3D statistical shape analysis of the face in Apert syndrome

Lara S. van de Lande

Great Ormond Street Institute of Child Health

University College London

This dissertation is submitted for the degree of

Doctor of Philosophy

London, September 2021

“I am among those who think that science has great beauty.”

– Marie Curie

DECLARATION OF ORIGINALITY

I, Lara Sophie van de Lande, 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. The algorithms used in this thesis were implemented by Dr. Athanasios Papaioannou, post-doctoral researcher, and Ms. Eimear O' Sullivan, PhD student at the Department of Computing, Imperial College London.

26th of September 2021

ABSTRACT

Timely diagnosis of craniofacial syndromes as well as adequate timing and choice of surgical technique are essential for proper care management. Statistical shape models and machine learning approaches are playing an increasing role in Medicine and have proven its usefulness. Frameworks that automate processes have become more popular. The use of 2D photographs for automated syndromic identification has shown its potential with the Face2Gene application. Yet, using 3D shape information without texture has not been studied in such depth. Moreover, the use of these models to understand shape change during growth and its applicability for surgical outcome measurements have not been analysed at length.

This thesis presents a framework using state-of-the-art machine learning and computer vision algorithms to explore possibilities for automated syndrome identification based on shape information only. The purpose of this was to enhance understanding of the natural development of the Apert syndromic face and its abnormality as compared to a normative group. An additional method was used to objectify changes as result of facial bipartition distraction, a common surgical correction technique, providing information on the successfulness and on inadequacies in terms of facial normalisation. Growth curves were constructed to further quantify facial abnormalities in Apert syndrome over time along with 3D shape models for intuitive visualisation of the shape variations. Post-operative models were built and compared with age-matched normative data to understand where normalisation is coming short.

The findings in this thesis provide markers for future translational research and may accelerate the adoption of the next generation diagnostics and surgical planning tools to further supplement the clinical decision-making process and ultimately to improve patients' quality of life.

IMPACT STATEMENT

For many craniofacial syndromes, diagnosis is done unnecessary late which can cause irreversible functional problems. Moreover, the timing and type of surgical correction to normalise the face are little understood. Apert syndrome, a craniosynostosis syndrome with a strong facial component, usually requires corrective surgery at some point in life. The indications and timings of surgery for (cosmetic) correction are mainly subjective as decision making and outcome analysis based on objective measurements are limited. Based on this subjective analysis, it has shown that facial normalisation for this patient population is extremely challenging due to complex pattern of the deformity. In addition, the natural development of the facial deformity is mildly understood making it impossible to predict the outcomes of facial growth over time, either with or without surgical correction.

To overcome this delay in diagnosis, lack of information on natural shape development during growth, and limited objective surgical outcome measurements, this thesis presents a framework that allows for automated syndrome identification and presents 3D statistical models to further understand the natural development of the Apert's syndromic face enabling objective surgical outcome analysis and to ultimately improve surgical corrective care of this population. It attempted to narrow the gap between technology and clinical practice. 3D models of soft tissue and bone were made for the face and skull and provided new insights in postoperative outcome analysis and valuable information on the natural development of Apert syndrome, which on its turn aided to understand the regions for improvement.

The presented models have an impact clinically, academically, and commercially. Clinically, it demonstrated the usefulness of 3D models for automated diagnosis and provided novel information on the natural development of Apert face growth. It also allowed for

objective surgical outcome analysis, which is a methodology that can be translated to numerous types of syndromes and surgical techniques. It eluded on areas where, facial bipartition distraction, a surgical technique used at our Unit lacks to achieve facial normalisation. Moreover, the algorithms used in this thesis were validated and could be implemented in commercial software. The automated component of the experiments performed in this thesis, do not only provide an interesting line of future research but also that such tools can have an impact in clinical decision-making and surgical planning. The models presented in this thesis will help to take a step closer to computer–assisted diagnostics and aimed to contribute to making surgical planning more accurate, faster, and personalised. Future access to these types of tools in clinic can help to guide discussions with caregivers and patients in the surgical work-up process.

It is hoped that the work presented in this thesis will influence craniofacial practice, both technically and clinically.

ACKNOWLEDGEMENTS

My PhD programme at ICH & UCL Great Ormond Street Hospital and my time in London has given me some of the best years of my life, it has been a great journey. I am very thankful for the people I was surrounded by. I would like to express my gratitude in particular to:

David Dunaway, Professor in Craniofacial surgery, Craniofacial Plastic Surgeon: *thank you for your brilliant ideas and the opportunity.* **Silvia Schievano**, Professor in Mechanical Engineer: *thank you for this unique opportunity and your support.* **Athanasios Papaioannou**, Computer Scientist: *thank you for caring.* **Noor U. Owase Jeelani**, Craniofacial Neurosurgeon: *thank you for your enthusiasm.* **Eimear O' Sullivan,** Computer Scientist: *thank you for being my ever-patient counterpart.* **Maarten Koudstaal**, Ass. Professor, Craniomaxillofacial Surgeon: *thank you for believing in me.* **Roman Khonsari**, Professor, Craniomaxillofacial Surgeon: *thank you for your trust in me.* **Alessandro, Sara, Selim, Prof. Hayward, Neil, Greg, Juling, Sohaib, Paul, Will, Anne-Jet, Alex, Xerxes, Karan, Khalid, Anoo, Aysima, Antonia, and Sebastiaan,** colleagues and students**:** *thank you for all your help, ideas, and research meetings.* **Stella**, friend: *thank you for your insights.* **Liana**, friend: *thank you for being there regardless of the countries we reside.* **Kiki**, friend: *thank you for showing me perspective.* **Floortje**, friend: *thank you and Sam for celebrating every single step of the way.* **Luca**, friend: *thank you for the laughter and great talks.* **Paulien**, friend: *thank you for the countless mind-clearing runs in Regent Park.* **Joanne**, friend: *thank you for showing me how to take things as they come.* **Linda,** friend, *thank you for your support.* **Mariana**, flatmate: *thank you for your kindness and support during lockdown.* **Matteo**, *thank you for motivating me, especially when I needed it most.* **My family:** *thank you for your eternal support and closeness despite living abroad.*

I am ever grateful for the generous support of the Great Ormond Street Hospital for Children's Charity through the FaceValue programme grant and the European Research Council grant. Many thanks to the patients and parents who gave permission for the use of their data.

You have helped me make this research possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
3DMM	Three dimensional morphable model
AED	Average Euclidean distance
AI	Artificial intelligence
AS	Asymmetric transcranial length
CH	Cranial height
CL	Cranial length
CMA	Convolutional mesh autoencoder models
COVID-19	Coronavirus disease 2019
CT	Computerised tomography
CW	Cranial width
DCNN	Deep convolutional neural network
DICOM	Digital imaging and communications in medicine
FDNA	Biotechnology company Boston, founders of Face2Gene
FGFR	Fibroblast growth factor receptor
FB-RED	Facial bipartition with rigid external distraction frame
GAN	Generative adversarial network
GOSH	Great Ormond Street Hospital for Children, London, United Kingdom
HI	Height index
ICP	Iterative closest-point algorithm
L	Left
LSFM	Large-scale facial model
LF2ZR	Le Fort II with zygomatic repositioning
LYHM	Liverpool-York head model
MRI	Magnetic resonance imaging

NEMH	Hôpital Necker-Enfants Malades, Paris, France
NICP	Non-rigid iterative closest point registration
OCLR	Oblique cranial length ratio
PCA	Principal component analysis
R	Right
RED	Rigid external distraction
t-SNE	t-Distributed stochastic neighbour embedding
SD	Standard deviation
SSM	Statistical Shape Modelling
UV-TPS	UV thin plate spine
UV-OF	UV optical flow

Author's note

In this PhD thesis the term 'normal' is used to define healthy individuals without any craniofacial deformities. 'Shape' is defined as the geometrical information that remains when location, scale, and rotational effects are filtered out from an object.

1. INTRODUCTION

1.1. Introduction

Apert syndrome (OMIM 101200) is a rare congenital craniofacial syndrome characterised by various craniofacial deformities and symmetric syndactyly of the hands and feet. (Figure 1.1-1.2). It was first described by the French paediatrician E. Apert in 1906 (6). It is a rare syndrome that affects males and females equally and is estimated to occur in 1 in 65,000 to 75,000 births, depending on the study cited (7, 8). The highest prevalence is found in Asia, where it is 1 in 45,000 live births. The incidence significantly increases with paternal age (9). The syndrome has complete penetrance, meaning that all individuals carrying the genetic mutation also manifest at the phenotypical level. Moreover, Apert has a variable expressivity, resulting in phenotypically mildly affected individuals, to severe deformities within the same family. It is an autosomal, dominant, inherited, craniosynostosis syndrome with a gain-of-function missense mutation of fibroblast growth factor receptor (FGFR) 2 on chromosome 10q (10).

Figure 1.1 Clinical photographs of Apert Syndrome. Two patients with Apert syndrome are presented here with frontal photographs. **A**) a 1-year-old unoperated female; **B**) an 8-year-old unoperated male.

Clinical features:

- Biconcave face
- Turribrachycephaly
- Hypertelorism
- Exorbitism due to shallow orbits
- Downslanting palpebral fissures
- Maxillary hypoplasia
- Narrow upper dental arch
- Class III malocclusion, anterior open bite, and dental crowding
- Cleft palate with high arched palate can occur
- Complex syndactyly of the hands and feet

Figure 1.2. Overview of clinical features. Here are the most common clinical features listed as described in literature.

Due to its complexity, Apert syndrome is generally diagnosed in infancy, followed by a 'life' long care management at a multidisciplinary, craniofacial Centre. Healthcare professionals involve in the care of these children comprise craniofacial surgeons, oral and maxillofacial surgeons, plastic surgeons, neurosurgeons, as well as orthodontists, radiologists, ophthalmologists, otolaryngologists, psychologists, specialised nurses, and speech and language therapists. The management is focused on protection of function in infancy and correction of deformity in childhood and adolescence. Functional problems seen in infancy require surgical management to protect from long term sequelae including raised intracranial pressure, airway issues, and eye exposure. Thus, early diagnosis of Apert syndrome is essential to deliver optimal care and long-term outcomes. However, screening pathways currently in place are inadequate and often result in delayed diagnosis. This, in turns, translates in

irreversible functional impairments, such as visual failure, neurocognitive defects, and airway problems that could be avoided by timely diagnosis and treatment. An automated method to identify children with Apert syndrome might speed up the diagnostic process and provide clinicians with a chance for delivering better management.

The dominant drivers for surgery in infancy are functional problems, whilst correction of facial abnormalities plays a larger role in childhood and adolescence. Several surgical techniques exist to correct the craniofacial deformities aiming to 'normalise' facial appearance, with choice differing between Centres and are governed by the surgeon's experience rather than objective outcome measurements.

To assess 'normalisation' of surgical intervention, various 2-dimensional (2D) and 3-dimensinal (3D) tools exist, mainly derived from traditional anthropometrics (11). Measurements for soft tissue or bone are inherently 2D in nature and are derived from sparse annotations – finite anatomical points on the skull or face – calculated using computerised tomography (CT) images, x-rays, or photographs (3, 12). While they can provide some insight into the shape of the human face, such analysis misses the extensive information that dense 3D analysis can offer. Dense 3D analysis, or statistical shape modelling (SSM), allows for the skull to be considered and analysed as a whole and can provide greater insight into the shape of the skull in detail (13, 14).

3D morphable models (3DMMs) are a common SSM approach and one of the most powerful statistical approaches for dense analysis and 3D shape modelling. These models are often constructed using principal component analysis (PCA), a dimensionality reduction algorithm which extracts shape modes by maximising the variance along orthogonal shape direction. Introduced in 1999, 3DMMs have proven to be adept at modelling the shape and variations within the human face and head shape, based on 3D photographs, thus relying on

texture (4, 5, 14, 15). They have also been used in the orthopaedic surgical field for bone and muscle models from CT scans and for the more complex structures such as the human hand and ear (14, 16-20). The Large-Scale Face Model (LSFM), based on more than 10,000 faces, has demonstrated the 3DMM power for facial representation. Experiments showed that the model can also be used to infer the age and gender of a given subject (5), for craniomaxillofacial diagnosis, and for surgical simulation in patients undergoing orthognathic surgery (jaw surgery) (21).

The recent introduction of convolutional mesh autoencoder models (CMAs), a deep neural network approach to 3D model construction, offers further potential for the construction of shape-based models (22, 23). These models learn to extract meaningful shape features from the input data and can consequently be used for classification tasks.

In this thesis, these algorithms and tools will be used and further developed to understand the 3D shape of Apert syndrome, to explore the possibilities of such technologies for automated syndrome identification for early diagnosis, and to quantify surgical results with the ultimate aim of improving outcomes and care for this group of patients.

All studies performed in this thesis are performed according to the Helsinki Act and have been approved by the Institutional Ethical Board; Joint Research and Development Office. These studies fall under R&D no. 14DS25.

1.2. **Aims and objectives**

The main aims of this thesis are:

- Automatic diagnosis of Apert syndrome cases from 3D facial photography.
- Analysis of the growing Apert's face and its facial skeletal abnormalities.
- Objective assessment of bipartition surgical outcomes in Apert's facial correction.
- Identification of areas in which the surgical technique might need improvement.

The main principal aim of this thesis is to apply 3D statistical shape methodologies to automatically identify Apert syndrome from 3D face images from genetically and phenotypically similar syndromes, and, once diagnosed, to evaluate the growing Apert's face and its facial skeletal abnormalities pre- and post-surgical correction – ultimately to improve current surgical practice.

To achieve these aims, the following objectives were identified:

- **Objective 1: To automatically identify Apert syndrome from a population of other FGFR-related craniosynostosis syndromes and from a normal population using 3D images and state-of-the-art 3DMM algorithms.**

  3DMMs of Apert syndrome, other FGFR-related conditions and an age matched paediatric normal population need to be constructed from 3D images, to set up clustering experiments. The diagnostic power needs to be assessed for genetically and phenotypically similar syndromes, i.e. FGFR-related craniosynostosis syndromes such as Muenke and Crouzon.

- **Objective 2: To quantify 3D shape development in the growing Apert's face and its facial skeletal abnormalities.**

3DMMs of the mean face of Apert will be compared to mean age-matched normal models to assess differences. 3DMMs of the normal paediatric facial skeleton and the Apert's facial skeleton will be constructed providing information of the Apert face that might give new insights in natural shape development over time and facial areas that worsen or improve with age.

- **Objective 3: Define and apply the optimal quantification method to assess surgical outcomes in Apert's corrective facial surgery.**

Once the models are built, the most accurate method to quantify shape differences will need to be determined. Once identified, the most accurate method is then applied to evaluate post-operative outcomes.

- **Objective 4: Evaluate facial normalisation after corrective surgery.**

The normal and Apert soft tissue models are used to evaluate the differences between the post-operative and age-matched healthy population. These differences can provide insight in the areas where the current surgical approach is successful and regions where normalisation might be lacking.

1.3. Outline of thesis

The remainder of this thesis has the following structure:

Chapter 2 provides background information on healthy craniofacial shape development and Apert syndromic abnormal growth. It presents a history of craniofacial morphometrics and describes the different types of surgical procedures known from literature to correct the facial deformities apparent in Apert syndrome.

Chapter 3 describes face features of Apert syndrome compared to the normal population, using 3DMMs and heat maps. This model introduces the potential power of state-of-the art algorithms for automated syndrome identification.

Chapter 4 demonstrates that genetically and phenotypically similar syndromes to Apert syndrome, i.e. FGFR-related craniosynostosis syndromes (Crouzon and Muenke syndrome) can be identified both from each other and from normal faces, based on shape alone, demonstrating the powerful diagnostic value of this information.

Chapter 5 presents bone 3DMMs for a normal paediatric population (below 4 years of age). The model is validated by evaluating intrinsic model characteristics and by comparing with gold standard data from literature.

Chapter 6 applies the bone model construction methodology from chapter 5 to an unoperated Apert population. In addition, growth curves on various midfacial and cranial measurements are provided and compared with normative data.

Chapter 7 evaluates the surgical outcomes of facial bipartition distraction, a common surgical technique to correct Apert's facial deformities, by analysing local shape changes of both soft tissue and bone.

Chapter 8 determines whether facial bipartition objectively normalises the Apert's face, by comparing the post-operative face with generated normal faces. The method highlights where facial bipartition distraction is successful in normalisation and in what facial regions the surgical technique may require improvement to achieve a more normal appearance.

Chapter 9 provides a summary of the main findings of this thesis, outlining the contribution to current understanding of facial shape changes in the growing Apert patients and their facial corrective surgical outcomes. Suggestions to improve clinical practice and for further research are presented, including some preliminary results of future studies.

2. BACKGROUND

Part of the work described in this chapter has been published in:

Journal of Orthodontics, 2019 (invited review)

Geometric morphometrics aided by machine learning in craniofacial surgery.

- **Lara S. van de Lande**, Athanasios Papaioannou, David J. Dunaway.

Plastic Surgery: Principles & Practice 1e, Farhadieh. Publisher: Elsevier, 2020. (book chapter)

Chapter 20: Craniosynostosis

- **Lara S. van de Lande**, Aina. V.H. Greig, David J. Dunaway

Aesthetic Surgery of the Facial Skeleton, Baker. Publisher: Elsevier, 2020. (book chapter)

Chapter 3: Anthropometrics.

- **Lara S. van de Lande**, Athanasios Papaioannou, David J. Dunaway

This chapter describes Apert syndrome and introduces other FGFR-related craniofacial syndromes. An overview of prior studies on shape analysis in Apert syndrome is presented along with current concepts in modern morphometrics. Finally, the chapter outlines routine surgical approaches to correct the Apert's face.

2.1. Apert syndrome and FGFR-related craniosynostosis

2.1.1. Genetical mutations

In Apert syndrome, most cases are sporadic, although autosomal dominant inheritance can occur. The cause is found on FGFR2 gene on chromosome 10q in 99% of the cases (10). There are two known main subtypes, distinct by affecting different amino acids, namely $FGFR2^{S252W}$ (Ser252Trp) and $FGFR2^{P253R}$ (Pro253Arg) (10). Besides these mutations, a deletion and insertion in exon IIIc of FGFR2 have also been described as a rare cause (24). All these genetic changes have a gain-of-function effect, leading to abnormal expression of the altered FGFR2 splice form in the mesoderm derived mesenchyme of the coronal suture, inducing accelerated proliferation and differentiation of the mesenchymal cells causing osteogenesis. As a result, the coronal suture fails in its function to facilitate skull growth, and synostosis of the frontal bone to the parietal bone occurs (25-27). While these two subtypes cause similar facial deformities, cleft palate is more common in Ser252Trp patients. Syndactyly of hands and feet are significantly more present in patients with Pro253Arg mutation (28).

2.1.2. FGFR-related craniosynostosis

Craniosynostosis is defined as premature fusion of cranial sutures, which can be due to primary or secondary causes. The premature closure of one or more sutures results in a distorted head shape owing to growth restriction along the affected suture and to compensatory perpendicular growth. FGFR-related craniosynostosis consists of FGFR1, FGFR2, and FGFR3 mutations. Depending on the amino acids affected, various types of syndromes result. The most common syndromes from the FGFR2 related mutations are Apert, Crouzon, and Pfeiffer syndrome. Muenke and Crouzon with acanthosis nigricans result from the FGFR3 mutation. Pfeiffer can also derive from FGFR1 mutations. Associated features include facial hypoplasia and dysmorphology, ocular exposure, hydrocephalus, hearing problems, mental retardation, delayed speech, feeding problems, extracranial anomalies, and many other functional problems deriving from this. Apert syndrome accounts for between 4 to 5% of all patients with craniosynostosis (29).

2.1.3. Face Appearance

FGFR-related syndromes are characterised by typical craniofacial features, however large phenotypical variety can be seen between and within different syndromes. Pfeiffer syndrome is characterised by brachycephalic headshape (short head shape), maxillary hypoplasia (retruded midface) and exorbitism (protrusion of the eyeball), Crouzon syndrome by brachycephalic headshape, maxillary hypoplasia, hypertelorism (increased inter-orbital width), exorbitism with retruded supraorbital, infra-orbital and lateral orbital rims, and class III malocclusion. Muenke syndrome is characterised by brachycephalic head shape (short head), mild midface hypoplasia, downslanting palpebral fissures (lowered position of the opening between the eye lids), and mild ptosis (protruding eye), and Apert syndrome is

characterised by a turribrachycephalic head shape (short and tall head), hypertelorism, exorbitism, vertical compaction toward the nasofrontal suture, downslanting palpebral fissures, maxillary hypoplasia, class III malocclusion (protruded lower jaw relative to the upper jaw), anterior open bite and dental crowding (Figure 2.1) (30).

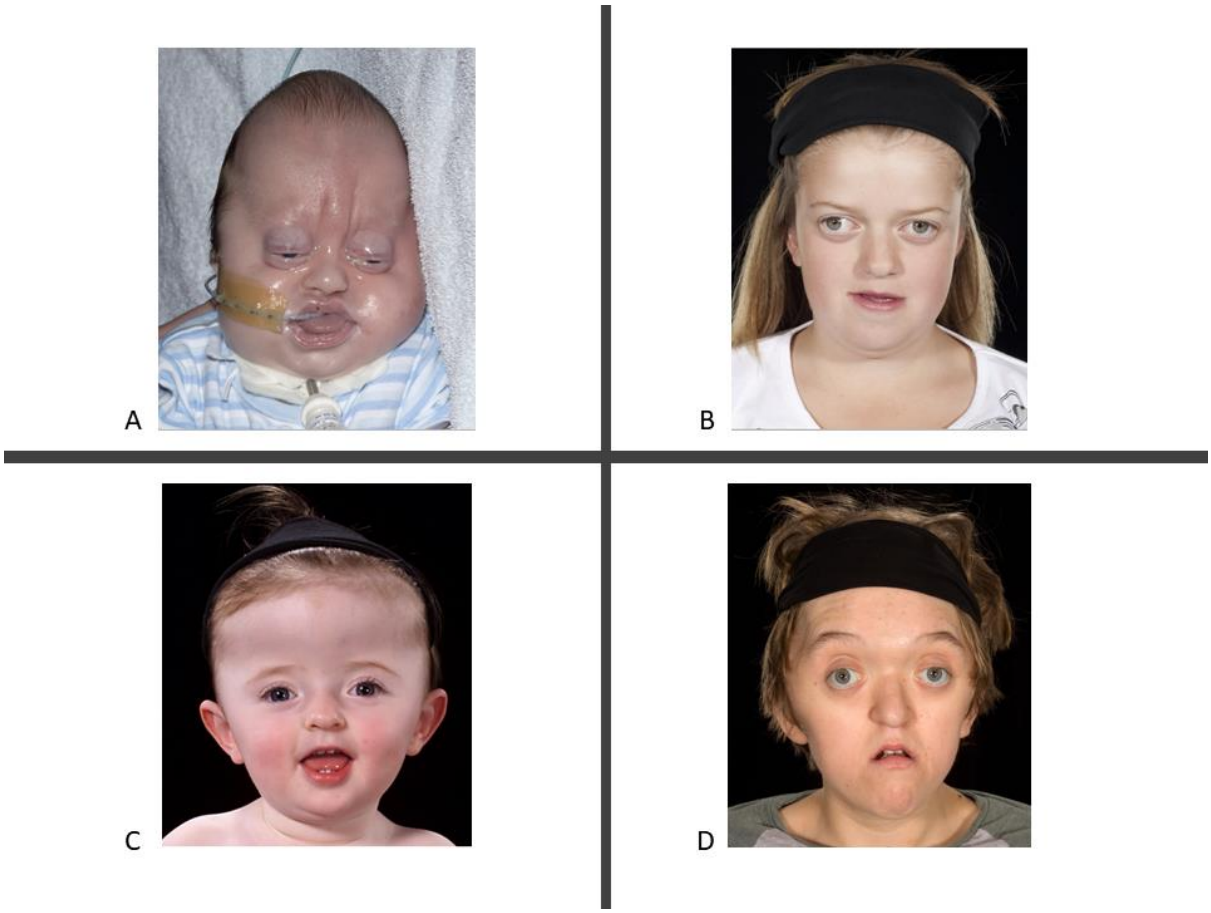


Figure 2.1. Clinical photographs of unoperated patients with FGFR-related craniosynostosis. A) a 30-day old male patient with Pfeiffer syndrome; **B)** a 7-year-old female patient with Crouzon syndrome; **C)** a 1.3-year-old female patient with Muenke syndrome; **D)** a 17-year-old male patient with Apert syndrome.

2.2. Current concepts in Craniofacial morphometrics

2.2.1. Traditional morphometric analysis

Traditionally, analysis of the craniofacial shape has been performed using linear measurements based on lengths, angles and ratios, often acquired from 2D projection images (11). These measurements are referred to as traditional morphometric analysis. Although manual acquisition of anthropometric data is time consuming, and requires great precision and a degree of expertise to record, traditional morphometric techniques form the basis of our understanding of the proportions of the human face and, thanks to the pioneering work of Leslie Farkas and others, large amounts of normative 2D data are available (31). Traditional morphometric analysis is useful in corrective surgery and has revolutionised orthognathic surgery and orthodontics with the cephalometric analysis. Despite its success, 2D analysis has numerous drawbacks and cannot capture the full complexity of the head and facial features in 3D.

2.2.2. Geometric morphometric analysis

More recently, geometric morphometric analysis has been introduced to study the human head shape, using Cartesian landmark coordinates in 3D and dense 3D analysis or statistical shape modelling (SSM). For this, a 3D shape of the head is required, acquired from CT scans, MRI, stereo photogrammetry, laser-scanning, handheld structured light scanners, or 3D photographs. Dense 3D analysis can provide extensive information, in contrast to sparse annotations, which, being defined by a finite number of anatomical measurements, can only offer limited insight into the shape of the human skull or face. Additionally, in dense 3D

analysis, the shape is defined as the geometrical information that remains when translation, rotation, and scale effects are removed, thus removing overall size effects which may be predominant in the growing child (13, 14).

2.2.3. 3D face models

3DMMs are a common SSM approach. Blanz and Vetter pioneered the use of SSM for 3DMMs in 1999 and have since then been applied to a variety of tasks (14). For over twenty years they have impacted 3D analysis with their ground-breaking approach and in the past few years, re-discovery has taken place in the context of deep learning.

A 3DMM is a generative model for face shape and appearance that is based on two key points: 1) all faces are in dense point-to-point correspondence and 2) the facial shape and colour are separated and disentangled from external factors such as illumination and camera factors. They have advantage of the availability of depth information. When a statistical model of the distribution of the faces is made, this is traditionally done using PCA and has included other learning techniques in subsequent work, such as autoencoders.

The leading assumption for the development of 3DMMs was that prior knowledge about object classes plays an important role in vision and helps to solve otherwise ill-posed problems. 3DMMs are designed to capture such prior knowledge and they are learned automatically from a set of examples. Prior to the time of Blanz and Vetter, Sirovich and Kirby in 1987 and Turk and Pentland in 1991 made an important influential paradigm shift with the Eigenfaces approach for 2D models. Grey-levels in the image domain were used and wrapping was performed using one single point, e.g. tip of the nose. PCA was performed and the eigenvectors could represent the main modes of variation in that space. However, with a drawback it being limited to a fixed pose and illumination and had no effective representation

of shape differences. The Eigenfaces approach was extended to the 3D face surfaces by Atick *et al.* 1996 to model shading variations in faces, having the same limitations (32). Other research groups proceeded by adding an Eigendecomposition of 2D shape variations between individual faces and used significantly more points for wrapping to obtain correspondence. Eigendecomposition is used to decompose a matrix into eigenvectors and eigenvalues, which can be used for PCA. In 1991, landmark-based face warping for image analysis was introduced by Craw and Cameron (33). The first statistical shape model was proposed in Active Shape models by Cootes *et al.* 1998 using about 200 landmarks (34). While this model used shape only, Active Appearance Models proposed a combination of shape and appearance and turned out to be very successful. Other groups computed dense pixel-wise image correspondences with optic-flow algorithms for modelling the facial shape variations (35, 36). In all these correspondence-based approaches, images are warped to a common template, and the appearance variation is then performed in the same way as the original Eigenfaces, but on the shape-normalized images. The shape model, on the other hand, provides a powerful and compact representation of shape differences by shifting pixels in the image plane. These 2D models were efficient to cover the shape variation for a fixed pose and illumination setting, but required many separate models, each limited to a small range of poses and illuminations. The introduction of 3D Computer Graphics in the 1990s demonstrated that variations in pose and illumination can easily be simulated. 3D models were used to manipulate existing images and 2D algorithms were applied on 3D surfaces, as at the time image-based models were rather elaborate. With the initial face scanner delivering surfaces in a 2D cylinder parameterisation, all those steps were performed in 2D, and were replaced with their 3D equivalent only many years later. Interestingly, after the development towards 3D, the approach came back to 2D representations by using deep learning, and has now again evolved to 3D, e.g., by integrating 3DMMs.

In the seminal work of Blanz and Vetter, they built a 3DMM of the human face by applying PCA to a dataset of 200 densely registered subjects, 100 males and 100 females of a similar ethnicity and age. Their dataset relied on laser scanning (37), which has a big disadvantage of acquisition time, as only very few samples are gathered at any given time - even at very high frame rates, such systems require the subjects to sit still for several seconds. Following the model proposed by Blanz and Vetter, various have followed. Table 2.1 provides an overview of all available models of the human shape.

Table 2.1. Overview on available shape and/or appearances models of human faces (38).

model	geometry	data	comment
Basel Face Model 2009 (39)	shape	200 individuals, each in neutral expression	includes separate models for facial parts
FaceWarehouse 2014 (40)	shape, expression	150 individuals, each with 20 expressions	
Global and local linear model (41)	shape	100 individuals	
Multilinear Wavelet model 2014 (42)	shape, expression	99 individuals, 25 expressions	
Multilinear face model 2015 (43)	shape, expression	2,500 scans (100 individuals, 25 expressions)	
Multilinear face model 2016 (44)	shape, expression	2,510 scans (205 individuals, up to 23 expressions	
Large Scale Facial Model (LSFM) 2016 (5)	shape	9,663 individuals	
Surrey Face Model 2016 (45)	shape, expression	169 individuals	multi-resolution
Liverpool-York Head Model (LYHM) 2017 (4)	shape	1212 individuals	full head (no hair, no eyes)
Faces Learned with an Articulated Model and Expressions 2017 (46)	shape, expression, head pose	3,800 individuals for shape, 8,000 for head pose, 21,000 frames for expression	female, male, gender neutral model, full head (no hair)

Basel Face Model 2017 (47)	shape, expression	200 individuals for shape and appearance, a total of 160 expression scans	full head and multi-resolution
York Ear Model 2018 (18)	shape	20 3D ear scans, augmented with 605 landmark-annotated 2D ear images	ear only
Multilinear autoencoder 2018 (48)	shape, expression	5,000 scans from 195 individuals, 500,000 after augmentation	
Convolutional Mesh Autoencoder 2018 (22)	shape, expression	12 individuals, 12 extreme expressions, 20,466 meshes in total	full head (no hair)
Combined Face & Head Model 2019 (15)	shape	Merged from LYHM and LSFM models	full head (no hair)
Morphable Face Albedo Model 2020 (49)	-	73 individuals (50 scanned + 23 from a 3D relightable clinical database (50)	

Over the past years, 3DMMs were applied beyond faces. Models were built for the surface of the human body (51, 52) and for other specific parts of the body like ears (18) and hands (20). Moreover for animals (53, 54) and even cars (55).

The process of building a 3DMM consists of collecting and standardising a training set of objects; aligning the object to a common template by removing translation, rotation, and scale (alignment); bringing every object into dense correspondence (registration); and, finally, analysing the statistical distribution of the registered shapes. These models are often constructed using some form of dimensionality reduction, as described - typically by applying PCA, though convolutional mesh autoencoder models and generative adversarial network (GAN) for 3D faces have also been used (22, 48, 56, 57). PCA is an algorithm that extracts

shape modes by maximising the variance along orthogonal shape directions. The construction of a 3DMM is feasible if and only if all meshes are re-parametrised into a consistent form where the number of vertices, the triangulation, and the anatomical meaning of each vertex are made consistent across all meshes. For example, if vertex *i* corresponds to the tip of the nose in one mesh, then in all other meshes the i^{th} vertex also corresponds to the tip of the nose. When the above criteria are met, the meshes are said to be in dense correspondence.

When 3D meshes are derived from CT scans, the triangulation, number of vertices, and the semantic meaning are inconsistent, which needs rectification prior to dense correspondence. This can be achieved by registering the index mesh with a mesh template of fixed topology. One of the most common approaches to achieve this is non-rigid iterative closest point algorithm (NICP) (Figure 2.2.), which has proven more accurate than other approaches such as UV-TPS (UV thin plate spine) or UV-OF (UV optical flow) (5, 58, 59). NICP can be achieved with or without landmarks. When NICP is performed without guidance of landmarks it makes use of non-linear transformation methods, such as Finite Elements, or define a non-linear transformation based on thin-plate splines of b-splines. Landmarked based NICP makes use of 3D landmarks that employ as a soft constraint to guide the registration process; several iterations are then required to achieve dense correspondence with the template mesh. NICP is a flexible version of rigid ICP. Rigid ICP is a method Horn in 1987 and Arun 1987 that makes use of two sets of corresponding landmarks that computes a rotation and translation by employing a linear algebra. It is faster than NICP but has also proven to be less accurate (16). To date, most commercial navigational or in silico 3D surgical planning softwares, such as BrainLab and Metronic, still use rigid registration methods as these are much faster.

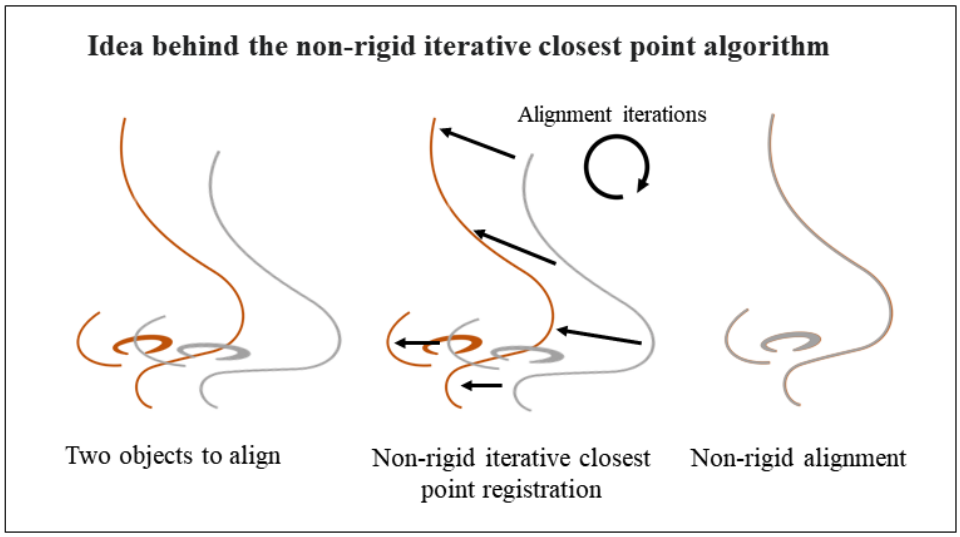


Figure 2.2. Simplified overview of non-rigid closest point algorithms. This illustration visualises landmark based NICP, the first image illustrates two objects one wants to align, step 2 makes use of a selected number of landmarks to guide the registration process to obtain that the tip of the nose will indeed be aligned with the tip of the nose of the template mesh. It requires multiple iterations before appropriate alignment is achieved, which is illustrated with the two noses in dense correspondence in the right image.

2.2.4. Large Scale Facial Models

To current data, LSFM is the largest 3DMM ever constructed (60). This model is trained on rich demographic information with a large variety of ages (the majority >4 years), gender (52% female), and ethnicity (82% White, 9% Asian, 5% mixed heritage, 3% Black, and 1% other). The population is described by the mean face and its principal components of shape variation (eigenvectors) which describe a multidimensional face space or face manifold. There are several thousand possible eigenvectors within a facial model (as many as there are subjects in the dataset from which the model is constructed), but nearly all the facial variations of the population are described within the first 100 components. As the dataset used to build the model is very large and extremely diverse, bespoke models can be formulated such as models

for females and males (Figure 2.3) or models tailored to different age-groups (16 to 21, 21 to 35, 35 to 50, and 50 to 90 years old) (Figure 2.4).

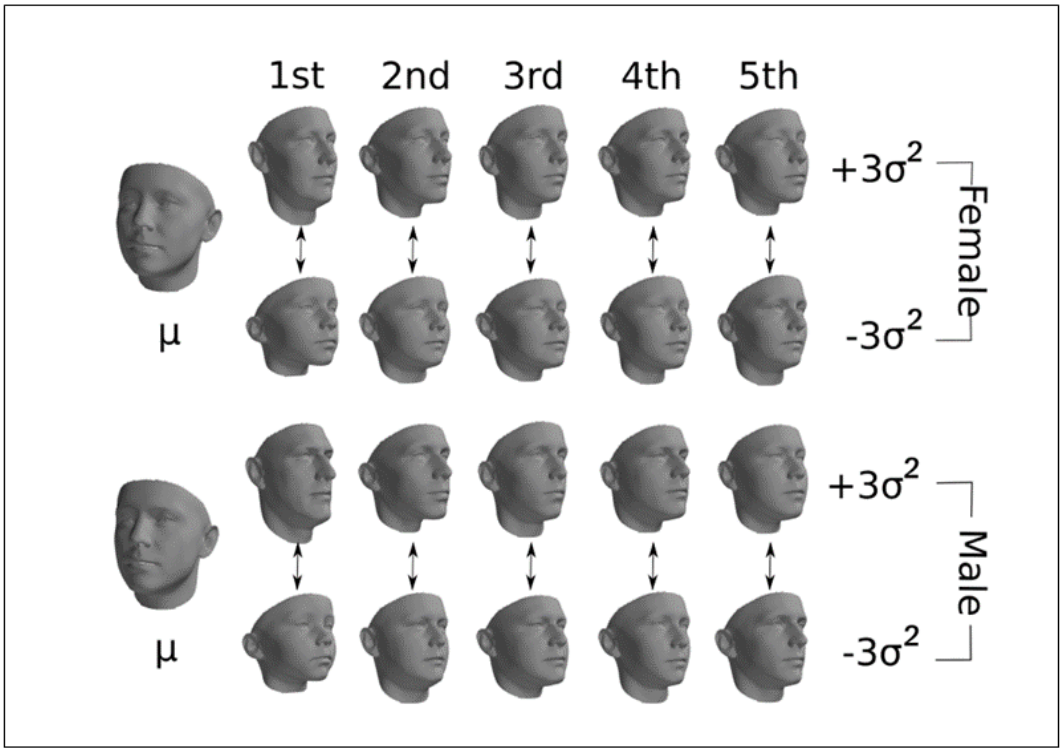


Figure 2.3. Bespoke LSFM models for gender. The mean (μ) and first five principal components with +3 and -3 standard deviation are illustrated for the female (top row) and male (bottom row) population computed from the Large-Scale Face Model (LSFM) database. It illustrates the largest differences for the study population (7-90 years old). Facial lengthening (component 1) and midfacial prominence (component 2) are some of the main findings.

