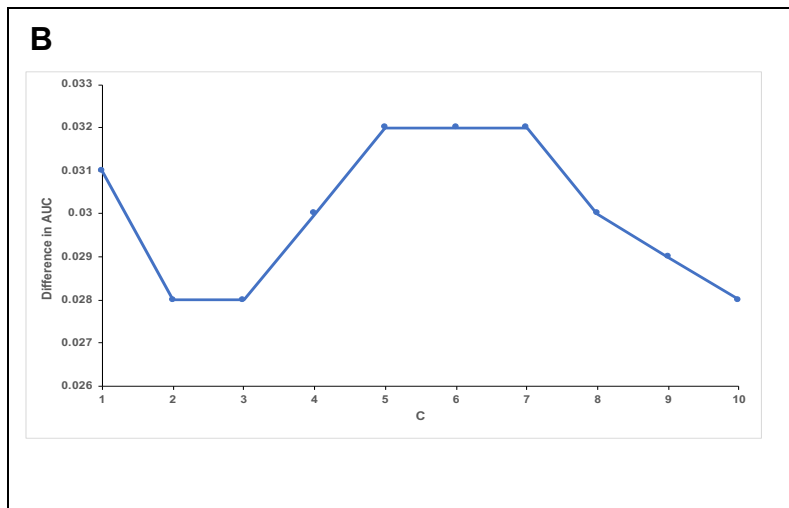
B

C	1st Marker	2nd Marker	3rd Marker
1	V=0.51	A=0.35	C=0.15
2	V=0.49	A=0.36	C=0.14
3	V=0.53	A=0.34	C=0.12
4	V=0.55	A=0.34	C=0.12
5	V=0.55	A=0.33	C=0.11
6	V=0.55	A=0.33	C=0.11
7	V=0.55	A=0.33	C=0.11
8	V=0.53	A=0.35	C=0.12
9	V=0.52	A=0.36	C=0.12
10	V=0.51	A=0.37	C=0.12

Table 4.6: Train and Test predictor importance for the ten CVA Linear SVM models created using the training portion.
For the three-marker CVA combination the most important marker is VGF. CGA is the least important marker across all the model.

4.5.6 Train and Test analysis of CVA RBF SVM models

Train and test analysis of CVA RBF SVM models were performed using 120 combinations of C parameter and gamma to assess the impact of these parameters on performance in the training and validation portions and the extent of overfitting (Figure 4.24). The results for this marker combination were mixed. For both portions, increasing the C parameter value tended to improve performance, although only up C=3 and 4 for the training and validation portions, respectively. Increasing the gamma value had no consistent effect on either portion. Moreover, the effect of changing either parameter was small (<0.01 units). Specifically, for the three-marker CVA combination the training

AUCs ranged from (0.876 to 0.883) and the validation portion the AUCs ranged from (0.845 to 0.852). Again, like with the Linear SVM, this meant that overfitting was not of major concern with these models across the C and gamma range.

Figure 4.24: RBF SVM 3 marker CVA combination Train and Test approach using SPSS Modeler. Data were min-max scaled. A C range of 1 to 10 was used and a gamma range of 0.4 to 0.95 was used. In total 120 combinations of C and gamma were examined. **(A)** Training portion (n=53) results. **(B)** Validation portion (n=53) results.
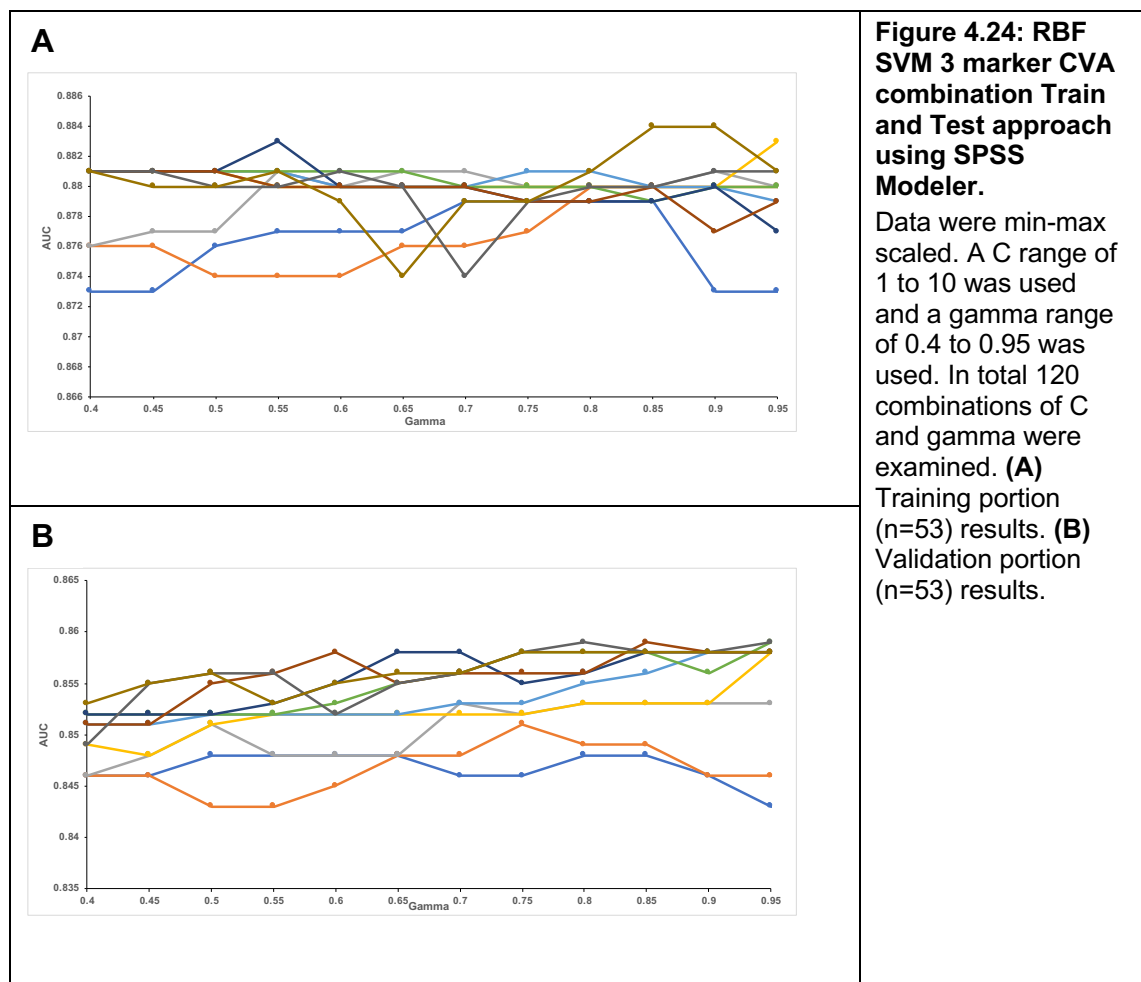
4.5.7 Train and Test analysis of CV, CA and VA Linear SVM models

The CV, CA and VA Linear SVM models were analysed using a C parameter range of 1 to 10. For the CV and the CA models, the training portion AUCs were higher than those seen for the validation portion across all C values, as expected (Figures 4.25 to 4.27). More specifically, the AUCs of CV training models ranged from 0.813 to 0.824, while the AUCs for the validation models ranged from 0.704 to 0.710 (Figure 4.25). The poorest performing validation models were seen at C=4, 5 and 8 (AUC = 0.704) and the best was at C=9 (AUC = 0.710). The AUCs of CA training models ranged from 0.833 to 0.846, and those of the validation portion ranged from 0.789 to 0.795 (Figure 4.26).

However, for the VA models the validation portion AUCs of 0.822 to 0.825 were higher than the training portion AUCs of 0.816 to 0.819 (Figure 4.27). For the VA training models the best AUC of 0.819 was seen at C=1, 9 and 10, while for VA validation models the best AUC of 0.825 was seen at C=1, 2, 8 and 9. The lower performance of training compared to validation models indicated a poorer ability of the VA marker data in the training set to produce linear SVM models compared to the validation set.

For the CV combination the difference in AUC was lowest at C=2, 6 and thus the impact of overfitting is less at these C values (Figure 4.25); for the CA combination this occurred at higher C parameter values of 7 to 10 (Figure 4.26); and for the VA combination, the training and validation model performance was most similar at C=10 (Figure 4.27). At the two-marker level the general trend of increased C value leading to a greater overfitted isn't necessarily the case. Finally, as the highest AUC of 0.846 obtained for all three two-marker linear SVM models was smaller than the lowest AUC of 0.876 obtained for the three-marker (CVA) linear SVM model in the training portion (n=53),thus it is clear that reducing to a two-marker combination leads to a drop in performance at the training level at C values across 1 to 10 and thus suggests that dropping to a two-marker combination may not be optimal.
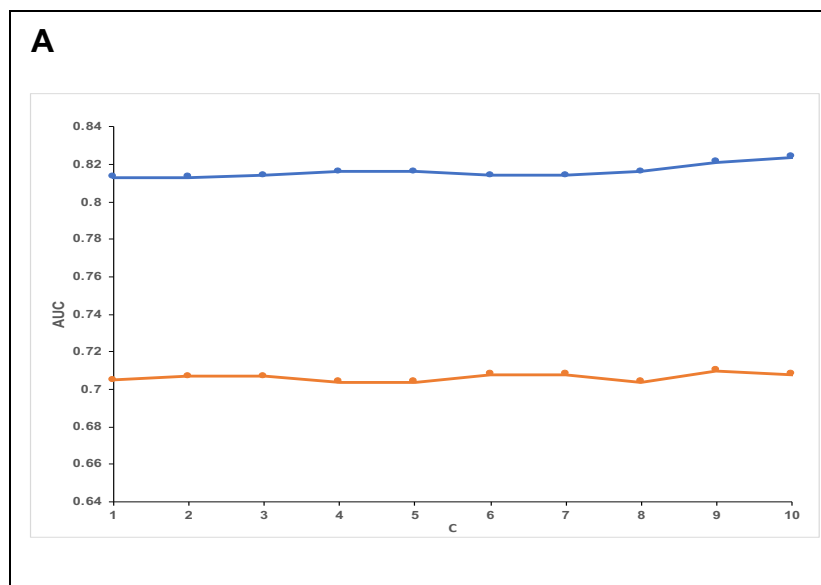
Figure 4.25: Linear SVM two-marker CV combination Train and Test approach using SPSS Modeler. Data were min-max scaled. A C range of 1 to 10 was used. 10 models were generated in total. **(A)** Linear SVM results comparison between the training portion (n=53) and validation portion (n=53) dataset results. **(B**) Difference in AUC between training and validation portion.
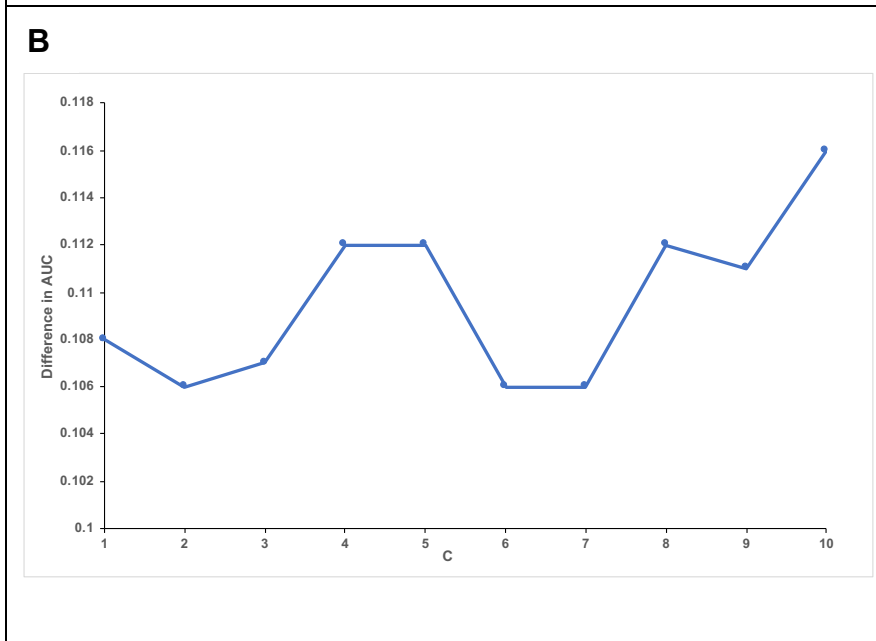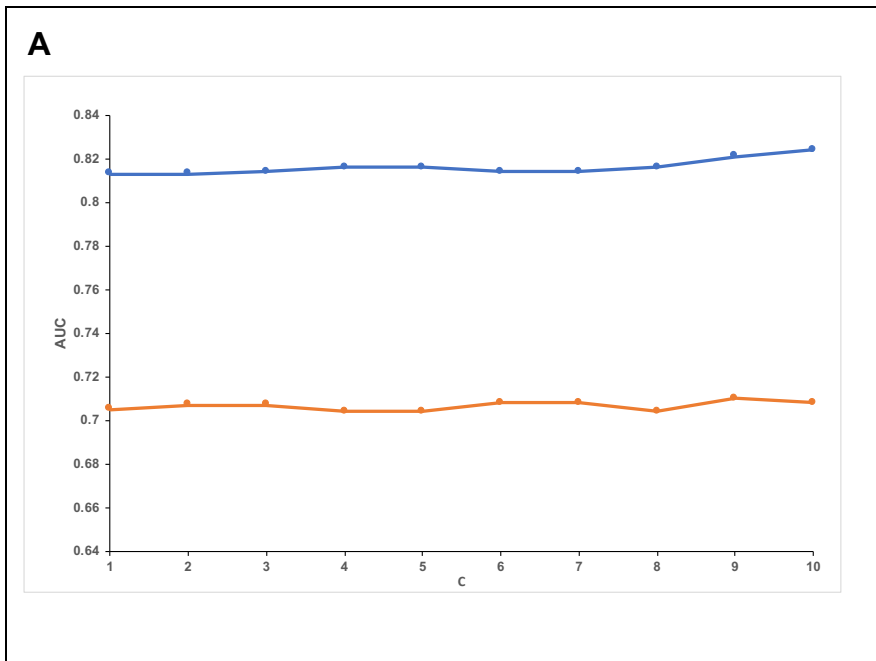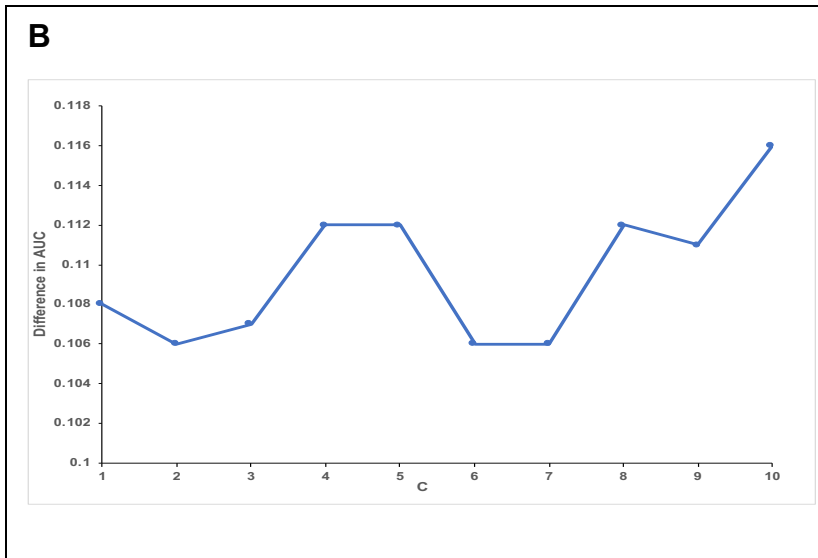
Figure 4.26: Linear SVM 2 marker CA combination Train and Test approach using SPSS Modeler. Data were min-max scaled. A C range of 1 to 10 was used. 10 models were generated in total. **(A)** Linear SVM results comparison between the training portion (n=53) and validation portion (n=53) dataset results. **(B)** Difference in AUC between training and validation portion.
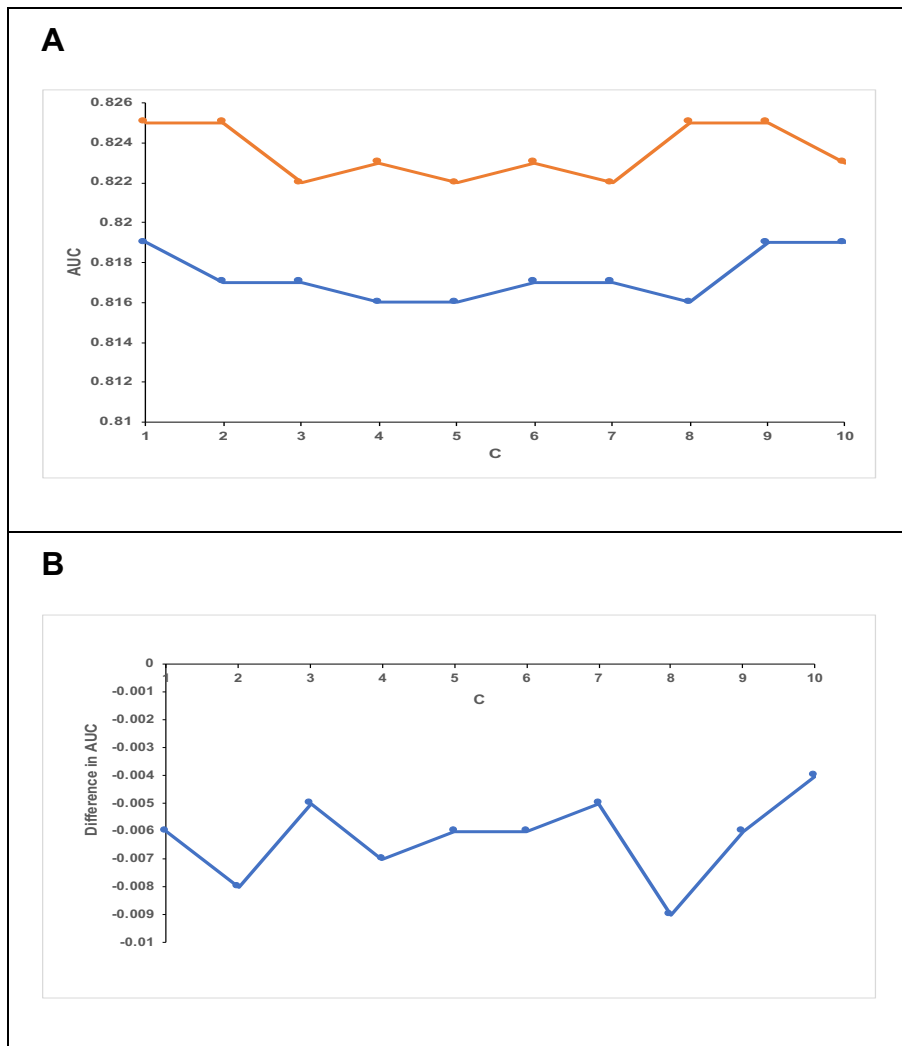
Figure 4.27: Linear SVM 2 marker VA combination Train and Test approach using SPSS Modeler. Data was min-max scaled. A C range of 1 to 10 was used. 10 models were generated in total**. (A)** Linear SVM results comparison between the training portion (n=53) and validation. **(B)** Difference in AUC between training and validation portion

C	1st marker	2nd marker
1	V=0.80	C=0.20
2	V=0.80	C=0.20
3	V=0.80	C=0.20
4	V=0.78	C=0.22
5	V=0.78	C=0.22
6	V=0.79	C=0.21
7	V=0.79	C=0.21
8	V=0.78	C=0.22
9	V=0.75	C=0.25
10	V=0.73	C=0.27

Table 4.7: Train and Test predictor importance for the ten CV Linear SVM models created using the training portion.
VGF-NGF was of greater importance for the model compared to CgA.

C	1st Marker	2nd Marker
1	A=0.69	C=0.31
2	A=0.69	C=0.31
3	A=0.67	C=0.33
4	A=0.67	C=0.33
5	A=0.68	C=0.32
6	A=0.69	C=0.31
7	A=0.69	C=0.31
8	A=0.69	C=0.31
9	A=0.69	C=0.31
10	A=0.69	C=0.31

Table 4.8: Train and Test predictor importance for the ten CA Linear SVM models created using the training portion.
ANG2 was of greater importance for the models compared to CgA.

C	1st Marker	2nd Marker
1	A=0.50	V=0.50
2	A=0.57	V=0.43
3	A=0.55	V=0.45
4	A=0.57	V=0.43
5	A=0.57	V=0.43
6	A=0.56	V=0.44
7	A=0.53	V=0.47
8	A=0.52	V=0.48
9	V=0.50	A=0.50
10	V=0.51	A=0.49

Table 4.9: Train and Test predictor importance for the ten VA Linear SVM models created using the training portion.
ANG2 was of greater importance for the model compared to VGF.

4.5.8 Train and Test analysis of CV, CA and VA RBF SVM models

The two-marker RBF SVM models were tested using 120 combinations of the C parameter and gamma with grid-based optimisation (Section 2.6.2). As found for the CV and CA linear SVM models, the corresponding training model AUCs were higher than the validation model AUCs across all C values (Figure 4.28 and Figure 4.29); the VA validation model AUCs were higher than those of the training models (Figure 4.30); and all training model AUCs were lower than those of the corresponding three-marker CVA models.

Specifically, the CA training model AUCs ranged from 0.834 to 0.844, while the validation model AUCs ranged from 0.787 to 0.793 (Figure 4.29); the CV training model AUCs ranged from 0.811 to 0.829 and validation model AUCs ranged from 0.704 to 0.717 (Figure 4.28); and the VA training model AUCs ranged from 0.816 to 0.819 and validation model AUCs ranged from 0.822 to 0.833 (Figure 4.30). Thus, for the RBF SVM, reduction to a two marker level leads to performances inferior to that seen at the three-marker level in the training portion and again like with the Linear SVM, it suggests that removal of an additional marker impacts performance in the training portion.
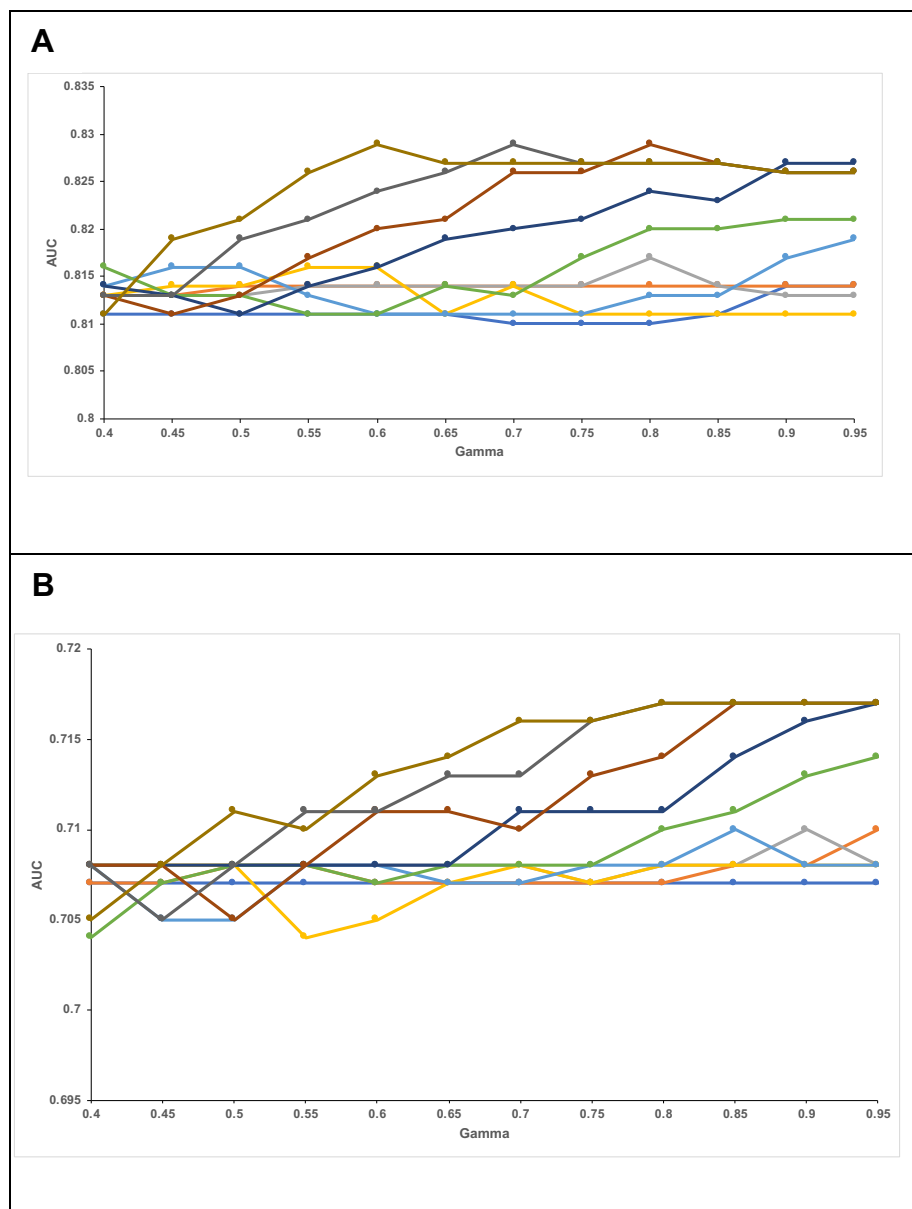
Figure 4.28: RBF SVM 2 marker CV combination Train and Test approach using SPSS Modeler. Data were min-max scaled. A C range of 1 to 10 was used and a gamma range of 0.4 to 0.95 was used. In total 120 combinations of C and gamma were examined. **(A)** Training portion dataset (n=53) results. **(B)** Validation portion dataset (n=53) results.
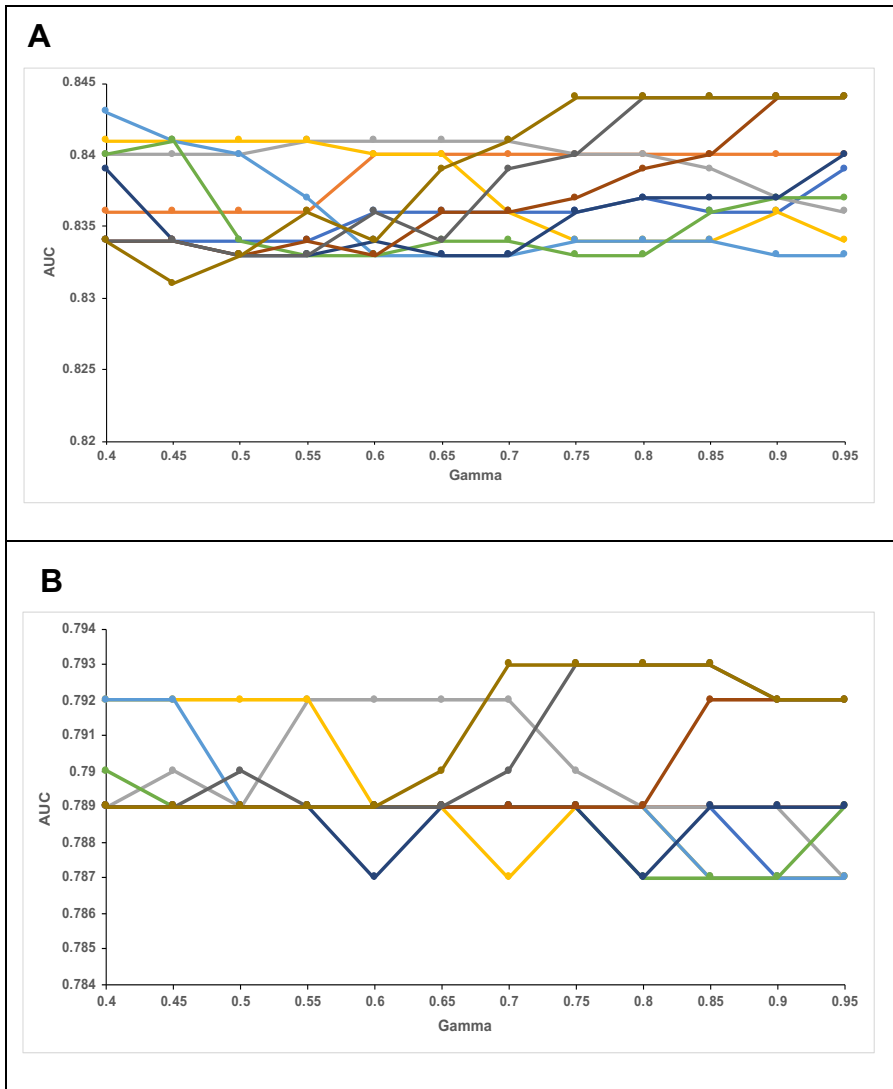
Figure 4.29: RBF SVM 2 marker CA combination Train and Test approach using SPSS Modeler. Data were min-max scaled. A C range of 1 to 10 was used and a gamma range of 0.4 to 0.95 was used. In total 120 combinations of C and gamma were examined. **(A)** Training portion dataset (n=53) results. **(B)** Validation portion dataset (n=53) results.
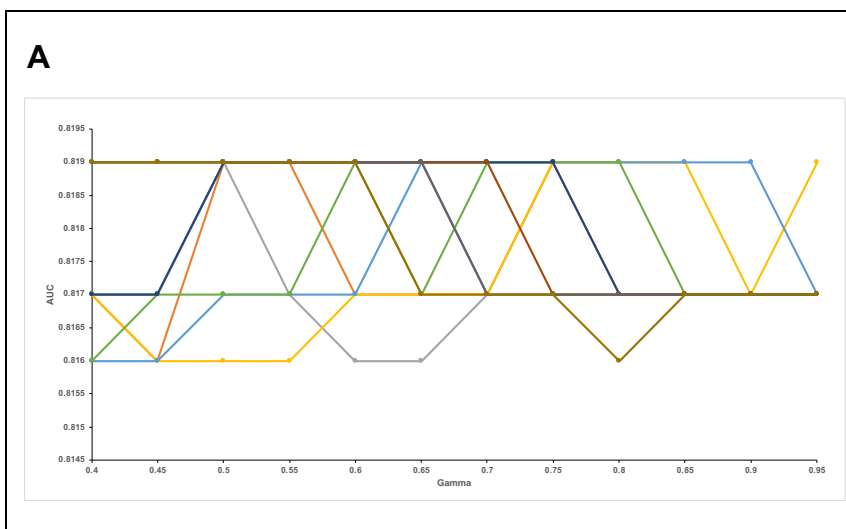
Figure 4.30: RBF SVM 2 marker VA combination Train and Test approach using SPSS Modeler Data were min-max scaled. A C range of 1 to 10 was used and a gamma range of 0.4 to 0.95 was used. In total 120 combinations of C and gamma were examined. **(A)** Training portion

B
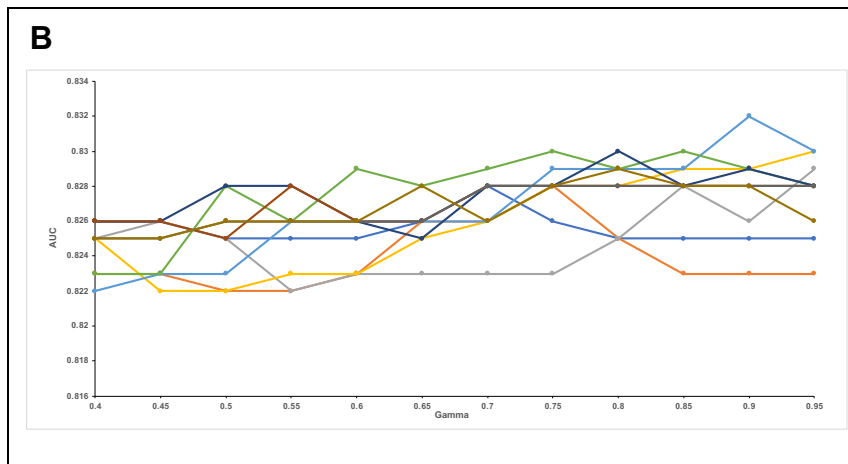
dataset (n=53) results. **(B)** Validation portion dataset (n=53) results.

4.6 Validation of boosted C5.0 decision trees using 10-FCV

Overfitting by boosted C5.0 decision trees was assessed using a 10FCV approach to assess performance when data for all seven markers were entered. This approach was taken to give an indication of the most important markers as well as which markers were redundant in the resulting models, thereby allowing a direct comparison with the results in Chapter 3. It also allowed me to analyse the impact of boosting on cross validation.

Consequently, instead of exploring each marker combination individually, ten individual cross validation runs of the seven-marker data were carried out as described in Section 2.8. The results from ten runs also provided insight into the variation between models produced from each run.

The results revealed that each run produced consistent high AUC values ranging from 0.917 and 0.93 (Table 4.10). Notably, the ACVM and ACV combinations were each used in five iterations, indicating the importance of the ANG2, CGA and VGF markers. Moreover, all ten runs produced C5.0 boosted decision trees that performed better than the C5.0 non-boosted decision tree results using either the accuracy or generality settings, showing that boosting enhanced C5.0 decision tree performance.

The extent of boosting differed between models as revealed by the percentage boosting, although the amount of boosting had little impact on the selection of markers or model performance. For example, runs 1 and 3 both used the three markers ANG2, CGA and VGF with boosting percentages of 70% and 90%, respectively. Nevertheless, each model achieved the same AUC (0.917) and the same predictor of importance for the ACV combination (A=0.37, C=0.36 and V=0.27). Similarly, for runs 1 and 7 the same marker combination was used

with the same boosting percentage (70%). The improved model performance (AUC=0.921) appeared to arise instead from relative changes in the relative contribution of each marker (C=0.41, A=0.30 and V=0.29). Finally, comparing runs 2, 4, 6, 9 and 10, which all used the 4 markers ANG2, CGA, VGF and MAC2BP, showed that while the boosting for these runs varied between 60 and 80%, the AUCs remained unchanged at 0.930.

Inclusion of the MAC2BP marker seemed to improve model performance as the AUC of 0.930 (mean accuracy 73.0-78.6%) exceed the range of 0.917-0.921 (mean accuracy 69.3-73.8%) achieved by models in which MAC2BP was not used.

	Markers entered	Markers used	Predictor Importance	Boost	AUC	Mean Accuracy (%)	Standard error (%)
Run 1	CTAMNVP	ACV	A=0.37 C=0.36 V=0.27	70%	0.917	73.0	3.7
Run 2	CTAMNVP	ACMV	A=0.30 C=0.27 M=0.22 V=0.20	60%	0.930	73.0	5.1
Run 3	CTAMNVP	ACV	A=0.37 C=0.36 V=0.27	90%	0.917	69.3	3.9
Run 4	CTAMNVP	ACMV	A=0.30 C=0.27 M=0.22 V=0.20	70%	0.930	78.6	3.0
Run 5	CTAMNVP	ACV	A=0.37 C=0.36 V=0.27	80%	0.917	73.8	3.3
Run 6	CTAMNVP	ACMV	A=0.30 C=0.27 M=0.22 V=0.20	70%	0.930	74.8	3.9
Run 7	CTAMNVP	CAV	C=0.41 A=0.30 V=0.29	70%	0.921	72.8	3.2
Run 8	CTAMNVP	ACV	A=0.37 C=0.36 V=0.27	80%	0.917	72.6	3.6
Run 9	CTAMNVP	ACMV	A=0.30 C=0.27 M=0.22 V=0.20	70%	0.930	73.0	5.4
Run 10	CTAMNVP	ACMV	A=0.30 C=0.27 M=0.22 V=0.20	80%	0.930	74.7	3.7

Table 4.10: 10FCV analysis of boosted C5.0 decision trees.
Models generated from seven marker inputs and ten iterations are compared. Each run resulted in a unique model.

The non-boosted C5.0 decision trees explored in Chapter 3 for the seven marker entered models, utilised ANG2 and CGA (Section 3.2). This was different to what was seen for the boosted cross validated C5.0 decision trees, which additionally utilised VGF and even MAC2BP in certain runs. Interestingly the AUC's obtained for the boosted 10FCV C5.0 decision trees were better than that obtained for non-boosted, non-cross validated C5.0 decision trees. Based on the ten runs obtained for the boosted C5.0 decision tree, the best model was the Run 9 and Run 10 models which utilised MAC2BP and VGF, and obtained an AUC of 0.930.

4.7 K-fold cross validation (k-FCV) of SVM models

K-FCV was used to evaluate overfitting of Optimised Linear and Optimised RBF models created in MCLA software. Building k-FCV models is a feature in MCLA and thus as Optimised SVM and Optimised RBF models were built using MCLA previously as described in Chapter 3, Optimised SVM models were explored for k-FCV. As the SVM algorithm is a distance-based algorithm, the data was normalised as described in Section 2.7.1. K values of 5, 10 and 20 were employed to assess the impact of fold number on performance and to compare the best performing models at each fold level. Optimal C and gamma parameters were assessed using MCLA because it provided a larger range of values than SPSS Modeler. Models were created for the seven-marker, six-marker (-TB4, -NSE, or -TIMP1), four-marker (CVAM), three-marker (CVA), and two-marker CV, CA and VA combinations. This was to be consistent with the approach taken in Chapter 3 and to compare the same combinations. Model performance was assessed using AUC and accuracy (percentage of correctly classified observations) values determined at each value of k.

Results from the k-FCV of Linear SVM models produced from all seven markers revealed AUCs of 0.83, 0.83 and 0.84 for the 5-, 10- and 20-fold models, respectively, indicating that increasing the fold number had little impact on performance.

The best Optimised Linear SVM 5FCV model was obtained using three different marker combinations: the six-marker (-TB4) model, the four-marker CVAM model, and the three-marker CVA model. All three models produced an AUC of 0.85. Hence, the 5FCV revealed the simplest optimal marker combination was

CVA. The best performing 10FCV and 20FCV models were both the three-marker CVA combination with an AUC of 0.86. Thus, across all k values, the CVA combination consistently produced the highest AUC values. Reducing to the two-marker combinations of CV, CA or VA produced lower AUC values of 0.74 to 0.83.

The best Optimised RBF SVM 5FCV model was obtained from the six-marker (-NSE) and the seven-marker combination with an AUC of 0.82. The best 10FCV model was obtained from the CVAM combination, which produced an AUC of 0.84. The best 20FCV was obtained from four marker combinations: the seven marker, six-marker (-NSE), CVAM and CVA combinations, which each produced an AUC of 0.81. The CV models had noticeably poorer performance, with AUCs of less than 0.8 across all k values. Thus, from these results across Optimised Linear and Optimised RBF SVM for each combination assessed fold number generally had little impact on AUC and performance.

Optimised Linear SVM Model	Fold Validation	Box constraint	AUC	Accuracy (%)
7 marker	5	215.4435000	0.83	72.6
7 marker	10	46.4159000	0.83	76.4
7 marker	20	46.4159000	0.84	76.4
6 marker (-NSE)	5	215.4435000	0.81	73.6
6 marker (-NSE)	10	215.4435000	0.81	76.4
6 marker (-NSE)	20	1000.0000000	0.82	74.5
6 marker (-TIMP1)	5	46.4159000	0.84	75.5
6 marker (-TIMP1)	10	46.4159000	0.84	78.3
6 marker (-TIMP1)	20	215.4435000	0.85	79.2
6 marker (-TB4)	5	215.4435000	0.85	74.5
6 marker (-TB4)	10	215.4435000	0.84	76.4
6 marker (-TB4)	20	46.4159000	0.84	75.5
4 marker CVAM	5	10.0000000	0.85	75.5
4 marker CVAM	10	10.0000000	0.85	78.3
4 marker CVAM	20	10.0000000	0.85	76.4
3 marker CVA	5	46.4159000	0.85	75.5
3 marker CVA	10	0.4641600	0.86	77.4
3 marker CVA	20	0.4641600	0.86	78.3
2 marker CV	5	0.0215440	0.74	64.2
2 marker CV	10	0.4641600	0.77	69.8
2 marker CV	20	0.4641600	0.78	73.6
2 marker CA	5	0.0046416	0.82	77.4
2 marker CA	10	0.0046416	0.83	75.5
2 marker CA	20	0.0010000	0.81	76.4
2 marker VA	5	0.1000000	0.80	74.5
2 marker VA	10	0.4641600	0.81	75.5
2 marker VA	20	2.1544000	0.80	75.5

Table 4.11: Summary of the Optimised Linear SVM k-fold cross validation analysis using MCLA.

Optimised RBF SVM Model	Fold Validation	Box constraint	Kernel Scale	AUC	Accuracy (%)
7 marker	5	10.00000	2.15440	0.82	73.6
7 marker	10	10.00000	2.15440	0.82	71.7
7 marker	20	46.41590	2.15440	0.81	72.6
6 marker(-NSE)	5	10.00000	2.15440	0.82	75.5
6 marker(-NSE)	10	10.00000	215.44350	0.82	71.7
6 marker(-NSE)	20	10.00000	2.15440	0.81	73.6
6 marker (-TIMP1)	5	46.41590	0.10000	0.77	73.6
6 marker (-TIMP1)	10	215.44350	0.10000	0.78	74.5
6 marker (-TIMP1)	20	215.44350	0.10000	0.78	75.5
6 marker (-TB4)	5	10.00000	0.10000	0.72	73.6
6 marker(-TB4)	10	10.00000	0.10000	0.75	73.6
6 marker (-TB4)	20	10.00000	0.10000	0.76	73.6
4 marker CVAM	5	0.46416	10.00000	0.80	70.8
4 marker CVAM	10	215.44350	2.15440	0.84	73.6
4 marker CVAM	20	2.15440	0.10000	0.81	75.5
3 marker CVA	5	0.46416	0.10000	0.81	80.2
3 marker CVA	10	0.46416	0.10000	0.80	80.2
3 marker CVA	20	0.46416	0.10000	0.81	79.2
2 marker CV	5	0.46416	0.10000	0.71	71.7
2 marker CV	10	10.00000	0.46416	0.74	70.8
2 marker CV	20	2.15440	0.10000	0.69	68.9
2 marker CA	5	2.15440	0.10000	0.78	74.5
2 marker CA	10	2.15440	0.10000	0.79	74.5
2 marker CA	20	2.15440	0.10000	0.80	74.5
2 marker VA	5	0.46416	0.10000	0.78	78.3
2 marker VA	10	215.44350	2.15440	0.80	72.6
2 marker VA	20	215.44350	2.15440	0.79	72.6

Table 4.12: Summary of the Optimised RBF SVM k-fold cross validation analysis using MCLA.

4.7 Discussion

In the work described in this chapter I explored different cross-validation approaches to assess the suitability of the different biomarkers and types of algorithms identified in Chapter 3. This assessment had a particular focus on understanding the level of overfitting via cross validation. The cross-validation approaches explored in this chapter were the train and test approach and k-FCV. Both the train and test approach and the k-FCV approaches generated models via splitting the UOL dataset and are different from the models described in Chapter 3 which were derived from the whole dataset. Thus, this chapter explored the utility of the UOL dataset itself for ML model derivation using train and test as well as k-FCV approaches.

The train and test analysis of selected LR models (Table 4.1) revealed that the relative performance of each model performed, as expected, more poorly in the validation compared to the training stage, because models were derived exclusively from the training set data. However, most LR models were found to be robust to overfitting, because the difference in training and validation AUCs was small (<10%). In general, the single marker LR models performed worse than multiple marker models, which again illustrates that a multiple marker approach improves performance, as seen before in Chapter 3. While the VA and A models had the smallest drop in AUC of 1%, the CVA model performed best overall with a training and validation AUCs of 0.894 and 0.849, a difference of only 5%. However, the inclusion of MAC2BP in the CVAM LR model did not improve either the training or validation AUC value, indicating that model performance can also be reduced by adding markers.

In summary of the train and test analysis, the CVA LR model was the most suitable one to take forward. This supported my findings in Chapter 3 regarding the utility of CgA, VGF and ANG2 markers as the optimal biomarker panel for pNET detection.

The utility of LR as a suitable ML algorithm utilising a train and test approach is illustrated in a study by Lee et al. (2021), which involved developing a multi-biomarker panel for the diagnosis of pancreatic cancer. The algorithm was developed using the three biomarkers leucine rich alpha-2-glycoprotein 1 (LRG1), transthyretin (TTR) and CA-19-9. A multi-centre cohort of normal, other cancer, pancreatic benign disease and PDAC patients was used for this analysis. This involved a training set created by extracting 70% of normal and PDAC type data and a test dataset combining the remaining 30% of normal and PDAC data and all the other cancer and pancreatic benign cancer data. They found that when looking at normal vs PDAC patients the training data had a sensitivity of 93.81% and a specificity of 90.86% and a sensitivity of 94.7% and 91.21% in the test data[473]. Thus, illustrating that the LR algorithm is robust to overfitting. Thus, providing further support to my findings that LR as an algorithm is robust to overfitting.

When assessing the C5.0 non-boosted decision tree, all C5.0 non-boosted decision trees appeared to have been overfitted to their training data as the differences in AUC were all high, ranging from 13 to 26% (Table 4.2). These results supported the idea that C5.0 decision trees tend to overfit[397]. Compared

to the C4.5 decision tree, the C5.0 algorithm produces smaller decision trees, and the training dataset is classified with high accuracy compared to the C4.5 tree[477]. The overfitting seen in the non-boosted C5.0 decision trees results was most likely exacerbated by the small size of the dataset.

Instead of carrying out ten runs as used before to generate models from the whole UOL training set analysis in Chapter 3, train and test analysis of RFs employed a single run to assess overfitting of models produced from ten different marker combinations. For all marker combinations, the training data performed better than the validation data, as expected (Table 4.3). However, based on the very high training AUC values (0.993-1.000) and the large differences from the validation AUCs (17-37%) all RF models appeared to have been overfitted to the training data, thus suggesting limitations of taking these generated models further. A study involving the development and validation of a RF Diagnostic model for Acute Myocardial Infarction based on ferroptosis-related genes had strong diagnostic performance in both validation and test set with these being 0.8550 and 0.7308 respectively[478] and hence in these results the impacts of overfitting are less based on the difference in the AUCs obtained. The Yifan et al., study constructed a RF diagnostic model of acute MI through eight differentially expressed ferroptosis-related genes. The study used k-FCV with n=15 to improve generalisation power. They found that the variability of AUCs between the validation and train sets was small enough to be satisfying. They also compared the performance of three supervised ML algorithms, that included LR, SVM and RF. After evaluating comprehensively Kolmogorov-Smirnov (KS), accuracy, TPR and AUC of all these three algorithms, the RF model showed good diagnostic performance AUC of 0.8550 and validated in different data sets AUC=0.7308. They also found that the model had a strong ability to eliminate false negative interference, which is critical for myocardial infarction which has a high fatality rate.

The train and test analysis of the Linear and RBF SVM models again showed that training performance was consistently better than the validation performance, with the single exception of the VA model. As superior validation performance of the VA model was seen in both Linear and RBF SVM models, this phenomenon was not kernel specific. For the validation portion the best AUC of 0.795 was seen at C=4. For the three marker CVA combination using the RBF SVM in the training and validation portion, the trend was that at lower

C and gamma values the AUC is lower compared to that seen at higher AUC and C values. This suggests that at lower C values there is potential underfitting and that the algorithm was not capturing patterns within the training data , however increasing the C value leads to over-optimistic models. Moreover, for a given C value, increasing the gamma parameter caused an increase in AUC and for a given gamma value, increasing the C value caused an increase in the AUC. Dropping to a two-marker combination, either CA, CV or VA caused a general drop in AUCs. A drop in performance between the train and test suggests overfitting, whilst underfitted models, are likely to be models that at training are not performing well reflected by lower AUCs. Thus, the ideal model, would have good performance at training but also maintain this performance at validation without substantial drop. However, as the train and test approach in this chapter utilised the same dataset which was split, with a smaller training number whilst useful in understanding the impact that parameters may be having on a model, using an external validation approach as discussed in Chapter 5 and assessing if the top performing models utilising the whole training data (larger n number) as opposed to the split data as discussed in this chapter and that impact of a drop in performance would provide stronger evidence as to whether the SVM model approach to a diagnostic pNET algorithm was suitable.

Thus, while the train and test analysis provided an insight into the potential underfitting and overfitting, a limitation of the Linear and RBF SVM train and test approach in SPSS Modeler was that a narrow range of C parameter and gamma values were selected based on recommendations for the use of SPSS Modeler[479], however there was no indication that the most optimal values lay within these ranges. Hence, this time-consuming process provided no guarantee of producing an optimal model. Thus, for SVM models the Optimised Linear and Optimised RBF SVM k-FCV are more reliable than the Linear and RBF train and test approach results as they address the issue of Optimising the SVM models that were created.

Validation in SPSS Modeler employed C and gamma parameter values from 1 to 10 and 0.4 to 0.95, respectively, which were larger than SPSS Modeler recommendations. Nevertheless, Optimised Linear and RBF SVM models created in MCLA had a far larger default range of 0.001 to 1000 for the gamma (kernel scale) and C (box constraint) values. Thus, SPSS Modeler was useful

for analysing trends across C and gamma values and to assess overfitting of SVM models but did not necessarily produce the best performing model. In contrast MCLA allowed the best optimised SVM models to be selected, albeit with an increased risk of overfitting.

k-FCV differs from train and test validation as the data used for training are also used for the validation of the model, thus there is a recycling of the data. Data are divided according to a predetermined number of folds (5, 10 or 20 in this study) and k-1 folds are then used for testing. 5, 10 and 20 folds were selected as these fold numbers would allow for sufficient amounts of data to consist in each fold based on the size of the dataset and also provide a range of folds numbers to explore. The k-FCV results for all the selected Linear SVM Optimised models revealed similar 5FCV, 10FCV and 20FCV AUC values (Table 4.11 and Table 4.12). This therefore implies that fold number did not have an impact on AUC and that model performance was stable across all folds assessed. However, the accuracy of all the selected Optimised Linear SVM models was higher with 20FCV than 5FCV. Increasing the k number reduces the bias but increases the variance and there is therefore a need to balance both. A higher the k number this means that each model is trained on a larger training set and then tested on a smaller test fold. This therefore in theory should lead to a lower prediction error due to the model seeing more of the available data. However, a lower k value means that the model is trained on a smaller training set and then tested n a larger test fold. Thus, the potential for the data distribution in the test fold to be different form the training set is larger and thus a higher prediction error is expected[480].

A study exploring Type 2 diabetes mellites prediction using LR, SVM, ANN , KNN and LR based on lncRNA expression used a stratified 10FCV approach for evaluation. The stratified 10-FCV ensured that each fold contained the same proportion of healthy and diabetic individuals. The results for the stratified 10-FCV for LR had an AUC across the folds ranging from 0.95 to 1.00 which was a strong consistent performance[481]. A stratified k-fold approach would be an additional approach to consider for the validation of pNET biomarkers and algorithms in the future, as this will allow for each fold in the dataset to have the same proportion of observations of a given label i.e., in this case pNET or control. However, this would ultimately also require a lot more samples.

Vablas et al. carried out a study exploring ML validation approaches by using Gaussian noise as data in which they simulated a situation in which robust validation should produce two class classification accuracy that approaches theoretical chance level of 50%. They tested five validation approaches including a Train/Test split, k-FCV, nested and two types of partially nested cross validation. They also performed the simulations using different sample sizes and tested what other factors aside from sample size influence overfitting. They found that k-FCV provided optimistically-biased performance estimates and was not sufficient to control overfitting. They suggest a nested-CV approach is a better approach as this provided an unbiased performance estimate. Likewise completely separating testing and training data was enough to obtain unbiased performance estimates regardless of sample size[433]. The study found that factors which influence bias include data dimensionality, hyperparameter space and the number of CV folds. They demonstrated that complex models are capable of fitting random noise in the data and thus if the data which was used for validation is also used in parameter tuning, the performance is inflated due to the models fitting the noise not only in the training but also in the validation data[433]. This therefore strengthens the importance of using an external validation approach as discussed in Chapter 5 to prevent over overoptimistic performance.

The usage of k-FCV as an approach in biomarker development using the RF algorithm was explored by Langmead et al. (2021). The study assessed predictive markers for persistent organ failure in AP with a panel of five circulating cytokines (ANG2, hepatocyte growth factor, interleukin-8, resistin and soluble necrosis factor receptor 1A in a prospective verification cohort with a RF classifier. The study utilised 10FCV approach to estimate how well the RF model would generalise to an independent dataset. The results revealed an accuracy of 89% and AUC of 0.91[482], which outperformed individual cytokines, laboratory tests and clinical scores. The strong performance of the five-marker RF model indicated that this algorithm with the validation approach used has the potential to be used for biomarker development, however a limitation is the lack of external validation for the panel, with this needing to be done to further support the findings.

The effect of fold number in k-FCV on algorithm performance was studied by Liang et al. (2020), who assessed gene-level copy number alteration using a deep learning network algorithm to predict the tissue of origin for cancers of unknown primary site. This algorithm was compared to other ML algorithms including RF and eXtereme Gradience Boosting (XGBoost). XGBoost is a type of boosted tree learning algorithm[483]. For all the assessed algorithms they found the general trend was for increasing k value to improve classifier performance. The larger the k value was, the more samples that there were in the training set, and the fewer samples that there were in the test set and vice versa. The k values used ranged from five to thirty. In their analysis a small test set had a negative impact on model evaluation[484]. However, clarifying what small is in practice is likely dependent on the size of the data. Thus, this study illustrates the potential importance of fold number on performance. In my k-FCV Optimised Linear and RBF SVM models developed for the pNET detection, AUC values were broadly similar across all folds suggesting that the k values chosen did not have an impact on performance and were most likely suitable. Different biomarker studies use different types of validation. Factors that commonly influence the choice of validation method include sample availability for internal validation methods, and access to independent cohorts for external validation methods. In situations where access to independent cohorts for external validation is limited or the cohort is too small to split into train and test cohorts for internal validation, an internal validation approach such as k-FCV would be favoured. Using unseen data to test a ML model entails the collection of a substantial amount of data, which is particularly rare in research involving human participants due primarily to restrictions on sample availability, time and cost[433].

Certain studies have a train, validation and test approach such as a study by Zhang et al. (2013) in which an SVM algorithm was developed to predict breast cancer[485]. In this study the 130 patients were randomly split into roughly equal training and testing groups. They had a training and validation group which consisted of 32 healthy samples and 34 cancer samples and a testing set consisting of 34 cancer samples and 31 healthy samples. This study adopted a three-way split comprising training, validation, and independent testing groups to try and reflect a real application outcome based on the performance of the testing groups.k-FCV was carried out on the training group to find the optimal

parameters for the SVM classifier. The training group was used for learning to fit the parameters of the classifier, the validation was used to tune the parameters of the classifier and the testing set was used to assess the performance of the fully trained classifier only. They found that the best model from the training group may not predict the best performance in the testing data due to inconsistencies between training and testing data. One limitation from their work was the utility of three-way data split is the sample size, as when they split a small sample size into three ways, the minimal data in each set meant that the analysis would lack power. With the train and test approach used for the pNET algorithm development, this reduced the original training size and meant that amount of data with which the algorithm was trained was less than that seen in Chapter 3 whereby the whole training set was used, and this was not split as seen in this chapter.

A particular disadvantage of using internal compared to external validation is that the same data are used to develop and evaluate a model. As a result, internal validation tends to be over-optimistic, even when performed appropriately, and particularly when large numbers of biomarkers are involved[486]. Thus, for my work in the development of a pNET algorithm it would be important to carry out external validation in model development (Chapter 5), to prevent over optimistic assessments.

From the work carried out in Chapters 3 and 4 there is rationale to take forward the markers Chromogranin A, VGF and ANG2 for external validation (Chapter 5). In Chapter 3 whereby the whole training dataset was used for model creation, the McNemar test results showed the reduction to a three-marker combination of CVA from CVAM was not significant for the LR algorithm, Linear SVM (C=4) and RBF SVM (C=10 and gamma=0.95) (P=1.00). Moreover, performance for the LR, Linear SVM (C=4) and RBF SVM (C=10 and gamma=0.95) models was above 0.85, with AUCs being 0.873, 0.863 and 0.868. A strong performance was also seen for the Optimised Linear SVM and Optimised RBF SVM models for the CVA combination using the whole training dataset which had AUCs of 0.88 and 1.00 respectively. The RF algorithm for CVA using the whole training dataset also performed with an average AUC of 1.00. Thus, across a range of algorithms, when using the whole training dataset, the CVA combination was suitable to take forward. This has been further supported by the approaches of internal validation in Chapter 4, with the

RF model performing with an AUC of 1.00 in the training portion and 0.828 in the validation portion. The RBF SVM models assessed using the train and test approach had training portion AUCs ranging from 0.876 to 0.883 with validation portion AUC ranging from 0.845 to 0.852. For the Optimised Linear and RBF SVM using the k-FCV approach, the AUCs for 5FCV, 10FCV and 20FCV Optimised Linear SVM were 0.85, 0.86 and 0.86 respectively and for 5FCV, 10FCV and 20FCV Optimised RBF SVM the AUCs obtained were 0.81, 0.80 and 0.81 respectively. Thus ,k-FCV provided further support for the suitability of these markers and algorithm. Finally, for the 10FCV boosted decision tree model when all seven markers were entered, for 5 out of 10 runs the model utilised Chromogranin A, VGF and ANG2 with AUCs above 0.9 across all these runs.

Thus, based on the results obtained in Chapters 3 and 4, it was evident that the combination of CHGA, VGF and ANG2 provided the best panel to take forward for external validation and to assess using the C5.0 decision tree, SVM, RF and LR algorithm. In Chapter 5, these markers and algorithms are evaluated using an independent cohort of pNET samples collected for this purpose at the Royal Free Hospital. Additionally, confounding factors and associations between marker levels and clinical characteristics are evaluated.

Chapter 5 - External Validation of the Biomarker Panel

5.1 Introduction to External Validation

In Chapter 4, validation approaches using the UOL cohort were assessed. These approaches included train and test and k-FCV. However, a more revealing evaluation of an algorithm's performance requires an independent cohort. Validation of biomarkers using independent cohorts can be seen in various studies in the literature across multiple disorders ranging from Parkinson's disease[487] to endometriosis[488] and to cancer[489]. While many biomarker discovery studies have generated large amounts of omic data, few novel biomarkers have reached clinical use. The importance of an independent validation cohort has been discussed in the context of the development of a metabolomics signature for Parkinson's disease[490]. However, the similar principles apply to biomarker discovery work in general. The authors concluded that independent cohorts are needed to address the problem of biomarkers that do not stand up to clinical use due to the overfitting of classifier models[490]. Factors that could differ between sample-collection centres include sample handling and other technical variations that could cause unwanted differences in marker levels between cohorts. They also suggested that studies 'merit caution' if they have not used an independent cohort for validation[490]. Moreover, as biomarker discovery, validation and implementation form a long process, taking candidate markers and models forward for clinical assessment without validation in independent cohorts is likely to be inefficient and wasteful of resources. Hence independent cohorts are an important part of the validation process.

