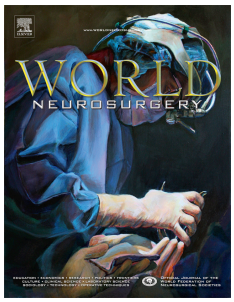
Journal Pre-proof



Utilising Natural Language Processing to Identify Brain Tumour Patients for Clinical Trials: Development and Initial Evaluation

James Booker, MRCS, Jack Penn, FRCS, Kawsar Noor, PhD, Richard J.B. Dobson, PhD, Naomi Fersht, PhD FRCR, Jonathan P. Funnell, MRCS, Ciaran S. Hill, PhD FRCS, Danyal Z. Khan, MRCS, Nicola Newall, MBChB, Tom Searle, PhD, Siddharth Sinha, MRCS, Lewis Thorne, FRCS, Simon C. Williams, MRCS, Michael Kosmin, PhD FRCR, Hani J. Marcus, PhD FRCS

PII: S1878-8750(25)00263-3

DOI: https://doi.org/10.1016/j.wneu.2025.123907

Reference: WNEU 123907

To appear in: *World Neurosurgery*

Received Date: 20 October 2024

Revised Date: 8 March 2025

Accepted Date: 10 March 2025

Please cite this article as: Booker J, Penn J, Noor K, Dobson RJB, Fersht N, Funnell JP, Hill CS, Khan DZ, Newall N, Searle T, Sinha S, Thorne L, Williams SC, Kosmin M, Marcus HJ, Utilising Natural Language Processing to Identify Brain Tumour Patients for Clinical Trials: Development and Initial Evaluation, *World Neurosurgery* (2025), doi: https://doi.org/10.1016/j.wneu.2025.123907.

Utilising Natural Language Processing to Identify Brain Tumour Patients for Clinical Trials: Development and Initial Evaluation

James Booker[1,2] MRCS, Jack Penn[1,2] FRCS, Kawsar Noor[3,4] PhD, Richard J.B. Dobson[3,4,5,6,7] PhD, Naomi Fersht[8] PhD FRCR, Jonathan P. Funnell[1,9] MRCS, Ciaran S. Hill[2,10] PhD FRCS, Danyal Z. Khan[1,2] MRCS, Nicola Newall[1,2] MBChB, Tom Searle[6] PhD, Siddharth Sinha[1,2] MRCS, Lewis Thorne[2] FRCS, Simon C. Williams[1,11] MRCS, Michael Kosmin[4,8] PhD FRCR & Hani J. Marcus[1,2] PhD FRCS

[1]Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London, UK

[2]Victor Horsley Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London, UK

[3]Institute for Health Informatics, University College London, London, UK

[4]NIHR Biomedical Research Centre, University College London Hospitals NHS Foundation Trust, London, UK

[5]Health Data Research UK London, University College London, London, UK

[6]NIHR Biomedical Research Centre, South London and Maudsley NHS Foundation Trust and King's College London, London, UK

[7]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK

[8]Department of Oncology, National Hospital for Neurology and Neurosurgery, London, UK

[9]Department of Neurosurgery, King's College Hospital NHS Foundation Trust, London, UK

[10]UCL Cancer Institute, University College London, London , UK

[11]Department of Neurosurgery, The Royal London Hospital, London, UK

Corresponding author:

James Booker, MRCS
James.booker.19@ucl.ac.uk
ORCID: 0000-0001-7588-2827
National Hospital for Neurology and Neurosurgery,
London,
WC1N 3BG
Tel: 020 3456 7890

Funding and disclosures

No specific funding was received for this study. The authors have no conflicts of interest to disclose.

Acknowledgements

JB, JP, JPF, DZK, NN, SS, SCW & HJM are supported by the Wellcome (203145Z/16/Z) EPSRC (NS/A000050/1) Centre for Interventional and Surgical Sciences, University College London. JP is supported by the HEE Topol Digital Fellowship. DZK is supported by a NIHR Academic Clinical Fellowship. CSH, KN, RJBD, & HJM are supported by the NIHR Biomedical Research Centre, University College London.

Author contributions

JB, JP, KN, RJBD, NF, JPF, CH, DZK, NN, SS, LT, SCW, MK and HJM contributed to conceiving and designing the study. JB and JP contributed to data extraction, curation, and analysis. KN and RJBD contributed to model design and development. JB, JP, JPF, DZK, NN, SS, SCW and HJM drafted the manuscript. CH, LT, NF, MK and HJM provided supervision of the study. All authors were involved in the writing and approval of the final version of the manuscript.

Title: **Utilizing Natural Language Processing to Identify Brain Tumor Patients for Clinical Trials: Development and Initial Evaluation**

Abstract

Background

Identifying patients eligible for clinical trials through eligibility screening is time and resource intensive. Natural Language Processing (NLP) models may enhance clinical trial screening by extracting data from Electronic Health Records (EHR).

Objective:

We aimed to determine whether an NLP model can extract brain tumor diagnoses from outpatient clinic letters and link this with ongoing clinical trials.

Methods

This retrospective cohort study reviewed outpatient neuro-oncology clinic letters, to detect brain tumor diagnoses. We used an NLP model to perform named-entity-recognition + linking algorithm that identified medical concepts in free text and linked them to a Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) ontology, which we used to search a clinical trials database. Human annotators reviewed the accuracy of the concepts extracted and the relevance of recommended clinical trials. Search results were shown on a notification dashboard accessible by clinicians and patients on the EHR. We report the model's performance using precision, recall and F1 scores.