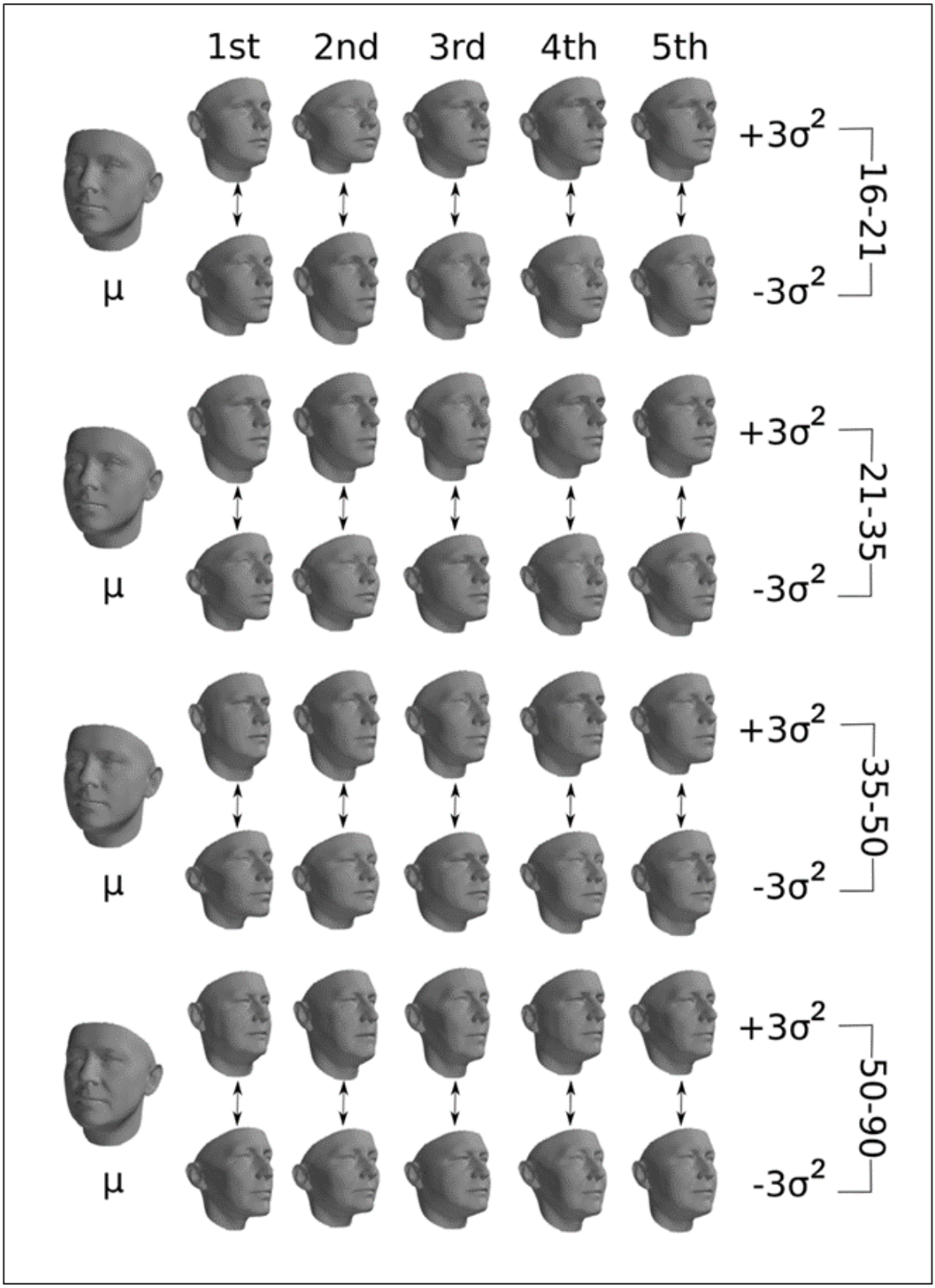


Figure 2.4. Bespoke LSFM models for age-groups. The mean (μ) and first five principal components with +3 and -3 standard deviation are illustrated for the 4 age-groups. These models include both genders and illustrate the main variations.

Various methods to analyse this 3D information exist, of which t-Distributed Stochastic Neighbour Embedding (t-SNE) is one of the most common and popular ones. t-SNE is a dimensionality reduction technique that allows the high dimensional shape vectors to be embedded in a lower dimensional space and can reveal hidden structures of the data (61). It gives each datapoint a location on a 2D map and produces good visualisation by reducing the tendency to crowd points together in the center of a map. When compared to other non-parametric techniques for dimensionality reduction, it is shown to be better at creating maps that reveals structure at many different scales. A study by Maaten *et al*. compared t-SNE with 7 other techniques, which can be found in (61).

For the LSFM model clustering can be seen for age and ethnicity on the t-SNE plot (Figure 2.5). A clear trend from younger subjects grouped on the far left of the embedding, with a tendency towards older ages when moving further to the right and towards the edges of the main cluster. Distinct clusters are also observed for Asian and Black participants. These findings suggest a potential for the use for the use of 3DMMs, such as LSFM, for diagnostics.

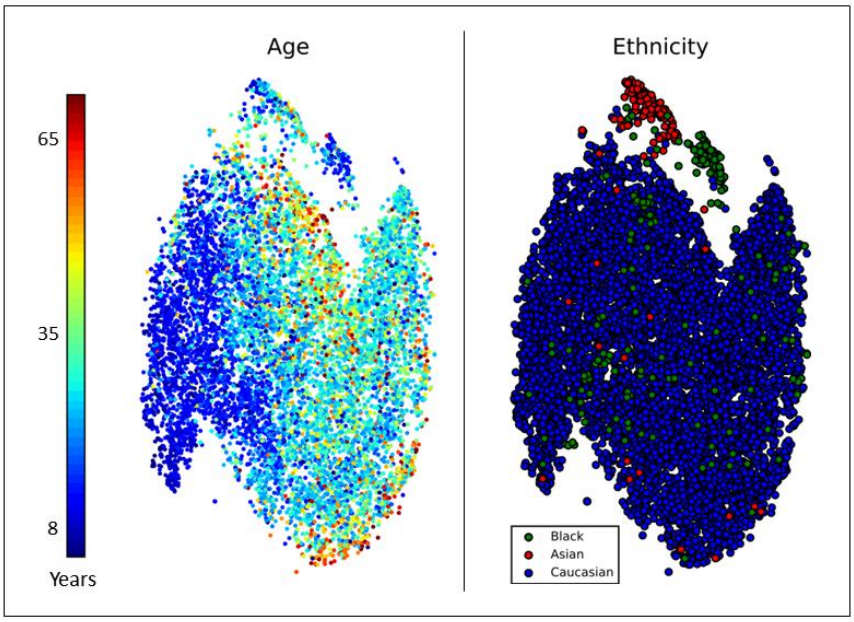


Figure 2.5. t-SNE embedding of the high-dimensional face manifold clustering for age and ethnicity. Two embeddings are shown in this figure, left: age embedding, demonstrating a clear trend for age, right: ethnicity embedding where three main groups are seen for different ethnicities.

2.2.5. Shape analysis of the unoperated Apert's face

Some of the earlier works analysing the Apert's head and face are studies by Goldberg *et al.* 1981 and Cohen *et al.* 1993, which provided several growth charts for cranial measurements. Head circumference measurements and cranial configurations such as cephalic length, breadth, and cranial height were acquired from adult lateral and frontal cephalometric tracings and compared with the dimensions from a normal population (62, 63). Later, some CT-based studies were published utilising cephalometric measurements to analyse Apert populations ranging from 8 – 25 patients. Reitsma *et al.* 2012 reported larger SNB angles (the angle between sella, nasion, and B-point (point of the deepest concavity anteriorly on the mandibular symphysis), showing the anterior limit of the mandibular basal arch in relation to the anterior cranial base) for Apert syndrome than for controls, with decreasing differences in adulthood; the ANB angles (the angle between A-point (the deepest point on the curvature of the maxillary alveolar process), nasion, and B-point, showing the relationship between the maxilla and mandible) increased over time as the Apert population developed a more severe Class III malocclusion (64). The inclination of palatal plane to the anterior cranial base was smaller than controls and the ratio of the lower anterior facial height and the total anterior facial height ratio was larger at all ages, with differences increasing in boys and decreasing in girls with age (Figure 2.6). In concordance with this study, a shorter anterior cranial fossa without widening of the anterior cranial fossa was also reported by Antonio Jorge *et al.* 2014. A shorter maxilla was noted with widths similar to that in controls (65).

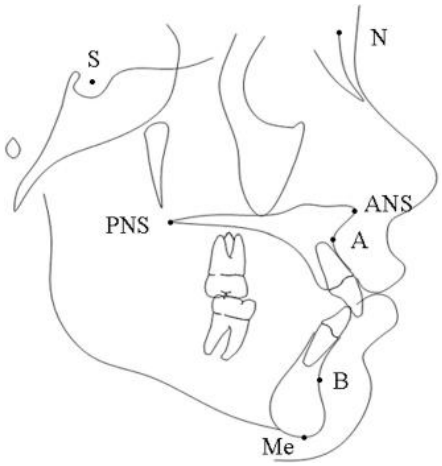


Figure 2.6. Description of cephalometric measurements. Lateral skull with landmark points. ANS = anterior nasal spine, Me = menton, N = nasion, PNS = posterior nasal spine, Point A = subspinale, the most posterior midline point in the concavity between ANS and prosthion (the most inferior point on the alveolar bone overlying the maxillary incisors); Point B = supramentale, the most posterior midline point in the concavity of the mandible between the most superior point on the alveolar bone overlying the mandibular incisors (infradentale) and pogonion (the most anterior point on the chin); S = sella, the geometric center of the pituitary fossa.

A work by Posnick *et al*. 2007 studied CT scans from 8 patients (7 female) under the age of 2 years and demonstrated a widened inter-coronal distance (110%) and cranial width, widened anterior interorbital distance (117%), increased lateral interorbital distance (113%) increased intertemporal distance (122%), larger distance between zygomatic buttresses and inter zygomatic arch distance. Yet, a shorter maximal cranial length, shorter medial orbital wall distance (92%), and a shorter zygomatic arch length (79%) shortened was reported (66). Another study utilising cephalometric measurements by Lu *et al.* 2018 reported a retruded zygoma, a shorter maxilla, and an increased transverse width of the zygoma with a normal growth of the maxilla in transverse and vertical directions (67). Of these, the retruded zygoma was the most severely affected area, with effects persisting into adulthood.

Using statistical shape analysis, Heuzé *et al.* 2014 studied the facial skeletal shape variations in FGFR-related craniosynostosis syndromes, including Apert syndrome (n=20) (68). In this study, statistical shape analysis was performed on skull anatomical units, such as the upper facial skeleton. The shape variations between $FGFR2^{252W}$ and $FGFR2^{P253R}$ could not

be identified from each other, based on the mean skull shapes or based on the facial skeleton. Non-significant differences were noted for the deformity for $FGFR2^{P253R}$ as compared to $FGFR2^{252W}$, which was in concordance with a much earlier conducted study by Park *et al.* 1995 (69).

2.3. Surgical correction of Apert's facial deformities

2.3.1. Transcranial approaches

To address the midface concavity and hypertelorism seen in Apert syndrome, facial bipartition with or without a rigid external distraction (RED) is often the surgical technique of choice. This procedure is the product of a number of historical developments in craniofacial surgery. In 1987, Ortiz-Monasterio *et al.* built on previous descriptions by Firmin *et al.* in 1974 and Tessier to develop the monobloc advancement, which has proven to be effective in the correction of midfacial hypoplasia and for increasing the orbital volume across the FGFR2 family of craniofacial dysostoses (70-72). The monobloc allows for advancement of the frontal bone and midface en-bloc. It was traditionally a static advancement procedure; however, with a paradigm shift in cranio-maxillofacial surgery for distraction osteogenesis, the monobloc advancement with internal/external advancement was developed, and, using distraction, a more stable movement without the need for extensive grafting could be facilitated (73-75). (Figure 2.7)

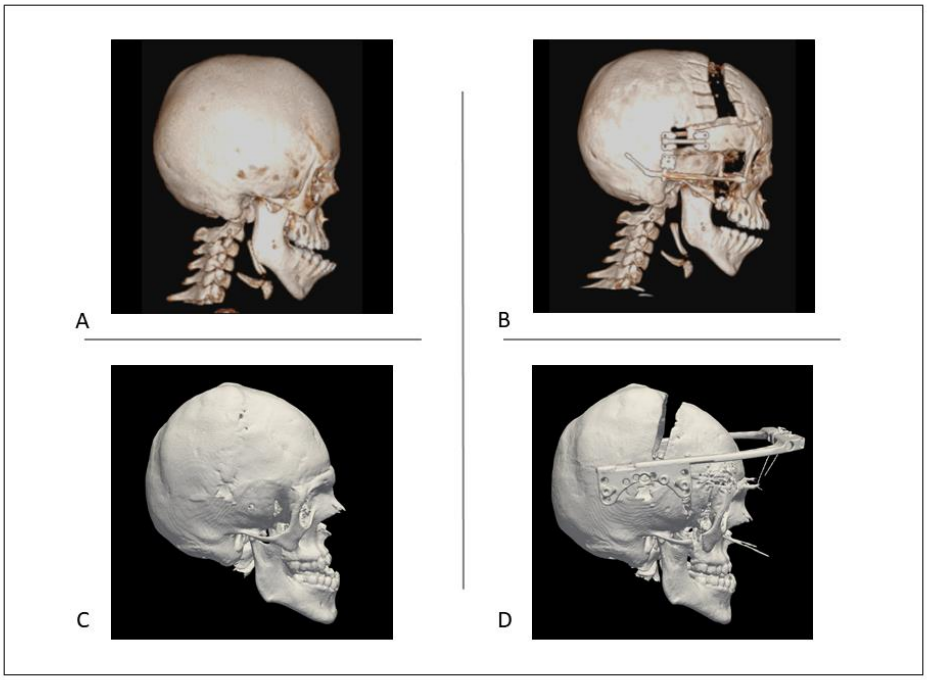


Figure 2.7. Internal distraction versus external distraction. A = pre-operative 3D object constructed from CT scan for a patient undergoing internal monobloc distraction; B = during distraction using internal distractors; C = pre-operative 3D construction from CT scan for patient undergoing external monobloc distraction; D = during distraction using external monobloc distraction.

Monobloc distraction, however, failed to reduce the interorbital width seen in Apert syndrome and alternative surgeries were introduced. The median faciotomy, described by van der Meulen in 1979 and refined by Tessier, became the facial bipartition operation. This addresses the interorbital issue, providing the ability to rotate the orbits medially (76, 77) (Figure 2.8.). Originally used in the correction of hypertelorism, this approach suits the management of the complex Apert phenotypes. After the introduction of craniofacial distraction osteogenesis, the facial bipartition distraction with RED-frame was born. This procedure has the functional capacity to potentially relieve airway disfunctions, such as sleep

apnoea, and provide corneal protection, with the aesthetic component focussing on reduction of the interorbital width and addressing the midfacial biconcavity (78).

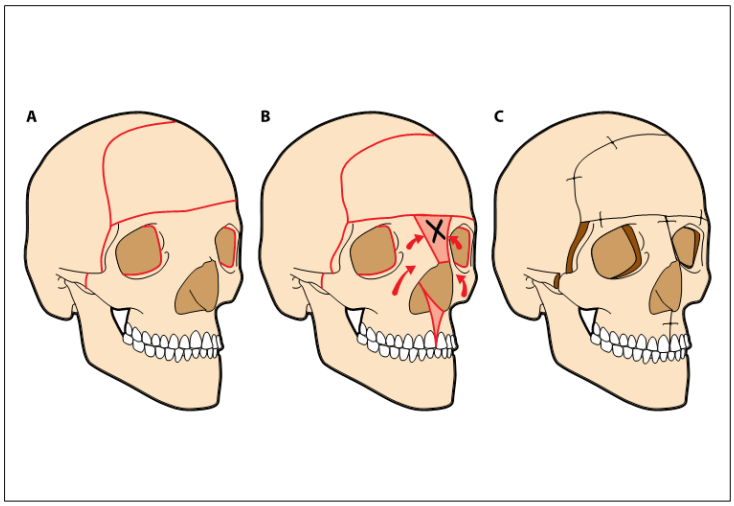


Figure 2.8. Simplified overview of facial bipartition. A = pre-operative skull with simplified osteotomy markings in red; B = peri-operative skull, the red arrows reflect the movement of the osteotomised segments, the 'X' illustrates the interorbital bony wedge selected for removal; C = post-operative skull with the newly positioned bone segments fixated with wires/plates and screws.

2.3.2. Subcranial approaches

In addition to the above transcranial approaches, subcranial techniques have also been described to treat Apert patients. Transcranial approaches have the benefit of simultaneous advancement of the frontal bone and midface; however, they are prone to complications related to the osteotomy that crosses the anterior cranial fossa floor creating a communication with the nasal cavity (79). Subcranial approaches for the correction of facial abnormalities derive from landmark studies on fractures of the skull performed by Rene Le Fort in 1901 (80). The Le Fort III osteotomy for the correction of craniofacial abnormalities was described by Tessier in 1967 and derived from the classification from Rene Le Fort. It was initially introduced for functional problems, e.g. improving airway, occlusion, and globe protection, and later also for aesthetic

purposes - leaving the frontal bone undisturbed (81-86). A Le Fort III essentially moves the midface en bloc into a more ‘normal’ position. (Figure 2.9. and Figure 2.10)

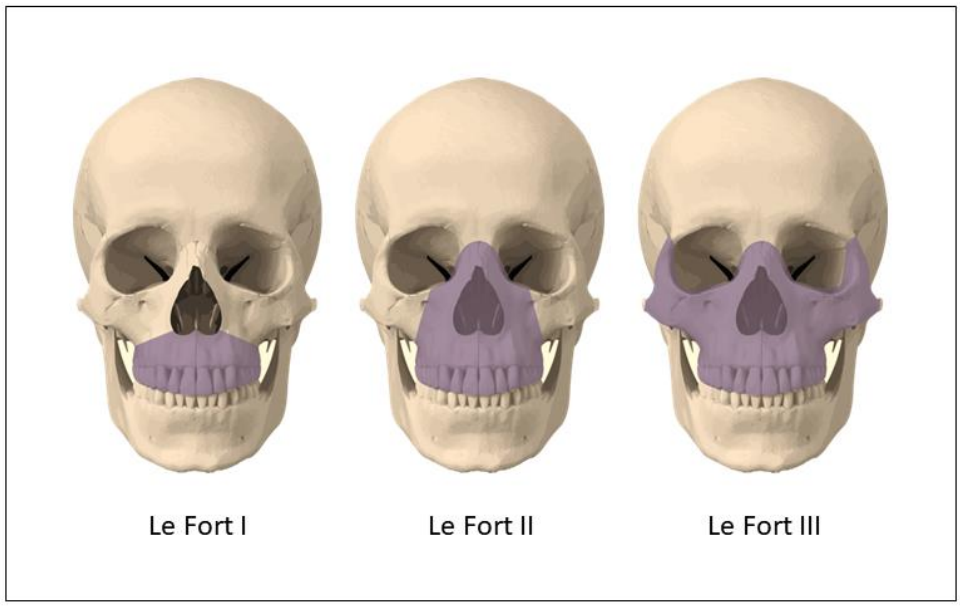


Figure 2.9. Design of Le Fort I, Le Fort II, and Le Fort III osteotomies.

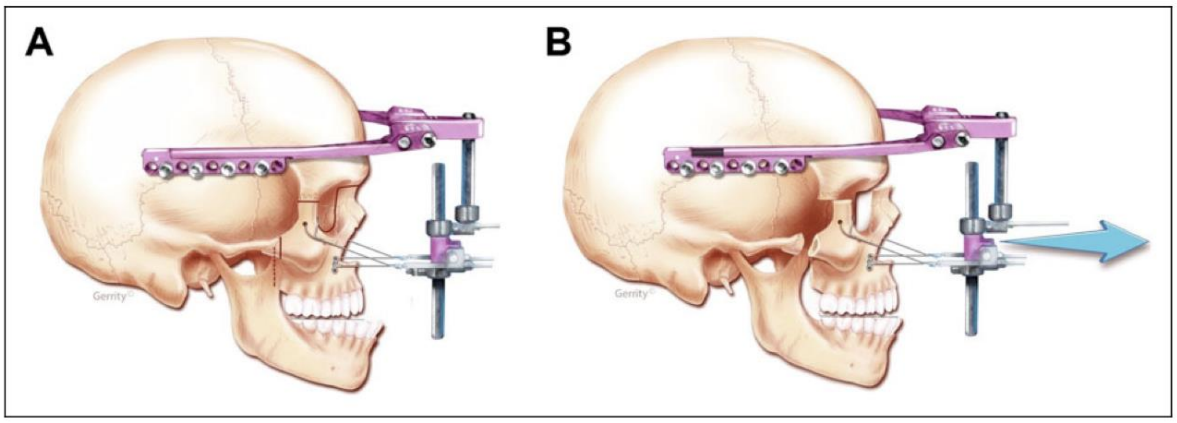


Figure 2.10. A Le Fort III advancement with rigid external distraction frame. A) illustrates clearly an hypoplastic midface and malocclusion, B) shows the how midface distraction using a Le Fort III osteotomy allows for facial profile normalisation and correction the occlusion. This illustration was used from the original work of (1)

Important to note is that Apert syndrome is characterised by an abnormal face in an abnormal position; and therefore, the Apert’s face does not sufficiently benefit from an en-bloc repositioning alone. To solve this, a Le Fort II with zygomatic repositioning (LF2ZR) was

introduced by Hopper (2, 87) (Figure 2.11). It was designed to correct severe functional midface hypoplasia in which there are discrepancies in degree between the central and lateral components. It allows for correction of exorbitism and the midface in different magnitudes – less advancement is needed to correct the exorbitism as to correct the central midface, with anterior and superior repositioning of the zygomas for exorbitism correction, and central midface distraction for sagittal correction and vertical lengthening. The differential movement of the central midface has a beneficial effect on the medial canthus position, which is typically vertically compacted toward the nasofrontal junction in these conditions.

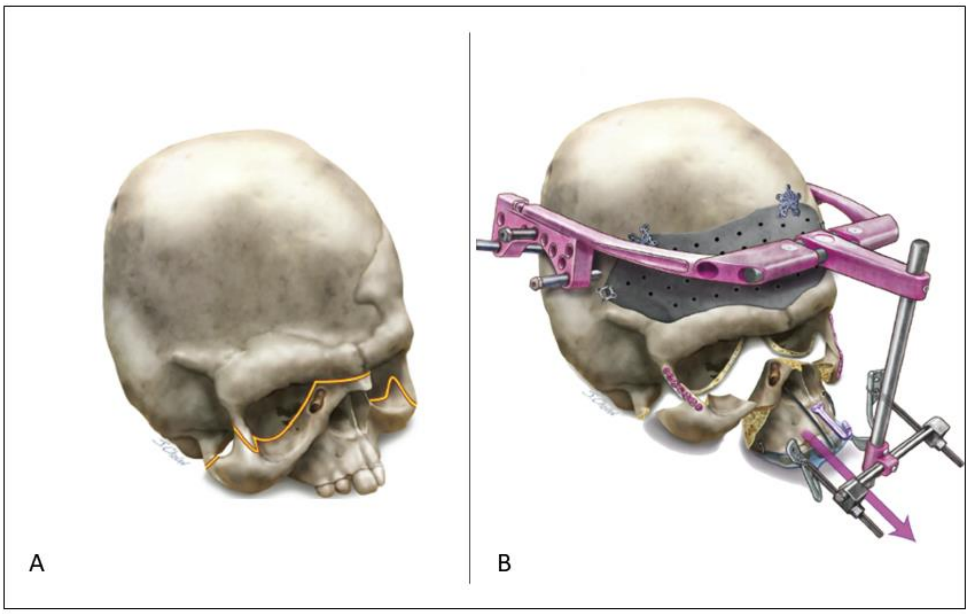


Figure 2.11. Drawing of a Le Fort II midface distraction with external distraction with simultaneous zygomatic repositioning. A) a skull with orbital deformity and retruded midface, the planned osteotomy cuts are illustrated in orange, B) following osteotomies and insertion of rigid external distraction frame, clearly is demonstrated that the central segment has can undergo larger distraction than the lateral parts (2).

2.3.3. Post-surgical outcomes

Various studies have attempted to evaluate surgical outcomes for the correction of the Apert's face. The majority of these are based on measurements on cephalometric films or

acquired from CT scans with the aim to analyse advancements and convexity changes in the facial region following surgery. Given the small number of patients included in these studies, it is impossible to draw conclusions on the most appropriate surgical approach; moreover, studies that assess facial normalisation in 3D are lacking. A study by Oberoi *et al.* 2012, using measurements on cephalometric films concluded that Apert patients undergoing Le Fort III are corrected with great reduction of the deformity. However, only information on the advancement was presented and no analysis was performed on facial normalisation, nor objective calculations on real improvement were performed (88). Some studies demonstrated significant advancement of the midface following surgery such as Le Fort III, LF2ZR or facial bipartition distraction, which can be expected from a distractive approach (66, 87, 89). A study performed by Hopper *et al.* 2013 evaluated the differences between Le Fort III and LF2ZR, using cephalometric type of measurements on CT scans, and found a significant difference in analysing the concavity results (87). This is an interesting study that demonstrated 'normalisation' of facial convexity on worm's eye view (view from below upwards from clinical photographs) of LF2ZR as compared to Le Fort III, which is a technique that translocates the surgical unit en-bloc. Le Fort III is a successful technique to advance the midface, however, it does not correct local shape abnormality and the patient is left with undesirable concavity of the midface and vertical compression. According to this study, the differential movements caused by LF2ZR resulted into facial ratios similar the normal controls. The study claimed that facial ratios were normalised, and that the surgery resulted in levelling of the palpebral fissure by inferior displacement of the medial canthi relative to the fixed lateral canthi as part of the independent inferior Le Fort II distraction vector. A study from our department at *Great Ormond Street Hospital for Children* (*GOSH*) by Ponniah *et al.* 2007 demonstrated that facial bipartition distraction results in significant skeletal unbending of the face, which was measured on pre- and post-operative axial, coronal, and sagittal slides of CT

scans (89). Another study by *GOSH* from Crombag *et al.* 2014 used geometric morphometrics to assess the corrective effects of facial bipartition distraction in a similar study population, which demonstrated midface advancement and inward rotation of the orbits, however, remaining differences were seen when comparing normative data to postoperative Apert scans. In their study they noted that the overall shape of the skull became wider, including the orbits, but the position of the orbits remained the same (90).

A modified technique – a Le Fort III segment correcting corneal-malar relationship and a Le Fort I segment undergoing rotation-advancement to treat the anterior open-bite deformation – described by Takishima *et al.* 2006 was applied to two patients. The orbital contour deformity was preserved and moved as one unit and resulted in independent occlusal correction without orbital distortion. This study was performed using cephalometric measurements (91).

Many surgical techniques and modifications exist; yet no true 3D shape analyses have been performed to analyse post-operative outcomes and to evaluate facial normalisation for both soft tissue and bone, ideally, in order to guide decision-making on the surgical technique of choice.

2.4. Summary

Apert syndrome is a complex syndrome affecting the face and head, in addition to other deformities. Enhancement of early diagnosis, understanding of the natural development of Apert's face and facial skeleton, and objective analysis of the surgical correction compared to normative data are the main topics of this thesis. The next chapter will focus on building a

3DMM of the Apert face to describe its 3D characteristics and compare the syndromic features with those from the normal population.

Key points:

- Apert syndrome is a heterogenous, complex syndrome requiring early diagnosis for optimal outcomes.
- 3DMMs might play an important role in enhancing automated diagnosis.
- Surgical planning can benefit from better understanding of natural development of the growing Apert's face.
- No consensus exists for the most adequate surgical approach to correct the Apert's facial deformities as it is governed by experience rather than objective outcome measurements.

3. APERT FACIAL ANALYSIS

Part of the work described in this chapter has been published in:

Journal of Plastic and Reconstructive Surgery (abstract)

3D Statistical Face Model for Syndromic Craniofacial Patients

- Athanasios Papaioannou, **Lara S. van de Lande**, Stefanos Zafeiriou, David J. Dunaway.

Part of this work was presented at:

- *The 18th Congress of International Society of Craniofacial Surgery (ISCFS), 16-19 September 2019, Paris, France*

As introduced in chapter 1 and 2, 3DMMs may have clinically useful applications within craniomaxillofacial specialties, such as diagnostics. In this chapter, it is hypothesised that automated Apert syndrome identification can be performed using facial shape information alone. This chapter describes the methodology and proof-of-concept study for automated diagnosis of Apert syndrome, starting from the construction of a soft tissue 3DMM of the Apert's face. The model is used to describe the Apert mean face and its variations in 3D, and to compare Apert features with those from normal population, or unaffected, individuals.

3.1. Introduction

The rarity, heterogeneity, and severity of craniofacial syndromes, including Apert, may make clinical diagnosis problematic. Delayed diagnosis can occur in phenotypically mild cases, risking irreversible functional impairments, such as visual failure, neurocognitive defects, and airway problems that can be avoided by timely diagnosis and treatment (92-96). The inadequacy of current screening paradigms, the importance of early triage and identification, and the phenotypical characteristics of different syndromes, make syndromic craniosynostosis a prime candidate for computer assisted diagnosis and referral (97). 3DMMs applied to the normal population have already demonstrated their potential to identify clusters of individuals with similar facial shape features. Knoops *et al.* 2019 demonstrated the use of 3DMMs (LSFM) to identify patients undergoing upper jaw surgery (21). These models constructed from syndromic populations might therefore have potential for automated diagnostics and can be used to understand mean shape differences. Thus, the purpose of this chapter is to construct an Apert soft tissue face model and evaluate the mean differences between Apert and an unaffected population. This information can be used to explore the potential for automated identification of the Apert's face from a normal population using only

shape information as a proof-of-concept for automated diagnostic applications. As described in chapter 2, LSFM is a strong model with a large variety of ages and ethnicity, however, lacks data of young children. Craniofacial syndromes are commonly diagnosed in infancy. Therefore, to test the diagnostic capacity of the state-of-the-art algorithms, a model of the paediatric face needs to be constructed.

3.2. Methodology

3.2.1. Data

Apert syndrome

All patients diagnosed with Apert syndrome at the *Craniofacial Unit at GOSH* and *Department of Maxillofacial Surgery at Hôpital Necker – Enfants Malades, Paris, France* (*NEMH*) between 2005 and 2018 were reviewed retrospectively for pre-operative 3D imaging. CT head scans were selected as the most suitable image modality for the assessment of the craniofacial anatomy and were collected following the inclusion and exclusion criteria reported in Table 3.1. Two subgroups were made: Apert <4 years old and Apert 4 – 17 (including 17) years old. Baseline characteristics were collected from corresponding medical charts. In addition, subtype gene mutation was recorded.

Normative

At *NEMH* CT scans of patients <4 years old without a history of craniofacial anomalies were assessed. Patients indicated for a CT-scan between 2011 and 2018 due to headaches, trauma or epilepsy were reviewed for inclusion. The scans were evaluated by two independent

reviewers: 1) a paediatric radiologist and 2) me, to exclude any scans with abnormalities such as fractures, brain tumours, brain damage or craniofacial anomalies. Henceforth, this database will be referred to as ‘normal <4’.

For the normative data of 4-17 years of age, the original LSFM database was addressed, which derives from 3D photographs. 101 samples between the ages of 4 and 17 years were randomly selected to serve as a healthy control subgroup. This group will be referred to as ‘normal 4 -17’.

Table 3.1. Inclusion and exclusion criteria for dataset. Face models were constructed using CT scans for the Apert <4, Apert 4-17, and <4 normal group. Randomly selected meshes from LSFM were used for the 4-17 normal population. All Apert and normal < 4 samples were acquired via CT scan. The LSFM database was obtained using 3dMD™ photometric stereo capture device set-ups (4, 5)

	Inclusion	Exclusion
CT head scans (2005-2018)	• Apert syndrome diagnosis, no age limits	• Incomplete scans (full face required) • Movement artefacts or squished faces due to gel pads used for scanning. • Insufficient CT slices for 3D construction • Scans after any type of craniomaxillofacial surgery
(2011-2018)	• Normal scans aged <4 years.	• CT visible abnormalities • Craniofacial diagnosis/ abnormalities • Incomplete scans (full face required) • Movement artefacts or squished faces due to gel pads used for scanning. • Insufficient CT slices for 3D construction
LSFM	• 3D meshes aged 4–17 years, randomly selected	• n/a

3.2.2. Image post-processing

DICOM files of all Apert and control CT scans were exported and converted to 3D soft tissue meshes by applying standardised skin setting using Horos (98), an open source code (FOSS) program that is distributed free of charge under the LGPL license at Horosproject.org and sponsored by Nimble Co LLC d/b/a Purview in Annapolis, MD USA. The meshes were exported at standard tessellation language (STL) files (Figure 3.1.). The STL files were imported into Autedesk Meshmixer (99), an open source 3D visualisation software for working with triangle meshes, to undergo a cleaning process where redundant objects, such as draping and gel pads, back of the CT scanner, and lines and tubes, were removed. The files were saved as Object files (Obj.). Infants with a pacifier in their mouth at time of CT scan were imported to Materialise Mimics Inprint 3.0, a 3D software that allows to clean up rough data, (100) and underwent a semi-automated process to remove the pacifier from the mesh (Figure 3.2.). The files were saved as Obj. files. Patients with movement artefacts and/or squished faces due to external forces of gel pads used during CT scanning to constrain movement of the child (Figure 3.2.), were excluded from the dataset.

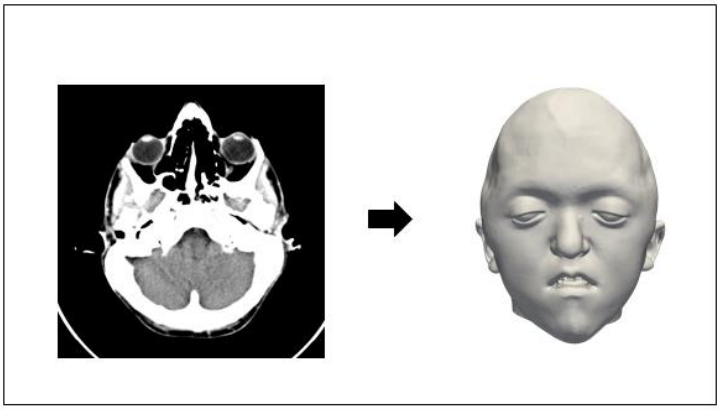


Figure 3.1. DICOM-files to 3D mesh conversion. The DICOM-files (left) were segmented to facial 3D Obj. files (right).

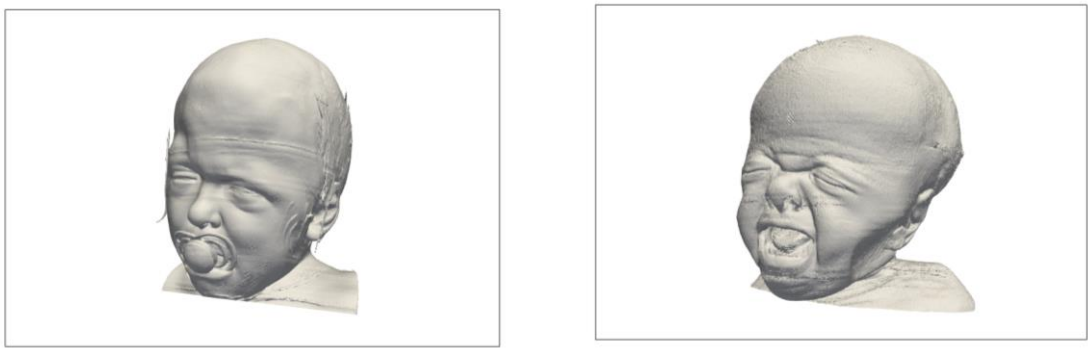


Figure 3.2 Mesh construction issues. Some CT scans needed pre-processing or had to be removed from the database. **Left** shows a mesh where the patient is wearing a pacifier during scanning, this pacifier was removed using Mimics Software, this mesh was then added to the dataset for registration. **Right** is an example of a mesh that was excluded from the database, the mesh is disturbed by movement artefacts, squished soft tissue due to external pressure of gel pads, and a strong facial expression.

3.2.3. Model building and intrinsic evaluation

Soft tissue face 3DMMs models were constructed for the Apert population: 1) all ages, 2) <4 years, and 3) 4-17 years. In addition, a new model was constructed for the unaffected <4-year group and one for the 4-17 normal population.

All meshes from all groups were manually landmarked with 68 facial landmarks derived from a cropped facial template from (60) using a landmarker server (see link 1. Appendix B2 for the webserver used for landmarking) to guide the dense correspondence process. Then, an optimal similarity alignment between mesh in question and the annotated template was performed using Procrustes analysis. Procrustes analysis is a form of statistical shape analysis used to analyse the distribution of a set of shapes. It is a rigid shape analysis that uses isomorphic scaling, translation, and rotation to find the 'best' fit between two or more landmarked shapes (101). NICP was then used to deform the template to take the shape of the input mesh, with the landmarks acting as a soft constraint, to achieve dense correspondence. For the Apert's faces an extra step was needed to achieve dense correspondence as NICP was not sufficient because the shapes of Apert's face differ greatly both from each other, and from

the samples used for the construction of facial 3DMM's. As such, Gaussian Process regressions were applied to the Apert face template mesh. Gaussian Processes is a machine learning algorithm that allows for increase of deformation flexibility and improve the quality of the correspondences obtained. The resulting deformed templates are re-parameterized versions of each subject which are in correspondence with one another (Figure 3.3). Then, all models were derived by applying PCA to the corresponding datasets.

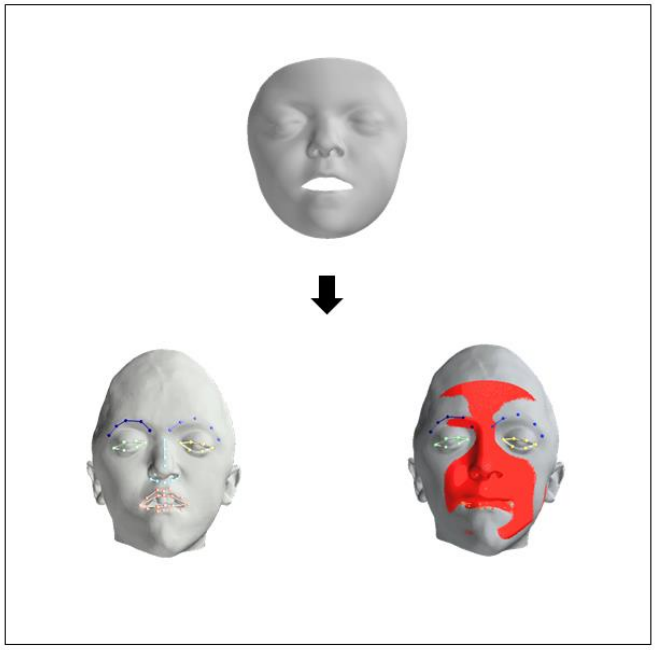


Figure 3.3. mesh annotation to guide dense correspondence. All meshes (top) were landmarked using 68 landmarks (left) to guide the dense correspondence process (right). The example in this illustration is an Apert patient.

The 3DMMs were evaluated for intrinsic characteristics using *compactness, generalisation, and specificity* (41, 43, 102).

Compactness measures the percentage of shape variation explained by the model when a certain number of principal components are retained.

Generalisation measures the ability of a model to represent novel instances of face shapes that were not encountered during training. Given the relatively small size of the dataset,

generalisation was evaluated using a leave-one-out strategy. To calculate the generalisation error for a sample in the test set at a given number of model principal components, the average Euclidean distance (AED) between the sample and its corresponding model projection was computed on a per-vertex basis:

$$AED = \frac{\sum_{i=1}^{n} \sqrt{(x_{i,A} - x_{i,B})^2 + (y_{i,A} - y_{i,B})^2 + (z_{i,A} - z_{i,B})^2}}{n}$$

Here, n is the number of mesh vertices, and x, y, and z represent the Cartesian coordinates of meshes A and B. The overall generalisation error at each principal component is then calculated as the mean per-vertex error over all meshes.

Model *specificity* evaluates the validity of novel instances generated by the model. Again, this was evaluated using a leave-one-out system. To measure the model specificity, 1,000 samples were randomly synthesised (generated) for each of the model principal components. The distance between each synthesised face and its nearest neighbour in the test set was then calculated as the average Euclidean distance over all mesh vertices. Figure 3.4 provides a simplified overview on specificity and generalisation.

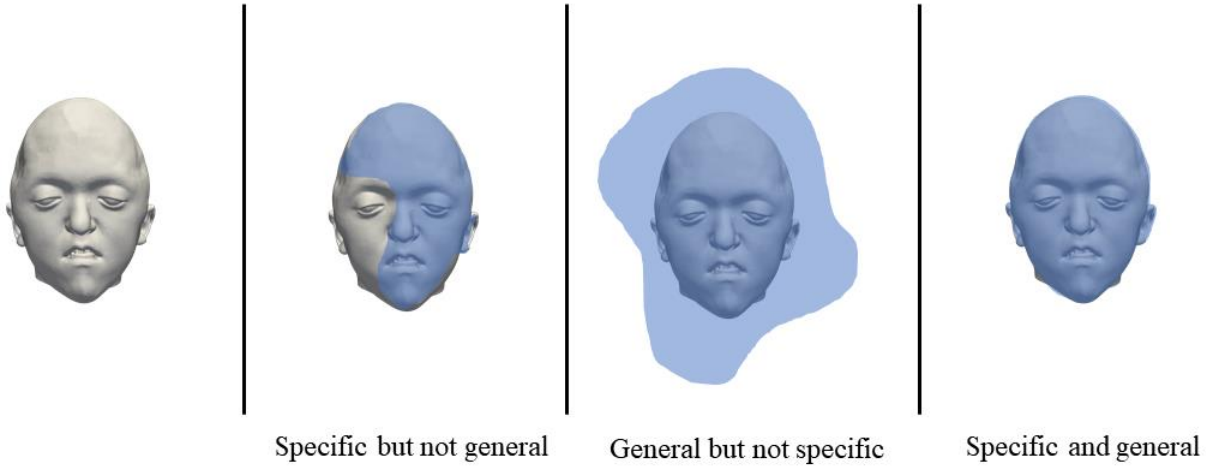


Figure 3.4. Simplified illustration for specificity and generalisation. From left to right, the first image illustrates the reference 3D object. The blue area represents how well the index 3D object corresponds with the reference 3D object. In the second image the blue area fits the 3D object well,

however, does not cover the full area. In the third image, the blue selection covers the 3D object generously, but it is not specific. The final image illustrates a perfect fit, it is specific and general.

