Cephalometric Landmark Detection:

Artificial Intelligence vs Human Examination

Suvarna Indermun

A mini-thesis submitted to the Faculty of Dentistry, University of the Western Cape, in partial fulfilment of the requirements for the degree Magister Scientiae in the Department of Diagnostics and Radiology, Faculty of Dentistry, University of the Western Cape.

November 2021

Supervisor: Dr Shoayeb Shaik

Co-supervisors: Dr Clement Nyirenda and Dr Keith Johannes

KEYWORDS

Accuracy

Artificial intelligence

Automated identification

Cephalometric analysis

Cephalometry

Computer-assisted

Deep learning

Dentistry

Detection

Error

Human examination

Landmarks

Machine learning

Orthodontics

Precision

Radiology

ABSTRACT

Background: Cephalometric landmark detection is important for accurate diagnosis and treatment planning. The most common cause of random errors, in both computer-aided cephalometry and manual cephalometric analysis, is inconsistency in landmark detection. These methods are time-consuming. As a result, attempts have been made to automate cephalometric analysis, to improve the accuracy and precision of landmark detection whilst also minimizing errors caused by clinician subjectivity.

Aim: This mini-thesis aimed to determine the precision of two cephalometric landmark identification methods, namely an artificial intelligence programme (BoneFinder®) and a computer-assisted examination software (Dolphin Imaging™).

Methods: This was a retrospective quantitative cross-sectional analytical study. The dataset comprised of 409 cephalograms obtained from a South African population. 19 landmarks were selected and detected using a computer-assisted approach and an automatic approach. The *x,y* coordinates for each landmark per system was recorded and the Euclidean distance was calculated. Precision was determined by calculating the standard deviation and standard error of the mean.

Results: The primary researcher acted as the gold standard and was calibrated prior to data collection. The inter- and intra-reliability tests yielded acceptable results. There were variations present in several landmarks between Dolphin and BoneFinder; however, they were statistically insignificant. The computer-aided approach was very sensitive to several variables. Attempts were made to draw valid comparisons and conclusions.

Conclusion: There was no significant difference between the artificial intelligence programme (BoneFinder®) and the computer-assisted human examination (Dolphin Imaging™) regarding the precision of landmark detection.

DECLARATION

I declare that **"*Cephalometric Landmark Detection: Artificial Intelligence vs Human Examination*"** is my own work, that it has not been submitted before for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged as complete references.

Suvarna Indermun 04 November 2021

Signed: ……………………………………

ACKNOWLEDGEMENTS

There are many people that I need to acknowledge. As always, the greatest debt one owes is to one's family, friends, and colleagues. In my case, my debt is overwhelming. To everyone mentioned here – you have my infinite respect and gratitude.

First and foremost, I thank God, the true author of wisdom and knowledge, for the countless gifts provided, in the most mysterious of ways.

To my parents, Molly and Vinesh, who continued to ask "when will you be done?" at the best and worst of times - this thesis belongs to all of us. Thank you for your prayers, love and your continued support and encouragement throughout all my moods.

To my sister and brother, Sunaina and Shival, thank you for being constant sources of inspiration and motivation. You both have gone beyond the roles of siblings – to include that of proofreaders and advisors.

In writing this thesis, I have been fortunate to be assisted by experts in the fields of radiology, computer science and orthodontics. This could not have been done without the help of my supervisors:

- My deepest thanks to my supervisor - Dr Shoayeb Shaik (Department of Oral and Maxillofacial Radiology). Thank you for always bringing calm to my mind and for bringing this topic to my attention. Thank you for your mentorship and support these past years. I am who am today, because of you. I've learnt things from you that I never could have learnt from textbooks.
- To my co-supervisor, Dr Clement Nyirenda (Department of Computer Sciences) – thank you for your unwavering support. I truly appreciate every session where you patiently explained AI complexities in the simplest way. It was a pleasure collaborating with you.
- To my co-supervisor, Dr Keith Johannes (Department of Orthodontics; South Africa Authorized Distributor for Dolphin Imaging™) - thank you for your advice and

assistance with using Dolphin Imaging™. Your time was the greatest thing you have given me throughout this project.