Thus, in this chapter I describe external validation of the best models identified in Chapter 3, which were produced by combining CgA, VGF and ANG2 marker data. External validation of these biomarkers and the model algorithms was carried out using an independent cohort of pNET patients at the RFH (Royal Free Hospital), who kindly allowed me to collect blood samples. Healthy control serum samples were obtained from the UKCTOCs biobank, collected by colleagues at UCL[491]. Plasma CgA levels were available for the RFH pNET patients as this is an established NET marker, and I compared these data with

my serum CgA results to determine the level of correlation and assess whether plasma CgA results were interchangeable with serum data in my models.

In Section 5.2 to 5.4 I describe and discuss my external validation of ANG2, VGF and CgA, respectively, as individual pNET markers. These sections also explore the association of specific clinical parameters with each marker and potential confounding factors. Based on the external validation results of the three candidate markers, I selected specific models developed in Chapter 3 for external validation. Finally, in Section 5.6 I consider whether more accurate models could be derived from a combination of the training and external validation datasets.

5.2 ANG2 External Validation

ANG2 was validated using an external case control cohort. The cases for this cohort consisted of pNET serum samples from the RFH (n=60) and the controls consisted of control samples from the UKCTOCs biobank (n=51). The ELISA results showed that ANG2 levels were elevated in cases compared to controls (Figure 5.1, Table 5.1).
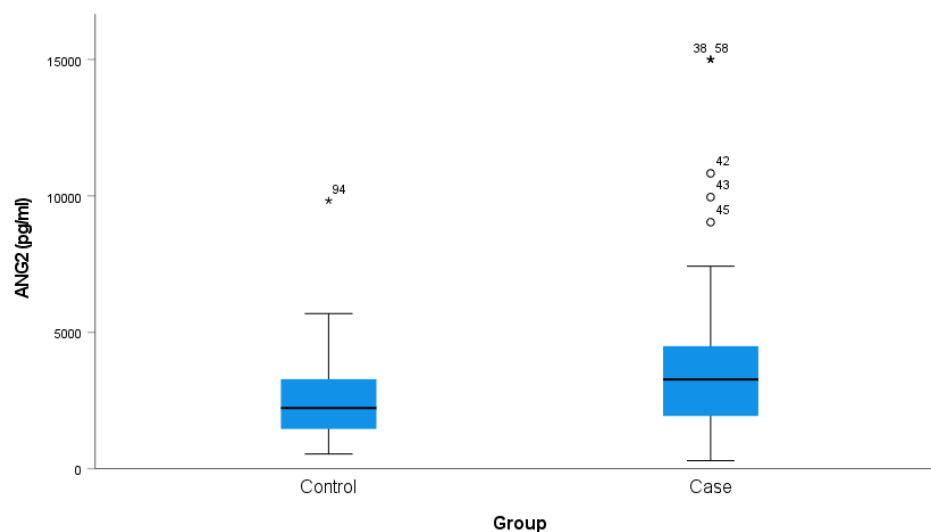
Figure 5.1: ANG2 levels in RFH pNET cases (n=60) and UKCTOCs controls (n=51).
ANG2 levels were assessed by ELISA in RFH pNET case and UKCTOCs control serum samples. The results are presented as a box plot with the median shown by the thick horizontal line, the top and bottom lines in the box show the 1st and 3rd quartiles, with the whiskers denoting the maximum and minimum values, with the exception of outlier values in SPSS which are denoted by ° with extreme outlier values by *. Extreme outlier values are considered in SPSS to lie in the 3rd quartile + (3 x interquartile range) or the 1st quartile – (3 x interquartile range).

	Case	Control
Mean ANG2 level ± standard deviation (pg/mL)	3999.74 ± 3354.01	2610 ± 1545.31

Table 5.1: Mean level of ANG2 in RFH pNET and healthy control samples.
ANG2 levels were elevated in pNET patients (n=60) compared to healthy controls (n=51).

The null hypothesis that the distribution of ANG2 was similar across cases and controls was tested using an independent-samples Mann-Whitney U test which produced a p value of 0.018. Hence, the null hypothesis was rejected, and I concluded that there was a statistically significant difference between case and control ANG2 levels, as expected.

5.2.1 Analysis of potential confounding factors for ANG2

Potential confounding factors for ANG2 in the cases and control samples were examined. Oral contraceptive pill (OCP) and BMI were examined in the control group as potential confounding factors for ANG2 as this information was available for the control group. For the pNET patients, PPIs, presence of another cancer, age, gender and diabetes status were all examined as potential confounding factors as this information was available for the pNET patients. The results revealed a Pearson correlation coefficient of 0.159 (p=0.266), indicating that there was no association between ANG2 level and BMI in the control group. Mann Whitney U test results revealed that there was also no significant relationship between ANG2 and OCP (no OCP usage = 22; OCP usage = 29) (p=0.313) was obtained in the control group.

For the pNET patients the null hypothesis that the ANG2 distributions were similar across different categories of PPI was examined through an independent-samples Kruskal-Wallis Test. The categories examined were PPI (n=21), omeprazole (n=11), lansoprazole (n=9), pantoprazole (n=1), no PPI (n=37) and not known (n=2). The results revealed a test statistic of 0.340 and p value of 0.844 and the null hypothesis was therefore not rejected. Thus, PPI usage did not appear to be a confounding factor.

ANG2 confounding factor analysis was also carried out for diabetes. Two samples were excluded, one with previous diabetes and one with borderline diabetes, leaving a total of 58 samples (no diabetes = 42, diabetes = 16). The null hypothesis that the distribution of ANG2 was similar across the two

categories of diabetic status was assessed using an independent-samples Mann-Whitney U test. The null hypothesis was accepted as no significant difference was found (p=0.639), indicating diabetes was not a confounding factor.

Gender was also examined as a potential confounding factor for RFH pNET ANG2 levels (n=60) using an independent-samples Mann-Whitney U test to test the null hypothesis that the distribution of ANG2 was similar between gender categories (male = 30, female = 30). As the result was not statistically significant (p=0.947), gender was not found to be a confounding factor.

Pearson Correlation was used to assess age as a potential confounding factor, and the resulting coefficient of -0.123 (p=0.349) indicated age was not a confounding factor.

The presence of another cancer as a confounding factor for ANG2 levels in pNET patients (n=60) was also assessed (no other cancer = 47, another cancer = 13) using an independent-samples Mann-Whitney U test and also found to be not significant (p=0.302).

Hence, none of the assessed factors appeared to confound the ANG2 data.

5.2.2 Analysis of potential ANG2-associated clinical characteristics

Association of ANG2 levels with various clinical characteristics of the RFH pNET cohort was also examined to explore whether pNET specific clinical characteristics were associated with ANG2 levels in the pNET patients. These characteristics included MEN1 status, metastasis status, tumour location, and tumour grade.

To examine whether ANG2 levels were associated with MEN1 status an independent-samples Mann-Whitney U Test was carried out for the RFH pNET ANG2 levels (n=57). Three samples were excluded because one was from a patient with Zollinger Ellison syndrome and MEN1, another was from a patient with Zollinger Ellison syndrome, and the third was from a patient with unconfirmed MEN1. The results showed that MEN1 status (MEN1 = 4, non-MEN1 = 53) was not associated with ANG2 level (p=0.729) in this this cohort.

Association with metastasis was investigated using an independent-samples Mann-Whitney U test to analyse the RFH pNET data (metastasis = 30, no metastasis = 29), with one locally-advanced pNET excluded. The results showed that metastasis was not associated with ANG2 level (p=0.111).

To examine whether tumour location was associated with ANG2 level in the RFH pNET data, an independent-samples Kruskal-Wallis Test was carried out. For this analysis seven samples were excluded because there was not a clearly defined location of body, neck or head. The remaining cohort (n=53) comprised body (n=8), head (n=17), neck (n=4), not known (n=12), and tail (n=12) samples. The results showed that ANG2 level was not associated with tumour location (test statistic 1.378, p=0.848).

Association between ANG2 and grade was examined. For this analysis one grade-3 sample was excluded because there were no other grade-3 samples. Three remaining samples (n=59) were divided into four groups: G1 (n=10), G2 (n=13), well differentiated (n=24), and not known (n=12). Independent-samples Kruskal-Wallis Test results showed that ANG2 level was not associated with grade (test statistic 2.583, p=0.461).

Finally, ANG2 level and QCancer® score correlation was carried out using a Pearson correlation test. The resulting coefficient value of 0.077 (p=0.558) indicated no significant correlation. ANG2 level and QCancer®Pancreas score correlation was also assessed using Pearson Correlation and the resulting coefficient of 0.133 (p=0.310) indicated no significant correlation.

Thus, none of the clinical characteristics examined were found to significantly associate with ANG2 level and they were therefore excluded as confounding factors.

5.2.2 Analysis of potential ANG2-associated biochemical markers

The association of ANG2 levels with other biochemical parameters that I obtained for the RFH pNET cohort (n=60) from their clinical records was assessed to examine whether there was a relationship between these biochemical markers and ANG2 levels in the pNET patients. A total of 60 pNET samples for which ANG2 levels and CgA levels were looked at with these additional biochemical parameter. However, the additional biochemical parameters were not available for all 60 pNET samples and only available results were used for the analysis. The additional biochemical parameters included CEA, carbohydrate antigen (CA19-9), C-reactive protein (CRP), creatinine and bilirubin. CEA, CA-19-9, CRP, creatinine and bilirubin levels were assessed in these patients as part of their routine clinical care. Results for these

markers showed that CA19-9 alone was significantly associated with ANG2 level (Table 5.2).

Biochemical marker	Available results (n)	Pearson correlation coefficient	Significance (2 tailed)
CEA	23	-0.30	0.892
CA 19-9	28	0.790	*<0.01
CRP	43	0.046	0.768
Creatinine	60	0.098	0.454
Bilirubin	60	0.138	0.294

Table 5.2: Correlation between ANG2 and biochemical markers
For the RFH pNET samples that had a CgA sera and an ANG2 result, CEA, CA 19-9, CRP, creatinine and bilirubin were assessed for a correlation with ANG2 level.

Association between ANG2 level and serum or plasma CgA levels was assessed (Fig. 5.2), to see if there was a significant correlation in pNET patients for both assessed biomarkers. CgA plasma levels were determined for the RFH pNET patients as part of routine clinical care. 8 samples were excluded from the n=60 because CgA plasma levels were not available for these patients. Additional samples were also excluded which were at the lower or upper level of detection of the ANG2 sera and CGA sera assay. Thus, a total of 39 samples were examined.
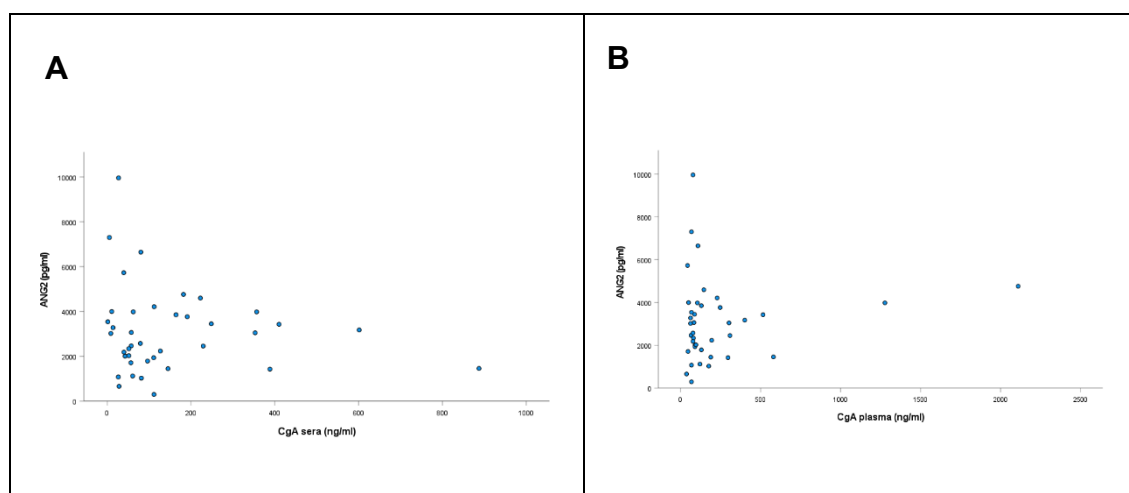
Figure 5.2: Association of ANG2 level with serum and plasma CgA level
ANG2 levels were compared to serum CgA levels obtained by ELISA using the CisBio assay and to plasma CgA levels determined as part of routine clinical care using the Diasource Assay. (A) ANG2 and serum CgA association (n=39) (B) ANG2 and plasma CgA association (n=39).

Pearson correlation analysis of ANG2 association with plasma CgA produced a slight positive correlation coefficient of 0.114 (p=0.491), while analysis of ANG2 association with serum CgA produced a slight negative correlation coefficient of

-0.129 (p=0.433). Thus, ANG2 level was not statistically significantly correlated with either serum or plasma CgA level. In summary, ANG2 levels were found to associate only with CA 19-9 levels in RFH pNET samples.

5.3 VGF External Validation

VGF was validated in RFH pNET and UKCTOCs control samples. VGF was evaluated in the same RFH pNET and UKCTOCS control serum samples as those assessed for ANG2 and CgA validation with the inclusion of the RFHpNET sample that had not been assayed for CgA. VGF was unexpectedly found to be higher in controls compared to cases (Figure 5.3 and Table 5.3), although Mann Whitney U test results for the null hypothesis that case (n=30) and control (n=35) VGF levels were similar was not rejected as the difference was not significant (p=0.324).

OCP and BMI usage in the control cohort were explored as potential confounding factors that might explain the increased levels of VGF seen in the control group as this information was available for the control group. However, Pearson correlation analysis of the control VGF and BMI data revealed a coefficient of 0.146 (p=0.403), indicating no significant association. In contrast, OCP usage was found to be a confounding factor for the control cohort as an independent-samples Mann-Whitney U test revealed a significant difference of -2.360 (standardised test statistic) (p=0.017) between the OCP usage (n=18) and no-OCP-usage (n=17) groups. Although the levels of VGF were higher in the controls who were not on OCP compared to those who were. OCP as a confounding factor for pNET patients was not assessed due to this information not specifically collected for the pNET patients, thus comparing pNET patients not on OCP and controls was not carried out.

VGF and grade was assessed in the RFHpNET patients. For this analysis one G3 RFHpNET sample was excluded due to insufficient numbers. The categories that were assessed were G1 (n=4), G2 (n=7), Not known (n=4) and Well differentiated (n=14). Independent Samples Kruskal Wallis Test results indicated that the distribution of VGF was the same across categories of grade was to be rejected as p=0.027. VGF and PPI usage was assessed in the RFHpNET samples. The distribution of VGF across the categories of PPI were assessed using an Independent-Samples Mann Whitney U Test, no (n=20) and PPI usage (n=10). Results were found to be not significant p=0.422. VGF and

metastasis was also assessed using Independent-Samples Mann-Whitney U test. No metastasis (n=14) and metastasis (n=15), one locally advanced sample was excluded. Results were found not to be significant with p=0.146. VGF and MEN1 was also assessed with one sample excluded using an Independent Samples Mann Whitney U Test MEN1 (n=3) no (n=26) with U=42.00 and p=0.866 and thus the null hypothesis was retained. VGF and location of tumour was assessed using Independent-Samples Kruskal-Wallis Test Body (n=3), Head (n=10), Neck (n=2), Not known (n=8), Tail (n=6).Test statistic 6.018 and p=0.198.
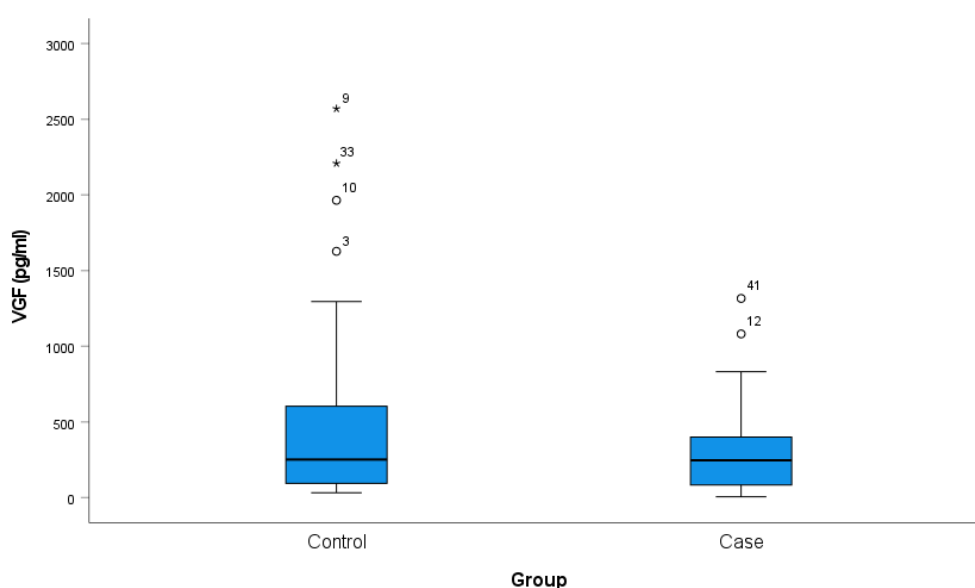
Figure 5.3: VGF levels in RFH pNET and control serum.

VGF levels of RFH pNET (n=30) and UKCTOCs control (n=35) samples were determined by ELISA. The results are presented as a box plot as described in the legend to Fig. 5.1.

	Case	Control
Mean VGF level ± standard deviation (pg/mL)	310.817 ± 314.736	523.471 ± 656.15

Table 5.3: Mean level of VGF in pNET cases and healthy controls

VGF levels were lower in RFH pNET patients (n=30) compared to healthy controls (n=35).

5.4 CgA External Validation

Case-control differences in CgA level were also validated using the external RFH case and UKCTOCS control cohorts. The results showed that CgA levels were elevated in cases compared to controls (Figure 5.4, Table 5.4). The null hypothesis that case and control CgA levels were similar was assessed using an independent-samples Mann-Whitney U test, which revealed that case levels

were significantly higher than control levels (p<0.001) and the null hypothesis was rejected, as expected.
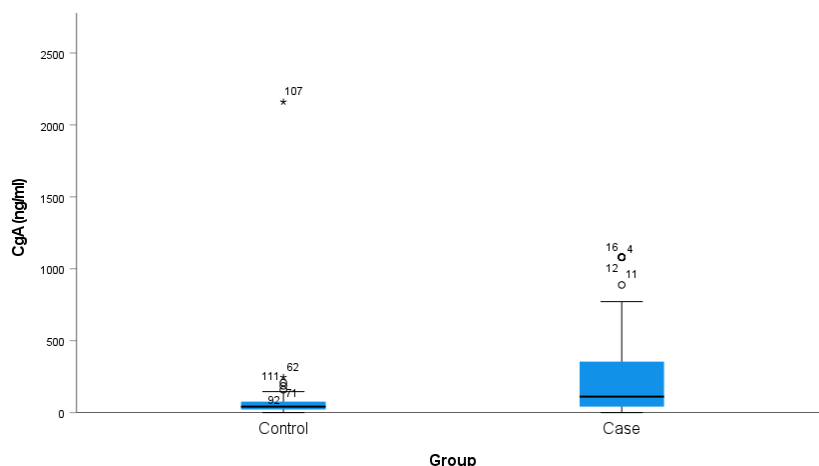
Figure 5.4: CgA levels in RFH pNET and UKCTOCs controls.
CgA levels in RFH pNET (n=60) and UKCTOCs control (n=51) samples were assessed by ELISA. The results are presented as a box plot as described in the legend to Fig. 5.1.

	Case	Control
Mean CgA level ± standard deviation (ng/mL)	277.765±362.558	100.996±299.132

Table 5.4: Mean level of CgA in RFH pNET and healthy control samples.
CgA levels were elevated in RFH pNET patients (n=60) compared to healthy controls (n=51).

5.4.1 Analysis of potential confounding factors for CgA

As previously described for control ANG2 and VGF levels, OC usage and BMI were explored as potential confounding factors for control CgA levels within the control cohort (n=51). The results revealed no significant correlation between CgA and BMI as assessed by Pearson corelation (coefficient of 0.134, p=0.350) or between CgA and OCP as assessed using an independent-samples Mann-Whitney U test (OCP usage = 29, no OCP usage = 22; (p=0.746).

PPI usage, diabetes, presence of another cancer, age and gender were also assessed as potential confounding factors for pNET CgA levels in the same way as performed for ANG2 and VGF.

An independent-samples Kruskal-Wallis test was carried out for the three PPI groups (no PPI use, 37 samples; not known, 2 samples; PPI use, 21 samples). PPI use included lansoprazole (n=9), omeprazole (n=11) and pantoprazole (n=1). However, dissimilar to pNET ANG2 levels, the null hypothesis was

rejected (test statistic = 6.082, p=0.048) and PPI usage was determined to be a confounding factor for CgA levels in pNET serum. Figure 5.5 illustrates the levels of CgA in PPI users, non-PPI users and unknown status. PPI users had higher CgA levels then non-PPI users, although the two patients with not known PPI status had the highest median CgA levels (Figure 5.5).
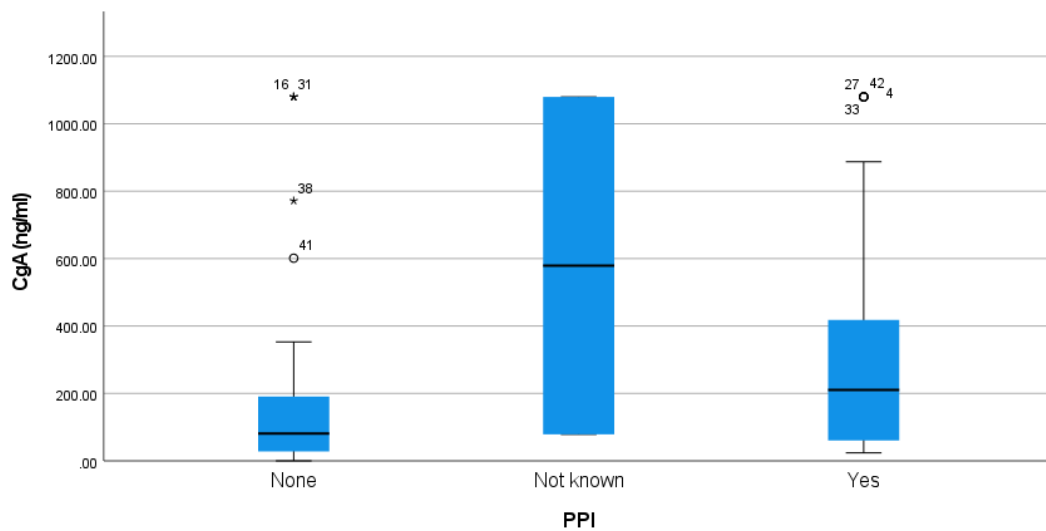
Figure 5.5: CgA levels in the RFHpNETs for PPI users and non-users.
CgA levels were compared in the RFHpNET patients who were on PPIs (n=21), and those not on PPIs (n=37). 2 patients PPI status was not known. The results are presented as a box plot as described in the legend to Fig. 5.1.

To assess whether diabetes was a confounding factor for CgA, an independent-samples Mann-Whitney U Test was performed to compare two groups (no diabetes, 42 samples; type II diabetes, 16 samples). As for ANG2, the null hypothesis that the levels of CgA in the two groups were similar was not rejected (p=0.958) and diabetes was deduced not to be a confounding factor for pNET CgA levels.

To assess whether the presence of another cancer was a confounder for pNET CgA levels an independent samples Mann-Whitney U test was carried out (no other cancer, 47 samples; other cancer, 13 samples). The results indicated that another cancer was not significant (p=0.535). Gender was also assessed using an independent-samples Mann-Whitney U test (males, 30 samples; females, 30 samples) and found not to be a confounding factor for pNET CgA level (U=452.50, p=0.970). Age was also found not to be a confounding factor for pNET CgA level as the Pearson correlation coefficient was 0.125 (p=0.343).

In summary, PPI use was the only cofounding factor identified for CgA levels in the RFH pNET cohort.

5.4.2 Analysis of potential CgA-associated clinical characteristics

Association between CgA and selected clinical characteristics (metastasis, MEN1, grade, location of the tumour, and QCancer® score) of the RFH pNET cohort was assessed.

For the assessment of an association between metastasis and pNET CgA level, one sample that was a locally advanced pNET was excluded, and the two groups used in the analysis were no metastasis (n=30) and metastasis (n=29). An independent-samples Mann-Whitney U test showed no significant difference between groups (U=515.00, p=0.225).

As mentioned previously, three samples were excluded for the MEN1 analysis because one sample was from a patient with Zollinger Ellison syndrome and MEN1, another was from a patient with Zollinger Ellison syndrome, and the third was from a patient with unconfirmed MEN1. The results showed that MEN1 status (MEN1 = 4, non-MEN1 = 53) were excluded from the assessment of an association between MEN1 and CgA level. An independent-samples Mann-Whitney U test was carried out to assess whether the distribution of CgA levels was similar between the MEN1 (n=4) and non-MEN1 (n=53) samples. No significance was found (U= 96.00, p=0.775) and the null hypothesis was retained.

For the assessment of association between grade and CgA levels in the RFH pNET samples, the grade 3 group was excluded as there were insufficient samples (n=1). The other 59 samples were distributed between G1 (n=10), G2 (n=13), not known (n=12) and well differentiated (n=24) groups. An independent-samples Kruskal-Wallis test found no significant difference between these groups (test statistic of 2.493, p=0.477) and the null hypothesis was retained.

Similarly, association between tumour location and CgA level was assessed using an independent-samples Kruskal-Wallis test, which found no significant difference (test statistic of 2.458, p=0.652) between body (n=8), head (n=17), neck (n=4), not known (n=12) and tail (n=12), and the corresponding null hypothesis was again retained.

Pearson correlation analysis of QCancer® scores and CgA for the RFH pNET

samples revealed a coefficient of 0.102 (p=0.440) indicating no significant correlation. In contrast, correlation analysis of QCancerPancreas® scores and CgA levels revealed a coefficient of 0.258 (p=0.047), indicating a significant, weak correlation. Figure 5.6 illustrates the correlation between QCancerPancreas and CgA seen in the RFHpNET patients. In summary, the only clinical characteristic that was found to be significantly associated with CgA level was QCancerPancreas® score.
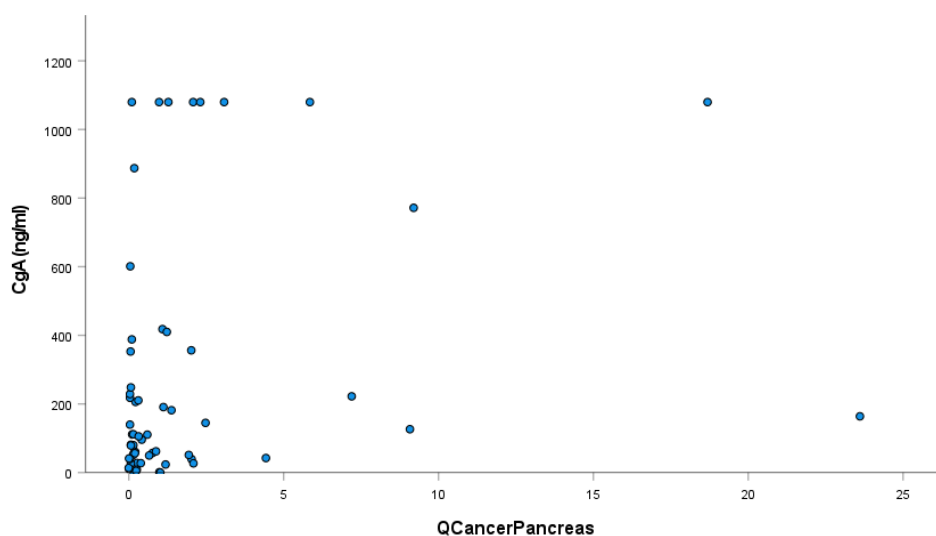
Figure 5.6: CgA levels and QCancerPancreas score
CgA levels and QCancerPancreas scores were compared in RFHpNET patients (n=60).

5.4.2 Analysis of potential CgA associated biochemical markers

CgA levels were assessed for association with biochemical parameters that were available in the patient records, including CEA (n=23), CA19-9 (n=28), CRP (n=43), creatinine (n=60) and bilirubin (n=60). The results showed that none of these biochemical marker levels was significantly correlated with serum CgA level (Table 5.5).

Biochemical Parameter	Available results	Pearson correlation coefficient	Significance (2 tailed)
CEA	23 out of 60	-0.49	0.823
CA 19-9	28 out of 60	0.240	0.219
CRP	43 out of 60	0.217	0.162
Creatinine	60 out of 60	0.174	0.184
Bilirubin	60 out of 60	-0.037	0.781

Table 5.5: Association of CgA levels of the RFH pNET samples with biochemical marker levels

The association of CgA plasma levels and CgA sera levels in the RFH pNETs was assessed for which both results were available. Samples that were at the higher limit of the ANG2 (sera assay) and the higher and lower limit of the CgA (sera assay) were excluded, thus a total of 39 RFHpNET samples were assessed for this analysis (Figure 5.7). A Pearson correlation coefficient of 0.278 was obtained p=0.086 indicating positive although not statistically significant correlation between CgA plasma and CgA serum levels, although plasma levels were generally higher than serum levels.
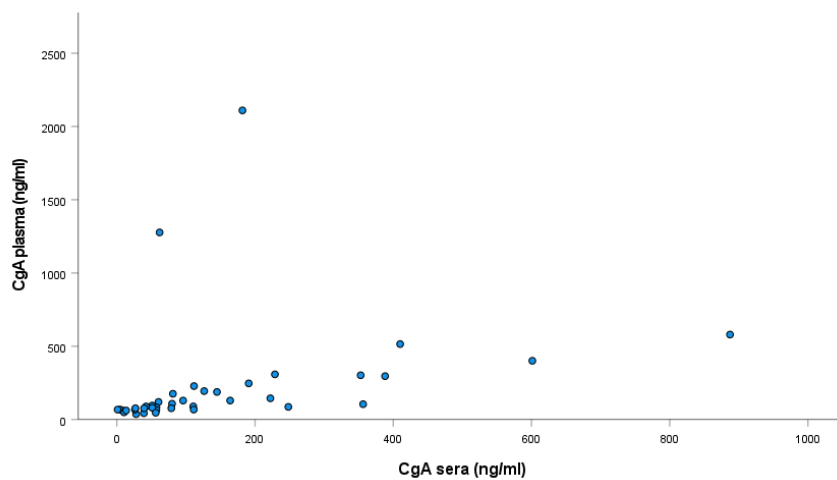
Figure 5.7: Comparison of plasma and serum CgA levels in RFH pNET samples.
Serum CgA levels obtained by ELISA using the CisBio assay were compared to plasma CgA levels determined as part of routine clinical care using the Diasource Assay (n=39).

In summary, the results for the three individual markers revealed that ANG2 and CgA levels are significantly elevated in the RFHpNET cases compared to the UKCTOCs controls (Table 5.6). However, this was not the case for VGF-NGF.

Marker	Independent-samples Mann-Whitney U test (p)
ANG2	0.018
VGF	0.324
CgA	<0.001

Table 5.6: Summary of results for ANG2, VGF and CgA levels in RFHpNETs vs UKCTOCs controls.

When comparing clinical parameters and assessing confounding factors for ANG2, only ANG2 and CA 19-9 levels were shown to be significant (Table 5.7).

For CgA, QCancer Pancreas score and CgA as well as CgA and PPI usage were found to be statistically significant (Table 5.7).

	Statistical Test	**Result**
QCancerPancreas score and CgA	Pearson Correlation	Correlation coefficient of 0.258 (p=0.047)
CgA and PPI usage	Independent Kruskal-Wallis	p=0.048
ANG2 and CA 19-9	Pearson Correlation	Correlation Coefficient of 0.790 p*<0.01

Table 5.7: Summary of statistically significant results seen for CgA and ANG2

5.5 Algorithm Validation

Models created using the UOL training dataset were externally validated using PNET sera in the RFH cohort. Due to the results in Section 5.3 demonstrating that VGF-NGF (V) was not a suitable marker, this marker was not assessed further and was not used in the external validation of algorithms. ELISA results for CgA and ANG2 samples that were at the upper and lower limit of detection were used for ML as well, as these results were likely not to be influenced by the thresholds used by the ML algorithms. Algorithms constructed from the AC combination, including C5.0 non-boosted decision tree, LR, SVM and RF models, using the UOL training dataset as described in Chapter 3 were evaluated. The validation results obtained for these models were compared to the corresponding results obtained for the A and C single-marker models. Sections 5.5.1 to 5.5.3 describe the results obtained for the AC, C and A models, respectively.

5.5.1 AC Model Validation

The AUC data obtained for the external validation cohort revealed that the best performing model was the RBF SVM with an AUC of 0.725 (Table 5.8, Figure 5.8). This was followed by the LR model with an AUC of 0.724 (Table 5.9, Figure 5.9) and the Linear SVM model with an AUC of 0.719 (Table 5.10, Figure 5.8). The worst performing models in the external validation were the RF models (Table 5.11) and C5.0 decision tree (Table 5.12, Figure 5.10).

AC Model	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
RBF SVM	0.725	65.00	64.70	35.30	35.00

Table 5.8: External validation of the AC RBF SVM algorithm.
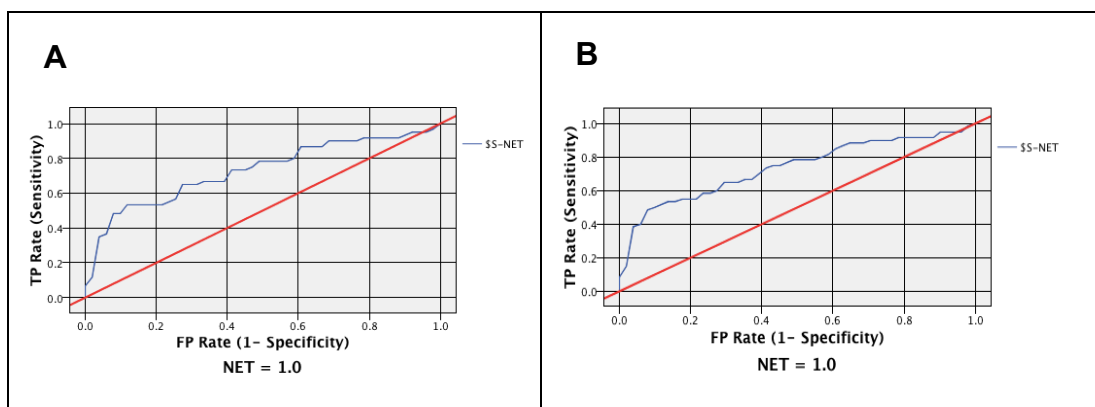
Figure 5.8: ROC curve analysis for the Linear and RBF SVM for the AC combination.
(A) ROC curve obtained for the Linear SVM algorithm with an AUC of 0.719. (B) ROC curve obtained for the RBF SVM algorithm with an AUC of 0.725.

AC Model	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
LR	0.724	88.30	31.40	68.60	11.70

Table 5.9: External validation of the AC combination using LR.
The AUC, TPR, TNR, FPR and FNR are shown for the external validation of this algorithm (n=111).
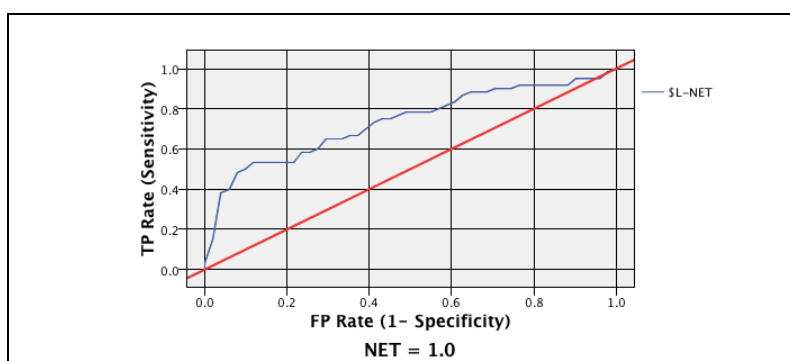
Figure 5.9: ROC curve for the external validation of the AC combination using LR.
ROC curve for the LR algorithm with an AUC of 0.724.

AC Model	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Linear SVM	0.719	66.70	64.70	35.30	33.30

Table 5.10 External validation of the AC Linear SVM algorithm.

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.500	80.00	27.50	72.50	20.00
2	0.613	78.30	27.50	72.50	21.70
3	0.600	86.70	25.50	74.50	13.30
4	0.616	85.00	21.60	78.40	15.00
5	0.596	85.00	19.60	80.40	15.00
6	0.597	78.30	31.40	68.60	21.70
7	0.558	83.30	25.50	74.50	16.70
8	0.576	83.30	25.40	74.50	16.70
9	0.608	81.70	23.50	76.50	18.30
10	0.560	81.70	27.50	72.50	27.50

Table 5.11: External validation of the AC combination RF models.
The models created for the 10 runs of the CA combination based on the UOL training dataset (n=106) were externally validated using the RFH cohort (n=111) of UKCTOCs controls and RFH pNETs.

C5.0 Decision Tree AC	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Non-boosted General setting	0.543	91.70	15.70	84.30	8.30
Non-boosted Accuracy setting	0.543	91.70	15.70	84.30	8.30

Table 5.12: External validation of the AC combination for the non-boosted C5.0 decision tree using both the general and accuracy settings.
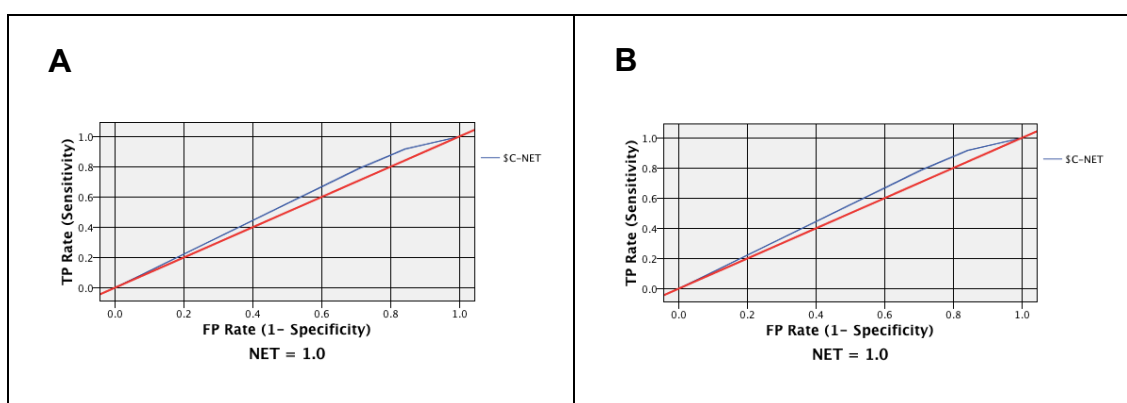
Figure 5.10: ROC curves for the external validation of the AC combination C5.0 Decision non-boosted decision tree.
(A) ROC curve for the general C5.0 non-boosted decision tree with an AUC of 0.543 (n=111). (B) ROC curve for the accuracy C5.0 non-boosted decision tree with an AUC of 0.543 (n=111).

The RF models had an AUC ranging from 0.500 to 0.616 (Table 5.11), while the C5.0 non-boosted decision tree had an AUC of 0.543 (Table 5.12) for both the general and accuracy models. The AC RF was evaluated across 10 runs, with

10 runs being chosen to assess whether there was a variation in AUC, sensitivity and specificity across the runs. The results for RF for the AC model in external validation show some variation across the AUCs, however this is to be expected due to the randomness involved in the generation of the tress of the RF. The TPRs across the runs ranged from 78.30% to 85%, the TNRs across the runs ranged from 19.60% to 27.50%, FPRs ranging from 68.60% to 80.40% and FNRs ranging from 15.00% to 27.50% (Table 5.11). This therefore suggests that across runs there is some variation. However, as the RF model is a black box model, understanding this in more detail is challenging. The external validation results across the different model algorithms (LR, C5.0 decision tree, SVM and RF) for ANG2 and CgA also demonstrated that performance at the validation level was lower than at the development level for all models.

The sensitivity and specificity levels set within each model were also assessed to allow a more detailed understanding model performance. The C5.0 decision tree model had the highest sensitivity (91.7%), but the associated specificity of this model was just 15.70% (Table 5.12). Similarly, the RF and LR models had high sensitivity (78.30 to 85.00 % and 88.30%, respectively), but poor specificity (19.60% to 31.40% and 31.4%, respectively;) (Table 5.11 and Table 5.9). In contrast, similar levels of sensitivity and specificity were apparent for the Linear SVM model (66.7% and 64.7%, respectively; Table 5.10) and the RBF SVM model (65.7% and 64.7%, respectively; Table 5.8). The best CA models based on AUC, sensitivity and specificity were therefore the SVM models. The TPR (true positive rate), TNR (true negative rate), FPR (false positive rate) and FNR (false negative rate) were also assessed for these models to reveal the impact of each model on each outcome group. The RF model had high TPRs (78.3 to 86.7% across the 10 runs) and low TNRs (19.6% to 31.4%, Table 5.11), but these models suffered from a high FPRs (68.6 to 80.4%, Table 5.11). The C5.0 non-boosted decision tree and LR models also had a high FPR (68.6% and 84.3%, respectively; Table 5.12 and Table 5.9). Notably, the SVM models had the lowest FPR of 35.30%, albeit with the highest FNRs of 33.3 to 35.0% (Tables 5.8 and 5.10).

5.5.2 C Model Validation

The single C marker RF, LR, and C5.0 decision tree models created using the UOL training dataset as described in Chapter 3 were externally validated using the independent sample cohort described in Section 5.4. SVM models cannot be generated for a single marker, as the simplest Linear SVM model requires two variables (x and y) to be separated by a hyperplane. For two input features the hyperplane is a line. The best-performing model based on external validation AUC was the LR model, which had an AUC of 0.698 (Table 5.13) with this model had a sensitivity of 60% and a specificity of 58.80%. This was followed by the C5.0 decision tree model, which had an AUC of 0.619 (Table 5.14, Figure 5.11) with a sensitivity of 76.70% and specificity of 47.10%. Unexpectedly, the C5.0 decision tree performed better at validation than the AC C5.0 decision tree (Section 5.5.1), suggesting that the addition of ANG2 data reduced performance.

C model	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
LR	0.698	60.00	58.80	41.20	40.00

Table 5.13: Externally validation of the single C marker LR model.
Results for the validation (n=111) of the LR model are shown.

C model	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Non-boosted General setting	0.619	76.70	47.10	52.90	23.30
Non-boosted Accuracy setting	0.619	76.70	47.10	52.90	23.30

Table 5.14: External validation of the single C marker C5.0 decision tree model.
Results for the validation (n=111) of the non-boosted C5.0 decision tree models generated using the general and accuracy settings.
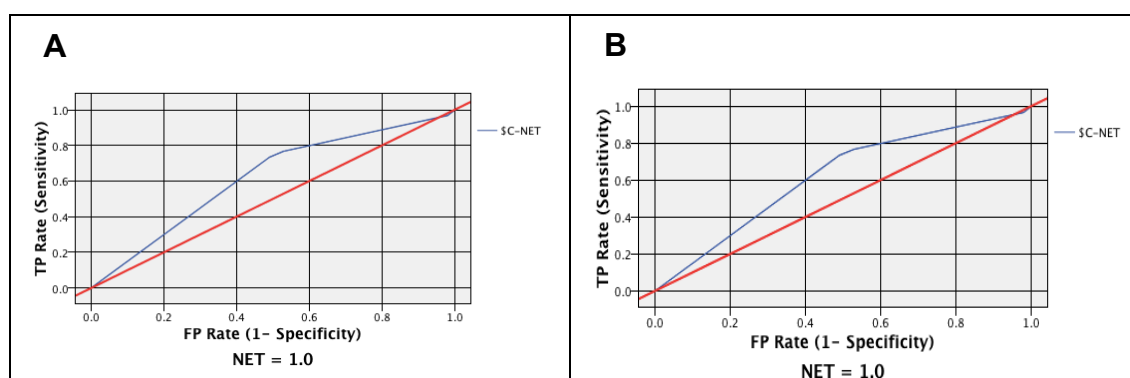
Figure 5.11: External validation ROC curves for the single C marker non-boosted C5.0 decision tree models.
(A) ROC curve for the external validation of the general C5.0 decision tree (n=111) with an AUC of 0.619.
(B) ROC curve for the accuracy model (n=111) with an AUC of 0.619.

The RF models performed the poorest, as found for AC model validation (Section 5.5.1) and produced AUCs across the ten runs ranging from 0.548 to 0.609 (Table 5.15), suggesting high degrees of overfitting in these models. When looking at the sensitivities and specificities across the ten runs, for the RF model, the sensitivities ranged from (70% to 80%) and specificities ranged from (29.40% to 49.00%) Thus, there was variation across the runs in sensitivity and specificities as well as AUCs for RF. The best validation TPR (80.0%) was seen for one of the RF runs, however the associated TNR was just 31.4% (Table 5.15). As seen for AC models, the FPR was high for the validation of all C models, ranging from 51.0% to 70.6% for the RF models (Table 5.15), 52.9% for the C5.0 decision tree (Table 5.14), and 41.2% for the LR model (Table 5.13). Nonetheless, the FPR values for C model validation were lower in general than those seen for CA models (except for SVM models).

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.590	78.30	29.40	70.60	21.70
2	0.596	70.00	49.00	51.00	30.00
3	0.584	76.70	31.40	68.60	23.30
4	0.635	78.30	31.40	68.60	21.70
5	0.586	76.70	29.40	70.60	23.30
6	0.609	80.00	31.40	68.60	20.00
7	0.584	78.30	29.40	70.60	21.70
8	0.604	76.70	31.40	68.60	23.30
9	0.598	76.70	33.30	66.70	23.30
10	0.548	78.30	33.30	66.70	21.70

Table 5.15: External validation of C marker RF models.
RF models created for the 10 runs of the C marker data in the UOL training dataset (n=106) were validated using the RFH cohort (n=111) of UKCTOCs controls and RFH pNETs.

5.5.3 A Model Validation

The single A marker RF, LR, and C5.0 decision tree models created using the UOL training dataset as described in Chapter 3 were externally validated using the independent cohort described in Section 5.2.

The best AUC (0.631) was seen for the LR model (Table 5.16, Figure 5.12), which was followed by the RF models which had AUCs ranging from 0.526 to 0.575 (Table 5.17), and the C5.0 decision tree which had an AUC of 0.539 (Table 5.18, Figure 5.13).

A Model	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
LR	0.631	80.00	29.40	70.60	20.00

Table 5.16: External validation of the A LR algorithm
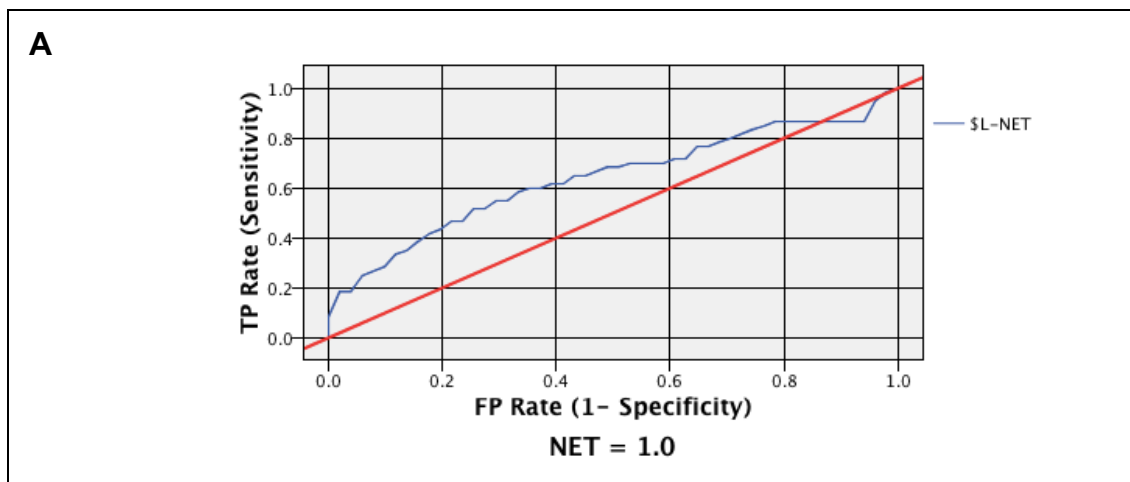
Figure 5.12: LR external validation for the A single marker
(A) ROC curve obtained for the LR for A external validation with an AUC of 0.631.

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.572	66.70	39.20	60.80	33.30
2	0.526	66.70	39.20	60.80	33.30
3	0.551	70.00	37.30	62.70	30.00
4	0.563	68.30	37.30	62.70	31.70
5	0.573	68.30	33.30	66.70	31.70
6	0.573	80.00	27.50	72.50	20.00
7	0.572	75.00	35.20	64.70	25.00
8	0.564	66.70	35.30	64.70	33.30
9	0.575	75.00	31.40	68.60	25.00
10	0.559	66.70	35.30	64.70	33.30

Table 5.17: Results for the external validation of the A single marker RF algorithm.

The models created for the 10 runs of the A single marker based on the UOL training dataset (n=106) were externally validated using the RFH cohort of UKCTOCs controls and RFH pNETs total of (n=111).

C5.0 Decision Tree A model	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Non-boosted General setting	0.539	78.30	29.40	70.60	21.70
Non-boosted Accuracy setting	0.539	78.30	29.40	70.60	21.70

Table 5.18: External validation of the A single marker non-boosted C5.0 decision tree using both the general and accuracy settings.
Results for the external validation (n=111) of the non-boosted C5.0 decision tree model for the general and accuracy setting. AUC, TPR, TNR, FPR and FNR are shown for both models.
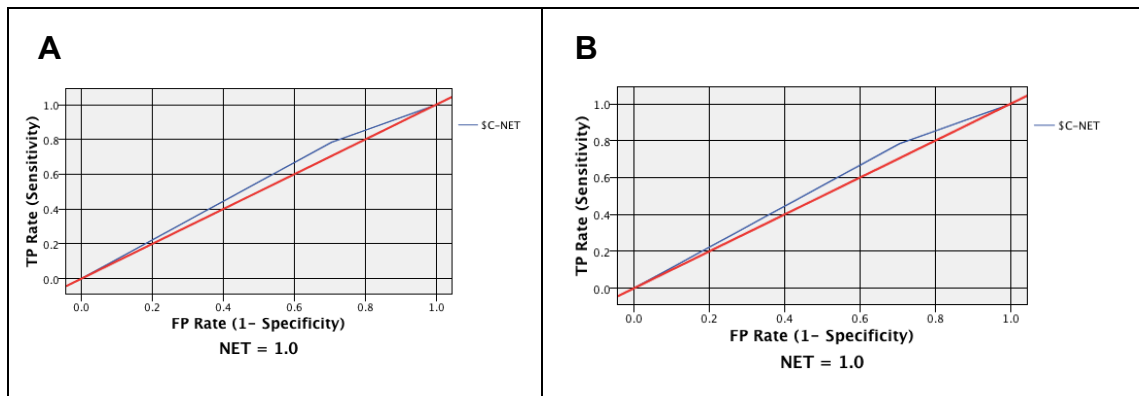
Figure 5.13: C5.0 Decision Tree external validation for the A single marker using the non-boosted ROC curve

(A) *ROC curve for the general C5.0 decision tree external validation (n=111) with an AUC of 0.539.* *(B)* *ROC curve for the accuracy settings (n=111) with an AUC of 0.539.*

The sensitivity and specificity for the LR model was 80% and 29.40%, respectively (Table 5.16), and 78.30% and 29.40%, respectively, for the C5.0 decision tree model (Table 5.18). The RF models had sensitivities across runs ranging from 66.70% to 80% and specificities ranging from 27.50% to 39.20% (Table 5.17).

The best validation TPR (80.0%) was seen for the LR model, which had an associated TNR of 29.4% (Table 5.16), and the RF model run 6, which had an associated TNR of 27.5% (Table 5.17). The FPR was high for all three model types, ranging from 60.8% to 72.5% for the RF models (Table 5.17), 70.6% for the C5.0 decision tree model (Table 5.18) and 70.6% for the LR model (Table 5.16).

A Model	AC	A	C
LR	AUC: 0.724 Sensitivity:88% Specificity:31%	AUC: 0.631 Sensitivity:80% Specificity:29%	AUC: 0.698 Sensitivity: 60% Specificity:59%
C5.0 Decision Tree (Generality)	AUC: 0.543 Sensitivity: 92% Specificity:16%	AUC: 0.539 Sensitivity: 78% Specificity:29%	AUC:0.619 Sensitivity: 77% Specificity: 47%
RF	Run 9 AUC:0.608 Sensitivity:82% Specificity:24%	Run 9 AUC:0.575 Sensitivity:75% Specificity31%	Run 4 AUC: 0.635 Sensitivity:78% Specificity:31%
Linear SVM C=10	AUC: 0.719 Sensitivity: 67% Specificity:65%	NA	NA
RBF SVM C=10 and gamma=0.7	AUC: 0.725 Sensitivity:65% Specificity:65%	NA	NA

Table 5.19: Summary table for the External validation of the algorithms developed for the AC combination using SPSS Modeler.

CA model	UOL Training cohort	RFH and UKCTOCs External validation cohort
LR	AUC: 0.849 Sensitivity: 72% Specificity:80%	AUC: 0.724 Sensitivity:88% Specificity:31%
C5.0 decision tree (Non-boosted Generality)	AUC: 0.835 Sensitivity:89% Specificity:76%	AUC: 0.543 Sensitivity: 92% Specificity: 16%
RF (Run 1)	AUC: 1.00 Sensitivity: 100% Specificity: 100%	AUC: 0.500 Sensitivity: 80% Specifcity:28%
Linear SVM C=10	AUC: 0.818 Sensitivity:74% Specificity:76%	AUC: 0.719 Sensitivity: 67% Specificity:65%
RBF SVM C=10 and gamma=0.7	AUC:0.821 Sensitivity:77% Specificity:74%	AUC: 0.725 Sensitivity:65% Specificity:65%

Table 5.20: UOL training and RFH and UKCTOCs External validation cohort algorithm performance comparison for the 2 marker CA combination.

C model	UOL Training cohort	RFH and UKCTOCs External validation cohort
LR	AUC:0.750 Sensitivity:58% Specificity:86%	AUC: 0.698 Sensitivity:60% Specificity:59%
C5.0 Decision Tree (Non-boosted Generality)	AUC: 0.813 Sensitivity:83% Specificity:76%	AUC:0.619 Sensitivity:77% Specificity:47%
RF (Run 6)	AUC: 1.00 Sensitivity:100% Specificity:98%	AUC: 0.609 Sensitivity:80% Specificity:31%

Table 5.21: UOL training and RFH and UKCTOCs External validation cohort algorithm performance comparison for the single marker C.

A model	UOL Training cohort	RFH and UKCTOCs External validation cohort
LR	AUC:0.758 Sensitivity:72% Specificity:76%	AUC: 0.631 Sensitivity:80% Specificity:29%
C5.0 Decision Tree (Non-boosted Generality)	AUC: 0.767 Sensitivity:72% Specificity: 80%	AUC: 0.539 Sensitivity: 78% Specificity:29%
RF (Run 3)	AUC: 0.996 Sensitivity:95% Specificity:98%	AUC: 0.551 Sensitivity:70% Specificity:37%

Table 5.22: UOL training and RFH and UKCTOCs External validation cohort performance comparison for the single marker A.

5.6 Further Algorithm Development

The external validation results described in Section 5.5 showed that the models generally had high FPR rates (>60%) and low specificity (<60%) with some model exceptions. Moreover, only certain algorithms, such as the LR and SVM

algorithms, achieved an AUC above 0.70, with AUCs of > 0.70 seen in the UOL cohort, and certain models such as RF dropping by large amounts at external validation, thus highlighting the inability of these markers and models to perform well in external validation. As these poor validation results indicated model performance varied widely with different patient data, a Combined dataset comprising the UOL dataset (Chapter 3) and the External Validation dataset (Chapter 5) was created in order to develop new algorithms. The hypothesis was that a larger dataset would produce better performing models that could be used and further tested on other independent cohorts in the future. The greater number of pNETs and controls within this cohort compared to each individual cohort was expected allow algorithms to better capture the heterogeneity of pNETs. To this end, C5.0 decision trees, LR model, RF and SVM (2-marker only) models were explored. C5.0 decision tree, LR and RF models were developed using SPSS Modeler, whereas Optimised Linear and RBF SVM models were developed using MCLA using methods previously described. Models were created for the ANG2 and CgA combination, and the CgA and ANG2 single markers.

The best ANG2 models developed using the Combined dataset were obtained for RF models, which produced an AUC between 0.978 and 0.994 (Table 5.23), followed by the LR model with an AUC of 0.663 (Figure 5.14). C5.0 decision trees obtained using the generality and the accuracy setting both produced an AUC of 0.656 (Figure 5.15).