3.2.4. Mean facial shape comparison.

Mean faces were computed for each group (Apert <4, Apert 4-17 years, normal <4, and normal 4-17 years). The mean faces of corresponding age groups were compared by superimposition of the meshes to visualise shape differences using distance heatmaps:

- The mean face of the normal <4 was superimposed with the mean Apert <4 years face.
- The mean face of normal 4-17 was superimposed with the mean Apert 4-17 years face.

3.2.5. Manifold Visualisation

To assess the diagnostic capacity of the models, t-SNE was applied to the high dimension latent vector encodings for the Apert patients and unaffected controls. t-SNE embedding was done for the Apert all ages, normal <4, and normal 4-17 allowing the global manifold of these vectors to be embedded in a 2D space for visualisation. The aim of this experiment is to uncover distinct groupings, or clusters. For this experiment, all Apert cases were put into one group, whereas the normal subgroups remained separated as they derive from different datasets, CT for <4 and 3DMD for 4-17 years. All t-SNE embeddings were created using a perplexity of 30 and run for 1,000 iterations. PCA was used to reduce the dimensionality to 30 to speed up the computation of pairwise distances between the data points and suppress some noise without severely distorting the interpoint distances. Each of the dimensionality reduction technique was used to convert the 30-dimensional representation to a 2D map and show the results as a scatterplot. For all the data sets, there is information about the class of each datapoint, but the class information is only used to select a colour for the map points (label). In this case the data was labelled with: Apert (green), normal <4 (yellow), normal 4-17 (red).

3.3. Results

In total, 46 Apert patients could be included for model construction after applying the inclusion and exclusion criteria (Table 3.1.). Genetic subtype information was available for 32 patients: 19 of these were affected by Ser252Trp mutation and 13 by Pro253Arg mutation. Gender and age information on the subgroups are tabulated in Table 3.2.

Table 3.2. Population overview for face model construction.

	n	males	Average age	Age range
Apert all	46	27	6.1 ± 6.2 years	48 days – 17 years
Apert <4	25	16	1.1 ± 1.0 years	48 days – 4 years
Apert 4-17	21	11	12.2 ± 3.7 years	4 years – 17 years
Normal <4	142	79	1.9 ± 1.2 years	1 day – 4 years
Normal 4-17	101	51	10.5 ± 4.0 years	4 years – 17 years

For Apert, models were constructed for Apert all ages, Apert <4, and Apert 4-17. Due to relatively low numbers for the subgroups Apert <4 and Apert 4-17, I will present the intrinsic model evaluations and visualisations of the mean shape and the first components for the Apert all ages model only. For the normal population, models were constructed for normal <4 and normal 4-17. As the normal 4–17-year data derived from the LSFM, which is described at length in chapter 2.2.4. and in (5), I will present the intrinsic model evaluations and visualisations of the mean shape and the first components for the normal <4 only.

3.3.1. Intrinsic model evaluation

The Apert all ages model can be considered reasonably compact; 90% of the model shape variance is captured within the first 10 principal components and 95% within the first 15 components. The model reported low reconstruction errors when assessing accuracy and specificity. The generalisation error was 1 mm when 20 principal components were included, and 0.5 mm when all model components were utilised, indicating that the model generalises well to unseen face instances. Specificity values of <1.2 mm demonstrate that novel face instances generated by the model are realistic (Figure 3.5).

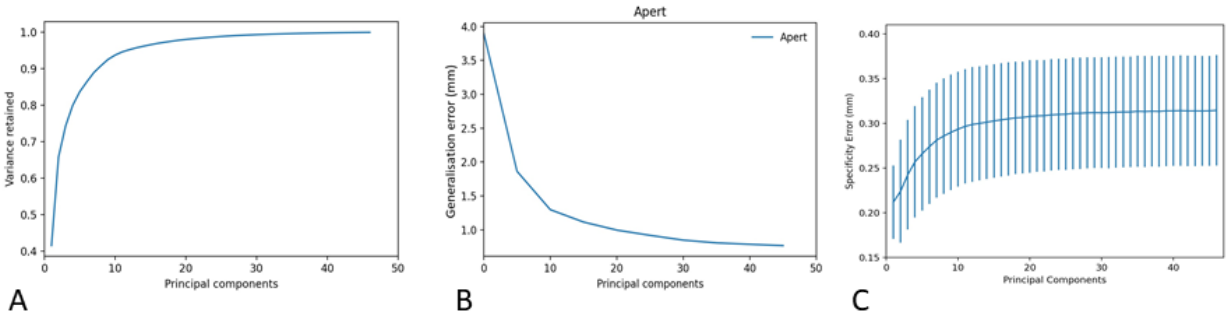


Figure 3.5. Intrinsic model evaluation for the Apert face 3DMM. The presented are plots for generalisation, compactness, and specificity evaluation. **A)** generalisation plot, demonstrates the ability to describe the faces that were not used to construct the original model, and at 20 components is 1.0 mm; **B)** compactness plot, the amount of variance retained for a certain number of principal components, is 94% at 10 components; **C)** specificity plot, measures how well synthetic faces resemble real faces, and is 0.30 mm ± 0.07 mm at 15 components.

For the normal <4 model, 90% of the model shape variance is captured within the first 20 principal components and 95% within the first 25 components. Low reconstruction errors were seen for accuracy and specificity. The generalisation error was 1 mm when 30 principal components were included, and 0.4 mm when all model components were utilised, indicating

that the model generalises well to unseen face instances. Specificity values of <1.0 mm demonstrate that novel face instances generated by the model are realistic (Figure 3.6).

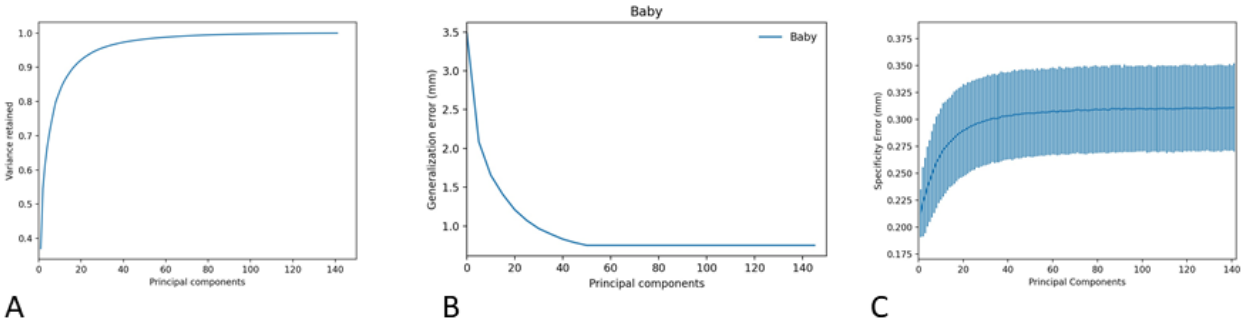


Figure 3.6. Intrinsic model evaluation for the normal <4 face 3DMM. The presented are plots for generalisation, compactness, and specificity evaluation. **A)** generalisation plot, demonstrates the ability to describe the faces that were not used to construct the original model, and at 30 components is 1.0 mm; **B)** compactness plot, the amount of variance retained for a certain number of principal components, is 90% at 20 components; **C)** specificity plot, measures how well synthetic faces resemble real faces, and is 0.30 mm ± 0.05 mm at 40 components.

3.3.2. Mean facial shape comparison.

Mean shape differences of Apert <4 with normal <4, and Apert 4-17 with normal 4-17 are illustrated in (Figure 3.7). Looking at both age groups, differences for Apert are seen with increased facial width, (pre)maxillary/midfacial retrusion, inward positioning of the nasofrontal junction, upward positioned nasal tip, differences in canthus positioning, and globe protrusion can be noted. The magnitude of these differences is well presented in the heatmaps.

For the <4 comparison, differences over 8 mm were computed for facial width and central midface retrusion (premaxillary hypoplasia). Yet, a large proportion of the face remains within differences of 1-2 mm. Very little differences are seen for the medial canthus (<1 mm) and large differences (up to 8 mm) for the lateral canthus.

For the 4-17 comparison, for Apert, the nasofrontal junction is increased (5-6 mm) due to restricted growth as compared to the <4 model (2-4 mm). Differences for the facial width remained, with an increase of differences in eye lid positioning. On average, the Apert's medial

canthus differs 5-6 mm from the normal and the lateral canthus differences remain similar (up to 8 mm). The nasal tip positioning normalises slightly in the 4-17 group (nasal tip differences in the <4 group are 2-5 mm, however, are 0-2 mm for the 4-17 group). Whereas the premaxilla differs on average more in the <4 group (6-8 mm) as compared the 4-17 group (5-6 mm). Although the lateral aspects of the forehead are severely wider in the <4 group as compared to the normal <4 group, generalised increase in shape differences can be noted for the 4-17 group for the facial upper third (Figure 3.8).

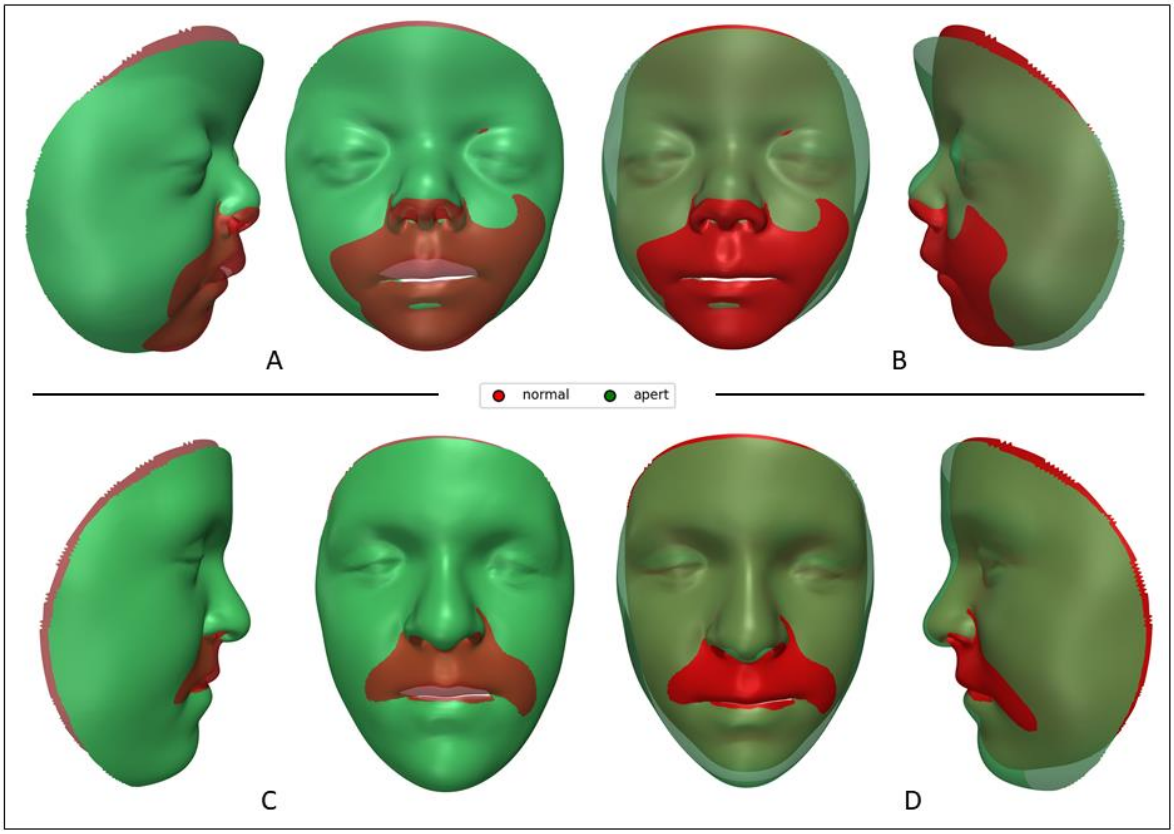


Figure 3.7. Mean facial shape comparison for >4 and 4-17. Shape differences are illustrated for <4 (top) and 4-17 (bottom) age-groups. Two mean facial meshes of each subgroup were superimposed to allow for assessment of the average differences apparent in the Apert <4 vs normal <4 and the Apert 4-17 group vs normal 4-17. **A)** normal <4 mean facial mesh (red, 30% transparency) superimposed with Apert <4 mean face mesh (green), **B)** normal <4 mean facial mesh (red) superimposed with Apert <4 mean face mesh (green, 30% transparency), **C)** normal 4-17 mean facial mesh (red, 30% transparency) superimposed with Apert 4-17 mean face mesh (green), **D)** normal 4-17 mean facial mesh (red) superimposed with Apert 4-17 mean face mesh (green, 30% transparency).

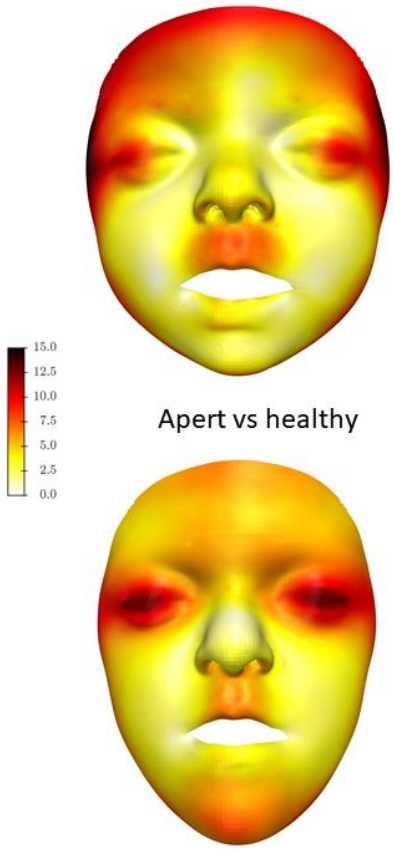


Figure 3.8. Heatmap mean facial shape comparisons. NICP was applied to create heatmaps. The colours that correspond to the differences between the meshes are in millimeters. Top) heatmap of index mesh of mean face normal <4 compared to Apert <4; bottom) heatmap of Apert 4-17 compared with normal 4-17. Natural development of the Apert's face demonstrate mean worsening of the eyelid positioning and overall forehead deformity, slight less differences seem apparent for the (pre)maxilla retrusion.

3.3.3. Manifold visualisation

The t-SNE embedding shows distinct clustering for all data sets, when samples were labelled for Apert (all ages) in green, normal <4 in yellow, and normal <4-17 in red (Figure 3.9.) No false positive or false negative results were identified, suggesting a strong diagnostic capacity of the model. Distinct clustering between normal 4-17 and normal <4 was noted which was expected due to age and different imaging modality used, 3DMD and CT scanning.

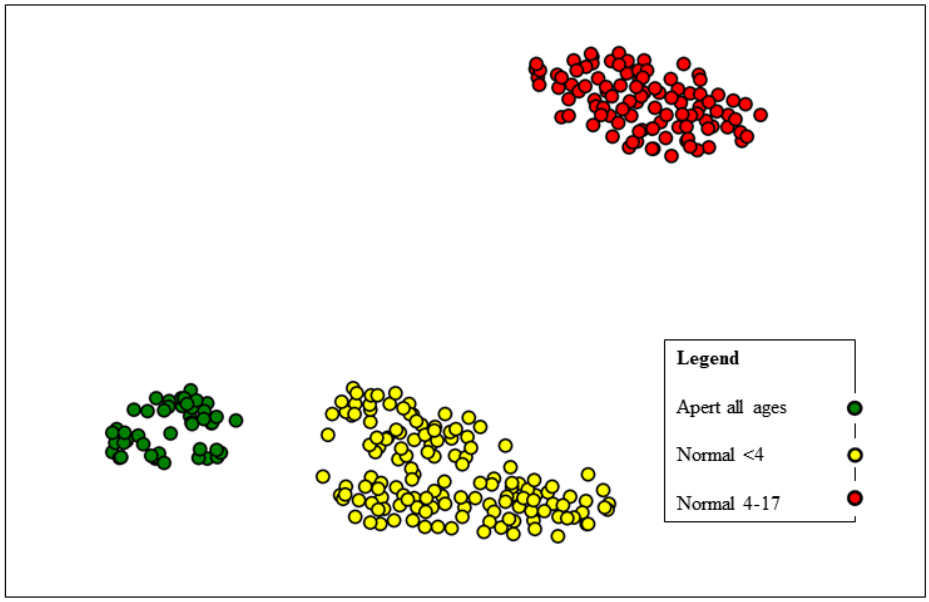


Figure 3.9. t-SNE embedding of the high-dimensional face manifold clustering. The t-SNE embedding in two dimensions was generated with randomly sampled LSFM faces, for visualisation purposes, and labelled according to normal 4-17 (red, n = 101), normal <4 (yellow, n = 178), and Apert all ages (green, n = 46) faces. Two distinct groups are identified without any false positive of false negative results.

3.4. Discussion

This study successfully constructed soft tissue 3DMMs of the Apert face and normal population and allowed to analyse the mean faces of the <4 and 4-17 age groups. As expected, large differences are seen between the mean normal and mean Apert's faces

persisting into adulthood. While facial deformities can become milder over time for some craniofacial syndromes, this was not seen for Apert as increase of severity of most affected areas was noted. Interestingly, the forehead protrusion / retrusion of the nasofrontal sutural region is more extreme in the 4-17 group. Large differences are noted for the eyelids (positioning of the canthus). In concordance with literature, downslanting palpebral fissures are seen in both age groups (30). However, for the Apert <4 group the medial canthus is positioned within normal limits and the lateral canthus lower for up to 8mm, whereas for the Apert 4-17 group the medial canthus is not within normal limited anymore and is positioned higher than the medial canthus of the normal 4-17 group, resulting in a more abnormal palpebral fissure orientation. The findings of the relatively higher positioning of the medial canthus with age are not clearly stated in current literature. Differences of the premaxilla are noted less in the 4-17 group, suggesting slight normalisation with age, yet still significantly out of normal limits. The natural developments of the soft tissue with age have great value for diagnosis and surgical planning. The intrinsic model characteristics validated the accuracy and can therefore be considered to give true information about the anatomical differences. The deformity changes seen with age, of which some have not been described in literature before, may have implications for surgical correction. Moreover, the proposed normative models can be used to further understand where surgical correction might be lacking which will be addressed in chapter 7.

The second part of this study applied a state-of-the-art framework to confirm the hypothesis that machine learning approaches can be successfully used to automatically identify Apert syndrome from an unaffected population. Machine learning techniques work best using large dataset, however craniofacial disorders rarely occur. Apert syndrome for example, has a prevalence of 1:65,000 to 75,000 births (8). Therefore, limited data is

available for model construction. Yet, using only 46 Apert cases to construct the face 3DMM with a non-uniform distribution, promising results for automated syndrome identification were presented in this chapter. This indicates that automated syndrome identification might be possible using 3DMMs. Most craniofacial syndromes severely affect the face shape, such as Apert syndrome; however, they often manifest as head shape deformities caused by craniosynostosis. Therefore, the next chapter will focus on the expansion of this approach to genetically and/or phenotypically similar patient groups to re-evaluate the diagnostic power in a larger setting and will include also head shape information for a complete analysis. Moreover, the presented accurate statistical model in this chapter opens the possibilities for automatic diagnosis assessment of surgical outcomes and surgical design, which will be addressed in the upcoming chapters.

3.5. Summary

The Apert's facial abnormalities are noted to increase with age. The algorithms and framework by Booth et al. (5) demonstrated excellent clustering for Apert from the unaffected population with potential for automated identification for a wider syndromic population and opens possibilities for assessment of surgical outcomes and design.

Key points:

- Mean abnormalities of the Apert's face increase with age when compared to an unaffected population which may have implications for surgical correction.
- Excellent clustering of Apert from the normal using described algorithms suggesting a potential for diagnostic application.

4. AUTOMATED SYNDROME IDENTIFICATION

Part of the work described in this chapter has been published in:

Nature Scientific Reports (original article, published)
Craniofacial syndrome identification using Convolutional Mesh Autoencoders

- Eimear O' Sullivan†, **Lara S. van de Lande**†, Athanasios Papaioannou, N. Owase Jeelani, Christian Duncan, Allan Ponniah, Silvia Schievano, Roman H. Khonsari, Stefanos Zafeiriou§, David. J. Dunaway§
  † / § These authors contributed equally to this work.

Aesthetic Surgery of the Facial Skeleton, S. Baker. Publisher: Elsevier (book chapter)
Chapter 3 Anthropometrics

- **Lara S. van de Lande**, Athanasios Papaioannou, David J. Dunaway

Rights from Elsevier for publication automatically granted under author permission.

Part of this work was presented at:

- *The 24th International Conference on Oral & Maxillofacial Surgery (ICOMS), 21-25 May 2019, Rio de Janeiro, Brazil.*
- *The 18th Congress of International Society of Craniofacial Surgery (ISCFS), 16-19 September 2019, Paris, France.*
- *The 25th Congress of the European Association for Cranio Maxillo Facial Surgery (EACFMS), 14-16 July 2021, Paris, France (Virtual Event).*

Chapter 3 described the construction of the Apert face 3DMM and proved its potential for automated identification on face shape information. Chapter 4 will apply this pipeline to genetically and phenotypically similar syndromes, i.e. FGFR-related craniosynostosis syndromes and include head shape information in addition to the face shape.

4.1. Introduction

The recent introduction of convolutional mesh autoencoder models (CMAs), a deep neural network approach to 3D model construction, offers further potential for the construction of shape-based models (4, 5, 14, 22, 23). This way of building morphable models is currently a very active area of research. These models learn to extract meaningful shape features from 3D meshes and can consequently be used for classification tasks. Neural networks are in use to solve a range of problems. For example, they can be used for object detection – feed the neural network an image and it will be able to identify locations of important objects in that image, language translation – feed a neural network with a sentence in English and it can deliver the equivalent in French, or audio classification – feed a neural network with a soundwave and it will determine the object that produces that sound, such as a dog barking. Other types of neural networks can generate information, which are called generative models, such as variational autoencoders. A better and more 'famous' example that revolutionize the discipline is generative adversarial network (GAN). A generative model learns to generate data from the same distribution of the training set. A generative model can be trained by large datasets and is able to generate unlimited outcomes based on the input data. Autoencoders play a role in this process. Autoencoders consists of two parts: the encoder network and the decoder network. The encoder takes an input sample and converts its information into some vector (a set of numbers), the decoder takes this vector and expands it out to reconstruct the input sample. The

reason for doing this – generating the same output as the input by using autoencoders – is not because of the output samples, but for the vector constructed in the process. This vector is a representation of the input data with less dimensions (for example a 3D mesh defined by 50,000 points, can be represented by a vector of 512 features) and but it can be fed to complex architectures to solve specific problems. An example of this is the interference of a location of a person based on that's person's social media post. In our case it might be usable for classifying the FGFR-related craniosynostosis. Among the more prominent examples of studies using deep convolutional neural networks is DeepGestalt, a facial identification tool based on tens of thousands of 2D images to identify facial phenotypes for genetic disorders (103) . Whilst such systems demonstrate impressive results, they are unable to take advantage of the rich geometric information in the face and cranium that may give critical insight into the phenotypical variations associated with different syndromes. Advances in 3D modelling and geometric deep learning have resulted in the introduction of a more shape-based approach craniofacial analysis (4, 5, 14, 22). Fernandez Abrevaya *et al.* 2018 was the first encoder-decoder architecture to model the 3D geometry of faces (48), projecting the 3D face to a 2D image. Ranjan *et al.* 2018 proposed the first autoencoder architecture for the geometry of faces that performs convolutions in a 3D mesh space directly instead of going through a 2D image representation (22). This work was in 2019 extended by Zhou *et al*. 2019 to encode both texture and shape information jointly (104).

A study by Hallgrimsson *et al.* 2020 attempted automated syndrome identification using parametric and machine learning approaches to phenotype the face; in this study, the authors included over 7,000 subjects (53% normal) and nearly 400 syndromes, including Apert (n≈25). They could distinguish syndromic from unaffected in 80% of their population using facial shape information. The overall classification rate to the correct syndrome was 71.8% and around 75% for Apert specifically (105). Hallgrimsson *et al.* 2020 also studied unaffected

relatives of syndromic cases and found that 23% of these relatives were classified as syndromic. After further investigation, most of those relatives turned out to be undiagnosed or had incomplete penetrance of the mutation.

This chapter reports on a geometric deep learning approach using convolutional autoencoders for the characterisation and identification of FGFR-related craniosynostosis syndromes, i.e. Apert, Muenke and Crouzon from the normal. As in most craniofacial syndromes not only the face, but also the head shape is affected, models including also the calvarium were investigated to identify the most effective diagnostic tool for FGFR classification.

4.2. Methodology

4.2.1. Data

Syndromic

In addition to the Apert patients that were included in the model from chapter 3, all patients diagnosed with Crouzon or Muenke syndrome at the *GOSH* were reviewed retrospectively for pre-operative CT imaging. All CT scans were exported and included after applying the inclusion and exclusion criteria (Table 4.1.). Additionally, baseline characteristics were collected from the medical charts.

Normative

LSFM was used as a healthy control group for the subpopulation between the ages of 4 and 17 years old. Of the subjects who met the desired age criteria, 196 samples were randomly

selected to provide this age matched reference (7 males, 7 females for each age). The <4 group described in chapter 3 was used for the normal <4 group.

In addition to the face 3DMM constructed in chapter 3, the model described in this chapter also included the cranium. For the <4 and syndromic datasets, several samples used for the face models were omitted when constructing the head models as not all scans used for the face model captured the full cranium. Scans with incomplete cranium or those with previous calvarial surgery, such as posterior vault remodelling/expansion, were excluded.

For additional head meshes that could not be provided from the LSFM dataset, data from the LYHM was used (4). This database consists of approximately 1,200 individuals between the age of 2 and 90. From LYHM, scans from those aged 4 to 17 were extracted to provide an age-matched reference for the full head scans from the syndromic dataset. See Table 4.1. for inclusion and exclusion criteria for all datasets used.

Table 4.1. Inclusion and exclusion criteria for data sources. Three different models were constructed: only face, only head (cranium), and face+head. As control data, external data sources were consulted. For the face model LSFM was used to generate meshes, for the head model LYHM was consulted to obtain age-matched normal controls.

	Inclusion	Exclusion
Face		
CT head scans 2005-2018	• Apert diagnosis, no age limits • Crouzon diagnosis, no age limits • Muenke diagnosis, no age limits	• Incomplete scans (full face required) • Movement artefacts or squished faces due to gel pads used for scanning. • Insufficient CT slices for 3D construction • Scans after any type of craniomaxillofacial surgery

	• Normal scans aged <4 years.	• CT-visible abnormalities • Craniofacial diagnosis/ abnormalities • Incomplete scans (full face required) • Movement artefacts or squished faces due to gel pads used for scanning. • Insufficient CT-slices for 3D construction
LSFM	• 3D meshes aged 4–17 years, randomly selected 7 male and 7 female per year-group.	
Cranium		
CT head scans (2005-2018)	• Apert diagnosis, no age limits • Crouzon diagnosis, no age limits • Muenke diagnosis, no age limits	• Incomplete scans (full face required) • Movement artefacts or squished faces due to gel pads used for scanning. • Insufficient CT slices for 3D construction • Scans after any type of craniomaxillofacial surgery
	• Normal scans aged <4 years.	• CT-visible abnormalities • Craniofacial diagnosis/ abnormalities • Incomplete scans (full cranium required) • Movement artefacts or squished faces due to gel pads used using CT scan. • Insufficient CT-slices for 3D construction
LYHM	• Head 3D meshes aged 4-17	

4.2.2. Image pre-processing

CT scan pre-processing and registration

CT datasets underwent the pipeline described in chapter 3.2. for model construction. Three types of mesh templates were used: a) face, b) head, and c) face+head. (Figure 4.1.) The cropped face template from (60) was used for construction of the face models as done in the previous chapter. For the head model registration, a template from (106) with 55 additional landmarks were used to encourage good correspondence around the ears. For the face+head models the template from (15) was used.

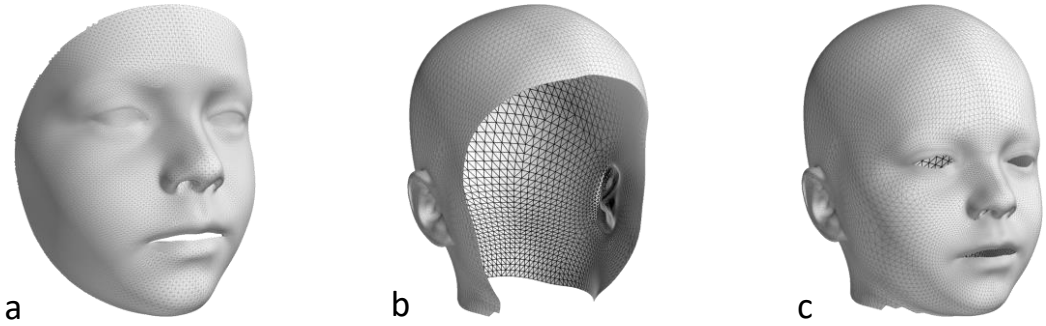


Figure 4.1.: Mesh templates. These are the templates used to guide the correspondence process. **a)** face template, **b)** head template, and **c)** face+head template.

Dense correspondence for the Syndromic dataset

Dense correspondence was achieved as described in chapter 3.2.2.

4.2.3. 3D Mesh Autoencoder Construction

Once dense correspondence was achieved for all meshes (i.e. Crouzon, Muenke, Apert, and normal) (100%), the 3D models were created using mesh autoencoders. Nine models in total were constructed, – face, head, face+head - to assess the role of facial and cranial shape in the diagnosis of FGFR-related craniosynostosis:

- <4: a model trained using all samples up to and including the age of three.

- 4-17: a model trained using all normal samples between 4-17 years of age.
- All ages: a model constructed from all available data.

The autoencoder architecture applied here is similar to that described in (22) and for in-depth explanation of this methodology I would like to refer to (107). In simplified words, 4 convolutional and down-sampling/up-sampling layers were used for the encoder and decoder for construction of the head and face+head models. Five such layers were used for the face model as the template mesh for this model is far denser (28,431 vertices for the face model, compared to 7,505 vertices in the head model, and 17,039 vertices in the face+head). Encoder convolutional filter sizes of [16, 16, 32, 32, 32] were used for the face model, while encoder filter sizes of [16, 16, 32, 32] were used for the head and face+head models. These filter sizes were chosen based on the method used in (22). In both cases, decoder filter sizes were the mirror of the encoder. Each convolutional layer in the encoder was followed by a mesh down sampling layer by a factor of 4 - in the decoder, this was replaced with a layer to up sample the mesh by a factor of 4. An additional convolutional layer was added to the decoder to allow for the reconstruction of the 3D shape coordinates. An ELU (Exponential Linear Unit) activation function was applied after each convolutional layer, which is an activation function for neural networks. Model weights were initialised using Xavier initialisation and Adam optimisation was used. All models were trained for 300 epochs using a batch size of 16, similar to the approach of (22).

4.2.4. Intrinsic model evaluation

The model was validated using the intrinsic model characteristic validation methodology evaluation compactness, specificity, and generalisation as described in chapter 3.2.3 (41, 43, 102).

4.2.5. Manifold Visualisation

t-SNE was applied to the high dimension latent vector encodings for the syndromic patients all ages and normal all ages, normal 4-17 and normal <4 normal meshes, to assess the diagnostic capacity of the models. Samples were labelled according to their syndromic class (normal, Apert, Crouzon, or Muenke) with the aim of uncovering distinct groupings, or clusters. In addition, a t-SNE embedding was applied to identify two subtypes of Apert. Samples were labelled with subtype: Ser252Trp, Pro253Arg, subtype unknown, and Crouzon (which was used as a control syndromic group). In addition, the age at time of scan was assessed to determine whether age played a role in the clustering behaviour of the two genetic subtypes. All t-SNE embeddings were created using a perplexity of 30 and run for 1,000 iterations.

4.2.6. Classification

Autoencoders are often utilised for their ability to compress data into a much more compact format. This manifests as the latent vectors of the model. These latent vectors provide a natural means by which we can attempt to classify the data and determine its applicability as a diagnostic tool. Classification was performed using a Support Vector Machine (SVM) with linear kernel and balanced class weighting. A stratified data split with an 80%:20% train:test

proportion was used. The scikit-learn SVM implementation with default gamma and regularization parameters (C=1.0) was employed. The mean accuracy, specificity, and sensitivity were calculated following a Monte-Carlo cross-validation system where the training and test sets were randomly selected 10,000 times.

4.3. Results

A summary of the demographics for the subjects used in this study is provided in Table 4.2. The models include data from four databases:

1. CT data from 122 FGFR-affected patients (47 Apert, 61 Crouzon, and 14 Muenke), at a mean age of 5.0 ± 5.1 years
2. CT data from 142 healthy infants at a mean age of 1.9 ± 1.2 years.
3. Stereophotogrammetric face data from 196 healthy subjects from the LSFM dataset with a mean age of 10.5 ± 4.0 years (60).
4. Stereophotogrammetric head data from 139 healthy subjects from the LYHM database at a mean age of 10.9 ± 3.8 years (4).

Table 4.2. Overview of the face and cranium dataset of the included Syndromic Craniosynostosis and normal samples. All syndromic and infant samples were acquired via CT scan. The LSFM and LYHM databases were obtained using 3dMD™ photometric stereo capture device set-ups (4, 5)

Type of SC	Number of subjects	Age at scan, years	Age range at scan, years	Male (%)
Face				
Normal 4-17	196	10.5 ± 4.0	4 – 17	98 (50%)
Normal <4	142	1.9 ± 1.2	0 – 3.9	79 (56%)

Apert	47	6.1 ± 6.2	0.1 – 20	28 (60%)
Crouzon	61	5.3 ± 4.4	0.08 – 17	35 (58%)
Muenke	14	1.6 ± 2.1	0 – 8	7 (50%)
Total	**460**		**0 – 20**	**247 (54%)**
Head				
Normal 4-17	139	10.9 ± 3.8	4 – 18	76 (55%)
Normal <4	111	1.8 ± 1.1	0 – 3.9	59 (53%)
Apert	39	6.5 ± 6.3	0.1 – 20	22 (56%)
Crouzon	53	5.4 ± 4.4	0.4 – 17	30 (57%)
Muenke	11	1.7 ± 2.3	0 – 8	6 (55%)
Total	**353**		**0– 20**	**193 (55%)**

4.3.1. Intrinsic Model Evaluation

All ages models

Using the available databases, all 9 models were built, with low reconstruction errors in terms of accuracy and specificity. For the face, head, and face+head models, the error values were 1.4 ± 1.2 mm, 3.8 ± 3.1 mm, and 2.9 ± 2.5 mm, respectively. Reconstruction error was higher for models that included the head shape, likely due to the greater degree of variation between subjects in this region. Model specificity assessment returned values of 2.7 mm, 4.3 mm, and 3.9 mm for the face, head, and face+head models, respectively, indicating that the samples generated are realistic.

<4 models

Low reconstruction errors were recorded for all models. For the face, head, and face+head models, these values were 1.9 ± 1.6 mm, 4.3 ± 3.4 mm and 3.0 ± 2.6 mm, respectively. Model specificity values of 2.9 mm for the face, 5.2 mm for the head, and 4.2 mm for face+head models. These values are higher than those recorded for the all ages model; this is expected, given the reduced number of samples used for model construction.

≥4 models

The reconstruction error for the face (1.3 ± 1.2 mm), head (3.0 ± 2.3 mm), and face+head models (2.3 ± 1.9 mm) were notably lower than the reconstruction errors observed for the complete model even though a reduced number of samples was used for model construction. This might be due to the reduced shape variations within the older individual dataset. Specificity values of 1.8 mm for the face, 3.6 mm for the head, and 3.4 mm for face+head models are again lower than values for the all ages models.

4.3.2. Manifold Visualisation

For the face model, clear clusters emerge between the normal and syndromic groups (Apert, Crouzon, and Muenke). For the normal groups, the samples from <4 and 4-17 are clustering separately, which can be caused due to age and/or due to different imaging modalities, <4 derives from CT scans and 4-17 from 3D photography. Within the syndromic cluster, the different syndromes are further subclustered. The clusters formed for the head embeddings are not as distinct as those observed for the face cases; however, groups for Apert, Crouzon, Muenke, and normal individuals do emerge. When considering the model constructed

using the face+head template, again clear clusters can be seen forming for the different subgroups in the dataset.

In all cases, even though the syndromic samples tend to group more tightly, the syndromes themselves seem relatively disentangled. The proximity of the Crouzon cluster to the normal cases in each of the embeddings indicates that this phenotype has a milder manifestation than either Apert or Muenke syndrome. For both the head and face+head embeddings, it can be noted that several Muenke and Crouzon samples cluster closer to the group of healthy cases.

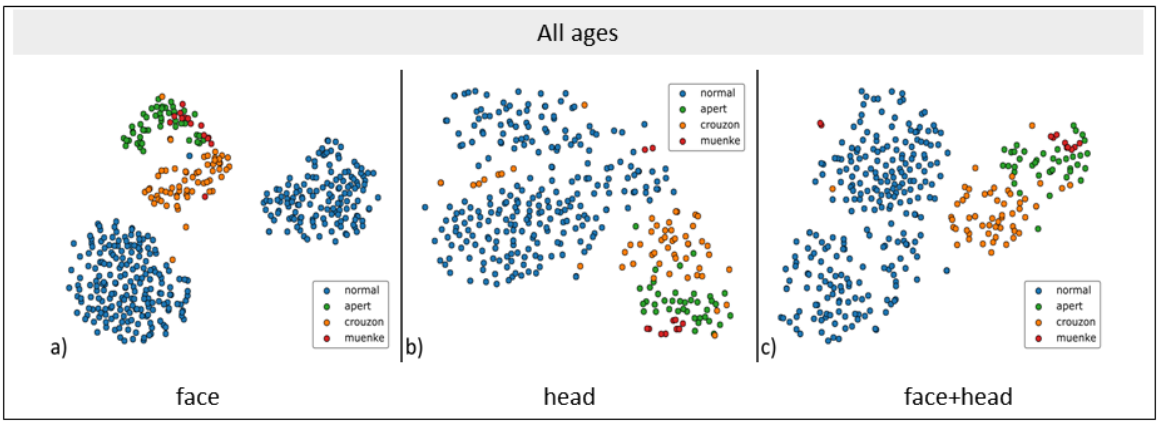


Figure 4.2. t-SNE embedding of the high-dimensional manifold clustering for the complete model. The t-SNE embedding in two dimensions was performed for a) face, b) head, and c) face+head models. Distinct clustering is seen for all models, with the least performance for the head only model.

For the <4, distinct groups are formed for the syndromic and the normal population for each model. In all cases, the Crouzon patients cluster closest to the normal population, indicating that the physical manifestation of Crouzon syndrome is milder than that of either Apert or Muenke syndrome.