Results

The model recognized 399 concepts across 196 letters with macro-precision=0.994, macro-recall=0.964 and macro-F1=0.977. Linking the model results with a clinical trials database identified 1,417 ongoing clinical trials, of these 755 were highly relevant to the individual patient, who met the eligibility criteria for trial recruitment.

Conclusions

NLP can be used effectively to extract brain tumor diagnoses from free-text EHR records with minimal additional training. The extracted concepts can then be linked to ongoing

clinical trials. While further analysis is required to assess the impact on clinical outcomes, these findings suggest a potential application for integrating NLP algorithms into clinical care.

(257 words)

Running Title

NLP to improve screening for clinical trials.

Keywords

Natural language processing, machine learning, clinical trials, recruitment, brain tumors, neuro-oncology.

Abbreviations

EHR, Electronic Health Records; NLP, Natural Language Processing; NER+L, Named Entity Recognition + Linking; SNOMED-CT, Systematized Nomenclature of Medicine Clinical Terms.

Manuscript Text (3194 words)

Introduction

A major challenge to the successful completion of clinical trials is patient recruitment. It is the most common cause of clinical trials discontinuation, with 53% of publicly funded randomized control trials (RCTs) not reaching recruitment targets.[1] This has a number of repercussions: (1) completed studies are underpowered; (2) if the study is extended to increase recruitment this raises costs;[2] (3) if the study is discontinued it is a waste of research resources. Therefore, the cost of patient recruitment is the main contributor to the $1 billion cost of new drug development.[3] A major barrier to enrolling patients into clinical trials is eligibility screening, which is a labor-intensive process involving the manual review of patient medical histories and has a large time and financial cost.[4] This restricts the number of patients that can be screened for a trial and increases the risk of an insufficient study size.

As more healthcare providers move towards using electronic health records (EHR) the possibility to leverage the digitized patient data to tackle this problem becomes possible. The EHR data represents a rich data asset since it combines structured data (e.g., demographic data) with unstructured data (e.g., clinical free-text notes). NLP, is a subfield of artificial intelligence (AI) concerned with using computers to interpret and process free text, can be used to extract data from clinical documents.[5] The primary advantage of NLP lies in its ability to rapidly and consistently analyze large volumes of free text without the risk of fatigue or performance degradation, which can occur with human reviewers during repetitive tasks.[6] In previous applications, NLP has been used to identify characteristic features of under-recognized disease to reduce the time to diagnosis. In a recent study, NLP analyzed over one million clinical documents including clinical letters, referrals and radiological reports were analyzed to identify shared characteristics of patients with shunt-responsive normal pressure hydrocephalus.[7] This information is vital to improving patient outcomes in this under-treated cohort.

For some time, there has been interest to use this data to power various downstream tasks, including clinical trial recruitment.[5] NLP has previously been shown to dramatically reduce the time required for eligibility screening for clinical trials of common conditions when

provided with large numbers of clinical documents per patient.[8–11] In addition, NLP has been used in cancer trial enrollment with comparable performance to manual screening.[12] However, previous cancer trials have had access to a wealth of clinical data and the results are skewed by cancers with high prevalence such as lung, breast and prostate. It is unclear if an NLP can be as effective when used in an outpatient setting to link a heterogenous cohort of patients with brain tumors to clinical trials based on limited clinical information. We aim to evaluate the utility of NLP to screen clinical notes to identify a brain tumor diagnosis. Secondly, we aim to use the NLP output to automatically search a clinical trial database and recommend suitable trails for enrolment.

Methods

Study overview, design, and setting

This study presents a retrospective review of patients' EHRs who sequentially presented to a tertiary neurosurgery center in the United Kingdom (UK) and were given a histological diagnosis of a brain tumor. Collected data was extracted using the unstructured data retrieval platform CogStack.[13] The data was then analyzed by an NLP model within CogStack known as the Medical Concept Annotation Toolkit (MedCAT). Ground truth data was established by two neurosurgical doctors manually identifying and documenting specific brain tumor diagnosis concepts within the text of a clinic letter. Subsequently, the model was tasked to extract brain tumor diagnosis concepts, and the results were compared to the ground truth data. Identified brain tumor diagnoses were then linked to ongoing clinical trials by searching a widely used clinical trial database.

Participants

Patients were identified using a prospectively maintained EHR system Epic (Epic systems corporation, Wisconsin, USA). Inclusion criteria were patients diagnosed with a brain tumor (primary or metastatic) from June to October 2022 via resection or biopsy. A broad inclusion criterion was used to ensure generalizability of the results. The timeframe allowed for 200 patients to be included, encompassing, new and recurrent/progressive cases. This target was set with a data engineer to allow adequate evaluation of the NLP model that had previously undergone validation in neurosurgical datasets.[7,14,15] Patients were excluded if aged below 18 years at admission or histological analysis showed non-neoplastic tissue.

AI System

The NLP model is an open source Named Entity Recognition + Linking (NER+L) algorithm that is trained to detect medical concepts in free text and link to a Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) ontology.

The NLP model used in this study is an open-source algorithm designed for Named Entity Recognition and Linking (NER+L). The model can automatically identify important medical

terms, such as diseases, symptoms, or treatments, from unstructured text. Once the model finds these medical concepts, it connects them to a standardized medical database called SNOMED-CT (Systematized Nomenclature of Medicine - Clinical Terms). SNOMED-CT is a comprehensive medical dictionary that helps ensure the terms are universally understood and consistent across healthcare systems. Through this process, the model translates the free text into structured medical information that can be used for further analysis or decision-making. For further info on model, see Supplement 1.