Run	AUC
1	0.990
2	0.981
3	0.994
4	0.986
5	0.988
6	0.989
7	0.991
8	0.978
9	0.981
10	0.984

Table 5.23: Performance of the RF model developed from ANG2 data in the Combined dataset (n=217)
A total of 10 runs were carried out with the AUC is shown for the RF model developed by each run.
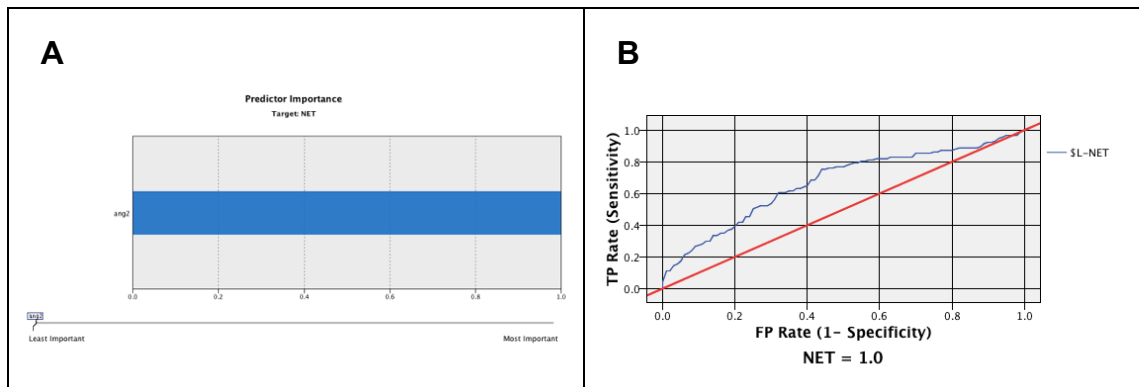
Figure 5.14: LR model Predictor Importance and ROC curve developed from ANG2 data in the Combined dataset (n=217)

(A) Predictor Importance=1.00 *(B)* ROC curve with an AUC of 0.663
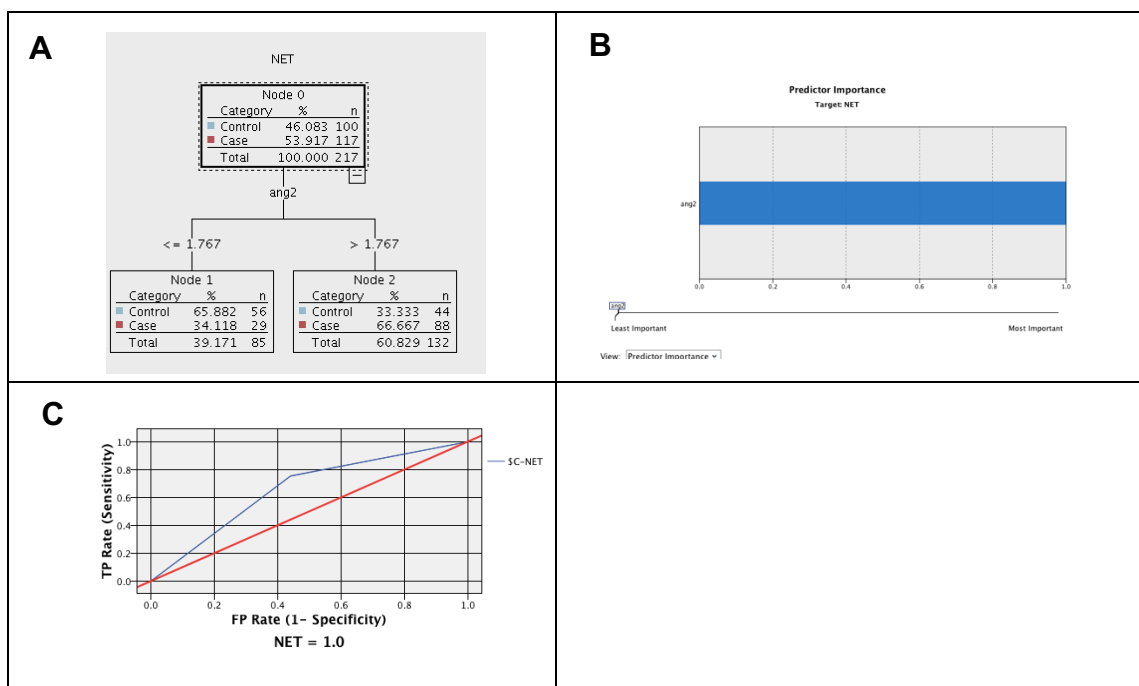
Figure 5.15: C5.0 decision tree model, Predictor Importance and ROC curve developed from ANG2 data in the Combined dataset (n=217)

(A) C5.0 Decision Tree model *(B)* Predictor Importance A=1.00 *(C)* ROC curve with an AUC of 0.656
The same model was obtained for the accuracy setting

Similarly, the best performing CgA models were the RF models which had an AUC ranging from 0.986 to 0.998 (Table 5.24), followed by the LR model with an AUC of 0.718 (Figure 5.16), and then the C5.0 decision tree model which had an AUC of 0.694 (Figure 5.17).

Run	AUC
1	0.986

2	0.987
3	0.998
4	0.987
5	0.988
6	0.993
7	0.992
8	0.987
9	0.986
10	0.987

Table 5.24: Performance of the RF models developed from CgA data in the Combined dataset (n=217)
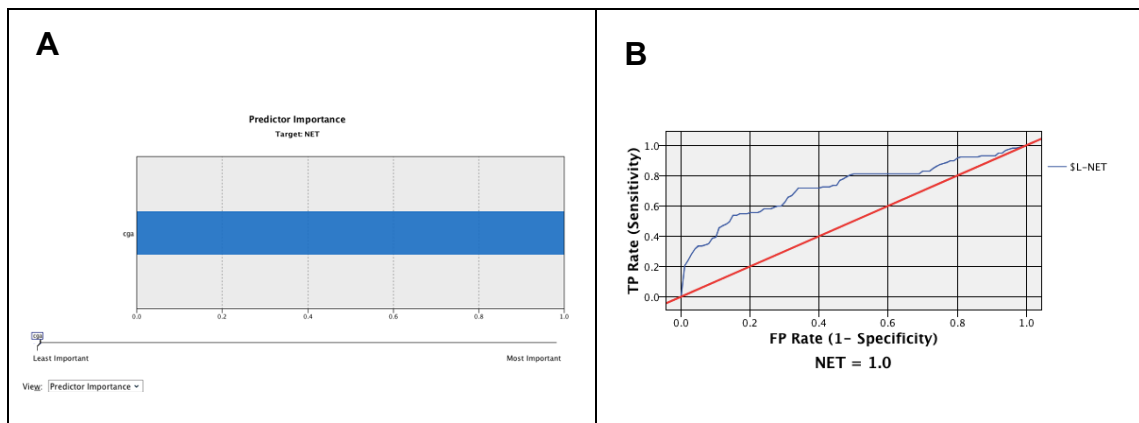
Figure 5.16: LR model and ROC curve for CgA using the combination dataset
(A) Predictor Importance C=1.00 *(B)* ROC curve with an AUC of 0.718.
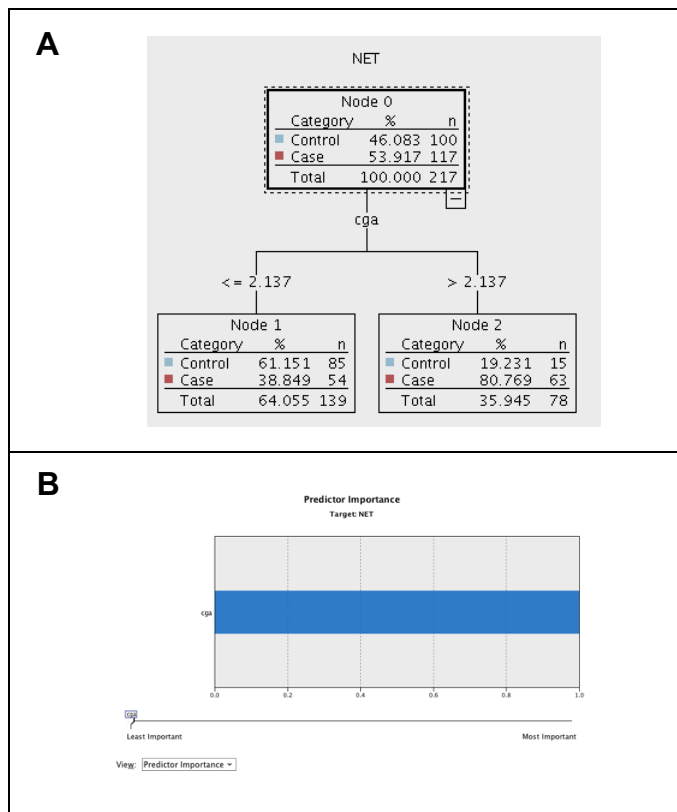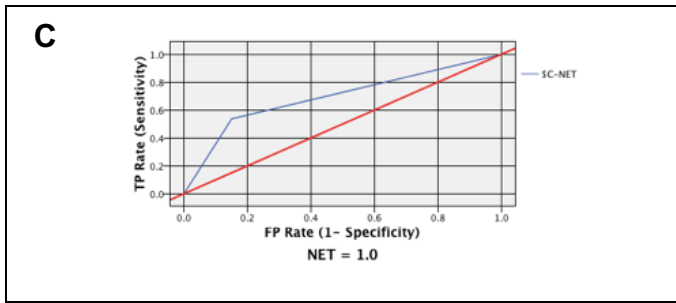
Figure 5.17: C5.0 decision tree model, Predictor Importance and ROC curve developed from CgA in the Combined dataset.
(A) *C5.0 decision tree model* *(B)* *Predictor Importance with C=1.00* *(C)* *ROC curve with an AUC of 0.694 The same decision tree and AUC were obtained for the accuracy setting.*

The Optimised RBF SVM and RF models were the best performing models developed from the CgA and ANG2 data taken together, as determined by their AUC values of 1.00 (Figure 5.18) and 0.996-1.000 (Table 5.25), respectively. The C5.0 decision tree was the next best performing model with an AUC of 0.777 (Figure 5.19), followed by the LR model with an AUC of 0.758 (Figure 5.20). The Optimised Linear SVM model was the worst performing model with an AUC of 0.660 (Figure 5.18).
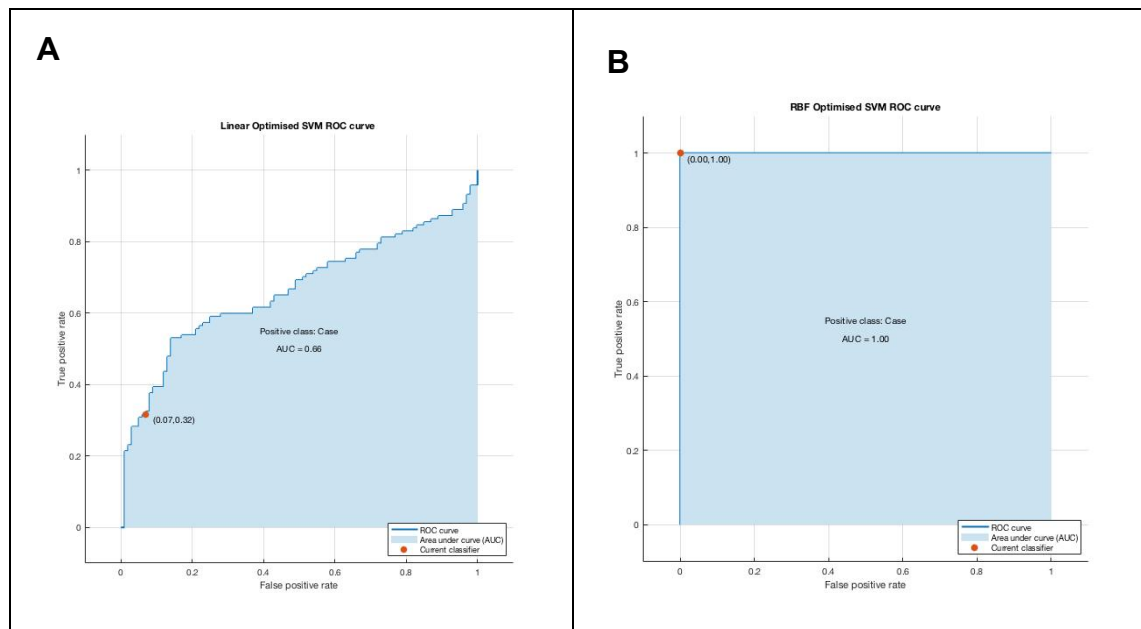
Figure 5.18: ROC curves for the Optimised Linear and RBF SVM models developed from AC data in the Combined dataset.
(A) *Optimised Linear SVM ROC curve with an AUC of 0.660* *(B)* *Optimised RBF SVM ROC curve with an AUC of 1.000.*

238

Run	AUC
Run 1	0.996
Run 2	0.999
Run 3	0.998
Run 4	1.000
Run 5	0.999
Run 6	0.999
Run 7	0.999
Run 8	0.997

Table 5.25: Performance of RF models developed from CA data in the Combined dataset.
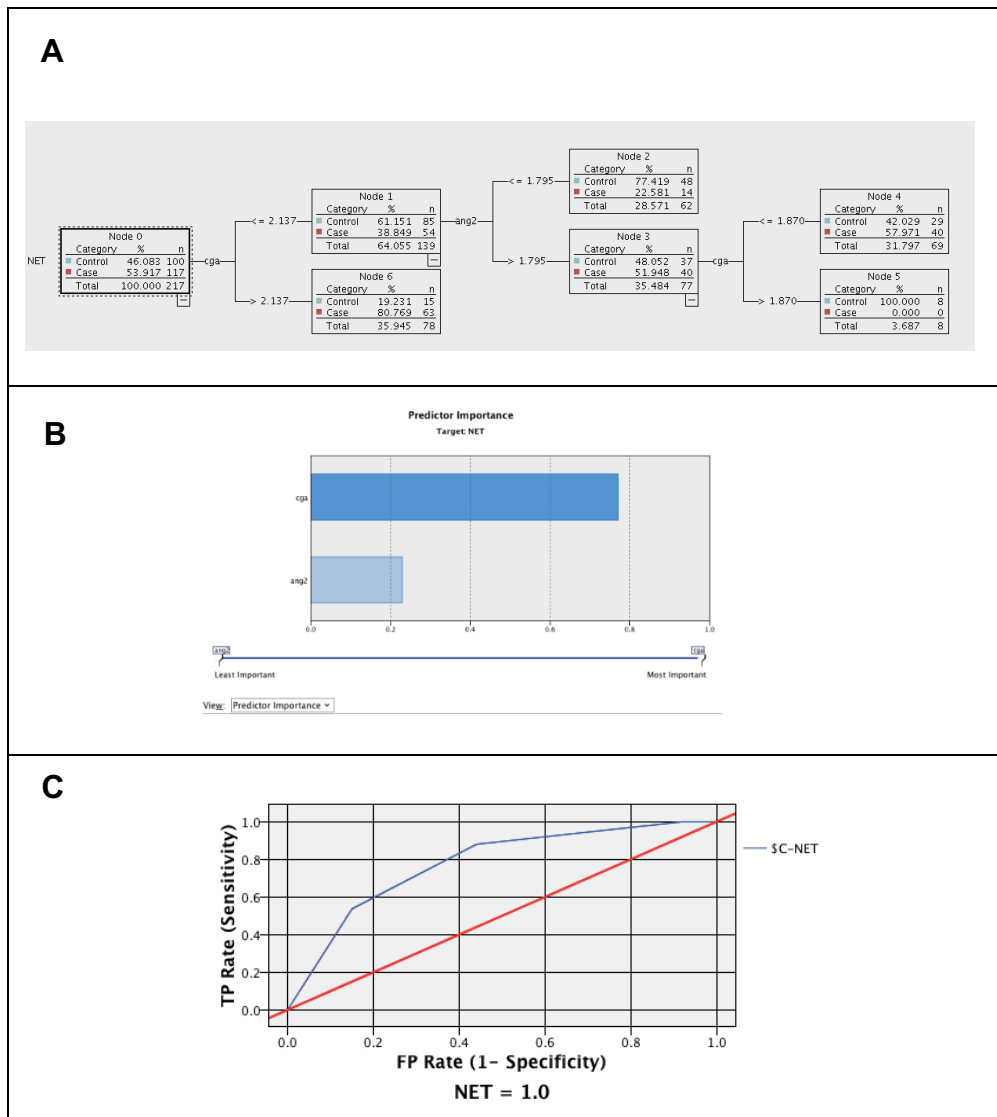
Figure 5.19: C5.0 decision tree model, Predictor Importance and ROC curve developed from CA data in the Combined dataset.

(A) C5.0 decision tree model **(B)** Predictor Importance C=0.77 and A=0.23 **(C)** ROC curve with an AUC of 0.777
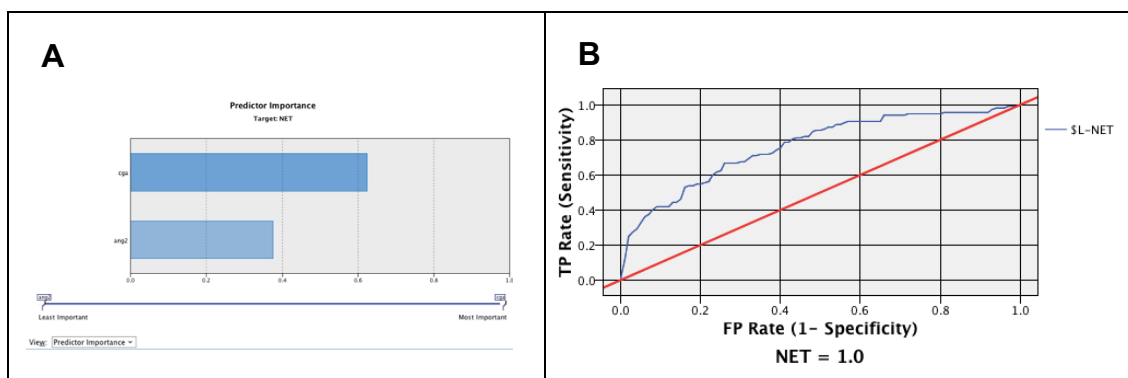
Figure 5.20: Performance of the LR model developed from CA data in the Combined dataset.

(A) Predictor Importance C=0.62 A=0.38 (B) ROC curve with an AUC of 0.758.

In summary, AUC and sensitivity were higher for all two-marker combined dataset LR, C5.0 decision tree and RF models compared to single-markers combined dataset models (Table 5.26). However, specificity did not similarly improve for all the algorithms (Table 5.26). The high performance of RF models was similar to that seen with the UOL training data (Chapter 3) and merited caution due to the high likelihood of overfitting. When comparing the results for AC vs the single markers of just A and C using the combined dataset, the two marker models performed better compared to the single markers of A and C across LR, C5.0 decision tree and RF. A sensitivity of >80% (88.03%) was seen for the C5.0 decision tree with ANG2 and CgA which was higher than that seen for the single marker ANG2 and CgA decision trees (Table 5.26). For the LR model, sensitivity didn't increase for the LR AC model compared to the A model, however sensitivity did increase for the LR AC model compared to the LR C model 59.80% vs 51.28% (Table 5.26). When comparing the model AUC results for the combined dataset for the 2 marker ANG2 and CgA combination and 1 marker ANG2 and CgA (Table 5.26) these AUC results were generally lower than the UOL training dataset (Tables 5.20, 5.21, 5.22).

	AC	A	C
LR	AUC: 0.758 Sensitivity: 59.80% Specificity: 79.00%	AUC: 0.663 Sensitivity: 63.25% Specificity: 63.00%	AUC:0.718 Sensitivity: 51.28% Specificity: 85.00%
C5.0 Decision Tree	AUC:0.777 Sensitivity: 88.03% Specificity: 56.00%	AUC: 0.656 Sensitivity: 75.20% Specificity: 56.00%	AUC:0.694 Sensitivity: 53.85% Specificity: 85.00%
RF	Run 4 AUC:1.000 Sensitivity: 99.15% Specificity: 100%	Run 3 AUC:0.994 Sensitivity: 92.31% Specificity: 99.00%	Run 3 AUC: 0.998 Sensitivity: 96.58% Specificity: 100%

Optimised Linear SVM	AUC:0.660 Sensitivity: 31.62% Specificity: 93.00%	NA	NA
Optimised RBF SVM	AUC:1.00 Sensitivity: 100% Specificity: 100%	NA	NA

Table 5.26: Summary of the results obtained for the performance of models developed from AC, A and C marker data in the Combined dataset.

5.7 Discussion

External validation using an independent cohort is important to detect overfitting to training data and to assess the generalisability of training-set models and markers. An example drawn from pancreatic cancer is the study by Melby et al., who developed a biomarker signature using a Scandinavian cohort and validated it using a cohort from the USA. They found that the biomarker signature was able to discriminate samples derived from patients with stage 1 and 2 pancreatic cancer from controls with a AUC of 0.96 and that the performance of the signature in the independent patient cohort was also 0.96. As their test performed well in both cohorts, there was no indication of overfitting and it was suitable to take forward for prospective validation[492]. Independent cohorts have also been used in the development of a marker panel for small intestine NETs (si-NETs). The panel was refined in training cohorts of plasma samples from 40 metastatic si-NET patients and 40 controls, and then validated in independent cohorts of 120 metastatic si-NET patients and 120 controls and found that elevated circulating levels of miR-21-5p and miR-22-3p and low levels of miR-150-5p are characteristic in patients who have metastatic si-NETs[493]. Independent cohorts were also used in a study of the prognostic relevance of the pNET markers Ubiquitin C-terminal hydrolase L1 (UCH-L1) and α-internexin and found that concurrent expression of both UCH-L1 and α-internexin were correlated with clinicopathological features that included prognosis in each independent collective and in the combination of the two cohorts[406]. As a final example, the NETest was independently validated using a cohort of 67 pNET and 44siNET patients and 63 controls, and was shown to correlate with imaging[494]. Hence the importance of external validation using independent cohorts has been highlighted across the field and is an important step in the pathway to biomarker implementation.

The work described in this chapter sought to validate the top three markers identified using the approaches to pNET biomarker model development outlined in Chapter 3 (UOL training) and Chapter 4 (UOL internal validation). From this the three most suitable markers identified to take forward to external validation were CgA, ANG2 and VGF. However, results described in this chapter for the external validation tests using the RFH pNET cohort, revealed that VGF levels were not elevated in cases compared to controls and it was therefore excluded as a pNET marker (Table 5.3). This finding highlighted the importance of external validation, as a marker that seemed to hold promise in the training stages was shown not have any value for an independent cohort. The reasons underlying the difference in VGF levels in the two pNET cohorts were not identified, but merit discussion.

VGF is a member of the chromogranin/secretogranin family. It is stored in large dense core vesicles located in neuroendocrine and neuronal cells,[161] and is localised to a number of neuroendocrine tissues including pancreatic endocrine cells[160]. Matsumoto et al. identified VGF in the conditioned medium of cells derived from a large cell neuroendocrine carcinoma via high-throughput mass spectrometry and thus proposed VGF to be a novel maker of NETs[495]. However, the mechanisms that elevate VGF expression in NETs and the biological effects of elevated VGF expression remain to be established. In pNENs, VGF binds to long non-coding RNA to activate the PI3K/AKT/CREB signalling pathway and promote the malignant phenotype and growth of pNENs[166]. Consequently, VGF was expected to be a marker for pNETs. While VGF has previously been shown to be elevated in pNEN tissue[166], it has not been identified as a pNET serum marker. However, despite VGF showing promise as a pNET marker for the UOL training cohort, this was not sustained in the external validation cohort. Further issues concerning VGF as a potential pNET marker arise from its role in a range of other conditions including Alzheimer's disease[496], Parkinson's disease[497], major depressive disorder[164], bipolar disorder[164], neuropathic pain[498], circadian rhythm[499], body fluid balance[500], diabetes[377], obesity[501], infertility[499] and sexual dysfunction[499]; as well as its elevation in NETs of the breast[502], large cell lung cancer[495], glioblastoma[503] and ovarian cancer[504].

The UKCTOCS healthy control external validation cohort had a mean serum VGF level of 523 ± 656.15 pg/m, while the RFH pNET external validation cohort had a mean serum VGF level of 311 ± 314.736 pg/ml (Table 5.3). These results could be interpreted as the control group value being much higher than expected or the pNET group value being much lower than expected, or both. To understand this further, other healthy control group data would need to be compared. The UOL training cohort consisted of healthy controls with a mean serum VGF level of 4500 ± 4711.56 pg/ml and pNET cases with a mean value of 8423 ± 7438.48 pg/ml. Thus, within the UOL cohort (n=106) the control and pNET groups had much higher mean VGF levels than seen at external validation. In the published literature control groups have mean VGF levels that are lower than that determined for the UOL training pNET groups but higher than that obtained for the external validation groups. For example, Chen et al. compared VGF levels in MDD patient, bipolar patient and healthy control sera[164]. The healthy controls used in the Chen et al. study had a mean VGF level of 1041 ± 54 pg/ml, compared to 994 ± 57 pg/ml for the MDD group and 1155.27± 45.72 pg/ml for the BD group. A study by Li et al., which compared healthy control to high- and low-risk suicide groups, found mean VGF concentrations of 1107 ± 155, 883 ± 140, and 1021 ± 132 pg/ml, respectively[505]. Thus, the mean levels of VGF in control groups were similar in the Chen et al. and Li et al. studies, but different from those I obtained for the UOL training cohort and UKCTOCS external validation cohort. Finally, there is no established reference range for healthy individuals. Taken together, these observations highlight the variability in control cohort results, but they do not shed light on the underlying reason for these differences. The VGF ELISA kit used for the training ELISAs was the VGF (Cloud-Clone Corp) which was the same kit used for validation ELISAs, thus kit variation is unlikely to be a factor for VGF. However, training and validation ELISAs were performed several years apart.

A limitation of the external validation cohort is that many of the patients were receiving different NET-related treatments, which could have influenced VGF levels. Some of the patients had previously received chemotherapy or were on NET-related medications such lanreotide or octreotide. Ideally treatment-naïve patients would have been recruited to minimise the impact of treatments, but this was not possible as patients were recruited at different stages of their care,

and at different disease stages. This reflects the real diversity of patients presenting with pNETs.

Further studies are needed to establish VGF mean values and reference ranges in pNET cohorts, and to identify confounding factors before VGF can be taken forward as a pNET marker. Gender is not a known confounding factor for VGF levels as its role as a confounder has not been explored in certain studies[164,505], and as this remains a possibility it would have been better for the all-female UKCTOCs controls to have matched the male and female RFH pNET group. Moreover, other currently unidentified confounders could have contributed to the higher VGF level in controls compared to cases, independent of pNET status. Identification of these confounding factors will be important to the future assessment of the suitability of VGF as a pNET marker. Ultimately, this could only be verified with further independent analyses with a much larger group of controls to allow the effects of multiple other conditions such as obesity, bipolar disorder or pain disorders, on VGF level to be assessed. Indeed, a relationship between VGF and obesity has been indicated by Rahimi et al., who found high VGF levels in the plasma of obese individuals with type 2 diabetes and of lean individuals (BMI<25)[377]. However, in the UKCTOCs controls used for the external validation, I found no association of VGF levels with BMI. A limitation for the UKCTOCs controls was that limited clinical data were available, including OCP usage, BMI, gender (all female) and ethnicity (all white). OCP usage was found to be a confounding factor in the controls and thus future studies should employ OCP usage as an exclusion factor. The number of controls on OCPs examined for VGF were 18 samples, with 17 samples from patients not on OCPs, thus exclusion of the OCP samples would not be possible due to the number involved being over half of the examined samples. The impact of OCP usage on pNET samples could not be explored as information on OCP usage by these patients was not specifically collected. Functionality of a pNET may also have influenced VGF levels, similar to the effect on CgA. However, functionality information was not available for the RFH pNETs, precluding this analysis. In conclusion I decided that VGF was not a suitable marker to take forward until further validation can be performed using additional, larger cohorts of pNET and healthy control samples.

ANG2 is a growth factor involved in angiogenesis, the well-studied and vital process of increased vascularisation in cancer growth maintenance and metastasis[470]. The external validation results reported in this chapter demonstrated that ANG2 levels were significantly elevated in pNET compared to control samples (p=0.018), with mean levels of 3999 ± 3354.01 and 2610 ± 1545.31 pg/ml, respectively (Table 5.1). This finding is consistent with work carried out by Srirajaskanthan et al., which found that serum ANG2 levels were significantly elevated in 47 GEP-NET patients compared to controls (median values of 4756 and 2495 pg/ml, respectively; P<0.001)[470]. However, while Melen-Mucha et al. found that plasma ANG2 levels were higher in 36 GEP-NET patients compared to controls, the difference did not reach significance[140]. It is also worth noting that these two published studies included different types of GEP-NETs; for example, the Srirajaskanthan et al. cohort had 17 pNET samples (36%), while the Melen-Mucha et al. cohort had 2 (6%).

While specific clinical parameters, including metastasis, grade and location of the tumour, did not significantly associate with ANG2 levels in the external validation results, it is worth noting that metastasis patients had a higher mean ANG2 level compared to non-metastatic patients in the external validation. I inferred from these results that ANG2 level may be related to tumour size and that studies with larger patient cohorts may allow significance to be reached. Srirajaskanthan et al. found that serum ANG2 levels were statistically significantly elevated in GEP-NET patients with distant metastases compared to those without metastasis (median 5080 vs 3360 pg/ml P=0.01)[470]. Their study also found a significant association between ANG2 level and volume of liver metastases (P=0.014), and time to disease progression was worse in patients with serum ANG2 levels >4756 pg/ml (P=0.04)[470]. A relationship between metastasis and ANG2 is also supported by Melen-Mucha et al., who found that plasma ANG2 levels were significantly higher in GEP-NET patients with metastatic disease than those with localised disease[140]. The results I obtained for grade in the external validation cohort indicated no significant association but were compromised by the presence of only one G3 pNET sample, which was too low to be included in my analysis.

My external validation results showed that BMI and OCP usage were not confounding factors of ANG2 level in the control samples. PPIs, having another cancer, diabetes and gender were not confounding factors for ANG2 level in the

external validation pNET samples. Similarly, there was no significant correlation between ANG2 and either QCancer® or QPancreas® score. My analysis of the CEA, CA19-9, CRP, creatinine and bilirubin biochemical markers, which were taken as part of the RFH pNET patient clinical work up, showed that only CA19-9 level ($p<0.01$) correlated significantly with serum ANG2 level (Table 5.2). This finding was interesting as CA19-9 has not previously been reported to be a pNET diagnostic marker. However, a study by Luo et al., indicated that CA-19-9 was a prognostic biomarker in pNETs with a cut off value of CA19-9 > 16U/ml being an adverse prognostic factor for patients overall survival[506]. They also found that CA19-9 > 16U/ml group had a statistically higher proportion of stage III or stage IV patients compared with CA 19-9<16U/ml. Thus, suggesting that CA19-9 is a prognostic biomarker for pNETs with this reflecting aggressiveness and severity of disease. The statistically significant correlation seen between CA19-9 and ANG2 in the RFH pNET patients in my study could be a reflection of increased aggressiveness and severity of disease. However, CA19-9 is an established PDAC marker[507]. Although CA19-9 has been suggested to accelerate PDAC angiogenesis[508], and while PDACs are characterised by an extremely high potential for invasion and metastasis with angiogenesis playing a crucial role in this process[509], ANG2 is not known to be a PDAC marker. Consequently, the correlation between ANG2 and CA19-9 levels seen in pNET patient sera appears not to arise from any similarity with PDAC patients.

A limitation on ANG2 usage clinically is the large number of confounding factors associated with this marker (Section 1.5.5) and implementation of this marker clinically may be challenging as a result.

External validation of higher CgA levels in pNET relative to healthy control serum (Table 5.4) showed that case levels (mean 278 ± 362.558 ng/ml) were indeed elevated compared to control levels (mean 101 ± 299.132 ng/ml). These results were similar to those obtained for the UOL training cohort, for which the mean CgA serum levels in the UOL pNET and control cases were 237 ± 434.48 ng/ml and 48.68± 49.42 ng/ml, respectively. Hence, both cohorts provided support for CgA as a pNET marker, which was expected based on the volume of similar published evidence.

My assessment of confounding factors in the external validation cohort showed that OCP usage and BMI were not statistically significantly correlated with CgA sera levels in UKCTOCs control samples, while PPI usage was a confounding

factor for the CgA level of pNET samples (p=0.048), consistent with previous studies[462,510,511]. This is an important limitation of CgA as a pNET marker as patients would need to stop taking PPIs to allow reliable CgA analysis. Clinical characteristics including MEN1 genotype, metastasis status, and grade were found not to associate significantly with CgA level. However, as previously mentioned, there were limitations to the grade analysis. The only clinical characteristic found to significantly associate with CgA serum levels was the QPancreas® score, which is a specific value developed from the QCancer® questionnaire by Professor Julia Hippisley-Cox. This QCancer® questionnaire aims to provide a prediction of cancer risk (a QCancer® score) based on symptoms and risk factors with specific questionnaires for males and females. The QPancreas® score is one of the additional scores provided in the QCancer® questionnaire that assesses the risk of pancreatic cancer in patients based on specific symptoms and risk factors. Thus, it is particularly interesting that CgA serum levels in the RFH pNET cohort associated with QPancreas® scores. This was not seen with QCancer® scores, suggesting that a specific combination of symptoms and risk factors that associate with pancreatic cancer are also associated with increased CgA level. Further study into this, including a more granular analysis of individual symptoms, would be useful in the future.

A limitation of the QPancreas® score is that it does not distinguish between PDAC and pNET patients. Serum CgA association with other biochemical parameters including CEA, CA19-9, creatinine, CRP and bilirubin in the external validation pNETs was not significant. When assessing the relationship between CgA sera and CgA plasma a Pearson correlation coefficient of 0.278 was obtained p=0.086 indicating a positive although not statistically significant correlation between CgA plasma and CgA serum levels, although plasma levels were generally higher than serum levels. CgA plasma levels could only be assessed in the RFHpNETs due to this test being routinely carried out for the patients in the RFH clinic, however this could not be assessed in the UKCTOCs controls and this is a limitation for the analysis. Assessment of whether plasma levels of CgA were higher compared to sera irrespective of disease status could not be examined as a result. However, a study by Lim et al., found that serum and plasma samples from self-reported healthy doors had a reference range determined to be 6-95ng/ml in serum and 20-146 ng/ml in plasma[512], as measured by the CisBio CgA ELISA kit. Hence, suggesting that plasma levels

of CgA are higher than sera levels regardless of disease status. The higher sera levels of CgA compared to plasma levels of CgA is an important assessment and has implications as to whether sera or plasma CgA ML models are more optimal to develop, if the ideal model should have generalizability to both sera and plasma results. ML models built using plasma levels of CgA are likely to have thresholds that are higher than that used for sera CgA to discriminate between pNETs and controls, meaning that sera CgA pNET data if used on these developed models is likely to be lower than the threshold used for plasma and thus these patients would not be picked up by the plasma developed ML models. However, models developed using sera CgA would likely have discriminant thresholds that between case and control that are lower, hence these generated models would still be able to identify pNET patients based on CgA plasma results.

CgA is currently the most useful NET marker for establishing a diagnosis. It has also been shown to have value in the prediction of disease recurrence, patient outcomes and efficacy of treatment[513]. However, there is no Food and Drug administration (FDA)-approved CgA assay for NET diagnosis[512]. CgA can be measured using different antibody-dependent assays including ELISA, immunoradiometric assay (IRMA), radioimmunoassay (RIA), and a more recent immunofluorescence assay based on Time-Resolved Amplified Cryptate Emission (TRACE)[514]. The RFH uses the Health Services Laboratories (HSL labs) for plasma CgA analysis as part of their routine care of NET patients, and HSL labs use the DiaSource CgA ELISA assay. The UOL training dataset was produced using the DAKO CgA ELISA assay, whereas the RFH external validation dataset was produced using the CisBio ELISA assay. The use of different ELISA assays may have been a limitation in comparing the training and validation data due to a lack of standardisation, however as the UOL cohort and RFH cohort samples were not tested with both ELISA kits, the impact that this could have had is not known.

In this regard, a study by Lim et al. found that different CgA ELISA kits for the same sample type (sera) obtained different results, as serum levels measured by the CisBio ELISA were 8-9 times higher than those obtained by a reference laboratory (Quest Diagnostics) CgA ELISA assay[512]. If CgA is to be employed widely as a NET marker, CgA assays will need to be standardised. Other issues remaining for CgA assay measurements and their utility, which include

determining the upper limit of normal in a population, sample stability (which is the temperature at which the sample is maintained prior to assay)[515] and developing systems that can measure the presence and absence of specific CgA peptide fragments[516]. Study design issues also remain for CgA including the lack of age and gender matching and addressing known confounding factors for CgA. This also complicates algorithm development and threshold selection. Thus, ideally longitudinal analyses should use the same CgA assay and a patient should be re-baselined if the CgA assay changes.

A study by Hijoka et al., found that the mean CgA levels of patients with pNET was significantly higher than controls, with this being 407.8 ± 984.6 ng/ml (n=69) compared to 62.5 ± 48.3 ng/ml (n=112)[517]. Another study by Qiao et al., found in pNET patients that sera levels of CgA in insulinomas (n=57) (median 64.8 ng/ml, range 25-164) were slightly higher than in healthy controls (n=86) (median 53.4 ng/ml range 39-94) but lower than the non-insulinomas (n=32) (median 193ng/ml range 27-9021)[90]. The RFpNETs from the external validation in my study had a mean CgA of 277.765 ± 362.56 ng/ml compared to 100.996 ± 299.132 ng/ml in the UKCTOCs controls. Thus, the elevations seen in CgA in pNETs was consistent to that of other pNET cohorts. The UOL pNET cohort from the training cohort in my study had a mean CgA of (237.48 $\pm$ 434.48 ng/ml) (n=57) and the UOL training healthy controls had a mean CgA of 48.68 $\pm$ 49.42 ng/ml (n=49). Thus, the findings from the UOL and RFH cohorts are consistent with other pNET cohorts which reported elevations in CgA compared to controls.

Moreover, CgA has been suggested not to be a good clinical marker for NETs. Factors for CgA not being a suitable marker include confounding factors associated with CgA. CgA is known to be elevated in various malignancies as mentioned in Section 1.5.1 of the thesis. CgA elevations in other malignancies means that it hampers utility of CgA as a marker for NETs in patients. However, it isn't also malignancies in which CgA levels have been shown to be elevated in, but in cardiac inflammatory and renal insufficiency[449]. Moreover, background levels are variable in different populations and a study has shown that postprandial serum CgA levels increase significantly in patients with ECL cell hyperplasia secondary to long term PPI usage as well as similar trend observed in normal controls. Thus, when CgA is measured in NET patients it should be measured in blood samples from fasting patients[518]. Moreover a study by Tseng

et al., showed that plasma CgA level was associated with tumour size, metastasis and tumour stage and that for early stage pNETs CgA exhibited a limited role in diagnosis[519].

A study examining NF-pNETs with low metastatic burden showed that preoperative CgA levels were elevated in only 10 out of 47 patients[520]. In a study of Japanese pNETs, CgA ELISA showed sensitivities and specificities of 53.6% and 78.6%[517].

When comparing algorithm performance between the UOL training dataset and the RFH external validation cohort for the CA combination, C and A single markers, the training AUC was always greater than the validation AUC. This was to be expected as the models were based on the training data. However, the impact of overfitting was higher for certain models than for others. The impact for the CA combination were particularly evident for the RF and C5.0 decision tree models, for which the training AUCs of 1.00 and 0.835, respectively, dropped in external validation to 0.500 and 0.543, respectively. The poor performance of the CA C5.0 decision tree could be explained by the thresholds specified by the tree. Specifically, the higher FPR and lower TNR rates suggest the ANG2 threshold of 1.758 ng/ml and CgA threshold of 1.894 nmol/L were too low for the external validation cohort. An explanation for the poorer performance of the CA RF model at external validation could be based upon the original model being too highly complex in that it focussed more on noise as opposed to real patterns within the data and thus had an inability to generalise to an independent dataset. The CA LR, Linear SVM and RBF SVM models proved to be more robust. The LR CA model had a training AUC of 0.849 which dropped to 0.724 in the external validation. The CA LR model had a training sensitivity and specificity of 72% and 80% with the sensitivity and specificity in external validation for this model being 88% and 31% respectively. Thus, showing the specificity was impacted at external validation and not the sensitivity. The CA Linear SVM model which had a training AUC of AUC of 0.818 dropped to an AUC of 0.719 at external validation. The training Linear SVM CA model had a sensitivity and specificity of 74% and 76% respectively with both the sensitivity and specificity dropping at external validation to 67% and 65%. RBF SVM CA model which had a training AUC of 0.821 with this dropping to an AUC of 0.725 in external validation. The sensitivity and specificity were also impacted, with the sensitivity and specificity at the UOL

training cohort for this model being 77% and 74% compared to 65%. Thus, when factoring AUC, sensitivity and specificity, the RBF SVM CA model performed the best in the external validation.

Amongst the CgA single-marker models, the LR model was the most robust with an AUC of 0.750 in training which decreased to 0.698 in the validation cohort. For the C5.0 decision tree and RF models, despite having stronger performances in the UOL training cohort of 0.813 and 1.00, respectively, these dropped to 0.619 and 0.609, respectively. Consequently, the LR model had the lowest difference between training and validation AUCs, indicating it was the least impacted by overfitting. The poorer performance of the C5.0 decision tree model at external validation could be explained by the thresholds that were used in the training decision tree model. Specifically, the CgA C5.0 decision tree training model contained splitting nodes with threshold CgA values of 1.305 nmol/L followed by 1.126 nmol/L, followed by 1.026 nmol/L. When looking at the nodes of the CgA non-boosted decision tree that was created using the training data, the additional smaller threshold splits at 1.126 nmol/L and 1.026nmol/L were utilised to further separate controls and cases due to the initial 1.305nmol/L threshold not being sufficient to do this. These thresholds for identifying pNETs appeared to be not high enough for the RFH external validation. To explore what thresholds would be suitable for the RFHpNET and UKCTOCs external validation cohort, this cohort of n111 was used as a training cohort with a C5.0 non boosted decision tree created for CgA to explore what threshold would be suitable based on that cohort. From this, it was seen that the C5.0 non-boosted decision tree that was used a threshold of 5.815nmol/L which was much higher than that seen based on the UOL training data and illustrating that discriminating against case and control based on lower thresholds would not be suitable for the RFH and UKCTOCs data and highlights the challenges in model development.

Additionally, the more nodes a decision tree has the more likely that the tree has been overfitted to the training data, reducing its ability to generalise to other datasets. The more complex a decision tree model is, the less reliable it would be to predict future records, with an extreme scenario for this being a very complex decision tree model that spreads wide in making the records in each lead node 100% pure (all assessed records having the target outcome)[397]. However, such a decision tree would be overfitted to the used observations and

have few records in each leaf and thus not reliably predicting future cases which means it would have poor generalisability[397]. Prevention of this involves incorporating stopping rules with stopping rules including the minimum number of records in a leaf, the minimum number of records in a node prior to splitting and the depth (number of steps of any leaf from the root node[397].

A similar trend was seen for the ANG2 single-marker models, with the LR model being least impacted by overfitting, while the RF and C5.0 decision tree models were substantially impacted. Consequently, the threshold of 1.758 ng/ml in the ANG2 decision tree training model was likely lower than that required for the external validation ANG2 data, resulting in high FPR and low TNR values. To assess the threshold of the C5.0 non-boosted decision tree when using the external validation cohort for decision tree creation, a decision tree model was not shown in the software and the resulting AUC of the C5.0 decision tree model was 0.500, suggesting that based on this one model type that ANG2 based on the external validation data used for training was poor and not a suitable marker. The AC C5.0 non-boosted decision tree model created using the external validation cohort reproduced the same decision tree threshold model when just CgA with a threshold of 5.815 nmol/L used (Appendix E). This further highlighted that ANG2 did not have any utility for C5.0 non boosted decision tree models and highlighted different training cohorts could reveal different markers that were more suitable.

The low validation AUCs of CgA single-marker models relative to corresponding training AUC values reveals a generally poor validation performance of CgA as a single pNET marker. These findings are similar to those seen in other case-control studies of CgA level in pNET patients. For example, a study by Pulvinetti et al., which assessed CgA levels in 99 pNET patients and 21 healthy controls, revealed an AUC of 0.77[462], while Miki et al. found an AUC of 0.78 in their study of 91 patients and 104 healthy controls[521]. ANG2 single-marker models performed worse in external validation than CgA models, achieving AUC values of 0.631 (LR model), 0.539 (C5.0 decision tree model) and 0.575 (RF model). This similarly highlights the unsuitability of ANG2 as a single marker for pNET detection.

The CA two-marker approach performed slightly better in external validation than the respective single markers. The CA models produced the highest AUC values at validation for the RBF SVM and LR models, with AUCs of 0.725 and

0.724, respectively. However, these AUCs were still low <0.80, and the LR model despite having a high sensitivity of 88% suffered from a poor specificity of 31%. The Linear SVM model and RBF SVM model had higher sensitivities of 67% and 65%, respectively with specificities of 65%. Thus, suggesting that better performing models and markers were needed.

Thus, based on the findings from the external validation of single CgA and ANG2 models and CA combination models across model types, these model performances were not good enough when compared to training performance. Thus, an alternative approach was created, whereby the training dataset and the external dataset were combined to create a larger dataset. The rationale for this being that this larger dataset could allow for more suitable thresholds and models to be created, that encompass the variations that are seen between the training and the validation cohort.

To improve model performance by using a larger training dataset, new C, A and CA models were derived from a combination of the UOL and RFH cohorts. These will require future validation using independent, large pNET cohorts, which was beyond the scope of my thesis. The same types of algorithm as those explored in Chapter 3 (UOL training), namely Optimised Linear and Optimised RBF SVM, non-boosted C5.0 decision tree, LR and RF algorithms were used for model development. Linear and RBF SVM using the SPSS Modeler approach of manual grid optimisation were not carried out due to the disadvantages to this approach compared to the MCLA approach, disadvantages of the SPSS Modeler approach including the inability to optimise directly within the software and a parameter selection which may not necessarily contain the best performing model, and the time consuming nature of a manual grid optimisation in SPSS Modeler in comparison to an automated grid search approach in MCLA. The best-performance was seen for the RF model (Run 4) with an AUC of 1.000, and the Optimised RBF SVM model which had an AUC of 1.00 for the CA combination, while the LR and the C5.0 decision tree models had lower AUCs of 0.758 and 0.777, respectively. The performance of CgA LR, C5.0 decision tree and RF models was again better than the corresponding ANG2 models. When looking at the combined dataset results, the AUCs were lower than that seen for when the UOL training set individually was used for the CA marker for LR and C5.0 decision trees. The UOL training dataset had AUC for the CA markers of 0.835 for the C5.0 decision tree and

0.849 for the LR model, thus comparing this to the LR and C5.0 decision tree model results obtaining the combined dataset, the performance of these markers is poorer in the larger combined training set compared to the smaller UOL alone training set. Thus, based on these results, CgA and ANG2 in combination may not be the most suitable model to take forward and thus, other markers would also need to be analysed in conjunction. The higher performances seen for RF and Optimised SVM were similar to that seen with the UOL alone training cohort with the UOL training cohort having an average AUC for the CA markers of 0.998 and the Optimised SVM model performing the same in the combined training dataset and the UOL training dataset with this being an AUC of 1.00.

A limitation for my work, is that only certain ML algorithms were explored, including C5.0 decision trees, RF, LR and SVM. There are other known supervised ML algorithms which haven't been explored as part of this work, which may have performed better in comparison to these other algorithms. Examples of other supervised ML algorithms which have been assessed in the context of other biomarker research include neural networks used in the development of an early diagnosis biomarker model for Alzheimer's disease [522], naïve bayes and k-nearest neighbour with both modes used in the identification of MDD markers[523] and LDA[524]. However, these algorithms also have disadvantages, such as with neural networks being black box[525] and LDA assuming the data to be normally distributed[526]. Thus, as with the algorithms used within the thesis, the disadvantages of other supervised ML algorithms would need to be considered before these are assessed in this work. Moreover, future studies should include other markers, such as TIMP1 and MAC2BP, in this larger cohort to assess the performance of models derived from combinations of these markers with CgA, ANG2, or both.

Information on functionality will be important to understanding variations between different types of pNETs, including non-functional and functional types as well us functional subtypes, as a lack of functionality data limited my work, and it is likely that different types of pNET have increased or decreased levels of a marker. For example, it is known that CgA levels are not a good predictor of insulinomas[90].

An important limitation of my pNET cohorts was their size, and large multicentre pNET cohorts are needed in the future for biomarker and algorithm

development and validation. Studies such as the Malczwewska et al., study of the NETest for the diagnosis of pNETs had a small cohort of pNETs n=67[494], and the Lv et al., study of CgA and NSE in NF-pNETs had a cohort of NF-pNETs n=167[468]. Hence pNET marker studies have variations in the number of pNET samples used in their assessment. This variation in sample size is perhaps due to the rarity of pNETs and availability of pNET samples. Thus, large cohort studies for biomarkers being more challenging. When the sensitivity of a diagnostic test is not known a range of sensitivities would need to be considered in order to cover a range of potential sensitivities of the diagnostics test for example from 60% to 90%, across a range of d (maximum marginal error values) for example 0.1 to 0.2. When the sensitivity of test is lower the of pNETs based on sensitivity needed would be higher than that needed if the diagnostic test was 90% sensitive vs 60% sensitive. The higher the maximum margin of error the fewer samples that are required. Moreover, having large multicentre cohorts would be important particularly for generating large training and validation cohorts and to assess biomarkers based on pNET functionality, stage and grade as it is important that the impact of biomarkers on these characteristics can be assessed without limitations of insufficient numbers of a particular grade or type of pNET based on the available samples and hence having a larger number of samples allows for a greater granular assessment to be made.

Further understanding of the heterogeneity of pNETs is needed for better marker development in the future. pNETs themselves are extremely diverse, due to the differences in functionality (functional or non-functional) types, differences in clinical presentation and prognosis which is due to functional status, genetic associations and variability in disease aggressiveness. Thus, studies exploring biomarker utility for diagnosis of pNETs should include cohorts which can fully explore these factors in a detailed way, which could lead to certain markers being identified for certain types of pNETs compared to others, as seen in the case of insulinomas and CgA not being a suitable marker[90]. When understanding the impact of the findings at validation compared to training the cohort heterogeneity is something considered. The UOL training cohort was diverse in terms of tumour grade and type of pNET. The cohort mirrored the heterogeneity of pNETs that are most encountered in clinical

practice but was not identical to the RFH external cohort in terms of certain characteristics such as staging. Staging for the UOL cohort was limited as previously described and a limitation for this cohort. For the external validation RFH pNET for ANG2 and CgA a similar number of metastatic and non-metastatic pNETs were present within the cohort. Understanding the staging composition of the cohort is important as a training cohort with more metastatic cases and less early-stage cases is different to a training cohort with which there are less metastatic and more early-stage cases, and the composition of such a cohort on which the models are developed is a consideration. Hence the composition of the different pNET cohorts themselves could have impacted performance at validation. However incomplete staging for the UOL cohort hinders further understanding of this. The UOL cohort contained a similar number of 4 confirmed insulinomas (n=4), gastrinomas (n=3) and MEN1 disease (n=4). Therefore, the number of confirmed insulinomas, gastrinomas and MEN1 related pNETs were similar and insulinomas did not make up a large portion of the group within this cohort. For the RFH pNET cohort for CgA and ANG2, the cohort consisted of 4 MEN1 cases but there was also Zollinger Ellison syndrome and MEN1 (n=1), Zollinger Ellison syndrome (n=1), and unconfirmed MEN1 (n=1). There were a similar number of MEN1 cases in both the RFH and UOL cohort. However, the lack of functional information for the RFH pNETs was a limitation. Having complete staging, functionality, MEN1 status and grading is important when comparing both cohorts and to assess the impact of biomarkers. The incompleteness of clinical characteristics in the cohorts is a limitation. As markers and models taken forward are based on the types of pNETs that were part of the UOL cohort and hence if the RFH cohort is different to the UOL cohort in terms of the heterogenous composition this could be a factor that influences the performance of algorithms at validation that were designed on a specific training cohort.

The external validation results illustrate the challenges of biomarker discovery and validation in terms of cohort size, detailed clinical information, control matching, and marker and assay quality, as well as the need for better algorithms. In the next chapter, the levels of CgA, ANG2 an VGF-NGF were determined in the ADEPTs cohort to compare levels of these markers in pNET,

PDAC, CP and AP patients. This part of my thesis aimed to assess if these markers could distinguish pNETs from common pancreatic conditions.

Chapter 6 - Application of the biomarker panel to the ADEPTs pancreatic diseases cohort

6.1 Introduction

A patient who has a pNET may need to be distinguished from a patient who has PDAC or a benign pancreatic condition when they present to a clinician. Advanced PDAC patients symptoms include abdominal pain (78-82%), anorexia (64%), early satiety (62%), jaundice (56-80%), sleep disorders (54%), weight loss (66-84%), diabetes (97%), back pain (48%), nausea and weight loss (50-86%)[527]. PDAC can also be identified asymptomatically, with asymptomatic PDAC characterised by smaller tumour size, early disease stage and higher resectability than those of symptomatic PDAC and associated with better prognosis compared to symptomatic PDAC[528]. pNET symptoms can vary depending on the type of functioning pNET that is present (VIPoma, gastrinoma, insulinoma, glucagonoma, somatostatinoma, ACTHoma, PTHRPomas and PPoma), with the associated symptoms dependent on the hormone that is secreted. For VIPomas associated symptoms include (watery diarrhoea, dehydration, hypokalaemia, abdominal pain, bloating and a flushed face), for gastrinomas (chest and or abdominal pain, acid reflux, heartburn, diarrhoea and tiredness), for insulinomas (dizziness, light headedness, sweating, hunger, confusion and irritability), for glucagonomas (distinct type of skin rash on the face stomach bottom and feet, diabetes, diarrhoea, weight loss, change in mood, anaemia, blood clots and sore mouth and tongue), for somatostainomas (gallstones, steatorrhea, anaemia, abdominal pain, high blood sugar levels and jaundice), for ACTHomas (weight gain, easy bruising, anaemia, depression, increased infection risk and darkened skin), for PTHRPomas (high calcium levels, abdominal pain, constipation, vomiting, bone pain, osteoporosis and fatigue) and for PPomas (diarrhoea, abdominal pain and weight loss)[529].

NF-pNETs tend to be diagnosed on imaging due to a lack of symptoms.

As pNETs are a rarer type of cancer, there is still no large scale study characterising symptomatology of pNETs nor are there studies that comprehensively compare the symptomatology of PDAC and PNEN[66]. The Liao et al., study tried to address this and found that of the symptoms identified twenty-three were significantly associated with PDAC, and nine symptoms with pNEN. They identified two alarm symptoms for both pNETs and PDAC which were jaundice and gastrointestinal bleeding. They also found that the risk of unintentional weight loss may be longer than two years before the diagnosis of pNEN[66].

The problem that clinicians face is that such patients can present with similar symptomology or imaging findings, which hampers making a diagnosis. For example, conditions mimicking pNETs in CT imaging include pancreatic metastases, acinar cell carcinoma, pancreatoblastoma, solitary fibrous tumour, pancreatic haematoma, serous adenoma, intrapancreatic spenules, exophytic gastrointestinal stromal tumours and peripancreatic paragangliomas[530]. Similarly, pathologies that can mimic PDAC in CT imaging include inflammatory conditions such as the various forms of pancreatitis (chronic-focal, mass-forming, autoimmune and groove pancreatitis), pNETs, solid pseudopapillary tumours (SPNs), metastasis (solid non-lymphomatous and hematologic), congenital variants (annular pancreas) peripancreatic lesions (accessory spleen, adrenal masses, duodenal masses, lymph nodes and vascular lesions), as well as certain rare pancreatic tumours (acinar cell tumours, solid serous tumours, hamartoma and solitary fibrous tumours)[531]. The number of differential potential diagnoses highlights some of the challenges that a clinician faces when trying to accurately diagnose a patient.

Diagnosing SPNs is difficult due to the non-specific clinical presentation and highly variable radiological and pathological features[532]. The predominantly solid lesions can mimic PDAC, however, younger patient age, identification of intratumoral haemorrhage or a capsule assist the differential diagnosis of the two types of tumour[531]. Another condition known as pancreatic abscesses are a rare condition seen in patients with pancreatic inflammation and patients with pancreatitis, with the condition being complicated by pseudocyst formation, in which pancreatic tissue necrosis, liquefaction and bacterial infiltration result in abscess formation[533]. In terms of symptoms, patients with pancreatic abscesses may have abdominal pain, fever, chills, nausea and vomiting or may present

with an inability to eat. Alternate symptomology for this condition is extremely unusual[533]. These symptoms overlap with those that can present for other pancreatic conditions including pNETs and PDAC. Case reports have been published in which a patient had pancreatic abscesses (infected pancreatic necrosis) which would have been narrowly misdiagnosed as a pancreatic malignancy[534,533,535]. Inflammatory conditions that commonly mimic PDAC at imaging include mass forming CP, focal autoimmune pancreatitis and paraduodenal pancreatitis[536]. Moreover, obstructive CP with the characteristics of diffuse ductal dilatation and diffuse pancreatic parenchymal atrophy can mimic ampullary masses or an intraductal papillary mucinous neoplasm (IPMN)[536]. Some patients with chronic inflammatory masses undergo the Whipple procedure for symptom management, whilst others undergo pancreatic resection because of a high clinical suspicion of malignancy and concern that biopsies have yielded false-negative results. Moreover, in 5-35% of Whipple procedures, the final pathologic diagnosis is a non-neoplastic inflammatory disease[537,538]. In patients suspected of having pancreatic cancer at an operable stage, surgical resections may be preferred in the absence of any other invasive pathologic diagnosis[539]. However, in clinical practice there is a risk of misdiagnosis and subsequent unnecessary resection of benign pancreatic disease[539]. 5-21% of pancreatectomies for presumed cancer were actually for benign pancreatic disease[540,541]. These problems with differential diagnosis highlight the clinical need for biomarker tests that could be utilised in the context of accurately discriminating pancreatic cancer from benign conditions. This would lead to quicker diagnoses and reduce unnecessary interventions.

AP, CP and pancreatic cancer are distinctive pancreatic diseases that have different prognoses and treatment options[542]. However, despite these three conditions being separate disease entities, inflammation and cancer can co-exist. Inflammation can be a cause of cancer, and cancer can be a cause of inflammation, making it challenging to discriminate one from the other[542]. In imaging, pancreatic inflammation may mimic pancreatic cancer by appearing as a focal mass on imaging[542]. Differentiation of AP, CP and pancreatic cancer may be helped by specific features like duct-penetrating signs and the duct to parenchyma ratio[542].

AP has been defined as an acute condition typically presenting with abdominal pain and usually associated with raised pancreatic enzymes in blood or urine as a result of an inflammatory disease of the pancreas[538].

In contrast, CP has been defined as a continuing inflammatory disease of the pancreas that is characterised by irreversible morphological changes, which typically cause pain and loss of exocrine and endocrine pancreatic fucntion[543,538]. Common causes of CP include alcohol, gall stones, pancreatic duct strictures, cystic fibrosis, chronic renal failure, hypercalcaemia, hyperlipidaemia, autoimmunity, smoking, pancreatic trauma, and hereditary/genetic pancreatitis, as well as idiopathic CP[538].

AP may also be a first symptom of pancreatic cancer, particularly in patients aged between 56 and 75 who have had a concomitant diagnosis of new onset diabetes mellitus or patients with a history of CP[544]. In such patients, additional imaging is warranted, with CT and EUS being prefered[542]. CP can also lead to pancreatic cancer through oncogenic mutation; this is mostly being seen in hereditary CP patients and in patients with risk factors for pancreatic cancer, such as nicotine and alcohol abuse[542]. Additionally, patients with PRSS1-mediated CP and patients with a history of autosomal dominant hereditary CP without any known genetic mutations may be considered for pancreatic cancer surveillance[542].

In addition to the risk of CP progressing to PDAC, there is also a risk of PDAC being misdiagnosed as CP because discrimination between these diseases can be difficult. Indeed, a retrospective study found that 5% of PDAC patients had been misdiagnosed with CP[545]. Moreover, patients with CP often have elevated levels of the commonly used PDAC biomarker, serum CA19-9, which may not be indicative of malignancy but a result of pancreatic inflammation[546]. This example illustrates the need for better biomarker tests to discriminate PDAC from CP.