- Royston Johannes, thank you for your technical expertise (Dolphin Imaging™).
- Prof Angela Harris, thank you for your initial guidance.
- Dr Claudia Lindner, the primary author of BoneFinder® software, who confirmed my belief that unexpected kindness is the most powerful, least costly and underrated agent of humanity. Thank you for letting me depend on the kindness of strangers, and for your willingness to help and for the generous advice. Your insight has been beyond valuable.
- Dr Riaan Mulder – for stepping in when you did, and offering statistical advice with amusing analogies, patience and kindness.
- Dr Faheema Kimmie – thank you for your continued advice and reassurance.
- Zulfah Smith, you know what role you played behind the scenes. Honestly, you are the rock-star of administration.

⬧⧫⬧

To my colleagues in Radiology: Jaco, Lisa, Leila, Naazia, Hoda, Florence and Tineke – I have not forgotten the subtle ways in which you have influenced my work, my progress and my ideas. Your teachings weren't explicit, but throughout this journey, I have learnt something from each of you.

My dear colleague and friend – Khaled Beshtawi – thank you for always being unselfish with imparting knowledge and advice. Your dedication to this field has always left me in awe and inspired.

My dear teachers – Prof Nortje, Prof Parker, Ms Crombie and Prof Myburgh thank you for laying the foundation that started this journey.

The staff in the Radiology Department – whose daily interactions made this journey so memorable and enjoyable. During the most stressful of days, you all could bring laughter to any situation.

Dr Ridwaana Carim and Dr Qaanita Isaacs – thank you both for being so supportive with your unlimited willingness to help. Finally, thank you to my "faculty mum", Dr Razia Adam, my constant cheerleader who always "told me so".

DEDICATION

⬩◆⬩

This thesis has brought to me
“one of those moments of perfect tiredness, of having conquered not only the work at hand
but the night *(and lockdown)* who had blocked the way…”

A global pandemic made for an interesting time for research, it both helped and burdened me. However, if this pandemic never happened this version of me and this version of this thesis would never have existed.

I took on this topic because of the exciting prospects AI is currently making in the field of radiology and orthodontics. I would be lying if I said I was fully confident about taking this project on, but I have learnt many valuable lessons (non-topic related) whilst compiling this.

⬩◆⬩

I dedicate this thesis to my brother and best friend, Shival Indermun,
who continues to teach *me* despite the order of our birth.
Thank you for always teaching me about the importance of hard work and
giving me the strength to persevere.
You taught me to question and solve problems using sound logic and scientific methods.
You can be a great source of inspiration when you aren’t annoying me.

⬩◆⬩

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS & ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
AI	Artificial intelligence
ANB	Angle formed by A-point, Nasion, B-point
ANS	Anterior nasal spine
Ceph	Cephalogram
Co	Condylion
DICOM	Digital Imaging and Communications in Medicine
Go	Gonion
Gn	Gnathion
ICC	Intra-class correlation
JPEG	Joint Photographic Experts Group
LLOA	Lower limit of agreement
Me	Menton
Mm	Millimetres
N	Nasion
Or	Orbitale
PNS	Posterior nasal spine
Po	Porion
SD	Standard deviation
SEM	Standard Error of the Mean
SNA	Sella–nasion–A point angle
SNB	Sella–nasion–B point angle
STROBE	Strengthening the Reporting of Observational studies in Epidemiology
ULOA	Upper limit of agreement

GLOSSARY

Accuracy: how close a sample estimate is to a gold standard or a true (or accepted) value, i.e. how nearly correct it is.

Artificial intelligence: branch of computer science assigned to the development of computer algorithms to accomplish tasks traditionally associated with human intelligence.