Data sources and measurements

Clinical documentation was reviewed from EHR after patients received a histological brain tumor diagnosis. Outpatient clinic letters were used because they are key communication tools between physicians and patients. Integrating the NLP model into these letters will aid in discussing trial recruitment. Two neurosurgical doctors selected the most detailed letter for each patient, from diagnosis to repeat surgery or death. Baseline characteristics, including age, gender, ethnicity, comorbidities, and histological diagnosis using the 2021 World Health Organization Classification of tumors[16] were manually collected.

Model evaluation

Patients' clinical documentation was manually reviewed and brain tumor diagnoses as SNOMED-CT concepts were identified. Following this, the NLP was tested on its ability to identify brain tumor diagnoses on a clinical documentation of patients from a prespecified list of SNOMED-CT terms (Supplement 2). Manual assessors reviewed the NLP outputs and either 'rejected' or 'accepted' the results using the MedCAT annotation tool.[17] This was compared against a ground truth histological diagnosis identified by the manual assessors. Each concept identified was mutually exclusive. In addition, brain tumor diagnosis terms that were not identified by the model were recorded as false negatives against the relevant SNOMED-CT concept. To ensure the NLP system could identify variations in language used to describe the same medical diagnosis, we incorporated specific "rules" into the program (Supplement 3). These rules help the NLP model recognize synonyms or variations of medical terms that might be used in doctors' notes. For example, "GBM" and another uses "glioblastoma multiforme" the system would know that both refer to the same condition and

link them to the same medical term in the SNOMED-CT database. Comparison between the NLP and ground truth manual assesses was done to create a precision, recall and F1 score for each discrete histological diagnosis. Precision measures the accuracy of the positive predictions made by the NLP model. It is calculated as the ratio of true positive diagnoses (correctly identified by the model) to the total number of positive predictions (both true and false positives). Recall (or sensitivity) measures the model's ability to correctly identify all relevant cases. It is calculated as the ratio of true positive diagnoses to the total number of actual positive cases (true positives plus false negatives). The F1 score is the harmonic mean of precision and recall, providing a single metric that balances the trade-off between them. Macro average metrics were calculated to assess the overall performance of the model.

Finally, histological diagnoses identified by the model were linked with ongoing brain tumor trials by searching a multinational, clinical trial database – *clinicaltrials.gov*. This was chosen because it is regularly updated, includes global trials, and has quality-reviewed mandatory data fields, including eligibility criteria.[18] Using this single source ensures comprehensive, efficient, and appropriate trial linking for the study cohort. The filters: study status = recruiting or not yet recruiting studies, location = UK and Age = Adult (18-64) or older adult (65+). This was done to identify relevant clinical trials for the study cohort. Results of the search were displayed on a notification dashboard developed within the EHR. Manual assessors reviewed the linked trials for relevance by reviewing the EHR of each patient. Linked trials were classified as relevant or not relevant based on the inclusion and exclusion criteria for the trials. Precision scores were calculated for each patient and an overall average precision was calculated for the study cohort. See Figure 1 for a study workflow diagram.

Ethical approval

The study was registered as part of a service evaluation and approved by the Clinical Governance Committee. Informed consent was not required for this study.

Reporting guidelines

DECIDE-AI[19] reporting checklist was used.

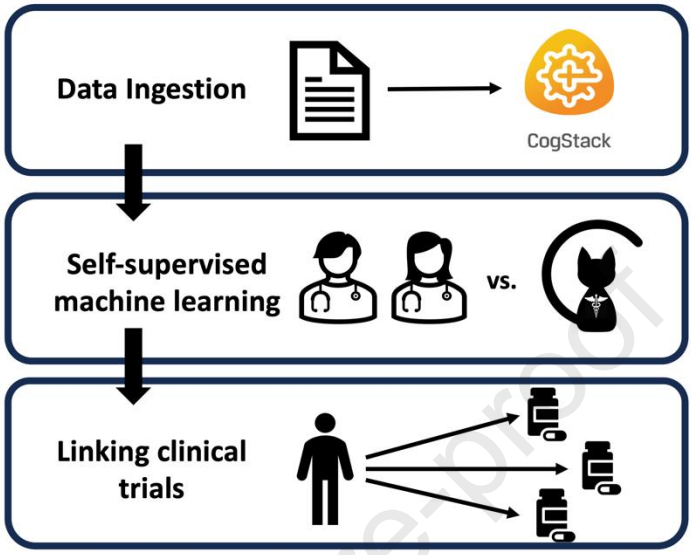


Figure 1. Study Workflow.

This figure illustrates the three key phases of the study workflow. In the 'data ingestion' phase, unstructured data from clinic letters and structured demographic data was retrieved from electronic health records using the CogStack platform. The 'self-supervised machine learning' stage involved the analysis of clinic letters separately using the MedCAT natural language learning tool and expert assessors, to identify brain tumor diagnoses. The brain tumor diagnoses extracted were compared using the expert assessors as ground truth. Finally, in the 'linking clinical trials' stage, each identified diagnosis was linked with relevant clinical trials through queries to the clinicaltrials.gov database.