Thus, a test that could distinguish patients who have a pNET from PDAC, AP and CP is a current unmet clinical need. The biomarkers CgA, ANG2 and VGF-NGF were identified as the most suitable biomarker combination in Chapter 3, with these findings being confirmed in internal validation as seen in Chapter 4. Thus, these markers were then further tested using external validation as described in Chapter 5. From the external validation results, VGF-NGF was shown to be unsuitable as a marker to discriminate pNET cases from controls in

the external validation cohort, whereas CgA and ANG2 were suitable markers to take forward to assess in the algorithms developed in Chapter 3. Section 6.2 describes the sample size calculation used to decide the optimal number of ADEPTs samples, Section 6.3 outlines the clinical characteristics of the ADEPTs pancreatic diseases cohort. The ADEPTs (Accelerated Diagnosis of neuroendocrine and pancreatic tumours) cohort (previously TRANSBIL) consisted of blood samples from patients attending endoscopy procedures and two week wait gastroenterology clinics. Thus, pNET, PDAC, AP and CP samples identified from samples within this cohort were used for my study. In Section 6.4 CgA and ANG2 levels in the ADEPTS cohort were first assessed individually for their ability to discriminate pNETs from PDAC, AP and CP. VGF-NGF levels were also assessed in this cohort to test my findings from Chapter 5 regarding the levels of VGF in pNETs. In Section 6.5 the best algorithms developed for ANG2 and CgA individually and in combination based on the UOL cohort (Chapter 3) were tested for their ability to distinguish pNETs from other pancreatic diseases in the ADEPTS cohort. Models developed using the UOL training cohort and the RFH external validation cohort were then similarly assessed. Finally, models were developed using just the ADEPTs pancreatic disease cohort were assessed.

6.2 Sample Size calculation results for the ADEPTs pancreatic disease cohort analysis

Sample size calculations were carried out to estimate the optimum number of AP, CP, PDAC and pNET samples needed for the application of the biomarker panel.   This was based on varying the sensitivity between 60 and 90% and the maximum margin of error between 0.1 and 0.25. A detailed methodology for this is in Section 2.11.2 and was based on the methodology described by Hajian et al.,[452]. Table 6.1 and Figure 6.1 show the resulting number sensitivity, which refers to the number of samples needed to achieve each specific sensitivity and maximum margin of error value. The number sensitivity was then divided by 4 to provide the number of samples for each of the disease groups (AP, CP, pNET and PDAC). The results showed that the number per group decreased as the sensitivity or maximum margin of error value increased. As the sensitivity for this test was not known a range of sensitivities were employed. Based on these

results, a suitable number of samples were chosen for each group. Retrospectively, the combinations that worked for the available pNET sample number was using a maximum margin error of 0.25 and a sensitivity of 0.9 or 0.8 or a maximum margin error of 0.20 and a sensitivity of 0.9 because of the low number of available pNET samples (n=11) (Table 6.1). Thus, whilst carrying out these calculations was important in order to assess how varying sensitivity and maximum margin of error impacts the number of samples needed per group in practice only a certain number of samples were available thus highlighting a limitation of such a calculation due to practical considerations.

Sensitivity	Maximum margin of error	Number sensitivity	Number per group sensitivity
0.6	0.10	369	92.25
0.6	0.15	164	41
0.6	0.20	92	23
0.6	0.25	59	14.75
0.7	0.10	323	80.75
0.7	0.15	143	35.75
0.7	0.20	80	20
0.7	0.25	52	13
0.8	0.1	246	61.5
0.8	0.15	109	27.25
0.8	0.2	62	15.5
0.8	0.25	39	9.75
0.9	0.1	138	34.5
0.9	0.15	62	15.5
0.9	0.2	35	8.75
0.9	0.25	22	5.5

Table 6.1: Summary of power calculations
The effect of altering the d value and sensitivity in order to calculate the number of samples needed per group.
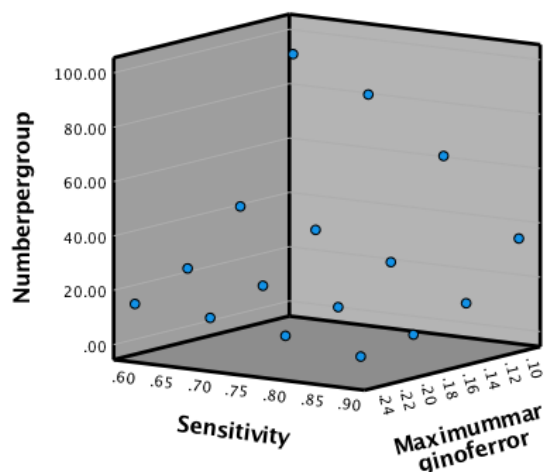
Figure 6.1: Graphical illustration of the relationship between sensitivity, maximum margin of error and number per group.

6.3 Clinical characteristics of the ADEPTs pancreatic disease cohort

For the ADEPTs cohort of samples used for this work, the inclusion criteria were patients who were diagnosed with pNET, PDAC, CP and AP. An exclusion criterion that was originally proposed was pregnancy, patients diagnosed with another cancer, patients on PPIs, patients with any mental health conditions (diagnosed depression, schizophrenia and bipolar disorder), however in practice an exclusion criteria including depression and PPI usage was not possible with the available samples. A range of symptoms was seen in the patients including abdominal pain, anaemia, bloating, jaundice, night sweats, rectal bleeding, reflux, and vomiting. More patients had abdominal pain with AP (8/16) compared to CP (5/16), although some AP (4/16), CP (5/16) and pNET (2/11) patients were also asymptomatic (Table 6.2). The most common symptom seen in pNETs was weight loss/unintentional weight loss (5/11), for

Symptom	AP (n=16)	CP (n=16)	pNET (n=11)	PDAC (n=16)
Abdominal pain	8	5	4	6
Abnormal LFTs	3	0	0	2
Anaemia	0	1	0	0
Asymptomatic	4	5	2	0
Bloating	1	1	1	0
Breast pain	0	0	1	0
Change in bowel habit	2	4	1	2
Constipation	0	0	0	1
Dyspepsia	0	0	1	0
Dysphagia	1	2	0	0
Epigastric pain	0	1	0	0
Gastritis	0	0	1	0
Heartburn	0	0	2	0
Jaundice/obstructive	1	0	0	8
Loss of appetite	0	0	2	0
Nausea	0	0	0	1
Not known	0	0	1	0
Night sweats	0	0	1	0
Rectal bleeding	0	0	0	1
Reflux	1	1	0	0
Vomiting	0	0	1	3
Weight loss/unintentional weight loss	2	4	5	8

Table 6.2: Symptoms reported for the ADEPTs pancreatic disease cohort.

CP and AP it was abdominal pain (5/16 and 8/16, respectively), and for PDAC it was weight loss/unintentional weight loss (8/16) and jaundice (8/16) (Table 6.2). The variation in non-specific symptoms combined with the lack of pNET-specific symptoms highlights a challenge in diagnosing pNET patients (Sections 1.2.4 and 6.6).

The pNET samples also varied in their respective tumour size, grade and location, and stage information was incomplete (Table 6.3). For most AP patients the cause of the AP was not known (8/16), and there were lower numbers of other aetiologies (Table 6.4). Functionality status was not available for the pNETs in the study, which was a limitation.

pNET sample	Grade	Size (cm)	Location	Ki67	Stage
1	NK	4	TOP	NK	NK
2	G1-2	1	NOP	NK	NK
3	G1	14	NOP	0.02	pT3 N1 (2/16) Mx R0
4	G2	3	HOP	0.1	pT3 N1 (1/20) R0
5	G3	5	TOP	0.23	NK
6	G3	3	HOP	0.3	NK
7	G1	8	NOP	<3	NK
8	G1	1	NOP	NK	NK
9	G1-2	2	TOP	<2	NK
10	G1	NA	HOP	<1	PT1 N0 Mx
11	G2	6	BOP/TOP	NK	PT3 N0 Mx

Table 6.3: pNET subgroup (n=11) clinical characteristics.

Gall stones	Autoimmune	PRRS-1	Idiopathic	Necrotising	Not known
4	1	1	1	1	8

Table 6.4: AP subgroup (n=16) clinical characteristics.

The CP cohort (n=16) mostly had unknown aetiologies (n=8), with other causes including alcohol related, IgG4, PRRS1 and gall stones (Table 6.5).

Necrotising Gall stones	Alcohol	IgG4	PRRS-1	Not known
2	2	1	1	10

Table 6.5: CP subgroup (n=16) clinical characteristics.

For the PDAC patients (n=16) most staging information was unavailable (Table 6.6).

PDAC sample	Tumour characteristics	Staging
1	Unknown	Unknown
2	HOP	Unknown

3	Unknown	Unknown
4	Unknown	Inoperable vascular involvement
5	Unknown	Unknown
6	HOP	Candidate for chemotherapy
7	Unknown	Locally Advanced PDAC arising from IPMN, metastatic
8	BOP 40mm	Unknown
9	BOP	Unknown
10	HOP	Dilated pancreatic duct
11	NOP	Technically resectable
12	Unknown	Unknown
13	HOP (5-6cm), NOP	SMV involvement
14	HOP (3.3 cm)	Locally Advanced
15	Unknown	Unknown
16	HOP	Unknown

Table 6.6: PDAC subgroup (n=16) clinical characteristics.
The tumour characteristics include HOP (Head of Pancreas), BOP (Body of pancreas), NOP (Neck of Pancreas), unknown. Staging information included whether there was vascular involvement, a candidate for chemotherapy, locally advanced, unknown, superior mesenteric vein (SMV), and resectable.

6.4 Analysis of biomarkers within the ADEPTs pancreatic disease cohort

Levels of the VGF, ANG2 and CgA biomarkers were determined for the pancreatic diseases cohort samples that had been selected from the ADEPTs cohort. For VGF a main question concerned whether my findings in Chapter 3 regarding non-elevation of VGF levels in pNET samples was supported in this independent cohort. Sections 6.4.1 to 6.4.3 describe the results of the VGF-NGF, ANG2 and CgA assays, respectively.

6.4.1 Analysis of VGF-NGF levels in the ADEPTs pancreatic disease cohort

VGF levels were analysed in the selected ADEPTs AP, CP, pNET and PDAC serum samples using the ELISA method described in Section 2.12.4. The hypothesis that the distribution of VGF is similar across the four disease groups was assessed using an Independent-Samples Kruskal Wallis Test, which indicated a statistically significantly difference (p=0.01) between these groups. The mean ± standard deviation values for AP (n=11), CP (n=12), PDAC (n=8) and pNET (n=6) samples were 98.53 ± 54.06, 1129.22 ± 3192.74, 374.45 ± 291.71 and 91.46 ± 42.18 pg/ml, respectively (Figure 6.2).
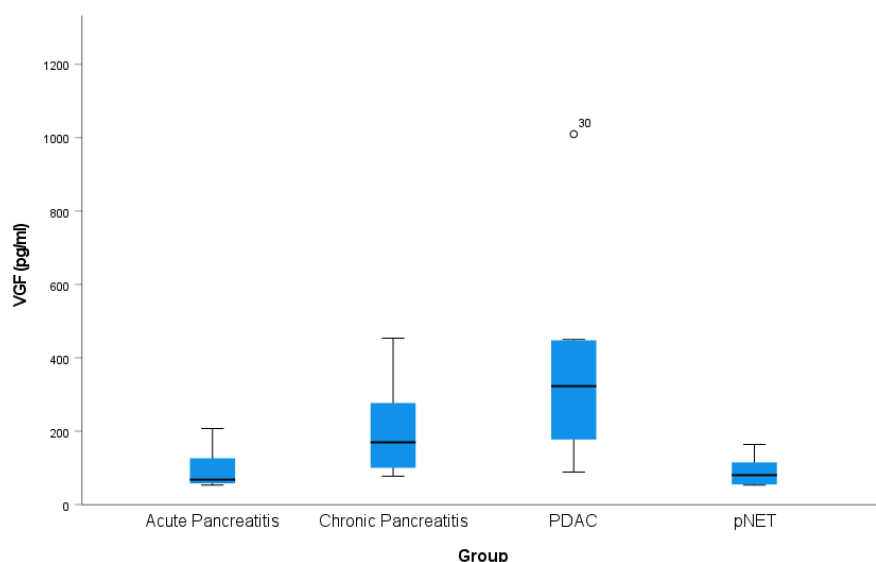
Figure 6.2: VGF levels in the pancreatic disease cohort.
The box plot shows the minimum and maximum (box whiskers) and the median (horizontal line) values for AP,CP, PDAC and pNET. One CP value which was particularly high was excluded when plotting the graph. A PDAC outlier is illustrated in the graph.

Statistically significant differences were identified between the pNET and CP, pNET and PDAC, AP and CP, and AP and PDAC groups (Table 6.7). pNET values for VGF were not elevated which was consistent to what was seen in Chapter 5, however not compared to healthy control levels here in Chapter 6. When assessing the pairwise comparisons for the different conditions for VGF, AP and PDAC were revealed to be significantly different from each other (Table 6.7).

Comparison	Significance	Adjusted Significance
pNET vs AP	0.741	1.000
pNET vs CP	0.019	0.113
pNET vs PDAC	0.002	0.014
AP vs CP	0.016	0.096
AP vs PDAC	0.002	0.009
CP vs PDAC	0.307	1.000

Table 6.7: Statistical analysis of VGF levels in the pancreatic disease cohort.
Adjusted significance values using the Bonferroni correction are shown.

6.4.2 Analysis of ANG2 levels in the ADEPTs pancreatic disease cohort

ANG2 levels in sera were analysed in the ADEPTs pancreatic disease cohort using the ELISA approach described in Section 2.12.2. The mean ± standard deviation levels for AP (n=16), CP (n=16), pNET (n=11), and PDAC (n=16) were 2990.99 ± 984.72, 3130.94 ± 1462.55, 2602.92 ± 760.27 and 3980.25 ± 2879.04 pg/ml, respectively (Figure 6.3). An Independent-Samples Kruskal-

Wallis Test revealed that the null hypothesis was retained (p=0.742). Moreover, this analysis implies that the usefulness of ANG2 as a pNET marker is poor based on these pNET samples (n=11) due to it having the lowest ANG2 median in the context of comparison with the other conditions (AP, CP and PDAC).
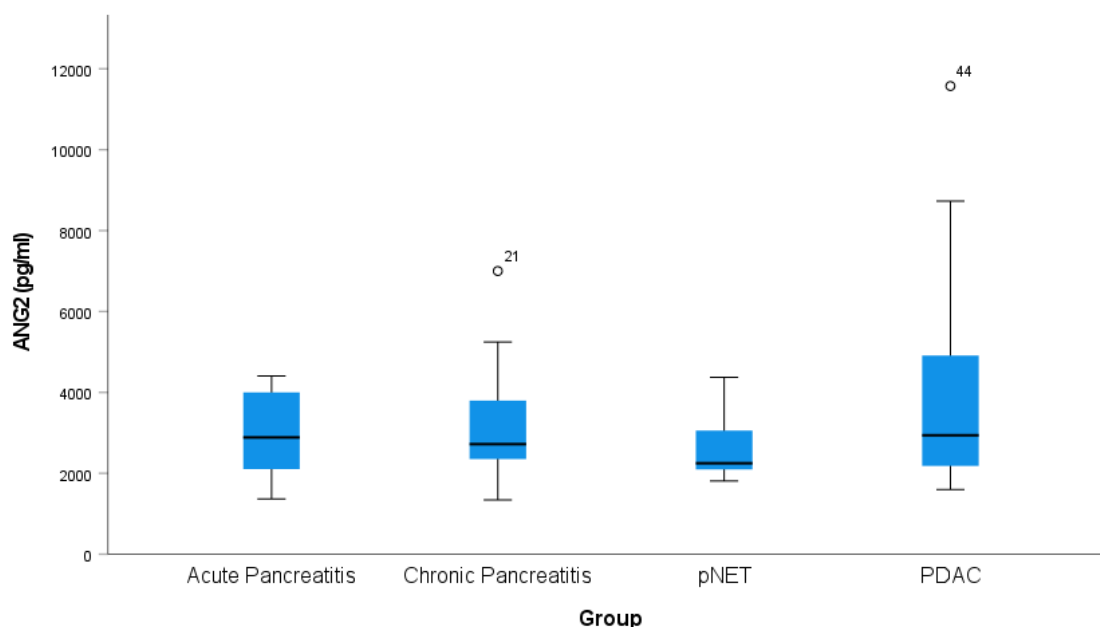
Figure 6.3: ANG2 levels in the pancreatic disease cohort.
The box plot shows the minimum and maximum (box whiskers) and the median (horizontal line) values. Outlier values are shown for Chronic Pancreatitis and PDAC.

6.4.2.1 Analysis of confounding factors for ANG2

Potential confounding factors for ANG2 within the ADEPTs pancreatic disease cohort were explored. These included gender, diabetes, symptomatic status (asymptomatic (n=11), symptomatic (n=47); one patient was excluded for symptomatic status analysis because symptomatic status was not available for this patient, diabetes medication and PPI usage. Gender (female (n=24), male (n=35)), diabetes (no (n=40), type II (n=19)), and symptomatic status were not confounding factors as assessed by Independent-Samples Mann Whitney U Test (p=0.829, 0.770 and 0.913, respectively). PPI status (Lansoprazole (n=7), not known (n=13), no (n=25), Omeprazole (n=14)) was found not to be a confounding factor as assessed by an Independent-Samples Kruskal-Wallis test (p=0.376). Thus, no confounding factors were found for ANG2 in the ADEPTs pancreatic disease cohort.

6.4.3 Analysis of CgA levels in the ADEPTs pancreatic disease cohort

Serum CgA levels were also assessed in the ADEPTs pancreatic disease cohort using ELISAs (Section 2.12.3). The mean ± standard deviation values for AP (n=16), CP (n=16), pNET (n=11) and PDAC (n=16) sample CgA levels were 295.47 ± 405.94, 229.51 ± 311.90, 39.75 ± 30.89 and 165.10 ± 223.20 ng/ml, respectively (Figure 6.4). The pNET levels of CgA were lower compared to AP,
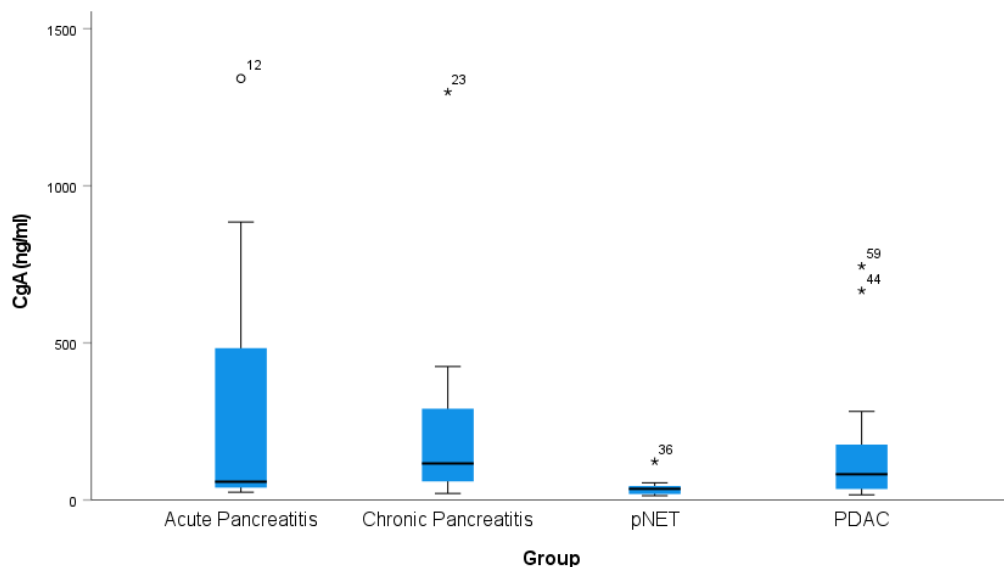
Figure 6.4: Comparison of CgA levels in the pancreatic disease cohort.
The box plot shows the minimum and maximum (box whiskers) and the median (horizontal line) values.

CP, PDAC. This was unexpected due to CgA being an established pNET marker. However the number of pNETs in this study are just (n=11) and with a lack of functional information available for the pNETs in the ADEPTs cohort it may be possible that the ADEPTs pNET cohort contain a large number of insulinomas for which CgA is known to be a poor marker[90]. CgA levels are also known to be lower in early stage pNETs and lower grade pNETs. However, the 11 pNETs contain four G1 tumours, two with G1-2 tumours, two with G2 tumours, two with G3 tumours and one with a tumour of unknown grade. Thus, the grade of tumours in the group was diverse and did not consist of lower grade tumours. However, none of the tumours in the pNET group were known to be metastatic which could also be a factor to explain these results. However, the CgA levels in the pNET group were particularly low, even lower than that seen in the UKCTOCs healthy controls which were assessed in Chapter 5 for CgA. A limitation also in the ADEPTs cohort analysis is that healthy controls were not collected as part of the ADEPTs study and thus a group of healthy

controls collected and stored in the same way could not be used in my analysis and as an additional fifth group. An Independent Samples Kruskal Wallis Test led to rejection of the null hypothesis (p=0.09), and a post hoc Dunn's test revealed statistically significantly differences between the pNET and PDAC, pNET and AP, and pNET and CP groups. Significance and adjusted significance levels using Bonferroni corrections are shown for each pair-wise comparison (Table 6.8). The results from the pair-wise comparison, further illustrated that CgA was much lower in the pNET group compared to these other groups and suggested that based on these 11 pNET samples that CgA was not a useful pNET marker

Comparison	Significance	Adjusted Significance
pNET vs PDAC	0.021	0.128
pNET vs AP	0.006	0.034
pNET vs CP	0.001	0.007
PDAC vs AP	0.610	1.000
PDAC vs CP	0.291	1.000
AP vs CP	0.585	1.000

Table 6.8: Statistical analysis of VGF levels in the pancreatic disease cohort.
Adjusted significance values using the Bonferroni correction are shown.

6.4.3.1 Confounding factor analysis for CgA

Gender, PPI status, diabetes status, and symptomatic status were assessed as confounding factors for CgA in the ADEPTs pancreatic disease cohort. Gender (female (n=24), male (n=35)), diabetes status (no (n=40), type II diabetes (n=19)), and symptomatic status (asymptomatic (n=11), symptomatic (n=47); one sample was excluded as the symptomatic status was not known) were assessed using an Independent-Samples Mann Whitney U Test. Gender was found to be a confounding factor for CgA levels (p=0.039), but diabetes status and symptomatic status were not significantly associated with CgA levels (p=0.256 and 0.766, respectively). PPI usage (Lansoprazole (n=7), Omeprazole (n=14), not known (n=13), and no (n=25)) was assessed using an Independent-Samples Kruskal-Wallis Test and was found to be a confounding factor for CgA levels (p=0.011). The pairwise comparison of PPI and CgA revealed that the no PPI and Omeprazole group were statistically significantly different from each other with the adjusted significance of p=0.014 (Table 6.9).

Comparison	Significance	Adjusted significance

No vs NK	0.328	1.000
No vs Lansoprazole	0.031	0.185
No vs Omeprazole	0.002	0.014
NK vs Lansoprazole	0.209	1.000
NK vs Omeprazole	0.077	0.465
Lansoprazole vs Omeprazole	0.843	1.000

Table 6.9: Pairwise comparison of PPI and CgA

Adjusted significance values using the Bonferroni correction are shown.

6.5 Application of ML algorithms to the ADEPTs pancreatic disease cohort

Two sets of ML algorithms (RF, C5.0 decision tree, and LR as well as SVM for the two-marker combination described in Chapters 3 and 5) that I had developed for the single markers CgA and ANG2, and for the combination of CgA and ANG2, were assessed against the ADEPTs pancreatic disease cohort. The ML application described in Section 6.5 was carried out to assess whether the models developed using a case and healthy control cohort as described in previous chapters could be used to identify pNETs. The ELISA results obtained for CgA (Figure 6.4) and ANG2 (Figure 6.3) for the ADEPTs cohort for the pNETs behaved in a way that was unexpected due to the low levels of these markers in the pNETs (n=11). Thus, based on this finding, it was likely that the models would not perform particularly well on the ADEPTs cohort for the identification of pNETs, however, applying the ML models, allowed for insights such as FPR, TPR, FNR and TNR and assisted in the understanding of where specifically models had drawbacks based on the ADEPTs cohort. This would also help refine model development in the future.

Both sets of algorithms had been created by comparing pNET to healthy control data and it was therefore important to determine whether these algorithms could also differentiate pNET from CP, AP and PDAC patients, which present with similar symptoms (Section 6.1), or whether newer algorithms or biomarkers were needed to distinguish these conditions. These analyses were carried out using SPSS Modeler with previously described methodology (Section 2.1) with the CP, AP and PDAC groups defined as controls.

Separately, LR, RF, C5.0 decision tree models (SPSS Modeler) and Optimised Linear and Optimised RBF SVM (MATLAB Classification App) were created using the ADEPTs pancreatic disease CgA and ANG2 data in order to assess

whether such models could differentiate pNETs from the other conditions. This was carried out using methodology described in Section 2.1.

6.5.1 Application of the UOL algorithms to the ADEPTs pancreatic disease cohort

All the assessed models (RF, C5.0 decision tree, LR and SVM) for the ANG2 and CgA combination in the UOL dataset performed poorly in the ADEPTs pancreatic diseases cohort (Tables 6.9 – 6.12). This was as expected based on the ELISA results obtained from the pancreatic diseases for ANG2 and CgA (Figure 6.3 and Figure 6.4). More specifically, the RF algorithms produced AUCs ranging from 0.349 to 0.516 (Table 6.10), suggesting that these algorithms were unsuitable for discriminating pNETs from PDAC, CP and AP. This also suggests that the thresholds used in the ML models developed based on case and control data were not suitable for use in an instance where AP, CP and PDAC were used as a control group, as it is likely that these patients have elevated levels of these markers compared to healthy controls, thus the threshold was too low. When assessing each individual run, the TPR rates (sensitivity, defined as the ability to identify a pNET sample correctly) were high across the 10 runs (72.70% to 91.00%) however the FPR rates were also high across all the runs (85.40% to 93.80%). This therefore illustrates that the other conditions were incorrectly being identified as positives and thus suggesting that the RF model was not suitable.

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.477	72.70	14.60	85.40	27.20
2	0.394	90.90	14.60	85.40	9.00
3	0.498	72.70	12.50	87.50	27.30
4	0.349	91.00	10.40	89.60	9.00
5	0.418	81.80	6.30	93.80	18.20
6	0.365	81.80	10.40	89.60	18.20
7	0.473	81.80	14.60	85.40	18.20
8	0.516	90.90	8.30	91.70	9.00
9	0.441	81.80	8.33	91.70	18.20
10	0.465	72.70	14.60	85.40	27.30

Table 6.10: Assessment of the UOL ANG2 and CgA combination RF models using the pancreatic disease cohort.
The models created for the 10 runs of the CA combination based on the UOL training dataset (n=106) were externally validated using the pancreatic disease cohort (n=59).

Although the C5.0 decision tree models performed slightly better with an AUC of 0.573 (Table 6.11), this performance was still poor. Similar to the RF models, the TPR (100%) and FPR (100%) were both maximal, meaning that the test classified all the ADEPTs samples as pNETs, and again indicating that the threshold values were too low for this cohort (Table 6.11). However, a problem is that the ADEPTs pNET marker data is different to the training data, and thus based on the ADEPTs pNET results the model performance was poor and unsuitable, however if the ADEPTs results are not representative widely of pNETs then the models itself may have some utility in other pNET cohorts.

AC C5.0 Decision Tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Non-boosted General	0.573	100.00	0.00	100.00	0.00
Non-boosted Accuracy	0.573	100.00	0.00	100.00	0.00

Table 6.11: Assessment of the ANG2 and CgA UOL C5.0 non boosted decision tree models using the pancreatic diseases cohort.

The LR and SVM models performed the worst with AUC values of 0.203 to 0.216 (Tables 6.12 and 6.13), highlighting the unsuitability of these models for differential diagnosis.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.203	90.90	2.00	97.90	9.00

Table 6.12: Assessment of the UOL ANG2 and CgA combination LR model using the pancreatic disease cohort.

SVM	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Linear SVM C=10	0.216	18.20	43.80	56.30	81.80
RBF SVM C=10 gamma=0.7	0.203	18.20	39.60	60.40	81.80

Table 6.13: Assessment of the UOL ANG2 and CgA combination Linear SVM and RBF SVM models using the pancreatic disease cohort.

All the assessed models (RF, C5.0 decision tree and LR) for the single ANG2 marker performed poorly in the ADEPTs pancreatic disease cohort (Tables 6.14 - 6.16).

For the RF models the AUC values across the runs ranged from 0.466 to 0.583, the TPR ranged from 81.80% to 90.90%, and the FPR ranged from 68.75% to 77.10%. The high FPR again indicated these models were unsuitable for differential diagnosis (Table 6.13). Hence, illustrating the underlying problem with the ADEPTs data being different from that of the UOL data. In the

pancreatic diseases there was no significant difference between ANG2 levels of any of the disease types. This then would imply that the ML models created whereby there was a difference in case and control ANG2 levels (UOL training) are likely not to perform well for this single marker in the ADEPTs cohort.

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.513	90.90	31.30	68.75	9.00
2	0.519	90.90	29.20	70.80	9.00
3	0.466	90.90	29.20	70.80	9.00
4	0.583	90.90	25.00	75.00	9.00
5	0.568	90.90	25.00	75.00	9.00
6	0.509	81.80	23.00	77.00	18.00
7	0.528	90.90	27.10	72.90	9.00
8	0.547	90.90	22.90	77.10	9.00
9	0.538	90.90	25.00	75.00	9.00
10	0.487	90.90	29.20	70.80	9.00

Table 6.14: Assessment of the UOL ANG2 RF models using the pancreatic disease cohort.

The ANG2 LR and C5.0 non-boosted decision tree models also performed poorly with AUC values of 0.393 (Table 6.15) and 0.573 (Table 6.16), respectively. Thus, none of these models was able to discriminate pNET from PDAC, CP and AP samples.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.393	100.00	8.30	91.70	0.00

Table 6.15: Assessment of the UOL ANG2 LR model using the pancreatic disease cohort.

C5.0 Decision Tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Non-boosted General	0.573	100.00	14.60	85.40	0.00
Non-boosted Accuracy	0.573	100.00	14.60	85.40	0.00

Table 6.16: Assessment of the UOL ANG2 non-boosted C5.0 decision tree models using the pancreatic diseases cohort

The RF, C5.0 decision tree and LR models for CgA also performed poorly in discriminating pNETs from PDAC, CP and AP samples (Table 6.17 - 6.19). The RF models had AUC values ranging from 0.380 to 0.486, with TPRs from 54.50% to 81.80% across the runs (Table 6.17).

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.439	81.80	31.30	68.80	18.20
2	0.414	54.50	29.20	70.80	45.50
3	0.434	72.70	29.20	70.80	27.30

4	0.421	72.70	29.20	70.80	27.30
5	0.38	81.80	29.20	70.80	18.20
6	0.486	81.80	27.10	72.90	18.20
7	0.366	63.60	27.10	72.90	36.40
8	0.449	63.60	33.30	66.70	36.40
9	0.414	72.70	31.30	68.80	27.30
10	0.488	72.70	31.30	68.80	27.30

Table 6.17: Assessment of the UOL CgA RF models using the pancreatic diseases cohort.

The CgA non-boosted C5.0 decision tree and LR models also performed poorly with AUC values of 0.286 (Table 6.18) and 0.184 (Table 6.19), respectively, again illustrating the unsuitability of the case-control models for the ADEPTs pancreatic disease cohort.

C5.0 Decision Tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Non-boosted General	0.286	36.40	22.90	77.10	63.60
Non-boosted Accuracy	0.286	36.40	22.90	77.10	63.60

Table 6.18: Assessment of the UOL CgA non-boosted C5.0 decision tree models using the pancreatic disease cohort.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.184	9.00	43.80	56.30	90.90

Table 6.19: Assessment of the UOL CgA LR model using the pancreatic disease cohort.

6.5.2 Application of the combined UOL and RFH cohort algorithms to the ADEPTs pancreatic disease cohort

Models were also created in Chapter 5, for the larger combined UOL and RFH cohort for ANG2 and CgA. Though performance of these combined training models compared to the UOL training cohort was generally poorer, assessment of these models on the pancreatic disease ADEPTs cohort was carried out as an additional assessment of the ADEPTs cohorts performance using these models.

The combined UOL and RFH cohort (n=217) LR, C5.0 non-boosted decision tree and RF algorithms for ANG2, CGA, and the ANG2 and CGA combination were applied to the ADEPTs pancreatic disease cohort to assess model performance.

The algorithms performed poorly for the ANG2 and CGA combination as reflected by the resulting AUC values of 0.199 to 0.301 for the RF models

(Table 6.20), 0.189 for the LR model (Table 6.21) and 0.275 for the C5.0 decision tree models (Table 6.22).

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.248	27.30	25.00	75.00	72.70
2	0.228	18.20	22.90	77.10	81.80
3	0.224	27.30	25.00	75.00	72.70
4	0.236	27.30	25.00	75.00	72.70
5	0.239	36.40	27.10	72.90	63.60
6	0.205	45.50	29.20	70.80	54.50
7	0.301	27.30	39.60	60.40	72.70
8	0.199	27.30	27.10	72.90	72.70
9	0.337	45.50	27.10	72.90	54.50
10	0.283	27.30	31.30	68.80	72.70

Table 6.20: Assessment of the combined UOL and RFH external cohort ANG2 and CgA combination RF model using the pancreatic disease cohort.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.189	36.40	20.80	79.20	63.60

Table 6.21: Assessment of the combined UOL and RFH external cohort ANG2 and CgA combination LR model using the pancreatic disease cohort.

C5.0 decision tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
General non-boosted	0.275	100.00	0.00	100.00	0.00
Accuracy non-boosted	0.275	100.00	0.0	100.00	0.00

Table 6.22: Assessment of the combined UOL and RFH external cohort ANG2 and CgA combination C5.0 non-boosted decision tree model using the pancreatic disease cohort.

For ANG2 alone the AUC values were 0.579 to 0.620 for the RF model (Table 6.23), 0.393 for the LR model (Table 6.24) and 0.573 for the C5.0 decision tree models (Table 6.25).

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.607	81.80	39.60	60.40	18.20
2	0.586	72.70	35.40	64.60	27.30
3	0.597	81.80	43.80	56.30	18.20
4	0.600	72.70	41.70	58.30	27.30
5	0.618	81.80	39.60	60.40	18.20
6	0.579	72.70	47.90	52.10	27.30
7	0.620	72.70	43.80	56.30	27.30
8	0.615	72.70	43.80	56.30	27.30
9	0.581	81.80	39.60	60.40	18.20
10	0.588	81.80	45.80	54.20	18.20

Table 6.23: Assessment of the combined UOL and RFH external cohort for the ANG2 RF model using the pancreatic disease cohort.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.393	72.70	18.80	81.30	27.30

Table 6.24: Assessment of the combined UOL and RFH external cohort ANG2 LR model using the pancreatic disease cohort.

C5.0 Decision Tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
General non-boosted	0.573	100.00	14.60	85.40	0.00
Accuracy non-boosted	0.573	100.00	14.60	85.40	0.00

Table 6.25: Assessment of the combined UOL and RFH external cohort ANG2 C5.0 non-boosted decision tree model using the pancreatic disease cohort.

For CgA alone the AUCs were 0.310 to 0.391 for the RF model (Table 6.26), 0.184 for the LR model (Table 6.27), and 0.275 for the non-boosted C5.0 decision tree models (Table 6.28).

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.313	45.50	33.33	66.70	54.50
2	0.310	36.40	35.40	64.60	63.60
3	0.343	45.50	33.30	66.70	54.50
4	0.344	63.60	37.50	62.50	36.40
5	0.333	36.40	33.30	66.70	63.60
6	0.391	45.50	37.50	62.50	54.50
7	0.310	45.50	35.40	64.60	54.50
8	0.324	45.50	31.30	68.80	54.50
9	0.333	36.30	27.10	72.90	63.60
10	0.315	27.30	33.30	66.70	72.70

Table 6.26: Assessment of the combined UOL and RFH external cohort CgA RF model using the pancreatic disease cohort.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.184	9.00	45.80	54.20	90.90

Table 6.27: Assessment of the combined UOL and RFH external cohort CgA LR model using the pancreatic disease cohort.

C5.0 Decision Tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
General non-boosted	0.275	9.10	45.80	54.20	90.90
Accuracy non-boosted	0.275	9.10	45.80	54.20	90.90

Table 6.28: Assessment of the combined UOL and RFH external cohort CgA C5.0 non-boosted decision tree model using the pancreatic disease cohort.

From these results it was clear that models trained on the pancreatic disease cohort itself were needed to classify the four disease types, and that models created using healthy control and pNET data were not generalisable to this cohort. The FPR levels across all the models were consistently high, meaning

that PDAC, CP and AP cases were incorrectly identified as pNET cases. Thus, based on the findings from Section 6.5.1 and 6.5.2 newer algorithms are required.

6.5.3 Creation of algorithms from the ADEPTs pancreatic disease cohort data

When case-control cohort models were tested on the ADEPTs pancreatic disease cohort, the results showed high FPR values, but the models also tended in general to have high sensitivity. C5.0 non-boosted decision trees, RF and LR algorithms were created for the ADEPTs pancreatic disease cohort (n=59) itself using the ANG2, CGA, and ANG2 and CGA combination. This approach was undertaken to assess whether suitable models could be built to better differentiate pNET from PDAC, AP and CP patients and to assess the FPR, TPR, FNR and TPR. There are some limitations to this approach, including that the numbers of the PDAC (n=16), AP (n=16), CP (n=16) and pNET (n=11) patients are particularly small, and the results generated from this analysis would not be particularly meaningful without subsequent validation. The issues for certain algorithms such as RF being susceptible to overfitting remain. However, creation of models using the ADEPTs cohort would help to further analyse the data further and provide insights into thresholds, FNR, TPR, TNR and FPR which provide an overview of how the ELISA results can be used in practice. However, based on the ELISA data that was obtained for the ADEPTs pancreatic diseases cohort, the CgA (Figure 6.4) and ANG2 (Figure 6.3) levels were lower for the pNETs in this cohort compared to AP, CP and PDAC. Thus, it is likely that the ML models that would be created for the ADEPTs pancreatic disease cohort with the aim of identifying pNETs would adopt an approach of lower marker levels as opposed to higher marker levels of CgA and ANG2 when compared to these other conditions. However, this would be different to that seen in the UOL cohort, as higher marker levels of CgA and ANG2 were indicative of pNETs as opposed to lower levels of these markers. The results revealed that the RF models had the best performance with AUC values ranging from 0.994 to 1.000 for the ANG2 and CGA combination. These models also had TPR rates ranging from 63.60% to 100% and a TNR of 100% (Table 6.29). However, the issue of overfitting remains, and without validation the impacts of this would not be known.

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.999	90.90	100.00	0.00	9.10
2	1.000	72.70	100.00	0.00	27.30
3	0.999	81.80	100.00	0.00	18.20
4	0.999	90.90	100.00	0.00	9.00
5	1.000	100.00	100.00	0.00	0.00
6	1.000	63.60	100.00	0.00	36.40
7	1.000	100.00	100.00	0.00	0.00
8	0.997	81.80	100.00	0.00	18.20
9	0.994	81.80	100.00	0.00	18.20
10	1.000	81.80	100.00	0.00	18.20

Table 6.29: Performance of the RF models created using the ANG2 and CgA markers in the pancreatic disease cohort.

The ANG2 and CgA combination LR model had an AUC of 0.856, however this model suffered from a poor TPR of 27.30% despite a TNR of 100% (Table 6.30). Thus, despite having a strong AUC, these results suggest that this model was not sensitive enough for pNET detection and thus not suitable.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.856	27.30	95.80	4.2	72.70

Table 6.30: Performance of the LR model created using the ANG2 and CgA markers in the pancreatic disease cohort.

The ANG2 and CgA combination C5.0 non-boosted decision tree models had an AUC of 0.841, a TPR of 27.30% and TNR of 100% (Table 6.31), again suggesting the test was highly specific but not sensitive enough for pNET detection. Interestingly the C5.0 decision tree, utilised just CgA and not ANG2, suggesting that ANG2 as a marker was not important. Two thresholds for CgA were used in the decision tree, the first was a threshold of CgA <1.558nmol/L, and a second of CgA <0.438 nmol/L for the identification of the pNETs (Figure 6.5).
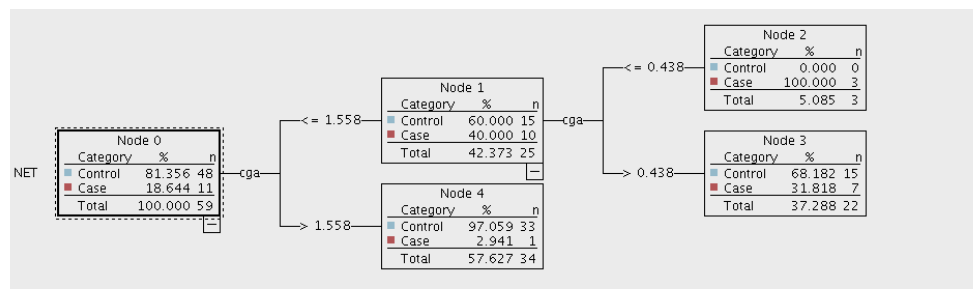
Figure 6.5: General C5.0 non boosted tree for ADEPTs pancreatic diseases cohort which only utilised CgA

C5.0 decision tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
General non-boosted	0.841	27.30	100.00	0.00	72.70
Accuracy non-boosted	0.841	27.30	100.00	0.00	72.70

Table 6.31: Performance of C5.0 non-boosted decision tree models created using the ANG2 and CgA markers in the pancreatic disease cohort.

The Optimised SVM results for the Linear and RBF SVM for the two-marker combination of ANG2 and CgA revealed results of an AUC of 0.84 for the Optimised Linear SVM and 1.00 for the Optimised RBF SVM. The TPR and TNR were 100% for the Optimised RBF, however for the Optimised Linear SVM, the TPR was 0% with a FNR of 100%. Thus, suggesting that the Optimised Linear SVM model is unsuitable due to a FNR of 100% meaning that the pNETs were incorrectly classified as control (Table 6.32).

SVM model	Kernel Scale	Box constraint	Number of grid divisions	Maximum training time (s)	Accuracy (%)	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
Optimised Linear	n/a	0.46416	10	300	81.4	0.84	0.00	100	0.00	100
Optimised RBF	0.0046416	10	10	300	100.00	1.00	100.00	100.00	0.00	0.00

Table 6.32: Performance of SVM models created using the ANG2 and CgA markers in the pancreatic disease cohort.

For the single marker ANG2, the RF models had AUC values ranging from 0.978 to 0.999, TPR values of 63.60% to 81.80%, and TNR values of 97.90% to 100.00% (Table 6.33).

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.995	81.80	97.90	2.00	18.20
2	0.993	72.70	100.00	0.00	27.30
3	0.991	81.80	97.90	2.00	18.20
4	0.978	81.80	97.90	2.00	18.20
5	0.995	90.90	100.00	0.00	9.00
6	0.999	81.80	100.00	0.00	18.20
7	0.986	72.70	100.00	0.00	27.30
8	0.986	81.80	100.00	0.00	18.20
9	0.999	81.80	100.00	0.00	18.20
10	0.999	63.60	100.00	0.00	36.40

Table 6.33: Performance of the RF models created using the ANG2 single marker in the pancreatic disease cohort.

The ANG2 LR and C5.0 non-boosted decision tree (general and accuracy) models performed poorly with AUC values of 0.607 (Table 6.34) and 0.500 (Table 6.35), respectively. Their specificities of 95.80% and 100%, respectively

(Table 6.34 and 6.35), highlighted their strong ability to identify the control (AP, CP and PDAC) as a single group accurately. However, both models were unsuitable as classifier algorithms as their TPRs were both 0%, showing a complete inability to identify pNET cases. Moreover, for the ANG2 C5.0 decision tree, that model itself outlining the thresholds was not provided by the software.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.607	0.00	100.00	0.00	100.00

Table 6.34: Performance of the LR model created using the ANG2 single marker in the pancreatic disease cohort.

C5.0 decision tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
General non-boosted	0.500	0.00	100.00	0.00	100.00
Accuracy non-boosted	0.500	0.00	100.00	0.00	100.00

Table 6.35: Performance of the C5.0 non-boosted decision tree model created using the ANG2 single marker in the pancreatic disease cohort.

For the CgA RF model AUCs ranged from 0.995 to 1.000 with TPR ranging from 81.80% and 100.00% and TNR ranging from 97.90% and 100.00% (Table 6.36).

Run	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
1	0.995	90.90	97.90	2.10	9.10
2	0.997	81.80	100.00	0.00	18.20
3	0.996	81.80	97.90	2.10	18.20
4	1.000	100.00	100.00	0.00	0.00
5	0.995	90.90	97.90	2.10	9.10
6	1.000	100.00	100.00	0.00	0.00
7	0.997	100.00	97.90	2.10	0.00
8	1.000	100.00	100.00	0.00	0.00
9	0.996	81.80	100.00	0.00	18.20
10	1.000	81.80	100.00	0.00	18.20

Table 6.36: Performance of RF models created using the CgA single marker in the pancreatic disease cohort.

The LR model had an AUC of 0.816 a TPR of 0% (Table 6.37), while the non-boosted C5.0 decision tree had an AUC of 0.84 and a TPR of 27.30% (Table 6.38). The thresholds for the CgA C5.0 decision tree were the same as seen when the two markers ANG2 and CgA were entered to create a decision tree (only CgA used), the thresholds used for pNET detection in the decision tree were firstly CgA <1.558 nmol/L, and secondly CgA <0.438 nmol/L for the

identification of the pNETs (Figure 6.6). Due to the poor performance in identifying pNETs these models were both unsuitable for pNET detection.

AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
0.816	0.00	100.00	0.00	100.00

Table 6.37: Performance of the LR model created using the CgA single marker in the pancreatic disease cohort.

C5.0 Decision Tree	AUC	TPR (%)	TNR (%)	FPR (%)	FNR (%)
General non-boosted	0.841	27.30	100.00	0.00	72.70
Accuracy non-boosted	0.841	27.30	100.00	0.00	72.70

Table 6.38: Performance of the C5.0 non-boosted decision tree model created using the CgA single marker in the pancreatic disease cohort.
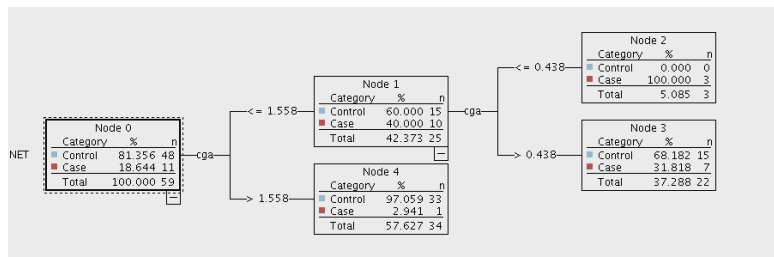
Figure 6.6: General C5.0 non boosted tree for ADEPTs pancreatic diseases cohort for CgA.

6.5.4 Summary of cohorts

The UOL cohort, RFH external validation cohort and ADEPTs training cohort had showed varying distributions of CgA in pNET patients (Figure 6.7). CgA
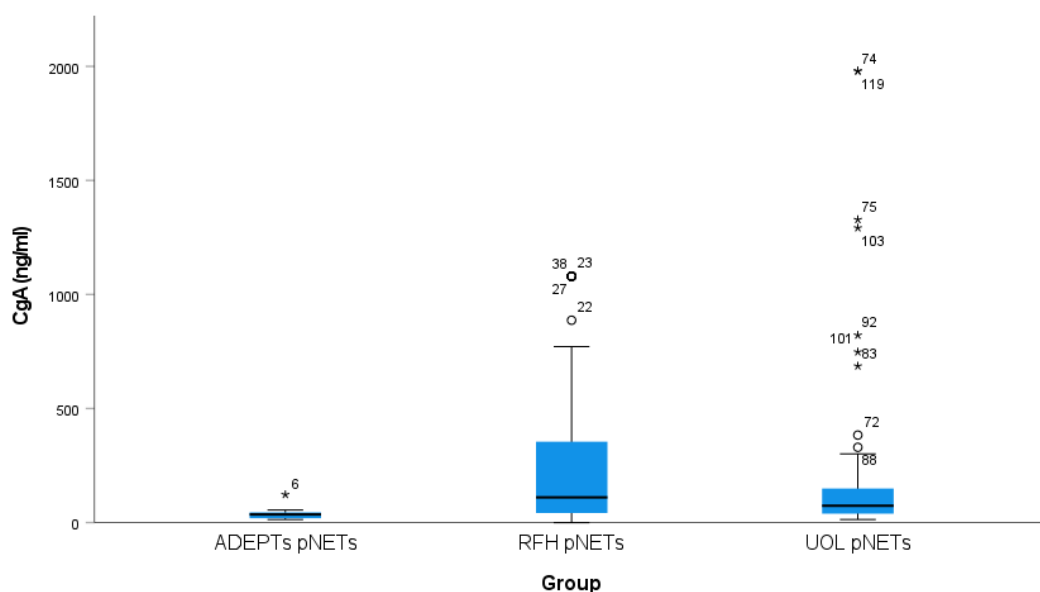
Figure 6.7: Comparison of the CgA distributions of the different pNET cohorts

RFHpNETs (n=60) UOLpNETs (n=57) and ADEPTs pNETs (n=11)

distributions across the different the control cohorts were also varied (Figure 6.8).
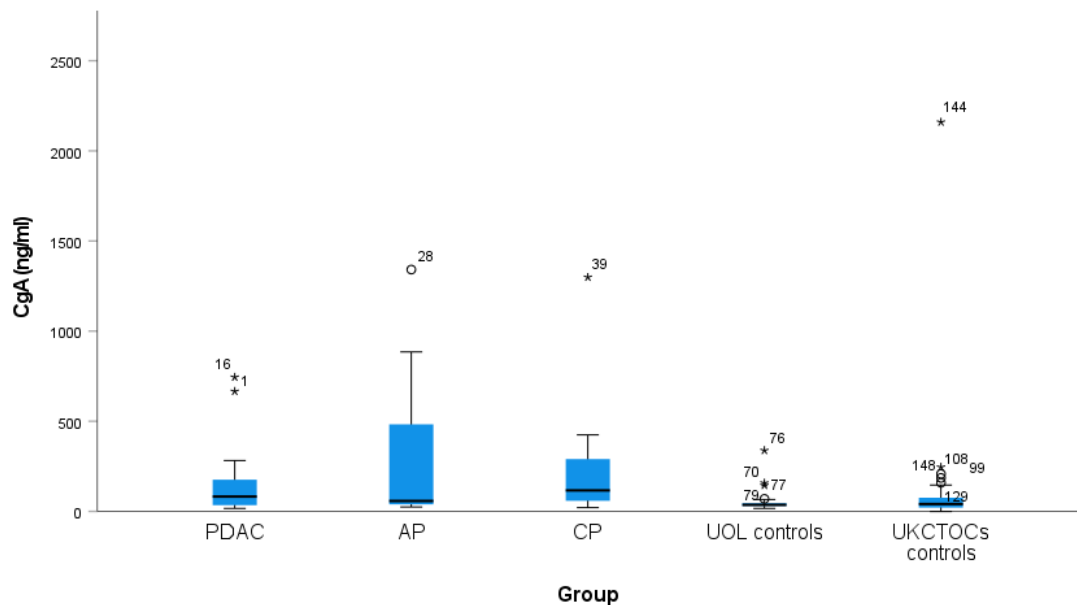
Figure 6.8: Comparison of the CgA distributions of the different control cohorts.

The mean pNET ANG2 levels were highest in the RFH external validation, followed by the ADEPTs pNET cohort and lowest in the UOL pNET cohort (Table 6.39). The pNET CgA distributions of the different. The mean pNET CgA level was found to be highest in the RFH external cohort, followed by the UOL training cohort and the ADEPTs cohort had the lowest mean CgA levels (Table 6.40).

	UOL training dataset	RFH external validation cohort	ADEPTs training cohort
pNET mean Ang2 ± SD	2260.60 $\pm$ 1074.27 pg/ml (n=57)	3999.74 $\pm$ 3354.01 pg/ml (n=60)	2602.92 ± 760.27 pg/ml (n=11)

Table 6.39: Summary of UOL (n=57), RFH external validation (n=60) and ADEPTs (n=11) pNET ANG2 results for pNETs.

	UOL training dataset	RFH external validation cohort	ADEPTs training cohort
pNET mean CgA ± SD	237.48 ± 434.48 ng/ml (n=57)	277.77 ± 362.558 ng/ml (n=60)	39.75 ± 30.89 ng/ml (n=11)

Table 6.40: Summary of UOL (n=57), RFH external validation (n=60) and ADEPTs pNETs (n=11) CgA results for pNETs.

Groups	Significance	Adjusted Significance
ADEPTs pNETs vs UOL pNETs	0.006	0.019
ADEPTs pNETs vs RFH pNETs	0.001	0.004
UOL pNETs vs RFH pNETs	0.369	1.000

Table 6.41: Pairwise comparison between the ADEPTs pNETs (n=11), UOL pNETs (n=57) and RFHpNETs (n=60) CgA distribution.

The cohorts were revealed to be significantly different for CgA in the pNETs of the ADEPTs and UOL cohorts, which has an adjusted significance of p = 0.019, and the ADEPTs pNETs and RFH pNET cohorts, which had an adjusted significance of p=0.004 (Table 6.41). Whereas the distributions between the CgA pNET levels of the UOL and RFH cohorts were not statistically significantly different from each other (Table 6.41).

The ANG2 distributions were varied across the pNET patient cohorts (Figure 6.9). The ANG2 distributions also varied across the control groups (Figure 6.10). The lowest mean control ANG2 level was seen in UOL cohort (Figure 6.42).
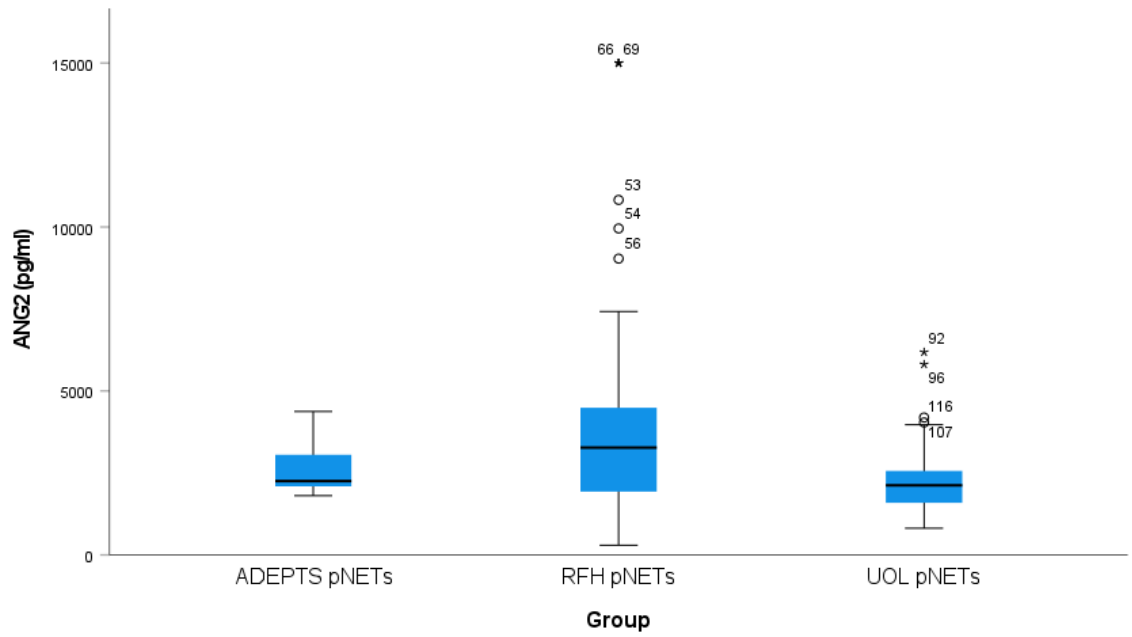
Figure 6.9: Comparison of the ANG2 distributions of the different pNET cohorts
RFHpNETs (n=60) UOLpNETs (n=57) and ADEPTs pNETs (n=11)
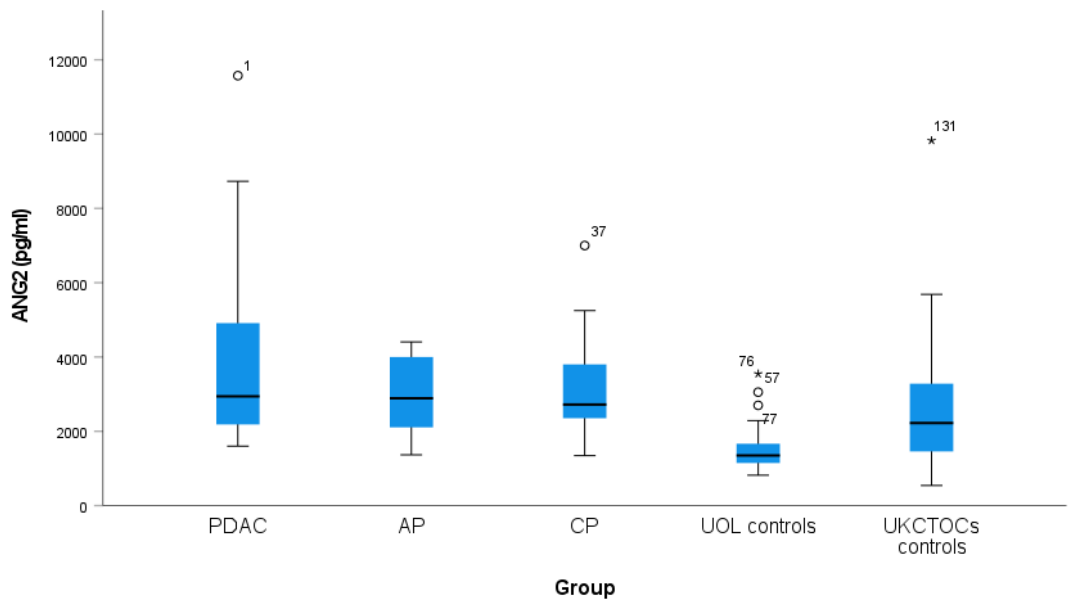
Figure 6.10: Comparison of the ANG2 distributions of the different control cohorts.