Cephalogram: profile radiograph of the skull and soft tissues, used to assess the relationship of the teeth in the jaws, the relation of the jaws to the skull and the relation of the soft tissues to the teeth and jaw

Cephalometry: a study of the measurements of the head using radiography

Deep learning: the subfield of representation learning which relies on multiple processing layers to learn representations of data with multiple layers of abstraction

Landmarks: an anatomical location on a lateral skull cephalogram used as a point of orientation or reference point in locating other structures

Machine Learning: part of research on AI that seeks to provide knowledge to computers through data, observations, and algorithms without being explicitly programmed

McNamara Analysis: cephalometric analysis composed of eight linear and three angular measurements; developed by Dr James A. McNamara

Precision: how close the sample estimates from *different samples are likely to be to each other*, i.e. the "spread" of the measurements or how close they are together

Random Error: an error in measurement caused by factors that vary from one measurement to another

Reference Frame: the coordinate system used whereby the origin, orientation and scale are defined by a set of reference points

Steiner Analysis: cephalometric analysis consisting of Skeletal, Dental and Soft Tissue Analysis; developed by Dr Cecil C. Steiner

Systematic Error: an error that is not determined by chance but is introduced by an inaccuracy (as of observation or measurement) inherent in the system

Wits Appraisal: Wits "appraisal of jaw disharmony" is a cephalometric analysis method whereby the severity or degree of anteroposterior jaw dysplasia is measured; it was established by Dr Alexander Jacobsen

CHAPTER 1 : INTRODUCTION

1.1 Background

In orthodontics, the role of oral maxillofacial radiology is integral to both treatment planning and monitoring (Machado, 2015). Traditional radiographic imaging modalities have included panoramic views, cephalograms (cephs), occlusals, bitewings and periapicals. These two-dimensional (2D) imaging techniques, whilst familiar to all, are not without limitations such as magnification, distortion and superimposition (Kapila *et al.*, 2011; Agrawal *et al.*, 2013; Noar and Pabari, 2013; Machado, 2015). Today, with the continuous advancement of imaging modalities; diagnosis and treatment planning has also been refined. The fairly recent introduction of cone-beam computed tomography (CBCT) to dentistry has transformed how orthodontists confirm diagnoses, develop and modify treatment plans, and monitor progress (Kapila *et al.,* 2011; Machado, 2015).

A game-changer has recently launched dental radiography into a new era. The use of artificial intelligence (AI) in dentistry is a concept that may border science-fiction. However, AI is a very genuine reality as it has now made its way from computer science through to health, dentistry and radiology (Durão *et al*., 2015; Tang *et al*., 2018; Tadinada, 2019).

Despite the adolescence of AI, orthodontists and radiologists have shown keen interest in using AI to detect cephalometric landmarks accurately. A fresh enquiry is underway involving whether AI could replace the conventional method of human detection of cephalometric landmarks (Lindner *et al*., 2016; Tang *et al.*, 2018; Park *et al*., 2019; Tadinada, 2019).

1.2 Motivation

The proposed research aimed to compare the precision of cephalometric landmark detection in automated systems and the conventional method of human examination. To the best of the researcher's knowledge, the literature regarding the precision and accuracy of automated cephalometric landmark detection is limited and needs further exploration. Therefore, the proposed comparative study intended to explore and describe the comparison of automated cephalometric landmark detection with human examination within a South African context.

According to the author's knowledge, no studies have been done using a South African population. Therefore, knowledge on this topic is desirable. It is the hope that coming soon to orthodontic practices is fully automated cephalometric landmark detection programmes that will assist workflow and improve treatment planning with increased precision and accuracy.

CHAPTER 2 : LITERATURE REVIEW

The following section presents the most recent literature, synthesizing the current knowledge and identifying the available methods with proven validity and relativity. The most widely accepted definitions of key concepts of AI and cephalometric landmarks has also been ascertained. By reviewing the existing literature, a better understanding of AI in cephalometric landmark detection can be acquired.

2.1 Field of Orthodontics

The field of orthodontics, as defined by Houston (cited in Mitchell *et al.*, 2011), is the "branch of dentistry concerned with the growth of the face, development of the dentition, and the prevention and correction of occlusal anomalies." Malocclusion and craniofacial anomalies have a direct effect on a patient's aesthetics, thus affecting the patient's quality of life. The field of orthodontics can either prevent or correct malocclusions to provide not only aesthetic advantages but can also boost self-esteem and improve function (American Academy of Oral and Maxillofacial Radiology, 2013).

Comprehensive examination and records of the craniofacial complex are imperative during orthodontic treatment. Records that are routinely taken include impressions for plaster models, radiographs and photographs. Radiographic imaging in orthodontics is crucial in all phases of treatment, i.e. diagnosis, treatment planning, growth assessment, and progress assessment. Imaging is used to evaluate the occlusal relationships, the growth of the craniofacial skeleton and the soft tissues (American Academy of Oral and Maxillofacial Radiology, 2013).