Results

Descriptive data

Two-hundred patients who underwent either a brain biopsy or a brain tumor resection between June and October 2022 with a histological diagnosis of a brain tumor, had their outpatient clinic letters analyzed. Four patients during this time-period were excluded due to the absence of neoplastic disease (Rathke's cleft cyst n=2, Tuberculoma n=1, lymphangioma n=1), leaving 196 patients included in the analysis. Glioblastoma was the most common histological diagnosis, followed by pituitary adenoma and meningioma (Table 1).

Table 1. Cohort demographics.

Characteristic	**N = 196**[1]
Age	54 (42, 67)
Sex	
Female	93 (47%)
Male	103 (53%)
Histological Diagnosis	
Glioblastoma, IDH wildtype	60 (31%)
Pituitary adenoma	47 (24%)
Meningioma	29 (14.8%)
Secondary malignant neoplasm of the brain	19 (9.7%)
Astrocytoma, CNS WHO Grade 1-3	11 (5.5%)
CNS Lymphoma	6 (1.0%)
Oligodendroglioma, CNS WHO Grade 2+3	6 (3.1%)

Characteristic	N = 196[1]
Chordoma	4 (2.0%)
Craniopharyngioma	3 (1.5%)
Medulloblastoma	3 (1.5%)
Germinoma	2 (1.0%)
Neuroepithelial tumor, CNS WHO grade 1	2 (1.0%)
Other	4 (2.0%)

[1] Median (IQR); n (%)

Main Results

Using the model, 195/196 (99%) letters contained diagnosis concepts that could be extracted and linked with clinical trials. The letter without a diagnosis concept contained a description of a diagnosis of glioma without reference to a specific SNOMED-CT term. The untuned model extracted 311 concepts in clinic letters and linked them with SNOMED-CT terms (precision=0.991, recall=0.735, F1=0.844).

Variations in language used to describe diagnoses existed in the clinic letters. Clinicians would use abbreviations of a diagnosis or synonym e.g., GBM instead of Glioblastoma. This resulted in some concepts being missed as false negatives. Therefore, model tuning was instituted by manually identifying language variations and creating rules to link them with the correct SNOMED-CT term (Supplement 2). In addition, previous brain tumor diagnoses written in the clinic letter were extracted as SNOMED-CT concepts even if not relevant to new histological diagnosis; this led to false positives.

Across the 196 letters, 399 individual concepts were recognized by the tuned model and linked to SNOMED-CT terms. Metrics for the top five most common SNOMED-CT

concepts, which comprised of n=303 (76%) of identified concepts, is shown in Figure 2. Overall macro averages across all concepts are shown in Table 2.

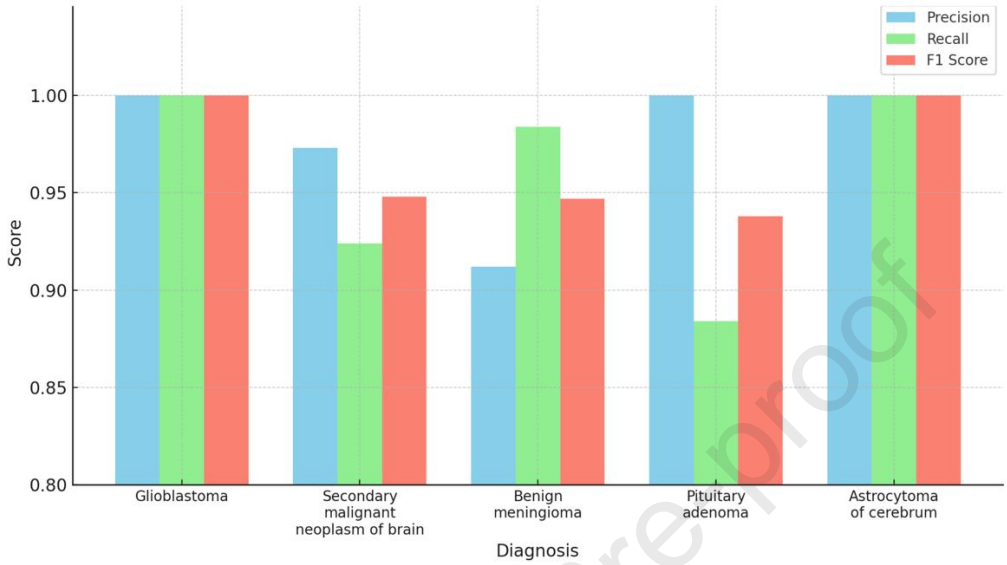


Figure 2. Model performance metrics per SNOMED-CT concept.

Bar chart plot of the precision, recall and F1 scores for the five most common brain tumor diagnoses.

Table 2. Model macro average performance metrics

Count	True positive	False positive	False negative	Precision	Recall	F1
399	390	9	21	0.994	0.964	0.977

Linking extracted terms with clinical trials

A notification dashboard was developed offline within the Epic to display the search results of the identified brain tumor diagnosis on clinicaltrials.gov. Each patient had a unique

dashboard, which displayed basic patient info, patient letters, extracted SNOMED-CT brain tumor diagnoses and linked clinical trials. The dashboard contained several tabs displaying different information clinicians are given the option to either 'accept' or 'reject' the clinical trial recommendation.