284

Controls	Mean $ANG2$ $(pg/ml) \pm SD$	Mean CgA $(ng/ml) \pm SD$
UOL (n=49)	1494.42 ± 546.94	48.68 ± 49.42
UKCTOCs (n=51)	2610.65 ± 1545.31	101.00 ± 299.13
ADEPTs CP (n=16)	3130.94 ± 1462.55	229.51 ± 311.90
ADEPTs AP (n=16)	2990.99 ± 984.72	295.47 ± 405.94
ADEPTs PDAC (n=16)	3980.25 ± 2879.04	165.10 ± 223.20

Table 6.42: Summary of control group mean ANG2 and CgA. UOL (n=49), UKCTOCs (n=51), ADEPTs CP (n=16), ADEPTs AP (n=16), ADEPTs PDAC (n=16).

The RFH pNETs were statistically significantly different from each other adjusted significance p=0.001, whereas the ANG2 pNET distributions between the ADEPTs and UOL, and the UOL and RFH cohorts were not statistically significant from each other (Table 6.43).

Groups	Significance	Adjusted Significance
ADEPTs pNETs vs UOL pNETs	0.231	0.694
ADEPTs pNETs vs RFH pNETs	<0.001	0.001
UOL pNETs vs RFH pNETs	0.406	1.000

Table 6.43: Pairwise comparison between the ADEPTs pNETs, UOL pNETs and RFHpNETs ANG2 distribution

When comparing the different control groups with the pairwise comparison for ANG2, it was seen that the UOL controls and the UKCTOCs controls were statistically significantly different from each other (p=0.000) (Table 6.44), similarly the UOL controls and the ADEPTs CP, AP and PDAC groups were statistically significantly different from each other (p=0.000) for ANG2 (Table 6.44).

Groups	Significance	Adjusted Significance
UOL controls vs UKTCTOCs controls	<0.001	0.000
UOL controls vs ADEPTs CP	<0.001	0.000
UOL controls vs ADEPTs AP	<0.001	0.000
UOL controls vs ADEPTs PDAC	<0.001	0.000
UKCTOCs controls vs ADEPTs CP	0.119	1.000
UKCTOCs controls vs ADEPTs AP	0.087	0.868
UKCTOCs controls vs ADEPTs PDAC	0.065	0.650
ADEPTs CP vs ADEPTs AP	0.902	1.000
ADEPTs CP vs ADEPTs PDAC	0.817	1.000
ADEPTs AP vs ADEPTs PDAC	0.915	1.000

Table 6.44: Pairwise comparison between the different control groups (UOL controls, UKCTOCs controls, ADEPTs CP, ADEPTs AP and ADEPTs PDACs) for ANG2 distribution.

When comparing the different control group distributions for CgA using the pairwise comparison, the UOL controls and the ADEPTs CP group were statistically significantly different from each other (p=0.002) (Table 6.45), and the UKCTOCs controls and the CP group were statistically significantly different from each other p=0.015 (Table 6.45).

Groups	Significance	Adjusted Significance
UOL controls vs UKCTOCs controls	0.436	1.000
UOL controls vs ADEPTs PDAC	0.016	0.162
UOL controls vs ADEPTs AP	0.003	0.033
UOL controls vs ADEPTs CP	<0.001	0.002
UKCTOCs controls vs ADEPTs PDAC	0.061	0.613
UKCTOCs controls vs ADEPTs AP	0.016	0.160
UKCTOCs controls vs ADEPTs CP	0.002	0.015
ADEPTs PDAC vs ADEPTs AP	0.664	1.000
ADEPTs PDAC vs ADEPTs CP	0.294	1.000
ADEPTs AP vs ADEPTs CP	0.539	1.000

Table 6.45: Pairwise comparison between the different control groups (UOL controls, UKCTOCs controls, ADEPTs CP, ADEPTs AP and ADEPTs PDACs) for CgA

6.6 Discussion

This chapter sought to assess whether the biomarkers ANG2 and CgA could be used to distinguish pNETs from other pancreatic conditions that share similar presentations and symptoms. ELISA data for CgA (Figure 6.4) revealed that there were significant differences in CgA level between these conditions. The mean CgA level of pNET samples was significantly lower than that of PDAC (p=0.021), CP (p=0.001) and AP (p=0.006) samples, and the mean CgA levels of PDAC, AP and CP samples were not significantly different from each other (Table 6.8). Moreover, there was a high level of variation within the disease groups as reflected by the standard deviations obtained (Section 6.4.3). As CgA is an established pNET marker, it was surprising that this marker had the lowest

mean levels of 39.75 ± 30.89 ng/ml (mean ± standard deviation) in the ADEPTs pNET samples (n=11).

In the absence of an accepted reference level for CgA, a comparison of the mean CgA level of ADEPTs pNET samples with the mean levels in my other pNET cohorts revealed that it was lower than that obtained for the RFH (277.77 ± 362.558 ng/ml, n=60) and the UOL pNET cohorts (237.48 ± 434.48 ng/ml, n=57) and when comparing whether the pNET cohort distributions of CgA were significantly different from one another, CgA distributions in the pNETs of the ADEPTs and UOL cohorts were statistically significantly different, adjusted significance of p=0.019 and in the ADEPTs pNETs and RFH pNET cohorts, adjusted significance p=0.004 (Table 6.41), whereas this was not the case for the UOL and RFH pNETs. In fact, the mean CgA level of the ADEPTs pNET cohort fell between the levels of the UOL (48.68 ± 49.42) ng/ml (n=49) and UKCTOCs (100.996 ± 299.132 ng/ml, (n=51)) healthy control cohorts, with the UOL CgA levels lower than that seen in the UKCTOCs healthy controls and in other studies. The Spadaro et al., study found healthy control levels of 48 ±18 ng/ml[182], the Qiao et al., study found healthy controls had a range of CgA of 39-94 ng/ml[90]. When comparing the distributions of CgA in the UOL cases vs the controls, the control CgA levels were statistically significantly different than the controls, and the mean levels of CgA in the UOL cases vs controls was higher (237.5 vs 48.7 ng/ml). Likewise, when comparing the distributions of CgA in the RFH pNET cases vs the UKCTOCs controls, these were statistically significantly p<0.001, and the mean levels of CgA in the RFHpNET cases (n=60) was higher 277.765 ± 362.558 ng/ml than the UKCTOCs controls (n=51) 100.996 ± 299.132 ng/ml.

In terms of the ELISA kits used in this work for assaying CgA, two different types of kits were used; the CgA CisBio and CgA DAKO ELISA kit. The UOL training cohort was the DAKO CgA ELISA kit, the combination training dataset (UOL and RFH) consisted of CgA assayed using DAKO and CisBio ELISA kit however for the ADEPTs cohort the CisBio ELISA kit was used. The use of two different ELISA kits is a limitation in this work due to the lack of CgA standardisation. Moreover, there is no FDA approved CgA test. The same CgA ELISA kit should have been used for both the UOL training cohort and the ADEPTs cohort however this was not possible. In terms of the kits themselves, the CgA Chromoa CisBio ELISA assay targets the core of the molecule (145-

245) to measure intact and fragmented CgA (CisBio). The DAKO CgA ELISA kit is specific for the 23kDa C-terminal fragment, so it will measure whole molecules and fragments[547].

The ADEPTs cohort for AP CgA levels were 295.47 ± 405.94 ng/ml (n=16), for CP 229.51 ± 311.90 ng/ml (n=16), PDAC 165.10 ± 223.20 ng/ml (n=16) and pNET 39.75 ± 30.89 ng/ml (n=11). Amongst the ADEPTs disease groups only the pNET group had a mean CgA level lower than that the mean levels seen in healthy control groups, while the other disease groups all had mean CgA levels higher than the mean levels of healthy controls. The higher CgA levels in the other pancreatic conditions indicated the assay itself was not compromised. The CgA level in the ADEPTs pNET group (n=11) was therefore determined to be low compared to the other pNET groups (RFH and UOL) and similar to healthy control levels, indicating that the ADEPTs pNET samples were somehow unusual. Moreover, the number of ADEPTs samples was low and therefore prone to sampling error and the SD of the disease groups was very high and there is a higher chance of sampling bias and the samples not being most representative of the population.

The 11 ADEPTs pNET samples were from four patients with G1 tumours, two with G1-2 tumours, two with G2 tumours, two with G3 tumours and one with a tumour of unknown grade. The tumour size was also diverse and ranged from 1-14cm with one not known. Ki67 data were only available for seven of the tumours and ranged widely from 0.02 to <3. Staging information was only available for four of the pNET samples and revealed that none of the four tumours was metastatic (Table 6.3).

Hijoka et al. found that factors associated with elevated CgA levels in pNET patients were tumour classification, tumour size and presence of liver metastases based on univariate analyses, and PPI use and liver metastases based on a multivariate (LR) analysis[517]. The functionality of the pNET samples could have been one factor that contributed to the unexpectedly low CgA levels, as CgA has been reported to have a lack of utility for the diagnosis of insulinomas compared to other types of pNET[90]. However, functionality data for the ADEPTs pNETs were unavailable. An absence or low number of metastatic disease samples was another factor that potentially contributed to the unexpectedly low CgA levels as the UOL and RFH cohorts contained metastatic and non-metastatic pNET patients. It is also possible that the ADEPTs cohort

contained an unusual number of small tumours compared to other cohorts and that these patients were earlier in the clinical pathway compared to the UOL and RFH patients, however this information is not available for comparison. Unfortunately, the incomplete clinical data and the small number of pNETs within the ADEPTs cohort precluded further assessment of these factors.

The CgA levels of patients with pancreatic diseases have previously been examined by Hijioka et al.[517], who found that the mean levels of pNET, CP, PDAC, and autoimmune pancreatitis (AIP) patients as well as controls were 407.8 ± 984.6 (n=69), 93.6 ± 57.5 (n=50), 91.8 ± 101.8 (n=50), 69.9 ± 52.4 (n=20), and 62.5 ± 48.3 (n=112) ng/ml, respectively. The highest levels were therefore found in pNET samples, followed by modestly elevated levels in AIP, CP and PDAC samples. These results are consistent with my own results, with the exception of the ADEPTs pNET samples.

The Hijoka et al. study also found that PPI usage increased CgA levels in PDAC, CP and AIP patients, which indicated that serum CgA level could only differentiate between pNET, other pancreatitis disease and control samples from patients who do not take PPIs. Interestingly when limiting the analysis to patients who were not on PPIs, the levels of CgA in patients with PDAC or CP were not significantly different from the controls. The issue of PPI usage as a confounding factor is of particular importance to my study, because PPI use was found to be a confounding factor for CgA level (p=0.011) in the ADEPTs pancreatic disease cohort. Specifically, the no vs Omeprazole group with an adjusted significance of p=0.014. PPI usage elevates CgA levels in pNET patients in general, however the confounding nature of the PPI usage could have elevated the other disease groups relative to the pNET group. This therefore highlights a limitation and ideally, patients not on PPIs should have been a part of the cohort, however this was not possible due to the availability of samples. This result, combined with the inability to stratify the pNET patients by PPI use due to the small number of patients, led to PPI usage by patients in the ADEPTs cohort compromising my analysis. To more accurately distinguish pNET from other pancreatic disease samples, PPIs should have been discontinued or replaced by Histamine Receptor antagonists for two weeks before blood sampling[510, 547].

ANG2 levels varied less than CgA levels between the different pNET cohorts (Table 6.39 and Table 6.40) with ANG2 levels seen in the UOL pNETs cohort being 2260.60 $\pm$ 1074.27 pg/ml (n=57) in the RFH cohort pNETs 3999.74 $\pm$ 3354.01 pg/ml (n=60) and in the pNETs of the ADEPTs pancreatic diseases cohort 2602.92 ± 760.27 pg/ml (n=11). For the ANG2 variation seen between the cohorts, the ELISA kits used is not a factor, due to the same ELISA kit being used for the training cohort, RFH pNET cohort and the ADEPTs cohort (Quantikine).

The UOL healthy control samples had ANG2 levels of 1494.42 ± 546.94 pg/ml (n=49), the UKCTOCs healthy controls had ANG2 levels of 2610.65 ± 1545.31 pg/ml (n=51). ANG2 levels in AP were 2.990 ± 0.984 ng/ml (n=16), in CP 3.130 ± 1.462 ng/ml (n=16), in pNET 2.602 ± 0.760 ng/ml (n=11), in PDAC 3.980 ± 2.879 ng/ml (n=16).

AP, CP and PDAC samples all had higher mean ANG2 levels than those observed in the UOL and UKCTOCs healthy control samples (Figure 6.10) and the distributions of the UOL controls compared to the ADEPTs CP, AP and PDAC samples were significantly different from each other with significance of p<0.001 and adjusted significance p=0.000 (Table 6.44). Hence, ANG2 may be a marker for PDAC, CP and AP, however further studies with larger cohorts would be required to explore this further. Interestingly, the pNET samples in the UOL cohort had the lowest mean ANG2 levels (Table 6.40). ANG2 is an angiogenic marker associated with vascularisation, that drives lymphatic metastasis of pancreatic cancer. In a study by Juusola et al., immunohistochemistry scores and circulating ANG2 levels of 196 PDAC patients were compared to clinical data and patient outcomes. They found that a higher than median level of circulating ANG2 (>2.72ng/ml) was associated with shorter disease-specific survival. Neither circulating nor tumour ANG2 expression was impacted by age or sex and there was no relationship with tumour size, stage, grade or lymph node metastasis, but a negative association between serum ANG2 level and tumour ANG2 expression was seen[548]. In another study, PDAC patients had an overall median survival of 28.4 months, but only 7.7 months in those with circulating ANG2 levels >75th percentile (P=0.0005). Moreover, in this study elevated circulating ANG2 levels correlated with the extent of lymphatic metastasis, and orthotopic PDAC xenografts with forced expression of ANG2, but not ANG1, displayed increased blood and

lymphatic vessel density and an enhanced rate of lymphatic metastasis. Interestingly in this study, patients with PDAC without lymph node involvement presented with ANG2 serum levels similar to controls, but different from patients with lymph node metastasis, suggesting that circulating ANG2 may serve as a non-invasive diagnostic indicator of tumour stage[549].

Serum ANG2 levels of CP patients have not been assessed in any published studies. However, serum ANG2 levels of AP patients has been assessed by Huang et al[330], who showed that AP patients had a significantly higher serum ANG2 concentration (6.51 ± 8.73 ng/ml) than controls (5.11 ± 1.74 ng/ml). ANG2 were also found to be more elevated among patients categorised as having severe AP (15.45 ± 13.60 ng/ml) compared to those having mild AP (4.88 ± 2.34 ng/ml)[330]. Moreover, ANG2 predicted poor prognosis and adverse events including mortality in AP patients.

The ANG2 models trained using either the UOL dataset (Tables 6.14 - 6.16) or the UOL RFH combined dataset (Tables 6.23 - 6.25) all suffered from high FPRs when applied to the ADEPTs dataset. This was due to the pNET data being lower than the UOL training dataset. This highlighted the unsuitability of these ML models and ANG2 as a discriminatory marker. In contrast, the ANG2 models created using just the ADEPTs dataset performed very differently to each other. Specifically, the LR and non-boosted C5.0 decision tree models suffered from a complete lack of discrimination between different disease types (Tables 6.34 and 6.35), while the RF model performed with extremely high sensitivity and specificity across runs (Table 6.33). Nonetheless, it is likely that the RF model was overfitted to the training data and would not generalise and thus, this model alone would not be reliable to take forward when considering that the C5.0 and LR models performed so poorly in comparison.

As seen with CgA and ANG2, the lowest levels of VGF (91.46 ± 42.18 pg/ml) were in the pNET samples of the ADEPTs cohort. The highest VGF levels were seen in CP and PDAC samples (1129.22 ± 3192.74 pg/ml and 374.45 ± 291.71 pg/ml, respectively). The low VGF levels in the ADEPTs pNET samples were consistent with the level in RFH pNET samples that were lower than controls but differed from the level in UOL pNET samples that were higher than all other cohorts in this study.

VGF and VGF-derived neuropeptides are biologically associated with pain, with rapid upregulation of VGF in sensory neurones after nerve and injury and inflammation and activation of microglial p38 by VGF peptides[550]. TLQP-21 which is a VGF-derived peptide's role in nociception was explored in mice and the formalin test revealed a significant increase of pain related licking response in mice following injection of TLQP-21 and an increase in licking response was also detected when TLQP-21 was injected alone without formalin[551]. Thus, the study illustrated that VGF-derived peptides may be involved in inflammatory pain in vivo. AP is defined as an acute condition that typically presents with symptoms of abdominal pain and raised pancreatic enzymes in the blood or urine due to an inflammatory disease of the pancreas[538], whereas CP has been defined as a continuing inflammatory disease of the pancreas characterised by irreversible morphological changes. These changes typically cause pain and loss of exocrine and endocrine pancreatic function. Pain is the predominant symptom in CP and 80-90% of CP patients present with pain as the primary symptom either at the first attack of AP or as the main reason for hospital readmission in the following months and years as a result of the disease progressing to what could be defined as CP[538]. VGF elevations in PDAC patients could be perhaps be explained by VGFs role in perineural invasion (PNI), which is a characteristic feature of PDAC associated with tumour recurrence, poor prognosis and the generation of pain[552]. However, the extent of perineural invasion in the PDAC patients assessed in the ADEPTs cohort is unknown. Despite AP also being associated with pain, VGF levels were not high in the AP ADEPTs patients and were less than that seen in CP ADEPTs patients. Hence although pain could be a factor in the increase in VGF levels seen for CP ADEPTs patients, it does not provide an explanation for the lower levels of VGF in the AP compared to the CP and PDAC ADEPTs patients. One explanation could be that the chronic nature of the pain seen in CP patients and the persistence of the pain over time could cause the VGF levels to be higher in these patients compared to AP patients, however this is to be determined. Moreover, pain is also a factor in pNET patients, with some pNET patients having abdominal pain as a symptom within this ADEPTs cohort (4/11), this was compared to 8/16 AP patients, 5/16 CP patients and 6/16 PDAC patients in the cohort. Thus, based on this observation, VGF levels would be expected to be the highest in the AP patients but this was not shown to be the case in these

patients. Hence, the elevations in VGF seen in the PDAC and CP patients may not be completely related to pain.

A limitation of the ADEPTs pNET cohort was the small pNET sample size (n=11). Despite the power calculation suggesting that a group size of ~11 could be used if the sensitivity of the test was 80% (maximum margin of error = 0.25), not knowing the actual sensitivity at which each test would perform in the ADEPTs pancreatic disease cohort setting precluded targeting an adequate sample number. The problem of sample number is difficult to address in practice due to the rarity of pNETs (prevalence/incidence) and the low rate of diagnosis. pNETs consist of less than 5% of all pancreatic tumours and 7% of all NETs[553,78]. They themselves are the second most common pancreatic neoplasm with an overall incidence of approximately 5 per 1,000,000 new cases/year[78,1]. Assuming a test sensitivity of 60% and a maximum margin error of 0.1 then the number of samples needed would be 92 per group. However, for the models that were developed using the ADEPTs cohort itself, the TPR varied greatly for the different models ranging from (0% to 100%), this highlights a particular challenge in adequately choosing a sensitivity by which most models would be covered.

A second limitation in this part of the study was applying algorithms designed to distinguish pNET from healthy control samples to distinguish pNET from AP, CP and PDAC samples. AP, CP and PDAC groups were found not to have CgA and ANG2 levels similar healthy controls; instead, CgA and ANG2 levels in the pancreatic diseases were elevated relative to healthy controls and covered a much wider range. This would therefore require a different strategy in the future, with one stream focussed on differential markers for pNETs vs other pancreatic conditions and another stream focussed on markers that could distinguish pNETs from healthy controls. Moreover, based on findings from this chapter neither ANG2 nor CgA perform particularly well in the context of a ML model to identify pNETs from these other conditions, however due to the unexpected findings from this work particularly due to the low levels of CgA and ANG2 in the pNETs (n=11) examined, a larger cohort of pNET samples would be required to assess this further.

 Thus, the results obtained for the ADEPTs cohort highlight the need for a much larger cohort of pNET samples, ideally stratified according to grade, stage, functionality and metastatic stage, for the creation of newer models that can

discriminate between pNET, AP, CP, PDAC and healthy samples. Work by Srirajaskanthan et al., and Melen Mucha et al illustrated that ANG2 levels were statistically significantly elevated in patients with metastasis compared to those without metastasis[140,470]. Multivariate analysis from the Srirajaskanthan et al., study also revealed that higher ANG2 levels were associated with higher grade tumours[470]. For CgA, pNETs with metastasis have been shown to have significantly higher levels compared to pNETs with no metastasis[90]. Moreover, insulinomas have shown to serum CgA levels that were lower than non-insulinoma pNETs[90]. In a study by Pulvineti et al., increased CgA was identified in 7/19 of G1 cases, 22/35 G2 cases and 6/6 G3 cases, illustrating the role that grade can have on CgA levels[462].

Moreover, PDAC, CP and AP sample stratification may also be needed as it is likely with a marker such as ANG2 that more severe disease will correlate with increased marker level. Different algorithms and different marker combinations may be needed to distinguish pNET from AP, CP, PDAC and healthy samples. In addition, ANG2 or CgA may each have more relevance as markers for certain types of pNET, potentially evidenced by their stronger performance as discriminatory markers for the UOL and RFH pNET cohorts and poor performance in the ADEPTs pNET cohort. The larger UOL and RFH pNET cohorts appeared to better represent of the heterogeneity of pNETs compared to the ADEPTs pNETs used in this study, but larger, stratified cohorts will be required to assess this issue.

Overfitting by the RF and optimised RBF SVM models, and the lack of sensitivity of the LR, decision tree and optimised linear SVM models indicate that models developed using the ANG2 and CgA combination are likely to require additional markers including previously assessed pNET markers, as well as newer markers to be identified through secretome studies, to differentiate pNETs from other pancreatic diseases. These additional markers may include markers previously assessed in Chapter 3 including MAC2BP. From the current evidence based on this chapter ANG2 and CgA are not suitable for discriminating pNET from pancreatic diseases, however there are important limitations to this cohort including low sample number, and unexpectedly low levels of these markers than was expected based on the UOL and RFH data and hence, removal of these markers at this stage would not be the best approach, but instead further studies with larger cohorts of AP, CP, PDAC and

pNET patients are needed to assess this issue before CgA and ANG2 are not kept. Studies have widely sought to assess markers that can differentiate pNETs from healthy controls[90,445], but the differential diagnosis of pNETs from other pancreatic conditions that requires further work and a one size fits all approach for a marker panel and associated and ML algorithm may not be the best approach as different thresholds are likely to be present for each of these conditions for CgA and ANG2, thus, treating these other pancreatic conditions in the most simple way as similar to healthy controls is likely not the best approach.Future studies should also include important aspects of the clinical data that were absent in the ADEPTs data, such as the functional and metastatic status of each pNET.

Chapter 7 - Wider Discussion

My work described in this thesis began with the development of ML classifier algorithms derived from a dataset comprising the levels of seven previously identified GEP-NET markers in pNET cases and healthy controls (UOL cohort). Three important conclusions were drawn from this training stage. First, strong performances were seen for RF algorithms derived from all combinations containing three to seven markers, with no clear preference for any specific marker combination, as discussed in Chapter 3. The extremely high AUCs of these RF algorithms were thought to be caused by overfitting and this inference was confirmed by validation of these algorithms in a train-and-test approach using the UOL cohort data, as discussed in Chapter 4. This result illustrated the value of algorithm validation, albeit using a less informative internal validation method. Second certain markers (TB4 and NSE) were not found to be suitable markers for the assessed algorithms (Chapter 3). Third, the three markers CgA, VGF-NGF and ANG2 performed well across all algorithms and this result provided the rationale to take these three markers forward for external validation as explored in Chapter 5.

Further confirmation that CgA, VGF and ANG2 were the most suitable of the seven markers that were initially employed for pNET detection was evident from the performance of different seven-marker algorithms in kFCV and train and test validation (Chapter 4). This was particularly clear for Optimised Linear and RBF SVM algorithms assessed using kFCV, which showed that the CVA combination performed strongly at 5FCV, 10FCV and 20FCV. In addition to SVM model performance, boosted C5.0 decision trees generated from seven-marker data showed that only CgA, VGF and ANG2 alongside MAC2BP were utilised across all 10FCV runs.

One particular finding from the work from this thesis was that VGF was not shown to be a suitable marker in the external validation and ADEPTs pNETs. The same VGF ELISA kit producer and product were used for all training and validation cohorts, providing methodological consistency and reducing experimental error. However, during external validation levels of VGF were not elevated in the independent RFH pNET cohort compared to the UKCTOCs healthy controls (Chapter 5), and VGF was therefore discounted as a suitable pNET marker.

Post protein translation VGF undergoes proteolytic processing in the regulated secretory pathway to produce active VGF-derived peptides. VGF peptides can be classified by their cleavage sites in VGF_{1-615}, the responsible protease, their mechanism of action and cellular function their physiological function and how they were identified., with at least 12 VGF derived peptides identified[554]. Thus, a deeper understanding of these VGF derived peptides is also needed, as some VGF-derived peptides might be of greater relevance than others for pNET detection.

The VGF results illustrated one of the difficulties with biomarker research, namely that markers which perform well in initial stages can fail at validation[555]. This result also emphasises the importance of eliminating markers at an early stage via external validation due to practical, time and cost implications. VGF-NGF is a neuroendocrine tissue marker[556], however its role as a serum marker has not previously been investigated, hence the reasons for VGF-NGF not performing as well within this cohort are unclear. Potential areas for further investigation include tumour heterogeneity, and clinical and demographic differences between the cohorts. The finding that VGF was not a suitable pNET marker was further supported by the results from the ADEPTs pNET cohort (Chapter 6), which showed no elevation over AP, CP and PDAC levels. Further analysis of VGF in other pNET cohorts are needed to confirm these findings. External validation using an independent cohort (Chapter 5) allowed the generalisability of algorithms trained on CgA, ANG2 and VGF levels in UOL samples (Chapter 3) to be assessed with reduced bias compared to internal validation methods. The results revealed that levels of ANG2 and CgA were higher in pNET cases compared to controls, however this was not seen for VGF. Despite the higher levels of CgA and ANG2, the performance of LR, SVM, C5.0 decision tree and RF algorithms based on the AC combination dropped compared to that seen at training. While LR sensitivity increased from 72 to 88%, AUC fell from 0.849 to 0.724, and specificity fell from 80% to 31%. Decreased AUC, sensitivity and specificity were seen for the other assessed models (C5.0 decision tree, RF, Linear SVM and RBF SVM) with no model having a specificity >65% in the external validation. The poor validation performance across different algorithms indicated differences between the training and validation cohort datasets and pointed to a need for larger cohorts.

Thus, newer models were trained on the combined training and external validation cohorts for ANG2 and CgA (this treated as a single training set). This combined dataset has not been internally validated and thus, internal validation of this joint cohort would provide a further insight into the suitability of these markers utilising a larger train and test portion and future validation of this joint cohort will require access to large independent cohorts. Moreover, other previously assessed markers such as MAC2BP and TIMP1 in the UOL training cohort (Chapter 3), should be assessed using external validation. For example, the CgA and MAC2BP C5.0 non boosted decision tree had an AUC of 0.850 using the training UOL dataset and the APMV general non boosted decision tree which had an AUC of 0.888 in the training UOL dataset. The CVAM combination performed well across SVM, LR and RF in the UOL training dataset assessment (Chapter 3), thus for these algorithms exploring the addition of MAC2BP in an external validation cohort would also be useful to assess the impact of MAC2BP addition on performance.

The ADEPTs cohort contained pNET, PDAC, AP and CP patient samples, selected to resemble a typical clinical challenge of discriminating suspected pNET patients from patients with other conditions that present with similar symptoms. Accordingly, my specific aim was to use CgA, VGF and ANG2 levels to discriminate pNETs from these other conditions. Higher levels of CgA and ANG2 were expected based on my UOL and RFH cohort findings, but the expected level of VGF was uncertain as it was higher than controls in the UOL cohort but lower in the RFH cohort. However, within the ADEPTs cohort, the levels of all three markers were lower in the pNET group than the other disease groups. The unexpected marker levels could have occurred due to (i) random sampling variation due to the low pNET sample number (n=11); (ii) a lack of advanced stage pNETs (Section 6.3); (iii) a potential prevalence of unusual pNET subtypes and (iv) use of different CgA ELISA kit for the UOL cohort to that employed for the RFH and ADEPTs cohorts (although the same type of ELISA kits were used throughout for ANG2 and VGF). The lack of data concerning ADEPTS pNET stage and subtype precluded further analysis. Nonetheless, further studies with far larger pNET cohorts consisting of a wide range of pNET grades, metastatic status and functionality are needed to better assess these variables.

7.1 Clinical value of point-of-care pNET detection

In clinical practice the diagnostic work-up for pNETs depends on whether the pNET is suspected to be functional or non-functional. When a gastrinoma is suspected, a serum gastrin level which is 10-times greater than the upper limit of the normal range along with a gastric pH <2 is diagnostic of a gastrinoma[557]. When an insulinoma is suspected, the serum levels of insulin and C-peptide along with glucose during prolonged fasting (up to 72h) are useful for the diagnosis of insulinomas if symptoms of hypoglycaemia are present or plasma glucose level is below 49mg/dL. In patients with insulinomas, insulin levels are not decreased during hypoglycaemia whereas C-peptide levels are elevated during hypoglycaemia. The diagnosis of VIPoma and somatostatinoma is confirmed through the fasting level of VIP and somatostatin in patients who have suspected symptoms[557]. Functional pNETs, (including insulinomas, gastrinomas, and somatostatinomas) and NF-pNETs can be associated with MEN1 mutation, which typically includes pituitary adenoma and parathyroid hyperplasia and often also includes duodenal gastrinoma. In general, when MEN1 involvement is suspected, serum prolactin, parathyroid hormone and gastrin levels are evaluated[557].

NF-pNETs are more common than their functional counterparts. They cause a lack of symptoms that would otherwise make a patient present to a clinician, and therefore they tend to be identified at a later stage. A single point-of-care test for F- and NF-pNETs would both reduce time to diagnosis and spare patients from undergoing multiple different examinations, particularly for NF-pNET patients.

There is an unmet clinical need for diagnosing pNETs at an earlier stage, when treatment is more likely to provide better outcomes with lesions being surgically resected when possible as this is the only potentially curative therapy; however, patients tend to be diagnosed at a metastatic stage. Thus early diagnosis of pNET patients is problematic[558]. For example, a study of UK pNET patients revealed a median diagnosis time to diagnosis of 24 months after initial presentation although respondents saw their GPs regarding their symptoms five times over a median period of 18 months[65]. A primary reason for late referral for further investigation and diagnosis is that GPs face a considerable challenge in distinguishing patients with suspected malignancy from multiple benign conditions. Moreover, it is extremely unlikely that a typical GP will encounter

many patients with pNETs in a year, due to the low incidence 1/100,000 per year of pNETs.[559]. There is also the challenge of a lack of specific pNET symptoms. For example, the Liao et al. study examined the symptomology of PDACs and pNETs with the aim of identifying PDAC- and pNET-specific symptoms[66]. However, no pNET-specific symptoms were found. Thus, an affordable test that could be used on patients with non-specific symptoms would be most valuable in a primary care or MDC setting. The feasibility of molecular tests in these settings is supported by the existence of several point-of-care molecular tests that have been developed including the GRAIL test, CancerSEEK test and the NET test. These tests work with a multiplex approach for cancer diagnosis, and if in cooperated into routine clinical practice, offer a route for pNETs to be diagnosed.

The Galleri (GRAIL) test is currently being trialled in the NHS as a multi cancer screening tool, with the NHS-Galleri study enrolling over 140,000 people invited from the general population of England aged 50-77 who did not have or were not being investigated for cancer[560]. Approximately 70,000 participants have been either assigned to having the GRAIL test or not having the test (controls). For participants in which a signal is detected from the test, the results are reported back to the participant and they are referred to a diagnostic pathway, with either cancer diagnosed, or no cancer diagnosed, with the no cancer diagnosed participants returning for 1 and 2 year follow up visits. For the control participants who are not provided with a GRAIL test, the sample is stored for potential future testing. The NHS-Galleri trial is the first randomised controlled trial, that is statistically powered to assess the clinical utility, including harms and benefits, of a blood based multi-cancer early detection test. However, the trial has some limitations, including the test only being evaluated only in a population of participants aged 50-80 years, the trial not being powered to evaluate the benefit of screening separately in specific socioeconomic, or ethnic subgroups, or not evaluating the clinical utility of each individual cancer type detected or the relative benefits of different screening schedules such as biannual or biennial screening. The trial is also not designed to answer whether a multicancer early diagnosis test could replace or change the frequency of other current single cancer screening programmes. There is also the concern that early diagnosis may not ultimately alter the disease course and participants with cfDNA detectable amounts still die at the same point in the absence of this

intervention. The scope for overdiagnosis however is limited with the test signal being based upon circulating tumour fraction which is the proportion of ctDNA in cfDNA. The GALLERI test has been shown to detect pancreatic cancer with a sensitivity of 63%, with the sensitivities ranging across stages (stage I= 63%, stage II= 83%, stage III= 75% and stage IV 100%)[79]. Thus, in the future, if this test were to be rolled out into the NHS routinely, pNETs are likely to be identified from this test and thus implementation of this test in a diagnostic pathway would provide another route for pNET diagnosis in the future.

The CancerSEEK test is another test that could be used in the context of diagnosing patients who may have non-specific symptoms. Developed in the US for the detection of eight cancers (liver, stomach, ovary, pancreas, oesophagus, colon, lung and breast), through assessment of mutation in cfDNA and levels of circulating proteins, the test was found to have sensitivities ranging from 69% to 98% at a fixed specificity of 99% and localised the site of cancer in a median of 83%[82]. The features of the test that were most important to the algorithm were the presence of a ctDNA mutation followed by elevations of CA-125, CEA, CA19-9, pro-lactin, hepatocyte growth factor, osteopontin, myeloperoxidase and TIMP1 protein levels. The CancerSEEK study found that the sensitivity for the earliest stage cancers (Stage 1) was highest for liver cancer (100%) and lowest for oesophageal cancer (20%). However, a major limitation of the study was that the patient cohort consisted of participants with known cancer and the true extent of false positives could not be explored, because the controls were limited to healthy individuals, and in reality non-cancer patients would exist including patients with inflammatory or other diseases which could have produced a greater proportion of false positives than what has been identified in this study. They also used cross validation an internal validation approach and did not independently externally validate the test. The portion of cancers of each type in the cohort was purposefully not representative of those in the US due to the study wanting to evaluate at least fifty examples of each cancer type with the resources available. CancerSEEK have also considered costs, with the estimation of the test costing less than 500 dollars which is comparable and lower than other screening tests for single cancers such as colonoscopy. Thus, the utility of CancerSEEK as a multi-cancer diagnostic test used in a similar way to the Galleri test for patients, might

provide another opportunity for diagnosis for pNET patients with non-specific symptoms.

The NETest which utilised 51 different RNAs for NET detection, has been found to perform well as a pNET diagnostic test with an AUC of 0.939 (n=87) for pNETs[83]. However, despite this test being comparatively better to CgA, costs, reproducibility and practicalities means that better diagnostics for pNET detection are needed for a point of care test.

7.2 Performance of the selected pNET markers

Biomarkers for both F-pNETs and NF-pNETs include CgA[93], but the use of this marker is compromised by the absence of an approved FDA test[512], with CgA levels varying between assaying methods, the type of biological material used (plasma or serum) or even between specific ELISA tests[512]. In this respect, the UOL cohort used a different CgA ELISA kit to that employed for the RFH and ADEPTs cohorts. This lack of standardisation limits the use of CgA for diagnosis, comparison of different cohorts, and studies of this marker. The use of CgA is further limited in man by non-pathological, day-to-day factors including food intake, and by many confounding factors, including malignancies, non-malignant conditions and certain medications (Section 1.5.1). Thus, if CgA were to be included in a biomarker panel, these factors would need to be controlled to reduce false positive results. When assessing CgA in the external validation cohort assessing for clinical for confounding factors, it was found that that CgA was associated with PPI usage, p=0.048. Thus, this is definitely a factor that limits its utility as a biomarker for pNETs and supports findings from other studies[202,510].

An algorithm that is developed for CgA should ideally be able to use plasma or serum interchangeably with a suitable quantitative adjustment built into the algorithm as plasma CgA levels are generally higher than serum levels. In this thesis the thresholds used for the serum algorithms were expected to be applicable to pNET plasma but may have been too low for control plasma. Whether to use serum or plasma for future algorithm development will likely depend primarily on the clinical availability of serum or plasma from pNET patients and tissue banks.

NSE is a second marker that has been assessed previously as a F- and NF-pNET marker, but it has proven to be less accurate than CgA in identifying GEP-NETs[561]. For example, in a study by Baudin et al. serum NSE and CgA levels were measured in 128 NET patients and 53 controls with non-endocrine tumours to compare their sensitivity and specificity. In all groups of NET patients, CgA proved to be more sensitive than NSE. NSE and CgA had a specificity of 73% and 68%, respectively and were elevated in 48 (38%) and 74 (59%) of the patients[562]. Raised NSE levels were significantly associated with poor tumour differentiation as raised levels were found in 12/19 patients with poorly-differentiated or intermediate GEP-NETs compared to 23/71 patients with well-differentiated GEP-NETs[562]. They also found that among six NET patients who were followed for 11-37 months, CgA appeared to be a better marker of tumour evolution. Indeed, the authors suggested that CgA ought to be the only general marker used to monitor NET patients[562]. However, when Lv et al. assessed 784 serum samples using a test that combined serum CgA and NSE levels, they found an improved power of prognosis for NF-pNET patients[468]. The Lv et al., study focussed on the impact that a combination of NSE and CgA and CgA alone would have on diagnosis of a more focussed group of pNET patients (NF-pNETs), whereas the Baudin et al., study explored CgA and NSE separately in a NET patient cohort. Thus, the differences in the findings for the suitability of NSE, could be due to specific characteristics of a NF-pNET cohort compared to a more broader NET cohort group.

From my assessment of NSE as a pNET single marker in the UOL cohort in Chapter 3, it was found to be one of the markers that across C5.0 non boosted decision trees and LR performed poorly. Moreover, when assessed as a combination with CgA for the C5.0 non boosted decision tree, the model created omitted NSE and instead created a model with just CgA suggesting that NSE was not important for that model (Chapter 3). For the LR algorithm NSE individually was one of the poorer single markers with an AUC of 0.564, with CgA individually in comparison having an AUC of 0.750. The NSE and CgA LR model for the UOL training cohort had an AUC of 0.767, thus compared to CgA alone, the combined model performed slightly better with the addition of NSE, however other CgA two marker combinations performed better such as the CgA and ANG2 combination which had an AUC of 0.849 for LR in the UOL training

cohort. Thus, suggesting that addition of other markers to CgA compared to NSE was better.

ANG2 was identified in the UOL case control analysis as a pNET marker, which was supported by the RFH case control analysis. However, ANG2 levels were not significantly different for pancreatitis, PDAC and pNET patients in the ADEPTS cohort, indicating that ANG2 utility may be limited in a clinical setting although larger cohort analyses are required to confirm this point. ANG2 levels in pNET plasma samples should also be assessed, because if, like CgA, plasma levels for these markers are higher than in sera, it is likely that plasma ANG2 would be more suitable for future algorithm development than serum. ANG2 also has many confounding factors, including a range of cancers, non-malignant conditions and certain medications (Section 1.5.5).

Thus, the use of ANG2 and CgA for pNET diagnosis in a primary care setting would be challenging, although ML models may be able to employ fine tuning of the thresholds used for the for pNET detection combined with confounding factor assessment to allow successful diagnosis.

7.3 Discovery of early detection biomarkers using pNET patients

Cancer biomarker discovery has progressed slowly with the currently clinically used serological tumour markers being discovered at least 30 years ago. Since that time no new serological tumour markers have been brought to the clinic[563]. Biomarkers that are identified when a person has symptomatic cancer are likely to reflect changes that occur a long time after the initiation of cancer and hence have little value for the early detection of cancer. This is a major problem with early biomarker research, typically due to the practical issue that diagnosed cohorts tend to be representative of later rather than early-stage disease. The ideal early biomarker test would have been developed using specimens from patients with no symptoms of the target cancer at the time of collection and who later developed the target cancer. In practice this is only feasible for either extremely large, untargeted specimen collections such as the UKCTOCS and UK Biobank collections, or for high-risk groups carrying specific gene mutations, such as MEN1. The very low incidence of pNETs favours the latter approach, but it remains to be seen whether MEN1 serum markers can also be used to detect other pNETs.

For the early detection of cancer with a biomarker, there should be a moderate or preferably high TPR associated with a low FPR. Low FPR is important to prevent unnecessary further investigations and patient anxiety; however, a low FNR is important to prevent patients from not being picked up by the test. Another limitation of cancer biomarker research is that little attention has been given to the frequency of specimen collection and storage[564]. For practical reasons, specimens are often collected infrequently, typically only once or twice from a patient[564]. Thus, longitudinal changes in a patient cannot be detected and biological changes in patients cannot be tracked over time. These biological changes could allow the optimum point for detecting different types of pNET to be assessed for different markers. A longitudinal study with a statistically adequate number of pNET patients would be useful for improving biomarker tests but has never been carried out.

Carcinogenesis is centred on two different theories, somatic mutation and tissue organisation field theory. Somatic mutation is centred at the cellular level of biological organisation and states that carcinogenesis is a problem of cell proliferation control. However the tissue organisation field theory considers carcinogenesis to be a process that is similar to organogenesis that has gone wrong and is centred at the tissue level[565]. Under the somatic mutation theory cancer begins with an initial mutation which proceeds through subsequent mutations[565]. Thus, infrequently collected specimens could still identify long-term irreversible genetic changes that would lead to a good biomarker test for the early detection of cancer[564]. Under the tissue organisation field theory[565], the cause of cancer is a disruption of cell communication that precedes irreversible changes associated with cancer. Biomarkers that are related to the disruption of cell communication may not detect cancer early because such disruptions may be reversible or not sufficient in duration to initiate cancer[564]. If tissue organisation field theory is correct then frequent specimen collection is recommended in order to identify biomarkers that arise in the small time window[564].

7.4 Discovery of early detection biomarkers using pNET models

pNET early diagnosis biomarker discovery has been hindered due to the lack of suitable biological models available for pNETs. Establishing experimental models that recapitulate the genesis and progression of pNETs is important to

increase understanding of the biology of this disease, but models are typically limited by their inability to recapitulate the genetic mutations and features of human neoplasia. pNET models include cell lines[566], mouse models[567], patient-derived xenografts (PDXs)[568], spheroids[569] and islet-like tumoroids[570]. GEP-NET cell lines have the advantages of low cost, empirical reproducibility, availability and being well characterised. However, a limitation is the genetic modification of the primary cell line over time[571]. Each model type has its associated advantages and disadvantages. 2D cell culture models of pNET cell lines, such as BON1, QGP1 and NT3, do not recapitulate the tumour well and have limitations. Kaku et al. established the QGP-1 cell line from a somatostatin-producing islet cell carcinoma[572], while the BON1 cell line was established in 1991 from a lymph node metastasis of pNET patient[573]. The QGP1 and BON1 cell lines have been widely used but they do not display a well differentiated neuroendocrine phenotype[566]. Moreover, they do not serve as a model to study radionuclide imaging and SSTR expression[574]. For the BON1 cell line the doubling time is less than 48 hours, which means that this cell line does not recapitulate characteristically slow growing NETs[575]. The NT3 cell line[566] is well differentiated with a low proliferation rate, but it was isolated from a lymph node metastasis rather than the primary tumour, a functional insulinoma. While candidate pNET markers have been identified from these cell lines[107] , better biological models of early pNETs are required to support the identification of new, early GEP-NET and pNET markers.

PDX models have been developed for pNETs, and used to evaluate two mTOR inhibitor drugs, everolimus and sapanisertib[568]. The PDX-pNETs maintained pNET morphology and pNET specific gene expression signature with serial passage and harboured mutations in genes that were previously associated with pNETs including MEN1 and PTEN, as well as activation of the mTOR pathway which was detected by Gallium PET-CT. Moreover, treatment of the PDX-pNETs with either everolimus or sapanisertib inhibited growth, as well as in some PDX-pNETs there was resistance to everolimus. Work by Lines et al., in the development of a pNET mouse model which is under temporal control allowed to study the impact of early tumorigenic events and overcome limitations of other MEN1 knockout models, whereby menin is lost from conception. They developed the pancreatic β cell specific NET model under temporal control (MEN1$^{L/L}$/RIP2-CreER) mouse model. These mice at

approximately 3 months old were given tamoxifen in the diet for 5 days, and pancreata then harvested, at 2-2.5, 2.9-3.5 and 4.5-5.5 months later. The control mice did not express Cre and did not receive tamoxifen, and the tamoxifen treated MEN1^{L/L}/RIP2-CreER mouse compared to the control mice showed at all ages, loss of menin in all islets, increased islet area, increased proliferation of insulin immunostaining β cells and decreased proliferation of glucagon immunostaining alpha cells. Their mouse model, of a time-controlled loss of menin may assist in providing a useful model for investigating early genetic and molecular mechanism that may occur in pancreatic islet cells, before, during and after menin loss. Hence development of such models is important in further understanding the early-stage biology of pNETs. Genetically engineered mouse models have the advantage of the tumour being present within the context of an intact immune system, an established causal relationship with genetics and phenotypes, and suitability for drug screening applications. However, limitations include biological differences between human and mice, and familial GEP-NET models may not be suitable for drug development, as most GEP-NETs are not familial, hence findings from familial GEP-NET models that harbour germline mutations may not be generalisable to sporadic pNETs. PDXs have the advantage of close resemblance of pathology to human samples including stromal interaction, however a limitation is that the engraftment rate of GEP-NENs is less than 10%[571]. Organoids have associated advantages including the ability to generate organoid libraries, be utilised for drug screening and for testing of personalised medicine, and to offer the potential to model the cancer pathway of GEP-NENs through genetic engineering, however they also have disadvantages including a lack of native environments which includes the tumour stroma, vasculature and immune system, a loss of heterogeneity and there being insufficient data on GEP-NENs[571]. Advantages of spheroids include being inexpensive, high efficiency and keeping physical interactions which are more closely reflective of behaviour in the 3D native tissue. Whilst disadvantages include having variable diameter and size, intense work and the diffusion gradient depending on the size (oxygen, nutrient, paracrine factor) that decreases inwards[576]. Tumoroids which are derived from patient derived tumour cells, have the advantage of recapitulating the tumour architecture in vivo and resembling the heterogeneity of the original tumour, however a disadvantage of spheroids is difficulty in

producing a large number of homogenous tumoroids for high-throughput drug testing[577].

The initial seven GEP-NET markers used as the basis for exploration in this thesis were based on secretomic analysis of the pNET BON1 cell line as well as the lung NET cell lines NCI-H727 and SHP77 NET. In this analysis a pNET-targeted approach was not initially adopted when these seven markers were selected. Markers more suitable specifically for pNET diagnosis thus may have been excluded. Future work should focus on pNET specific cell lines only, including the QGP1 cell line and the newer NT3 cell line, for pNET-specific marker identification using secretomics. When comparing whether to use a 2D or 3D of these cell lines for the identification of newer markers, factors such as cost, difficulty, reproducibility and likeness to in-vivo are to be considered. 2D cultures have the advantage of being low cost, replicable and interpretable, however a 3D culture system offers a greater insight and has more similarities to in vivo. Advantages of the 3D system is that the natural cell shape is preserved and that cells grow into 3D aggregates/spheroids with the spheroids containing multiple layers. Additionally, cell junctions are common and allow for cell-to-cell communication whereas in 2D culture cell junction are less common and less accurately represent real junctions. In 2D cell lines, cell differentiation is poor, whereas in 3D cell culture cells are well differentiated. Drug sensitivity in 2D cells, are that cells often have little resistance to drugs compared to in 3D whereby cells often have more resistance to drug treatment, thus giving a more accurate representation of the drug's effects. Cell proliferation is also at an unnatural pace in 2D cell cultures compared to in 3D cell cultures which are more realistic and can be high or low dependent on technique and types of cells. Finally, expression levels, with gene and protein expression levels in 2D cell cultures often vastly different compared to in vivo models, whereas with 3D models gene and protein expression levels resemble levels found in cells in vivo[578]. Thus, adopting a 3D culturing approach for these pNET cell lines would be better.

3D culture systems have been developed and used in the context of pNET research[569,579]. The Bresciani et al., study evaluated spheroid 3D culture methods to study the BON1 cell line used as a model alongside sunitinitb[569]. The authors compared different 3D scaffold-free culture systems, including 96 well hanging drop plates (HD plates), 24-well plates with a cell repellent surface

and ultra-low attachment 96 well plates with clear round bottom (ULA plates) in order to understand which methodology was the best option in the study of pNETs in terms of easiness of culture method and reproducibility. They found that the ULA plates method allowed to obtain the most reproducible results when assessing perimeter evaluation in BON1 spheroids compared to other assessed methods. A study by Gulde et al., explored the single and combination treatment of buparlisbib (PI3K inhibitor) and ribociclib (CDK4/6 inhibitor) in different preclinical models[579]. This included using cell lines representative of well differentiated (INS-1E and NT3) and poorly differentiated (BON1) pancreatic neoplasms grown in 2D and 3D, primary 3D microtissues from MEN1 knockout and control mice and patient derived primary 3D tumoroid cultures (specifically 3D tumoroids from four human pNET tissues (two primary tumours and two metastases)). The findings across the model types indicated that the combined inhibition of PI3K and CDK4/6 pathways is a potentially effective therapeutic option for pNETs. They also demonstrated that the combination treatment inhibits the viability of primary islets from a genetic animal model of pNETs (MEN1-dffecient mice) without significantly affecting viability and function of primary islets from wild type mice. Hence, this study has adopted multiple pre-clinical pNET model approaches to explore buparlisbib and ribociclib and utilised models other than and along with 2D cell lines to explore this.

Conditioned media from different, early stage, primary pNET models could help to identify commonly secreted proteins in both the cell lines and the organoids which would aid the selection of markers to take forward for the development of a pNET detection test. However, an issue is that early-stage cell lines are not available. Thus, the lack of these cell lines, limits the opportunities for early biomarker identification. It is possible that some of the markers explored in this thesis would be identified in these analyses, which would support their inclusion in a biomarker panel. A more informed and targeted method for selecting pNET markers is likely to improve the ML approach and algorithms compared to those developed in this thesis.

7.5 Challenges in biomarker development

Numerous biomarkers have been combined in several cancer tests such as GRAIL[560], NETest[83] and CancerSEEK[82], but in practice, simple biomarker tests

are more likely to be employed early in clinical pathways on the basis of cost and point of care use, if of course their classification accuracy is adequate[564]. The work described in this thesis aimed to address this need by developing an algorithm that could perform well with a few biomarkers. The ANG2 and CgA combination achieved this with the UOL training cohort and when externally validated using the RFH cohort. However, an unresolved issue arose with ANG2 and CgA concerning its low levels in the ADEPTs pNET cohort, meaning that further studies are needed with a much larger group of pNETs, as used in the UOL and RFH cohorts.

The work described in this thesis also illustrates several more general challenges in biomarker discovery and validation. First, the training cohorts used for model development were selected to target discrimination between cases and healthy controls, similar to most published studies on NET biomarkers. However, as the main end goal of a pNET test would not be just to differentiate pNET from healthy individuals, but to differentiate pNET cases from individuals with similar presentation caused by other diseases, the composition of the training cohort is critical to achieve this goal, and the samples that were collected for this work and many other research studies do not reflect this. Although, case and control models remain relevant, as models need to be able to make a distinction between both in a clinical setting. Second, an early diagnosis test requires patients to be picked up earlier in the clinical pathway, hence biomarker discovery and model development ideally need to be carried out using a cohort of low stage pNET patients. However, as pNET patients are typically diagnosed later[580], with the SEER registry from 1973-2000 characterising out of 1483 patients with pNETs that 60.2% of patients were either metastatic and 20.7% regionally advanced at the time of diagnosis. Thus, if these patients provided samples for research, the samples accumulated for pNET research samples would tend to be more of a late stage and thus biomarkers that are identified and researched would tend to be more suitable for advanced disease and prone to failure in downstream validation for early-stage detection. This issue is a clear hindrance to biomarker discovery and validation that needs to be addressed in future studies aiming to diagnose pNET patients at an earlier stage. It would also be ideal to train on real clinical cohorts which are drawn retrospectively from primary care.

A third major consideration is the inappropriate selection of controls, which is a major cause of systemic bias. In prospective studies, samples taken from apparently healthy volunteers or from patients with a short follow up period that does not lead to a diagnosis may be used as controls. The ADEPTs cohort of PDAC, AP, CP patients which were used as 'controls' compared to the pNET group (Chapter 6), illustrates this issue of lack of follow up, as the patients have not been followed up for this work. Without longer follow-up and assessment of disease status, a study may therefore underestimate a biomarker's capacity to discriminate between individuals with and without disease. In this regard, very limited information was available regarding controls in the UOL control cohort.

Fourth, potential confounding factors for the biomarkers explored in this thesis were identified in the literature as illustrated in Section 1.5. Some confounding factors explored such as PPI usage were found to be confounding for CgA in my work which is in keeping with the literature. However, in practice, when having an inclusion and exclusion criteria for patients, it is not possible to control for all potential confounding factors. This is due to the reality that patients exist with multiple co-morbidities, a major challenge to health care systems around the world[581]. Patients with multiple comorbidities, also tend to be on different medications and thus there are polypharmacy comorbidities, including increased drug-disease and drug-drug interactions in these patients[581]. Also, for biomarkers, not all potential confounders would have been researched and identified. Exclusion of such participants would be important to assess the clear relationship between the disease of interest and the biomarker assessed, and to minimise the risk of additional confounders. However, exclusion of such patients from studies, means that the cohort size may be too small. Moreover, exclusion of such patients in a real-world scenario clinical scenario is not possible. Clinicians would need to be able to use a biomarker test on all suspected patients. Hence, in the future, algorithms would need to be developed that can accommodate their inclusion. This could potentially be via altered thresholds taking into account known confounding factors.

The importance of identifying confounding factors has been noted in many other biomarker studies. For example, a researcher group that aimed to develop an algorithm to detect prostate cancer found that their decision algorithm was unsuccessful in discriminating prostate cancer from controls with any predictive utility due to confounding factors arising from disparities in sample storage[582].

When developing a test for pNETs the biomarker should identify a wide range of pNETs with minimal false positives. For functioning NETs which are often clinically silent for many years with signs and symptoms often only becoming clinically manifest when metastases are already present, however among the many bioactive substances secreted by NETs some may be included in the diagnostic process[583]. However, given the rarity of NETs and the confounding factors associated with them screening for NETs would be challenging. NET markers can be divided into two types categories either specific or non-specific. With specific NET biomarkers produced by functioning NETs and non-specific being produced by all virtually all NETs. Non-specific markers for NETs include CgA, NSE, pancreatic polypeptide, human chorionic gonadotropin and alpha-fetoprotein. Specific NET markers include gastrin, serotonin, insulin, glucagon, somatostatin and vasoactive intestinal peptide, however these would be specific for a particular type of functional pNET and not broad enough to detect other types of pNET. Ultimately a pNET test that is developed should be sensitive enough to identify all types of pNET, with specific NET markers only suitable for functional types. Despite non-specific markers being suitable for detecting a wider range of NETs. Measurement of biologically active amines is not suggested for screening, with generic tumour markers such as the non-specific marker CgA being increased in several clinical conditions which could possibly lead to false positives. Hence leading to its unsuitability for a use also as a marker for detecting pNETs in a screening context.

In addition, although a test may show reasonable sensitivity and specificity during development, the test itself may not be adopted clinically if the positive predictive value (PPV) and negative predictive value (NPV) are below clinically defined thresholds[555]. The PPV of a test is the chance of having a disease if the test is positive. The PPV is dependent not only on the sensitivity and specificity of the test but also on the disease prevalence among the test population. For rarer cancers such as pancreatic and ovarian cancer, for which prevalence in tested populations is <0.05%, an adequate screening test is characterised by a very high sensitivity (>90%) and even higher specificity (>99%) to yield a PPV of >10%[555]. Investigators can report sensitivities at 90%, 95% of 98% these high thresholds may still not be enough or screening of less common cancer such as pancreatic and ovarian. An example is if the cancer prevalence in the screened population is 1:400 a test with 100% sensitivity and 95% specificity

yields a PPV of 5% which means that there is a 5% chance of a patient truly having cancer if the test is positive[555]. Hence screening tests must have extremely high specificity to ensure high PPV and avoidance of false positives. This principal would also apply to pNETs, which are rarer with an overall incidence of 5 per 10,000,000 and have a prevalence of 1 in 100[1,78]. Thus, any test developed for screening for pNETs would need to have an extremely high specificity to ensure high PPV.

Finally, tumour heterogeneity is also important to consider and can explain some of the reasons as to why promising markers fail at validation. Cancer is not a single disease, and cancer genomics has revealed that there are many more forms of cancer than is apparent by traditional classifications based on morphology[584]. With Genomic methods such as whole exome, whole genome and single cell sequencing, a much-improved understanding of the vast landscape of tumour heterogeneity has become established over recent years[555]. With increasing evidence for high variability in tumour composition among patients with similar histological cancer types[585], a newer approach to biomarker research is perhaps needed. Instead of the orthodox approach of discovering biomarkers that are common for a cancer type, we might instead analyse serum samples from individual patients to identify the most informative set of markers for their unique tumour, effectively making the approach more personalised[563].

Hundreds to thousands of biomarkers have been reported in the literature to show high specificity but low sensitivity and therefore have not developed further. Thus, the approach of compiling a collection of rare tumour markers which are informative for only 5-15% of patients has been suggested to offer a more robust tool with which to screen patients and to provide information on their unique tumour[563]. Indeed, the authors postulated that physicians would soon be able to submit serum samples from newly diagnosed cancer patients to centralised laboratories which would screen 100-1000 candidate markers with the aim of identifying 1-5 rare tumour biomarkers that would be most informative for these patients[563]. A similar approach could be adopted for pNET patients to address the heterogeneity of tumours. Moreover, this personalised approach could also find applicability in selecting therapies, response to therapies and to monitor for recurrence. A personalised approach towards pNET biomarker discovery and validation is something that future work within this field should

aim to address. This may also allow for the barrier for the low number of pNET patients to be overcome.

7.6 Strengths and limitations of my approach

Amongst the strengths of the work described in this thesis, the first is the development of ML algorithms using the UOL cohort assay data combined with different validation approaches. These include train and test internal validation using samples separated into training and testing portions; kFCV in which the data used for training is re-used for validation; and external validation using a different cohort. The variety of validation approaches allowed for the suitability of the markers and algorithms to be evaluated more thoroughly than using a single approach. In addition, the use of the independent RFH-UKCTOCS cohort provided a more robust assessment of the general use of the marker and algorithms than using a single cohort.