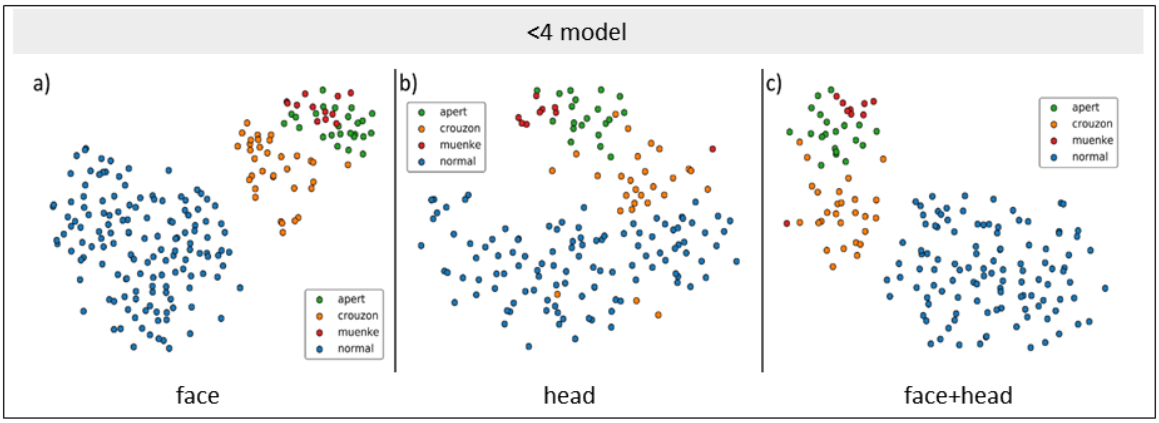


Figure 4.3. t-SNE embedding of the high-dimensional manifold clustering for the under-4 model. The t-SNE embedding in two dimensions was performed for a) face, b) head, and c) face+head models. Distinct clustering is seen for all models, with the least performance for the head only model.

Also for the >4 model, again clusters are formed for each of the included groups (Figure 4.4.). This is clearest for the face model; however, syndromic samples do still cluster separately from the unaffected samples in both the head and face+head models. Though there are few Muenke samples in this cohort (n=2), these are still separated from the normal population.

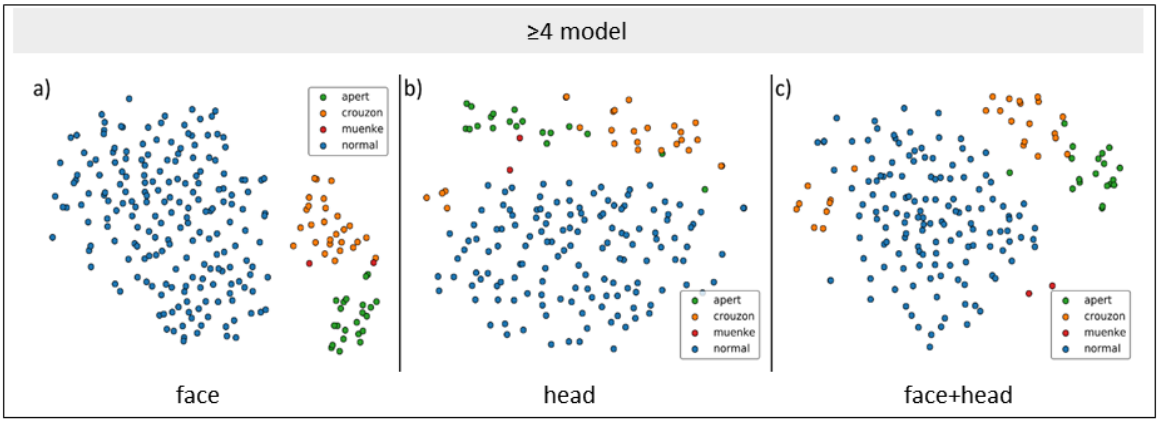
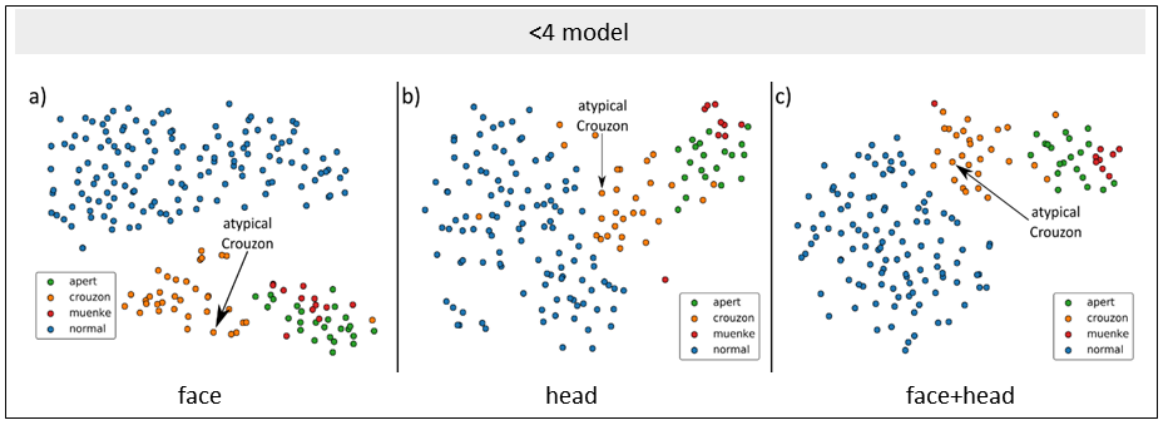


Figure 4.4. t-SNE embedding of the high-dimensional manifold clustering for the ≥4 model. The t-SNE embedding in two dimensions was performed for a) face, b) head, and c) face+head models. Distinct clustering is seen for all models, with the least performance for the head only model.

This method may outperform clinical expert assessment as proven by a case in this chapter (Figure 4.5): a 2-year-old male sibling of a male Crouzon patient was considered unaffected by the clinical team during a consultation when the relative accompanied the patient to clinic. However, the brother proved to have Crouzon syndrome as well, which was confirmed on genetic testing that was carried out as part of routine genetic testing for relatives from affected patients. For the sibling, a CT scan was performed, and this case was used as a sample to test whether the algorithm could pick up this seemingly unaffected case. The model was able to identify the sibling (labelled as atypical Crouzon in Figure 4.5) as part of the Crouzon cluster in the t-SNE plot. The atypical Crouzon was clustered as Crouzon in all three models, suggestion that model can identify clinically undetectable variations from the norm.

Figure 4.5. t-SNE embedding of the high-dimensional manifold clustering for atypical Crouzon case. The t-SNE embedding in two dimensions was performed for a) face, b) head, and c) face+head models. Correct diagnostic clustering for this atypical and clinically undetected patient can be noted.



4.3.4. Identification of Apert's genetical subtypes

No distinct grouping is seen for the Apert genetic subtypes. Across the Apert groups, age might play a larger role than the genetical mutation in terms of shape clustering (Figure 4.6.). The ages at time of CT scan are labelled and suggests being an influencing factor for the clustering behaviour. The lower left cluster in Figure 4.6. is dominated by cases around the age

of 2-18 months with the majority under the age of 57 month. Two exceptions can be seen: one case of 214 months (Pro253Arg) and one case of 115 months (Ser252Trp). The top-center cluster in the t-SNE embedding is dominated by cases between 76 and 240 months of age (6-18 years).

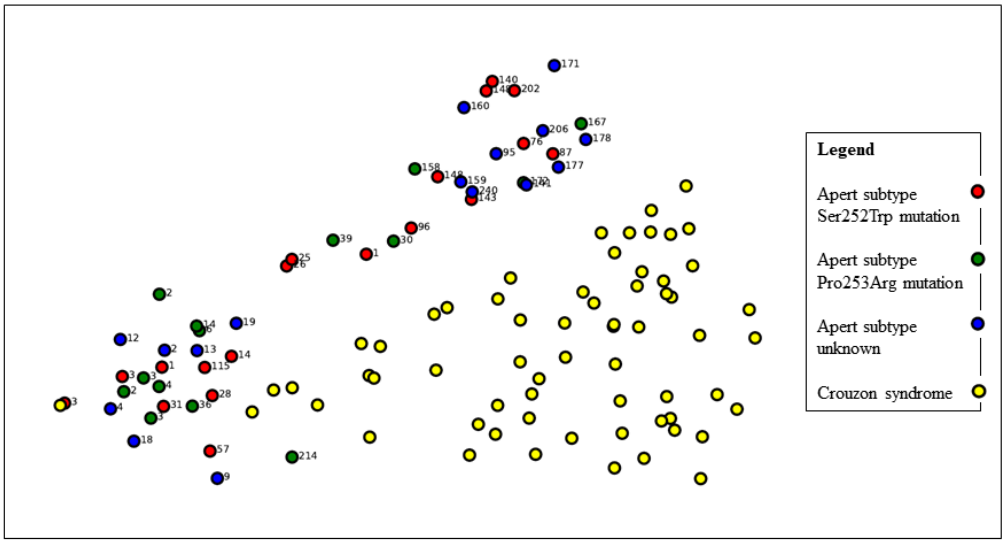


Figure 4.6. t-SNE embedding of the high-dimensional manifold clustering of Apert subtype. The t-SNE embedding in two dimensions was performed for the face-only models and demonstrate clear clustering between Crouzon and Apert. No distinct groups are identified for the subtypes. The labelled numbers correlate with the age in months at time of scan. Grouping seems more driven by age than genetical subtype.

4.3.3. Classification

All ages models

Classification was performed with all syndromic and unaffected scans. The mean sensitivity, specificity, and accuracy over all iterations for each of the assessed regions in the binary classification experiment for the complete model is presented in Table 4.3. The highest accuracy is seen for the face model (99.98%) with a specificity of 100%. The lowest accuracy of the three models (99.09%) was recorded for the head model.

Table 4.3. Classification results for the binary classification experiments for the all ages model. Model sensitivity, specificity and accuracy was calculated for all models. The face model outperformed the head model and the face+head model.

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)
Face	99.95	100.00	99.98
Head	98.36	99.41	99.09
Face+head	99.82	100.00	99.95

Accuracies of >99% was seen for the assessment whether an individual belonged to either the syndromic group or normal group. The high sensitivity of the models indicates that very few syndromic cases were misidentified as normal (1:1000 for the face model, 16:1000 for the head model, and 2:1000 for the face+head models). The inverse is also true; few healthy cases were seldom, if ever, misidentified as having a craniofacial syndrome as indicated by the high model specificity values (Figure 4.5.).

The multi-class classification model endeavoured to predict whether a patient belonged to either the non-syndromic, Apert, Crouzon, or Muenke categories. Accuracies of 98.3% for face, 97.9% for head, and 98.2% for face+head model were observed. When considering the face model, Muenke patients were the most likely to be misdiagnosed; they also have the fewest instances of this syndrome in the database. When the head shape was considered for classification, Crouzon and Apert patients were most likely to be misidentified as each other. As with the binary classification, the poorest performance was seen for the head model. These results would indicate that the facial region contains valuable shape information for the correct diagnosis of FGFR-related craniosynostosis (Figure 4.7.).

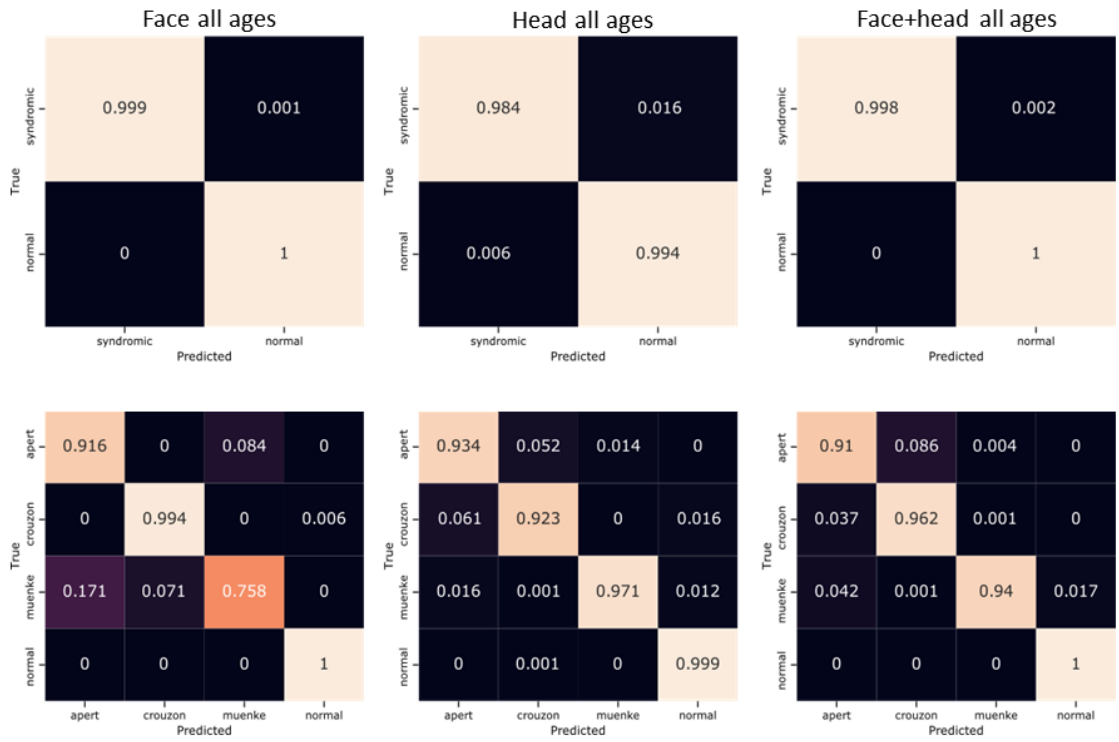


Figure 4.7. Confusion matrices for the all ages models. Confusion matrices were performed for the face, head, and face+head models. **Top row** shows the confusion matrices for the binary classification. **Bottom row** shows the multi-class classification.

<4 models

For the <4 model, accuracies for binary syndrome classification yields >99% in all cases (Table 4.4.). As with the all ages models, the high model sensitivity indicates that few syndromic cases are misidentified as normal. The high specificity values demonstrate that the normal individuals are also highly unlikely to be misidentified as a syndromic patient.

Multi-class classification accuracies of 97.8% (face model), 97.1% (head model), and 97.7% (face+head model) are observed. Confusion matrices for the binary and multi-class classification for all <4 models are shown in Figure 4.8.

Table 4.4. Classification results for the binary classification experiments for all <4 models. Model sensitivity, specificity and accuracy was calculated for the <4 models. The face+head model outperformed the head model and the face model.

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)
Face	99.71	100.00	99.90
Head	99.53	99.99	99.84
Face+head	99.95	100.00	99.98

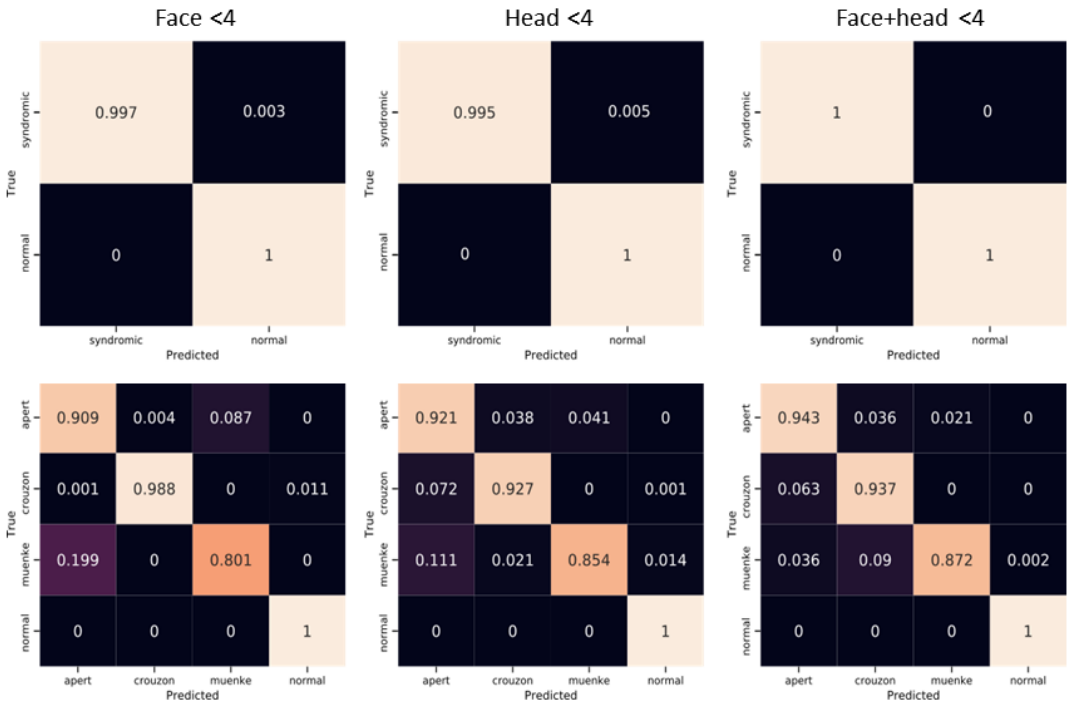


Figure 4.8. Confusion matrices for the <4 models. Confusion matrices were performed for the face, head, and face+head models. **Top row** shows the confusion matrices for the binary classification. **Bottom row** shows the multi-class classification.

≥4 models

As with the all ages, and <4 models, binary classification accuracies for the ≥4 model are greater than 99%. Again, high sensitivity and specificity values are recorded (Table 4.5.).

Multi-classification experiments demonstrate accuracies of 99.9%, 95.8%, and 96.2%. From these results, it is again observed that models that incorporate the face yield higher

classification accuracies, further indicating the importance of the facial region for the classification of FGFR-related craniosynostosis syndromes. Confusion matrices and multi-class classification matrices for the ≥4 models are shown in Figure 4.9.

Table 4.5. Classification results for the binary classification experiments for all ≥4 models. Model sensitivity, specificity and accuracy was calculated for the <4 models. The face and the head models demonstrated 100% for all classification experiments.

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)
Face	100.00	100.00	100.00
Head	100.00	100.00	100.00
Face+head	99.15	100.00	99.77

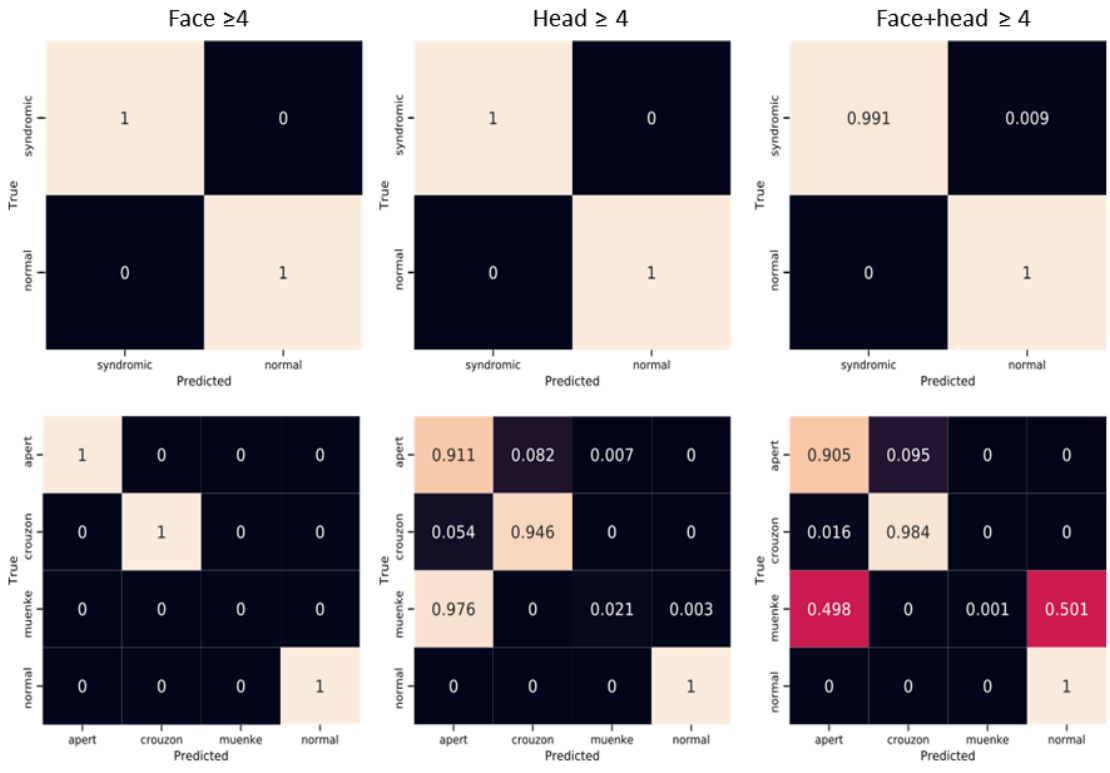


Figure 4.9. Confusion matrices for the ≥4 models. Confusion matrices were performed for the face, head, and face+head models. **Top row** shows the confusion matrices for the binary classification. **Bottom row** shows the multi-class classification.

4.4. Discussion

This chapter demonstrates the use of AI for FGFR-related syndrome identification with a state-of-the-art geometric deep learning approach. Machine diagnosis showed an accuracy of 99.98%, sensitivity of 99.95%, and specificity of 100%. The diagnostic precision of this technique supports its potential inclusion in clinical decision support systems and indeed the algorithm outperformed expert clinical diagnosis in one example. Needing only a 3D input allows potentially deceptive texture information to be omitted, enabling the architecture to focus on the extraction of characteristic shape-based features to return an accurate diagnosis. Although, many more syndromes exist, these are encouraging results in a field where clinical diagnosis requires expert knowledge and a time when automated diagnosis is set to play an increasing role in healthcare.

Utilising 3D topography in our models rather than the surface texture analysis, which is central to most other facial analysis techniques, lends itself to a unique integration with many forms of conventional diagnostic 3D imaging technologies, such as CT, MRI and ultrasound scans. The presented technique facilitates automatic segmentation of surfaces as well as syndrome identification and could therefore be used as a machine learning diagnostic tool to aid radiological diagnosis. Integration with ultrasound imaging is of particular interest, as this presents an opportunity for foetal detection of genetic disorders (108). The rising availability of 3D scanning applications and cameras on mobile devices presents further possibilities to introduce such a framework in primary and secondary care. In a discipline where timely diagnosis is necessary for appropriate management, the use of such technologies to detect syndromic children and other conditions will streamline assessment and could be pivotal to improving long-term health outcomes.

The model presented can be applied for binary classification (syndromic vs. healthy) and multi-class classification (Apert, Crouzon, Muenke, and healthy). The high sensitivity and specificity of the model make it suitable as a diagnostic aid in primary and secondary care settings ensuring reliable diagnosis with few false positive results. Muenke was most misdiagnosed, but also had the least number of patients included in their dataset (n=14). Increasing the study population size will most likely lead to increased accuracy for this patient group. The algorithm suggests outperformance over the expert's human eye, which was demonstrated with an example case of an atypical Crouzon patient. However, the algorithm could not distinguish the Apert subtypes. One reason for this might be that the sample size was too small (Ser252Trp mutation, n = 10 and Pro253Arg, n = 13). Further, age seems to have a more dominant effect on the shape differences than genetics. Given these results, we conclude that there are no facial differences between the 2 subtypes, or, if there are, the facial differences are too subtle and the sample sizes too small for the algorithm to pick them up.

A finding of note is that the face region seems to be more diagnostically relevant than the head shape for automated identification of FGFR-related craniosynostosis. While it might be suspected that the shape of the head contains the most valuable diagnostic information for children with craniosynostosis, and indeed this may be the case with non-syndromic craniosynostosis, for those subjects included in the analysis, much of the information required for accurate diagnosis appears to manifest in the facial region. One suspected cause for this is that Apert, Crouzon, and Muenke syndromes can be accompanied by the closure of similar sutures, and might, therefore, result in similar shaped heads, whilst the genetic differences between these syndromes may be more evident in the facial features.

This framework was tested and validated for FGFR-related craniosynostosis syndromes, including Apert, Crouzon, and Muenke syndrome. The excellent classification results are

encouraging for further research into extending such models to craniofacial syndromes affected by other genes. Syndromes that also affect the facial shape could be included, such as craniofacial microsomia or Treacher Collins syndrome, or even to a wider population with syndromes that have very subtle facial deformities, as attempted in Hallgrimsson *et al.* 2020 (105). At present, an easy-to-use toolbox is not yet available – nor in the assessments performed in this study, and neither in similar published studies (105). A future platform should consist of a more-user friendly open-source interface, removing the need for any programming expertise. Indeed, Face2Gene is one of the forerunners in this. Face2Gene is an application developed by FDNA and employs a 'deep gestalt' framework to extract features from 2D images and compares them to pre-learned 2D facial 'gestalts' constructed from 2D images of patients with known genetic disorders (109). Based on similarity to these gestalts, it prioritises candidate diseases for the individual. In contrast, the 3D shape analysis presented in this chapter are intended to facilitate diagnosis and outperform 2D analysis. Mattews *et al.* 2021 very recently presented 3D growth curves; the authors did not aim at automated diagnosis, yet their work might be of use for future purposes by extracting confounding variation due to age and gender from training images and the new subject images being assessed (110).

A limitation of this study is the lack of comparison to other methods that are currently available, either neural or non-neural paradigms. Other deep learning approaches such as GAN for 3D face modelling (56) could provide perhaps even better outcomes but are not described in this thesis. Further, the statistics of 3DMMs are limited to the face and do not include information on the eyes, intra-oral cavity, or hair. Combining various parts in one model is attempted by Ploumpis *et al.* 2019, however, turns out not to be straightforward (15). A strong limitation of this study is the lack of information on racial background of the included individuals. From other models, it is known that a strong racial bias towards white exists. This could be ameliorated by using data from across the world, or by allowing to generate and add

synthetic data to biased datasets. Another existing open challenge is to learn from inhomogeneous data, many datasets have different quality and resolutions, coverage, noise characteristics, and so on. In the studies presented in this thesis, the included data were limited to scans with >100 slides and <1 mm, however, it would be beneficial if also data with less quality could be included, considering the rarity of the study population. Lui *et al.* 2019 attempts to address this problem by constructing models based on diverse raw scan data (111). Finally, the 2 groups of clustering noted between <4 and 4-17 may be influenced by the different data sources used – CT and surface scanning, respectively, however this has not been tested specifically.

Future work is described in chapter 8 as the proposed models can be used for post-surgical assessment.

4.4. Summary

In summary, an autoencoder framework was presented and demonstrated its diagnosis potential to assist in the identification of FGFR-related syndromic craniosynostosis using supervised machine learning. This framework can be applied to give a binary output indicating whether the presence of a syndromic craniosynostosis is likely and can be further applied for the identification of specific genetic mutations. Although the focus of this chapter was on three distinct mutations in the FGFR gene, namely Apert, Crouzon, and Muenke syndrome, given sufficient data, this framework could be readily extended to a greater variety of craniofacial syndromes. Extending the model to larger patient cohorts and a greater number of syndromes may lead to new diagnostic tools, facilitating low-cost analysis and identification of craniofacial disorders, and the subsequent management of such conditions, at an earlier stage.

Future applications could include diagnostics of post-operative outcome, surgical planning, and surgical outcome prediction.

Key points:

- An autoencoder framework might have a valuable future role in automated syndrome diagnosis.
- These models relay on shape information only and could therefore be easily integrated with other types of imaging modalities, such as MRI, CT scan and ultrasound.
- Facial abnormalities play a larger role than head shape information for automated identification.
- Key differences in Apert face are suggested to be driven by age over genetical subtype mutation.

5. A SKULL MORPHABLE MODEL – based on a healthy paediatric population

Part of the work described in this chapter has been published in:

Journal of Bone Reports (published)
A Statistical Model of the Paediatric Skull

- *Eimear O' Sullivan†, **Lara S. van de Lande†**, Anne-Jet C. Oosting, Athanasios Papaioannou, N. Owase Jeelani, Maarten J. Koudstaal, Silvia Schievano, Roman H. Khonsari, David J. Dunaway, Stefanos Zafeiriou*
  † These authors contributed equally to this work

The Journal of Craniofacial Surgery (abstract)
A 'normal' paediatric skull 3d morphable model

- Eimear O' Sullivan, **Lara Sophie van de Lande**, Anne-Jet C. Oosting, Athanasios Papaioannou, Maarten J. Koudstaal, Silvia Schievano, Roman Khonsari, David J. Dunaway.

The Journal of Craniofacial Surgery (abstract)
A paediatric soft tissue 3d morphable model

- **Lara Sophie van de Lande**, Eimear O' Sullivan, Athanasios Papaioannou, Silvia Schievano, Stefanos Zafeiriou, Roman Hossein, David J. Dunaway. Abstract publication.

Part of this work was presented at:

- *The 25th Congress of the European Association for Cranio Maxillo Facial Surgery (EACFMS), 14-16 July 2021, Paris, France (Virtual Event).*

In this chapter, the construction and validation of a normal paediatric skull model is presented. The soft tissue models, described in chapter 3, gave insight into the facial shape of various ages, and provided a comparison of the mean Apert's face with an unaffected face. Yet, the 3D abnormalities of the architecture underneath the soft tissue are rather unknown. To further investigate the Apert's face statistically, chapter 5 aims to first construct a normal skull model and validate this with published normative 2D and 3D measurements. With this validated normal model, an Apert skull model is constructed, described in chapter 6. The term 'skull' used in chapter 5 and 6 are defined as the skeletal structures of the midface and calvarium without the mandible.

5.1. Introduction

Accurate modelling of the skull has grown to play an important role in surgical planning for craniomaxillofacial surgery, craniofacial diagnostics, surgical outcome analysis, and provides invaluable insight into growth and developmental patterns (21, 112-114). However, 3D models of the paediatric normal skull are scarce and most normative datasets derive from 2D information and measurements (3). Waitzman *et al*. 1992 and Delye *et al*. 2015 studied craniofacial skeletal measurements derived from CT scan data to provide relatively large datasets of normative data for a paediatric population (3, 115). These measurements include cephalic index, cephalic length, cephalic width, oblique cranial length ratio, lateral orbital distance, anterior orbital distance, intertemporal distance, inter-zygomatic buttress distance, zygomatic arch length, and inter-zygomatic arch distance. However, at present, no statistical model exists for the skull of the human child providing detailed 3D shape information that has generative capability. The human skull displays major structural and geometrical changes

during childhood, most prominently in the first years of life, and therefore 2D measurements and growth charts do not suffice to understand its natural development accurately (116).

3D modelling of the craniofacial skeleton presents several challenges as the skull is a complex structure comprising many different facial and cranial bones. The cranial bones are loosely connected at birth and these expand and fuse as the infant grows, further complicating the modelling process. 3D models of the human adult skull exist and have commonly been applied for the reconstruction of the facial shape given only the skull (116). Lüthi *et al.* 2008 have applied skull models to assist in the segmentation of the human skull from MRI images (117).

A statistical model of the human paediatric skull could be used to understand normal shape change in childhood and provide normative measurements; it can tailor the pre-operative (in silico) surgical plan towards patient-specific normative target values and can help in the design of osteotomies and cranioplasties, particularly when mirroring of the skull is not possible (118, 119). Studying growth patterns in children is important due to the timing and rate of growth of different parts of the skull. The way this leads to shape change, might have implications for the timing of surgery. To be valid and effective, in silico trials still relay on cohorts of real, clinical data (120). However, collecting and annotating large quantities of data to train and test the computational models is expensive and time-consuming, thus restricting the choice of evaluation and modelling algorithms that can be applied. Ethical considerations further limit the purposes for which datasets can be used. Conversely, generated synthetic data are easy to manipulate, fully user controlled and carry limited ethical concerns. The ability to generate new samples means that cohort size is not a limiting factor, overcoming statistical power issues and the burden of enrolling large cohorts of real patients (121). As such model, could also be used to compare post-surgical results with a normal age-matched skull and

understand where the current techniques is successful and where normalisation is lacking (122, 123).

This chapter describes the construction and validation of a normative skull model for children <4 years of age and it provides a paediatric skull model that can be used for many applications, and in this thesis, for comparison with Apert syndrome.

5.2. Material and methodology

5.2.1. Data

The <4 dataset of chapter 3 was used for the construction of the healthy skull model (see methods chapter 3.2). All CT scans that met the inclusion and exclusion criteria listed in Table 5.1 were included.

Table 5.1. Eligibility criteria for CT-data. The <4 dataset was used for model construction; this table presents and overview of inclusion and exclusion criteria.

	Inclusion	Exclusion
CT head scans 2011-2018	Normal scans aged 0–4 years. >150 slices and/or ≤1mm.	• CT-visible abnormalities • Craniofacial diagnosis/ abnormalities • Incomplete scans (full skull required) • Artefacts due to movement or metal (e.g. braces) • Insufficient CT-slices for 3D construction

5.2.2. Image processing

All CT scans underwent the following workup to convert the DICOM files to skull meshes usable for model construction (Figure 5.1.):

1. DICOM files were imported in Mimics Inprint 3.0 for segmentation, carried out by thresholding with default bone-setting grey scale values (I).
2. The skull was isolated from other structures captured by CT scanning, i.e. mandible, vertebrae, and extra-cranial objects such as the back of the CT scanner, using a foreground and background tool (II), leaving the mesh as one continuous object (III).
3. The segmented skull was converted to 3D and saved as Obj. file. (IV).

Figure 5.1. 3D mesh construction. This is an overview providing the segmentation process. After the DICOM files are imported the first step consists of thresholding using bone default setting (**I**), then the mandible is omitted using foreground and background tools (**II**), then the mesh is isolated (**III),** and finally converted to a 3D object (**IV)** and saved as Obj. file.

5.2.3. Skull template

For the 3D craniofacial skeletal model construction, all skull meshes needed to be in dense correspondence; in a similar process to the soft tissue model construction described in chapter 3.2. A normalised mesh template of fixed topology was used to put all skulls in dense correspondence using NICP (58) with a set of 29 anatomical landmarks. (Figure 5.2. and Table 5.2.) A literature review was conducted using PubMed/MEDLINE database to review articles describing validated landmarks with little error. A search strategy was performed using Medical Subject Headings (MeSH) terms on the 19th of November 2019. See Appendix B.1. for the search term used and Appendix B.2. for the included articles with their published inter- and intraclass correlations for each landmark.

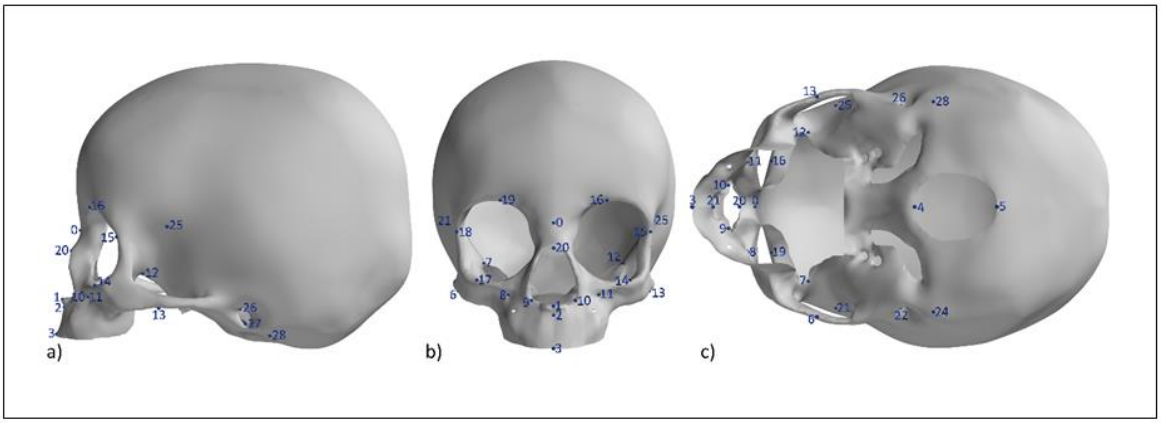


Figure 5.2. Annotation template for model construction. A total of 29 landmarks were used to guide dense correspondence. The landmarks are shown from **a)** lateral view, **b)** frontal view and **c)** cranial view.

Table 5.2. Landmark definitions. These landmarks were used to guide the correspondence process. **R** indicates the right-hand side of the skull, whereas **L** indicates the left-hand side.

Landmark number	Anatomical reference	Definition
0	Nasion	Midpoint of the nasal frontal suture
1	Anterior nasal spine (ANS)	Most anterior mid-point of the pointed projection formed by the protrusion of the maxilla at the nose base
2	A point	Point of the maximum concavity anteriorly of the maxillary alveolar process in the midline
3	Prosthion	Most anterior point of the maxillary alveolar process in the midline
4	Basion	Most inferior point of the midpoint of the anterior curvature of the foramen magnum
5	Opisthion	Most inferior point of the midpoint of the posterior curvature of the foramen magnum
6	Zygomatic Arch R	Most latero-inferior point on the zygomaticotemporal suture R
7	Jugale R	Intersection of the zygoma, maxilla and sphenoid R
8	Foramen infra-orbitale R	Midpoint on the superior margin of the infraorbital foramen R
9	Nasal cavity R	Latero-inferior point in the curvature of the nasal cavity R
10	Nasal cavity L	Latero- inferior point in the curvature of the nasal cavity L
11	Foramen infra-orbitale L	Midpoint of the superior margin of the infraorbital foramen L
12	Jugale L	Intersection of the zygoma, maxilla and sphenoid L
13	Zygomatic Arch L	Most latero-inferior point on the zygomaticotemporal suture L
14	Orbitale L	Most antero-inferior point on the inferior orbital margin L
15	Frontozygomatic suture L	Mid-anterior point on the lateral curve of the orbit on the intersection of the zygoma and frontal bone L
16	Supra-orbitale L	Anterolateral point of the supra-orbital notch of the supra-orbital rim L
17	Orbitale R	Most anteroinferior point on the inferior orbital margin R
18	Frontozygomatic suture R	Mid-anterior point on the lateral curve of the orbit on the intersection of the zygoma and frontal bone R
19	Supraorbitale R	Antero-lateral point of the supraorbital notch of the superior orbital rim R
20	Rhinion	Most antero-inferior point of the nasal bone
21	Pterion R	Most antero-inferior point of the parietal bone on the intersection of the frontal, parietal, and sphenoid bones R
22	External auditory meatus superior R	Most latero-superior point of the external auditory canal R

23	External auditory meatus inferior R	Most latero-inferior point of the external auditory canal R
24	Mastoid process R	Most inferior point of the mastoid process R
25	Pterion L	Most antero-inferior point of the parietal bone on the intersection of the frontal, parietal, and sphenoid bones R
26	External auditory meatus superior L	Most latero-superior point of the external auditory canal L
27	External auditory meatus inferior L	Most latero-inferior point of the external auditory canal L
28	Mastoid process L	Most inferior point of the mastoid process L

The dataset included many samples with an open fontanelle and varying levels of suture closure, as expected at this age. To prevent problematic registrations for these samples, a "stiffness" factor was added to the template mesh in the region of the opening to facilitate the registration process and prevent inwards collapse of the template mesh in this region. This prevented inwards collapse of the template mesh in this region.

5.2.4. Skull model validation

The 3D mean anatomical skull shape was computed based on all the densely registered meshes. PCA was applied to build the 3DMM and detect key contributors to 3D shape variability in the population. The first 5 principal components, representing the key 3D shape features of the input population, were visualised as -3SD to +3SD deformations from the mean shape.

The intrinsic characteristics of 3DMMs were evaluated using compactness, generalisation, and specificity as described in chapter 3.2. This was achieved by randomly generating 1,000 skull samples from the skull model using all principal components.

The mean shape was validated via a 10-fold cross validation approach – to ensure that the final shape was not overly influenced by adding of leaving out a specific subject – and using a geometric approach – to demonstrate that the 3D mean shape was representative of the study cohort (124). For the 10-fold cross validation, the study population was divided randomly into 10 subset folds. The mean shape was then computed using 9 of the folds, until each fold had been omitted once. The mean and standard deviation vertex difference between the original 3D mean shape and cross validated shapes was then calculated.

For geometric approach, the distances from Waitzman *et al.* 1992 (Table 5.2) and the additional 2D morphometric lengths and ratios (transcranial length (AS1, AS2), cephalic index, height index, oblique cranial length ratio) measured from the 3DMM mean shape were compared to the mean values calculated from the input population; deviations <5% were considered acceptable for the mean shape to represent the population with a good approximation. All measurements and ratios were then calculated for each of the synthesised skull samples. A two-tailed t-test was used to compare the mean of the generated samples to the mean values of the real <4 dataset for each of these metrics. A value of 0.05 was considered for statistical significance.

5.2.5. Manifold visualisation

t-SNE was applied to the high dimension latent vector encodings for the study population to assess age from shape. The population was labelled for age. The t-SNE embedding was created using a perplexity of 30 and run for 1,000 iterations.

5.3. Results

To construct the skull model, 178 unique scans of healthy children <4 years old were included, 55% were male. Scans were taken at mean age of 20.3 ± 12.9 months, Figure 5.3).