Using the clinical trials dashboard 1,417 trials were linked to 176 patient diagnoses. There was large variation in the number of trials that were linked to diagnostic SNOMED-CT terms, ranging from 54 for B-cell lymphoma to zero for medulloblastoma. Of the linked trials, 755 were highly relevant to individual patient, who met the eligibility criteria for trial recruitment. Linked trials were at a variety of translational stages including observational studies and phase I-III. There were 662 trials that were linked, but the individual patient did not meet the eligibility criteria. Clinical trial linking precision was highest for meningioma (average precision 1.00) and lowest for pituitary adenomas (average precision 0.331). Poor performance status, unresectable tumor and tumor recurrence were common reasons for patients being non-eligible. The average precision of this clinical trial linking stage using the *clinicaltrial.gov* search engine was 0.578.

Discussion

This study demonstrated the effectiveness of NLP in screening clinical notes for the identification of brain tumor diagnoses. Additionally, the study utilized the NLP-generated outputs to automatically query a clinical trial database and recommend appropriate trials for patient enrollment. The study represents the first attempt in streamlining eligibility screening in brain tumor trials using outpatient letters.

Principal Findings

Firstly, the model showed high performance metrics on recognizing concepts and linking them with SNOMED-CT terms, despite not being specifically trained in this domain. The NLP had high performance metrics (macro-precision=0.994, macro-recall=0.964 and macro-F1=0.977), despite only using free-text information from a single clinical letter. This surpasses previous performance from the MedCAT model which has used a large number of clinical documents from multiple sources.[20] This is testament to the continued performance improvement seen in the development of the MedCAT model. This demonstrates that the NLP is fit-for-purpose as a clinical tool to aid clinicians in increasing patient access to clinical trials. It is also promising for the integration of the NLP into multiple use cases across different clinical domains because a high performance was achieved without a dedicated training stage.

Secondly, the NLP had a limited amount of unstructured data per patient, using one outpatient letter per patient. This is representative of clinical practice; at our center each patient given a new diagnosis of a brain tumor will receive a ‘diagnosis appointment’ with a clinician who will orchestrate the future management. The clinic letter written from this encounter will have detailed unstructured data pertaining to the brain tumor diagnosis, which was used in this study. Additionally, information from pathology reports and imaging studies is condensed and included in the letter by the clinician. Therefore, including other documents would have led to duplication of information without any marginal gain. In contrast, previous effective published NLPs have used a wealth of clinical data, which may not always be available, especially in patients with glioblastoma in which 82% of patients have an operation within a month of their first scan, having previously been fit and well.[21] Also, the model is locally tunable using the MedCAT user interface for specific scenarios and domains. The

authors added ‘rules’ to the model to recognize synonyms of diagnoses which further improve the performance metrics of the model, most notably the recall. The performance metrics presented in this study are a conservative representation of what is possible with refinement of the model. Through an iterative process it would be possible to achieve even higher performance metrics by adding a vocabulary of ‘rules’ linked to corresponding SNOMED-CT concepts to identify all synonyms of all brain tumor diagnoses. The multi-domain performance demonstrates that the results observed in this study will be generalizable to other clinical settings, where similar performance would be expected. Importantly this technology is highly scalable and can be integrated into a clinical pathway in multiple use cases to increase the assess of patients to clinical trials.

Thirdly, the retrieval of clinical trials for each patient was displayed using a notification dashboard on the EHR to allow clinicians and patients to see when a clinical trial in their area was recruiting patients with the same brain tumor diagnosis. Quantitative evaluation of the clinical trial dashboard identified that 755 trials across 176 patients were highly relevant to the individual patient met the eligibility criteria for recruitment. The clinical translation of this work is to increase trial recruitment once integrated into clinical practice. The benefit of this process is the NLP works synergistically with humans and allows clinicians to reject the suggestion if they are not relevant. We propose integrating this into clinical practice, with research nurses reviewing clinical trial dashboards for patients seen weekly in clinic and highlighting patients for trial recruitment to physicians. There were 662 trials that were recommended on the clinical trials dashboard, but the individual patient did not meet the eligibility criteria. We found that diagnoses with diverse histological subtypes, such as pituitary adenomas, led to poor precision in trial linking. Whereas, diagnoses with homogenous histological subtypes such as benign meningioma led to broad inclusion criteria and high precision scores. Therefore, physicians acting as the final review of recommended trials limits any unintended harm associated with incorrect trial linking. In addition, the simple pipeline used to search an online international clinical trials database may be integrated in other healthcare settings.

Safety and errors

The main source of errors using the model is the reliance on the SNOMED-CT ontology to link recognized concepts. Large biomedical concepts databases such as SNOMED-CT do contain acronyms, abbreviations, and synonyms of concepts, but they are not exhaustive. Therefore, when certain abbreviations were used the model was unable to identify the concepts and link them to SNOMED-CT terms. For example, the presence “pit adenoma” instead of “pituitary adenoma” in clinical letters. In addition, although this was not observed in the study, rare subtypes of brain tumors may not be listed in the SNOMED-CT ontology. However, the MedCAT user interface allows for rules to be added to the model, which was used effectively in this study, increasing recall performance from 0.735 to 0.964. In addition, standardizing the diagnostic terms used by clinicians would also lead to improve performance of the model.

Another source of observed errors was that the MedCAT model cannot differentiate between new and previous diagnoses of brain tumors. The SNOMED-CT terms linked to written text can be filtered to just brain tumor diagnoses, which reduces the chance of overlapping concepts being identified. However, when a patient has had multiple brain tumor diagnoses it led to false positives in the performance metrics and decreased precision scores. Also, for meningiomas a higher false positive rate occurred because the model identified patients with atypical meningiomas as having benign meningiomas. Future improvements in the NER+L algorithm underlying the NLP to identify past medical history and subtypes of brain tumor will improve model performance.