Another strength of this work includes the assessment of four different types of algorithm, namely the simpler and easier to interpret LR and C5.0 decision tree models and the more complex SVM and RF models. Understanding how the markers were being utilised in the simpler models will allow future comparison between different models and consequent fine tuning. The more complex types of models are black box nature and pose questions as to how such algorithms could be protected commercially for adoption into clinical practice.

Limitations from this work were that only three (CgA, ANG2 and VGF) of the seven starting markers were taken forward for external validation using the RFH and ADEPTs cohorts. The initial selection of CgA, VGF and ANG2 was based on the rationale that CgA, ANG2 and VGF were the top performing markers for the UOL cohort, and this was supported by the internal validation results discussed in Chapter 4. However, as the external validation results obtained for these three markers showed widely different results, particularly for VGF, a re-evaluation of the other four markers using the external validation cohorts may be informative and algorithms utilising one or more of these markers might be better suited to these cohorts. Moreover, the other four markers not assessed at external validation may have potential as discriminatory markers, when comparing pNETs with AP, CP and PDAC and thus this should be explored in the future. The Luminex platform is a multiplex platform which requires a very

small amount of serum. Utility of high-throughput assays which require minimal sera offer a useful way forward for this work in the future compared to an ELISA approach if the required target assays are available. Table 7.1 illustrates the different learning points and future areas of work from this thesis.

Learning points	Future areas of work
• The importance of using different validation approaches to evaluate biomarkers including train and test internal validation using samples separated into training and testing portions; kFCV in which the data used for training is re-used for validation; and external validation using a different cohort.	• Conditioned media from different, early stage, primary pNET models could help to identify commonly secreted proteins in both the cell lines and the organoids which would aid the selection of markers to take forward for the development of a pNET detection test.
• Ultimately markers can fall through at different points and ideally all markers should be assessed using external validation.	• Assess whether markers assessed in this thesis were identified in analyses of early-stage cell lines.
• Utility of different ML approaches allows for performance to be evaluated and suitable algorithms to be identified.	• Use of 3D culture systems for biomarker discovery.
• Complete staging, grading, and functional characteristics for understanding the impacts of these characteristics on marker levels.	• Development of early-stage cell lines for biomarker discovery is needed.
• Understanding confounding factors for certain biomarkers which can impact their utility clinically.	• Assessment of early-stage markers identified in early-stage models and assessed in sera using a multiplex platform such as the Luminex platform and development of ML algorithms and early-stage markers.
• The importance of having cohorts that are reflective of early-stage disease for assessing markers for early-stage utility.	• Validation of early-stage markers in large early-stage clinical cohorts to assess suitability of markers using independent cohorts.

Table 7.1: Learning points and future areas of work from this thesis.

7.7 Future directions for clinical pNET detection using molecular markers

Circulating biomarkers have been studied by various groups for diagnosing and following up NET patients. These include protein, monoamine, CTC[586], microRNA[85] and metabolomic[89] approaches.

With functional NENs it is possible to analyse secreted hormonal markers in blood or urine samples, but for non-functional tumours only more general markers can be used. As most pNETs are non-functional, the use of hormone levels alone for pNET diagnosis is not reliable.

The currently used secretory protein and monoamine biomarkers, CgA, NSE, pancreastatin and serotonin, are mono-analyte measures that lack sensitivity, specificity, and predictive capacity. Consequently, none of these meet NIH metrics for clinical usage[587]. The best-established non-specific marker for NET diagnosis and prognosis is CgA and the pressing need for additional pNET markers is exemplified by the results of the EXPLAIN study which found that only 57% of pNET patients had CgA levels >ULN and CgA did not classify as one of the top 10 biomarkers for pNET detection[588]. Additionally, it was only the 7th most important biomarker for discriminating between pNETs and si-NETs in the best models. Serum NSE has been studied previously as a GEP-NET marker, however work from this thesis. illustrates that NSE has limited role in diagnosing pNETs. Pancreastatin which is a derived peptide of CgA was assessed in a cohort of 92 NET patients and 11 non-NET patients[589]. The radioimmunoassay analysis of pancreastatin found that mean pancreastatin levels were significantly higher in the 92 NET patients compared to the 11 non-NET patients (p<0.05). Pancreastatin had a sensitivity of 64% and specificity of 100%[589]. The study found that pancreastatin showed significant diagnostic value as a biomarker to identify patients the a known NET where CgA was found to be normal[589]. Pancreastatin was also been shown to predict survival in NETs, with higher pancreastatin levels associated with worse PFD and OS in pNETs[590]. The study also found thar small bowel NETs had lower pancreastatin than pNETs[590]. Excessive serotonin secretion is associated with carcinoid syndrome and the ENETs Consensus Guidelines suggest measuring urinary serotonin and serum CgA in all patients with NETs both as part of follow up and diagnosis[591]. Despite good sensitivity and specificity (70% and 90% respectively) urinary serotonin levels might be normal in non-metastatic patients[583]. Serotonin levels are also influenced by malabsorption, celiac disease and tryptophan rich foods, hence the issue of false positives[592,593]. Serotonin is also impaired in other conditions such as renal failure and haemodialysis which could result in falsely negative concentrations[583]. Moreover serotonin secreting NETs of the pancreas are rare, and thus the need for better biomarkers for diagnosing pNETs[594].

The traditional approach of using monoanalytes for NET diagnosis have variable sensitivity and specificity in diagnosing different NETs which is mainly due to the high heterogeneity of these tumours[595]. Moreover circulating

indicators of tumour secretory functionality (amines or peptides) represent the minority of NETs[494]. There is also the limitation of CgA not being a suitable marker for certain types of pNET (insulinoma)[90]. Thus, a multianalyte approach would likely have a broader reach compared to a single marker approach, thus multianalyte approaches have been developed more broadly for multiple types of cancer and NETs specifically. These include the NETest, CTCs, CancerSEEK and Galleri test.

CTCs have been assessed in studies of NETs[586,596] A study by Khan et al., found that for midgut and unknown primary NETs, 47% had CTCs detectable and 68% of whom had greater than 5CTCs/7.5ml[596]. These levels were found to be comparable with other tumour types, however these were not as high as seen in metastatic breast and prostate cancers, likely reflecting the indolent nature of most NETs[597]. When looking at pNETs in this study, despite pNETs being EpCAM positive, only a small proportion had CTCs detected compared with midgut NETs[596]. Explanations for this finding include loss of EpCAM expression, slow shedding of CTCs from pNETs or unidentified factors particular to the pancreatic sample[596]. Hence, CTCs have more diagnostic utility for certain NET subtypes compared to others, and for pNETs they may not be as useful based on this study. Another issue with CTCs includes a technological aspect of the measurement[587]. At present there are many different detection techniques, not one single technique can detect all different types of CTCs[598]. In order to improve the accuracy of CTC tests, a standard set of performance assessment criteria is needed to be developed, including enrichment, cell viability and release efficient capture efficiency and purity, this will therefore improve the sensitivity and specificity of CTC detection. Moreover, despite CTCs having potential for real-time monitoring, prognosis and diagnosis of tumours, CTCs alone cannot be used as an effective indicator to guide the formulation of clinical treatment plans for tumours and would need to be used with other biomarkers for this[598]. However, CTC and ctDNA detection has shown to demonstrate greater sensitivity and accuracy in monitoring of breast cancer progression compared to serum biomarkers CA15-3 with the study conducted by Dawson et al demonstrating a superior sensitivity of ctDNA quantification (96%) in identification of patients with metastasises compared to CA 15-3 measurement (78%)[599,600]. Protein biomarkers are known to remain in

circulation for weeks and be impacted by many factors leading to misleading results, however ctDNA is estimated to have a short half-life of approximately 2h[601].

The NETest is a 51 different circulating ribonucleic acids (RNAs) for NET detection[83]. The NETest multianalyte biomarker approach was developed by Wren Laboratories in the USA[595]. The NETest output is a score that is scaled from 0-100 which represents the risk of NET disease. The normal score is <20, stable disease 21-40, progressive disease 41-100[602]. The NETest was developed based on the identification of individual genes from tumour cells and whole blood samples from patients with the finalised NETest now a standardised and reproducible biomarker platform[595]. The NETest as opposed to CgA or other monoanalyte peptides and hormones works using a multi-analyte molecular signature which represents biological information related to the clinical neuroendocrine disease course[494]. A study of the NETest in 13 si-NETs patients who underwent surgical resection of the primary tumour and or mesenteric mass found that the blood NETest scores accurately identified si-NETs and that the NETest levels were significantly decreased by curative surgery[603]. Hence, this study suggested that monitoring NETest post-operatively may facilitate management of these patients by identifying the presence of residual/progressive disease. The NETest also has utility in diagnosing gastric NETs, with the NETest being increased in gastric NETs (23±11) vs controls (7±4 p<0.0001)[604]. In histology positive NETs, the NETest accuracy was 100%. Additionally, in histology and image negative disease, elevated NETest could reflect increased neuroendocrine gene expression of hypergastrinemia-induced neoplastic transformation of enterochromaffin-like cells to tumour status. However, the study had found a positive NETest being evident in 3% of controls[604]. The NETest has also been found to be effective for the diagnosis of pNETs and si-NETs. The study found that the NETest was significantly increased in 111 NETs compared to controls[494]. Patients were either image positive (n=75) or image negative (n=11) with the image positives consisting of 42 pNETs and image negatives consisting of 8 pNETs. The NETest was found to be significantly higher in image positive disease (36±22) compared to image negative disease (8±7 P<0.0001). Moreover, the NETest achieved high accuracy, sensitivity and specificity which were 97%, 99% and 95% respectively[494]. Hence this study demonstrated that a blood-based multigene

biomarker provides accurate information that is concordant with imaging. Utility of the NETest information obtained by venipuncture from patients could provide a point of care basis for monitoring disease, hence reducing patient exposure to radiation[494].

Despite the NETest strong performance there are still limitations to its utility. Firstly, that the data available is still limited with larger multi centre cohorts evaluating NETest performance needed. Additionally, future prospective studies are needed to evaluate NETest performance on specific NET types as well as different grades and stages. However, an important issue with the NETest is cost. However, cost effectiveness for the NETest has been assessed in a study looking at NET tumour recurrence[602]. Moreover, using the NETest to stratify postoperative imaging resulted in a cost saving of 42% and for pNETs specifically (n=42) there was a cost saving of 46.2% but further studies are required to evaluate the cost effectiveness of the NETest in the detection of postoperative recurrent disease[602]. Additionally, the MEN1 patient group have not had the NETest extensively studied[605]. Currently existing serum biomarkers do not have the necessary NPV to replace imaging as a way of detecting MEN1 patients at risk for progression or metastatic disease. Fahrmann et al., have assessed a group of three polyamines N-acetylputrescine, acetylspermidine and diacetylspermidine to distinguish MEN1 patients with metastatic duodenal-pNETs from controls who were defined as MEN1 indolent duodenal-pNETs or MEN1 with no duodenal-pNETs and found that the signature for distinguishing metastatic duodenal-pancreatic NETs from the controls with 66.7% sensitivity and 95% specificity respectively and an AUC of 0.84[606].

The Delphi technique aims to develop a consensus about an issue from a panel of experts or stakeholders. This approach has been applied to determine which types of biomarker provided the most utility for NET diagnosis[587]. The assessed biomarkers in the Delphi-consensus included mono-analytes (CgA, pancreastatin and serotonin), CTCs, miRNAs, metabolomics and mRNAs. However, a limitation is that cfDNA was not considered, and specific multi-analyte protein marker approaches were not considered. However, the need for multianalyte biomarkers was mentioned. These are also areas of research interest for biomarker development. A consensus level of 75% was used to indicate clear evidence of a majority opinion, with voting anonymised and discussion when there was no consensus. The consensus was negative for

CTCs (70%), and current mono-analyte blood markers including CgA, serotonin and pancreastatin, were considered inadequate (80%). However, it was positive for miRNA (67%), metabolomics (75%) and circulating mRNA (80%)[587]. Metabolomics investigations were considered of interest because functional and non-functional tumour are readily separated ($R^2=0.98$)[89]. For miRNA, the group agreed that the state of current technology meant that it could not support clinical usage. It was concluded from the meeting that a critical requirement was the development of a multi-analyte molecular tool that could identify disease status and define treatment responses. The use of circulating RNA as a biomarker, as in the NETest, was confirmed to be better than the effectiveness of standard mono-analyte biomarkers with potential for clinical applicability. Additionally, the overall consensus was that adjunct biomarker tools should be developed with the ability to provide synergistic information with imaging and facilitate the assessment of therapy. There was agreement that a better understanding of tumour biology would expedite the development of appropriate therapeutic biomarkers. At the Delphi meeting a consensus was reached in 89% of questions that genomics technology had significant potential to identify novel tissue biomarkers, and that insufficient specific mutations and treatment-targetable mutations had been identified. Thus, circulating DNA was not considered a viable option for the development of a biomarker. However since the Delphi consensus of 2016[587], cfDNA detection has been explored in the context of pNETs and si-NETs[607].

Other multiplex cancer tests have been developed including the Galleri blood test, which was developed by GRAIL and is based on analysis of panel of >100,000 cfDNA methylation regions[79]. Although a pre-specified set of 23 cancer types were examined in this study, it was unclear how many NETs were identified using the Galleri test. It would also be useful to know if the Galleri test can identify NETs at an early stage. A UK pilot randomised control trial began in 2021 and patients identified through this trial are being sent for further referral[80].The CancerSEEK test developed at Johns Hopkins University is smaller and aims to detect eight types of cancer by assessing the levels of nine circulating proteins and mutations in defined regions of 16 genes present in cfDNA[82]. However, despite good performance, limitations include the cost of the test (estimated to be $500) and that the cohort consisted only of patients with a cancer diagnosis based on the symptoms of disease. Hence, the cohort lacked

early-stage cancers in the truest sense. Advantages to the OLINK approach is that the protein assay technique has high sensitivity and specificity based on two specific antibodies and the PCR assay requires only a drop of dried blood on paper. Thus, this procedure allows easy collection and transport of samples. Cell free DNA (cfDNA) has also been assessed in patients with si-NETs and pNETs[607]. The study had a cohort of si-NETs (n=50) an pNETs (n=20). Plasma cfDNA were higher in both si-NETs and pNETs compared to previously established healthy controls (p<0.0001). There was no effect of PRRT treatment on cfDNA levels and no difference in cfDNA in patients with and without progressive disease after PRRT. However, they found that cfDNA levels were significantly higher in never smokers and previous smokers compared to in current smokers (p=0.026). Their findings suggests that cfDNA levels are not associated with the disease course in low grade NETs in contrast to other malignancies[607]. There is no gold standard method to assess cfDNA, with the lack of standardised protocols being one of the hurdles in hampering application of cfDNA analysis in routine clinical laboratories[608].

A recently published study aimed to develop a multiplex biomarker test for pNET detection utilising ML algorithms[588]. The EXPLAIN study published in May 2022 by the Nordic NET biomarker group analysed multiple plasma-protein biomarker levels using supervised ML to improve the diagnosis of pNETs and to differentiate pNETs from si-NETs[588]. The study had 29 pNET patients, 135 si-NET patients and 144 controls. Case inclusion criteria were patients who were >18 years old and had metastatic non-resectable NETs (WHO Grade 1 or 2, up to 20% Ki-67). The exclusion criteria for the study included patients with other malignant diseases, chronic inflammatory diseases, reduced kidney or liver function, and previous treatment with anti-proliferative treatments for NET disease (including somatostatin analogues, peptide receptor radionuclide therapy). Patients who had undergone primary tumour surgery but retained residual metastatic disease were included. Patients on PPIs were not excluded, nor was PPI treatment stopped before blood sampling.

The EXPLAIN study included a control group (n=144), age and sex matched with the patient population, and free of malignant disease, chronic inflammatory disease, several renal or hepatic dysfunction. To validate the algorithms a stratified 3FCV was performed for all the models. For this, the complete dataset was partitioned into equally (or near equally) sized folds or segments. This gave

an approximate split between training and validation sets of 80% and 20% respectively. A 3FCV approach was chosen to avoid having smaller groups of patients within the validation sets and carried out for all the models and classifiers. The study itself used supervised ML techniques including boosted decision trees, LDA and SVM for developing models to discriminate between pNETs and controls, and between pNETs and si-NETs. The SVM approach used a Linear SVM model with a C parameter of 1.

The EXPLAIN study used ELISA to assay CgA levels, which were alone able to identify pNETs with a sensitivity of 41%, a specificity of 94%, a PPV of 0.64, and a NPV of 0.84. In addition, a panel of 92 plasma proteins were assayed using the Proseek Oncology-II (OLink) panel, and the results used either with or without CgA to develop different ML models. They assessed each marker's contribution to fitting a model based on a likelihood ratio chi square proportion above zero in each of the three rounds of cross validation. In this way they identified the top biomarkers out of the 92 cancer-related proteins to discriminate pNETs from controls as well as pNETs from si-NETs. Boosted decision tree and SVM models classified pNETs and controls with similar accuracy values, with accuracy in ML meaning the number of correct predictions made by the model in relation to the total number of predictions that are made, and this was 0.91 and 0.94, respectively. These results were the same for the SVM approach whether or not CgA was included alongside the 92 plasma biomarkers. The accuracy of the CgA inclusive model alongside the 92 plasma biomarkers was better for the SVM model compared to the decision tree, thus CgA inclusion seemed to have a negative effect on the decision tree model and illustrates the variability that a particular marker can have between algorithm types. Corresponding LDA models constructed with the 92 plasma biomarkers had a worse predictive performance with accuracy values of 0.82 either with or without CgA. The top boosted decision tree with the ten identified top performing markers and CgA being excluded comprised a sensitivity of 84%, specificity of 98%, PPV of 0.92, NPV of 0.95, accuracy of 0.94 and a ROC curve AUC of 0.99. The top boosted decision tree with the top seven biomarkers identified from the 92 plasma biomarkers along with CgA comprised of a sensitivity of 74%, specificity of 98%, PPV of 0.91, NPV of 0.93, accuracy of 0.92 and AUC of 0.99. Thus, the top marker boosted decision tree performed

the best for the CgA excluded developed model. Importantly, the top markers boosted decision tree with CgA and without CgA utilised certain common markers between them but also had different markers. For example, both used cyclin-dependent kinase inhibitor 1 but the top marker boosted decision tree with CgA used methionine aminopeptidase 2 whereas the top marker boosted decision tree without CgA did not. Thus, a different overall combination of top markers were used for the CgA not included top biomarker boosted decision tree model and the CgA excluded top marker boosted decision tree model.

Moreover, the accuracy result obtained for the top biomarkers boosted decision tree created without CgA achieved an accuracy of 0.94 which was the same as seen for when all 92 biomarkers were included for SVM with or without CgA and better than that seen for the boosted decision tree including all 92 plasma biomarkers with or without CgA. Hence, illustrating that not all 92 biomarkers nor CgA were needed for the best accuracy, and that a top biomarkers boosted tree approach without CgA was suitable pNET vs control diagnosis.

The best performance for pNET versus si-NET classification was achieved using a SVM model, which achieved an accuracy of 0.91 when CgA was included, and 0.93 when CgA was excluded. Again, the inclusion of CgA had a negative impact for the SVM model, which illustrates that the marker was not one of the most important markers for this model, type and that other markers from the 92 additional markers assessed were of greater importance. Similar accuracies were obtained for boosted decision tree and LDA models whether or not CgA was included. Again, highlighting that CgA was not important for discriminating pNETs from Si-NETs. The boosted tree model using the top eight biomarkers and CgA developed to discriminate between pNETs and si-NETs had a sensitivity of 61%, specificity of 96%, PPV of 0.83, NPV of 0.90, accuracy of 0.88 and a ROC curve AUC of 0.98. The boosted tree model using the top biomarkers but not CgA had a sensitivity of 55%, specificity of 97%, accuracy of 0.88 and AUC of 0.97.

The EXPLAIN study echoes other studies about the limitations of CgA as a pNET marker, with only 57% of pNET patients in the present study having CgA levels >ULN and CgA itself not classified as one of the top ten biomarkers for the detection of pNETs. In terms of the role of CgA for discriminating between pNETs and si-NETs it was found to be the seventh most important marker in the

best models, and the authors found that other proteins such as carboxypeptidase E (CPE) and mothers against decapentaplegic homolog 5 (MAD homolog 5) were stronger contributors for detection of pNETs and discriminating pNETs from si-NETs. The top protein markers identified in the EXPLAIN study will undoubtedly be of considerable interest to future studies. Proteins identified as top biomarkers for pNET versus si-NET discrimination included CPE, MAD homolog 5, transmembrane glycoprotein NMB (GPNMB), tyrosine-protein kinase Lyn (LYN), interleukin-6 (IL-6), secreted protein acidic and cysteine rich (SPARC), glypican-1, S100A4 protein and granzyme B. Proteins identified as top biomarkers for pNETs vs control discrimination included MAD homolog 5, methionine aminopeptidase 2 (MetAP2), cyclin dependent kinase inhibitor 1 (CDKN1A), CPE, kallikrein-8 (KLK8), vimentin, LYN, integrin beta-5 (ITGB5), GDSL esterase/lipase, carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5), vimentin, TNF-related apoptosis-inducing ligand (TRAIL).

The proteins identified in the EXPLAIN study as top markers for identification of pNETs from controls and discrimination of pNETs and si-NETs, have wider roles in cancer and certain explored in the context of NETs such as CPE. Low expression of CDKN1A was found to be more common in SiNETs (62.8%) compared to pNETs (12%)[609]. TRAIL expression and it's receptors apart from osteoprotegerin was found frequently in PDAC and normal tissue whereas expression of osteoprotegerin was only detected in PDAC[610]. Moreover cancer cell death by TRAIL ranged from 37% to 77% in all the PDAC cell lines[610]. SPARC has been shown to promote proliferation and migration of PANC-1 and SW1990 cell lines invitro[611]. IL-6 is known to promote pancreatic cancer cell migration through activation of the small GTPase cell division cycle 42 in human PDAC cells[612] and sera levels of IL-6 are elevated in pancreatic cancer patients[613]. CPE is a metallo-carboxypeptidase involved in the biosynthesis of peptide hormones, and CPE is highly expressed and secreted from NETs[614]. Whilst MAD homolog 5 is a transcription regulatory protein involved in the signalling pathway by which TGF beta inhibits proliferation and Lyn is a tyrosine protein kinase belonging to a group of proteins targeted often for cancer treatment. Upregulation of KLK8 predicting poor prognosis in pancreatic cancer, with KLK8 overexpression having pro-proliferative and anti-apoptotic functions in pancreatic cancer cells through EGF-signalling dependent activation of the

PI3K/Akt/mTOR pathway[615]. Hence activation of the EGF signalling pathway through overexpression of KLK8 would be likely to have relevance to pNETs and further explored in the future. High vimentin expression with E-cadherin expression loss correlated with lymph node metastasis, distant metastasis, disease progression and poor prognosis in patients with G1 and G2 pNETs who underwent resection[616]. CEACAM5 expression was enriched in neuroendocrine prostate cancer and achaete-scute family BHLH Transcription Factor 1 (ASCL1) promotes neuroendocrine transdifferation of prostate cancer increasing chromatin accessibility of the core promoter of CEACAM5[617]. The mechanism of CEACAM5 regulation by ASCL1 may also be conserved in other types of NECs, however further studies exploring this are needed. MetAP2 has been explored in the context of prostate cancer, with MetAP2 being associated with shorter time to biological recurrence in prostate cancer that is treated surgically[618]. Integrin beta 5 is a prognostic biomarker in glioblastoma[619]. Glycoprotein NMB promotes tumour formation as well as progression of laryngeal squamous cell carcinoma[620]. S100A4 expression has been shown to be important for the invasiveness of pancreatic cancer, with 48 out of 83 pancreatic cancer patients positively expressing S100A4[621]. Thus, certain proteins identified in this study may have roles not specific for pNETs but in conditions that symptomatically can mimic pNETs such as PDAC, and thus the utility of such proteins to be able to distinguish pNETs from PDAC and other pancreatic benign conditions would need to be examined. In a study of the si-NET tumour microenvironment, in tumours several areas of T lymphocyte aggregation were observed which were not associated with any morphological features observable by H&E staining and did not express granzyme B to a larger extent compared to other T lymphocytes[622]. Expression of glypican 1 was found in normal tissues but the magnitude increased strongly in well differentiated NETs, however the expression of this protein dramatically decreased in high grade NETs with undetectable levels, hence the downregulation in high frae rumours suggested that glypican 1's role in cancer development is more complex and context dependent manner.[623]

When discriminating between pNET and si-NET populations, the results were more variable and had lower performance metrics compared to discriminating between pNETs and controls. This is likely due to pNET and Si-NET populations being harder to distinguish from each other due to similar

underlying NET biology in the two conditions compared to a healthy control vs NET, and any underlying differences between pNET and Si-NET would be more subtle in comparison to a healthy vs pNET comparison. Despite the SVM models having slightly better performance levels than boosted decision tree models in terms of AUC and accuracy, the latter were chosen to identify biomarkers with the highest contribution to the classification model. This was because one of the drawbacks of SVM is that it does not provide any explanation or justification for why the model changes as it learns, which limits the practical application of SVM models.

While the results indicate a good level of test accuracy, there are several important imitations to the EXPLAIN study. The first is the small size of the pNET group, when similarly large group sizes would be preferable to improve the power of the study. Second, it is not possible to assess at present how this test compares to the NETest, because this comparison has not been made. Third, the plasma proteins identified as the top markers used in this study could be elevated in other diseases and further studies are needed to explore whether this strategy can discriminate pNETs from other diseases including malignancies, benign diseases and other diseases of the pancreas. Fourth, the pNET cohort lacked early stage pNETs and most patients had a significant tumour burden that was easily detected on imaging; consequently, it is unclear how well the test would perform for patients with a small primary or with modest regional lymph node involvement. Fifth, external validation has not been carried out with an independent pNET cohort, which is a strength of the work carried out in this thesis. Other unknowns include how the test would perform when applied after presumed curative surgery in order to detect recurrence. Thus, how this model would work in a clinical routine setting requires further work. Due to patients with pNETs presenting with non-specific, vague symptoms, patients are facing delays to their diagnosis which impacts the stage at which diagnosis and in turn survival. Thus, it is important that pNETs are being diagnosed at an early stage. Point of care tests provide an opportunity for patients to be identified at an early stage, and the development of rapid non-invasive tests for accurate diagnosis of pNETs and a test that can discriminate pNETs from similar conditions is needed. The development of multi-analyte biomarker tests as an approach combined with ML algorithms presents promise for pNET diagnosis, however identification and in cooperation of confounding

factors is important for these tests in order for them to be accurate, as well as identifying early-stage markers which are likely to detect patients at an earlier stage. The NHS long term plan[624] aims for 75% of people with cancer to be diagnosed at an early stage (stage 1 and 2)[624]. One way this may be achieved is through the GRAIL (GALLERI) test which is now being trialled within the NHS[560], and if this is successful at identifying patients with cancer and identifying them at an early stage, could contribute to this NHS long term plan target being achieved. It is likely that pNETs will be identified through the NHS trial of the GALLERI test and thus, the number of pNETs identified from this trial will provide insight into the effectiveness of pNET identification through this route. Although due to pNETs being rare, the numbers that are identified may not be high. Additionally, the NHS faster diagnosis programme. which aims to have a 100% roll out of the non-specific pathway for diagnosis across the UK by 2024 provides an opportunity for identifying patients for whom a standard diagnosis pathway doesn't exists such as pNET patients. For pNET diagnosis, multiple avenues of tests have been explored including the NETest[625], but with CTCs there is no gold standard protocol, and there needs to be balance between expensive and complicated DNA tests with the potential of having a simple and affordable test for early referral in primary care. The QCancer tool developed by Professor Julia Hippsely Cox as of March 2021 is available for GPs in England and Wales[626]. Hence development of such tools is also needed alongside the development of biomarkers for pNET diagnosis. Hence despite biomarker development forming an important part in the effort for early pNET diagnosis, other areas such as patient awareness of possible pNET symptoms, appropriate pathways for referral and clinical decision tools to aid GP diagnosis will all need to be considered alongside biomarker development to maximise diagnosis of pNET patients at an early stage.

In conclusion, the work described in this thesis sought to find a pNET biomarker panel that would perform better than the current gold standard of CgA. Although this was achieved, particularly in the UOL training cohort across many marker combinations and ML model types, independent validation of the three markers that were taken forward (CgA, ANG2 and VGF) suggested that VGF was not a suitable pNET marker. This finding was particularly important because it illustrated the importance of the independent validation process as markers that show promise at training may not be as promising once externally

validated. Moreover, model performance for the addition of ANG2 to CgA didn't improve model performance seen in the external validation, suggesting that newer markers, alongside larger training and validation cohorts were needed. Additionally, greater information regarding the functionality of pNETs as well as a comprehensive understanding of confounding factors is required before a biomarker panel can be implemented. Further work is needed to develop a panel of markers that could discriminate pNETs, AP, CP and PDAC, with high sensitivity and specificity. ANG2 and CgA levels in the ADEPTs pNETs were found to be unexpectedly low, thus larger cohorts of pNET patients similar to those used in the RFH and UOL cohorts are needed for the development of ML models to discriminate pNETs from AP, CP and PDAC. Training model development should also focus on creating independent models which would then be evaluated on a validation cohort consisting of pNET, AP, CP and PDAC samples before the suitability of CgA and ANG2 for discrimination of pNETs and other pancreatic diseases can be decided. Further discovery of new pNET markers is needed as these additional markers could enhance current ML diagnosis models. Finally, biomarkers for identifying early-stage disease based on new experimental models that better recapitulate pNET early-stage tumours are likely to have a great impact in terms of survival of pNET patients. This targets an unmet clinical need which future work should aim to primarily address.

References

1. Yao, J. C. *et al.* One Hundred Years After "Carcinoid": Epidemiology of and Prognostic Factors for Neuroendocrine Tumors in 35,825 Cases in the United States. *Journal of Clinical Oncology* **26**, 3063–3072 (2008).

2. de Herder, W. W., Rehfeld, J. F., Kidd, M. & Modlin, I. M. A short history of neuroendocrine tumours and their peptide hormones. *Best Practice & Research Clinical Endocrinology & Metabolism* **30**, 3–17 (2016).

3. Tsoucalas, G., Karamanou, M. & Androutsos, G. The eminent German pathologist Siegfried Oberndorfer (1876-1944) and his landmark work on carcinoid tumors. 3.

4. Kaltsas, G. A., Besser, G. M. & Grossman, A. B. The Diagnosis and Medical Management of Advanced Neuroendocrine Tumors. *Endocrine Reviews* **25**, 458–511 (2004).

5. Oronsky, B., Ma, P. C., Morgensztern, D. & Carter, C. A. Nothing But NET: A Review of Neuroendocrine Tumors and Carcinomas. *Neoplasia* **19**, 991–1002 (2017).

6. Genus, T. S. E. *et al.* Impact of neuroendocrine morphology on cancer outcomes and stage at diagnosis: a UK nationwide cohort study 2013–2015. *Br J Cancer* **121**, 966–972 (2019).

7. Das, S. & Dasari, A. Epidemiology, Incidence, and Prevalence of Neuroendocrine Neoplasms: Are There Global Differences? *Curr Oncol Rep* **23**, 43 (2021).

8. Sonbol, M. B. *et al.* Survival and Incidence Patterns of Pancreatic Neuroendocrine Tumors Over the Last 2 Decades: A SEER Database Analysis. *The Oncologist* **27**, 573–578 (2022).

9. Ito, T., Igarashi, H. & Jensen, R. T. Pancreatic neuroendocrine tumors: Clinical features, diagnosis and medical treatment: Advances. *Best Practice & Research Clinical Gastroenterology* **26**, 737–753 (2012).

10. Yang, G. *et al.* Surgery management for sporadic small (≤2 cm), non-functioning pancreatic neuroendocrine tumors: A consensus statement by the Chinese Study Group for Neuroendocrine Tumors (CSNET). *International Journal of Oncology* **50**, 567–574 (2017).

11. Zheng, K. *et al.* Mutational landscape and potential therapeutic targets for sporadic pancreatic neuroendocrine tumors based on target next-generation sequencing. *Exp Ther Med* **21**, 415 (2021).

12. Kanthan, R., Senger, J.-L., Ahmed, S. & Kanthan, S. C. Pancreatic Neuroendocrine Tumors in the 21<sup>st</sup> Century—An Update. *Journal of Cancer Therapy* **08**, 1194–1233 (2017).

13. Falconi, M. *et al.* ENETS Consensus Guidelines Update for the Management of Patients with Functional Pancreatic Neuroendocrine Tumors and Non-Functional Pancreatic Neuroendocrine Tumors. *Neuroendocrinology* **103**, 153–171 (2016).

14. Pea, A., Hruban, R. H. & Wood, L. D. Genetics of pancreatic neuroendocrine tumors: implications for the clinic. *Expert Review of Gastroenterology & Hepatology* **9**, 1407–1419 (2015).

15. Mete, O. & Chetty, R. Pancreatic endocrine neoplasia: familial syndromes. *Diagnostic Histopathology* **23**, 378–385 (2017).

16. Thakker, R. V. Multiple endocrine neoplasia type 1 (MEN1) and type 4 (MEN4). *Molecular and Cellular Endocrinology* **386**, 2–15 (2014).

17. Kamilaris, C. D. C. & Stratakis, C. A. Multiple Endocrine Neoplasia Type 1 (MEN1): An Update and the Significance of Early Genetic and Clinical Diagnosis. *Frontiers in Endocrinology* **10**, (2019).

18. Frost, M., Lines, K. E. & Thakker, R. V. Current and emerging therapies for PNETs in patients with or without MEN1. *Nature Reviews Endocrinology* **14**, 216–227 (2018).

19. Marx, S. Multiple Endocrine Neoplasia Type 1: Clinical and Genetic Topics. *Annals of Internal Medicine* **129**, 484 (1998).

20. Couvelard, A., Cazes, A. & Cros, J. Updates in histopathological classification and tissue biomarkers of digestive neuroendocrine neoplasms: What the clinician should know. *Best Practice & Research Clinical Endocrinology & Metabolism* **37**, 101795 (2023).

21. Rindi, G. *et al.* Overview of the 2022 WHO Classification of Neuroendocrine Neoplasms. *Endocr Pathol* **33**, 115–154 (2022).

22. Nagtegaal, I. D. *et al.* The 2019 WHO classification of tumours of the digestive system. *Histopathology* **76**, 182–188 (2020).

23. Panzuto, F. *et al.* Metastatic and Locally Advanced Pancreatic Endocrine Carcinomas: Analysis of Factors Associated With Disease Progression. *JCO* **29**, 2372–2377 (2011).

24. Shiba, S. *et al.* Pancreatic neuroendocrine tumors: A single-center 20-year experience with 100 patients. *Pancreatology* **16**, 99–105 (2016).

25. Ma, Z.-Y. *et al.* Pancreatic neuroendocrine tumors: A review of serum biomarkers, staging, and management. *World J Gastroenterol* **26**, 2305–2322 (2020).

26. Edge, S. B. & Compton, C. C. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* **17**, 1471–1474 (2010).

27. Partelli, S. *et al.* Management of Asymptomatic Sporadic Nonfunctioning Pancreatic Neuroendocrine Neoplasms (ASPEN) ≤2 cm: Study Protocol for a Prospective Observational Study. *Front Med (Lausanne)* **7**, 598438 (2020).

28. Partelli, S. *et al.* Systematic review of active surveillance *versus* surgical management of asymptomatic small non-functioning pancreatic neuroendocrine neoplasms: Surveillance *versus* surgery for asymptomatic small non-functioning pancreatic neuroendocrine neoplasms. *Br J Surg* **104**, 34–41 (2017).

29. Belotto, M., Crouzillard, B. do N. S., Araujo, K. de O. & Peixoto, R. D. PANCREATIC NEUROENDOCRINE TUMORS: SURGICAL RESECTION. *ABCD. Arquivos Brasileiros de Cirurgia Digestiva (São Paulo)* **32**, (2019).

30. Oleinikov, K. *et al.* Endoscopic Ultrasound-Guided Radiofrequency Ablation: A New Therapeutic Approach for Pancreatic Neuroendocrine Tumors. *The Journal of Clinical Endocrinology & Metabolism* **104**, 2637–2647 (2019).

31. Alberti, P., Martin, D., Gemenetzis, G. & Parks, R. Surgical management of pancreatic neuroendocrine neoplasms. *Laparoscopic, Endoscopic and Robotic Surgery* **6**, 83–90 (2023).

32. Dumlu, E. G., Karakoç, D. & Özdemir, A. Nonfunctional Pancreatic Neuroendocrine Tumors: Advances in Diagnosis, Management, and Controversies. *International Surgery* **100**, 1089–1097 (2015).

33. Minter, R. M. & Simeone, D. M. Contemporary Management of Nonfunctioning Pancreatic Neuroendocrine Tumors. *Journal of Gastrointestinal Surgery* **16**, 435–446 (2012).

34. Falconi, M. *et al.* ENETS Consensus Guidelines for the Management of Patients with Digestive Neuroendocrine Neoplasms of the Digestive System: Well-

Differentiated Pancreatic Non-Functioning Tumors. *Neuroendocrinology* **95**, 120–

134 (2012).

35. Nigri, G. *et al.* Treatment options for PNET liver metastases: a systematic review.

World Journal of Surgical Oncology **16**, (2018).

36. Merola, E. *et al.* Somatostatin Analogs for Pancreatic Neuroendocrine Tumors: Any

Benefit When Ki-67 Is ≥10%? *Oncologist* **26**, 294–301 (2021).

37. Wiedenmann, B., Pavel, M. & Kos-Kudla, B. From Targets to Treatments: A Review

of Molecular Targets in Pancreatic Neuroendocrine Tumors. *Neuroendocrinology*

94, 177–190 (2011).

38. Zandee, W. T. *et al.* Symptomatic and Radiological Response to 177Lu-DOTATATE

for the Treatment of Functioning Pancreatic Neuroendocrine Tumors. *J Clin*

Endocrinol Metab **104**, 1336–1344 (2019).

39. Raymond, E. *et al.* Sunitinib malate for the treatment of pancreatic

neuroendocrine tumors. *N Engl J Med* **364**, 501–513 (2011).

40. Ducreux, M. *et al.* Bevacizumab combined with 5-FU/streptozocin in patients with

progressive metastatic well-differentiated pancreatic endocrine tumours (BETTER

trial) – A phase II non-randomised trial. *European Journal of Cancer* **50**, 3098–

3106 (2014).

41. Yao, J. C. *et al.* Efficacy of RAD001 (Everolimus) and Octreotide LAR in Advanced

Low- to Intermediate-Grade Neuroendocrine Tumors: Results of a Phase II Study.

Journal of Clinical Oncology **26**, 4311–4318 (2008).

42. Lombard-Bohas, C. *et al.* Impact of Prior Chemotherapy Use on the Efficacy of

Everolimus in Patients With Advanced Pancreatic Neuroendocrine Tumors: A

Subgroup Analysis of the Phase III RADIANT-3 Trial. *Pancreas* **44**, 181–189 (2015).

43. Krug, S., Gress, T. M., Michl, P. & Rinke, A. The Role of Cytotoxic Chemotherapy in Advanced Pancreatic Neuroendocrine Tumors. *Digestion* **96**, 67–75 (2017).

44. Uri, I. & Grozinsky-Glasberg, S. Current treatment strategies for patients with advanced gastroenteropancreatic neuroendocrine tumors (GEP-NETs). *Clinical Diabetes and Endocrinology* **4**, (2018).

45. Aoki, T. *et al.* Streptozocin chemotherapy for advanced/metastatic well-differentiated neuroendocrine tumors: an analysis of a multi-center survey in Japan. *Journal of Gastroenterology* **50**, 769–775 (2015).

46. Zappa, M. *et al.* Liver-directed therapies in liver metastases from neuroendocrine tumors of the gastrointestinal tract. *Target Oncol* **7**, 107–116 (2012).

47. Kennedy, A. S. *et al.* Radioembolization for unresectable neuroendocrine hepatic metastases using resin 90Y-microspheres: early results in 148 patients. *Am J Clin Oncol* **31**, 271–279 (2008).

48. Klöppel, G. Tumour biology and histopathology of neuroendocrine tumours. *Best Practice & Research Clinical Endocrinology & Metabolism* **21**, 15–31 (2007).

49. Kasajima, A., Yazdani, S. & Sasano, H. Pathology diagnosis of pancreatic neuroendocrine tumors: ▪▪▪. *J Hepatobiliary Pancreat Sci* **22**, 586–593 (2015).

50. Antwi, K., Nicolas, G., Wild, D. & Christ, E. Molecular imaging for neuroendocrine tumours. *Swiss Med Wkly* (2019) doi:10.4414/smw.2019.20017.

51. Dromain, C. *et al.* Imaging of neuroendocrine tumors of the pancreas. *Diagnostic and Interventional Imaging* **97**, 1241–1257 (2016).

52. Kulaksiz, H. *et al.* Identification of somatostatin receptor subtypes 1, 2A, 3, and 5 in neuroendocrine tumours with subtype specific antibodies. *Gut* **50**, 52–60 (2002).

53. Kartalis, N., Mucelli, R. M. P. & Sundin, A. Recent developments in imaging of pancreatic neuroendocrine tumors. *Ann Gastroenterol* **28**, 193–202 (2015).

54. Guo, C. *et al.* The differentiation of pancreatic neuroendocrine carcinoma from pancreatic ductal adenocarcinoma: the values of CT imaging features and texture analysis. *Cancer Imaging* **18**, 37 (2018).

55. Bhosale, P. R. *et al.* Vascular pancreatic lesions: spectrum of imaging findings of malignant masses and mimics with pathologic correlation. *Abdom Imaging* **38**, 802–817 (2013).

56. Fidler, J. L. *et al.* Preoperative Detection of Pancreatic Insulinomas on Multiphasic Helical CT. *American Journal of Roentgenology* **181**, 775–780 (2003).

57. Kim, M. K. Endoscopic ultrasound in gastroenteropancreatic neuroendocrine tumors. *Gut Liver* **6**, 405–410 (2012).

58. Passaro, Jr, E., Howard, T. J., Sawicki, M. P., Watt, P. C. & Stabile, B. E. The Origin of Sporadic Gastrinomas Within the Gastrinoma Triangle: A Theory. *Arch Surg* **133**, (1998).

59. Gauger, P. G. *et al.* Role of endoscopic ultrasonography in screening and treatment of pancreatic endocrine tumours in asymptomatic patients with multiple endocrine neoplasia type 1. *Br J Surg* **90**, 748–754 (2003).

60. Basuroy, R., Bouvier, C., Ramage, J. K., Sissons, M. & Srirajaskanthan, R. Delays and routes to diagnosis of neuroendocrine tumours. *BMC Cancer* **18**, 1122 (2018).

61. Thakker, R. V. *et al.* Clinical Practice Guidelines for Multiple Endocrine Neoplasia Type 1 (MEN1). *The Journal of Clinical Endocrinology & Metabolism* **97**, 2990–3011 (2012).

62. Pancreatic cancer - NICE Pathways. https://pathways.nice.org.uk/pathways/pancreatic-cancer#content=view-node%3Anodes-diagnosis.

63. Chapman, D. *et al.* First results from five multidisciplinary diagnostic centre (MDC) projects for non-specific but concerning symptoms, possibly indicative of cancer. *Br J Cancer* **123**, 722–729 (2020).

64. Vedsted, P. & Olesen, F. A differentiated approach to referrals from general practice to support early cancer diagnosis – the Danish three-legged strategy. *Br J Cancer* **112**, S65–S69 (2015).

65. Basuroy, R. *et al.* Presenting Symptoms and Delay in Diagnosis of Gastrointestinal and Pancreatic Neuroendocrine Tumours. *Neuroendocrinology* **107**, 42–49 (2018).

66. Liao, W. *et al.* Identifying symptoms associated with diagnosis of pancreatic exocrine and neuroendocrine neoplasms: a nested case-control study of the UK primary care population. *Br J Gen Pract* **71**, e836–e845 (2021).

67. Halfdanarson, T. R. *et al.* Risk factors for pancreatic neuroendocrine tumors: a clinic-based case-control study. *Pancreas* **43**, 1219–1222 (2014).

68. Capurso, G. *et al.* Risk Factors for Sporadic Pancreatic Endocrine Tumors: A Case-Control Study of Prospectively Evaluated Patients. *American Journal of Gastroenterology* **104**, 3034–3041 (2009).

69. Giraldi, L. *et al.* Risk factors for pancreas and lung neuroendocrine neoplasms: a case–control study. *Endocrine* **71**, 233–241 (2021).

70. Hassan, M. M. *et al.* Risk factors associated with neuroendocrine tumors: A U.S.-based case-control study: Risk Factors Associated with Neuroendocrine Tumors: A U.S.-Based Case-Control Study. *Int. J. Cancer* **123**, 867–873 (2008).

71. Ben, Q. *et al.* Risk Factors for Sporadic Pancreatic Neuroendocrine Tumors: A Case-Control Study. *Sci Rep* **6**, 36073 (2016).

72. Choi, J. H. *et al.* The association between use of statin or aspirin and pancreatic ductal adenocarcinoma: A nested case-control study in a Korean nationwide cohort. *Cancer Med* **8**, 7419–7430 (2019).

73. Vitali, E. *et al.* A novel insight into the anticancer mechanism of metformin in pancreatic neuroendocrine tumor cells. *Molecular and Cellular Endocrinology* **509**, 110803 (2020).

74. Wang, L. *et al.* Insulin therapy contributes to the increased risk of colorectal cancer in diabetes patients: a meta-analysis. *Diagn Pathol* **8**, 180 (2013).

75. Wiedenmann, B., Franke, W. W., Kuhn, C., Moll, R. & Gould, V. E. Synaptophysin: a marker protein for neuroendocrine cells and neoplasms. *Proceedings of the National Academy of Sciences* **83**, 3500–3504 (1986).

76. Eriksson, B. *et al.* Chromogranins — New Sensitive Markers for Neuroendocrine Tumors. *Acta Oncologica* **28**, 325–329 (1989).

77. Hofland, J., Zandee, W. T. & de Herder, W. W. Role of biomarker tests for diagnosis of neuroendocrine tumours. *Nature Reviews Endocrinology* **14**, 656–669 (2018).

78. Lawrence, B. *et al.* The Epidemiology of Gastroenteropancreatic Neuroendocrine Tumors. *Endocrinology and Metabolism Clinics of North America* **40**, 1–18 (2011).

79. Liu, M. C. *et al.* Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology* **31**, 745–759 (2020).

80. NHS to pilot blood test that could detect over 50 different cancer types. *Cancer Research UK - Cancer News* https://news.cancerresearchuk.org/2020/11/27/nhs-to-pilot-blood-test-that-could-detect-over-50-different-cancer-types/ (2020).

81. GRAIL and University of Oxford to Present Results From First Prospective Study of Multi-Cancer Early Detection in a Symptomatic Patient Population at 2023 ASCO Annual Meeting. *GRAIL* https://grail.com/press-releases/grail-and-university-of-

oxford-to-present-results-from-first-prospective-study-of-multi-cancer-early-detection-in-a-symptomatic-patient-population-at-2023-asco-annual-meeting/.

82. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).

83. Modlin, I. M. *et al.* A multianalyte PCR blood test outperforms single analyte ELISAs (chromogranin A, pancreastatin, neurokinin A) for neuroendocrine tumor detection. *Endocrine-Related Cancer* **21**, 615–628 (2014).

84. Guo, X., Gao, S., Li, Z. & Hao, J. The role of biomarker in pancreatic neuroendocrine tumor: a narrative review. *Journal of Pancreatology* **Publish Ahead of Print**, (2021).

85. Thorns, C. *et al.* Global microRNA profiling of pancreatic neuroendocrine neoplasias. *Anticancer Res* **34**, 2249–2254 (2014).

86. Redova, M., Sana, J. & Slaby, O. Circulating miRNAs as new blood-based biomarkers for solid cancers. *Future Oncol* **9**, 387–402 (2013).

87. Walter, B. A., Valera, V. A., Pinto, P. A. & Merino, M. J. Comprehensive microRNA Profiling of Prostate Cancer. *J Cancer* **4**, 350–357 (2013).

88. Li, A. *et al.* MicroRNA Array Analysis Finds Elevated Serum miR-1290 Accurately Distinguishes Patients with Low-Stage Pancreatic Cancer from Healthy and Disease Controls. *Clinical Cancer Research* 12.

89. Kinross, J. M., Drymousis, P., Jiménez, B. & Frilling, A. Metabonomic profiling: A novel approach in neuroendocrine neoplasias. *Surgery* **154**, 1185–1193 (2013).

90. Qiao, X.-W. *et al.* Chromogranin A is a reliable serum diagnostic biomarker for pancreatic neuroendocrine tumors but not for insulinomas. *BMC Endocrine Disorders* **14**, (2014).

91. Bartolomucci, A. *et al.* The extended granin family: structure, function, and biomedical implications. *Endocr Rev* **32**, 755–797 (2011).

92. D'amico, M. A., Ghinassi, B., Izzicupo, P., Manzoli, L. & Di Baldassarre, A. Biological function and clinical relevance of chromogranin A and derived peptides. *Endocrine Connections* **3**, R45–R54 (2014).

93. Modlin, I. M. *et al.* Chromogranin A—Biological Function and Clinical Utility in Neuro Endocrine Tumor Disease. *Annals of Surgical Oncology* **17**, 2427–2443 (2010).

94. Banks, P. & Helle, K. The Release of Protein from the Stimulated Adrenal Medulla. *Biochemical Journal* **97**, 40C-41C (1965).

95. Smith, A. & Winkler, H. Purification and properties of an acidic protein from chromaffin granules of bovine adrenal medulla. *Biochemical Journal* **103**, 483–492 (1967).

96. Fischer-Colbrie, R. & Frischenschlager, I. Immunological Characterization of Secretory Proteins of Chromaffin Granules: Chromogranins A, Chromogranins B, and Enkephalin-Containing Peptides. *J Neurochem* **44**, 1854–1861 (1985).

97. Dam, G. *et al.* Prospective Study of Chromogranin A as a Predictor of Progression in Patients with Pancreatic, Small-Intestinal, and Unknown Primary Neuroendocrine Tumors. *Neuroendocrinology* **110**, 217–224 (2020).

98. Yao, J. C. *et al.* Chromogranin A and Neuron-Specific Enolase as Prognostic Markers in Patients with Advanced pNET Treated with Everolimus. *The Journal of Clinical Endocrinology & Metabolism* **96**, 3741–3749 (2011).

99. Rossi, R. E. *et al.* Chromogranin A as a predictor of radiological disease progression in neuroendocrine tumours. *Annals of Translational Medicine* **3**, 7 (2015).

100. Nobels, F. R. E. Chromogranin A as Serum Marker for Neuroendocrine Neoplasia: Comparison with Neuron-Specific Enolase and the -Subunit of Glycoprotein Hormones. *Journal of Clinical Endocrinology & Metabolism* **82**, 2622–2628 (1997).

101. Zhao, Y. *et al.* Surgical management of patients with insulinomas: Result of 292 cases in a single institution: Surgical Management of Insulinoma. *J. Surg. Oncol.* **103**, 169–174 (2011).

102. Tomassetti, P. *et al.* Diagnostic value of plasma chromogranin A in neuroendocrine tumours: *European Journal of Gastroenterology & Hepatology* **13**, 55–58 (2001).

103. Song, G. *et al.* TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *Journal of Experimental & Clinical Cancer Research* **35**, (2016).

104. Batra, J. *et al.* Matrix Metalloproteinase-10 (MMP-10) Interaction with Tissue Inhibitors of Metalloproteinases TIMP-1 and TIMP-2: BINDING STUDIES AND CRYSTAL STRUCTURE. *Journal of Biological Chemistry* **287**, 15935–15946 (2012).

105. Nalluri, S. *et al.* TIMP-1 Inhibits Apoptosis in Lung Adenocarcinoma Cells via Interaction with Bcl-2. *PLOS ONE* **10**, e0137673 (2015).

106. Li, G., Fridman, R. & Kim, H. R. Tissue inhibitor of metalloproteinase-1 inhibits apoptosis of human breast epithelial cells. *Cancer Res* **59**, 6267–6275 (1999).

107. Srirajaskanthan, R. *et al.* Identification of Mac-2-binding Protein as a Putative Marker of Neuroendocrine Tumors from the Analysis of Cell Line Secretomes. *Molecular & Cellular Proteomics* **9**, 656–666 (2010).

108. Ries, C. Cytokine functions of TIMP-1. *Cell. Mol. Life Sci.* **71**, 659–672 (2014).

109. Kopitz, C. *et al.* Tissue Inhibitor of Metalloproteinases-1 Promotes Liver Metastasis by Induction of Hepatocyte Growth Factor Signaling. *Cancer Research* **67**, 8615–8623 (2007).

110. Han, X., Sun, Y., Scott, S. & Bleich, D. Tissue Inhibitor of Metalloproteinase-1 Prevents Cytokine-Mediated Dysfunction and Cytotoxicity in Pancreatic Islets and -cells. *Diabetes* **50**, 1047–1055 (2001).

111. Roy, R. *et al.* Urinary TIMP-1 and MMP-2 levels detect the presence of pancreatic malignancies. *British Journal of Cancer* **111**, 1772–1779 (2014).

112. Holten-Andersen, L. *et al.* Saliva and plasma TIMP-1 in patients with colorectal cancer: a prospective study. *Scand J Gastroenterol* **47**, 1234–1241 (2012).

113. Slater, E. P. *et al.* LCN2 and TIMP1 as Potential Serum Markers for the Early Detection of Familial Pancreatic Cancer. *Translational Oncology* **6**, 99–103 (2013).

114. Bartsch, D. K. *et al.* Prevalence of familial pancreatic cancer in Germany: Familial Pancreatic Cancer in Germany. *International Journal of Cancer* **110**, 902–906 (2004).

115. KothsS, K., Halenbeck, R. & CasipitQ, C. Cloning and Characterization oaf Human Mac-2-bindingProtein, a New Member of the Superfamily Defined by the Macrophage Scavenger Receptor Cysteine-rich Domain. 5.

116. Sasaki, T. Mac-2 binding protein is a cell-adhesive protein of the extracellular matrix which self-assembles into ring-like structures and binds beta 1 integrins, collagens and fibronectin. *The EMBO Journal* **17**, 1606–1613 (1998).

117. Shirure, V. S., Reynolds, N. M. & Burdick, M. M. Mac-2 Binding Protein Is a Novel E-Selectin Ligand Expressed by Breast Cancer Cells. *PLoS ONE* **7**, e44529 (2012).

118. Cobanoglu, U. *et al.* Are Serum Mac 2-Binding Protein Levels Elevated in Esophageal Cancer? A Control Study of Esophageal Squamous Cell Carcinoma Patients. *Disease Markers* **2018**, 1–4 (2018).

119. Wu, C.-C. *et al.* Cancer cell-secreted proteomes as a basis for searching potential tumor markers: Nasopharyngeal carcinoma as a model. *PROTEOMICS* **5**, 3173–3182 (2005).

120. Stampolidis, P., Ullrich, A. & Iacobelli, S. LGALS3BP, lectin galactoside-binding soluble 3 binding protein, promotes oncogenic cellular events impeded by antibody intervention. *Oncogene* **34**, 39–52 (2015).

121. Tinari, N. *et al.* High expression of 90K (Mac-2 BP) is associated with poor survival in node-negative breast cancer patients not receiving adjuvant systemic therapies. *International Journal of Cancer* **124**, 333–338 (2009).

122. Capurso, G. *et al.* Molecular pathology and genetics of pancreatic endocrine tumours. *Journal of Molecular Endocrinology* **49**, R37–R50 (2012).

123. Rumilla, K. M., Erickson, L. A., Erickson, A. K. & Lloyd, R. V. Galectin-4 Expression in Carcinoid Tumors. *Endocrine Pathology* **17**, 243–250 (2006).

124. Naba, A., Clauser, K. R., Mani, D. R., Carr, S. A. & Hynes, R. O. Quantitative proteomic profiling of the extracellular matrix of pancreatic islets during the angiogenic switch and insulinoma progression. *Sci Rep* **7**, 40495 (2017).

125. Vizin, T. & Kos, J. Gamma-enolase: a well-known tumour marker, with a less-known role in cancer. *Radiology and Oncology* **49**, 217–226 (2015).

126. Jones, W. & Bianchi, K. Aerobic Glycolysis: Beyond Proliferation. *Frontiers in Immunology* **6**, (2015).

127. Vinores, S. A., Herman, M. M. & Rubinstein, L. J. Electron-immunocytochemical localization of neuron-specific enolase in cytoplasm and on membranes of

primary and metastatic cerebral tumours and on glial filaments of glioma cells. *Histopathology* **10**, 891–908 (2007).

128. Haque, A., Polcyn, R., Matzelle, D. & Banik, N. L. New Insights into the Role of Neuron-Specific Enolase in Neuro-Inflammation, Neurodegeneration, and Neuroprotection. *Brain Sciences* **8**, 33 (2018).

129. Haque, A., Ray, S. K., Cox, A. & Banik, N. L. Neuron specific enolase: a promising therapeutic target in acute spinal cord injury. *Metabolic Brain Disease* **31**, 487–495 (2016).

130. Hafner, A., Obermajer, N. & Kos, J. γ-Enolase C-terminal peptide promotes cell survival and neurite outgrowth by activation of the PI3K/Akt and MAPK/ERK signalling pathways. *Biochemical Journal* **443**, 439–450 (2012).

131. Yan, T. *et al.* Neuronal markers are expressed in human gliomas and NSE knockdown sensitizes glioblastoma cells to radiotherapy and temozolomide. *BMC Cancer* **11**, (2011).

132. Hafner, A., Obermajer, N. & Kos, J. Gamma-1-Syntrophin Mediates Trafficking of Gamma-Enolase towards the Plasma Membrane and Enhances Its Neurotrophic Activity. *Neurosignals* **18**, 246–258 (2010).

133. Zhu, M. *et al.* NSE from diffuse large B-cell lymphoma cells regulates macrophage polarization. *Cancer Management and Research* **Volume 11**, 4577–4595 (2019).

134. Dal Bello, M. G. *et al.* The role of CEA, CYFRA21-1 and NSE in monitoring tumor response to Nivolumab in advanced non-small cell lung cancer (NSCLC) patients. *Journal of Translational Medicine* **17**, (2019).

135. Wang, X. *et al.* Systematic Correlation Analyses of Circulating Tumor Cells with Clinical Variables and Tumor Markers in Lung Cancer Patients. *Journal of Cancer* **8**, 3099–3104 (2017).

136. Ishibashi, N., Maebayashi, T., Aizawa, T., Sakaguchi, M. & Okada, M. Serum tumor marker levels at the development of intracranial metastasis in patients with lung or breast cancer. *Journal of Thoracic Disease* **11**, 1765–1771 (2019).

137. Nakamura, Y. *et al.* Changes to levels of serum neuron-specific enolase in a patient with small cell carcinoma of the pancreas. *Journal of Hepato-Biliary-Pancreatic Surgery* **12**, 93–98 (2005).