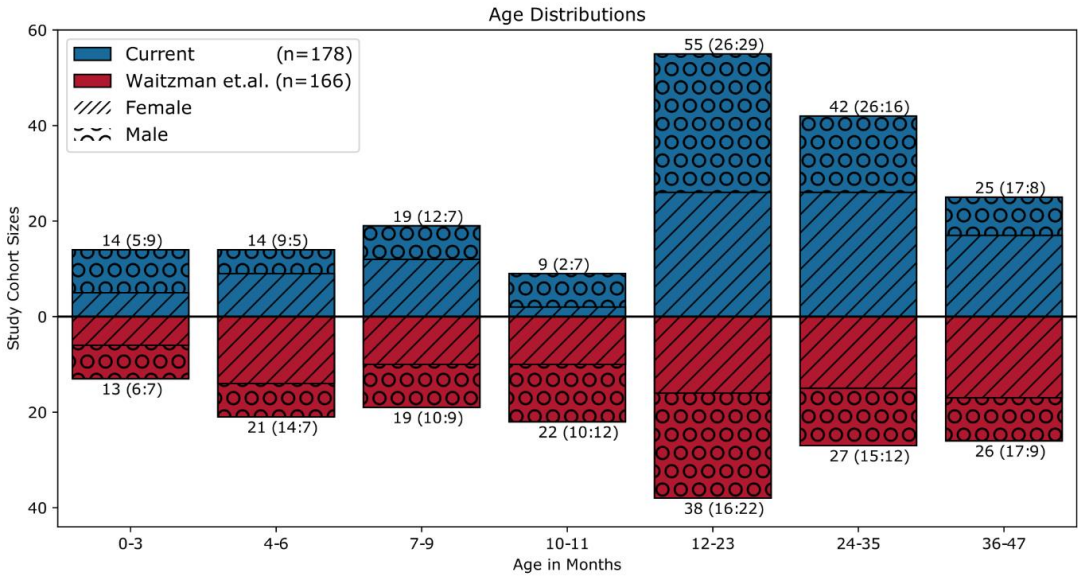


Figure 5.3. Study population for model construction and model validation. The <4 skull data (blue) was used for model construction. The model was validated using published data from Waitzman *et al.* 1992 (3). The age and gender distributions are presented in this figure.

5.3.1. Linear anthropometric measurements

The means and standard deviations of the <4 population for the assessed anthropometric measurements are shown in Table 5.3. where they are compared with data from Waitzman *et al.* 1992 (3), showing similar values and thus validating the selected cohort as representative of the <4 population. The greatest difference between the two datasets is observed for the inter orbital and temporal distance. Age group split downs on the included measurements are tabulated in Tables 5.4-5.6.

Table 5.3. Means and standard deviations for the anthropometric measurements. Measurements were acquired for real skulls and compared with samples generated using the model presented above. Values from the Waitzman *et al.* 1992 (W et al.) (3) were collected from their paper. AS= transcranial length, OCLR = oblique cranial length ratio.

Anthropometric Measure	W et al. [n=166]	<4 dataset [n=178]	Mean shape (% deviation from population mean)	Synthetic samples (<4 dataset vs. synthetic p values) [n=1,000]
Lateral orbital distance	74.18 ± 5.71	78.12 ± 5.33	78.77 (-0.83)	78.96 ± 3.54 (0.044)
Anterior inter orbital distance	18.30 ± 1.93	15.91 ± 1.52	16.06 (-0.92)	16.16 ± 1.52 (0.046)
Intertemporal distance	64.80 ± 5.21	75.97 ± 5.24	76.61 (-0.84)	76.77 ± 3.80 (0.054)
Inter-zygomatic buttress distance	70.20 ± 6.35	71.73 ± 6.58	72.14 (-0.57)	72.25 ± 2.25 (0.296)
Zygomatic arch length	42.64 ± 4.78	38.04 ± 4.76	38.05 (-0.03)	38.05 ± 1.88 (0.982)
Inter-zygomatic arch distance	86.86 ± 8.07	92.75 ± 8.82	93.24 (-0.53)	93.34 ± 2.64 (0.379)
Inter-coronal distance	101.11 ± 8.87	104.53 ± 8.42	105.31 (-0.74)	105.49 ± 5.09 (0.143)
Cranial length	157.79 ± 14.82	151.77 ± 15.73	152.01 (-0.15)	152.35 ± 6.51 (0.631)
Cranial width	121.93 ± 10.09	128.43 ± 10.60	128.38 (0.04)	129.79 ± 6.71 (0.102)
Cranial height	---	131.05 ± 11.24	130.56 (0.37)	132.18 ± 5.03 (0.187)
AS1	---	143.24 +- 12.41	143.889 (-0.453)	144.36 ± 5.60 (0.239)
AS2	---	141.85 +- 12.53	142.232 (-1.437)	142.75 ± 5.60 (0.346)
Cephalic index	---	85.04 ± 6.34	84.46 (0.68)	85.38 ± 6.25 (0.500)
Height index	---	86.60 ± 4.25	85.89 (0.82)	86.88 ± 4.16 (0.422)
OCLR	---	1.03 ± 0.03	1.01 (2.00)	1.03 ± 0.03 (0.503)

Table 5.4. Means and standard deviations for the cranial vault measurements. Measurements were acquired for samples generated from the skull model. Values from Waitzman *et al.* 1992 study were collected and collated for the desired age range. Skull model = data derived from current study; W et al. = data derived from Waitzman *et al.* 1992(3)

Age (months)	Inter-coronal distance, mm		Cephalic length, mm		Cephalic width, mm	
	<4 dataset	**W et al.**	**<4 dataset**	**W et al.**	**<4 dataset**	**W et al.**
0-3	96.36 ± 8.35	88.2 ± 6.5	119.07 ± 9.11	136.9 ± 6.7	105.60 ± 7.66	102.3 ± 5.5
4-6	104.06 ± 10.44	95.4 ± 6.6	129.98 ± 6.10	144.0 ± 7.4	119.14 ± 7.24	115.0 ± 5.8
7-9	105.17 ± 8.85	96.9 ± 5.0	142.08 ± 6.47	152.5 ± 9.9	122.80 ± 5.01	116.8 ± 6.7
10-11	102.68 ± 3.96	99.9 ± 4.4	140.30 ± 8.20	150.2 ± 6.3	126.26 ± 4.85	121.7 ± 7.0
12 – 23	105.81 ± 5.45	101.7 ± 6.6	153.98 ± 8.89	163.3 ± 8.5	128.36 ± 6.22	124.8 ± 7.3
24 – 35	112.28 ± 6.96	106.6 ± 9.3	160.84 ± 8.03	165.4 ± 15.8	135.09 ± 8.07	129.8 ± 4.9
36 – 47	112.06 ± 5.44	109.7 ± 5.0	164.83 ± 6.77	173.7 ± 7.7	135.26 ± 7.07	128.9 ± 7.1
Total	107.21 ± 8.21	102.61 ± 9.30	150.59 ± 15.56	160.51 ± 15.42	127.83 ± 10.56	123.31 ± 10.24

Table 5.5. Means and standard deviations for the orbital measurements. Measurements were acquired for samples generated from the skull model. Values from Waitzman *et al.* 1992 were collected and collated for the desired age range. Skull model = data derived from current study; W et al. = data derived from Waitzman *et al.* 1992 (3)

Age (months)	Lateral orbital distance		Anterior interorbital distance		Intertemporal distance	
	<4 dataset	**W et al.**	**<4 dataset**	**W et al.**	**<4 dataset**	**W et al.**
0-3	75.43 ± 7.76	65.4 ± 4.8	15.84 ± 2.09	17.4 ± 2.1	72.53 ± 6.15	58.5 ± 4.5
4-6	78.51 ± 7.50	69.9 ± 6.3	16.60 ± 1.90	17.9 ± 1.6	76.84 ± 6.42	63.1 ± 6.0
7-9	79.58 ± 5.64	72.9 ± 4.0	16.45 ± 1.64	18.6 ± 2.1	76.90 ± 6.09	64.0 ± 3.7
10-11	76.92 ± 3.17	73.5 ± 2.9	15.77 ± 1.38	17.8 ± 1.9	74.48 ± 3.77	64.1 ± 3.6
12 – 23	78.73 ± 3.58	74.7 ± 4.0	15.96 ± 1.47	18.2 ± 1.9	77.15 ± 3.50	65.1 ± 4.8
24 – 35	82.42 ± 3.16	77.7 ± 3.9	16.92 ± 1.16	18.4 ± 1.7	80.28 ± 3.47	66.5 ± 4.7
36 – 47	84.52 ± 3.66	79.1 ± 3.6	16.46 ± 1.66	19.3 ± 1.7	80.98 ± 3.82	68.3 ± 4.1
Total	80.13 ± 5.25	75.30 ± 6.18	16.32 ± 1.55	18.59 ± 2.13	77.92 ± 5.00	65.48 ± 5.46

Table 5.6. Means and standard deviations for the zygoma measurements. Measurements were acquired for samples generated from the skull model. Values from Waitzman *et al.* 1992 were collected and collated for the desired age range. Skull model = data derived from current study; W et al. = data derived from Waitzman *et al.* 1992 (3)

Age (months)	Inter-zygomatic buttress distance		Zygomatic arch length		Inter-zygomatic arch distance	
	<4 dataset	**W et al.**	**<4 dataset**	**W et al.**	**<4 dataset**	**W et al.**
0-3	64.98 ± 6.70	63.2 ± 6.9	32.30 ± 3.52	33.3 ± 3.4	82.21 ± 7.60	73.5 ± 5.6
4-6	68.97 ± 5.91	67.9 ± 7.4	33.59 ± 3.22	38.4 ± 3.0	88.02 ± 7.40	80.8 ± 6.7
7-9	71.53 ± 4.30	67.9 ± 4.8	36.02 ± 1.77	41.6 ± 3.6	90.87 ± 5.73	83.1 ± 3.2
10-11	69.64 ± 3.29	70.8 ± 4.4	35.36 ± 1.80	43.3 ± 3.3	90.79 ± 3.66	85.9 ± 4.2
12 – 23	72.33 ± 3.65	69.8 ± 5.5	39.05 ± 2.38	44.2 ± 3.3	94.18 ± 4.26	87.3 ± 5.6
24 – 35	77.20 ± 3.53	73.0 ± 4.5	41.92 ± 2.45	45.6 ± 2.9	100.45 ± 4.70	92.5 ± 5.6
36 – 47	80.08 ± 3.69	74.4 ± 5.5	43.87 ± 2.15	45.6 ± 2.8	103.35 ± 4.99	95.5 ± 4.8
Total	73.52 ± 5.94	71.27 ± 6.71	38.96 ± 4.30	43.22 ± 4.76	95.05 ± 7.90	88.76 ± 8.98

The greatest percentage difference between the cohort mean and the computed template was observed for OCLR (2.0%), though it is noted that the template is more symmetric in shape than the population average (Table 5.5.). The template was considered validated as a representative mean shape for the study cohort.

5.3.2. Skull 3DMM characteristics

The mean shape and the first five principal modes of shape variation are shown in Figure 5.4. The first principal component, accounting for 37.5% of the shape variation in the population (Figure 5.5.a) captures clear differences in overall skull length, width and height. The aerial images demonstrate the variations in the roundness of the skull, from a globally rounded shape to a more elongated in the antero-posterior direction. Large variations are seen for the midfacial length and protrusion and orbital shape, especially the lower orbital rim. The second component highlights shape variations localised mainly in the facial part with focus on the orbital size and zygoma and maxilla. Albeit variations in skull length are also visible. The

third and fourth component show an asymmetrical shape difference in the posterior portion of the cranium (aerial view). Moreover, moderate variations are seen for the midfacial width. The fourth components also capture variations in frontal bone positioning (frontal bossing/retruded forehead), notable in lateral view. The fifth component demonstrated more subtle differences in the skull height anterior / posterior proportions, orbital size and shape, and protrusion of the mid-face.

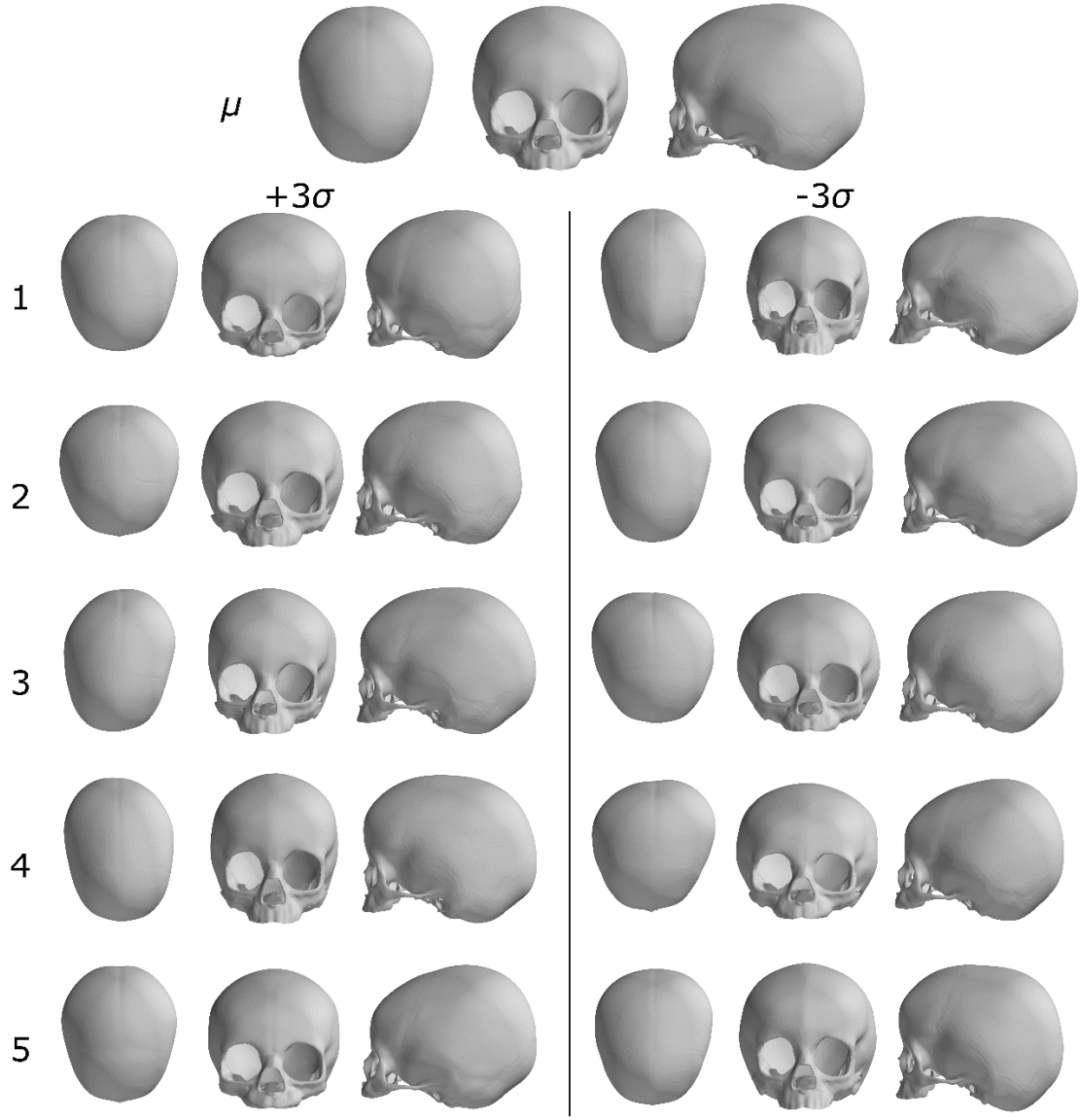


Figure 5.4. Visualisation of the <4 skull 3DMM. The mean shape, μ, and the first five principal components are shown. The principal components are visualised as either an addition or a subtraction from the mean shape with a weight of ±3σ, where σ_i is the standard deviation of the i^{th} principal component. Each model instance is shown at a 45° angle, from anterior, and from lateral.

Intrinsic model characteristics are shown in Figure 5.5. With 90% of the model shape variance captured within the first 10 principal components, and 95% of the variance captured within the first 20 principal components, the model can be considered reasonably compact. The generalisation error was 1 mm when 15 principal components were included, and of 0.47 mm when all model components were utilised, indicating that the model generalises well to unseen skull instances. Specificity values of less than 0.7 mm demonstrate that novel skull instances generated by the model are realistic.

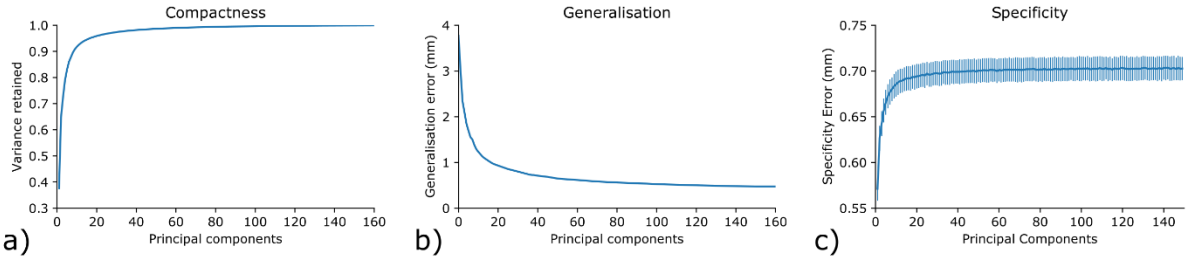


Figure 5.5. Intrinsic model evaluation for the Under-4 skull 3DMM. The presented are plots for compactness, generalisation, and specificity evaluation. **A)** compactness plot, the amount of variance retained for a certain number of principal components, is 95% at 20 components; **B)** generalisation plot, demonstrates the ability to describe the faces that were not used to construct the original model, and at 20 components is 1.0 mm; **C)** specificity plot, measures how well synthetic faces resemble real faces, and is 0.70 mm ± 0.02 mm at 40 components.

The generative abilities of the 3DMM are further validated by a good agreement across all measurements comparing the synthetic data with the real <4 population (Table 5.4.). Smaller standard deviations are observed for the generated samples; this can likely be attributed to the omission of size effects from the model. For the means of the real and synthetic data there are no significant differences but for the anterior interorbital distance and the lateral orbital distance, and it can be concluded that the means of the two populations not different. The mean

measurement of the assessed population demonstrates good agreement with those of the Waitzman *et al.* cohort, validating the study population as representative.

A trend for age can be noted on the t-SNE age embedding for the real population. A slight upper right distribution to lower left clustering can be seen in Figure 5.6. driven by increase in age. This means that the age of the patient can be identified by analysing the shape only.

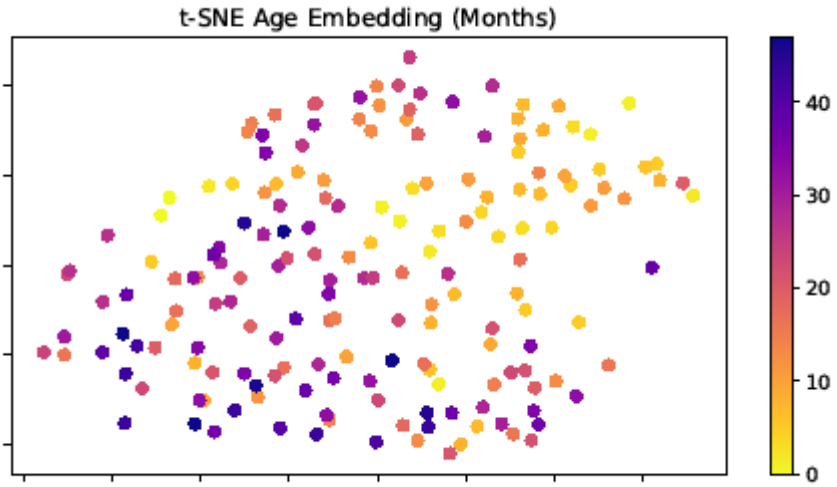


Figure 5.6. t-SNE embedding labelled for age in months. This embedding suggests a pattern for age based on skull shape information.

5.4. Discussion

The skull shape is constantly developing throughout life and these changes are related to numerous anatomical and functional factors such as skeletal growth, sinus development, and airway volume increase. This development is particularly pronounced during early childhood as the face and head undergo rapid growth. The pace of these skeletal changes, coupled with the difficulty of collecting sufficient normative 3D skull data for a young age group, have

presented challenges to large-scale shape analysis and therefore a true 3-dimensional understanding of skull-based shape changes of this group was previously unknown.

This chapter presented a 3DMM constructed from 178 scans of the <4 normative dataset and demonstrated its capability for generating valid novel skull instances. This model can be applied as an alternative to the limited availability of skull data for 3D analysis and might open doors to many applications where access to data is currently a limiting factor, such as forming a reference for surgical planning or diagnostics for rare craniofacial syndromes like Apert syndrome.

The 3DMM was validated using published reference data from the work of Waitzman *et al*. 1992, which is the most comprehensive literature available on bone measurements (i.e. cranial and midfacial standardised measurement) in a healthy paediatric population. Validation was performed for both the raw data meshes and for newly generated meshes from the 3DMM. A strong agreement between the CI, OCLR, and all assessed skull measurements is observed in both cases. These results, combined with the intrinsic evaluation of the 3DMM for compactness, generalisation, and specificity, confirm the model to be useful and robust.

This model, which contains information about the dominant 3D skull shape variations, can be used for a variety of applications including surgical planning, post-operative assessment, and syndrome identification. In cases where mirroring of a pre-operative scan for surgical planning is not an option due to bilateral craniofacial abnormalities, an age and gender matched surgical reference template derived from this model could be a realistic solution in the future. However, prior to this, gender matched references should be constructed, which is not included in this thesis.

No objective post-operative assessment tool exists to evaluate the aesthetic success rate or degree of ‘normalisation’ of children undergoing surgical correction of craniofacial

anomalies. At present, the outcome is judged by the operating surgeon and the patients'/caregivers' view which is generally biased and can be influenced by subjectivity. With the use of an age matched bone reference 3D object, an objective post-operative assessment could be performed. This model could also be used to compare the post-operative outcome with the "ideal" outcome, providing an outcome of the closest possible normalised skull to understand the minimum surgical change required. This analysis can inform the surgeon on the skeletal regions where the outcome matches the ideal 'normalised' outcome, and, perhaps more importantly, it can indicate which regions of the skull require greater changes to resemble the generated outcome more closely. These evaluations are currently performed in research settings and using the presented model, could be developed into a clinically useful software. Building an easy-to-use interface for the use of these algorithms in clinical practice should be part of the next steps, removing the need for programming expertise and allowing clinicians without a computer vision background or technical skills to use these tools.

The diagnostic experiments that demonstrated excellent results for the soft tissue models in chapter 3 and 4 could also be applied to the bone models and expended to various syndromes across a range of different age groups to enhance radiologic diagnostics. This would require the construction of bone models for all these specific groups. In addition to diagnostics, they could facilitate greater understanding of the growth and shape changes within an unoperated syndromic population. This might help to better indicate the timing of treatments and surgical correction of the abnormalities alongside indicators deriving from functional concerns. A 3DMM of Apert will be addressed in the next chapter to address the diagnostic power and increase understanding of natural skeletal development.

Craniofacial surgery relating to functional problems, such as midfacial advancement for airway issues, posterior vault expansion or cranial remodelling for raised intracranial

pressure, is typically performed in the first years of life. Conversely, surgical intervention for aesthetic reasons tends to occur later in life as this type of corrective surgery is ideally performed after maturation of the skeleton and eruption of dentition. These surgeries include upper jaw advancement, monobloc advancement, and midface advancement. A model useful for post-operative assessment for the latter indication would require an adult skull model. The construction of a model including patients above age 4 has been initiated but is not included in this thesis.

Lastly, this model could form the base of a comprehensive and a more detailed model. The existing soft tissue model could be fused for a robust combined model. Brain models could be made, the vascular system could be added, as well as a model of the mandible to enhance the usability, which on its own could be useful for pre- and post-operative analysis as well as analysis on mandible shape and airway problems. To date, it has been challenging to combine models from different regions, a first combined model for the face and head has been proposed by Ploumpis *et al.* 2019 (15).

A limitation of the model in its current state, is that there is no easy-to-use interface available yet. In order to use the model, some technical skills of the user are required. Moreover, the interpretability of the representations could be improved. PCA, the most commonly used method to perform statistics on 3D faces, as also used in this chapter, is an unsupervised method, the principal components do not coincide with attributes that humans would use to describe the face/skull. Moreover, the morphable models do not include size and would require to be re-scaled prior to surgical planning use. This could be addressed by building a user-friendly interface for the use of clinicians in their preoperative work-up.

5.5. Summary

A skull model was successfully constructed for a healthy population under 4 years of age. Linear anthropometric measurements were compared to gold standard data derived from literature to validate the model. The model provides large normative data and many potential applications for a generative skull model were outlined.

Key points

- Successful construction of a normative <4 skull model.
- Large normative data available due to generatability.
- Model applications include automated syndrome identification, in silico surgical planning, objective post-operative and normalisation assessments.
- Model is not user-friendly yet.

6. AN APERT SKULL MODEL

Part of the work described in this chapter has been published in:

The Journal of Craniofacial Surgery (abstract)
A 3D skull morphable model of apert syndrome

- Anne-Jet C. Oosting, **Lara S. van de Lande,** Eimear O' Sullivan, Athanasios Papaioannou, Maarten J. Koudstaal, Silvia Schievano, Stefanos Zafeiriou, David J. Dunaway

Part of this work was presented at:

- *The 25th Congress of the European Association for Cranio Maxillo Facial Surgery (EACFMS), 14-16 July 2021, Paris, France (Virtual Event).*

Following the introduction of a normative paediatric skull model (<4) that provides a large set of normative measurements, this chapter will focus on the Apert skull, describing the construction of a 3DMM and analysing growth curves and skeletal shape changes of the unoperated Apert population.

6.1. Introduction

Statistical modelling of the craniofacial skeleton has been pursued for a few craniofacial disorders, including Crouzon-Pfeiffer syndrome and Craniofacial Microsomia, where PCA of 3D meshes derived from CT scans was applied for skeletal assessment, facial skeleton and mandible respectively (114, 125). These models were utilised to understand the shape changes and variations of the syndrome with age. Some other studies compared syndromic with age-matched healthy individuals to those, however, analyses were based on small cohorts and mostly based on cephalic measurements on CT scans and lack information on growth patterns. The details of these studies are presented in more details in chapter 2.2.5 where the few studies that attempted to objectify the facial features were summarized (62-66). This chapter demonstrates the construction and validation of an Apert skull 3DMM and provides growth charts for an unoperated population to further understand the natural growth patterns. It also aims to test the radiological diagnostic capability for automated identification using facial bone shape information.

6.2. Methods

6.2.1. Data sources

All patients diagnosed with Apert Syndrome between March 1994 to February 2020 at *GOSH* and between August 2007 to July 2019 at *NEMH* were reviewed for preoperative head CT scans and were included following the inclusion and exclusion criteria (Table 6.1). Baseline characteristics were collected from (electronic) medical charts. In case a patient had multiple CT scans available, only one CT scan with the highest quality for 3D construction was selected. In case more than one scan had sufficient quality for 3D conversion, a selection took place based on age to obtain a consistent age-distribution throughout the dataset to minimise selection bias.

To match the normal <4 model, a subgroup of the Apert data was made consisting only scans <4 years. Another group was made that included Apert all ages, to compare with the Waitzman *et al.* 1992 (3) data that includes normal data from individuals 0-17 years of age.

6.2.2. Model construction

CT scans of the Apert were processed as described in chapter 5.2 for model construction. Two models were built: 1) Apert <4 and 2) Apert all ages. The intrinsic characteristics of 3DMMs were evaluated using compactness, generalisation, and specificity (41, 43, 102) as described in chapter 3.2.

Table 6.1. Inclusion and exclusion criteria for Apert skull model.

	Inclusion	Exclusion
CT-head scans	• Diagnosis Apert syndrome • High quality scans (≤1 mm or ≥150 slices)	• Incomplete scans (full head scan required, scans missing parts of the maxilla or cranium were omitted). • Scans with movement artefacts disturbing bone 3D construction. • Insufficient DICOM slices for 3D construction. • History of craniomaxillofacial surgery.

6.2.3. Intrinsic skull morphometrics

Registered mesh data were used for measurements and construction of growth charts. Growth charts were plotted for: 1) the Apert <4 against the normal <4, 2) the Apert all ages against normal <17 (i.e. Waitzman *et al.* 1992 (3) data.)

Based on the measurements considered in Waitzman *et al.* 1992 (3), the following distances were considered for comparison and construction of growth charts: zygomatic arch length, inter zygomatic arch width, inter zygomatic buttress distance, inter lateral orbital wall distance, inter inferior rim distance, cephalic width, cephalic length, and inter coronal distance. A template for automated distance calculations was constructed to mimic the anatomical points that were used to calculate the above named distances as described in Waitzman *et al*. 1992 (3) (Table 6.2 and Figure 6.1).

For the comparison of Apert <4 and normal <4, distances were automatically measured using the landmark template from Figure 6.1, and a logarithmic fit for the data was plotted.

For comparison of the Apert all ages group with normal <17, the data from Apert all ages were plotted using automated calculations of distances from the landmark template and a

logarithmic fit was plotted, for the measurements from Waitzman *et al.* 1992 the means and standard deviations were plotted with a line of fit between the means.

Table 6.2. Landmarks used for skull measurements. These landmarks were used for distance measurements. **R** indicates the right-hand side of the skull, whereas **L** indicates the left-hand side.

Landmarks	Definition
0-1	Inter lateral orbital walls distance
2-3	Inter anterior medial orbital walls distance
4-6	Zygomatic arch length R
5-7	Zygomatic arch length L
8-9	Inter zygomatic arch width
10-11	Inter coronal distance
12-13	Inter zygomatic buttress distance

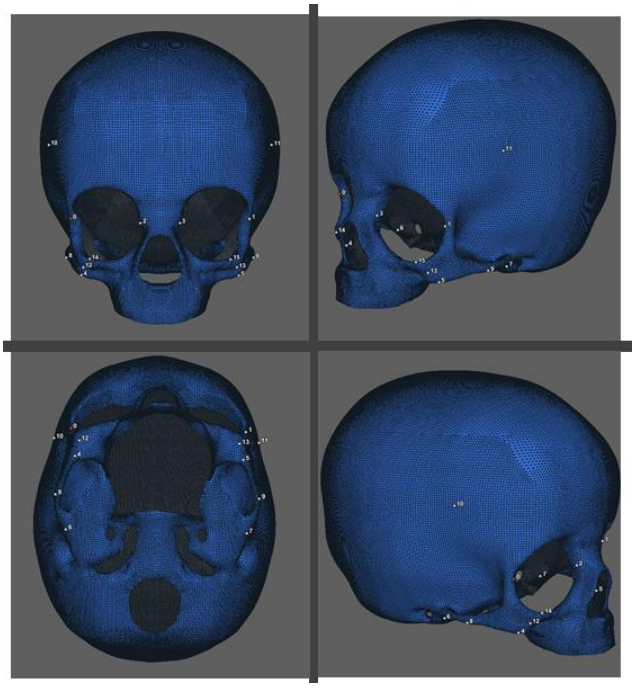


Figure 6.1. Landmark template for skull measurements. In this figure the landmarks indicated in Table 6.2 are visualised, this template was used for automated distance calculations for the registered meshes of Apert <4 and normal <4.

6.3.4. Manifold visualisation of Apert skull 3DMM

t-SNE was applied to the high dimension latent vector encodings for the Apert all ages and the normal <4 bone meshes. Due to lack of normative bone meshes of age >4 years, embedding was performed using only the normal <4 as control group. t-SNE embedding was created using a perplexity of 30 and run for 1,000 iterations.

6.3. Results

Fifty-two skull meshes from unoperated Apert patients were included for intrinsic skull measurements and model construction. The majority of the patients were <4 years of age (Figure 6.2) with a second subgroup notable between 11-17 years. More males (63.5%) than females were included.

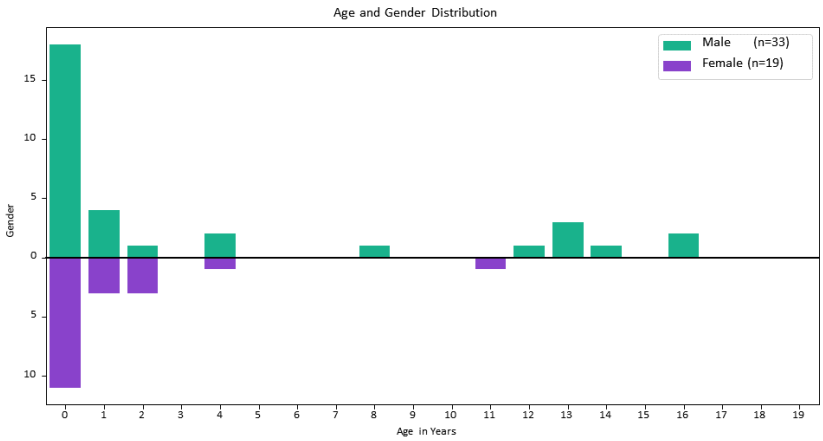


Figure 6.2. Age and gender distribution Apert skull 3DMM. A total of 52 unoperated Apert patients were included for intrinsic skull measurements and model construction.

6.3.1. Anthropometric linear measurements

Abnormal facial growth in Apert syndrome in childhood (<4)

Unambiguous differences are seen for the zygomatic arch length. The data suggest that the growth curve does not follow the curve of normality as it misses the initial growth spurt occurring within the first year of life. The width of the zygomatic arches, however, seems within the lower range of normality, resulting in an overall short zygomatic arch with a strong curvature. Moreover, the distance between the left and right zygomatic buttress is larger in Apert and increases with age. As expected, large differences are seen for the lateral orbital wall distance and the distance between the inferior orbital rims. Whereas the distance between the inferior orbital rim seem to flatten over time, the distance between the lateral orbital walls only increases more. Up to five months of age, the cephalic width in patients with Apert is similar

to that from normative data; however, a wider head is seen for the growing Apert child. This trend is similar to that of the inter coronal distances which can be noted in the plot. The cephalic lengths measured for the Apert patients start off in the lowest range of its normal reference and appear to follow the same growth pattern as the healthy population (Figure 6.3).

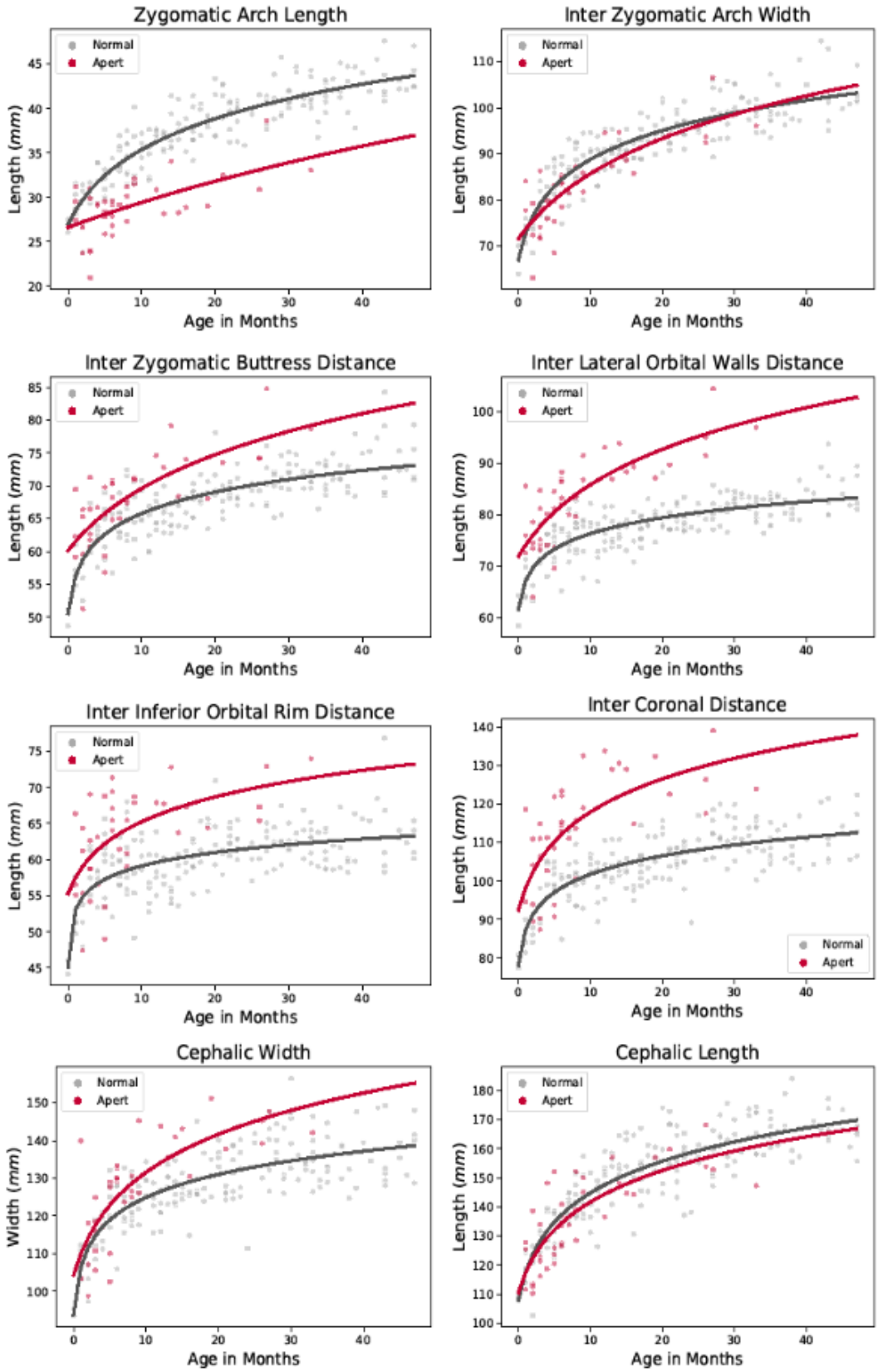


Figure 6.3. Comparison of craniofacial measurements for the Apert <4 with normal <4. Growth charts are provided for zygomatic arch length, inter zygomatic arch width, inter zygomatic buttress distance, inter lateral orbital wall distance, inter inferior orbital rim distance, inter coronal distance,

cephalic width, and cephalic length. The zygomatic arch length misses its initial growth spurt and remains up to 10 mm shorter than its age-matched normal. A wider maxilla is noted as demonstrated by the overall larger distances between the zygomatic buttresses. The inter orbital distances are 5-8 mm larger in the first weeks of life and deviate from the normal with age. The cephalic width increases more than the normal after 5 months of age, which is confirmed by a similar pattern for the inter coronal distances. The cephalic length starts off in the lowest range of normal and appears to normalise over time.

How does the abnormal growth persist into adulthood?

Clear differences are seen and were expected for the zygomatic arch length. As noted in more detail in the <4 age group, the zygomatic arch deviates from the normal curve starting at some of the youngest scans onwards. Whereas an Apert patient has an average length of 40 mm at age 13, this is 55 mm in the unaffected population. The zygomatic width is initially plotted on the lower border of normal, whereas it shifts to the wider border of normal at an older age. The zygomatic buttress distance persists to be on the larger range of normal. The distances between the lateral orbital wall are clearly different between normal and Apert, which is demonstrated from the youngest scans up to the oldest meshes included. In the normal data 80 mm is expected for a 2-year-old, where this is 93 mm in the Apert population, and later in life, at age 16 years, 95 mm for normal is measured whereas this is calculated as 110 mm for Apert. The anterior medial orbital wall is measured larger for the Apert population in the first 4 years of life, however, seem to vary extremely for the adult subgroup, with measures below and above the normal fit of means. Also, for the healthy population, large variations are seen for this parameter (note the large standard deviations in Figure 6.4). The cranium continues to widen more as compared to normal and although initially the cephalic length seems to follow a normal growth pattern, obvious shortening of the head is plotted for the complete dataset (Figure 6.4).