2.2 Cephalometry and its Applications

The customary extra-oral radiographs that are frequently prescribed in all orthodontic cases have included orthopantomograms/ panoramic views and lateral cephalograms (also referred to as lateral skull views) (Cattaneo *et al*., 2008; Kapila *et al*., 2011; Abdelkarim, 2012; Machado, 2015; Pereira *et al*., 2015).

Cephalometry is the study of craniofacial measurements in orthodontics and is used to assess the growth and development of the skull. Lateral cephalograms present the sagittal view of the skull, the soft tissue profile and dental structures. Traditional 2D cephalometry has been considered the "gold standard" diagnostic tool for evaluating craniofacial growth and skeletal deformities (Cattaneo *et al.*, 2008; Kapila *et al.*, 2011; Iannucci and Howerton, 2012).

Cephalograms serve two functions: (1) it displays dentoalveolar and skeletal relationships that cannot be otherwise viewed, and (2) it enables accurate monitoring of treatment progress and outcomes by comparing pre-, peri- and post-operative lateral cephalograms. This substantiates why lateral cephalograms are required for patients who have seemingly normal dental and skeletal relationships. Due to the years of usage, cephalometry in orthodontics has become second nature and many orthodontists have agreed that diagnosing and treating skeletal malocclusions devoid of cephalometric evaluation is a major blunder (Pereira *et al.*, 2015).

Cephalometric analyses identify specific anatomic landmarks, on both hard and soft tissues, that are joined to create lines and angles (de Lima Navarro *et al.*, 2013) (Figure 2.1). Determining the spatial and angular relationship between these landmarks enables the classification of the skeletal and dentoalveolar relationship (Cattaneo *et al.*, 2008; Manosudprasit *et al.*, 2017). Many cephalometric analyses utilize anatomic landmarks such as nasion, sella turcica, and basion, to obtain baselines such as sella-nasion, basion-nasion, and porion-orbitale. Over the years, numerous studies about cephalometry in orthodontics have been carried out, and thus normal values for the linear and angular measurements have been defined and established (Sadowsky, 2006), allowing cephalometric measurements of patients to be compared to the normal for age, gender and population group (Sadowsky, 2006; Nervina, 2012).

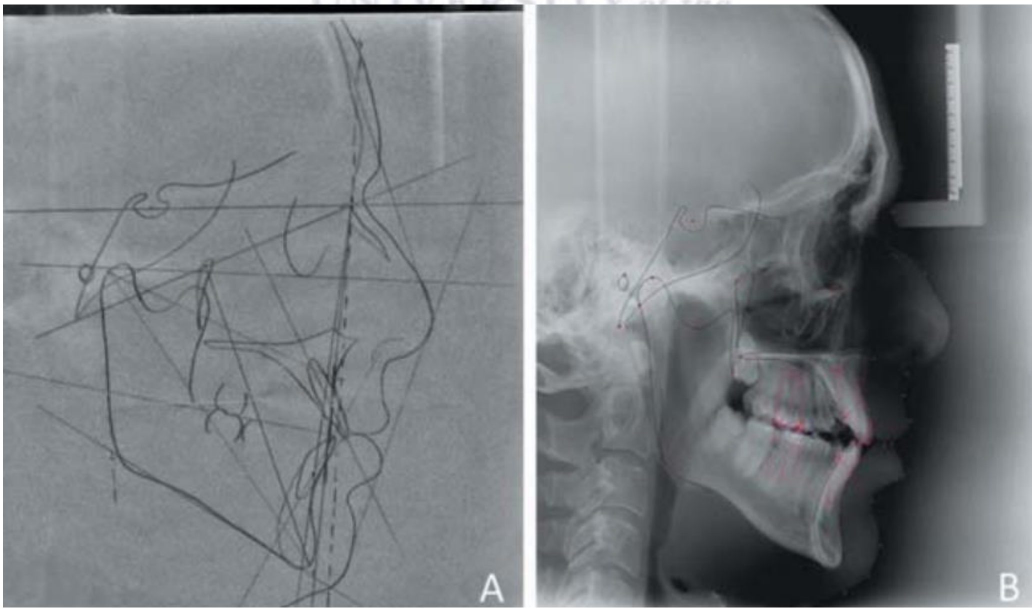


Figure 2.1: Examples of cephalometric tracings

The identification of specific anatomic landmarks, on both hard and soft tissues, are joined to create lines and angles. (A) Example of Manual tracing and (B) Digitized lateral cephalograms (de Lima Navarro *et al.*, 2013)