138. Srirajaskanthan, R. *et al.* Circulating angiopoietin-2 is elevated in patients with neuroendocrine tumours and correlates with disease burden and prognosis. *Endocrine-Related Cancer* **16**, 967–976 (2009).

139. Akwii, R. G., Sajib, M. S., Zahra, F. T. & Mikelis, C. M. Role of Angiopoietin-2 in Vascular Physiology and Pathophysiology. *Cells* **8**, 471 (2019).

140. Melen-Mucha, G. *et al.* Elevated Peripheral Blood Plasma Concentrations of Tie-2 and Angiopoietin 2 in Patients with Neuroendocrine Tumors. *International Journal of Molecular Sciences* **13**, 1444–1460 (2012).

141. Tanaka, S. *et al.* Biologic significance of angiopoietin-2 expression in human hepatocellular carcinoma. *Journal of Clinical Investigation* **103**, 341–345 (1999).

142. Oladejo, A. O. GASTROENTEROPANCREATIC NEUROENDOCRINE TUMORS (GEP-NETs) - APPROACH TO DIAGNOSIS AND MANAGEMENT. **7**, 5 (2009).

143. Gao, X. *et al.* Thymosin Beta-4 Induces Mouse Hair Growth. *PLOS ONE* **10**, e0130040 (2015).

144. Hannappel, E. Thymosin β4 and its posttranslational modifications: Thymosin β4 and its posttranslational modifications. *Annals of the New York Academy of Sciences* **1194**, 27–35 (2010).

145. Kaur, H. & Mutus, B. Platelet function and thymosin β4. *Biological Chemistry* **393**, (2012).

146. Theunissen, W. *et al.* Thymosin beta 4 and thymosin beta 10 expression in hepatocellular carcinoma. *Eur J Histochem* **58**, (2014).

147. Goldstein, A., Hannappel, E. & Kleinman, H. Thymosin β: actin-sequestering protein moonlights to repair injured tissues. *Trends in Molecular Medicine* **11**, 421–429 (2005).

148. Crockford, D., Turjman, N., Allan, C. & Angel, J. Thymosin β4: structure, function, and biological properties supporting current and future clinical applications: Scientific rationale and use of thymosin β4 in man. *Annals of the New York Academy of Sciences* **1194**, 179–189 (2010).

149. Yamaguchi, H. & Condeelis, J. Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1773**, 642–652 (2007).

150. Freeman, K. W., Bowman, B. R. & Zetter, B. R. Regenerative protein thymosin β-4 is a novel regulator of purinergic signaling. *The FASEB Journal* **25**, 907–915 (2011).

151. Kaur, H., Heeney, R., Carriveau, R., Sosne, G. & Mutus, B. Whole blood, flow-chamber studies in real-time indicate a biphasic role for thymosin β-4 in platelet adhesion. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1800**, 1256–1261 (2010).

152. Oh, J.-M. *et al.* Hypoxia-inducible transcription factor (HIF)-1α stabilization by actin-sequestering protein, thymosin beta-4 (TB4) in Hela cervical tumor cells. *Cancer Letters* **264**, 29–35 (2008).

153. Huang, D., Wang, S., Wang, A., Chen, X. & Zhang, H. Thymosin beta 4 silencing suppresses proliferation and invasion of non-small cell lung cancer cells by repressing Notch1 activation. *Acta Biochimica et Biophysica Sinica* **48**, 788–794 (2016).

154. Cha, H.-J. Role of Thymosin 4 in Tumor Metastasis and Angiogenesis. *CancerSpectrum Knowledge Environment* **95**, 1674–1680 (2003).

155. Bock-Marquette, I., Saxena, A., White, M. D., DiMaio, J. M. & Srivastava, D. Thymosin b4 activates integrin-linked kinase and promotes cardiac cell migration, survival and cardiac repair. **432**, 7 (2004).

156. Wei, C., Kumar, S., Kim, I.-K. & Gupta, S. Thymosin Beta 4 Protects Cardiomyocytes from Oxidative Stress by Targeting Anti-Oxidative Enzymes and Anti-Apoptotic Genes. *PLoS ONE* **7**, e42586 (2012).

157. Sosne, G., Qiu, P. & Kurpakus, M. Thymosin beta 4: A novel corneal wound healing and anti-inflammatory agent. *Clinical Ophthalmology* 8.

158. Wang, L. *et al.* Angiopoietin-1/Tie2 signaling pathway contributes to the therapeutic effect of thymosin β4 on diabetic peripheral neuropathy. *Neuroscience Research* (2018) doi:10.1016/j.neures.2018.10.005.

159. Young, J. D. *et al.* Thymosin β 4 sulfoxide is an anti-inflammatory agent generated by monocytes in the presence of glucocorticoids. *Nature Medicine* **5**, 1424–1427 (1999).

160. Salton, S. R. J. *et al.* VGF: A Novel Role for This Neuronal and Neuroendocrine Polypeptide in the Regulation of Energy Balance. *Frontiers in Neuroendocrinology* **21**, 199–219 (2000).

161. Zhao, Z. *et al.* Vgf is a novel biomarker associated with muscle weakness in amyotrophic lateral sclerosis (ALS), with a potential role in disease pathogenesis. *Int J Med Sci* **5**, 92–99 (2008).

162. Levi, A., Eldridge, J. D. & Paterson, B. M. Molecular cloning of a gene sequence regulated by nerve growth factor. *Science* **229**, 393–395 (1985).

163. Noda, Y. *et al.* VGF and striatal cell damage in in vitro and in vivo models of Huntington's disease. *Pharmacol Res Perspect* **3**, (2015).

164. Chen, S. *et al.* Higher serum VGF protein levels discriminate bipolar depression from major depressive disorder. *J Neurosci Res* **97**, 597–606 (2019).

165. Thakker-Varia, S. & Alder, J. Neuropeptides in depression: Role of VGF. *Behavioural Brain Research* **197**, 262–278 (2009).

166. Ji, M. *et al.* lncRNA H19 binds VGF and promotes pNEN progression via PI3K/AKT/CREB signaling. *Endocrine-Related Cancer* **26**, 643–658 (2019).

167. Mayeux, R. Biomarkers: potential uses and limitations. *NeuroRx* **1**, 182–188 (2004).

168. Wu, J. T., Erickson, A. J., Tsao, K. C., Wu, T. L. & Sun, C. F. Elevated serum chromogranin A is detectable in patients with carcinomas at advanced disease stages. *Ann Clin Lab Sci* **30**, 175–178 (2000).

169. Leone, N., Pellicano, R., Brunello, F., Rizzetto, M. & Ponzetto, A. Elevated serum chromogranin A in patients with hepatocellular carcinoma. *Clin Exp Med* **2**, 119–123 (2002).

170. Sobol, R. E. *et al.* Elevated serum chromogranin A concentrations in small-cell lung carcinoma. *Ann Intern Med* **105**, 698–700 (1986).

171. Szarvas, T. *et al.* High-soluble CGA levels are associated with poor survival in bladder cancer. *Endocr Connect* **8**, 625–633 (2019).

172. Parmiani, G. *et al.* A pilot Phase I study combining peptide-based vaccination and NGR-hTNF vessel targeting therapy in metastatic melanoma. *OncoImmunology* **3**, e963406 (2014).

173. Bianco, M. *et al.* Chromogranin A Is Preferentially Cleaved into Proangiogenic Peptides in the Bone Marrow of Multiple Myeloma Patients. *Cancer Res* **76**, 1781–1791 (2016).

174. Malaguarnera, M. *et al.* Elevated chromogranin A (CgA) serum levels in the patients with advanced pancreatic cancer. *Arch Gerontol Geriatr* **48**, 213–217 (2009).

175. Tropea, F. *et al.* Evaluation of chromogranin A expression in patients with non-neuroendocrine tumours. *Clin Drug Investig* **26**, 715–722 (2006).

176. Khan, M. O. & Ather, M. H. Chromogranin A--serum marker for prostate cancer. *J Pak Med Assoc* **61**, 108–111 (2011).

177. Lee, S. H. *et al.* Plasma Chromogranin A as a Prognostic Marker in Pancreatic Ductal Adenocarcinoma. *Pancreas* **48**, 662–669 (2019).

178. Hsiao, R. J., Seeger, R. C., Yu, A. L. & O'Connor, D. T. Chromogranin A in children with neuroblastoma. Serum concentration parallels disease stage and predicts survival. *J Clin Invest* **85**, 1555–1559 (1990).

179. Dottorini, M. E. *et al.* Multivariate analysis of patients with medullary thyroid carcinoma. Prognostic significance and impact on treatment of clinical and pathologic variables. *Cancer* **77**, 1556–1565 (1996).

180. Matar, S. *et al.* Blood Chromogranin A Is Not Effective as a Biomarker for Diagnosis or Management of Bronchopulmonary Neuroendocrine Tumors/Neoplasms. *Neuroendocrinology* **110**, 185–197 (2020).

181. Malaguarnera, M. *et al.* Elevated chromogranin A serum levels in ovarian carcinoma patients. *Indian J Cancer* **51**, 315–318 (2014).

182. Spadaro, A. *et al.* Serum chromogranin-A in hepatocellular carcinoma: diagnostic utility and limits. *World J Gastroenterol* **11**, 1987–1990 (2005).

183. Al-Risi, E. S., Al-Essry, F. S. & Mula-Abed, W.-A. S. Chromogranin A as a Biochemical Marker for Neuroendocrine Tumors: A Single Center Experience at Royal Hospital, Oman. *Oman Medical Journal* **32**, 365–370 (2017).

184. O'Connor, D. T. Chromogranin A: implications for hypertension. *J Hypertens Suppl* **2**, S147-150 (1984).

185. Sciola, V. *et al.* Plasma chromogranin a in patients with inflammatory bowel disease. *Inflamm Bowel Dis* **15**, 867–871 (2009).

186. Kurnatowska, I. & Nowicki, M. Serum chromogranin A concentration and intradialytic hypotension in chronic haemodialysis patients. *Int Urol Nephrol* **38**, 701–705 (2006).

187. Yao, Q., Tong, Y., Peng, R., Liu, Z. & Li, Y. Associations of serum chromogranin A with depressive symptoms in men with unipolar depressive disorder. *Gen Hosp Psychiatry* **66**, 120–124 (2020).

188. Estensen, M. E. *et al.* Prognostic value of plasma chromogranin A levels in patients with complicated myocardial infarction. *American Heart Journal* **152**, 927.e1-927.e6 (2006).

189. Wu, P. B. *et al.* Increased plasma CgA levels associated with nonalcoholic fatty liver disease. *Turk J Gastroenterol* **26**, 404–407 (2015).

190. Capellino, S. *et al.* Increased chromogranin A levels indicate sympathetic hyperactivity in patients with rheumatoid arthritis and systemic lupus erythematosus. *J Rheumatol* **35**, 91–99 (2008).

191. Tian, R. *et al.* Plasma PTX3, MCP1 and Ang2 are early biomarkers to evaluate the severity of sepsis and septic shock. *Scand J Immunol* **90**, (2019).

192. Herold, Z. *et al.* Serum chromogranin A level continuously rises with the progression of type 1 diabetes, and indicates the presence of both

enterochromaffin-like cell hyperplasia and autoimmune gastritis. *J Diabetes Investig* **11**, 865–873 (2020).

193. Kogawa, E. M. *et al.* Impact of glycemic control on oral health status in type 2 diabetes individuals and its association with salivary and plasma levels of chromogranin A. *Arch Oral Biol* **62**, 10–19 (2016).

194. Teggi, R. *et al.* Altered chromogranin A circulating levels in Meniere's disease. *Dis Markers* **2015**, 643420 (2015).

195. Mojiminiyi, O. A., Abdella, N. & George, S. Evaluation of serum cystatin C and chromogranin A as markers of nephropathy in patients with type 2 diabetes mellitus. *Scand J Clin Lab Invest* **60**, 483–489 (2000).

196. Dag, E. *et al.* Alteration in chromogranin A, obestatin and total ghrelin levels of saliva and serum in epilepsy cases. *Peptides* **31**, 932–937 (2010).

197. Al-Shoumer, K. A. S. & Vasanthy, B. A. Serum chromogranin A concentration in hyperthyroidism before and after medical treatment. *J Clin Endocrinol Metab* **94**, 2321–2324 (2009).

198. Domenici, E. *et al.* Plasma Protein Biomarkers for Depression and Schizophrenia by Multi Analyte Profiling of Case-Control Collections. *PLoS ONE* **5**, e9166 (2010).

199. Reshma, A. P. *et al.* Chromogranin A: Novel biomarker between periodontal disease and psychosocial stress. *J Indian Soc Periodontol* **17**, 214–218 (2013).

200. Moftaquir-Handaj, A., Barbé, F., Barbarino-Monnier, P., Aunis, D. & Boutroy, M. J. Circulating chromogranin A and catecholamines in human fetuses at uneventful birth. *Pediatr Res* **37**, 101–105 (1995).

201. Kalkan, Ç., Karakaya, F. & Soykan, İ. Factors associated with elevated serum chromogranin A levels in patients with autoimmune gastritis. *Turk J Gastroenterol* **27**, 515–520 (2016).

202. Pregun, I. *et al.* Effect of proton-pump inhibitor therapy on serum chromogranin a level. *Digestion* **84**, 22–28 (2011).

203. Sasaki, T. *et al.* Changes in chromogranin a serum levels during endocrine therapy in metastatic prostate cancer patients. *Eur Urol* **48**, 224–229; discussion 229-230 (2005).

204. von Hardenberg, J. *et al.* Influence of abiraterone acetate on circulating neuromediators in chemotherapy-naïve castration-resistant prostate cancer. *Prostate* **76**, 613–619 (2016).

205. Nickel, T. *et al.* Extreme exercise enhances chromogranin A levels correlating with stress levels but not with cardiac burden. *Atherosclerosis* **220**, 219–222 (2012).

206. Biondi, A. *et al.* Elevated serum levels of Chromogranin A in hepatocellular carcinoma. *BMC Surg* **12 Suppl 1**, S7 (2012).

207. Lugowska, I. *et al.* Serum markers in early-stage and locally advanced melanoma. *Tumor Biology* **36**, 8277–8285 (2015).

208. Lipton, A. *et al.* Serum TIMP-1 and Response to the Aromatase Inhibitor Letrozole Versus Tamoxifen in Metastatic Breast Cancer. *JCO* **26**, 2653–2658 (2008).

209. Lempinen, M. *et al.* Prognostic value of serum MMP-8, -9 and TIMP-1 in patients with hepatocellular carcinoma. *Annals of Medicine* **45**, 482–487 (2013).

210. Määtta, M., Talvensaari-Mattila, A., Turpeenniemi-Hujanen, T. & Santala, M. Matrix metalloproteinase-2 (MMP-2) and -9 (MMP-9) and their tissue inhibitors (TIMP-1 and TIMP-2) in differential diagnosis between low malignant potential (LMP) and malignant ovarian tumours. *Anticancer Res* **27**, 2753–2758 (2007).

211. Brand, R. E. *et al.* Serum biomarker panels for the detection of pancreatic cancer. *Clin Cancer Res* **17**, 805–816 (2011).

212. Terpos, E. *et al.* High levels of serum TIMP-1 correlate with advanced disease and predict for poor survival in patients with multiple myeloma treated with novel agents. *Leuk Res* **34**, 399–402 (2010).

213. Jumper, C., Cobos, E. & Lox, C. Determination of the serum matrix metalloproteinase-9 (MMP-9) and tissue inhibitor of matrix metalloproteinase-1 (TIMP-1) in patients with either advanced small-cell lung cancer or non-small-cell lung cancer prior to treatment. *Respir Med* **98**, 173–177 (2004).

214. Higashimoto, Y. Increased serum concentrations of tissue inhibitor of metalloproteinase-1 in COPD patients. *European Respiratory Journal* **25**, 885–890 (2005).

215. Fiedorczyk, M., Klimiuk, P. A., Sierakowski, S., Gindzienska-Sieskiewicz, E. & Chwiecko, J. Serum Matrix Metalloproteinases and Tissue Inhibitors of Metalloproteinases in Patients with Early Rheumatoid Arthritis. *The Journal of Rheumatology* 7.

216. Flisiak, I., Myśliwiec, H. & Chodynicka, B. Effect of psoriasis treatment on plasma concentrations of metalloproteinase-1 and tissue inhibitor of metalloproteinases-1. *J Eur Acad Dermatol Venereol* **19**, 418–421 (2005).

217. Wiercinska-Drapalo, A. Plasma matrix metalloproteinase-1 and tissue inhibitor of metalloproteinases-1 as biomarkers of ulcerative colitis activity. *WJG* **9**, 2843 (2003).

218. Zayani, Y. *et al.* Abnormal Circulating Levels of Metalloproteinase and Their Inhibitor in Hypertensive Patients. *Clin Lab* **62**, 527–533 (2016).

219. Hallgren, H. C. B., Eliasson, P., Aspenberg, P. & Adolfsson, L. E. Elevated plasma levels of TIMP-1 in patients with rotator cuff tear. *Acta Orthopaedica* **83**, 523–528 (2012).

220. Erman, H. *et al.* The association of vascular endothelial growth factor, metalloproteinases and their tissue inhibitors with cardiovascular risk factors in the metabolic syndrome. *Eur Rev Med Pharmacol Sci* **20**, 1015–1022 (2016).

221. Blaise, S. *et al.* Correlation of biomarkers of endothelium dysfunction and matrix remodeling in patients with systemic sclerosis. *J Rheumatol* **36**, 984–988 (2009).

222. Martins-Oliveira, A. *et al.* Specific matrix metalloproteinase 9 (MMP-9) haplotype affect the circulating MMP-9 levels in women with migraine. *J Neuroimmunol* **252**, 89–94 (2012).

223. Hästbacka, J. *et al.* Matrix metalloproteinases -8 and -9 and tissue inhibitor of metalloproteinase-1 in burn patients. A prospective observational study. *PLoS One* **10**, e0125918 (2015).

224. Nakamura, T. *et al.* Elevation of serum levels of metalloproteinase-1, tissue inhibitor of metalloproteinase-1 and type IV collagen, and plasma levels of metalloproteinase-9 in polycystic kidney disease. *Am J Nephrol* **20**, 32–36 (2000).

225. Eisner, L., Vambutas, A. & Pathak, S. The Balance of Tissue Inhibitor of Metalloproteinase-1 and Matrix Metalloproteinase-9 in the Autoimmune Inner Ear Disease Patients. *J Interferon Cytokine Res* **37**, 354–361 (2017).

226. Pereira, T. N. *et al.* Serum markers of hepatic fibrogenesis in cystic fibrosis liver disease. *J Hepatol* **41**, 576–583 (2004).

227. Chiang, T.-Y. *et al.* The circulating level of MMP-9 and its ratio to TIMP-1 as a predictor of severity in patients with community-acquired pneumonia. *Clin Chim Acta* **424**, 261–266 (2013).

228. Świdrowska, J., Smolewski, P., Stańczyk, J. & Smolewska, E. Serum Angiogenesis Markers and Their Correlation with Ultrasound-Detected Synovitis in Juvenile Idiopathic Arthritis. *J Immunol Res* **2015**, 741457 (2015).

229. Prokopchuk, O. *et al.* Elevated systemic levels of the matrix metalloproteinase inhibitor TIMP-1 correlate with clinical markers of cachexia in patients with chronic pancreatitis and pancreatic cancer. *BMC Cancer* **18**, 128 (2018).

230. Naito, K. *et al.* Measurement of matrix metalloproteinases (MMPs) and tissue inhibitor of metalloproteinases-1 (TIMP-1) in patients with knee osteoarthritis: comparison with generalized osteoarthritis. *Rheumatology (Oxford)* **38**, 510–515 (1999).

231. Kralisch, S. *et al.* Tissue inhibitor of metalloproteinase-1 predicts adiposity in humans. *Eur J Endocrinol* **156**, 257–261 (2007).

232. Gomes, V. A. *et al.* Oral contraceptive containing chlormadinone acetate and ethinylestradiol reduces plasma concentrations of matrix metalloproteinase-2 in women with polycystic ovary syndrome. *Basic Clin Pharmacol Toxicol* **111**, 211–216 (2012).

233. Cevik, C. *et al.* Rosuvastatin therapy does not affect serum MMP-13 or TIMP-1 levels in hypercholesterolemic patients. *Tex Heart Inst J* **38**, 229–233 (2011).

234. Trucco, E. *et al.* Plasma tissue inhibitor of matrix metalloproteinase-1 a predictor of long-term mortality in patients treated with cardiac resynchronization therapy. *Europace* **18**, 232–237 (2016).

235. Iacobelli, S. *et al.* Prognostic value of a novel circulating serum 90K antigen in breast cancer. *Br J Cancer* **69**, 172–176 (1994).

236. Park, Y. P. *et al.* Up-regulation of Mac-2 binding protein by hTERT in gastric cancer. *Int J Cancer* **120**, 813–820 (2007).

237. Marchetti, A. *et al.* Expression of 90K (Mac-2 BP) correlates with distant metastasis and predicts survival in stage I non-small cell lung cancer patients. *Cancer Res* **62**, 2535–2539 (2002).

238. Fornarini, B. *et al.* Adhesion to 90K (Mac-2 BP) as a mechanism for lymphoma drug resistance in vivo. *Blood* **96**, 3282–3285 (2000).

239. Iacovazzi, P. A. *et al.* Serum 90K/MAC-2BP glycoprotein in patients with liver cirrhosis and hepatocellular carcinoma: a comparison with alpha-fetoprotein. *Clin Chem Lab Med* **39**, 961–965 (2001).

240. Maekawa, T. *et al.* Serum Mac-2 binding protein is a novel biomarker for chronic pancreatitis. *WJG* **22**, 4403 (2016).

241. Kamada, Y. *et al.* Serum Mac-2 binding protein levels as a novel diagnostic biomarker for prediction of disease severity and nonalcoholic steatohepatitis. *Proteomics Clin Appl* **7**, 648–656 (2013).

242. Xie, H. *et al.* Elevated plasma levels of Mac-2 binding protein predict poor cardiovascular outcomes in patients with acute coronary syndrome. *Coronary Artery Disease* **28**, 683–689 (2017).

243. Sugiura, T. *et al.* Serum levels of Mac-2 binding protein increase with cardiovascular risk and reflect silent atherosclerosis. *Atherosclerosis* **251**, 192–196 (2016).

244. Xie, H. *et al.* Expression of Mac-2 binding protein in human carotid atheroma is associated with plaque instability and clinical manifestations. *Biomedicine & Pharmacotherapy* **110**, 465–472 (2019).

245. Hashimoto, Y. *et al.* Serum levels of mac-2 binding protein are associated with diabetic microangiopathy and macroangiopathy in people with type 2 diabetes. *BMJ Open Diabetes Res Care* **8**, (2020).

246. Cecamore, C. *et al.* 90K immunostimulatory glycoprotein in children with juvenile idiopathic arthritis. *Mod Rheumatol* **28**, 637–641 (2018).

247. Kamada, Y. *et al.* Serum Mac-2 Binding Protein Levels Associate with Metabolic Parameters and Predict Liver Fibrosis Progression in Subjects with Fatty Liver Disease: A 7-Year Longitudinal Study. *Nutrients* **12**, (2020).

248. Darlix, A. *et al.* Serum NSE, MMP-9 and HER2 extracellular domain are associated with brain metastases in metastatic breast cancer patients: predictive biomarkers for brain metastases? *Int J Cancer* **139**, 2299–2311 (2016).

249. Naito, A. *et al.* Prognostic significance of serum neuron-specific enolase in small cell carcinoma of the urinary bladder. *World J Urol* **35**, 97–103 (2017).

250. Chen, C. A. *et al.* Serum neuron-specific enolase levels in patients with small cell carcinoma of the uterine cervix. *J Formos Med Assoc* **93**, 81–83 (1994).

251. Yang, W. Y. *et al.* Analysis of specific serum markers of colon carcinoma using a Bhattacharyya-based support vector machine. *Genet Mol Res* **16**, (2017).

252. Luo, H. *et al.* Clinical significance of serum neuron-specific enolase in gastric adenocarcinoma. *Medicine (Baltimore)* **99**, e19829 (2020).

253. Yu, J. *et al.* Identification of potential serum biomarkers for simultaneously classifying lung adenocarcinoma, squamous cell carcinoma and small cell carcinoma. *Cancer Biomark* (2020) doi:10.3233/CBM-201440.

254. Fujiwara, H. *et al.* Clinical significance of serum neuron-specific enolase in patients with adult T-cell leukemia. *Am J Hematol* **71**, 80–84 (2002).

255. Yang, H. *et al.* Expression of neuron-specific enolase in multiple myeloma and implications for clinical diagnosis and treatment. *PLoS One* **9**, e94304 (2014).

256. Kuramitsu, M., Sawa, H., Takeshita, I., Iwaki, T. & Kato, K. Neuron-specific gamma-enolase derived from human glioma. *Neurochem Pathol* **4**, 89–105 (1986).

257. Adamiak, G., Jurkiewicz, D., Polberg, K. & Konieczna, M. [Differences in assay sensitivity of selected tumor markers in head and neck neoplasms]. *Otolaryngol Pol* **49 Suppl 20**, 99–101 (1995).

258. Szarvas, T. *et al.* Comprehensive analysis of serum chromogranin A and neuron-specific enolase levels in localized and castration-resistant prostate cancer. *BJU Int* **127**, 44–55 (2021).

259. Kamiya, N. *et al.* Pretreatment serum level of neuron specific enolase (NSE) as a prognostic factor in metastatic prostate cancer patients treated with endocrine therapy. *Eur Urol* **44**, 309–314; discussion 314 (2003).

260. Gao, X.-N., Tang, S.-Q. & Lin, J. [Clinical features and prognosis of advanced neuroblastoma in children]. *Zhongguo Dang Dai Er Ke Za Zhi* **9**, 351–354 (2007).

261. Tatekawa, Y., Kemmotsu, H., Mouri, T., Joe, K. & Ohkawa, H. A case of pediatric ovarian dysgerminoma associated with high serum levels and positive immunohistochemical staining of neuron-specific enolase. *J Pediatr Surg* **39**, 1437–1439 (2004).

262. Hirabayashi, K., Yasuda, M., Osamura, R. Y., Hirasawa, T. & Murakami, M. Ovarian nongestational choriocarcinoma mixed with various epithelial malignancies in association with endometriosis. *Gynecol Oncol* **102**, 111–117 (2006).

263. Xue, Y., Li, J. & Lu, X. A Novel Immune-Related Prognostic Signature for Thyroid Carcinoma. *Technol Cancer Res Treat* **19**, 153303382093586 (2020).

264. Windrichova, J. *et al.* Testing of a Novel Cancer Metastatic Multiplex Panel for the Detection of Bone-metastatic Disease - a Pilot Study. *Anticancer Res* **36**, 1973–1978 (2016).

265. Wiener, C. D. *et al.* Reduced serum levels of neuron specific enolase (NSE) in drug-naïve subjects with major depression and bipolar disorder. *Neurochem Res* **38**, 1394–1398 (2013).

266. Shang, X. *et al.* The serum levels of tumor marker CA19-9, CEA, CA72-4, and NSE in type 2 diabetes without malignancy and the relations to the metabolic control. *Saudi Med J* **38**, 204–208 (2017).

267. Dinc, B. *et al.* Evaluation of Preoperative and Postoperative S100β and NSE Levels in Liver Transplantation and Right Lobe Living-Donor Hepatectomy: A Prospective Cohort Study. *Transplant Proc* **53**, 16–24 (2021).

268. Li, J. *et al.* NSE, a Potential Biomarker, Is Closely Connected to Diabetic Peripheral Neuropathy. *Diabetes Care* **36**, 3405–3410 (2013).

269. Collazos, J., Esteban, C., Fernández, A. & Genollá, J. Measurement of the serum tumor marker neuron-specific enolase in patients with benign pulmonary diseases. *Am J Respir Crit Care Med* **150**, 143–145 (1994).

270. Xu, Y. *et al.* Clinical characteristics of patients with lung cancer and idiopathic pulmonary fibrosis in China. *Thorac Cancer* **3**, 156–161 (2012).

271. Wang, S.-Y. *et al.* Metabolic syndrome and its components with neuron-specific enolase: a cross-sectional study in large health check-up population in China. *BMJ Open* **8**, e020899 (2018).

272. Gönen, M. *et al.* S100B and neuron-specific enolase levels in episodic and chronic migraine. *Acta Neurol Scand* **143**, 298–302 (2021).

273. Li, Y. *et al.* A Pilot Study of the Use of Dexmedetomidine for the Control of Delirium by Reducing the Serum Concentrations of Brain-Derived Neurotrophic Factor, Neuron-Specific Enolase, and S100B in Polytrauma Patients. *J Intensive Care Med* **34**, 674–681 (2019).

274. Hoffmann, J. *et al.* Association between serum neuron-specific enolase, age, overweight, and structural MRI patterns in 901 subjects. *Transl Psychiatry* **7**, 1272 (2017).

275. Bergman, L. & Åkerud, H. Plasma Levels of the Cerebral Biomarker, Neuron-Specific Enolase, are Elevated During Pregnancy in Women Developing Preeclampsia. *Reprod Sci* **23**, 395–400 (2016).

276. Nam, S.-J. *et al.* Neuron-specific enolase as a novel biomarker reflecting tuberculosis activity and treatment response. *Korean J Intern Med* **31**, 694–702 (2016).

277. Du, W. *et al.* The Prognostic Value of Serum Neuron Specific Enolase (NSE) and S100B Level in Patients of Acute Spinal Cord Injury. *Med Sci Monit* **24**, 4510–4515 (2018).

278. Massabki, P. S., Silva, N. P., Lourenço, D. M. & Andrade, L. E. C. Neuron specific enolase concentration is increased in serum and decreased in platelets of patients with active systemic sclerosis. *J Rheumatol* **30**, 2606–2612 (2003).

279. Shen, Y. & Gao, H.-M. Serum somatostatin and neuron-specific enolase might be biochemical markers of vascular dementia in the early stage. *Int J Clin Exp Med* **8**, 19471–19475 (2015).

280. Sohn, J.-H., Kim, C.-H., Lee, S.-H., Kim, J. H. & Lee, J. J. Diagnostic Value of Serum Biomarkers for Differentiating Central and Peripheral Causes of Acute Vertigo. *Front Med (Lausanne)* **7**, 84 (2020).

281. Pandey, A., Shrivastava, A. K. & Saxena, K. Neuron Specific Enolase and C-reactive Protein Levels in Stroke and Its Subtypes: Correlation with Degree of Disability. *Neurochemical Research* **39**, 1426–1432 (2014).

282. Huang, H.-B. *et al.* Serum neuron-specific enolase: A promising biomarker of silicosis. *WJCC* **9**, 1016–1025 (2021).

283. Oka, S. *et al.* Serum biomarker analysis of collagen disease patients with acute-onset diffuse interstitial lung disease. *BMC Immunol* **14**, 9 (2013).

284. Maiti, R. *et al.* Effect of carbamazepine and oxcarbazepine on serum neuron-specific enolase in focal seizures: A randomized controlled trial. *Epilepsy Res* **138**, 5–10 (2017).

285. Bagnato, S. *et al.* Reduced Neuron-Specific Enolase Levels in Chronic Severe Traumatic Brain Injury. *J Neurotrauma* **37**, 423–427 (2020).

286. Koch, M., Mostert, J., Heersema, D., Teelken, A. & De Keyser, J. Plasma S100beta and NSE levels and progression in multiple sclerosis. *J Neurol Sci* **252**, 154–158 (2007).

287. Rinösl, H., Skhirtladze, K., Felli, A., Ankersmit, H. J. & Dworschak, M. The neuroprotective effect of magnesium sulphate during iatrogenically-induced ventricular fibrillation. *Magnes Res* **26**, 109–119 (2013).

288. Soresi, E. *et al.* Effect of octreotide on neuroenolase levels in patients with small cell lung cancer. *Tumori* **80**, 332–334 (1994).

289. Qiu, J. *et al.* Effects of fentanyl for pain control and neuroprotection in very preterm newborns on mechanical ventilation. *J Matern Fetal Neonatal Med* **32**, 3734–3740 (2019).

290. Yılmaz, F. M. *et al.* Serum biochemical markers of central nerve system damage in children with acute elemental mercury intoxicatıon. *Clin Toxicol (Phila)* **52**, 32–38 (2014).

291. Peng, M. *et al.* Report - Cerebral electrical impedance value reflects brain edema caused by cardiopulmonary bypass in infants. *Pak J Pharm Sci* **30**, 913–919 (2017).

292. Szarvas, T. *et al.* Serum Levels of Angiogenic Factors and their Prognostic Relevance in Bladder Cancer. *Pathol. Oncol. Res.* **15**, 193–201 (2009).

293. Li, P., He, Q., Luo, C. & Qian, L. Diagnostic and prognostic potential of serum angiopoietin-2 expression in human breast cancer. 5.

294. Quartarone, E. *et al.* Differential levels of soluble angiopoietin-2 and Tie-2 in patients with haematological malignancies. *Eur J Haematol* **77**, 480–485 (2006).

295. Maffei, R. *et al.* Endothelium-mediated survival of leukemic cells and angiogenesis-related factors are affected by lenalidomide treatment in chronic lymphocytic leukemia. *Experimental Hematology* **42**, 126-136.e1 (2014).

296. Pappa, C. A. *et al.* Prognostic impact of angiopoietin-2 in multiple myeloma. *J Cancer Res Clin Oncol* **140**, 1801–1805 (2014).

297. Terlikowska, K. M. *et al.* Clinical value of selected markers of angiogenesis, inflammation, insulin resistance and obesity in type 1 endometrial cancer. *BMC Cancer* **20**, 921 (2020).

298. Sallinen, H. *et al.* Preoperative angiopoietin-2 serum levels: a marker of malignant potential in ovarian neoplasms and poor prognosis in epithelial ovarian cancer. *Int J Gynecol Cancer* **20**, 1498–1505 (2010).

299. Engin, H. *et al.* Plasma concentrations of angiopoietin-1, angiopoietin-2 and Tie-2 in colon cancer. *Eur Cytokine Netw* **23**, 68–71 (2012).

300. Engin, H., Ustündağ, Y., Ozel Tekin, I. & Gökmen, A. Plasma concentrations of Ang-1, Ang-2 and Tie-2 in gastric cancer. *Eur Cytokine Netw* **23**, 21–24 (2012).

301. Kopczyńska, E. *et al.* Plasma concentration of angiopoietin-1, angiopoietin-2 and Tie-2 in cervical cancer. *Eur J Gynaecol Oncol* **30**, 646–649 (2009).

302. Ilhan, A. *et al.* Angiogenic factors in plasma of brain tumour patients. *Anticancer Res* **29**, 731–736 (2009).

303. Chuma, M. *et al.* Early Changes in Circulating FGF19 and Ang-2 Levels as Possible Predictive Biomarkers of Clinical Response to Lenvatinib Therapy in Hepatocellular Carcinoma. *Cancers* **12**, 293 (2020).

304. Attia, M. A., Hazzaa, S. M., Essa, S. A. & Seleim, M. F. Prognostic value of soluble angiopoietin-2 and soluble Tie-2 in Egyptian patients with acute myeloid leukemia. *Turk J Hematol* 7.

305. D' Angelo, S. P. *et al.* Alliance A091103 a phase II study of the angiopoietin 1 and 2 peptibody trebananib for the treatment of angiosarcoma. *Cancer Chemother Pharmacol* **75**, 629–638 (2015).

306. Kawaguchi, M. *et al.* Serum levels of angiopoietin-2, but not angiopoietin-1, are elevated in patients with erythrodermic cutaneous T-cell lymphoma. *Acta Derm Venereol* **94**, 9–13 (2014).

307. Akın Kabalak, P. *et al.* Prognostic significance of serum vascular endothelial growth factor and Angiopoietin-2 in patients with lung cancer. *Tuberk Toraks* **63**, 71–77 (2015).

308. Helfrich, I. *et al.* Angiopoietin-2 levels are associated with disease progression in metastatic malignant melanoma. *Clin Cancer Res* **15**, 1384–1392 (2009).

309. Lee, C.-H. *et al.* Correlative serum biomarker analyses in the phase 2 trial of lenvatinib-plus-everolimus in patients with metastatic renal cell carcinoma. *Br J Cancer* **124**, 237–246 (2021).

310. Sridharan, V. *et al.* Effects of definitive chemoradiation on circulating immunologic angiogenic cytokines in head and neck cancer patients. *J Immunother Cancer* **4**, 32 (2016).

311. Zhou, Y.-Z. *et al.* Role of serum angiopoietin-2 level in screening for esophageal squamous cell cancer and its precursors. *Chin Med J (Engl)* **120**, 1216–1219 (2007).

312. Chen, S.-M., Li, D., Xing, X. & Li, Z.-P. Higher serum angiopoietin 2 levels are independently associated with coronary microvascular dysfunction in patients with angina in the absence of obstructive coronary artery disease. *Chinese Medical Journal* **133**, 1662–1668 (2020).

313. Ben-Shoshan, J. *et al.* Sustained Elevation of Vascular Endothelial Growth Factor and Angiopoietin-2 Levels After Transcatheter Aortic Valve Replacement. *Can J Cardiol* **32**, 1454–1461 (2016).

314. Demir, V. *et al.* Relationship of serum calprotectin, angiopoietin-1, and angiopoietin-2 levels with coronary collateral circulation in patients with stable coronary artery disease. *Kardiol Pol* **77**, 1155–1162 (2019).

315. Siddiqui, K., Joy, S. S., Nawaz, S. S., Al Otaibi, M. T. & Al-Rubeaan, K. Angiopoietin-2 level as a tool for cardiovascular risk stratification in hypertensive type 2 diabetic subjects. *Postgraduate Medicine* **130**, 402–408 (2018).

316. Sahni, M. *et al.* Novel biomarkers of bronchopulmonary dysplasia and bronchopulmonary dysplasia-associated pulmonary hypertension. *J Perinatol* **40**, 1634–1643 (2020).

317. Ojeda-Fernandez, L. *et al.* Reduced plasma levels of Ang-2 and sEng as novel biomarkers in hereditary hemorrhagic telangiectasia (HHT). *Clin Chim Acta* **411**, 494–499 (2010).

318. Alqudah, M. *et al.* Changes of serum growth factors profiles in patients with venous thromboembolism. *Scandinavian Journal of Clinical and Laboratory Investigation* **77**, 595–600 (2017).

319. Chen, S., Guo, L., Chen, B., Sun, L. & Cui, M. Association of serum angiopoietin-1, angiopoietin-2 and angiopoietin-2 to angiopoietin-1 ratio with heart failure in patients with acute myocardial infarction. *Exp Ther Med* **5**, 937–941 (2013).

320. Patel, J. V., Lim, H. S., Varughese, G. I., Hughes, E. A. & Lip, G. Y. H. Angiopoietin-2 levels as a biomarker of cardiovascular risk in patients with hypertension. *Annals of Medicine* **40**, 215–222 (2008).

321. Iribarren, C. *et al.* Circulating Angiopoietins-1 and -2, Angiopoietin Receptor Tie-2 and Vascular Endothelial Growth Factor-A as Biomarkers of Acute Myocardial Infarction: a Prospective Nested Case-Control Study. *BMC Cardiovasc Disord* **11**, 31 (2011).

322. Höbaus, C. *et al.* Angiopoietin-2 and Survival in Peripheral Artery Disease Patients. *Thromb Haemost* **118**, 791–797 (2018).

323. Kerget, F., Özkurt, Z., Öztürk, N. & Yilmaz, S. The relationship with clinical course and prognosis of serum endothelin-1, angiopoietin-2, and tie-2 levels in Crimean-Congo hemorrhagic fever. *Turk J Med Sci* 6.

324. Nikolakopoulou, S. *et al.* Serum Angiopoietin-2 and CRP Levels During COPD Exacerbations. *COPD: Journal of Chronic Obstructive Pulmonary Disease* **11**, 46–51 (2014).

325. Uehara, M. *et al.* Impact of angiopoietin-1 and -2 on clinical course of idiopathic pulmonary fibrosis. *Respiratory Medicine* **114**, 18–26 (2016).

326. Ando, M. *et al.* Angiopoietin-2 expression in patients with an acute exacerbation of idiopathic interstitial pneumonias. *Respiratory Medicine* **117**, 27–32 (2016).

327. Takahashi, T. *et al.* Dynamics of serum angiopoietin-2 levels correlate with efficacy of intravenous pulse cyclophosphamide therapy for interstitial lung disease associated with systemic sclerosis. *Mod Rheumatol* **23**, 884–890 (2013).

328. Kathamuthu, G. R. *et al.* Tuberculous lymphadenitis is associated with altered levels of circulating angiogenic factors. *Int J Tuberc Lung Dis* **22**, 557–566 (2018).

329. Diamond, J. M. *et al.* Elevated plasma angiopoietin-2 levels and primary graft dysfunction after lung transplantation. *PLoS One* **7**, e51932 (2012).

330. Huang, Q. *et al.* Angiopoietin-2 Is an Early Predictor for Acute Gastrointestinal Injury and Intestinal Barrier Dysfunction in Patients with Acute Pancreatitis. *Dig Dis Sci* **66**, 114–120 (2021).

331. Hernández-Bartolomé, Á. *et al.* Angiopoietin-2/angiopoietin-1 as non-invasive biomarker of cirrhosis in chronic hepatitis C. *World J Gastroenterol* **22**, 9744–9751 (2016).

332. Holleran, G., Hussey, M., Smith, S. & McNamara, D. Assessment of serum angiogenic factors as a diagnostic aid for small bowel angiodysplasia in patients with obscure gastrointestinal bleeding and anaemia. *WJGP* **8**, 127 (2017).

333. Lefere, S. *et al.* Angiopoietin-2 Promotes Pathological Angiogenesis and Is a Therapeutic Target in Murine Nonalcoholic Fatty Liver Disease. *Hepatology* **69**, 1087–1104 (2019).

334. Koutroubakis, I. E. *et al.* Potential role of soluble angiopoietin-2 and Tie-2 in patients with inflammatory bowel disease. *Eur J Clin Invest* **36**, 127–132 (2006).

335. Khattab, F. M. & Said, N. M. Serum angiopoietin-2 level as a novel biomarker in atopic dermatitis. *Int J Dermatol* **59**, 3–7 (2020).

336. Kurosaka, D. *et al.* Clinical significance of serum levels of vascular endothelial growth factor, angiopoietin-1, and angiopoietin-2 in patients with rheumatoid arthritis. *J Rheumatol* **37**, 1121–1128 (2010).

337. Michalska-Jakubus, M., Kowal-Bielecka, O., Chodorowska, G., Bielecki, M. & Krasowska, D. Angiopoietins-1 and -2 are differentially expressed in the sera of patients with systemic sclerosis: high angiopoietin-2 levels are associated with greater severity and higher activity of the disease. *Rheumatology* **50**, 746–755 (2011).

338. Karampoor, S., Zahednasab, H., Ramagopalan, S., Mehrpour, M. & Keyvani, H. Angiogenic factors are associated with multiple sclerosis. *J Neuroimmunol* **301**, 88–93 (2016).

339. Wang, J.-M. *et al.* Serum levels and gene polymorphisms of angiopoietin 2 in systemic lupus erythematosus patients. *Sci Rep* **11**, 10 (2021).

340. Paczkowska, E. *et al.* Evidence for proangiogenic cellular and humoral systemic response in patients with acute onset of spinal cord injury. *The Journal of Spinal Cord Medicine* **38**, 729–744 (2015).

341. Oluboyo, A. O., Chukwu, S. I., Oluboyo, B. O. & Odewusi, O. O. Evaluation of Angiopoietins 1 and 2 in Malaria-Infested Children. *Journal of Environmental and Public Health* **2020**, 1–5 (2020).

342. Figueroa-Vega, N. *et al.* Serum levels of angiogenic molecules in autoimmune thyroid diseases and their correlation with laboratory and clinical features. *J Clin Endocrinol Metab* **94**, 1145–1153 (2009).

343. Breunis, W. B. *et al.* Disruption of vascular homeostasis in patients with Kawasaki disease: involvement of vascular endothelial growth factor and angiopoietins. *Arthritis Rheum* **64**, 306–315 (2012).

344. David, S. *et al.* Circulating angiopoietin-2 levels increase with progress of chronic kidney disease. *Nephrology Dialysis Transplantation* **25**, 2571–2579 (2010).

345. Graham, S. M. *et al.* A prospective study of endothelial activation biomarkers, including plasma angiopoietin-1 and angiopoietin-2, in Kenyan women initiating antiretroviral therapy. *BMC Infect Dis* **13**, 263 (2013).

346. Venter, P. C., Malan, L. & Schutte, A. E. Psychosocial stress but not hypertensive status associated with angiogenesis in Africans. *Blood Pressure* **23**, 307–314 (2014).

347. Daponte, A. *et al.* Angiopoietin-1 and angiopoietin-2 as serum biomarkers for ectopic pregnancy and missed abortion: a case-control study. *Clin Chim Acta* **415**, 145–151 (2013).

348. Gozal, D. *et al.* Angiopoietin-2 and soluble Tie-2 receptor plasma levels in children with obstructive sleep apnea and obesity. *Obesity (Silver Spring)* **25**, 1083–1090 (2017).

349. Xu, M. *et al.* Radiosurgery reduces plasma levels of angiogenic factors in brain arteriovenous malformation patients. *Brain Res Bull* **140**, 220–225 (2018).

350. Mapalagamage, M. *et al.* High Levels of Serum Angiopoietin 2 and Angiopoietin 2/1 Ratio at the Critical Stage of Dengue Hemorrhagic Fever in Patients and Association with Clinical and Biochemical Parameters. *J Clin Microbiol* **58**, e00436-19, /jcm/58/4/JCM.00436-19.atom (2020).

351. Giamarellos-Bourboulis, E. J., Kanellakopoulou, K., Pelekanou, A., Tsaganos, T. & Kotzampassi, K. Kinetics of angiopoietin-2 in serum of multi-trauma patients: correlation with patient severity. *Cytokine* **44**, 310–313 (2008).

352. Coban, M. & Inci, A. The association of serum angiogenic growth factors with renal structure and function in patients with adult autosomal dominant polycystic kidney disease. *Int Urol Nephrol* **50**, 1293–1300 (2018).

353. Hirokoshi, K. *et al.* Increase of Serum Angiopoietin-2 During Pregnancy Is Suppressed in Women With Preeclampsia. *American Journal of Hypertension* **18**, 1181–1188 (2005).

354. Crane, J. *et al.* Kaposiform lymphangiomatosis treated with multimodal therapy improves coagulopathy and reduces blood angiopoietin-2 levels. *Pediatr Blood Cancer* **67**, e28529 (2020).

355. Lieb, W. *et al.* Clinical and genetic correlates of circulating angiopoietin-2 and soluble Tie-2 in the community. *Circ Cardiovasc Genet* **3**, 300–306 (2010).

356. Spangenberg, A. *et al.* Bone Mineral Density and Body Composition are Associated with Circulating Angiogenic Factors in Post-menopausal Women. *Calcif Tissue Int* **99**, 608–615 (2016).

357. Loffredo, S. *et al.* Elevated plasma levels of vascular permeability factors in C1 inhibitor-deficient hereditary angioedema. *Allergy* **71**, 989–996 (2016).

358. Khalaf, N. *et al.* Role of Angiopoietins and Tie-2 in Diabetic Retinopathy. *Electron Physician* **9**, 5031–5035 (2017).

359. Hiremath, J. *et al.* Exercise improvement and plasma biomarker changes with intravenous treprostinil therapy for pulmonary arterial hypertension: A placebo-controlled trial. *The Journal of Heart and Lung Transplantation* **29**, 137–149 (2010).

360. van der Veldt, A. A. M. *et al.* Sunitinib-induced changes in circulating endothelial cell-related proteins in patients with metastatic renal cell cancer. *Int. J. Cancer* **131**, E484–E493 (2012).

361. Bessar, H., Kandil, A. H., Nasr, N. M. & Khattab, F. Itraconazole Versus Propranolol: Therapeutic and Pharmacologic Effect On Serum Angiopoietin-2 In Patients With Infantile Hemangioma. *Journal of Dermatological Treatment* 1–25 (2019) doi:10.1080/09546634.2019.1687822.

362. Yang, X., Wang, J. & Chen, C. Serum VEGF and Ang-2 Levels in Infants Before and After Laser Treatment for Retinopathy of Prematurity. *Fetal and Pediatric Pathology* 1–7 (2020) doi:10.1080/15513815.2020.1721625.

363. Liu, Y., Liu, M., Shi, Y. & Liu, Y. Serum Thymosin β4 Concentrations in Obstructive Sleep Apnea Syndrome Patients: Serum Thymosin β4 in OSAS Patients. *J. Clin. Lab. Anal.* **30**, 736–740 (2016).

364. Han, T. Serum thymosin β4 levels in patients with hepatitis B virus-related liver failure. *WJG* **16**, 625 (2010).

365. Song, R. *et al.* Association between serum thymosin β4 levels of rheumatoid arthritis patients and disease activity and response to therapy. *Clin Rheumatol* **31**, 1253–1258 (2012).

366. Belsky, J. B., Rivers, E. P., Filbin, M. R., Lee, P. J. & Morris, D. C. Thymosin beta 4 regulation of actin in sepsis. *Expert Opinion on Biological Therapy* **18**, 193–197 (2018).

367. Plavina, T., Hincapie, M., Wakshull, E., Subramanyam, M. & Hancock, W. S. Increased Plasma Concentrations of Cytoskeletal and Ca2+-Binding Proteins and Their Peptides in Psoriasis Patients. *Clinical Chemistry* **54**, 1805–1814 (2008).

368. Wen, Q. *et al.* Peptidomic Identification of Serum Peptides Diagnosing Preeclampsia. *PLoS One* **8**, e65571 (2013).

369. Huang, Q. *et al.* Comparative proteomic analysis of plasma of children with congenital heart disease. *Electrophoresis* **40**, 1848–1854 (2019).

370. Han, T. Serum thymosin β4 levels in patients with hepatitis B virus-related liver failure. *WJG* **16**, 625 (2010).

371. Drum, C. L. *et al.* Thymosin Beta-4 Is Elevated in Women With Heart Failure With Preserved Ejection Fraction. *JAHA* **6**, (2017).

372. Kim, H. *et al.* Identification of altered protein expression in major depressive disorder and bipolar disorder patients using liquid chromatography-tandem mass spectrometry. *Psychiatry Res* **299**, 113850 (2021).

373. Mutchnick, M. G., Lee, H. H., Hollander, D. I., Haynes, G. D. & Chua, D. C. Defective in vitro gamma interferon production and elevated serum immunoreactive thymosin beta 4 levels in patients with inflammatory bowel disease. *Clin Immunol Immunopathol* **47**, 84–92 (1988).

374. Wei, M., Duan, D., Liu, Y., Wang, Z. & Li, Z. Increased thymosin β4 levels in the serum and SF of knee osteoarthritis patients correlate with disease severity. *Regulatory Peptides* **185**, 34–36 (2013).

375. Yeşilay, A. B. *et al.* Thymosin beta4 levels after successful primary percutaneous coronary intervention for acute myocardial infarction. *Turk Kardiyol Dern Ars* **39**, 654–660 (2011).

376. Jiang, Y. *et al.* Association between thymosin beta4 and non-alcoholic fatty liver disease. *Rev Esp Enferm Dig* **111**, 308–313 (2019).

377. Rahimi, M. *et al.* Age-related obesity and type 2 diabetes dysregulate neuronal associated genes and proteins in humans. *Oncotarget* **6**, 29818–29832 (2015).

378. Magrabi, F. *et al.* Artificial Intelligence in Clinical Decision Support: Challenges for Evaluating AI and Practical Implications. *Yearb Med Inform* **28**, 128–134 (2019).

379. Ramesh, A. N., Kambhampati, C., Monson, J. R. T. & Drew, P. J. Artificial intelligence in medicine. *Ann R Coll Surg Engl* **86**, 334–338 (2004).

380. Zadeh, L. A. Fuzzy sets. *Information and Control* **8**, 338–353 (1965).

381. Goldberg, D. E. & Holland, J. H. [No title found]. *Machine Learning* **3**, 95–99 (1988).

382. Du, A. X., Emam, S. & Gniadecki, R. Review of Machine Learning in Predicting Dermatological Outcomes. *Front. Med.* **7**, 266 (2020).

383. Coiera, E. The fate of medicine in the time of AI. *The Lancet* **392**, 2331–2332 (2018).

384. Gorris, M., Hoogenboom, S. A., Wallace, M. B. & van Hooft, J. E. Artificial intelligence for the management of pancreatic diseases. *Dig Endosc* **33**, 231–241 (2021).

385. Sidey-Gibbons, J. A. M. & Sidey-Gibbons, C. J. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* **19**, 64 (2019).

386. Yu, D. & Deng, L. Deep Learning and Its Applications to Signal and Information Processing [Exploratory DSP. *IEEE Signal Process. Mag.* **28**, 145–154 (2011).

387. Jolliffe, I. Principal Component Analysis. in *Encyclopedia of Statistics in Behavioral Science* (eds. Everitt, B. S. & Howell, D. C.) bsa501 (John Wiley & Sons, Ltd, Chichester, UK, 2005). doi:10.1002/0470013192.bsa501.

388. Blei, D. M. Latent Dirichlet Allocation. 30.

389. Bzdok, D., Krzywinski, M. & Altman, N. Machine learning: supervised methods. *Nat Methods* **15**, 5–6 (2018).

390. Uddin, S., Khan, A., Hossain, M. E. & Moni, M. A. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak* **19**, 281 (2019).

391. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification And Regression Trees*. (Routledge, 2017). doi:10.1201/9781315139470.

392. Information on See5/C5.0. https://www.rulequest.com/see5-info.html.

393. Loh, W.-Y. & Shih, Y.-S. Split Selection Methods for Classification Trees. *Statistica Sinica* **7**, 24 (1997).

394. Kass, G. V. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics* **29**, 119 (1980).

395. Lee, C. Y. Representation of Switching Circuits by Binary-Decision Programs. *Bell System Technical Journal* **38**, 985–999 (1959).

396. Pang, S. & Gong, J. C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks. *Systems Engineering - Theory & Practice* **29**, 94–104 (2009).

397. Song, Y. & Lu, Y. Decision tree methods: applications for classification and prediction. **27**, 6 (2015).

398. Patel, N. & Upadhyay, S. Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA. *IJCA* **60**, 20–25 (2012).

399. IBM Docs. https://www.ibm.com/docs/en/spss-modeler/18.2.1?topic=trees-random-node (2021).

400. IBM Docs. https://www.ibm.com/docs/en/spss-modeler/18.2.1?topic=trees-decision-tree-nodes (2021).

401. Lever, J., Krzywinski, M. & Altman, N. Logistic regression. *Nat Methods* **13**, 541–542 (2016).

402. Mazzocco, T. & Hussain, A. Novel logistic regression models to aid the diagnosis of dementia. *Expert Systems with Applications* **39**, 3356–3361 (2012).

403. Ranganathan, P., Pramesh, C. S. & Aggarwal, R. Common pitfalls in statistical analysis: Logistic regression. *Perspect Clin Res* **8**, 148–151 (2017).

404. Hong, H.-C. *et al.* A panel of eight microRNAs is a good predictive parameter for triple-negative breast cancer relapse. *Theranostics* **10**, 8771–8789 (2020).

405. Chung, H. W., Kim, J. J., Choi, J. I., Lee, H. R. & Lim, J. B. A Disintegrin and Metalloproteinase 8 as a Potential Blood Biomarker for Early Diagnosis of Gastric Cancer. *Yonsei Med J* **60**, 713–719 (2019).

406. Song, H.-J. *et al.* Best serum biomarker combination for ovarian cancer classification. *Biomed Eng Online* **17**, 152 (2018).

407. Blyuss, O. *et al.* Development of PancRISK, a urine biomarker-based risk score for stratified screening of pancreatic cancer patients. *Br J Cancer* **122**, 692–696 (2020).

408. Cortes, C. & Vapnik, V. Support-vector networks. *Mach Learn* **20**, 273–297 (1995).

409. Liu, Z. & Xu, H. Kernel Parameter Selection for Support Vector Machine Classification. *Journal of Algorithms & Computational Technology* **8**, 163–177 (2014).

410. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I. Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal* **13**, 8–17 (2015).

411. Zoppis, I., Mauri, G. & Dondi, R. Kernel Methods: Support Vector Machines. in *Encyclopedia of Bioinformatics and Computational Biology* 503–510 (Elsevier, 2019). doi:10.1016/B978-0-12-809633-8.20342-7.

412. Amami, R., Ayed, D. B. & Ellouze, N. Practical Selection of SVM Supervised Parameters with Different Feature Representations for Vowel Recognition. 7.

413. Shawe-Taylor, J. & Cristianini, N. On the generalization of soft margin algorithms. *IEEE Trans. Inform. Theory* **48**, 2721–2735 (2002).

414. Yili Ren, Fuxiang Hu, & Hongping Miao. The optimization of kernel function and its parameters for SVM in well-logging. in *2016 13th International Conference on Service Systems and Service Management (ICSSSM)* 1–5 (IEEE, Kunming, China, 2016). doi:10.1109/ICSSSM.2016.7538563.

415. Malik, S., Khadgawat, R., Anand, S. & Gupta, S. Non-invasive detection of fasting blood glucose level via electrochemical measurement of saliva. *Springerplus* **5**, 701 (2016).

416. Yu, X., Yu, Y. & Zeng, Q. Support vector machine classification of streptavidin-binding aptamers. *PLoS One* **9**, e99964 (2014).

417. RBF SVM parameters. *scikit-learn* https://scikit-learn/stable/auto_examples/svm/plot_rbf_parameters.html.

418. Liu, W., Dong, J., Ma, S., Liang, L. & Yang, J. A Support Vector Machine Model Predicting the Risk of Duodenal Cancer in Patients with Familial Adenomatous Polyposis at the Transcript Levels. *Biomed Res Int* **2020**, 5807295 (2020).

419. Zhi, J., Sun, J., Wang, Z. & Ding, W. Support vector machine classifier for prediction of the metastasis of colorectal cancer. *Int J Mol Med* **41**, 1419–1426 (2018).

420. Angraal, S. *et al.* Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction. *JACC Heart Fail* **8**, 12–21 (2020).

421. Qiu, H. *et al.* Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. *BMC Med Inform Decis Mak* **20**, 83 (2020).

422. Huang, W.-C., Lee, P.-L., Liu, Y.-T., Chiang, A. A. & Lai, F. Support vector machine prediction of obstructive sleep apnea in a large-scale Chinese clinical sample. *Sleep* **43**, (2020).

423. Finkelstein, J. & Jeong, I. C. Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann N Y Acad Sci* **1387**, 153–165 (2017).

424. Wu, Y. *et al.* Use of radiomic features and support vector machine to distinguish Parkinson's disease cases from normal controls. *Ann Transl Med* **7**, 773 (2019).