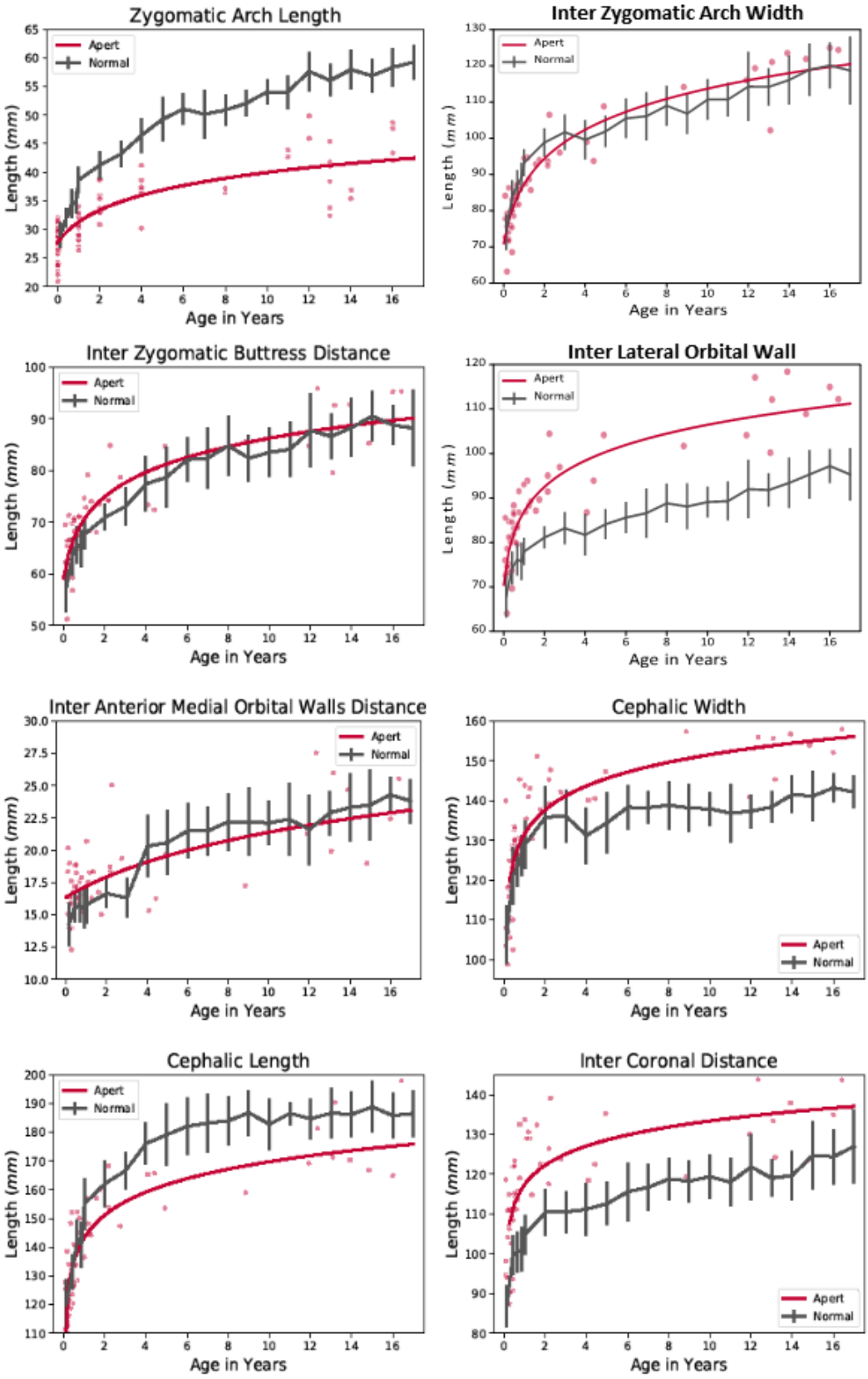


Figure 6.4 Comparison of craniofacial measurements of Apert all ages with normative data from literature (0-17 years). Growth charts are provided for zygomatic arch length, inter zygomatic arch width, inter zygomatic buttress distance, inter lateral orbital wall distance, inter inferior orbital rim distance, inter coronal distance, cephalic width, and cephalic length. The zygomatic arch length

continues to deviate from the normative curve and the distance between the lateral orbital walls continue to deviate from the normal curve as well. The cranium continues to widen more as compared to normal and although initially the cephalic length seems to follow a normal growth pattern, obvious shortening of the head is plotted for the complete dataset.

6.3.2. Apert 3DMM

Figure 6.5 (top) illustrates the mean and first 5 principal components of the Apert <4 and Apert all ages model. For the Apert <4 model the first component highlights large differences of the frontal bone with frontal bossing (pronounced bulging of the frontal bone) is most apparent in the -3SD of component one and +3SD of component 5. The orbital shape is strongly variated in shape and size, which is characterized by a more pronounced supra orbital rim vs retruded supra orbital rim in either SD, and variations in the position of the infra orbital rim. Moreover, the cranial shape shows large variations with severe brachycephalic head shape in the -3SD. The second component shows mainly variations of the orbital width, orbital size, and moderate variations of frontal bossing. The third component highlights variations in cranial shape asymmetry, moderate variations in brachycephaly, and underdevelopment of the supra orbital rim. Moreover, hypoplasia of the maxilla for the +3SD with a protrusion of the maxilla in the -3SD relative to the upper third of the facial skeleton. The fourth component illustrates variations of the cranial width and position of the vertex (highest point of the skull), also moderate variations can be noted for the supra orbital rim. The fifth component demonstrates variations of the frontal bone from severe frontal bossing to a declined frontal bone. Moreover, again, variations of the positioning and development of the supra and inferior orbital rim can be noted, in a lesser extent than the former components. In all components, a tall frontal bone is noted.

The Apert all ages model, illustrated in Figure 6.5 (bottom), where similar outcomes of the components can be noted as the Apert <4 model. However, with more extreme variations for the components.

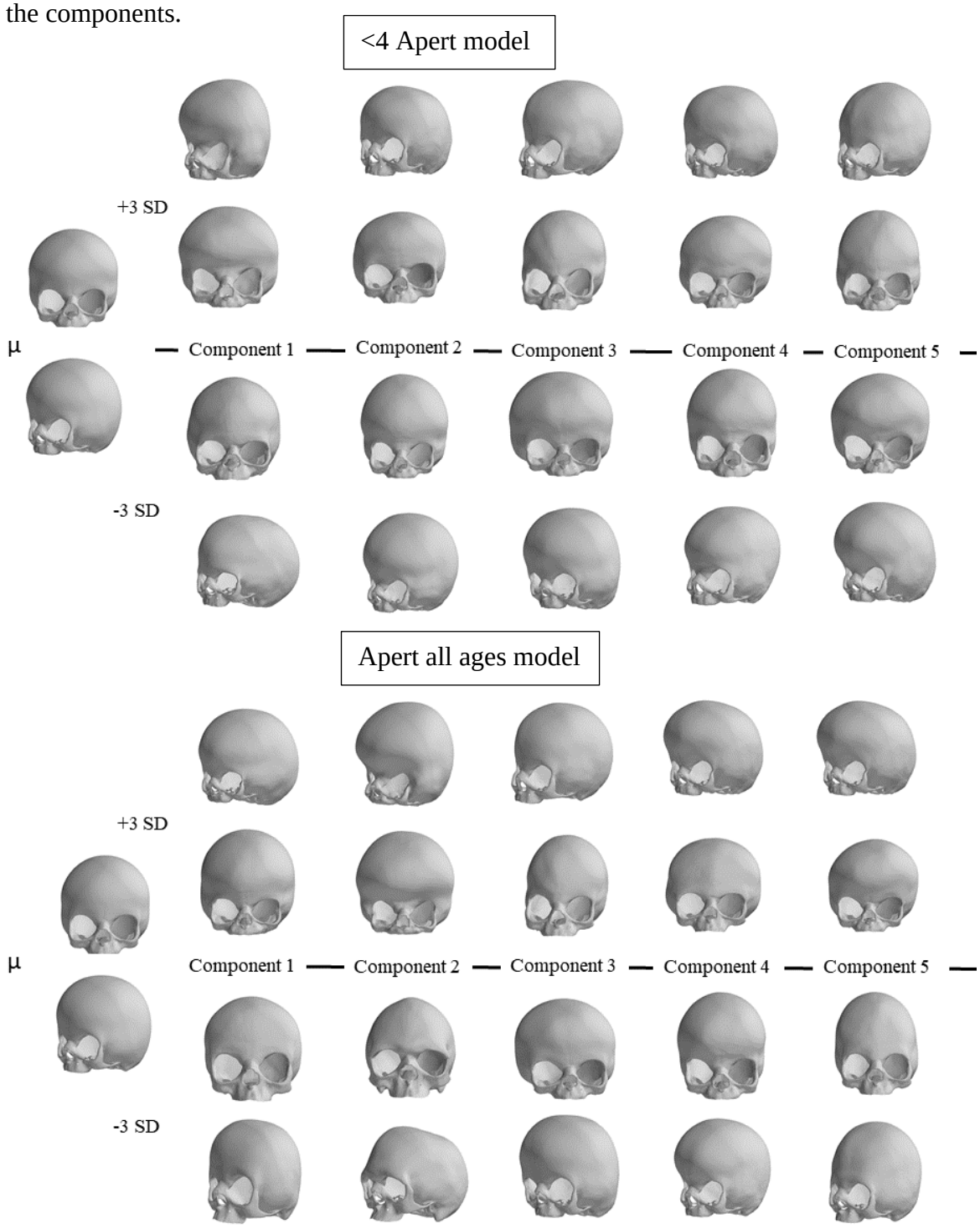


Figure 6.5. Visualisation of the Apert skull 3DMMs for <4 and all ages. The top model consists of the Apert <4 data, where large variations can be noted for the positioning (development) of the frontal bone, retrusion of the supraorbital rim, location of the infra orbital rim, maxillary protrusion/hypoplasia, cranial asymmetry, width and length. Similar outcomes are highlighted for the all ages model (bottom) in a more extreme extent.

The Apert <4 model has a compactness of 95% at 15 components, good generalisation of 2.5 mm at 20 components and a reasonable specificity of 1.2 mm ± 0.05 mm at 20 components. For the Apert all ages model, this was 95% at 15 components for compactness, 2.5 mm at 20 components for generalisation, and 1.3 mm ± 0.05 mm at 20 components for specificity. (Figure 6.6 -6.7)

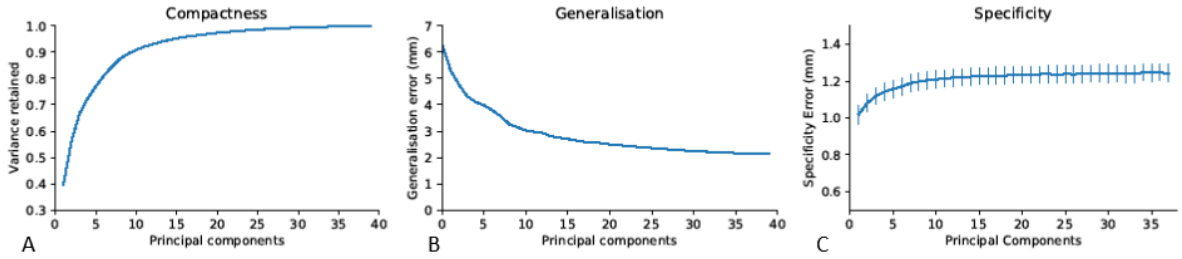


Figure 6.6 Intrinsic model evaluation for the <4 Apert skull 3DMM. The presented are plots for compactness, generalisation, and specificity evaluation. **A)** compactness plot, the amount of variance retained for a certain number of principal components, is 95% at 15 components; **B)** generalisation plot, demonstrates the ability to describe the faces that were not used to construct the original model, and at 20 components is 2.5 mm; **C)** specificity plot, measures how well synthetic faces resemble real faces, and is 1.2 mm ± 0.05 mm at 20 components.

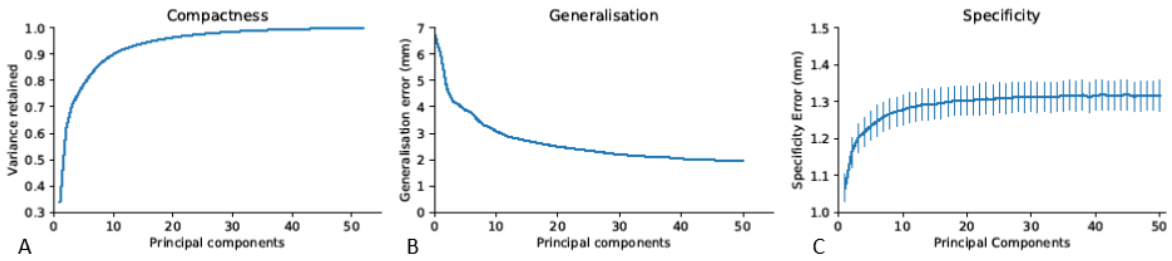


Figure 6.7 Intrinsic model evaluation for the 0-240 months Apert skull 3DMM. The presented are plots for compactness, generalisation, and specificity evaluation. **A)** compactness plot, the amount of variance retained for a certain number of principal components, is 95% at 15 components; **B)** generalisation plot, demonstrates the ability to describe the faces that were not used to construct the original model, and at 20 components is 2.5 mm; **C)** specificity plot, measures how well synthetic faces resemble real faces, and is 1.3 mm ± 0.05 mm at 20 components.

6.3.3. Classification

Clear clustering can be seen for the Apert all ages bone meshes following t-SNE embedding with the normal <4 meshes (Figure 6.8). Classification of Apert all ages versus normal demonstrate clear clustering following t-SNE embedding. This confirms that 3DMMs can not only be used for automated diagnosis for Apert for soft tissue as described in chapter 3, but also for identification based on bone meshes.

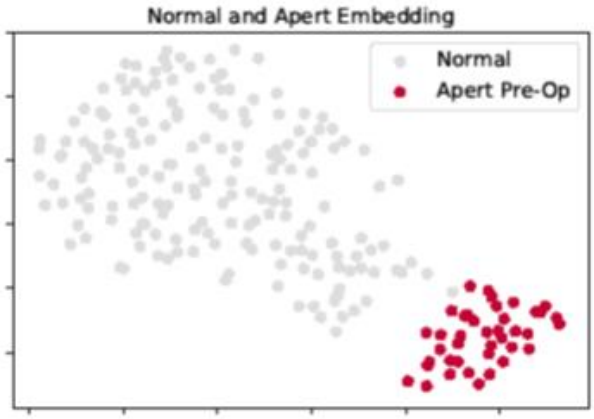


Figure 6.8 Normal and Apert t-SNE embedding. Clear clustering is seen for the unoperated Apert skulls from the normative data, suggesting automated diagnosis can also be made on a bony level.

6.4. Discussion

This chapter described the construction of an unoperated Apert skull model and gave insight into the natural and abnormal development of the skull size and shape utilising 3D analysis and growth curves. Comparison of Apert syndrome with normative data provided information on the natural development and reported standard craniofacial measurements.

Interestingly, the model for all ages demonstrated similar outcomes as the Apert <4 model for the components, however, in a larger degree. This more obvious increase in deformity over time was also seen for soft tissue (chapter 3) and contains information that has, to my knowledge, not been described in literature. This increase in abnormality might also teach us

something for the correction of the face and might perhaps open new discussions on the most suitable approach for surgical correction. For example, even though the zygomatic arch lacks growth from the very young, it does grow somewhat with age. Moreover, a clear hypertelorism is seen in youngsters, yet a very disperse variation is seen for those aged 12-17 years old, which could infer that some children might grow into an orbital distance within normal limits and others, maintain above the normal curve. These findings could have direct clinical implications by justifying or even encourage delay in surgery until maturation to identify the best surgical approach for that specific patient and thereby potentially avoid sequential reoperations.

Growth curves can be useful for surgical planning by understanding the natural course for this patient populations' shape development. In combination with the visualisation of the 3D shapes it provides even more information and a better understanding of the shape changes. Using this data, any type of measurements can be calculated on demand by simply defining the landmarks that captures the region of interest: the distances can be calculated automatically from all meshes included in the dataset. As mentioned in chapter 5, a future study has been initiated to expand the input data to adulthood, so this model could potentially replace the Waitzman *et al.* 1992 graphs as 3D data might provide more accurate and interactive growth charts (3). Ideally, an easy-to-use software with access to the meshes should be available for clinicians to select their measurements of choice.

Future additions to these measurements might include the use of 3DMMs for volumetric analysis based on airway dimensions, the mandible, and other craniofacial regions of interest. Moreover, constructing skull models for other types of craniofacial syndromes might benefit clinicians for the abovementioned reasons.

3DMM of the Apert population can potentially aid with automated diagnostics on a skeletal level; this might however be of less clinical value than automated diagnosis based on soft tissue

meshes as the goal would be to avoid radiating a vulnerable population such that of craniofacial patients. However, a combined model might provide insight into the relationship between bone and soft tissue in the growing Apert's face.

Limitations

As raw data was lacking for analysis of the normal >4, published means and standard deviations were used from Waitzman *et al.* 1992 (3), which is to date the gold standard due to lack of more recent normative databases including 3D information, for this age group. As mentioned, a study has been initiated to overcome this problem, but was not included in this thesis. Moreover, this study included little data for the age group between 5 and 12 years of age. Yet, it could still give some insight into the growth patterns of specific areas of the Apert's face and cranium, mainly based on the line of best fit. However, to fully study the growth and shape change of the Apert's face, more and better distributed data is needed. The reason for this unbalanced distribution of the data is due to the timing of surgery, which is either performed during infancy or around the age of 15-17 years. CT scans are generally acquired before surgery; therefore, no CT scans were available for the remaining age groups. Collaboration with other craniofacial units could provide an opportunity to saturate this model as they might perform scans more frequently or indicate surgery at a different age than done at *GOSH* and *NEMH*.

6.5. Summary

A 3DMM of an unoperated Apert population was presented. Growth chart were constructed and compared to a healthy population. Most measured parameters deviate from the normal

growth curve and increase in abnormality was noted. 3D analysis provided valuable insight into the shape differences of the growing Apert versus a normal model.

Key points

- Apert skull 3DMM was successfully constructed.
- Growth charts were provided with comparison for normative data.
- Increase in abnormal growth and shape deformity is apparent.
- Manifold visualisation demonstrated value for automated radiological identification.
- Potential future applications were outlined.

7. FACIAL SURGICAL OUTCOMES

Part of the work described in this chapter has been published in:

Journal of Orthodontics, 2019 (invited review)
Geometric morphometrics aided by machine learning in craniofacial surgery.

- **Lara S. van de Lande**, Athanasios Papaioannou, David J. Dunaway

The Journal of Craniofacial Surgery, 2020 (original article)
Local Soft Tissue and Bone Displacements following Midfacial Bipartition Distraction in Apert Syndrome – Quantification using a Semi-Automated Method.

- **Lara S. van de Lande**, Eimear O' Sullivan, Paul G.M. Knoops, Athanasios Papaioannou, Juling Ong, Greg James, N. Owase Jeelani, Silvia Schievano, David J. Dunaway

The Journal of Craniofacial Surgery, 2021 (abstract)
Why we should use non-rigid iterative closest point registration for surgical outcome measurements – demonstrated by midfacial bipartition distraction in apert syndrome.

- **Lara S. van de Lande**, Eimear O' Sullivan, Paul M. Knoops, Athanasios Papaioannou, Juling Ong, Noor O. Jeelani, Silvia Schievano, David Dunaway

Part of the work described in this chapter has been presented at:

- *18th Congress of International Society of Craniofacial Surgery (ISCFS), 16-19 September 2019, Paris, France.*
- *The 2nd International Craniofacial Morphometrics Meeting, 20-21 September 2019, London, UK*
- *The 25th Congress of the European Association for Cranio Maxillo Facial Surgery (EACFMS), 14-16 July 2021, Paris, France (Virtual Event).*

Chapters 3-6 addressed soft tissue and skull 3D models to understand shape, shape changes over time, and reported growth charts for the Apert and normal paediatric population. This chapter will focus on the local 3D shape changes induced in Apert patients by craniofacial surgery, with the aim to identify in which areas the surgical technique might be lacking to achieve facial normalisation.

7.1. Introduction

Adequate assessment of facial changes following craniofacial surgery is essential to improve surgical techniques, choice of procedure, and surgical planning. At *GOSH*, patients with Apert syndrome undergo mostly facial bipartition distraction with RED-frame (FB-RED), with or without a nasal bone graft to correct their facial deformity. The choice for FB-RED at *GOSH* is based on the unique facial deformity of Apert syndrome, where not only the midface is retruded, but also the intrinsic shapes are radically distorted as described in chapters 3 and 6. FB-RED makes use of osteotomies that change the shape of the midface and orbits in contrast to a Monobloc or Le Fort III distraction, which moves the osteotomised region en-bloc.

Facial bipartition distraction and Le Fort II with zygomatic osteotomies (described in chapter 2.3) both represent attempts to treat the intrinsic flatness of the Apert midfacial skeleton and address other abnormalities such as hypertelorism and the disproportionate midfacial height.(71, 75-77) These intrinsic changes in midfacial bone structure are difficult to quantify using conventional cephalometric and anthropometric measurements, several have attempted to do this as described in chapter 2.3. However, these methods rely on a few facial landmarks and do not allow for a holistic analysis of the shape changes (126). Analysis of facial shape change is further complicated by the fact that the overlying soft tissues are abnormal, such as the positioning of the canthi, and might not move in a conventional relationship with the

underlying bones. At *GOSH*, the choice of FB-RED for the correction of Apert syndrome is mainly driven by surgeon's experience and by outcomes of some studies that attempted to quantify the surgical outcomes.

Automated methodologies that do not rely on facial landmarks have also been used to analyse the soft tissue and skeletal deformations (127). Almukhtar *et al.* 2018 made use of the rigid ICP registration algorithm to align the pre- and post-operative 3D reconstructions for the analysis of the mandible following jaw surgery. The 3D shape displacements were then evaluated by determining the closest point-to-point correspondences between pre- and post-operative surface pairs, regardless of their anatomical correspondences. However, closest point correspondences have been reported not to accurately represent the true displacements of different anatomical features (128). To obtain the true anatomical point-to-point correspondences (i.e. anatomical correspondence, or dense correspondence), a more flexible version of rigid ICP, non-rigid ICP (described in chapter 3.2.) can be applied to the 3D bone and soft tissue reconstructions. NICP has been used for soft tissue analysis for patients following upper jaw surgery (Le Fort I) in a study by Knoops *et al.* 2019 (21), yet, it is rather new in the evaluation of surgical outcomes on a skeletal level for patients undergoing craniofacial surgery. In other fields, NICP has been used for intra-operative navigation for brain tumor resection or identifying the smallest region for intra-operative radiation in spinal surgery (129, 130). Introducing this technique for skeletal level evaluation could allow for not only the analysis of rigid movements (e.g., advancements and rotations due to the RED-frame distraction), but also the smaller, more localised, 'intrinsic' changes resulting from the distraction. Prior studies on this topic analysed the rigid displacements and rotations that occur following RED-frame distraction, leaving the intrinsic, local bony changes due to distraction unknown (78, 89, 131-133). While the soft tissue changes following these skeletal movements have been studied, the soft tissue response to the skeletal movement on a local level is unclear.

Some studies have highlighted the challenges in determining how soft tissue displacements are driven by the skeletal movements, and non-linear relationships between bone and soft tissue displacements have been described (21, 132, 134).

This chapter presents a 3D morphometric analysis using a NICP workflow methodology. It quantifies the complex changes in hard and soft tissue topography, giving the surgeon a better insight into the changes resulting from FB-RED in Apert syndrome.

7.2. Material and methodology

Pre- and post-operative CT scans of patients with Apert Syndrome who underwent FB-RED at *GOSH* between 2005 and 2019 were collected. CT scans of insufficient image quality, such as those with a limited number of slices (<100 slices), too thick slices (>1 mm), or movement artefacts, were excluded, as were patients who did not have both pre- and post-operative CT scans available. Baseline characteristics were collected, including gender, craniofacial surgical history, age at time of surgery, and time between pre-operative CT scan, FB-RED, and post-operative CT scan.

A pipeline was created for the semi-automated 3D quantification of bone and intrinsic soft tissue changes following craniofacial surgery. All 3D bone and soft tissue reconstructions from CT scans were constructed and cleaned using Mimics Inprint 3.0. as described in chapter 3.2.2. and 5.2.2. The 3D bone meshes were further processed to leave only the FB-RED surgical site, i.e. frontal bone, orbits, midface and zygomatic arch. Therefore, the mandible and vertebrae were semi-automatically removed by applying the foreground/background tool as described in chapter 5.2.2, whilst the posterior part of the calvarium was removed by a vertical plane cut

approximately 1 cm posterior from the coronal suture, for the pre-operative 3D bone mesh, and 1 cm posterior from the frontal bone osteotomy, for the post-operative 3D bone mesh. This was done using Meshmixer (99). The reason for removing the back of the head was to ensure only registration on the front of the face/skeleton, where the region of interest lies in this study.

Local changes

As this chapter aims to quantify the local bone and soft tissue changes following distraction, the post-operative 3D meshes were aligned (superimposed) on the pre-operative maxilla. This was done so that the overall midfacial distraction rigid displacement was removed from the analysis and the quantification could focus solely on the local changes of the bone and soft tissue. The maxilla was chosen as the alignment area for all pre- and post-operative 3D meshes, as this bone segment experiences the greatest displacement in terms of rigid advancement during facial bipartition distraction. Prior studies have also shown that the central part undergoes more distraction than the lateral areas (78, 89). The alignment was achieved using a rigid transformation (Figure 7.1). To obtain this rigid alignment, the maxilla (i.e. area for alignment) was landmarked with the 6 points shown in Figure 7.2 on both pre- and post-operative meshes using Materialise 3-matic software (135). The same rigid transformation used to align the bone meshes was applied to the soft tissue, so that bone and corresponding soft tissue 3D meshes moved as a single entity.

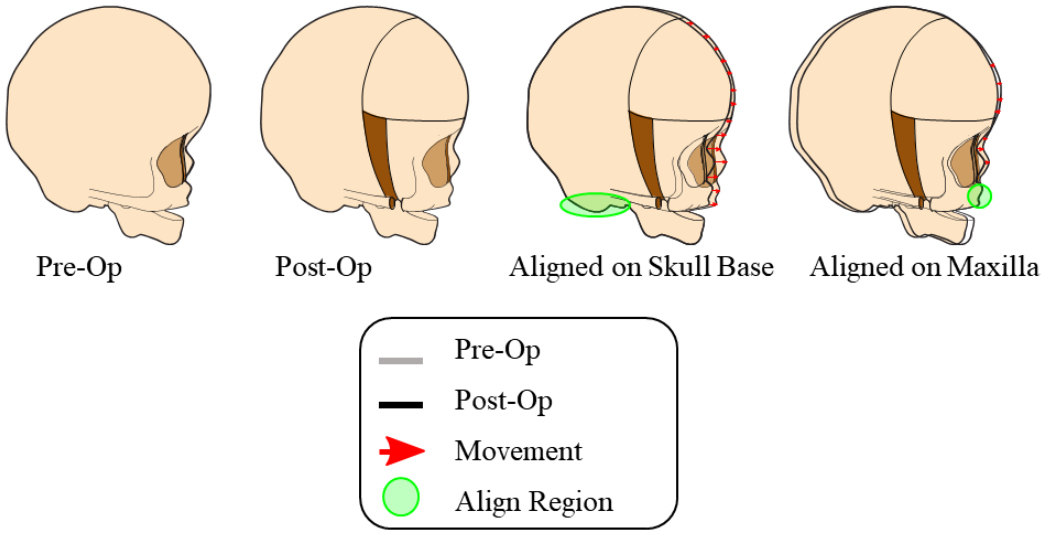


Figure 7.1. Illustration on rigid alignment. This illustration demonstrates the rigid alignments used in this study. Whereas with alignment on the skull base the facial bipartition distraction itself can be analysed (which has been done in prior studies), for this study the subjects were aligned on the maxilla (third image) to allow for evaluation of the local movements resulting from the procedure.

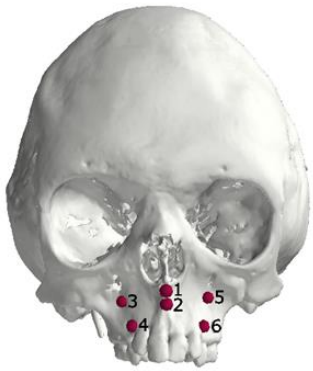


Figure 7.2 Landmarks used for rigid alignment on the maxilla. Six landmarks were used to superimpose the post-operative mesh on the pre-operative bone mesh, the soft tissue moves along as one entity.

Following rigid transformation, the meshes were landmarked to guide the NICP to obtain anatomical correspondence using the 16 landmarks presented in Figure 7.3.The corresponding soft tissue mesh followed the same non-linear transformation as the bone mesh.

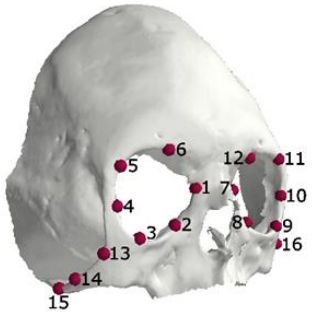


Figure 7.3 Landmarks used to guide the NICP. Sixteen landmarks were used to guide the NCIP registration. The soft tissue mesh followed with the same vector as the bone mesh.

Anatomical point-to-point distances were quantified and visualised using an arrow-map for the bone and a heatmap for the soft tissue. An arrow map was used to demonstrate the anatomical point-to-point movements of the bone as a result of the surgery in terms of local movements – as mentioned, the rigid displacements of the midface were removed from the analysis - where the arrows reflect the vector of the local differences from pre- to post-operative 3D Mesh. A heatmap was used to illustrate the absolute soft tissue displacements following surgery. Changes of less than 2 millimetres (mm) were considered negligible and are shown in green (Figure 7.5). As the nasal bone graft was not present in pre-operative skeletal 3D meshes, anatomical dense correspondence could not be achieved in this area, and the bone graft was therefore omitted from the analysis.

Measurements for displacements were automatically measured for bone and soft tissue according to the landmarks places as visualised in Figure 7.4.

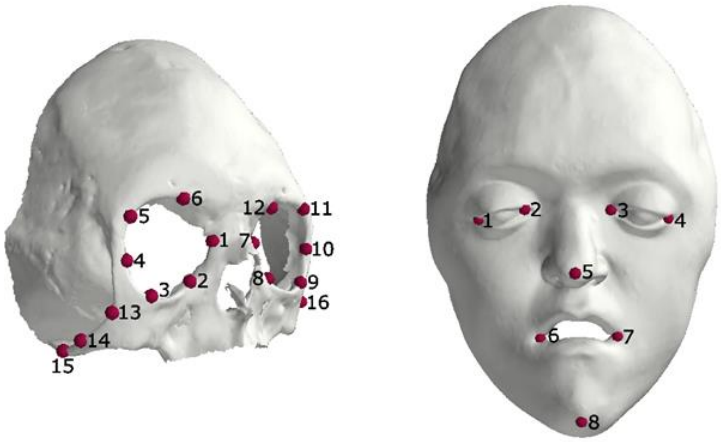


Figure 7.4. Landmarks used for intrinsic changes for bone (L) and soft tissue (R). Sixteen landmarks were used for bone and eight to determine the local changes.

7.3. Results

From a total of 23 patients that underwent FB-RED between 2005 and 2019, 9 patients met the inclusion criteria and were included in the construction and validation of the methodology. Eight patients underwent surgery at 17 +/- 4 years of age. An additional patient aged 2 years old was also included. The median time between pre-operative CT scan and post-operative CT scan was 477 days (range, 108 – 764), the median time between pre-operative CT scan and FB-RED (at time of insertion RED-frame), and FB-RED and post-operative CT scan was 158 days (range, 39 -566) and 198 days (range, 32 – 470), respectively (Table 7.1.).

Table 7.1. Study population FB-RED
SD = standard deviation, FB-RED = facial bipartition distraction with RED-frame, + = with, - = without

Study case	Gender	Age at time of FB-RED, years	FB-RED +/- nasal bone graft	No. of days between pre-operative CT scan and MBD-RED	No. of days between FB-RED and post-operative CT scan	No. of days between pre- and post-operative CT scan
1	Female	21	+	76	32	108
2	Female	23	-	158	56	214
3	Male	17	+	40	470	510
4	Female	13	-	223	421	644
5	Male	17	+	483	169	652
6	Female	12	+	132	345	477
7	Male	18	+	566	198	764
8	Female	13	+	389	56	445
9	Female	2	-	39	295	334
Total	**F:M**	**Median Age, years (age range, SD)**	**Nose bone graft, %**	**Median time, days (range, SD)**	**Median time, days (range, SD)**	**Median time, days (range, SD)**
	6:3	17 (2 – 23, 6.2)	5/9 (55.6%)	158 (39 – 566, 198)	198 (32 – 470, 164)	477 (108 – 764, 214)

7.3.1. Local movements

Skeletal movements

The mean absolute displacements for the 18 landmarks for each patient were calculated as 3.3 mm with SD 2.1 mm for the mildest case (patient 1), 6.9 mm with SD, 3.4 mm for the average case (patient 7) and 12.8 mm with SD 7.3 mm for the most severe case (patient 2). Landmarks 4,5,6 and 10,11,12 (right/left upper lateral orbital rim, superior orbital rim, and upper medial orbital rim, respectively) demonstrate the largest point-to-point movements. The zygomatic arches also underwent large point-to-point displacements with repositioning up to 17.8 mm observed at landmark 18 for patient 2 (Table 7.2). The local bone movements (after excluding the rigid facial advancements) show an overall upward and inward motion of the

supraorbital rim, and a similar upward rotation of the zygoma. Overall, the arrows report an inward motion of the frontal bone, with displacements ranging from mild to severe over the cohort patients. Patient 2, the most severe case of the study-population, underwent the largest local movements from pre-to-post-operative state (Figure 7.5).

Soft tissue deformations in response to the skeletal movements

The mildest changes were seen in patient 1 with displacements of 2.5 mm and 4.6 mm for the right and left lateral canthus; 1.7 mm and 1.4 mm for the right and left medial canthi, respectively, were observed. For the average case, patient 7, this was calculated as 8.3 mm and 8.0 mm, and 11.0 mm and 4.6 mm of displacement for left and right lateral canthi, and left and right medial canthi, respectively. As with the bone movements, the largest displacements were again observed for patient 2.

Although some of the largest local bone movements are seen in the zygomatic arch, the heatmaps show minor changes for the soft tissue in this region (Figure 7.5.) Thus, relatively large bone movements in the zygomatic arch do not directly translate to an equivalent change in the corresponding soft tissue. Figure 7.5 also demonstrates that certain areas, such as the cheekbone and peri-ocular region seem to translate in a different ratio from bone to soft tissue movement than other areas, such as the forehead, where a smaller ratio from bone to soft tissue movement is suggested.

For the most part, the largest soft tissue changes were observed at landmarks 1 and 4, i.e. the right and left lateral canthi, closely followed by the corresponding medial canthi. A larger degree of change was typically observed in the frontal bone; this is particular evident with patient 2, the most severe case in the study cohort.

Table 7.2. Overview bone landmarks and distance calculations. RLC = Right Lateral Cantus, RMC = Right Medial Cantus, LMC = Left Medial Cantus, LLC = Left Lateral Cantus, Nose = Nose Tip, RM = Right Mouth Corner, LM = Left Mouth Corner. SD = Standard Deviation mm = millimetre.

Landmark	Semantic	Distance (mm)								
		1	2	3	4	5	6	7	8	9
1	RLC	2.5	16.1	9.4	11.4	17.9	11.5	8.3	9.5	7.8
2	RMC	1.7	9.1	6.9	6.5	12.3	7.5	11.0	8.5	7.9
3	LMC	1.4	17.2	7.2	4.9	7.8	14.1	4.6	8.1	7.3
4	LLC	4.6	23.3	6.5	8.4	12.4	11.7	8.0	10.8	10.8
5	Nose	5.3	8.1	7.2	9.2	7.4	2.3	2.7	1.9	1.4
6	RM	4.6	3.8	8.6	6.2	11.5	4.1	4.4	2.9	0.9
7	LM	4.3	10.8	9.1	3.4	8.4	6.4	2.5	2.0	2.9
8	Chin	8.4	10.1	8.4	9.5	14.9	9.2	6.8	3.6	6.8
Mean		4.1	12.3	7.9	7.4	11.6	8.3	6.0	5.9	5.7
SD		2.3	6.2	1.1	2.7	3.7	4.0	3.0	3.7	3.6

Table 7.3. Overview soft tissue landmarks and distance calculations. RO A-F= Right Orbit A-F; LO A-F = Left Orbit A-F, RZ A-C =Right Zygoma A-C, LZ A-B = Left Zygoma A-C. SD =Standard Deviation; mm = millimetres.

Landmark	Semantic	Distance (mm) per patient								
		1	2	3	4	5	6	7	8	9
1	RO A	1.4	4.5	5.3	5.5	6.5	4.7	4.9	2.9	3.1
2	RO B	0.6	3.0	3.8	3.1	7.2	4.6	3.0	1.0	2.5
3	RO C	1.3	7.9	6.0	7.6	10.0	6.2	4.8	2.8	2.1
4	RO D	1.3	7.2	9.2	9.2	13.4	7.6	5.9	5.6	3.7
5	RO E	2.8	15.0	8.7	7.4	12.3	4.3	8.1	5.2	4.6
6	RO F	3.1	21.8	7.2	8.7	9.2	7.5	10.6	2.9	2.2
7	LO A	3.4	10.2	3.5	5.3	5.5	4.2	4.1	3.7	2.7
8	LO B	1.5	8.6	3.1	4.5	3.4	4.0	2.4	1.2	2.5
9	LO C	3.2	15.8	3.9	6.6	3.5	6.0	2.9	4.4	5.9
10	LO D	3.1	19.7	6.4	6.6	5.3	7.9	5.7	5.5	7.8
11	LO E	3.0	27.8	5.8	6.4	5.7	7.6	8.9	6.9	5.0
12	LO F	3.5	24.5	9.9	5.0	5.7	4.3	8.8	5.9	4.2
13	RZ A	1.6	0.8	6.4	7.8	12.2	4.6	3.6	3.4	5.5
14	RZ B	4.6	4.3	4.5	9.1	12.0	5.5	5.3	2.9	8.4
15	RZ C	7.0	6.6	7.9	11.2	9.1	10.1	13.8	5.1	10.5
16	LZ A	3.4	17.2	3.4	7.2	3.3	6.7	6.2	3.4	10.6
17	LZ B	7.5	17.6	6.2	7.3	6.1	6.5	12.8	1.8	13.9
18	LZ C	7.2	17.8	10.8	10.5	11.4	9.8	12.6	6.7	15.5
Mean		3.3	12.8	6.2	7.2	7.9	6.2	6.9	4.0	6.2
SD		2.1	7.3	2.2	2.2	3.8	2.8	3.4	2.6	3.9

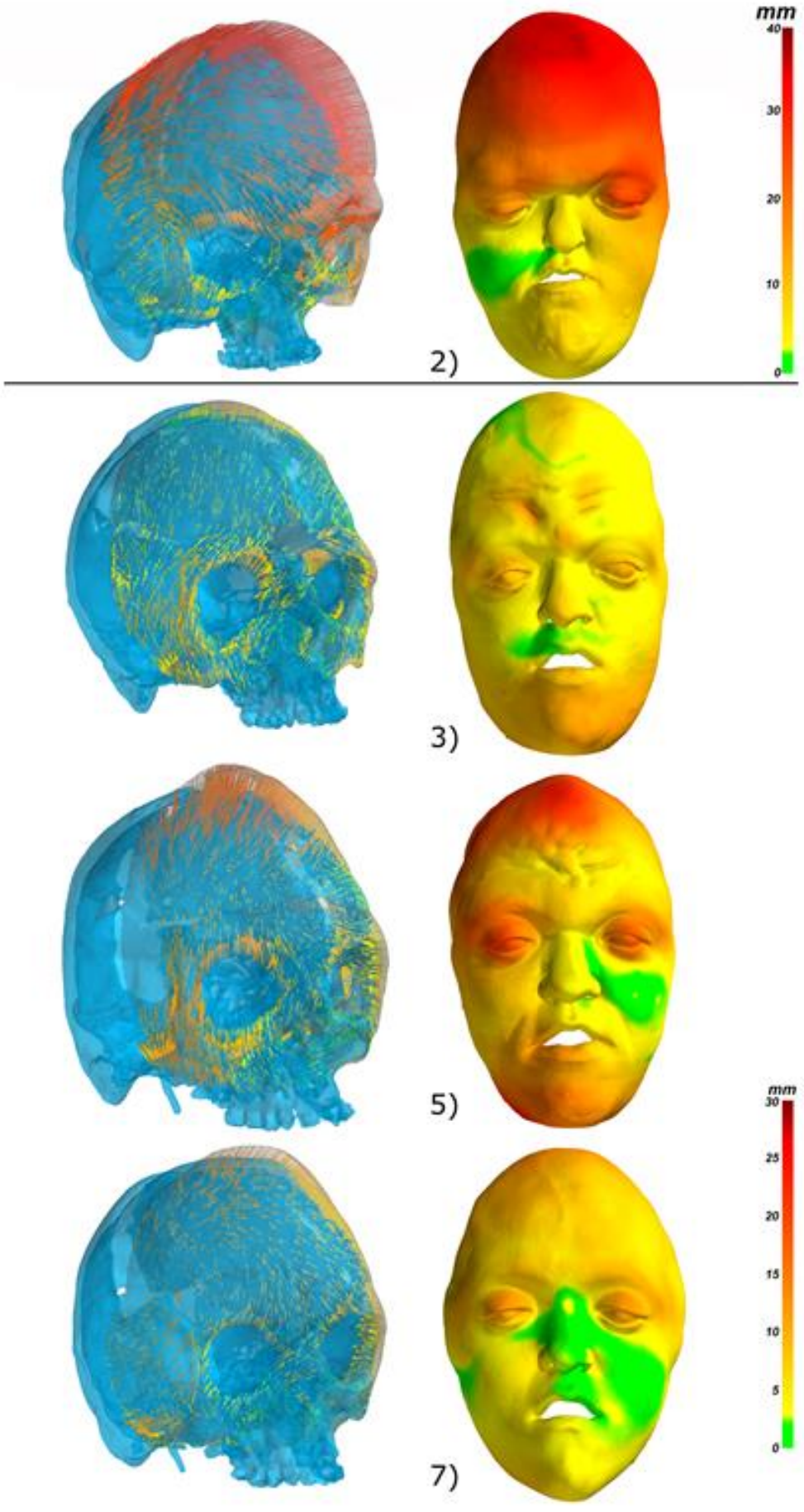

mm
40
30
20
10
0
2)
3)
5)
mm
30
25
20
15
10
5
0
7)

Figure 7.5. Heat and arrow maps of local bone and soft tissue changes following facial bipartition distraction. Four cases are presented here with various degrees of deformation severity. **Patient 2,** the most severe patient of the dataset is illustrated at the top with a personalised legend for the colours representing the distances. **Patient 3** demonstrates mild local changes and a homogenous change of the upper face **Patient 5** visualised relatively large posterior movement of the frontal bone /forehead and obvious changes of the periocular/peri-orbital region. Notably, is the asymmetric correction of the cheek on the right side, which has been advanced less than the patients' left side. **Patient 7** is the average case of the study set, demonstrating a unilateral movement of the right cheek with moderate changes of both eyelids, relative backward movement of the frontal bone is seen with remaining prominent supra-orbital rims. Notably, is the asymmetric correction of the cheek on the right side, which has been advanced less than the patient's left side.