During the clinical trial linking stage several errors arose, which caused a reduction in the precision. This reduction in precision was due to the *clinicaltrials.gov* search engine, which provides a broad search of clinical trials based on keywords. Therefore, although patients were linked to clinical trials investigating the correct diagnosis, exclusion criteria such as poor performance status, large tumor location, multifocal tumor, and tumor recurrence, meant that patients were not eligible for recruitment. In addition, we did account for trials which the patients would have been eligible and were not suggested by the clinical trials database**.** Future integration of more clinical data into NLP model and use of more specific search

parameters amongst several clinical trial databases will improve the clinical trial recommendations.

Strengths and limitations

The main strength of this study is the use of a multi-domain, self-supervised NLP which was used without a dedicated training stage and with access to limited unstructured data. Therefore, these results will be reproducible in other healthcare settings without a high upfront time cost of model training. Another strength is that the model has been deployed in several hospitals in the UK within sensitive data silos. This reduces the risk of data breaches, which are much higher with current large language models with external servers. This allows the model to adhere to the strict data governance rules of the National Health Service (NHS).

A recognized limitation of this study is its design as a retrospective review of clinical notes from a single neurosurgical center. Future work will assess the MedCAT model's effectiveness in a live clinical setting and its impact on clinical trial recruitment. The model was developed using Epic, the most widely used EHR, but results may not apply to other EHR systems, limiting generalizability.[22] Additionally, the model needs further validation with international datasets. Furthermore, as this study is an early evaluation of the model, only a single clinical trial database was used for trial linking. This resulted in an incomplete retrieval of clinical trials and no recall metric calculation. To overcome these limitations, future work aims to use live patient data, across multiple centers and combine multiple clinical trial databases during the trial linking phase.

Conclusions

This study presents an early evaluation of an NLP NER+L model using self-supervised learning to extract brain tumor diagnoses from outpatient clinic letters and link them with ongoing clinical trials. The model showed high performance metrics despite no dedicated training stage and is imminently scalable to other healthcare settings. The clinical translation

of this model is to automatically notify clinicians and researchers when a patient meets the criteria for a brain tumor trial, streamlining the recruitment process.

Code and Data availability

The code and data used in this study will be made available upon reasonable request by contacting the corresponding author by email.

References

1. Jacques RM, Ahmed R, Harper J, et al. Open access Recruitment, consent and retention of participants in randomised controlled trials: a review of trials published in the National Institute for Health Research (NIHR) Journals Library (1997-2020). *BMJ Open*. 2022;12:59230. doi:10.1136/bmjopen-2021-059230

2. Sully BGO, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. *Trials*. 2013;14(1):166. doi:10.1186/1745-6215-14-166

3. Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: a systematic review. *Health Policy*. 2011;100(1):4-17. doi:10.1016/j.healthpol.2010.12.002

4. Penberthy LT, Dahman BA, Petkov VI, DeShazo JP. Effort required in eligibility screening for clinical trials. *J Oncol Pract*. 2012;8(6):365-370. doi:10.1200/JOP.2012.000646

5. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. *Ann Surg*. 2018;268(1):70-76. doi:10.1097/SLA.0000000000002693

6. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: A review. *Trends in Anaesthesia and Critical Care*. 2021;38:4-9. doi:10.1016/j.tacc.2021.02.007

7. Funnell JP, Noor K, Khan DZ, et al. Characterization of patients with idiopathic normal pressure hydrocephalus using natural language processing within an electronic

healthcare record system. *J Neurosurg*. Published online November 18, 2022:1-9. doi:10.3171/2022.9.jns221095

8. Cai T, Cai F, Dahal KP, et al. Improving the Efficiency of Clinical Trial Recruitment Using an Ensemble Machine Learning to Assist With Eligibility Screening. *ACR Open Rheumatol*. 2021;3(9):593-600. doi:10.1002/acr2.11289

9. Ni Y, Wright J, Perentesis J, et al. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility Pre-screening for pediatric oncology patients. Published online 2015. doi:10.1186/s12911-015-0149-3

10. Beauharnais CC, Larkin ME, Zai AH, Boykin EC, Luttrell J, Wexler DJ. Efficacy and cost-effectiveness of an automated screening algorithm in an inpatient clinical trial. *Clinical Trials*. 2012;9(2):198-203. doi:10.1177/1740774511434844

11. Tissot HC, Shah AD, Brealey D, et al. Natural Language Processing for Mimicking Clinical Trial Recruitment in Critical Care: A Semi-Automated Simulation Based on the LeoPARDS Trial. *IEEE J Biomed Health Inform*. 2020;24(10):2950-2959. doi:10.1109/JBHI.2020.2977925

12. Chow R, Midroni J, Kaur J, et al. Use of artificial intelligence for cancer clinical trial enrollment: a systematic review and meta-analysis. *J Natl Cancer Inst*. 2023;115(4):365-374. doi:10.1093/jnci/djad013

13. Jackson R, Kartoglu I, Stringer C, et al. CogStack - experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital. *BMC Med Inform Decis Mak*. 2018;18(1):47. doi:10.1186/s12911-018-0623-9

14. Williams S, Noor K, Sinha S, et al. *Concept Recognition and Characterisation of Patients Undergoing Resection of Vestibular Schwannoma Using Natural Language Processing*.; 2024.