425. Park, J. H. *et al.* Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *NPJ Digit Med* **3**, 46 (2020).

426. Tong, F. *et al.* Classification of the urinary metabolome using machine learning and potential applications to diagnosing interstitial cystitis. **7**, 7 (2020).

427. Yokoyama, S. *et al.* Predicted Prognosis of Patients with Pancreatic Cancer by Machine Learning. *Clin Cancer Res* **26**, 2411–2421 (2020).

428. Zhao, D. *et al.* A reliable method for colorectal cancer prediction based on feature selection and support vector machine. *Med Biol Eng Comput* **57**, 901–912 (2019).

429. Speiser, J. L., Durkalski, V. L. & Lee, W. M. Random forest classification of etiologies for an orphan disease. 13 (2014).

430. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).

431. Speiser, J. L., Miller, M. E., Tooze, J. & Ip, E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications* **134**, 93–101 (2019).

432. Couronné, R., Probst, P. & Boulesteix, A.-L. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* **19**, (2018).

433. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLOS ONE* **14**, e0224365 (2019).

434. Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861–874 (2006).

435. Adedokun, O. A. & Burgess, W. D. Analysis of Paired Dichotomous Data: A Gentle Introduction to the McNemar Test in SPSS. **8**, 7 (2012).

436. Cancer incidence statistics. *Cancer Research UK* https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence (2015).

437. Screening for cancer. *Cancer Research UK* https://www.cancerresearchuk.org/about-cancer/screening (2015).

438. Breast cancer statistics. *Cancer Research UK* https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer (2015).

439. Breast screening | Breast cancer | Cancer Research UK. https://www.cancerresearchuk.org/about-cancer/breast-cancer/getting-diagnosed/screening/breast-screening.

440. Bowel cancer screening | Bowel cancer | Cancer Research UK. https://www.cancerresearchuk.org/about-cancer/bowel-cancer/getting-diagnosed/screening.

441. Hicks, G. *et al.* Using the faecal immunochemical test in patients with rectal bleeding: Evidence from the NICE FIT Study. *Colorectal Disease* codi.15593 (2021) doi:10.1111/codi.15593.

442. About cervical screening | Cervical cancer | Cancer Research UK. https://www.cancerresearchuk.org/about-cancer/cervical-cancer/getting-diagnosed/screening/about.

443. EUROPAC trial. *Pancreatic Cancer UK* https://clinicaltrialfinderdetails.pancreaticcancer.org.uk/europac/.

444. Modlin, I. M., Kidd, M., Bodei, L., Drozdov, I. & Aslanian, H. The Clinical Utility of a Novel Blood-Based Multi-Transcriptome Assay for the Diagnosis of Neuroendocrine Tumors of the Gastrointestinal Tract: *American Journal of Gastroenterology* **110**, 1223–1232 (2015).

445. Ramachandran, R. *et al.* Comparison of the Utility of Cocaine- and Amphetamine-Regulated Transcript (CART), Chromogranin A, and Chromogranin B in Neuroendocrine Tumor Diagnosis and Assessment of Disease Progression. *The Journal of Clinical Endocrinology & Metabolism* **100**, 1520–1528 (2015).

446. Stevenson, R., Libutti, S. K. & Saif, M. W. Prognostic and Predictive Biomarkers in Gastroenteropancreatic Neuroendocrine Tumors. 3 (2013).

447. Giandomenico, V. *et al.* Improving the Diagnosis and Management of Neuroendocrine Tumors: Utilizing New Advances in Biomarker and Molecular Imaging Science. *Neuroendocrinology* **98**, 16–30 (2013).

448. Vlasveld, L. T. False elevation of chromogranin A due to proton pump inhibitors. 1.

449. Hsiao, R. J., Mezger, M. S. & O'Connor, D. T. Chromogranin A in uremia: Progressive retention of immunoreactive fragments. *Kidney International* **37**, 955–964 (1990).

450. Malczewska, A. *et al.* The clinical applications of a multigene liquid biopsy (NETest) in neuroendocrine tumors. *Advances in Medical Sciences* **65**, 18–29 (2020).

451. Pembury Smith, M. Q. R. & Ruxton, G. D. Effective use of the McNemar test. *Behav Ecol Sociobiol* **74**, 133 (2020).

452. Hajian-Tilaki, K. Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of Biomedical Informatics* **48**, 193–204 (2014).

453. Jiang, F. *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* **2**, 230–243 (2017).

454. Ben-Hur, A. & Weston, J. A User's Guide to Support Vector Machines. in *Data Mining Techniques for the Life Sciences* (eds. Carugo, O. & Eisenhaber, F.) vol. 609 223–239 (Humana Press, Totowa, NJ, 2010).

455. Tong, W. *et al.* Serum biomarker panels for diagnosis of gastric cancer. *Onco Targets Ther* **9**, 2455–2463 (2016).

456. Meng, X.-H., Huang, Y.-X., Rao, D.-P., Zhang, Q. & Liu, Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *Kaohsiung J Med Sci* **29**, 93–99 (2013).

457. Denisko, D. & Hoffman, M. M. Classification and interaction in random forests. *Proc Natl Acad Sci U S A* **115**, 1690–1692 (2018).

458. Livingstone, D. J., Manallack, D. T. & Tetko, I. V. Data modelling with neural networks: advantages and limitations. *J Comput Aided Mol Des* **11**, 135–142 (1997).

459. Pohar, M., Blas, M. & Turk, S. Comparison of logistic regression and linear discriminant analysis: A simulation study. *Adv Meth Stat* **1**, 143–161 (2004).

460. Munos, B., Niederreiter, J. & Riccaboni, M. *Improving the Prediction of Clinical Success Using Machine Learning*. http://medrxiv.org/lookup/doi/10.1101/2021.02.01.21250864 (2021) doi:10.1101/2021.02.01.21250864.

461. Hong, L. *et al.* Chromogranin A: A Valuable Serum Diagnostic Marker for Non-Insulinoma Neuroendocrine Tumors of the Pancreas in a Chinese Population. *Med Sci Monit* **26**, e926635 (2020).

462. Pulvirenti, A. *et al.* Limited role of Chromogranin A as clinical biomarker for pancreatic neuroendocrine tumors. *HPB (Oxford)* **21**, 612–618 (2019).

463. Khan, M. S. & Caplin, M. E. The use of biomarkers in neuroendocrine tumours. *Frontline Gastroenterology* **4**, 175–181 (2013).

464. Svensson, E., Muth, A., Hedenström, P. & Ragnarsson, O. The incidence of insulinoma in Western Sweden between 2002 and 2019. *Ann Gastroenterol* **35**, 434–440 (2022).

465. Peltola, E. *et al.* Characteristics and Outcomes of 79 Patients with an Insulinoma: A Nationwide Retrospective Study in Finland. *Int J Endocrinol* **2018**, 2059481 (2018).

466. Marini, F., Giusti, F., Tonelli, F. & Brandi, M. L. Pancreatic Neuroendocrine Neoplasms in Multiple Endocrine Neoplasia Type 1. *Int J Mol Sci* **22**, 4041 (2021).

467. Cingam, S. R., Botejue, M., Hoilat, G. J. & Karanchi, H. Gastrinoma. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2022).

468. Lv, Y. *et al.* Combined test of serum CgA and NSE improved the power of prognosis prediction of NF-pNETs. *Endocrine Connections* **7**, 169–178 (2018).

469. Kövesdi, A. *et al.* Circulating miRNA Increases the Diagnostic Accuracy of Chromogranin A in Metastatic Pancreatic Neuroendocrine Tumors. *Cancers (Basel)* **12**, E2488 (2020).

470. Srirajaskanthan, R. *et al.* Circulating angiopoietin-2 is elevated in patients with neuroendocrine tumours and correlates with disease burden and prognosis. *Endocrine-Related Cancer* **16**, 967–976 (2009).

471. Nichols, J. A., Herbert Chan, H. W. & Baker, M. A. B. Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophys Rev* **11**, 111–118 (2019).

472. Charilaou, P. & Battat, R. Machine learning models and over-fitting considerations. *WJG* **28**, 605–607 (2022).

473. Lee, D. *et al.* Multi-biomarker panel prediction model for diagnosis of pancreatic cancer. *J Hepatobiliary Pancreat Sci* jhbp.986 (2021) doi:10.1002/jhbp.986.

474. Thompson, I. M. *et al.* Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. *J Natl Cancer Inst* **98**, 529–534 (2006).

475. Huh, J. W., Kim, S. C., Sohn, I., Jung, S.-H. & Kim, H. C. Serum protein profiling using an aptamer array predicts clinical outcomes of stage IIA colon cancer: A leave-one-out crossvalidation. *Oncotarget* **7**, 16338–16348 (2016).

476. Zhao, Z. *et al.* External validation of a biomarker and clinical prediction model for hospital mortality in acute respiratory distress syndrome. *Intensive Care Med* **43**, 1123–1131 (2017).

477. Moslehi, S., Rabiei, N., Soltanian, A. R. & Mamani, M. Application of machine learning models based on decision trees in classifying the factors affecting mortality of COVID-19 patients in Hamadan, Iran. *BMC Med Inform Decis Mak* **22**, 192 (2022).

478. Yifan, C., Jianfeng, S. & Jun, P. Development and Validation of a Random Forest Diagnostic Model of Acute Myocardial Infarction Based on Ferroptosis-Related Genes in Circulating Endothelial Cells. *Front Cardiovasc Med* **8**, 663509 (2021).

479. IBM Docs. https://www.ibm.com/docs/en/spss-modeler/18.2.1?topic=node-svm-expert-options (2021).

480. Multiple-k: Picking the number of folds for cross-validation https://cran.r-project.org/web/packages/cvms/vignettes/picking_the_number_of_folds_for_cross-validation.html.

481. Kazerouni, F. *et al.* Type2 diabetes mellitus prediction using data mining algorithms based on the long-noncoding RNAs expression: a comparison of four data mining approaches. *BMC Bioinformatics* **21**, 372 (2020).

482. Langmead, C. *et al.* A Novel 5-Cytokine Panel Outperforms Conventional Predictive Markers of Persistent Organ Failure in Acute Pancreatitis. *Clin Transl Gastroenterol* **12**, e00351 (2021).

483. Lu, H. & Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* **249**, 126169 (2020).

484. Liang, Y. *et al.* A Deep Learning Framework to Predict Tumor Tissue-of-Origin Based on Copy Number Alteration. *Front. Bioeng. Biotechnol.* **8**, 701 (2020).

485. Zhang, F., Deng, Y. & Drabier, R. Multiple biomarker panels for early detection of breast cancer in peripheral blood. *Biomed Res Int* **2013**, 781618 (2013).

486. Taylor, J. M. G., Ankerst, D. P. & Andridge, R. R. Validation of biomarker-based risk prediction models. *Clin Cancer Res* **14**, 5977–5983 (2008).

487. Mollenhauer, B. *et al.* Validation of Serum Neurofilament Light Chain as a Biomarker of Parkinson's Disease Progression. *Mov Disord* **35**, 1999–2008 (2020).

488. O, D. F. *et al.* Technical Verification and Assessment of Independent Validation of Biomarker Models for Endometriosis. *Biomed Res Int* **2019**, 3673060 (2019).

489. Klein, E. A. *et al.* Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol* **32**, 1167–1177 (2021).

490. Trezzi, J.-P., Hiller, K. & Mollenhauer, B. The importance of an independent validation cohort for metabolomics biomarker studies. *Mov Disord* **33**, 856 (2018).

491. Menon, U. *et al.* Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* **397**, 2182–2193 (2021).

492. Mellby, L. D. *et al.* Serum Biomarker Signature-Based Liquid Biopsy for Diagnosis of Early-Stage Pancreatic Cancer. *J Clin Oncol* **36**, 2887–2894 (2018).

493. Bowden, M. *et al.* Profiling of metastatic small intestine neuroendocrine tumors reveals characteristic miRNAs detectable in plasma. *Oncotarget* **8**, 54331–54344 (2017).

494. Malczewska, A. *et al.* NETest liquid biopsy is diagnostic of small intestine and pancreatic neuroendocrine tumors and correlates with imaging. *Endocrine Connections* **8**, 442–453 (2019).

495. Matsumoto, T. *et al.* A new possible lung cancer marker: VGF detection from the conditioned medium of pulmonary large cell neuroendocrine carcinoma-derived cells using secretome analysis. *Int J Biol Markers* **24**, 282–285 (2009).

496. Llano, D. A., Devanarayan, P., Devanarayan, V., & Alzheimer's Disease Neuroimaging Initiative (ADNI). VGF in Cerebrospinal Fluid Combined With Conventional Biomarkers Enhances Prediction of Conversion From MCI to AD. *Alzheimer Dis Assoc Disord* **33**, 307–314 (2019).

497. Virreira Winter, S. *et al.* Urinary proteome profiling for stratifying patients with familial Parkinson's disease. *EMBO Mol Med* **13**, (2021).

498. Lind, A.-L. *et al.* Spinal Cord Stimulation Alters Protein Levels in the Cerebrospinal Fluid of Neuropathic Pain Patients: A Proteomic Mass Spectrometric Analysis. *Neuromodulation: Technology at the Neural Interface* **19**, 549–562 (2016).

499. Hahm, S. *et al.* Targeted Deletion of the Vgf Gene Indicates that the Encoded Secretory Peptide Precursor Plays a Novel Role in the Regulation of Energy Balance. *Neuron* **23**, 537–548 (1999).

500. Mahata, S. K., Mahata, M., Fischer-Colbrie, R. & Winkler, H. In situ hybridization: mRNA levels of secretogranin II, VGF and peptidylglycine alpha-amidating monooxygenase in brain of salt-loaded rats. *Histochemistry* **99**, 287–293 (1993).

501. Koc, G., Soyocak, A., Alis, H., Kankaya, B. & Kanigur, G. Changes in VGF and C3aR1 gene expression in human adipose tissue in obesity. *Mol Biol Rep* **48**, 251–257 (2021).

502. Annaratone, L. *et al.* Search for neuro-endocrine markers (chromogranin A, synaptophysin and VGF) in breast cancers. An integrated approach using immunohistochemistry and gene expression profiling. *Endocr Pathol* **25**, 219–228 (2014).

503. Wang, X. *et al.* Reciprocal Signaling between Glioblastoma Stem Cells and Differentiated Tumor Cells Promotes Malignant Progression. *Cell Stem Cell* **22**, 514-528.e5 (2018).

504. Brait, M. *et al.* Association of promoter methylation of VGF and PGP9.5 with ovarian cancer progression. *PLoS One* **8**, e70878 (2013).

505. Li, X. *et al.* Reduced serum VGF levels are linked with suicide risk in Chinese Han patients with major depressive disorder. *BMC Psychiatry* **20**, 225 (2020).

506. Luo, G. *et al.* Carbohydrate antigen 19-9 as a prognostic biomarker in pancreatic neuroendocrine tumors. *Oncol Lett* (2017) doi:10.3892/ol.2017.7071.

507. Poruk, K. E. *et al.* Serum osteopontin and tissue inhibitor of metalloproteinase 1 as diagnostic and prognostic biomarkers for pancreatic adenocarcinoma. *Pancreas* **42**, 193–197 (2013).

508. Luo, G. *et al.* Roles of CA19-9 in pancreatic cancer: Biomarker, predictor and promoter. *Biochim Biophys Acta Rev Cancer* **1875**, 188409 (2021).

509. Zhang, Z. *et al.* Role of angiogenesis in pancreatic cancer biology and therapy. *Biomedicine & Pharmacotherapy* **108**, 1135–1140 (2018).

510. Korse, C. M., Muller, M. & Taal, B. G. Discontinuation of proton pump inhibitors during assessment of chromogranin A levels in patients with neuroendocrine tumours. *Br J Cancer* **105**, 1173–1175 (2011).

511. Baekdal, J. *et al.* Limited Diagnostic Utility of Chromogranin A Measurements in Workup of Neuroendocrine Tumors. *Diagnostics (Basel)* **10**, E881 (2020).

512. Lim, M. *et al.* Evaluation of CisBio ELISA for Chromogranin A Measurement. *J Appl Lab Med* **4**, 11–18 (2019).

513. Di Giacinto, P. *et al.* Chromogranin A: From Laboratory to Clinical Aspects of Patients with Neuroendocrine Tumors. *International Journal of Endocrinology* **2018**, 1–12 (2018).

514. Marotta, V. *et al.* Chromogranin A as circulating marker for diagnosis and management of neuroendocrine neoplasms: more flaws than fame. *Endocrine-Related Cancer* **25**, R11–R29 (2018).

515. Pedersen, L. & Nybo, M. Preanalytical factors of importance for measurement of Chromogranin A. *Clin Chim Acta* **436**, 41–44 (2014).

516. Kidd, M., Bodei, L. & Modlin, I. M. Chromogranin A: any relevance in neuroendocrine tumors? *Curr Opin Endocrinol Diabetes Obes* **23**, 28–37 (2016).

517. Hijioka, M. *et al.* Serum chromogranin A is a useful marker for Japanese patients with pancreatic neuroendocrine tumors. *Cancer Sci* **105**, 1464–1471 (2014).

518. Jianu, C. S., Fossmark, R., Syversen, U., Hauso, Ø. & Waldum, H. L. A meal test improves the specificity of chromogranin A as a marker of neuroendocrine neoplasia. *Tumor Biol.* **31**, 373–380 (2010).

519. Tseng, C. *et al.* Low accuracy of chromogranin A for diagnosing early-stage pancreatic neuroendocrine tumors. *Oncology Letters* (2018) doi:10.3892/ol.2018.8472.

520. Jilesen, A. P. J., Busch, O. R. C., van Gulik, T. M., Gouma, D. J. & Nieveen van Dijkum, E. J. M. Standard pre- and postoperative determination of chromogranin a in resectable non-functioning pancreatic neuroendocrine tumors--diagnostic accuracy: NF-pNET and low tumor burden. *Dig Surg* **31**, 407–414 (2014).

521. Miki, M. *et al.* Utility of chromogranin B compared with chromogranin A as a biomarker in Japanese patients with pancreatic neuroendocrine tumors. *Japanese Journal of Clinical Oncology* **47**, 520–528 (2017).

522. Wang, N. *et al.* Application of artificial neural network model in diagnosis of Alzheimer's disease. *BMC Neurol* **19**, 154 (2019).

523. Zhao, S. *et al.* Identification of Diagnostic Markers for Major Depressive Disorder Using Machine Learning Methods. *Front Neurosci* **15**, 645998 (2021).

524. Xu, F. *et al.* A Linear Discriminant Analysis Model Based on the Changes of 7 Proteins in Plasma Predicts Response to Anlotinib Therapy in Advanced Non-Small Cell Lung Cancer Patients. *Front. Oncol.* **11**, 756902 (2022).

525. Shao, Y. *et al.* Shedding Light on the Black Box: Explaining Deep Neural Network Prediction of Clinical Outcomes. *J Med Syst* **45**, 5 (2021).

526. Discriminant Analysis Classification - MATLAB & Simulink - MathWorks United Kingdom. https://uk.mathworks.com/help/stats/discriminant-analysis.html.

527. Sharma, C. Advances in diagnosis, treatment and palliation of pancreatic carcinoma: 1990-2010. *WJG* **17**, 867 (2011).

528. Takikawa, T. *et al.* Clinical features and prognostic impact of asymptomatic pancreatic cancer. *Sci Rep* **12**, 4262 (2022).

529. Pancreatic neuroendocrine cancers (PancNETs). *Pancreatic Cancer UK* https://www.pancreaticcancer.org.uk/information/just-diagnosed-with-pancreatic-cancer/pancreatic-neuroendocrine-tumours-pancnets/.

530. Raman, S. P., Hruban, R. H., Cameron, J. L., Wolfgang, C. L. & Fishman, E. K. Pancreatic imaging mimics: part 2, pancreatic neuroendocrine tumors and their mimics. *AJR Am J Roentgenol* **199**, 309–318 (2012).

531. Al-Hawary, M. M., Kaza, R. K., Azar, S. F., Ruma, J. A. & Francis, I. R. Mimics of pancreatic ductal adenocarcinoma. *Cancer Imaging* **13**, 342–349 (2013).

532. El Imad, T., Haddad, F. G., Kesavan, M., Deeb, L. & Andrawes, S. Solid Pseudopapillary Tumor of the Pancreas: An Unusual Cause of Abdominal Pain. *Cureus* **9**, e1252 (2017).

533. Shulik, O., Cavanagh, Y. & Grossman, M. Pancreatic Lesion: Malignancy or Abscess? *Am J Case Rep* **17**, 337–339 (2016).

534. Chase, M. P., Yarze, J. C., Gumustop, B. & Leach, R. P. Endoscopic ultrasound-guided aspiration and oral antibiotic therapy as definitive treatment of an asymptomatic pancreatic abscess. *Pancreas* **38**, 475–476 (2009).

535. Kim, M. J. *et al.* Pyogenic pancreatic abscess mimicking pancreatic neoplasm: a four-case series. *Korean J Gastroenterol* **65**, 252–257 (2015).

536. Wolske, K. M. *et al.* Chronic Pancreatitis or Pancreatic Tumor? A Problem-solving Approach. *RadioGraphics* **39**, 1965–1982 (2019).

537. Kennedy, T. *et al.* Incidence of benign inflammatory disease in patients undergoing Whipple procedure for clinically suspected carcinoma: a single-institution experience. *Am J Surg* **191**, 437–441 (2006).

538. Goulden, M. R. The pain of chronic pancreatitis: a persistent clinical challenge. *British Journal of Pain* **7**, 8–22 (2013).

539. Heo, J. Infected Pancreatic Necrosis Mimicking Pancreatic Cancer. *Case Rep Gastroenterol* **14**, 436–442 (2020).

540. Birnbaum, D. J. *et al.* Surgery for pancreatic neoplasms: How accurate are our surgical indications? *Surgery* **162**, 112–119 (2017).

541. Gerritsen, A. *et al.* Preoperative characteristics of patients with presumed pancreatic cancer but ultimately benign disease: a multicenter series of 344 pancreatoduodenectomies. *Ann Surg Oncol* **21**, 3999–4006 (2014).

542. Umans, D. S. *et al.* Pancreatitis and pancreatic cancer: A case of the chicken or the egg. *World J Gastroenterol* **27**, 3148–3157 (2021).

543. Sarner, M. & Cotton, P. B. Classification of pancreatitis. *Gut* **25**, 756–759 (1984).

544. Kirkegård, J., Mortensen, F. V., Heide-Jørgensen, U. & Cronin-Fenton, D. Predictors of underlying pancreatic cancer in patients with acute pancreatitis: a Danish nationwide cohort study. *HPB* **22**, 553–562 (2020).

545. Munigala, S., Kanwal, F., Xian, H. & Agarwal, B. New Diagnosis of Chronic Pancreatitis: Risk of Missing an Underlying Pancreatic Cancer. *American Journal of Gastroenterology* **109**, 1824–1830 (2014).

546. Furuya, N. *et al.* Comparative study of CA242 and CA19-9 in chronic pancreatitis. *Br J Cancer* **73**, 372–376 (1996).

547. Mosli, H. H., Dennis, A., Kocha, W., Asher, L. J. & Van Uum, S. H. M. Effect of short-term proton pump inhibitor treatment and its discontinuation on

chromogranin A in healthy subjects. *J Clin Endocrinol Metab* **97**, E1731-1735 (2012).

548. Juusola, M., Mustonen, H., Saharinen, P., Haglund, C. & Seppänen, H. The Dual Role of Angiopoietin 2 in Pancreatic Cancer. *HPB* **23**, S886 (2021).

549. Schulz, P. *et al.* Angiopoietin-2 drives lymphatic metastasis of pancreatic cancer. *FASEB J* **25**, 3325–3335 (2011).

550. Riedl, M. S. *et al.* Proteomic analysis uncovers novel actions of the neurosecretory protein VGF in nociceptive processing. *J Neurosci* **29**, 13377–13388 (2009).

551. Rizzi, R. *et al.* The VGF-derived peptide TLQP-21: A new modulatory peptide for inflammatory pain. *Neuroscience Letters* **441**, 129–133 (2008).

552. Alrawashdeh, W. *et al.* Perineural invasion in pancreatic cancer: proteomic analysis and in vitro modelling. *Mol Oncol* **13**, 1075–1091 (2019).

553. Oberg, K. & Jelic, S. Neuroendocrine gastroenteropancreatic tumors: ESMO Clinical Recommendations for diagnosis, treatment and follow-up. *Annals of Oncology* **19**, ii104–ii105 (2008).

554. Quinn, J. P., Kandigian, S. E., Trombetta, B. A., Arnold, S. E. & Carlyle, B. C. VGF as a biomarker and therapeutic target in neurodegenerative and psychiatric diseases. *Brain Communications* **3**, fcab261 (2021).

555. Ren, A. H., Fiala, C. A., Diamandis, E. P. & Kulasingam, V. Pitfalls in Cancer Biomarker Discovery and Validation with Emphasis on Circulating Tumor DNA. *Cancer Epidemiology, Biomarkers & Prevention* **29**, 2568–2574 (2020).

556. Yang, L. *et al.* Neuronal survival factor VGF promotes chemoresistance and predicts poor prognosis in lung cancers with neuroendocrine feature. *Intl Journal of Cancer* **151**, 1611–1625 (2022).

557. Lee, D. W., Kim, M. K. & Kim, H. G. Diagnosis of Pancreatic Neuroendocrine Tumors. *Clinical Endoscopy* **50**, 537–545 (2017).

558. McKenna, L. R. & Edil, B. H. Update on pancreatic neuroendocrine tumors. *Gland Surg* **3**, 258–275 (2014).

559. Hopper, A. D., Jalal, M. & Munir, A. Recent advances in the diagnosis and management of pancreatic neuroendocrine tumours. *Frontline Gastroenterol* **10**, 269–274 (2019).

560. Neal, R. D. *et al.* Cell-Free DNA–Based Multi-Cancer Early Detection Test in an Asymptomatic Screening Population (NHS-Galleri): Design of a Pragmatic, Prospective Randomised Controlled Trial. *Cancers* **14**, 4818 (2022).

561. Seregni, E., Ferrari, L., Bajetta, E., Martinetti, A. & Bombardieri, E. Clinical significance of blood chromogranin A measurement in neuroendocrine tumours. *Annals of Oncology* **12**, S69–S72 (2001).

562. Baudin, E. *et al.* Neuron-specific enolase and chromogranin A as markers of neuroendocrine tumours. *Br J Cancer* **78**, 1102–1107 (1998).

563. Kulasingam, V., Prassas, I. & Diamandis, E. P. Towards personalized tumor markers. *npj Precision Onc* **1**, 17 (2017).

564. Baker, S. G. Improving the biomarker pipeline to develop and evaluate cancer screening tests. *J Natl Cancer Inst* **101**, 1116–1119 (2009).

565. Sonnenschein, C. & Soto, A. M. Theories of carcinogenesis: an emerging perspective. *Semin Cancer Biol* **18**, 372–377 (2008).

566. Benten, D. *et al.* Establishment of the First Well-differentiated Human Pancreatic Neuroendocrine Tumor Model. *Molecular Cancer Research* **16**, 496–507 (2018).

567. Lines, K. E. *et al.* A MEN1 pancreatic neuroendocrine tumour mouse model under temporal control. *Endocrine Connections* **6**, 232–242 (2017).

568. Chamberlain, C. E. *et al.* A Patient-derived Xenograft Model of Pancreatic Neuroendocrine Tumors Identifies Sapanisertib as a Possible New Treatment for Everolimus-resistant Tumors. *Mol Cancer Ther* **17**, 2702–2709 (2018).

569. Bresciani, G. *et al.* Evaluation of Spheroid 3D Culture Methods to Study a Pancreatic Neuroendocrine Neoplasm Cell Line. *Front Endocrinol (Lausanne)* **10**, 682 (2019).

570. April-Monn, S. L. *et al.* Three-Dimensional Primary Cell Culture: A Novel Preclinical Model for Pancreatic Neuroendocrine Tumors. *Neuroendocrinology* **111**, 273–287 (2021).

571. Kawasaki, K., Fujii, M. & Sato, T. Gastroenteropancreatic neuroendocrine neoplasms: genes, therapies and models. *Dis Model Mech* **11**, dmm029595 (2018).

572. Kaku, M., Nishiyama, T., Yagawa, K. & Abe, M. Establishment of a carcinoembryonic antigen-producing cell line from human pancreatic carcinoma. *Gan* **71**, 596–601 (1980).

573. Evers, B. M. *et al.* Establishment and characterization of a human carcinoid in nude mice and effect of various agents on tumor growth. *Gastroenterology* **101**, 303–311 (1991).

574. Taelman, V. F. *et al.* Upregulation of Key Molecules for Targeted Imaging and Therapy. *Journal of Nuclear Medicine* **57**, 1805–1810 (2016).

575. Leu, F. P., Nandi, M. & Niu, C. The Effect of Transforming Growth Factor on Human Neuroendocrine Tumor BON Cell Proliferation and Differentiation Is Mediated through Somatostatin Signaling. *Molecular Cancer Research* **6**, 1029–1042 (2008).

576. Habanjar, O., Diab-Assaf, M., Caldefie-Chezet, F. & Delort, L. 3D Cell Culture Systems: Tumor Application, Advantages, and Disadvantages. *Int J Mol Sci* **22**, 12200 (2021).

577. Basnayake, B. W. M. T. J. *et al.* Head and neck cancer patient-derived tumouroid cultures: opportunities and challenges. *Br J Cancer* **128**, 1807–1818 (2023).

578. Jensen, C. & Teng, Y. Is It Time to Start Transitioning From 2D to 3D Cell Culture? *Front. Mol. Biosci.* **7**, 33 (2020).

579. Gulde, S. *et al.* Combined Targeting of Pathogenetic Mechanisms in Pancreatic Neuroendocrine Tumors Elicits Synergistic Antitumor Effects. *Cancers (Basel)* **14**, 5481 (2022).

580. Halfdanarson, T. R., Rabe, K. G., Rubin, J. & Petersen, G. M. Pancreatic neuroendocrine tumors (PNETs): incidence, prognosis and recent trend toward improved survival. *Annals of Oncology* **19**, 1727–1733 (2008).

581. Moffat, K. & Mercer, S. W. Challenges of managing people with multimorbidity in today's healthcare systems. *BMC Fam Pract* **16**, 129 (2015).

582. McLerran, D. *et al.* Analytical validation of serum proteomic profiling for diagnosis of prostate cancer: sources of sample bias. *Clin Chem* **54**, 44–52 (2008).

583. Sansone, A. *et al.* Specific and Non-Specific Biomarkers in Neuroendocrine Gastroenteropancreatic Tumors. *Cancers* **11**, 1113 (2019).

584. Benz, E. J. The Jeremiah Metzger Lecture Cancer in the Twenty-First Century: An Inside View from an Outsider. *Trans Am Clin Climatol Assoc* **128**, 275–297 (2017).

585. Davis-Marcisak, E. F. *et al.* Differential Variation Analysis Enables Detection of Tumor Heterogeneity Using Single-Cell RNA-Sequencing Data. *Cancer Research* **79**, 5102–5112 (2019).

586. Khan, M. S. *et al.* Circulating Tumor Cells As Prognostic Markers in Neuroendocrine Tumors. *JCO* **31**, 365–372 (2013).

587. Oberg, K. *et al.* A Delphic consensus assessment: imaging and biomarkers in gastroenteropancreatic neuroendocrine tumor disease management. *Endocrine Connections* **5**, 174–187 (2016).

588. Thiis-Evensen, E. *et al.* Plasma protein biomarkers for the detection of pancreatic neuroendocrine tumors and differentiation from small intestinal neuroendocrine tumors. *J Neuroendocrinology* (2022) doi:10.1111/jne.13176.

589. Rustagi, S., Warner, R. R. P. & Divino, C. M. Serum pancreastatin: The next predictive neuroendocrine tumor marker: Pancreastatin: The Next NET Biomarker? *J. Surg. Oncol.* **108**, 126–128 (2013).

590. Sherman, S. K., Maxwell, J. E., O'Dorisio, M. S., O'Dorisio, T. M. & Howe, J. R. Pancreastatin predicts survival in neuroendocrine tumors. *Ann Surg Oncol* **21**, 2971–2980 (2014).

591. Niederle, B. *et al.* ENETS Consensus Guidelines Update for Neuroendocrine Neoplasms of the Jejunum and Ileum. *Neuroendocrinology* **103**, 125–138 (2016).

592. O'Toole, D. *et al.* ENETS Consensus Guidelines for the Standards of Care in Neuroendocrine Tumors: biochemical markers. *Neuroendocrinology* **90**, 194–202 (2009).

593. Mashige, F. *et al.* Acidic Catecholamine Metabolites and 5-Hydroxyindoleacetic Acid in Urine: The Influence of Diet. *Ann Clin Biochem* **33**, 43–49 (1996).

594. Milanetto, A. C., Fassan, M., David, A. & Pasquali, C. Serotonin-Secreting Neuroendocrine Tumours of the Pancreas. *J Clin Med* **9**, 1363 (2020).

595. Kalligeros, M., Diamantopoulos, L. & Toumpanakis, C. Biomarkers in Small Intestine NETs and Carcinoid Heart Disease: A Comprehensive Review. *Biology* **10**, 950 (2021).

596. Khan, M. S. *et al.* Circulating Tumor Cells and EpCAM Expression in Neuroendocrine Tumors. *Clinical Cancer Research* **17**, 337–345 (2011).

597. Allard, W. J. *et al.* Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin Cancer Res* **10**, 6897–6904 (2004).

598. Zhang, H. *et al.* Detection Methods and Clinical Applications of Circulating Tumor Cells in Breast Cancer. *Front. Oncol.* **11**, 652253 (2021).

599. Cristofanilli, M. & Fortina, P. Circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **369**, 93 (2013).

600. Ivanova, E., Ward, A., Wiegmans, A. P. & Richard, D. J. Circulating Tumor Cells in Metastatic Breast Cancer: From Genome Instability to Metastasis. *Front. Mol. Biosci.* **7**, 134 (2020).

601. Diaz, L. A. & Bardelli, A. Liquid Biopsies: Genotyping Circulating Tumor DNA. *JCO* **32**, 579–586 (2014).

602. Modlin, I. M. *et al.* Molecular Genomic Assessment Using a Blood-based mRNA Signature (NETest) is Cost-effective and Predicts Neuroendocrine Tumor Recurrence With 94% Accuracy. *Annals of Surgery* **274**, 481–490 (2021).

603. Laskaratos, F.-M. *et al.* Evaluation of circulating transcript analysis (NETest) in small intestinal neuroendocrine neoplasms after surgical resection. *Endocrine* **69**, 430–440 (2020).

604. Malczewska, A. *et al.* The NETest liquid biopsy is diagnostic for gastric neuroendocrine tumors: observations on the blood-based identification of microscopic and macroscopic residual disease. *BMC Gastroenterol* **20**, 235 (2020).

605. Kassi, E. & Kaltsas, G. Plasma Polyamines as an Additional to Imaging Biomarker in MEN1 Patients With Duodenopancreatic Neuroendocrine Tumors. *The Journal of Clinical Endocrinology & Metabolism* **107**, e880–e882 (2022).

606. Fahrmann, J. F. *et al.* A Blood-based Polyamine Signature Associated With MEN1 Duodenopancreatic Neuroendocrine Tumor Progression. *The Journal of Clinical Endocrinology & Metabolism* dgab554 (2021) doi:10.1210/clinem/dgab554.

607. Oversoe, S. K., Sorensen, B. S., Tabaksblat, E. M., Gronbaek, H. & Kelsen, J. Cell-Free DNA and Clinical Characteristics in Patients with Small Intestinal or Pancreatic Neuroendocrine Tumors. *Neuroendocrinology* **112**, 43–50 (2022).

608. Cisneros-Villanueva, M. *et al.* Cell-free DNA analysis in current cancer clinical trials: a review. *Br J Cancer* **126**, 391–400 (2022).

609. Shi, Y. *et al.* Cell Cycle Protein Expression in Neuroendocrine Tumors: Association of CDK4/CDK6, CCND1, and Phosphorylated Retinoblastoma Protein With Proliferative Index. *Pancreas* **46**, 1347–1353 (2017).

610. Satoh, K. *et al.* Tumor necrosis factor-related apoptosis-inducing ligand and its receptor expression and the pathway of apoptosis in human pancreatic cancer. *Pancreas* **23**, 251–258 (2001).

611. Pan, K., Huang, X. & Jia, X. SPARC promotes pancreatic cancer cell proliferation and migration through autocrine secretion into the extracellular milieu. *Oncol Lett* **21**, 485 (2021).

612. Razidlo, G. L., Burton, K. M. & McNiven, M. A. Interleukin-6 promotes pancreatic cancer cell migration by rapidly activating the small GTPase CDC42. *Journal of Biological Chemistry* **293**, 11143–11153 (2018).

613. Okada, S. *et al.* Elevated Serum Interleukin-6 Levels in Patients with Pancreatic Cancer. *Japanese Journal of Clinical Oncology* **28**, 12–15 (1998).

614. Murthy, S. R. K. *et al.* Carboxypeptidase E promotes cancer cell survival, but inhibits migration and invasion. *Cancer Letters* **341**, 204–213 (2013).

615. Hua, Q. *et al.* Upregulation of KLK8 Predicts Poor Prognosis in Pancreatic Cancer. *Front Oncol* **11**, 624837 (2021).

616. Zhou, B. *et al.* High vimentin expression with E-cadherin expression loss predicts a poor prognosis after resection of grade 1 and 2 pancreatic neuroendocrine tumors. *BMC Cancer* **21**, 334 (2021).

617. DeLucia, D. C. *et al.* Regulation of CEACAM5 and Therapeutic Efficacy of an Anti-CEACAM5-SN38 Antibody-drug Conjugate in Neuroendocrine Prostate Cancer. *Clin Cancer Res* **27**, 759–774 (2021).

618. Xie, J. *et al.* In Vivo Imaging of Methionine Aminopeptidase II for Prostate Cancer Risk Stratification. *Cancer Res* **81**, 2510–2521 (2021).

619. Zhang, L. *et al.* Integrin Beta 5 Is a Prognostic Biomarker and Potential Therapeutic Target in Glioblastoma. *Front. Oncol.* **9**, 904 (2019).

620. Manevich, L. *et al.* Glycoprotein NMB promotes tumor formation and malignant progression of laryngeal squamous cell carcinoma. *Cancer Sci* **113**, 3244–3254 (2022).

621. Tsukamoto, N. *et al.* The expression of S100A4 in human pancreatic cancer is associated with invasion. *Pancreas* **42**, 1027–1033 (2013).

622. Hofving, T. *et al.* The Microenvironment of Small Intestinal Neuroendocrine Tumours Contains Lymphocytes Capable of Recognition and Activation after Expansion. *Cancers (Basel)* **13**, 4305 (2021).

623. García-Suárez, O., García, B., Fernández-Vega, I., Astudillo, A. & Quirós, L. M. Neuroendocrine tumors show altered expression of chondroitin sulfate, glypican 1, glypican 5, and syndecan 2 depending on their differentiation grade. *Front Oncol* **4**, 15 (2014).

624. Plan, N. L. T. NHS Long Term Plan. *NHS Long Term Plan* https://www.longtermplan.nhs.uk (2020).

625. Malczewska, A. *et al.* NETest liquid biopsy is diagnostic of small intestine and pancreatic neuroendocrine tumors and correlates with imaging. *Endocr Connect* EC-19-0030.R1 (2019) doi:10.1530/EC-19-0030.

626. Supporting GPs with electronic cancer decision support (CDS) tools. *NHS Transformation Directorate* https://transform.england.nhs.uk/key-tools-and-info/digital-playbooks/cancer-digital-playbook/supporting-gps-with-electronic-cancer-decision-support-tools/.

Appendix A: C5.0 non-boosted accuracy decision tree results

Biomarkers entered	Biomarkers used	Predictor Importance	AUC	Gini
C	C	C=1.00	0.813	0.627
A	A	A=1.00	0.767	0.534
V	V	V=1.00	0.681	0.361
M	None	Not Available	0.500	0.000
N	None	Not Available	0.500	0.000
T	None	Not Available	0.500	0.000
P	None	Not Available	0.500	0.000

Table A1: C5.0 accuracy non-boosted decision tree results for the 1 marker combination.
Results obtained when a single marker was entered to create a C5.0 accuracy non-boosted decision tree models using the whole training dataset (n=107). SPSS Modeler was used for model creation. Three models were created.

Biomarkers entered	Biomarkers used in the model	Predictor Importance	AUC	Gini
CM	CM	C=0.90 M=0.10	0.850	0.700
CA	AC	A=0.68 C=0.32	0.835	0.670
CN	C	C=1.00	0.813	0.627
CP	C	C=1.00	0.813	0.627
CT	C	C=1.00	0.813	0.627
AM	A	A=1.00	0.767	0.534
AN	A	A=1.00	0.767	0.534
AP	A	A=1.00	0.767	0.534
AT	A	A=1.00	0.767	0.534
AV	A	A=1.00	0.767	0.534
CV	C	C=1.00	0.751	0.534
NV	V	V=1.00	0.681	0.361
TV	V	V=1.00	0.681	0.361
MV	V	V=1.00	0.681	0.361
VP	V	V=1.00	0.681	0.361
MN	None	Not Available	0.500	0.000
MP	None	Not Available	0.500	0.000
MT	None	Not Available	0.500	0.000
PN	None	Not Available	0.500	0.000
TN	None	Not Available	0.500	0.000
TP	None	Not Available	0.500	0.000

Table A2: C5.0 accuracy non-boosted decision tree results for 2 marker combinations of markers.
Results obtained when 2 marker combinations were entered to create a C5.0 accuracy non-boosted decision tree models using the whole training dataset n=107. SPSS Modeler was used for model creation. The best performing 2 marker combination was the CM combination with an AUC of 0.850.

Biomarkers entered	Biomarkers used in the model	Predictor Importance	AUC	Gini
CMP	CM	C=0.90 M=0.10	0.850	0.700
CMT	CM	C=0.90 M=0.10	0.850	0.700
MNC	CM	C=0.90 M=0.10	0.850	0.700
PVC	CV	C=0.80 V=0.20	0.815	0.629

PAV	AVP	A=0.66 V=0.20 P=0.15	0.844	0.688
AMP	AMP	A=0.75 M=0.21 P=0.04	0.844	0.689
CAV	AC	A=0.68 C=0.32	0.835	0.670
CAM	AC	A=0.68 C=0.32	0.835	0.670
CAP	AC	A=0.68 C=0.32	0.835	0.670
CTA	AC	A=0.68 C=0.32	0.835	0.670
CAN	AC	A=0.68 C=0.32	0.835	0.670
MVC	CV	C=0.80 V=0.20	0.815	0.629
CTP	C	C=1.00	0.813	0.627
CNT	C	C=1.00	0.813	0.627
CNP	C	C=1.00	0.813	0.627
VAM	A	A=1.00	0.767	0.534
ATP	A	A=1.00	0.767	0.534
VAT	A	A=1.00	0.767	0.534
TNA	A	A=1.00	0.767	0.534
TAM	A	A=1.00	0.767	0.534
MNA	A	A=1.00	0.767	0.534
PAN	A	A=1.00	0.767	0.534
VAN	A	A=1.00	0.767	0.534
CVT	C	C=1.00	0.751	0.502
CNV	C	C=1.00	0.751	0.502
TNV	V	V=1.00	0.681	0.361
VTP	V	V=1.00	0.681	0.361
MNV	V	V=1.00	0.681	0.361
PNV	V	V=1.00	0.681	0.361
MVT	V	V=1.00	0.681	0.361
VMP	V	V=1.00	0.681	0.361
MTP	None	Not Available	0.500	0.000
TNP	None	Not Available	0.500	0.000
NMP	None	Not Available	0.500	0.000
NMT	None	Not Available	0.500	0.000

Table A3: C5.0 accuracy non-boosted decision tree results for 3 marker combinations of markers.

Results obtained when 3 marker combinations were entered to create a C5.0 accuracy non-boosted decision tree models using the whole training dataset n=107. The best performing model despite 3 markers being entered was for the CMP,CMT and MNC combinations where only 2 markers (C and M) were used the model with an AUC of 0.850.

Biomarkers entered	Biomarkers used in the models	Predictor Importance	AUC	Gini
TAMP	AMTP	A=0.79 M=0.14 T=0.06 P=0.01	0.867	0.734
AMVP	AVPM	A=0.59 P=0.25 V=0.14 M=0.02	0.864	0.729
CMPN	CM	C=0.90 M=0.10	0.850	0.700
CMTP	CM	C=0.90 M=0.10	0.850	0.700
CTMN	CM	C=0.90 M=0.10	0.850	0.700
AMPN	AMP	A=0.75 M=0.21 P=0.04	0.844	0.689
AVPN	AVP	A=0.66 V=0.20 P=0.15	0.844	0.688
TAPV	AVP	A=0.66 V=0.20 P=0.15	0.844	0.688
CATP	AC	A=0.68 C=0.32	0.835	0.670
CAPV	AC	A=0.68 C=0.32	0.835	0.670
CTAN	AC	A=0.68 C=0.32	0.835	0.670
AMVC	AC	A=0.68 C=0.32	0.835	0.670
AMCN	AC	A=0.68 C=0.32	0.835	0.670

ACPN	AC	A=0.68 C=0.32	0.835	0.670
CTAV	AC	A=0.68 C=0.32	0.835	0.670
CTAM	AC	A=0.68 C=0.32	0.835	0.670
CAMP	AC	A=0.68 C=0.32	0.835	0.670
CTAV	AC	A=0.68 C=0.32	0.835	0.670
AVCN	AC	A=0.68 C=0.32	0.835	0.670
CVTP	CV	C=0.80 V=0.20	0.815	0.629
CMNV	CV	C=0.80 V=0.20	0.815	0.629
TMVC	CV	C=0.80 V=0.20	0.815	0.629
CNVP	CV	C=0.80 V=0.20	0.815	0.629
CMVP	CV	C=0.80 V=0.20	0.815	0.629
CNTP	C	C=1.00	0.813	0.627
TNMA	A	A=1.00	0.767	0.534
TAVN	A	A=1.00	0.767	0.534
AMVT	A	A=1.00	0.767	0.534
AMVN	A	A=1.00	0.767	0.534
ATPN	A	A=1.00	0.767	0.534
CTVN	C	C=1.00	0.751	0.502
NMVP	V	V=1.00	0.681	0.361
TNVP	V	V=1.00	0.681	0.361
TMNV	V	V=1.00	0.681	0.361
TMPV	V	V=1.00	0.681	0.361
TNMP	None	Not Available	0.500	0.000

Table A4: C5.0 accuracy non-boosted decision tree results for 4 marker combinations of markers.

Results obtained when 4 marker combinations the biomarkers were entered to create a

C5.0 accuracy non-boosted decision tree models using the whole training dataset n=107.
SPSS Modeler was used for model creation. 35 combinations were carried out.

Biomarkers entered	Biomarkers used	Predictor Importance	AUC	Gini
AMNTP	APMT	A=0.69 M=0.13 P=0.10 T=0.08	0.871	0.742
VTAMP	APVM	A=0.59 P=0.25 V=0.14 M=0.02	0.864	0.729
VNAMP	APVM	A=0.59 P=0.25 V=0.14 M=0.02	0.864	0.729
TNPMC	CM	C=0.90 M=0.10	0.850	0.700
TAPVN	AVP	A=0.66 V=0.20 P=0.15	0.844	0.688
CTAPN	AC	A=0.68 C=0.32	0.835	0.670
VAPMC	AC	A=0.68 C=0.32	0.835	0.670
AVPNC	AC	A=0.68 C=0.32	0.835	0.670
ANPMC	AC	A=0.68 C=0.32	0.835	0.670
AVMNC	AC	A=0.68 C=0.32	0.835	0.670
CTAPV	AC	A=0.68 C=0.32	0.835	0.670
AVTNC	AC	A=0.68 C=0.32	0.835	0.670
CTAMP	AC	A=0.68 C=0.32	0.835	0.670
CTAMV	AC	A=0.68 C=0.32	0.835	0.670
CTAMN	AC	A=0.68 C=0.32	0.835	0.670
CTMPV	CV	C=0.80 V=0.20	0.815	0.670
TVMNC	CV	C=0.80 V=0.20	0.815	0.670
CTPVN	CV	C=0.80 V=0.20	0.815	0.670
VNPMC	CV	C=0.80 V=0.20	0.815	0.670
AMNTV	A	A=1.00	0.767	0.534
TMPVN	V	V=1.00	0.681	0.361

Table A5: C5.0 accuracy non-boosted decision tree results for 5 marker combinations of markers.

Results obtained when 5 marker combinations of the biomarkers were entered to create a C5.0 accuracy non-boosted decision tree models using the whole training dataset n=107.

SPSS Modeler was used for model creation. 21 combinations were carried out. The best

model obtained was when the AMNTP markers were entered and APMT were used.

The AUC associated with this was 0.871.

Biomarkers entered	Biomarkers used in the model	Predictor Importance	AUC	Gini
TVAMNP	AVPM	A=0.59 P=0.25 V=0.14 M=0.02	0.864	0.729
CVAMNP	AC	A=0.68 C=0.32	0.835	0.670
CATMNP	AC	A=0.68 C=0.32	0.835	0.670
CTAMNV	AC	A=0.68 C=0.32	0.835	0.670
CVMPTA	AC	A=0.68 C=0.32	0.835	0.670
CVNPTA	AC	A=0.68 C=0.32	0.835	0.670
CVMNPT	CV	C=0.80 V=0.20	0.815	0.629

Table A6: C5.0 accuracy non-boosted decision tree results for 6 marker combinations of markers.

Results obtained when 6 marker combinations of the biomarkers were entered to create a C5.0 accuracy non-boosted decision tree models using the whole training dataset n=107.

SPSS Modeler was used for model creation. 7 combinations in total were carried out. The best

model obtained was when the TVAMNP combination was entered, however only 4 markers AVPM were used in the model. The AUC for this model was 0.864.

Biomarkers entered	Biomarkers used in the model	Predictor Importance	AUC	Gini
CTAMNVP	AC	A=0.68 C=0.32	0.835	0.670

Table A7: Table illustrating the result obtained when the 7 marker combination

was entered to create the C5.0 accuracy non-boosted model.

SPSS Modeler was used for model creation. This model only used the AC markers and had an AUC of 0.835.

Appendix B: RF results for the 4 to 7-marker combinations

4 marker	Average AUC	STEDEV	Range
CVAM	1.000	0.000	0.999-1.000
CANM	0.999	0.001	0.997-1.000
CPNM	0.999	0.001	0.997-1.000
CVAP	0.999	0.001	0.996-1.000
CVNM	0.999	0.001	0.997-1.000
CVPN	0.999	0.001	0.997-1.000
TCAN	0.999	0.001	0.996-1.000
TCAP	0.999	0.001	0.997-1.000
TCNM	0.999	0.001	0.996-1.000
TCVA	0.999	0.001	0.996-1.000
TCVP	0.999	0.001	0.997-1.000
VPNM	0.999	0.002	0.995-1.000
CAPN	0.999	0.001	0.997-1.000
CVPM	0.999	0.001	0.996-1.000
TCVM	0.999	0.002	0.995-1.000
TCAM	0.999	0.002	0.994-1.000
APNM	0.998	0.002	0.994-1.000
CAPM	0.998	0.001	0.996-1.000
CVAN	0.998	0.002	0.992-1.000
TCPM	0.998	0.002	0.992-1.000
TNPM	0.998	0.003	0.992-1.000
TCPN	0.998	0.002	0.996-1.000
TCVN	0.998	0.001	0.997-1.000
TVAM	0.998	0.002	0.995-1.000
TVAN	0.998	0.001	0.996-1.000
VANM	0.998	0.003	0.989-1.000
VAPN	0.998	0.002	0.995-1.000
TANM	0.997	0.003	0.991-1.000
TVAP	0.997	0.003	0.990-1.000
TVNM	0.997	0.003	0.990-1.000
TVPM	0.997	0.004	0.996-1.000
TVPN	0.997	0.004	0.993-1.000
VAPM	0.997	0.002	0.993-1.000
TAPM	0.996	0.004	0.989-1.000
TAPN	0.996	0.002	0.992-1.000

Table B1: Results for the 4 marker combinations for the 7 markers ANG2 (A), MAC2BP (M), NSE (N), TIMP1 (P), TB4 (T), VGF (V), CGA (C) using RF for the training dataset n=106. SPSS modeler was used for the creation of the models. Average AUC across all the 4 marker AUCs was above 0.996. A total of 35 4 marker combinations were looked at.

5 marker	Average	STDEV	Range
TVAPM	0.999	0.001	0.996-1.00
TCVAN	0.999	0.001	0.998-1.00
CVANM	0.999	0.001	0.998-1.00
TCVPN	0.999	0.002	0.994-1.00
TAPNM	0.999	0.002	0.995-1.00
TCVAM	0.999	0.001	0.996-1.00
CVPNM	0.999	0.001	0.996-1.00
CVAPM	0.999	0.001	0.995-1.00
TCPNM	0.999	0.002	0.994-1.00
TCAPM	0.999	0.002	0.994-1.00
CVAPN	0.999	0.001	0.996-1.00

CAPNM	0.999	0.001	0.997-1.00
TCVPA	0.999	0.002	0.993-1.00
TVAPN	0.998	0.003	0.993-1.00
TVPNM	0.998	0.001	0.996-1.00
TCVPM	0.998	0.003	0.992-1.00
TCVNM	0.998	0.003	0.991-1.00
VAPNM	0.998	0.002	0.995-1.00
TCAPN	0.998	0.001	0.997-1.00
TCANM	0.998	0.002	0.994-1.00
TVANM	0.997	0.001	0.997-1.00

Table B2: Results for the 5 marker combinations for the 7 markers ANG2 (A), MAC2BP (M), NSE (N), TIMP1 (P), TB4 (T), VGF (V), CGA (C) using RF for the training dataset n=106. SPSS modeler was used for the creation of the models. 21 combinations were carried out and for each combination with 10 runs were carried out. The average AUC, standard deviation and range. Average AUCs across all 5 marker AUCs were above 0.997.

Combination	Average AUC	STDEV	Range
TCVAPN	1.000	0.001	0.996-1.000
TCAPNM	0.999	0.001	0.997-1.000
TCVPNM	0.999	0.001	0.997-1.000
TCVAPM	0.999	0.001	0.998-1.000
CVAPNM	0.999	0.001	0.998-1.000
TVAPNM	0.999	0.003	0.993-1.000
TCVANM	0.998	0.003	0.992-1.000

Table B3: Results for the 6 marker combinations of ANG2 (A), MAC2BP (M), NSE (N), TIMP1 (P), TB4 (T), VGF (V), CGA (C) using RF for the training dataset n=106. SPSS modeler was used for the creation of the models. 7 combinations with 10 runs per combination were carried out. The average AUC, standard deviation and range for each combination is shown. Average AUCs across all 6 marker AUCs were all above 0.998.

Combination	Average AUC	STDEV	Range
7 marker	0.999	0.001	0.997-1.00

Table B4: Results for the 6 marker combinations of ANG2 (A), MAC2BP (M), NSE (N), TIMP1 (P), TB4 (T), VGF (V), CGA (C) using RF for the training dataset n=106. SPSS modeler was used for the creation of the model 10 runs were carried out. The average AUC, standard deviation and range. Average AUC for the 7 marker combination was 0.999

Appendix C: Confusion Matrices for the CVA algorithms

	0.00	1.00
0.00	38	11
1.00	13	44

Table C1: Confusion Matrix obtained for the RBF SVM C=10 gamma=0.95 algorithm using SPSS Modeler for the CVA combination.
38 control were identified correctly as control, 44 cases identified correctly as case, 13 cases incorrectly identified as control and 11 controls incorrectly identified as cases.

	0.00	1.00
0.00	39	10
1.00	15	42

Table C2: Confusion Matrix obtained for the Linear SVM C=4 algorithm using SPSS Modeler for the CVA combination.
39 controls were identified correctly as control. 42 cases were identified correctly as cases, 15 cases were incorrectly identified as control and 10 controls were incorrectly identified as cases.

	0.00	1.00
0.00	41	8
1.00	15	42

Table C3: Confusion Matrix obtained for the LR algorithm obtained in SPSS Modeler for the CVA combination.
41 controls were identified correctly as control. 42 cases were identified correctly as cases. 15 cases were incorrectly identified as control and 8 controls were incorrectly identified as cases.

Correct	**106**	**100%**
Wrong	0	0%

Table C4: Results for the RF using Run 1 of 10 for the CVA marker combination.
The algorithm correctly identified all cases as cases and all controls as controls.

	1.00	0.00
1.00	**42**	**15**
0.00	**8**	**41**

Table C5: Confusion Matrix obtained for the Optimised Linear SVM for the CVA combination using the MLCA. 42 cases were correctly identified as case.
41 controls were correctly identified as controls. 8 controls were incorrectly identified as case and 15 cases were incorrectly identified as control.

	1.00	0.00
1.00	**57**	**0**
0.00	**0**	**49**

Table C6: Confusion Matrix obtained for the Optimised RBF SVM for the CVA combination using the MATLAB Classification Learner App.
57 cases were correctly identified as case and 49 controls were correctly identified as control.

Appendix D: Autoclassifier for C5.0 decision tree

| Model | Graph | Summary | Settings | Annotations |

| Sort by: | Use | | ⊙ Ascending | ○ Descending | | ✕ Delete Unused Models | View: | Training set ▾ |

Use?	Graph	Model	Build Time (mins)	Max Profit	Max Profit Occurs in (%)	Lift{Top 30%}	Overall Accuracy (%)	No. Fields Used	Area Under Curve
☑		C5 1	< 1	200.000	57	1.539	83.178	2	0.835
☑		Quest 1	< 1	152.857	56	1.454	75.701	7	0.762
☑		CHAID 1	< 1	170.0	67	1.452	78.505	2	0.8

Figure D1: C5.0 decision tree identified as most suitable decision tree model for UOL training data

Appendix E: C5.0 decision tree for external validation cohort used as a training cohort
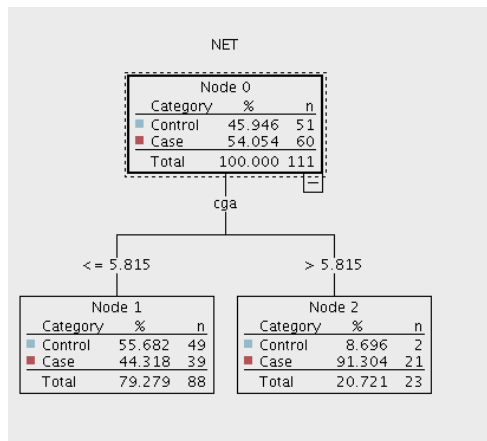

Figure E: General C5.0 decision tree for the AC combination whereby CgA was used for the external validation cohort used as a training cohort. Controls (n=51) and pNET cases (n=60).