7.4. Discussion

A semi-automated workflow was developed for the quantification of bone and soft tissue changes for craniofacial procedures. The constructed pipeline was applied to a dataset of 9 Apert patients who underwent facial bipartition distraction with RED-frame. To quantify the intrinsic bone changes and corresponding soft tissue motion, the overall forward displacements achieved by the surgery and RED-frame distraction were omitted from the analysis. This was achieved by aligning the 3D bone meshes rigidly on the maxilla and applying the same transformation to the soft tissue so that both moved together. The maxilla therefore acted as a reference for the regions of interest for quantification of the local movement, i.e. the zygomatic arches, nasal bone, orbits, and frontal bone. After removal of the facial advancement, the largest deformations were observed in the upper lateral to medial orbital rim on the skeletal level which coincides with the aim of FB-RED; namely to correct the hypertelorism and the counterrotated orbits seen in Apert syndrome. This methodology provides information on the vectors of the skeletal differential movements, visualised in the arrow map. Due to different positioning of the mandible at time of pre- and post-operative CT, this study did not describe displacements in this region. For two cases, clear asymmetric advancement was seen for the maxilla. One reason could be a registration issue, however this was not noted in the data processing phase, another reason could be that one part of the maxilla was advanced more than the other side,

which is more likely the case. This study did not analyse if this asymmetric movement resulted in a more symmetric outcome, i.e. correcting facial asymmetry by advancing one half of the maxilla more than the other half. However, the next chapter will dive further into normalisation analyses.

The relationship between bone and soft tissue movement varied. This might be due to various degree of decoupling of soft tissue as a result of dissection, ancillary procedures such as lateral canthopexy and midfacial suspension causing different shape changes post-operatively for bone as for soft tissue.

The time between pre- and post-operative CT scans varied from 108-764 days, and factors such as growth may therefore have played a role in the final measurements and point-to-point correspondences described in Table 7.2. At *GOSH*, patients currently undergo a standardised pre-operative and follow-up CT scan protocol, consisting of a CT scan module with high-quality imaging especially for craniofacial patients pre-operatively, one week post-operatively and only in case of clinical concerns, a subsequent CT scan as the aim is to limit radiation as much as possible in children. Many patients underwent FB-RED prior to introduction of this standardised protocol and therefore did not have ideal CT scans timings for the analysis of the bone and soft tissue movements. Others could not be included in this study due to the limited quality of the CT scan for 3D reconstruction. Henceforth, this study could only include a relatively small number of FB-RED cases as compared to the numbers that were in fact performed at *GOSH*. The several patients that did meet the quality and inclusion criteria were less preferable in terms of CT scan timings. The fact that surgery is performed at different ages and post-operative scans were performed at different intervals was a real challenge and from the point of view for this study has significantly downgraded the quality of the data. It is however a result of the *GOSH* philosophy on (fronto/mid)facial advancement surgery. Midfacial advancement is performed only when there is a clinical need or the expectation of

need for functional or psychological reasons rather than as part of an age dependant reconstructive protocol. Similarly, late post-operative CT scans are only undertaken to answer clinical questions in situations where they could alter management.

Literature suggests that the mean facial growth spurt corresponds to an age of 14.35 years for boys and 11.52 years for girls (136). The majority of the subjects included in this study population were beyond this growth peak and it is therefore unlikely that growth on its own played a large factor in the observed deformations. One case included in the analysis, however, was 2 years old at the time of FB-RED and had 334 days between their pre- and post-operative CT scans. The deformations quantified for this young case were inevitably not only due to the surgery but also attributed to skull growth during this period. As the new CT-protocol is now in place, future studies could benefit from more timely quantification.

The proposed methodology uses the absolute distances (vectors) between corresponding points to display the deformation in a heatmap. This is different from the approach presented in a prior study where 3 heatmaps were used per patient to illustrate the movements in the x (medio-lateral), y (vertical) and z (anterio-posterior) directions (127). With the methodology proposed in this chapter, the bone (arrow map) and soft tissue (heat maps) can be jointly considered and obtain a more intuitive understanding of the overall displacements.

In addition to solving functional issues – being the indication for surgery mostly in youngsters - the surgery studied here aims to ‘normalise’ the patient’s appearance and address psychosocial problems deriving from cosmesis concerns. Optimisation of the surgical approach would require full understanding of the shape changes by objective analysis of the true surgical movements and by comparison of the post-operative outcomes with a normal match reference. However, the ‘normal face’ that one might aim to achieve with surgery is not unique for a given individual, but falls within a wide range of acceptable shape features as demonstrated well by

LSFM (5, 21). Indeed, Knoops *et al.* 2019 described how LSFM can be used to describe the surgical outcome of pre- to post-operative state after Le Fort I surgery where the orthognathic patients were compared to an age matched group of normal peers (21). Using a similar approach, chapter 8 will analyse the effect of FB-RED in relation to an age-matched group of healthy peers. The combination of a 3D understanding of the local tissue and bone movements provided by the presented methods, with an understanding of how results compare to the healthy LSFM population can potentially play a role in a personalised and quantified treatment approach.

The flexibility of the presented methodologies lends itself to applications to other types of craniofacial surgery and could be used to facilitate real-time monitoring of a patient undergoing RED-frame distractive surgery at time of distraction. This approach could also be applied to other surgical techniques other than FB-RED, as it can provide additional information about how the bone is moving in response to the RED-frame and allow for the opportunity to simultaneously adjust the distraction protocol accordingly. Moreover, in case sequential post-operative imaging is available, this methodology could be used to determine post-operative relapse in terms of the quantity and direction of the relapse. This information could, in turn, be used for surgical decision making and planning, and might influence the direction or quantity of distraction.

Ultimately, this methodology can give us accurate insight into the localised changes resulting from surgery regarding hard tissue and soft tissue. Although it does not give direct answers on how to improve surgery, it provides information to understand how shape changes have occurred which may lead to alterations in technique. Some initial steps on this are described in the next chapter.

7.5. Summary

Local changes of the bone and soft tissue after facial bipartition distraction in Apert syndrome were evaluated in a semi-automated fashion and are characterised by an upward inward rotation of the orbits, upward rotation of the zygoma and relative posterior motion of the frontal bone. The largest soft tissue movements are seen at the lateral canthi. No hard conclusions can be drawn given the varied cohort.

Key points

- A semi-automated workflow to determine bone and soft tissue changes following FB-RED was introduced and can be readily applied to other craniofacial abnormalities and surgeries.
- Local changes are seen by means of upward inward rotation of the orbits, upward rotation of the zygoma and relative posterior motion of the frontal bone.
- Relatively large movements of the zygoma result in relatively small changes of the soft tissue.

8. CORRECTED BUT NORMALISED?

In chapter 3, the main differences of the Apert's unoperated face from the normal population face were highlighted, showing clear clustering between these two study populations. Chapter 7 analysed postsurgical outcomes and evaluated local face changes resulting from surgery in a 3D fashion. The next chapter will focus on evaluating the surgical results in comparison to the normal population, aiming to understand if surgery normalises the face and if not, in which areas this technique might be lacking.

8.1. Introduction

At *GOSH*, the correction of the Apert's face is commonly addressed with a facial bipartition distraction as described in chapter 7. To my knowledge, no in-depth studies exist on patient's aesthetical satisfaction, however, from clinical experience most patients and indeed most caregivers' mention to be pleased with the improvements. One study by Tovertjärn *et al.* 2012 followed up on 28 adult Apert patients and reported significantly lower education, social life including friendships, marital state, and sexual relationships. They stated that this could be the result of an abnormal appearance, however, could also be due to less practiced social skills because of the time-consuming treatments of these patients in childhood and adolescence (137). Moreover, to the best of my knowledge no studies exist on objective 3D normalisation analysis for soft tissue. Two studies mention evaluation of the bone abnormalities pre- and post FB-RED, using linear measurements and geometric analysis by means of rigid ICP (90, 138). Both studies were performed at *GOSH*. A study by Crombag *et al.* 2014 demonstrated inward rotation of the orbits and reduction of the intraocular distance and a forwardly moved maxilla (90). When compared to normative CT scans remaining deformities were seen, mainly the orbits seemed to have remained in the same position. Similar results were seen in a study by Glass *et al.* 2018 that concluded a partially correction of the midfacial width and correction of

the interorbital distance and inadequate correction of upper midfacial height (138). Further improvements were suggested such as a combination of facial bipartition with a LF2ZR, which would however increase the complexity of the surgery to an unacceptable degree and diminish stability and control. It was suggested that early manipulation of the skull base may address some of these issues. In their study the following was proposed: 1) bipartition distraction to correct midfacial retrusion and 'unbend the face', 2) further bony surgery to correct the occlusion and facial height, width, and asymmetry; a. Le Fort I or bimaxillary surgery, b. zygomatic osteotomies or onlays, and 3) final adjustments: a. rhinoplasty, b. soft tissue procedures. This suggested protocol would require a long-term surgical plan of many years with multiple surgeries.

FB-RED is a surgical procedure built on the monobloc distraction by adding a straightforward vertical osteotomy, which does not seem to increase operating time or morbidity (78). It is designed to correct hypertelorism, exorbitism, midfacial contour, and canthal positioning. This chapter aims to evaluate the Apert FB-RED's post-operative outcomes in comparison to normative data in order to understand if normalisation is achieved for soft tissue. It also aims to discuss the topics 'what we do well', 'what can we do better' and finally 'when is it good enough'.

8.2. Material and methodology

8.2.1. Data sources

The 23 patients that were initially considered for analyses in chapter 7 were addressed and those with pre- and post-operative scans available with sufficient quality for soft tissue mesh reconstruction were included. All DICOM-files from pre- and post-operative CT scans

underwent the pipeline described in chapter 3 (i.e. soft tissue mesh conversion, mesh clean-up, and annotation to guide NICP registration). Age matched meshes were selected from the LSFM dataset as perform as normative control group.

8.2.2. Normalisation experiments

Two types of experiments were performed:

- Mean face comparison of the pre-operative and post-operative face compared with a normal face.
- t-SNE embedding labelled for pre-operative, post-operative, and normal.

Mean face comparison

For the mean face comparison, three mean meshes were constructed:

1) A mean face mesh of the pre-operative Apert patients.
2) A mean face mesh of the post-operative Apert patient (following facial bipartition).
3) A mean face mesh of the LSFM dataset (aged matched with mean mesh 1) and 2)).

The mean pre-operative and the mean post-operative Apert face meshes underwent NICP registration with the mean normal face and heatmaps were computed allowing for visualisation of differences. This experiment aimed to give more insight into the facial areas where FB-RED was successful and areas where normalisation might be lacking.

Manifold visualisation

All included pre-operative and post-operative meshes were be labelled with 1) preop and 2) postop, and an equal number of normal meshes were selected and labelled with ‘normal’. t-SNE was applied to the high dimension latent vector encodings for the study population with the aim to evaluate if the post-operative subgroup would cluster within the normal population, thereby confirming normalisation following FB-RED. The t-SNE embedding was created using a perplexity of 30 and run for 1,000 iterations.

8.3. Results

Seventeen patients could be included for analysis. The patients had a mean age of 15.68 years (range, 12.18 – 20.44 years) at time of surgery, of those nine were female. Eight patients were indicated for surgery for solely aesthetic reasons, the remainder for (a combination of) functional problems. When indicated for aesthetic and functional concerns, these included respiratory concerns such as obstructive sleep apnoea and proptosis.

Mean face comparison illustrates differences up to 15 mm (dark red coloured) for the peri-ocular region and to 10 mm for the premaxilla. The white regions presented in the figures that reach 2 mm differences are considered clinically neglectable and include the nasal tip and nasal labial folds; they are therefore comparable to the mean normal pre-operative. Post-operatively, increase in abnormality is seen for these areas, reporting differences of 3-4 mm for the nasal tip and nasal labial folds, as well as differences up to 6 mm for the glabella, which was closer to mean normal before any treatment. The (pre)maxilla and the eyelid positions improved. Pre-operatively mean differences of 8-10 mm can be noted whereas post-operatively this was reduced to 2-4 mm. Significant improvements were observed for the medial canthi, while a lesser degree of normalisation was seen for the lateral canthi. The lateral aspect of the eyebrows

improved whereas medially worsening is apparent. The glabella and forehead were also further from the normal as compared to pre-operative analyses.

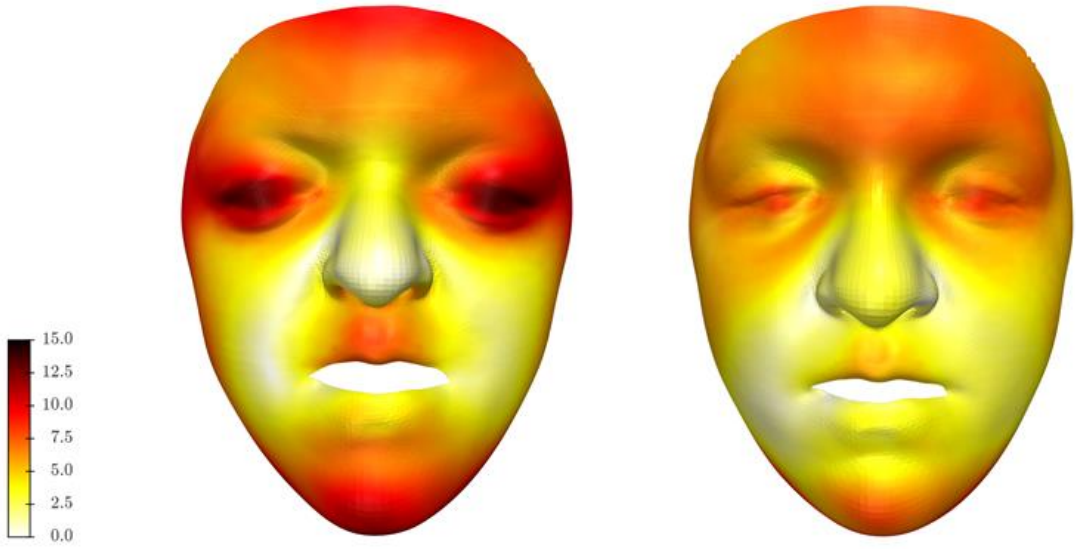


Figure 8.1. Heatmaps of Apert pre and postoperative means compared with mean normal. Left mean Apert pre-operative mesh superimposed with normative mean. Main differences are seen for the upper lip, periorbital region and the forehead. **Right** post-operative mean Apert face superimposed with normal mean. Main improvements are seen for the medial canthi, upper lip, and lateral aspect of the eyebrows. Worsening is seen for the nasal tip, forehead, and medial aspect of the eyebrow.

Figure 8.1 demonstrates that FB-RED does contribute towards normalisation based on soft tissue evaluation. Yet, after applying the t-SNE embedding, none of the post-operative cases was clustered within the normal population rejecting a hypothesis that FB-RED might achieve objective normalisation of the Apert's face (Figure 8.2).

Figure 8.2. t-SNE embedding of the high-dimensional manifold clustering for post-operative analysis. This plot demonstrates that post-operative Apert meshes (**black labels**) are not embedded within the normal population (**red labels**), the post-operative group however, clusters close to the pre-operative group (**green labels**) and seems to subgroup within the whole.

8.4. Discussion

FB-RED aims to normalise the face within existing anatomical restrictions. This chapter evaluated facial normalisation in a most objective manner and attempted to understand where this surgery is successful from a facial normalisation point of view.

Based on t-SNE clustering, it can be stated that FB-RED does not normalise the face from a shape analysis perspective. It also indicates that there is still a long way to go in order to achieve normalisation. An important question might be, if human perception for normalisation

assesses the same things as a statistical model does. Does human perception and geometric morphometrics agree on what a 'good' result is? Our perception might prioritise features such as facial symmetry and appearance of the eyes over other facial features, whereas the model might pick up more clear differences of the prominence of the brows and nose. To answer this question a study on human perception and geometric facial normalisation should be performed.

This study looked at soft tissue normalisation outcomes, and did therefore not analyse the type and size of the nasal bone graft that some might have had at time of FB-RED. The use of a nasal bone graft at *GOSH* is a decision that is usually made on table peri-operatively when a stronger facial profile is subjectively indicated by the operating surgeon. The size and position of insertion of this bone graft (which is generally not objectified prior to insertion) will have contributed to the shape variations seen for the nose on soft tissue evaluation. Thus, hard conclusions can be drawn on the variations of the nose postoperatively. It would be interesting to compare those that did not receive a bone graft with those that did to understand whether this contributes to normalisation.

In this chapter a significant larger number could be included for analysis than the study population presented in chapter 7 as usable soft tissue meshes can be constructed from lower quality meshes than those needed for bone meshes. Ideally, bone and soft tissue are analysed jointly, as was performed in chapter 7, enabling detailed analysis on the bone movements and its translation to soft tissue changes. The reasons why datasets with high quality data and 'right' timings for adequate assessment are often quite small were outlined in chapter 7.

Apert's facial deformities are challenging to treat due to the complex and commonly severe phenotype. As presented in the results of this chapters, FB-RED improves some of the facial shape but at the expenses of others. A study by von Gernet *et al.* 2000 recognised that the outcomes of fronto-orbital advancement in S252W patients are more challenging to correct the

shape and had better results when correcting the P253R mutation (139). They suggested that genotype should be considered for individual planning of the surgical procedure to achieve the best treatment for the patient. Although at *GOSH* an anterior-first approach is avoided and a posterior expansion would be the first treatment of choice to increase the cranial vault, it is an interesting finding that one subtype seemed more challenging to correct over the other. In this thesis, I did not find any facial shape differences for the two subtypes, however, it is possible that this is due to lack of sufficient data for adequate analysis. A future, potentially collaborative study to increase the dataset, could allow for not only the analysis of various techniques used in the correction of Apert syndrome, but also to sub divide the groups in their genetic subtypes and ascertain whether this plays a role in surgical outcomes.

8.4.1. What are we doing well?

As found by Cromberg *et al.* 2014 and Glass *et al.* 2018 FB-RED advances the midface well, decreases the inter orbital distance and rotates the orbits inwards as they studied on a skeletal level (90, 138). Chapter 7 and this chapter showed that relative backward movements are seen for the forehead, resulting in a less obvious frontal bone protrusion and better location for the eyebrows, especially for the lateral aspect. It showed an improved location of the canthi and correction of exorbitism. Moreover, an improved positioning of the upper lip was quantified. However, none of these regions fully reached the mean matched normal face.

8.4.2. What can we do better?

This thesis identified the facial region that are on average not fully corrected, however it was not studied which areas are in fact most important to achieve facial normalisation

objectively and subjectively. Information on what regions to prioritise in the surgical approach might give a better idea on how to design improved operative approaches addressing the key deformations. Although the FB-RED surgically removes a wedge of bone between the orbits and rotation of the bone segments results in improvement of the temporal region and lateral aspect of the brow, the surgery seems to have a lesser impact on the shape of the forehead, especially at the level of the frontonasal suture. Albeit that the lateral aspect of the brows seems to have improved, prominent brows remain and are most likely the direct results of the protruding supraorbital rims.

Interestingly, in chapter 6 it was noted that some Apert patients grow into inter-orbital distances that are within normal limits, or even below. Thus, awaiting for surgery until full maturation might avoid unnecessary medialisation, or in case, orbital medialisation is indeed indicated, can provide information on the 'right' amount required to achieve normalisation. The models demonstrated that large distances remain between the lateral orbital rims and they clearly visualised the deformity of the infra-orbital walls that are not corrected by FB-RED. To address this, newer techniques, such as personalised lay-ons could potentially be a solution.

Based on the heterogeneity seen within the Apert population a personalised surgical planning technique should be in place enabling a customized surgical approach. Future approaches could include the design and use of customised and improved types of distractors with pre-planned vectors and distraction distances per facial region.

8.4.3. When is it good enough?

There is no doubt that complete normalisation of the Apert's face is extremely complex and might be an impossible task to achieve. The studies presented in this thesis did not include any evaluation on patients' or caregivers' satisfaction, the wishes, and goals they had pre-

operatively, and their main areas of concern. This thesis focused on objective analysis of the abnormal shapes and critically assessed the current surgical approach in place at *GOSH* for its correction. Unfortunately, the studies presented did not include any other types of surgical interventions nor patients'/caregivers' viewpoint. It could be of interest to include these parameters in any future studies and collaborative studies with other craniofacial institutions utilising for example LF2ZR as pioneered by Hopper and its group at Seattle Children's Hospital, United Stated of America, or the Monobloc advancement as done by *NEMH* to truly assess whether other approaches might be more successful in correcting particular areas of the face. Clinical reconsiderations for the surgical technique of choice to correct the facial deformities might adept based on such objective analysis.

As technology becomes increasingly part of daily clinical practice, the voice of the patients should too. Indeed, psychologists are getting a more prominent place in the pre-operative work up for craniofacial patients wanting to undergo aesthetic facial corrective surgery. They play a significant role in psychosocial assessment to understand whether the patient's needs, and wishes can be reached by a surgical treatment. Unfortunately, no real-time surgical decision-making format exists which allows for the visualisation of outcomes of specific types of surgeries. For the patient to see how they might look like after surgery at a preoperative consultation would highly improve the decision-making process. Not included in these studies are complications occurring after surgery. These should also be part of this decision-making process so that a well-informed shared decision can be made which is tailored to every unique patient. Although this thesis mainly described mean faces and its variations from the Apert population, the end goal is to use this technology for a personalised care management approach by implementing patient specific parameters. Taking psychosocial factors into account, is perhaps nowadays more important than ever as young adults grow up in an era where they are exposed not only to social pressure at school or outside but are also

constantly exposed to social media and internet, which has shown to have an enormous impact on mental health (140).

8.5. Summary

Objective facial normalisation is not achieved with FB-RED described by the methodologies presented in this work. Improvement towards normalisation is noted and areas that might require surgical adjustments – where technically possible – were described.

Key points

- Facial bipartition changes the face of Apert syndrome, however, does not achieve normalisation of the face.
- It is unknown whether geometrical normalisations prioritise the same regions as human perception.
- There is need for patient specific surgical planning.

9. CONCLUSIONS

This chapter summarises the main findings and outcomes of this thesis, drawing from all previous chapters, and describes in detail how the application of computer vision has contributed to achieve the aim of this thesis. The limitations are discussed, as are directions for future research and final remarks.

9.1. Overview

The aim of this thesis was to apply 3D methodologies for automatic identification of Apert syndrome from genetically and phenotypically similar syndromes, and, once diagnosed, to evaluate the growing Apert's face and its facial skeletal abnormalities, and pre- and post-surgical correction with an ultimate aim of improving current practice. The first objective was to automatically identify Apert syndrome from other FGFR-related craniosynostosis syndromes and from the normal population using state-of-the-art algorithms. The second objective was to quantify shape development in the growing Apert's face and its facial skeletal abnormalities. The third objective was to define and apply the optimal quantification method to assess surgical outcomes in Apert's corrective facial surgery. The final objective was to evaluate facial normalisation after corrective surgery.

9.2. Detailed outcomes

Chapter 3 – Apert facial shape analysis

3DMMs of the Apert's face and one of a paediatric healthy population were constructed using state-of-the-art algorithms and tested for the application of automated diagnosis by employing embedding experiments. Excellent clustering was noted for the Apert's face from

the normal implying that this approach can have valuable clinical applications for automated syndrome identification. Moreover, a mean analysis was performed on two age sub-populations which demonstrated that the Apert facial abnormalities worsen with age. To summarise:

1. Successful construction of Apert's face 3DMM and healthy paediatric population was noted which might have potential for automated syndrome identification.
2. Apert's facial deformities increase with age when assessing the mean faces, however, need more detailed analysis on how the natural shape develops over time.

Chapter 4 – Automated syndrome identification

In chapter 4, the use of computer vision for automated identification was further explored for genetically and phenotypically similar syndromes to Apert. Head shape information was added and with the use of convolutional autoencoders 3 types of models were constructed, face, head, and face+head. Clustering experiments outperformed clinical observation and it was demonstrated that although head shape information is useful information, the model performs best for face shape only. Given sufficient data, this framework could be readily extended to a greater variety of craniofacial syndromes. Extending the model to larger patient cohorts and a greater number of syndromes may lead to new diagnostic tools, facilitating low-cost analysis and identification of craniofacial disorders. This could benefit in the subsequent management of such conditions, at an earlier stage. The main outcome of this study was:

3. Using the described framework, FGFR-related craniofacial syndromes can be automatically identified from one another and from the normal.

4. Face shape information outperformed face+head models, or head models for the given dataset. Although all models presented had excellent classification results, face shape information suffices for automated identification.

Chapter 5 – a skull morphable model

In chapter 5, following construction of soft tissue 3DMMs, a paediatric 3D skull model <4 years of age was constructed and validated with gold standard measurements from literature. In addition to the understanding of shape change and the information this model can provide, it also has applications for in silico surgical planning, post-operative analysis, and could become part of a more comprehensive model in the future. As all meshes are in dense correspondence, selected distances can easily be calculated from raw data, providing a large amount of normative growth curves for the facial skeleton. In summary:

5. A normative skull model of an <4 population was constructed, and its applications were outlined.
6. True and generated data can provide large normative measurements and growth curves which might benefit a clinician for understanding the natural course of development and for surgical planning.

Chapter 6 – an Apert skull model

Chapter 6 described the construction of an Apert skeletal model and presented the mean and eigenvectors for two bespoke models (<4 and all ages). Growth curves were constructed and demonstrated the main differences of the Apert's skull from the normal for various ages. In conclusion:

7. An Apert skull 3DMM was successfully constructed.
8. Growth charts were provided with comparison of normative data.
9. Most measured parameters deviate from the normal growth curves.
10. Potential future applications were outlined.

Chapter 7 – facial surgical outcomes

Chapter 7 showed the local changes made by facial bipartition distraction using a semi-automated approach. These changes are characterised by an upward inward rotation of the orbits, upward rotation of the zygoma and relative posterior motion of the frontal bone. The largest soft tissue movements were seen at the lateral canthi. The method to analyse these changes was demonstrated useful and can be applied to other types of surgical outcome analysis. To summarise the findings in this chapter:

11. Anatomical point-to-point correspondence (NICP) is a useful approach to analyse surgical outcomes in 3D.
12. The relationship of movement of bone relocation to the soft tissue changes were mostly noted for the zygomatic arch, where relatively larger movements of the zygoma result in small changes of the soft tissue.

Chapter 8 – corrected but normalised?

Chapter 8 objectively analysed facial normalisation following facial bipartition distraction. The mean faces of preoperative and postoperative Apert patients were compared with the mean face of normative data and demonstrated that the surgery improves but does not

normalise the face fully. Improvement towards normalisation is noted and areas that might require surgical adjustments – where technically possible – were described.

13. Objective facial normalisation is not achieved with FB-RED described by the methodologies presented in this work.

9.3. Limitations and future directions

9.3.1. Sample size and data

One of the major limitations of this thesis is the relatively small sample sizes in the syndromic databases. In any single center study of rare syndromes, small sample sizes are a limitation. This was experienced when using the *GOSH* data for chapter 3, 7 and 8. Chapter 6 included data from *NEMH* as well, which enlarged the dataset significantly. To provide further evidence for the findings in this thesis would require collaboration with multiple large craniofacial units, ideally where sequential imaging is routinely taken, including CT scans and 3D photography. Prospective large-cohort studies and overall standardization of data acquisition protocols would enable for larger studies and more homogenous studies. This would necessitate large scale sharing of information across international borders, which holds data sharing problems and was a in particular an issue during COVID-19 times.

An added difficulty to the rarity of the syndromes and therefore small sample sized studies is the reliance on data produced by ionising CT scans and the understandably conservative imaging protocols in place at most craniofacial institutions. Exposure to ionising radiation is kept to a minimum with CT scans taken pre-operatively to aid surgical assessment

and planning, but at *GOSH*, follow-up CT scans are often not taken unless there are clinical pressing reasons, such as complications or pre-operative assessment for further surgery. The lack of follow up imaging or unacceptable time lag between operative procedure and post-operative imaging led to the exclusion of a large number of patients. Low-quality imaging that was available for a part of the study population was another reason for exclusion. The advent of ultra-low dose 3D CT scanning and black bone MRI protocols could ameliorate these problems (141). However, the problem of repeated anaesthesia, which is usually needed for infants undergoing scanning, in the developing child remains. 3D photography might be the answer for soft tissue analysis, yet in silico pre-operative planning would still require 3D reconstructions derived from scanning slices that captured the bone.

9.3.2. Impact of COVID-19

Though most of this work could be performed digitally, COVID-19 did have an impact to what has been presented in this thesis. From a positive perspective, it encouraged for more frequent online meetings with the research team, yet international collaborations were restricted and data transfer between centers was impeded. In a field where one works with patient sensitive data, new collaborative studies could not be performed. Collaborative studies on for example comparing other types of surgical approached or adding more data to existing datasets would have strengthened the studies presented and allowed for further evaluation, potentially providing more answers to the questions proposed in this thesis. I am hoping to maintain the strong collaborative research teams that are in place and the reintroduction of onsite collaborative work as we were able to do prior to COVID-19 times.

9.3.3. Future directions

Combined model

A certain valuable next step would include the construction of a combined model. For the <4 group, bone and soft tissue models are built and readily available. Merging of these models would allow for a more intuitive evaluation of abnormality and in silico surgical planning. It would require expanding the included age group into adulthood and by adding in the mandible, which are studies that have been initiated. However, as described in the thesis and detailed in (38), combining models from different regions, various imaging modalities comes with many challenges.

Sub analysis of th

Surgical outcome prediction

We can learn from what has already happened. A preliminary study by our research group under the lead of Dr. Athanasios Papaioannou has demonstrated that retrospective data can be used to predict surgical outcomes (142). Using a machine learning methodology as described in (142) various types of surgical techniques could be analysed and various potential outcomes visualised which might aid the shared decision-making process. These projects are very much in the beginning stage, however, seem promising for future analysis.

Prenatal analysis

Dalll'Asta *et al 2017*, performed a quantitative analysis of fetal facial morphology using 3D ultrasound, and demonstrated potential for the use of face shape information to

enhance pre-natal diagnosis (108). A study between *GOSH* and *Boston Children's Hospital, Boston, Unites States* was commenced in 2020 to use prenatal MRI scans to understand the foetal face shape development in healthy and in syndromic patients, which could potentially enhance prenatal diagnostics. Due to COVID-19 restrictions, this study is currently on hold.

Model sharing

The scientific and clinical community might benefit from access to large normative databases with generative ability. Difficulties to access normative data is not uncommon and is an avoidable problem when forces are joint. A list of currently available models for scientific use are outlined in (38) and the collaborative use of these might enable more and better research.

9.4. Conclusions

This thesis presented successful approaches in which the use of modern technology aids to answer clinical questions. State-of-the-art algorithms were applied to evaluate automated syndrome identification and expanded its scope to skeletal models. The generative abilities of the presented models were proven, and various growth charts were presented, with the option to construct any type of measurements of the facial skeleton on demand by selecting facial landmarks on a template mesh. Using a modern approach, the post-surgical outcomes were analysed, and clinically valuable information was presented on facial normalisation analysis. Future lines of research were suggested, including the merging of the skull and soft tissue model for in silico operative planning purposes. While this thesis was focused on Apert

syndrome, the methodologies presented can be applied to any type of craniofacial syndrome and/or analysis of any surgical approach.

To conclude, this thesis has introduced the use of computer vision methodologies to enable automated syndromic identification, with the aim to enhance and speed up diagnosis. It provided large normative data and growth curves on the growing Apert's face, it also proposed a 3D methodology for accurate soft tissue and bone analysis and has presented some of the many potential applications that might positively change care management of craniofacial patients.

REFERENCES

1. Robertson, K. J., Mendez, B. M., Bruce, W. J., McDonnell, B. D., Chiodo, M. V., Patel, P. A. Le Fort III distraction with internal vs external distractors: a cephalometric analysis. *The Cleft Palate-Craniofacial Journal* 2018;55:721-727.
2. Hopper, R. A., Kapadia, H., Susarla, S. M. Le Fort II distraction with zygomatic repositioning: a technique for differential correction of midface hypoplasia. *Journal of Oral and Maxillofacial Surgery* 2018;76:2002. e2001-2002. e2014.
3. Waitzman, A. A., Posnick, J. C., Armstrong, D. C., Pron, G. E. Craniofacial skeletal measurements based on computed tomography: Part II. Normal values and growth trends. *The Cleft Palate-Craniofacial Journal* 1992;29:118-128.
4. Dai, H., Pears, N., Smith, W. A., Duncan, C. A 3d morphable model of craniofacial shape and texture variation. In *Proceedings of the IEEE International Conference on Computer Vision*2017.
5. Booth, J., Roussos, A., Zafeiriou, S., Ponniah, A., Dunaway, D. A 3d morphable model learnt from 10,000 faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*2016.
6. Apert, E. *Traité des maladies familiales et des maladies congénitales*: Baillière; 1907.
7. Cohen Jr, M. M., Kreiborg, S., Lammer, E. J., et al. Birth prevalence study of the Apert syndrome. *American journal of medical genetics* 1992;42:655-659.
8. Fearon, J. A. Treatment of the hands and feet in Apert syndrome: an evolution in management. *Plastic and reconstructive surgery* 2003;112:1-12.
9. Goriely, A., McVean, G. A., Röjmyr, M., Ingemarsson, B., Wilkie, A. O. Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. *Science* 2003;301:643-646.
10. Wilkie, A. O., Slaney, S. F., Oldridge, M., et al. Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome. *Nature genetics* 1995;9:165-172.
11. Marcus, L. F. Traditional morphometrics. In *Proceedings of the Michigan morphometrics workshop*1990.
12. Hutchison, B. L., Hutchison, L. A., Thompson, J. M., Mitchell, E. A. Quantification of plagiocephaly and brachycephaly in infants using a digital photographic technique. *The Cleft Palate-Craniofacial Journal* 2005;42:539-547.
13. Atmosukarto, I., Shapiro, L., Starr, J., et al. Three-dimensional head shape quantification for infants with and without deformational plagiocephaly. *The Cleft Palate-Craniofacial Journal* 2010;47:368-377.
14. Blanz, V., Vetter, T. A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*1999.
15. Ploumpis, S., Wang, H., Pears, N., Smith, W. A., Zafeiriou, S. Combining 3d morphable models: A large scale face-and-head model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*2019.
16. Amberg, M., Luthi, M., Vetter, T. Fully automated segmentation of the knee using local deformation-model fitting. In *MICCAI 2010 Workshop Medical Image Analysis for the Clinic—A Grand Challenge (SKI10)*2010.

17. Majeed, T., Fundana, K., Lüthi, M., Kiriyanthan, S., Beinemann, J., Cattin, P. C. Using a flexibility constrained 3D statistical shape model for robust MRF-based segmentation. In *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*2012.
18. Dai, H., Pears, N., Smith, W. A data-augmented 3d morphable model of the ear. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*2018.
19. Zolfaghari, R., Epain, N., Jin, C. T., Glaunes, J., Tew, A. Generating a morphable model of ears. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*2016.
20. Khamis, S., Taylor, J., Shotton, J., Keskin, C., Izadi, S., Fitzgibbon, A. Learning an efficient model of hand shape variation from depth images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*2015.
21. Knoops, P. G., Papaioannou, A., Borghi, A., et al. A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery. *Scientific reports* 2019;9:1-12.
22. Ranjan, A., Bolkart, T., Sanyal, S., Black, M. J. Generating 3D faces using convolutional mesh autoencoders. In *Proceedings of the European Conference on Computer Vision (ECCV)*2018.
23. Bouritsas, G., Bokhnyak, S., Ploumpis, S., Bronstein, M., Zafeiriou, S. Neural 3d morphable models: Spiral convolutional networks for 3d shape representation learning and generation. In *Proceedings of the IEEE International Conference on Computer Vision*2019.
24. Bochukova, E. G., Roscioli, T., Hedges, D. J., et al. Rare mutations of FGFR2 causing apert syndrome: identification of the first partial gene deletion, and an Alu element insertion from a new subfamily. *Human mutation* 2009;30:204-211.
25. Twigg, S. R., Wilkie, A. O. A genetic-pathophysiological framework for craniosynostosis. *The American Journal of Human Genetics* 2015;97:359-377.
26. Heuzé, Y., Singh, N., Basilico, C., Jabs, E. W., Holmes, G., Richtsmeier, J. T. Morphological comparison of the craniofacial phenotypes of mouse models expressing the Apert FGFR2 S252W mutation in neural crest-or mesoderm-derived tissues. *Bone* 2014;63:101-109.
27. Holmes, G., Basilico, C. Mesodermal expression of Fgfr2S252W is necessary and sufficient to induce craniosynostosis in a mouse model of Apert syndrome. *Developmental biology* 2012;368:283-293.
28. Slaney, S. F., Oldridge, M., Hurst, J. A., et al. Differential effects of FGFR2 mutations on syndactyly and cleft palate in Apert syndrome. *American journal of human genetics* 1996;58:923.
29. Wilkie, A. O. Craniosynostosis: genes and mechanisms. *Human molecular genetics* 1997;6:1647-1656.
30. Van De Lande, L. S., Greig, A. V., Dunaway, D. J. Craniosynostosis. *Plastic Surgery-Principles and Practice*: Elsevier; 2022:267-282.
31. Frakas, L. Anthropometry of the Head and Face in Medicine. Elsevier North Holland Inc, New York, 1981.
32. Atick, J. J., Griffin, P. A., Redlich, A. N. Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images. *Neural computation* 1996;8:1321-1340.
33. Craw, I., Cameron, P. Parameterising images for recognition and reconstruction. *BMVC91*: Springer; 1991:367-370.