15. Booker J, Penn J, Noor K, et al. Early evaluation of a natural language processing tool to improve access to educational resources for surgical patients. *Eur Spine J*. 2024;33(7):2545-2552. doi:10.1007/s00586-024-08315-5

16. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol*. 2021;23(8):1231-1251. doi:10.1093/neuonc/noab106

17. Kraljevic Z, Searle T, Shek A, et al. Multi-domain clinical natural language processing with MedCAT: The Medical Concept Annotation Toolkit. *Artif Intell Med*. 2021;117. doi:10.1016/j.artmed.2021.102083

18. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database--update and key issues. *N Engl J Med*. 2011;364(9):852-860. doi:10.1056/NEJMsa1012065

19. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. 2022;28(5):924-933. doi:10.1038/s41591-022-01772-9

20. Kraljevic Z, Searle T, Shek A, et al. Multi-domain clinical natural language processing with MedCAT: The Medical Concept Annotation Toolkit. *Artif Intell Med*. 2021;117:102083. doi:10.1016/j.artmed.2021.102083

21. Graus F, Bruna J, Pardo J, et al. Patterns of care and outcome for patients with glioblastoma diagnosed during 2008-2010 in Spain. *Neuro Oncol*. 2013;15(6):797-805. doi:10.1093/neuonc/not013

22. Chishtie J, Sapiro N, Wiebe N, et al. Use of Epic Electronic Health Record System for Health Care Research: Scoping Review. *J Med Internet Res*. 2023;25:e51003. doi:10.2196/51003

Figure legends

Figure 1. Study Workflow.

This figure illustrates the three key phases of the study workflow. In the 'data ingestion' phase, unstructured data from clinic letters and structured demographic data was retrieved from electronic health records using the CogStack platform. The 'self-supervised machine learning' stage involved the analysis of clinic letters separately using the MedCAT natural language learning tool and expert assessors, to identify brain tumor diagnoses. The brain tumor diagnoses extracted were compared using the expert assessors as ground truth. Finally, in the ‘linking clinical trials’ stage, each identified diagnosis was linked with relevant clinical trials through queries to the clinicaltrials.gov database.

Figure 2. Model performance metrics per SNOMED-CT concept.

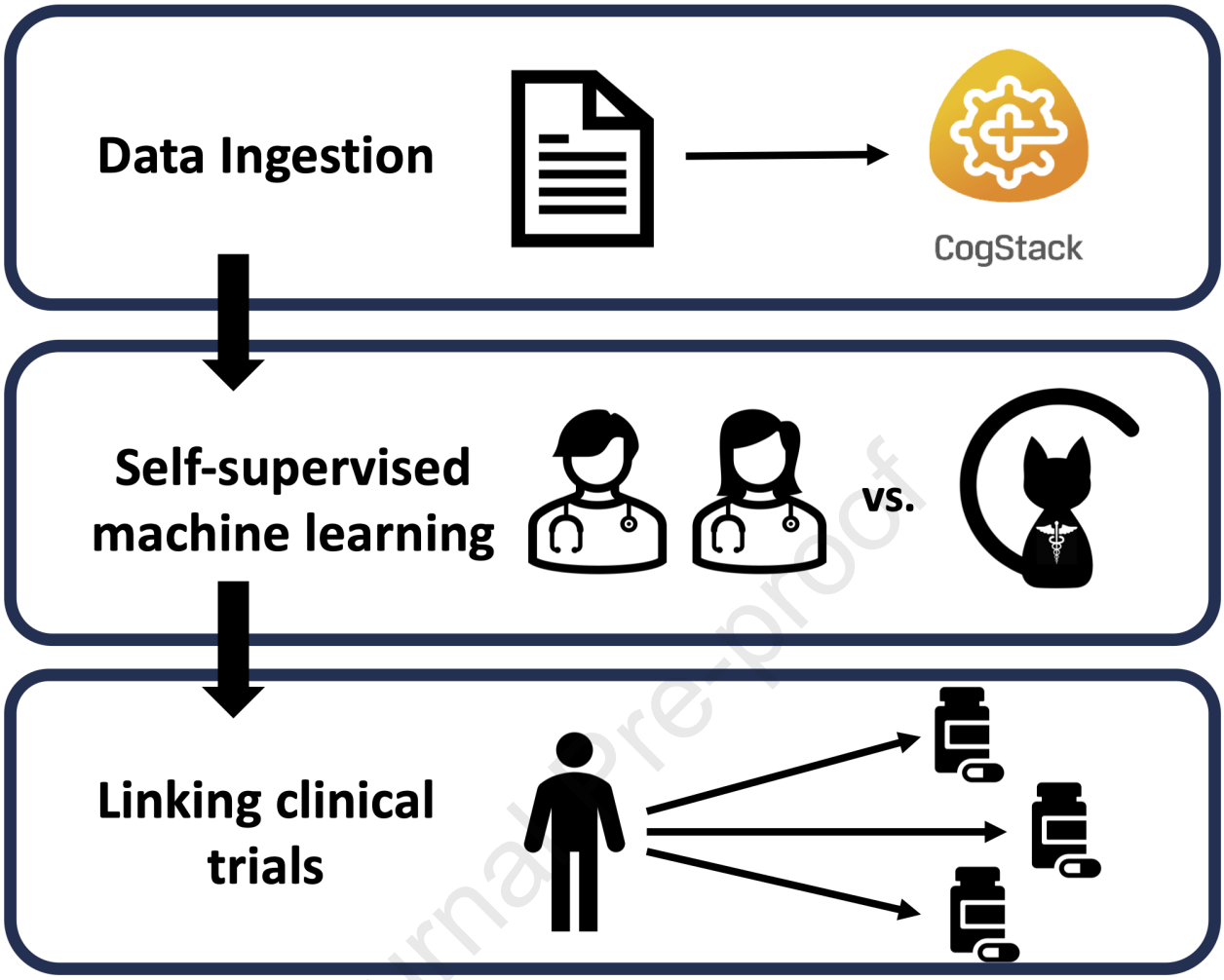
Bar chart plot of the precision, recall and F1 scores for the five most common brain tumor diagnoses.

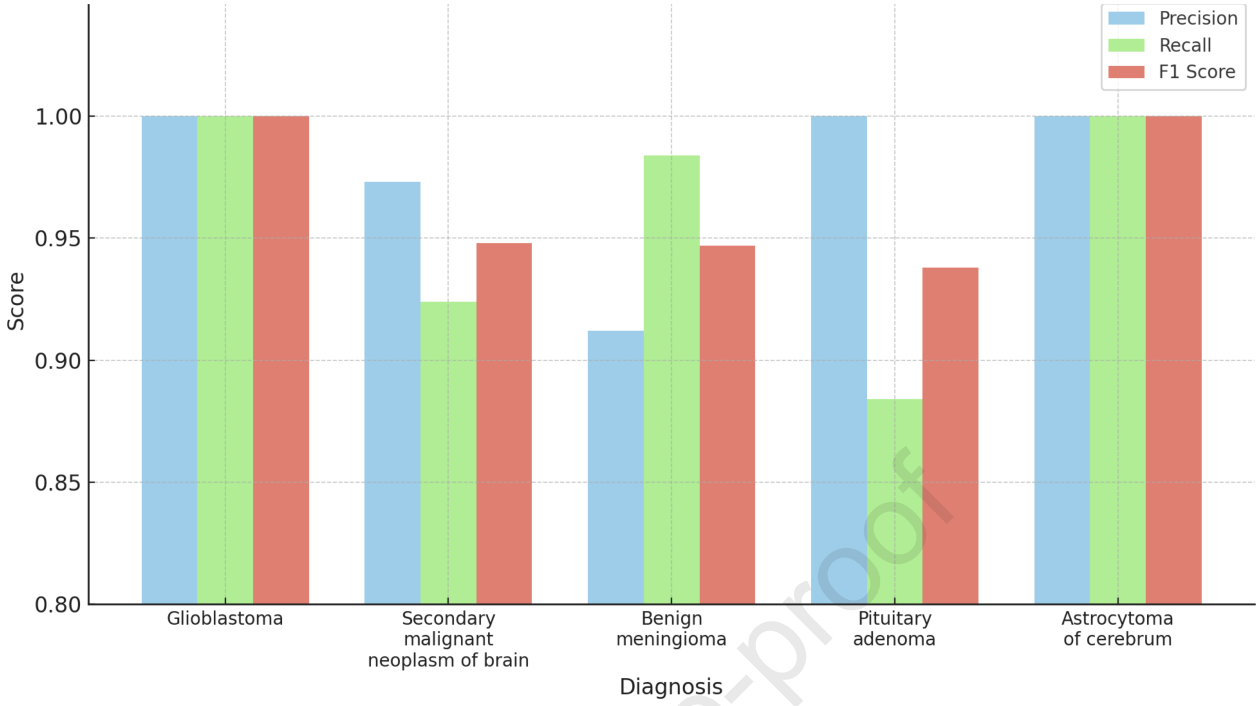
Supplement legends

Supplement 1. NLP model further information

Supplement 2. Prespecified list of SNOMED-CT terms

Supplement 3. MedCAT Rules





Abbreviations

EHR, Electronic Health Records; NLP, Natural Language Processing; NER+L, Named Entity Recognition + Linking; SNOMED-CT, Systematized Nomenclature of Medicine Clinical Terms.

Disclosure-Conflict of interest

Funding and disclosures
No specific funding was received for this study. The authors have no conflicts of interest to disclose.

Acknowledgements

JB, JP, JPF, DZK, NN, SS, SCW & HJM are supported by the Wellcome (203145Z/16/Z) EPSRC (NS/A000050/1) Centre for Interventional and Surgical Sciences, University College London. JP is supported by the HEE Topol Digital Fellowship. DZK is supported by a NIHR Academic Clinical Fellowship. CSH, KN, RJBD, & HJM are supported by the NIHR Biomedical Research Centre, University College London.

Author contributions
JB, JP, KN, RJBD, NF, JPF, CH, DZK, NN, SS, LT, SCW, MK and HJM contributed to conceiving and designing the study. JB and JP contributed to data extraction, curation, and analysis. KN and RJBD contributed to model design and development. JB, JP, JPF, DZK, NN, SS, SCW and HJM drafted the manuscript. CH, LT, NF, MK and HJM provided supervision of the study. All authors were involved in the writing and approval of the final version of the manuscript.

I confirm that all authors have met ICMJE criteria for authorship and have no financial interests to disclose. The manuscript is not being considered, in whole or in part, for publication at any other journal. All co-authors have read and approve of this submission. Please do not hesitate to contact me if you require further information or clarification.

Yours sincerely,

Dr James Booker, MD MRCS

Honorary Clinical Research Fellow
University College London

Neurosurgery Resident
Royal London Hospital
James.booker.19@ucl.ac.uk