34. Cootes, T. F., Edwards, G. J., Taylor, C. J. Active appearance models. In *European conference on computer vision*1998.
35. Hallinan, P. W., Gordon, G., Yuille, A. L., Giblin, P., Mumford, D. *Two-and three-dimensional patterns of the face*: AK Peters/CRC Press; 1999.
36. Jones, M. J., Poggio, T. Multidimensional morphable models. In *Sixth International Conference on Computer Vision (IEEE Cat No 98CH36271)*1998.
37. Levoy, M., Pulli, K., Curless, B., et al. The digital Michelangelo project: 3D scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*2000.
38. Egger, B., Smith, W. A., Tewari, A., et al. 3d morphable face models—past, present, and future. *ACM Transactions on Graphics (TOG)* 2020;39:1-38.
39. Paysan, P. Statistical modeling of facial aging based on 3D scans. University_of_Basel, 2010.
40. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics* 2013;20:413-425.
41. Brunton, A., Salazar, A., Bolkart, T., Wuhrer, S. Review of statistical shape spaces for 3D data with comparative analysis for human faces. *Computer Vision and Image Understanding* 2014;128:1-17.
42. Brunton, A., Bolkart, T., Wuhrer, S. Multilinear wavelets: A statistical shape space for human faces. In *European Conference on Computer Vision*2014.
43. Bolkart, T., Wuhrer, S. A groupwise multilinear correspondence optimization for 3d faces. In *Proceedings of the IEEE international conference on computer vision*2015.
44. Bolkart, T., Wuhrer, S. A robust multilinear model learning framework for 3d faces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*2016.
45. Huber, P., Hu, G., Tena, R., et al. A multiresolution 3d morphable face model and fitting framework. In *Proceedings of the 11th joint conference on computer vision, imaging and computer graphics theory and applications*2016.
46. Li, T., Bolkart, T., Black, M. J., Li, H., Romero, J. Learning a model of facial shape and expression from 4D scans. *ACM Trans Graph* 2017;36:194:191-194:117.
47. Gerig, T., Morel-Forster, A., Blumer, C., et al. Morphable face models-an open framework. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*2018.
48. Abrevaya, V. F., Wuhrer, S., Boyer, E. Multilinear autoencoder for 3d face model learning. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*2018.
49. Smith, W. A., Seck, A., Dee, H., Tiddeman, B., Tenenbaum, J. B., Egger, B. A morphable face albedo model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*2020.
50. Stratou, G., Ghosh, A., Debevec, P., Morency, L.-P. Effect of illumination on automatic expression recognition: a novel 3D relightable facial database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*2011.
51. Allen, B., Curless, B., Popović, Z. The space of human body shapes: reconstruction and parameterization from range scans. *ACM transactions on graphics (TOG)* 2003;22:587-594.
52. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J. Scape: shape completion and animation of people. *ACM SIGGRAPH 2005 Papers*; 2005:408-416.

53. Sun, Y., Murata, N. CAFM: A 3D morphable model for animals. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*2020.
54. Zuffi, S., Kanazawa, A., Black, M. J. Lions and tigers and bears: Capturing non-rigid, 3d, articulated shape from images. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*2018.
55. Shelton, C. R. Morphable surface models. *International Journal of Computer Vision* 2000;38:75-91.
56. Shamai, G., Slossberg, R., Kimmel, R. Synthesizing facial photometries and corresponding geometries using generative adversarial networks. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 2019;15:1-24.
57. Cheng, S., Bronstein, M., Zhou, Y., Kotsia, I., Pantic, M., Zafeiriou, S. Meshgan: Non-linear 3d morphable models of faces. *arXiv preprint arXiv:190310384* 2019.
58. Amberg, B., Romdhani, S., Vetter, T. Optimal step nonrigid ICP algorithms for surface registration. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*2007.
59. Liang, L., Wei, M., Szymczak, A., et al. Nonrigid iterative closest points for registration of 3D biomedical surfaces. *Optics and Lasers in Engineering* 2018;100:141-154.
60. Booth, J., Roussos, A., Ponniah, A., Dunaway, D., Zafeiriou, S. Large scale 3D morphable models. *International Journal of Computer Vision* 2018;126:233-254.
61. Maaten, L. v. d., Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* 2008;9:2579-2605.
62. Cohen Jr, M. M., Kreiborg, S. Growth pattern in the Apert syndrome. *American journal of medical genetics* 1993;47:617-623.
63. Goldberg, J., Enlow, D., Whitaker, L., Zins, J., Kurihara, S. Some anatomical characteristics in several craniofacial syndromes. *Journal of Oral Surgery (American Dental Association: 1965)* 1981;39:489-498.
64. Reitsma, J. H., Ongkosuwito, E. M., Buschang, P. H., Prahl-Andersen, B. Facial growth in patients with Apert and Crouzon syndromes compared to normal children. *The Cleft palate-craniofacial journal* 2012;49:185-193.
65. Forte, A. J., Alonso, N., Persing, J. A., Pfaff, M. J., Brooks, E. D., Steinbacher, D. M. Analysis of midface retrusion in Crouzon and Apert syndromes. *Plastic and reconstructive surgery* 2014;134:285-293.
66. Posnick, J. C., Lin, K. Y., Jhawar, B. J., Armstrong, D. Apert syndrome: quantitative assessment by CT scan of presenting deformity and surgical results after first-stage reconstruction. *Plastic and reconstructive surgery* 1994;93:489-497.
67. Lu, X., Forte, A. J., Sawh-Martinez, R., et al. Spatial and temporal changes of midface in Apert's syndrome. *Journal of plastic surgery and hand surgery* 2019;53:130-137.
68. Heuzé, Y., Martínez-Abadías, N., Stella, J. M., et al. Quantification of facial skeletal shape variation in fibroblast growth factor receptor-related craniosynostosis syndromes. *Birth Defects Research Part A: Clinical and Molecular Teratology* 2014;100:250-259.
69. Park, W.-J., Theda, C., Maestri, N. E., et al. Analysis of phenotypic features and FGFR2 mutations in Apert syndrome. *American journal of human genetics* 1995;57:321.
70. FIRMIN, F., COCCARO, P. J., CONVERSE, J. M. Cephalometric analysis in diagnosis and treatment planning of craniofacial dysostoses. *Plastic and reconstructive surgery* 1974;54:300-311.
71. Ortiz-Monasterio, F., Del Campo, A. F., Carrillo, A. Advancement of the orbits and the midface in one piece, combined with frontal repositioning, for the correction of Crouzon's deformities. *Plastic and reconstructive surgery* 1978;61:507-516.

72. Tessier, P. The definitive plastic surgical treatment of the severe facial deformities of craniofacial dysostosis: Crouzon's and Apert's diseases. *Annals of plastic surgery* 1987;18:330-351.
73. McCarthy, J., Schreiber, J., Karp, N., Thorne, C., Grayson, B. Lengthening the Human Mandible by Gradual Distraction. *Aesthetic Plastic Surgery in Asians: Principles and Techniques, Two-Volume Set* 2015:79.
74. McCarthy, J. G., Stelnicki, E. J., Mehrara, B. J., Longaker, M. T. Distraction osteogenesis of the craniofacial skeleton. *Plastic and reconstructive surgery* 2001;107:1812-1827.
75. Polley, J. W., Figueroa, A. A. Management of severe maxillary deficiency in childhood and adolescence through distraction osteogenesis with an external, adjustable, rigid distraction device. *The Journal of craniofacial surgery* 1997;8:181-185; discussion 186.
76. Van der Meulen, J. Medial faciotomy. *British journal of plastic surgery* 1979;32:339-342.
77. Tessier, P. Facial bipartition: A concept more than a procedure. *Craniofacial Surgery*: Springer; 1987:217-245.
78. Greig, A. V., Britto, J. A., Abela, C., et al. Correcting the typical Apert face: combining bipartition with monobloc distraction. *Plastic and reconstructive surgery* 2013;131:219e-230e.
79. Dunaway, D. J., Britto, J. A., Abela, C., Evans, R. D., Jeelani, N. O. Complications of frontofacial advancement. *Child's Nervous System* 2012;28:1571-1576.
80. Le Fort, R., Tessier, D. The classic reprint. Experimental study of fractures of the upper jaw. *Plast Reconstr Surg* 1972;50:497-506.
81. Ettinger, R. E., Hopper, R. A., Sandercoe, G., et al. Quantitative computed tomographic scan and polysomnographic analysis of patients with syndromic midface hypoplasia before and after Le Fort III distraction advancement. *Plastic and reconstructive surgery* 2011;127:1612-1619.
82. Fearon, J. A. The Le Fort III osteotomy: to distract or not to distract? *Plastic and reconstructive surgery* 2001;107:1091-1103; discussion 1104.
83. Hopper, R. A., Aspinall, C., Heike, C., et al. What the patients and parents do not tell you—recollections from families following external LeFort III midface distraction. *Plastic Surgical Nursing* 2009;29:78-85.
84. Shetye, P. R., Davidson, E. H., Sorkin, M., Grayson, B. H., McCarthy, J. G. Evaluation of three surgical techniques for advancement of the midface in growing children with syndromic craniosynostosis. *Plastic and reconstructive surgery* 2010;126:982-994.
85. Shetye, P. R., Grayson, B. H., McCarthy, J. G. Le Fort III distraction: controlling position and path of the osteotomized midface segment on a rigid platform. *Journal of Craniofacial Surgery* 2010;21:1118-1121.
86. Tessier, P. Total facial osteotomy. Crouzon's syndrome, Apert's syndrome: oxycephaly, scaphocephaly, turricephaly. In *Annales de chirurgie plastique*1967.
87. Hopper, R. A., Kapadia, H., Morton, T. Normalizing facial ratios in Apert syndrome patients with Le Fort II midface distraction and simultaneous zygomatic repositioning. *Plastic and reconstructive surgery* 2013;132:129-140.
88. Oberoi, S., Hoffman, W. Y., Vargervik, K. Craniofacial team management in Apert syndrome. *American Journal of Orthodontics and Dentofacial Orthopedics* 2012;141:S82-S87.

89. Ponniah, A. J., Witherow, H., Richards, R., Evans, R., Hayward, R., Dunaway, D. Three-dimensional image analysis of facial skeletal changes after monobloc and bipartition distraction. *Plastic and reconstructive surgery* 2008;122:225-231.
90. Crombag, G. A., Verdoorn, M. H., Nikkhah, D., Ponniah, A. J., Ruff, C., Dunaway, D. Assessing the corrective effects of facial bipartition distraction in Apert syndrome using geometric morphometrics. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 2014;67:e151-e161.
91. Takashima, M., Kitai, N., Murakami, S., et al. Dual segmental distraction osteogenesis of the midface in a patient with Apert syndrome. *The Cleft palate-craniofacial journal* 2006;43:499-506.
92. Spruijt, B., Joosten, K. F., Driessen, C., et al. Algorithm for the management of intracranial hypertension in children with syndromic craniosynostosis. *Plastic and reconstructive surgery* 2015;136:331-340.
93. de Jong, T., Bannink, N., Bredero-Boelhouwer, H., et al. Long-term functional outcome in 167 patients with syndromic craniosynostosis; defining a syndrome-specific risk profile. *Journal of plastic, reconstructive & aesthetic surgery* 2010;63:1635-1641.
94. Doerga, P. N., Spruijt, B., Mathijssen, I. M., Wolvius, E. B., Joosten, K. F., van der Schroeff, M. P. Upper airway endoscopy to optimize obstructive sleep apnea treatment in Apert and Crouzon syndromes. *Journal of Cranio-Maxillofacial Surgery* 2016;44:191-196.
95. David, D. J., Anderson, P., Flapper, W., Syme-Grant, J., Santoreneos, S., Moore, M. Apert syndrome: outcomes from the Australian Craniofacial Unit's birth to maturity management protocol. *Journal of Craniofacial Surgery* 2016;27:1125-1134.
96. Driessen, C., van Veelen, M., Joosten, K., et al. Apert syndrome: the Paris and Rotterdam philosophy. *Expert opinion on orphan drugs* 2017;5:599-605.
97. Gandolfi, B. M., Sobol, D. L., Farjat, A. E., Allori, A. C., Muh, C. R., Marcus, J. R. Risk Factors for Delayed Referral to a Craniofacial Specialist for.
98. Horos project. Available at: https://horosproject.org/about/?utm_medium=email&utm_source=transaction.
99. Autodesk Meshmixer (RRID:SCR_015736). Available at: https://www.meshmixer.com/.
100. Materialise Mimics Inprint 3.0. Available at: https://www.materialise.com/en/medical/mimics-innovation-suite/products-services.
101. Ross, A. Procrustes analysis. *Course report, Department of Computer Science and Engineering, University of South Carolina* 2004;26.
102. Davies, R., Twining, C., Taylor, C. *Statistical models of shape: Optimisation and evaluation*: Springer Science & Business Media; 2008.
103. Cohen Jr, M. M., Kreiborg, S. A clinical study of the craniofacial features in Apert syndrome. *International journal of oral and maxillofacial surgery* 1996;25:45-53.
104. Deng, J., Zhou, Y., Kotsia, I., Zafeiriou, S. Dense 3D face decoding over 2500FPS: joint texture and shape convolutional mesh decoders.
105. Hallgrímsson, B., Aponte, J. D., Katz, D. C., et al. Automated syndrome diagnosis by three-dimensional facial imaging. *Genetics in Medicine* 2020;22:1682-1693.
106. Zhou, Y., Zaferiou, S. Deformable models of ears in-the-wild for alignment and recognition. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*2017.
107. O'Sullivan, E., van de Lande, L. S., Papaioannou, A., et al. Craniofacial Syndrome Identification Using Convolutional Mesh Autoencoders. *Available at SSRN 3795325*.

108. Dall'Asta, A., Schievano, S., Bruse, J. L., et al. Quantitative analysis of fetal facial morphology using 3D ultrasound and statistical shape modeling: a feasibility study. *American journal of obstetrics and gynecology* 2017;217:76. e71-76. e78.
109. Gurovich, Y., Hanani, Y., Bar, O., et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nature medicine* 2019;25:60-64.
110. Matthews, H. S., Palmer, R. L., Baynam, G. S., et al. Large-scale open-source three-dimensional growth curves for clinical facial assessment and objective description of facial dysmorphism. *Scientific reports* 2021;11:1-12.
111. Liu, F., Tran, L., Liu, X. 3d face modeling from diverse raw scan data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*2019.
112. Farkas, L. G., Posnick, J. C., Hreczko, T. M. Anthropometric growth study of the head. *The Cleft Palate-Craniofacial Journal* 1992;29:303-308.
113. Mauler, F., Langguth, C., Schweizer, A., et al. Prediction of normal bone anatomy for the planning of corrective osteotomies of malunited forearm bones using a three-dimensional statistical shape model. *Journal of Orthopaedic Research* 2017;35:2630-2636.
114. Staal, F. C., Ponniah, A. J., Angullia, F., Ruff, C., Koudstaal, M. J., Dunaway, D. Describing Crouzon and Pfeiffer syndrome based on principal component analysis. *Journal of Cranio-Maxillofacial Surgery* 2015;43:528-536.
115. Delye, H., Clijmans, T., Mommaerts, M. Y., Sloten, J. V., Goffin, J. Creating a normative database of age-specific 3D geometrical data, bone density, and bone thickness of the developing skull: a pilot study. *Journal of Neurosurgery: Pediatrics* 2015;16:687-702.
116. Duan, F., Huang, D., Tian, Y., Lu, K., Wu, Z., Zhou, M. 3D face reconstruction from skull by regression modeling in shape parameter spaces. *Neurocomputing* 2015;151:674-682.
117. Lüthi, M., Lerch, A., Albrecht, T., Krol, Z., Vetter, T. A hierarchical, multi-resolution approach for model-based skull-segmentation in mri volumes. In *Conference Proceedings D*2009.
118. Khechoyan, D. Y., Saber, N. R., Burge, J., et al. Surgical outcomes in craniosynostosis reconstruction: the use of prefabricated templates in cranial vault remodelling. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 2014;67:9-16.
119. Mardini, S., Alsubaie, S., Cayci, C., Chim, H., Wetjen, N. Three-dimensional preoperative virtual planning and template use for surgical correction of craniosynostosis. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 2014;67:336-343.
120. Badano, A. In silico imaging clinical trials: cheaper, faster, better, safer, and more scalable. *Trials* 2021;22:1-7.
121. Faris, O., Shuren, J. An FDA viewpoint on unique considerations for medical-device clinical trials. *New England Journal of Medicine* 2017;376:1350-1357.
122. Marcus, J. R., Domeshek, L. F., Loyd, A. M., et al. Use of a three-dimensional, normative database of pediatric craniofacial morphology for modern anthropometric analysis. *Plastic and reconstructive surgery* 2009;124:2076-2084.
123. Swennen, G. R., Schutyser, F. Three-dimensional cephalometry: spiral multi-slice vs cone-beam computed tomography. *American Journal of Orthodontics and Dentofacial Orthopedics* 2006;130:410-416.
124. Bruse, J. L., Khushnood, A., McLeod, K., et al. How successful is successful? Aortic arch shape after successful aortic coarctation repair correlates with left ventricular function. *The Journal of thoracic and cardiovascular surgery* 2017;153:418-427.

125. Kaya, O., Pluijmers, B. I., Staal, F., et al. Describing the mandible in patients with craniofacial microsomia based on principal component analysis and thin plate spline video analysis. *International journal of oral and maxillofacial surgery* 2019;48:302-308.
126. Badiali, G., Roncari, A., Bianchi, A., Taddei, F., Marchetti, C., Schileo, E. Navigation in orthognathic surgery: 3D accuracy. *Facial Plastic Surgery* 2015;31:463-473.
127. Almukhtar, A., Khambay, B., Ju, X., Ayoub, A. Comprehensive analysis of soft tissue changes in response to orthognathic surgery: mandibular versus bimaxillary advancement. *International journal of oral and maxillofacial surgery* 2018;47:732-737.
128. Miller, L., Morris, D. O., Berry, E. Visualizing three-dimensional facial soft tissue changes following orthognathic surgery. *The European Journal of Orthodontics* 2007;29:14-20.
129. Archip, N., Clatz, O., Whalen, S., et al. Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery. *Neuroimage* 2007;35:609-624.
130. Chen, L., Zhang, X., He, Y., Wang, W., Zhang, F., Sun, L. A method of 3D-3D multi-stage non-rigid registration of the spine based on binocular structured light. *The International Journal of Medical Robotics and Computer Assisted Surgery* 2021:e2283.
131. Shetye, P. R., Boutros, S., Grayson, B. H., McCarthy, J. G. Midterm follow-up of midface distraction for syndromic craniosynostosis: a clinical and cephalometric study. *Plastic and reconstructive surgery* 2007;120:1621-1632.
132. Shetye, P. R., Caterson, E. J., Grayson, B. H., McCarthy, J. G. Soft-tissue profile changes following early Le Fort III distraction in growing children with syndromic craniosynostosis. *Plastic and reconstructive surgery* 2013;132:945-954.
133. Visser, R., Ruff, C. F., Angullia, F., et al. Evaluating the efficacy of monobloc distraction in the Crouzon-Pfeiffer craniofacial deformity using geometric morphometrics. *Plastic and reconstructive surgery* 2017;139:477e-487e.
134. Olate, S., Zaror, C., Mommaerts, M. Y. A systematic review of soft-to-hard tissue ratios in orthognathic surgery. Part IV: 3D analysis–Is there evidence? *Journal of Cranio-Maxillofacial Surgery* 2017;45:1278-1286.
135. Materialise 3-Matic Available at: https://www.materialise.com/en/software/3-matic.
136. Mellion, Z. J., Behrents, R. G., Johnston Jr, L. E. The pattern of facial skeletal growth and its relationship to various common indexes of maturation. *American Journal of Orthodontics and Dentofacial Orthopedics* 2013;143:845-854.
137. Tovetjärn, R., Tarnow, P., Maltese, G., Fischer, S., Sahlin, P.-E., Kölby, L. Children with Apert syndrome as adults: a follow-up study of 28 Scandinavian patients. *Plastic and reconstructive surgery* 2012;130:572e-576e.
138. Glass, G. E., Ruff, C. F., Crombag, G. A., et al. The role of bipartition distraction in the treatment of Apert syndrome. *Plastic and reconstructive surgery* 2018;141:747-750.
139. Von Gernet, S., Golla, A., Ehrenfels, Y., Schuffenhauer, S., Fairley, J. Genotype–phenotype analysis in Apert syndrome suggests opposite effects of the two recurrent mutations on syndactyly and outcome of craniofacial surgery. *Clinical genetics* 2000;57:137-139.
140. O'Keeffe, G. S., Clarke-Pearson, K. The impact of social media on children, adolescents, and families. *Pediatrics* 2011;127:800-804.
141. Eley, K. A., Watt-Smith, S. R., Sheerin, F., Golding, S. J. "Black Bone" MRI: a potential alternative to CT with three-dimensional reconstruction of the craniofacial skeleton in the diagnosis of craniosynostosis. *European radiology* 2014;24:2417-2426.

142. van de Lande, L., Papaioannou, A., Zafeiriou, S., Dunaway, D. S12-09 SESSION 12: FACIOCRANIOSYNOSTOSIS–PART II A MACHINE LEARNING APPROACH FOR OUTCOME PREDICTION OF MIDFACIAL BIPARTITION DISTRACTION IN APERT PATIENTS. *Plastic and Reconstructive Surgery–Global Open* 2019;7:155-156.
143. Koo, T. K., Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 2016;15:155-163.

Appendix A

LIST OF PUBLICATIONS

A.1. Peer reviewed journal articles and book publications directly related to this work.

1. **Lara S. van de Lande**, Athanasios Papaioannou, David J. Dunaway. Geometric morphometrics aided by machine learning in craniofacial surgery. *Journal of Orthodontics, 2019.*
2. **Lara S. van de Lande***, Eimear O' Sullivan*, Paul G.M. Knoops, Athanasios Papaioannou, Juling Ong, Greg James, N. Owase Jeelani, Silvia Schievano, David J. Dunaway. Local Soft Tissue and Bone Displacements following Midfacial Bipartition Distraction in Apert Syndrome – Quantification using a Semi-Automated Method. *The Journal of Craniofacial Surgery, 2020.*
3. **Lara S. van de Lande**, Aina. V.H. Greig, David J. Dunaway. Chapter 20 Craniosynostosis. *Plastic Surgery: Principles & Practice 1e, Farhadieh. Publisher: Elsevier, 2021.*
4. **Lara S. van de Lande**, Athanasios Papaioannou, David J. Dunaway. Chapter 3 Anthropometrics. *Aesthetic Surgery of the Facial Skeleton, Baker. Publisher: Elsevier, 2021.*
5. **Lara S. van de Lande***, Eimear O' Sullivan*, Athanasios Papaioannou, N. Owase Jeelani, Silvia Schievano, Roman Khonsari, David. J. Dunaway, Stefanos Zafeiriou. Craniofacial syndrome identification using 3D Morphable model-based machine learning. *Nature Scientific Reports, 2021.*
6. Eimear O' Sullivan***, Lara S. van de Lande***, Anne-Jet C. Oosting, Athanasios Papaioannou, N. Owase Jeelani, Maarten J. Koudstaal, Silvia Schievano, Roman H. Khonsari, David J. Dunaway, Stefanos Zafeiriou A Statistical Model of the Paediatric Skull. *Bone Reports, 2021*

A.2. Peer reviewed journal articles not directly related to this work.

1. **Lara S. van de Lande*,** Cornelia J.J.M Caron*, Britt I. Pluijmers, Koen F.M. Joosten, Marloes Streppel, David J. Dunaway, Maarten J. Koudstaal, Bonnie L. Padwa. Evaluation of Swallow Function in Patients with Craniofacial Microsomia: A Retrospective Study. *Dysphagia, 2017.*
2. **Lara S. van de Lande,** Britt I. Pluijmers, Cornelia J.J.M. Caron, Eppo B. Wolvius, Prof; David J Dunaway, Maarten J. Koudstaal, Bonnie L Padwa. Surgical correction of the Midface in Craniofacial Microsomia; Part 1: a systematic review. *Journal of Cranio-Maxillofacial Surgery, 2018.*
3. **Lara S van de Lande*,** Britt I. Pluijmers*, Cornelia J.J.M. Caron, DMD; Eppo B. Wolvius, Prof; David J Dunaway, Bonnie L Padwa, Maarten J. Koudstaal. Part 2: Is the maxillary canting and its surgical correction in patients with CFM correlated to the mandibular deformity? *Journal of Cranio-Maxillofacial Surgery, 2018.*
4. **Lara S. van de Lande**, Ben M. Eyck, Jelle J. Mooij, Hieronymus P. Stevens, Joris A. van Dongen. The Rainbow Scale for the assessment of the cervicomental angle: a validated scale. *Aesthetic Surgery Journal, 2019.*
5. Britt I. Pluijmers, Cornelia J.J.M. Caron, **Lara S. van de Lande,** Sontje Schaal, Irene M. Mathijssen, Eppo B. Wolvius, Neil Bulstrode, Robert D. Evans, Peter Ayliffe, Bonnie Padwa MD; Maarten J. Koudstaal, David Dunaway. Surgical correction of craniofacial microsomia: evaluation of interventions in 565 patients at 3 major craniofacial units. *Journal of Plastic and Reconstructive Surgery, 2019.*

6. Joris A. van Dongen, Mirte Langeveld, **Lara S. van de Lande**, Martin C. Harmsen, Hieronymus P. Stevens, Berend van der Lei. The effects of facial lipografting on skin quality: a systematic review. *Journal of Plastic and Reconstructive Surgery, 2019.*

7. Joris A. van Dongen, Mirte Langeveld, **Lara S. van de Lande**, Martin C. Harmsen, Hieronymus P. Stevens, Berend van der Lei. Reply to Letter: 'The effects of facial lipografting on skin quality: a systematic review' *Journal of Plastic and Reconstructive Surgery, 2020.*

8. S. Bozkurt, A. Borghi, **L.S. van de Lande**, N.U.O. Jeelani, D. J. Dunaway, S. Schievano. Computational Modelling of Patient-Specific Spring Assisted Lambdoid Craniosynostosis Correction. *Nature Scientific Reports, 2020.*

9. Karan R.R. Ramdat Misier, Richard. W.F. Breakey, **Lara S. van de Lande**, Benedetta Biffi, Paul G.M. Knoops, Silvia Schievano, Cornelia J.J.M. Caron, David J. Dunaway, Maarten J. Koudstaal, N.U. Owase Jeelani, Alessandro Borghi. Correlation Between Head Shape and Volumetric Changes Following Spring-Assisted Posterior Vault Expansion. *International Journal of Cranio-Maxillofacial Surgery, 2020.*

10. Alexander J. Rickart, **Lara S. Van de Lande**, Eimear O' Sullivan, Juling Ong, Daljit S. Gil, Robert D. Evans, Silvia Schievano, David J. Dunaway. Dental Arch Relapse Following Facial Bipartition: a Three-Dimensional Semi-Automated Quantification. *The Journal of Craniofacial Surgery, 2021.*

11. Mirte Langeveld*, **Lara S. van de Lande*,** Eimear O' Sullivan, Berend van der Lei, Joris A. van Dongen. View point: Skin measurement devices to assess skin quality. *Journal of Plastic and Reconstructive Surgery, 2021.*

12. Sohaib R. Rufai, Oliver R. Marmoy, **Lara S. van de Lande,** R. William Breakey, Catey Bunce, Frank A Proudlock, Vasiliki Panteli, Kemmie Schwiebert, David J. Dunaway, Dorothy Thompson, Irene Gottlob, Richard Bowman, Noor ul Owase Jeelani. Ophthalmologic Monitoring of Intracranial Pressure and Visual Outcomes in Surgical Patients with Craniosynostosis: A Longitudinal Study. *Journal Eye, 2021.*

13. R. William Breakey*, **Lara S. van de Lande***, Jai Sidpra, Paul M. Knoops, Alessandro Borghi, Justine O'Hara, Juling Ong, Greg James, Silvia Schievano, Richard Hayward, David J. Dunaway, N. Owase Jeelani. Spring-Assisted Posterior Vault Expansion – a Single Centre experience of 200 cases. *Journal of Child Nerve System, 2021.*

14. Mirte Langeveld*, **Lara S. van de Lande***, Eimear O' Sullivan, Berend van der Lei, Joris A. van Dongen. Skin measurement devices to assess skin quality: a systematic review on reliability and validity. *Skin Research and Technology, 2021.*

15. Cornelia J.J.M. Caron, Caroline A.A. Beaumont, Eline E.C.M. Elsten, **Lara S. van de Lande**, MD, Koen F.M. Joosten, Bonnie L. Padwa, David J. Dunaway, E.B. Wolvius, Maarten J. Koudstaal. The effect of cleft lip and palate in craniofacial microsomia on breathing, feeding and swallowing. *Journal of Craniomaxillo facial surgery, 2020* [under review].

16. J.J. Mooij*, **L.S. van de Lande***, S.M.W. Pool, H.P.J.D. Stevens, B. van der Lei, J.A. van Dongen. The Rainbow Scale for the Assessment of dermatochalasis of the upper eyelid: A validated scale. *Aesthetic Surgery Journal, 2021* [under review].

17. Anoopama Ramjeeawon, Pleun van der Plas, **Lara van de Lande**, Juling Ong, Michelle Wyatt, Francois Abel, Loshan Kangesu, Brian Sommerlad, Jamuna Navaratnarajah, Eppo Wolvius, Aidan Laverty, Neil Bulstrode. Airway and

Breathing Problems in Pierre Robin Sequence: a systematic review protocol. *Systematic Reviews,2021*.

18. Andrea White*, **Lara S. van de Lande***, Justine O'Hara, John Hartley, Richard Hayward, Greg James, N. Owase Jeelani, David J. Dunaway. Fronto-facial surgery: reducing infection with the development and 6-year outcome of a fronto-facial protocol. *Journal of Plastic and Reconstructive Surgery, 2021*.

19. Eimear O' Sullivan*, **Lara S. van de Lande***, Khalid El Ghoul, Maarten J. Koudstaal, Silvia Schievano, Roman H. Khonsari, David J. Dunaway, Stefanos Zafeiriou. Growth Patterns and Shape Development of the Paediatric Mandible – A 3D statistical model. *Bone reports, 2022*

20. Eimear O' Sullivan, Lara S. van de Lande, Anne-Jet C. Oosting, Athanasios Papaioannou, N. Owase Jeelani, Maarten J. Koudstaal, Silvia Schievano, Stefanos Zafeiriou, Roman H. Khonsari, David J. Dunaway. 3D shape changes of the human skull during the first 4 years of life. *Bone reports, 2022* [under review].

A.3. Peer reviewed conference publications directly related to this work

1. **Lara S. van de Lande***, Athanasios Papaioannou, Richard W.F. Breakey, David J. Dunaway. A new 3D statistical model for Treacher Collins Syndrome using Machine Learning Techniques to Quantify Facial Shape Differences from the Normal. Presented at *The 24th International Conference on Oral & Maxillofacial*

Surgery (ICOMS), 21-25 May 2019, Rio de Janeiro, Brazil. Published in *International Journal of Oral and Maxilofacial Surgery, May 2019.*

2. Athanasios Papaioannou*, **Lara S. Van de Lande**, Stefanos Zafeiriou, David J. Dunaway. 3D Statistical Face Model for Syndromic Craniofacial Patients. Presented at *The 18th Congress of International Society of Craniofacial Surgery (ISCFS),* 16-19 September 2019, Paris, France. Published in the *Journal of Plastic and Reconstructive Surgery, Sept 2019.*

3. **Lara S. van de Lande***, Athanasios Papaioannou, Stefanos Zafeiriou, David J. Dunaway. A Machine Learning Approach for Outcome Prediction of Midfacial Bipartitions Distraction in Apert Patients. Presented at *The 18th Congress of International Society of Craniofacial Surgery (ISCFS),* 16-19 September 2019, Paris, France. Published in the *Journal of Plastic and Reconstructive Surgery, September 2019.*

4. David J. Dunaway*, Athanasios Papaioannou, Alessandro Borghi, **Lara van de Lande**, Paul Knoops, Freida Angullia, Owase N. Jeelani, Silvia Schievano, Stefanos Zafeiriou. Statistical Shape Modeling And Related Techniques as A Method to Aid Diagnosis, Plan Surgery and Assess Outcome In Craniofacial Surgery. Presented at *The 18th Congress of International Society of Craniofacial Surgery (ISCFS),* 16-19 September 2019, Paris, France. Published in the *Journal of Plastic and Reconstructive Surgery, September 2019.*

5. Eimear O' Sullivan*, **Lara Sophie van de Lande**, Anne-Jet C. Oosting, Athanasios Papaioannou, Maarten J. Koudstaal, Silvia Schievano, Roman Khonsari, David J. Dunaway. A 'normal' paediatric skull 3d morphable model. Presented at *The 25th Congress of the European Association for Cranio Maxillo Facial Surgery*

(EACFMS), 14-16 July 2021, Paris, France (Virtual congress). Published in *The Journal of Craniofacial Surgery, July 2021.*

6. **Lara Sophie van de Lande***, Eimear O' Sullivan, Athanasios Papaioannou, Silvia Schievano, Stefanos Zafeiriou, Roman Hossein, David J. Dunaway. Abstract publication. A paediatric soft tissue 3d morphable model. Presented at *The 25th Congress of the European Association for Cranio Maxillo Facial Surgery (EACFMS), 14-16 July 2021, Paris, France (Virtual congress).* Published in *The Journal of Craniofacial Surgery, July 2021.*
7. Anne-Jet C. Oosting*, **Lara S. van de Lande,** Eimear O' Sullivan, Athanasios Papaioannou, Maarten J. Koudstaal, Silvia Schievano, Stefanos Zafeiriou, David J. Dunaway. A 3D skull morphable model of apert syndrome. Presented at *The 25th Congress of the European Association for Cranio Maxillo Facial Surgery (EACFMS), 14-16 July 2021, Paris, France (Virtual congress).* Published in *The Journal of Craniofacial Surgery, July 2021.*
8. **Lara S. van de Lande**, Eimear O' Sullivan, Paul M. Knoops, Athanasios Papaioannou, Juling Ong, Greg James, Noor O. Jeelani, Silvia Schievano, David. J. Dunaway. Abstract publication Why we should use non-rigid iterative closest point registration for surgical outcome measurements – demonstrated by midfacial bipartition distraction in Apert syndrome. Presented at *The 25th Congress of the European Association for Cranio Maxillo Facial Surgery (EACFMS), 14-16 July 2021, Paris, France (Virtual congress).* Published in *The Journal of Craniofacial Surgery, July 2021.*

Appendix B

SUPPLEMENTARY INFORMATION

B.1. Search terms for validated landmarks

(((((("Algorithms"[Mesh] AND "Anatomic Landmarks"[Majr]) AND "Image Processing, Computer-Assisted"[Majr]) AND "Imaging, Three-Dimensional/methods"[Majr]) AND "Reproducibility of Results"[Mesh]) AND "Skull/diagnostic imaging"[Majr]) OR ((((("Anatomic Landmarks/diagnostic imaging"[Mesh]) AND "Imaging, Three-Dimensional"[Mesh]) AND "Skull"[Mesh]) AND "Reproducibility of Results"[Mesh])) OR ((((("Cephalometry/methods"[Mesh] OR "Cephalometry/statistics and numerical data"[Mesh])) AND "Reproducibility of Results"[Mesh]) AND ("Imaging, Three-Dimensional/methods"[Mesh] OR "Imaging, Three-Dimensional/statistics and numerical data"[Mesh])) AND "Anatomic Landmarks"[Mesh]) OR ((((("Anatomic Landmarks"[Majr] AND ("Imaging, Three-Dimensional/methods"[Mesh] OR "Imaging, Three-Dimensional/statistics and numerical data"[Mesh])) AND "Skull"[Majr]) AND "Reproducibility of Results"[Mesh]) AND "Cephalometry"[Mesh])

B.2. Intra- and interclass correlations for skull landmarks

Intra- and interclass correlations with a value less than 0.5 indicate poor reliability, values between 0.5 and 0.75 moderate reliability, values between 0.75-0.9 good reliability and values greater than 0.9 indicate excellent reliability (143). Table B.2.1. provides an overview of mean intra- and interclass correlation in X-, Y-, Z- and XYZ-axis per landmark found in the 3 included studies:

Liberton *et al.*

Liberton DK, Verma P, Contratto A, Lee JS. Development and Validation of Novel Three-Dimensional Craniofacial Landmarks on Cone-Beam Computed Tomography Scans. J Craniofac Surg. 2019;30(7):e611-e5.

Da Neiva *et al.*

Neiva MB, Soares AC, Lisboa Cde O, Vilella Ode V, Motta AT. Evaluation of cephalometric landmark identification on CBCT multiplanar and 3D reconstructions. Angle Orthod. 2015;85(1):11-7.

Lemieux *et al.*

Lemieux G, Carey JP, Flores-Mir C, Secanell M, Hart A, Lagravere MO. Precision and accuracy of suggested maxillary and mandibular landmarks with cone-beam computed tomography for regional superimpositions: An in vitro study. Am J Orthod Dentofacial Orthop. 2016;149(1):67-75.

Table B.2.1. Overview of mean intra- and interclass correlations. The correlations are provided in X-, Y-, Z- and XYZ-axis per landmark: ICC <0.5 indicate poor reliability (**red**), 0.5 and 0.75 moderate reliability (**yellow**), 0.75-0.9 good reliability (**seagreen**) and > 0.9 indicate excellent reliability (**green**). L is left and R is right.

	Mean Intraclass Correlation (ICC)				**Mean Interclass Correlation (ICC)**			
Landmark	**X**	**Y**	**Z**	**XYZ**	**X**	**Y**	**Z**	**XYZ**
Nasion	0,99	1,00	0,89	**0,96**	0,96	0,99	0,89	**0,95**
Prosthion	0,74	0,91	0,65	**0,77**	0,81	0,97	0,80	**0,86**
A Point	0,90	0,95	0,78	**0,88**	0,91	0,94	0,79	**0,88**
ANS	0,88	0,91	0,94	**0,91**	0,93	0,93	0,96	**0,94**
Orbitale R	0,86	0,87	0,97	**0,90**	0,80	0,86	0,96	**0,87**
Orbitale L	0,76	0,85	0,97	**0,86**	0,77	0,87	0,95	**0,86**
Supraorbitale R	0,88	0,89	0,90	**0,89**	0,32	0,79	0,76	**0,62**
Supraorbitale L	0,62	0,83	0,87	**0,77**	0,20	0,51	0,84	**0,52**
Frontozygomatic suture R	0,95	0,96	0,95	**0,95**	0,94	0,98	0,94	**0,95**
Frontozygomatic suture L	0,95	0,93	0,91	**0,93**	0,89	0,96	0,94	**0,93**
Zygomatic arch R	0,96	0,86	0,94	**0,92**	0,98	0,87	0,96	**0,94**
Zygomatic arch L	0,96	0,70	0,95	**0,87**	0,98	0,78	0,96	**0,91**
Jugale R	0,75	0,88	0,85	**0,83**	0,67	0,93	0,87	**0,82**
Jugale L	0,71	0,87	0,79	**0,79**	0,73	0,89	0,89	**0,84**
Nasal cavity R	0,71	0,91	0,91	**0,84**	0,68	0,94	0,96	**0,86**
Nasal cavity L	0,58	0,90	0,85	**0,78**	0,77	0,95	0,92	**0,88**
Infradentale mandible	0,86	0,99	0,95	**0,93**	0,91	0,99	0,91	**0,94**
B point	0,92	0,98	0,89	**0,93**	0,94	0,92	0,89	**0,91**
Pogonion	0,90	0,97	0,94	**0,93**	0,94	0,92	0,93	**0,93**
Gnathion	0,94	0,91	0,68	**0,84**	0,94	0,81	0,69	**0,81**
Menton	0,93	0,97	0,98	**0,96**	0,94	0,92	0,95	**0,94**
Coronoid process R	0,96	0,97	0,98	**0,97**	0,95	0,95	0,98	**0,96**
Coronoid process L	0,97	0,97	0,97	**0,97**	0,98	0,95	0,98	**0,97**
Condylion R	0,45	0,62	0,96	**0,68**	0,49	0,60	0,98	**0,69**
Condylion L	0,75	0,95	0,97	**0,89**	0,80	0,96	0,98	**0,91**
Articulare R	0,82	0,99	0,92	**0,91**	0,66	0,98	0,93	**0,86**
Articulare L	0,62	0,98	0,91	**0,84**	0,76	0,97	0,95	**0,89**
Gonion R	0,76	0,77	0,93	**0,82**	0,76	0,74	0,96	**0,82**
Gonion L	0,98	0,96	0,94	**0,96**	0,96	0,95	0,95	**0,95**
Infraorbital foramen R	1,00	0,99	1,00	**1,00**	1,00	0,99	0,98	**0,99**
Infraorbital foramen L	1,00	1,00	1,00	**1,00**	1,00	0,99	1,00	**0,99**
Mental foramen R	1,00	1,00	1,00	**1,00**	1,00	1,00	0,99	**1,00**
Mental foramen L	1,00	1,00	0,98	**0,99**	1,00	0,99	0,99	**1,00**

B.4. Weblinks

Link B.4.1. This is the weblink used to for annotation of the soft tissue meshes for the construction of the face 3DMMs:

https://www.landmarker.io/#server=https%3A%2F%2F5c2ec5e8.ngrok.io