2.3 Shortfalls of Human examination

There are two ways to identify cephalometric landmarks: (1) manual approach and a (2) computer-aided approach (Leonardi *et al*., 2008; Miloro *et al.*, 2014) (Figure 2.1). The oldest and most widely used method is the manual approach, involving placing a sheet of tracing paper over the cephalometric radiograph, tracing prominent features, identifying landmarks, and making linear and angular measurements between landmark locations using a ruler, compass, and protractor (Leonardi *et al.*, 2008). The computerized cephalometric analysis can be done in two ways: (1) manual landmark detection, using a tracing of the radiograph to identify landmarks followed by the transfer of this tracing to a digitizer linked to a computer, or (2) direct digitization of the lateral cephalogram by scanning it into a computer and then locating landmarks on the monitor (Leonardi *et al.*, 2008). This computer method is still afflicted by inconsistencies caused by possible subjectivity in landmark identification. The accuracy and reproducibility of landmark identification using these different methods were studied extensively. However, the direct digitization of radiographs is reported to be the most reproducible, and therefore, the most accurate method, although the difference between methods is small and not statistically significant (Miloro *et al*., 2014).

The progression of manual cephalometry to computer assisted-cephalometric analysis is directed at improving the diagnostic value of cephalometric analysis by reducing any systematic or random errors and saving time. Random errors involve tracing, landmark identification and measurement errors. Identification errors are associated with landmark recognition. According to the literature, landmark detection is the major source of errors. The factors contributing to the detection error are examiner experience and subjectivity, landmark definition and interpretation, and the density and sharpness of the image (Ongkosuwito *et al.*, 2002).

Computer-assisted cephalometric analysis eliminates the mechanical errors when drawing lines between landmarks as well as those made when measuring with a protractor. However, the variation in landmark detection is still an important source of random errors both in computer-aided digital cephalometry and in manual cephalometric analysis. Both methods are time-consuming, thus resulting in efforts to automate cephalometric analysis, improving the accuracy of landmark identification and reducing the errors due to clinicians' subjectivity (Leonardi *et al.*, 2008).

Inconsistent and inaccurate landmark identification may have a ripple effect potentially resulting in inaccurate diagnoses and treatment plans. Detection of certain anatomical landmarks, such as the porion (Po), condylion (Co), orbitale (Or), basion, gonion (Go), anterior nasal spine (ANS), posterior nasal spine (PNS), and lower inferior apex (LIA), may be more susceptible to error due to overlapping structures superimposed on the landmark and its location. Similarly, the quality of radiographic images can interfere with the identification of some landmarks, such as Po, Co, Or, ANS, point B, the pogonion (Pog), Go, and the glabella (Durão *et al.*, 2015).

Researchers have also proposed that the level of an examiner's knowledge and his or her professional background play an important role in landmark identification (da Silveira and Silveira, 2006; Durão *et al.*, 2015). As put forth by Halazonetis (1994), reducing errors related to landmark identification is difficult and calls for a thorough definition of the anatomic landmarks, detailed knowledge of radiographic anatomy and cephalograms of high quality. A study to compare the accuracy of orthodontists and maxillofacial radiologists in identifying 17 commonly used cephalometric landmarks was carried out (Durão *et al.*, 2015). Gnathion (Gn) point was the least reliable landmark for orthodontists, while the least reliable landmark for maxillofacial radiologists was orbitale (Or). The least consistent was the condylion (Co)-Gn plane. It was established that the most consistently identified landmark in both groups was the lower incisor border, while the least reliable points were Co, Gn, Or, and the anterior nasal spine. Overall, a lower level of reproducibility in the identification of cephalometric landmarks was observed among orthodontists (Durão *et al.*, 2015). Whilst this study makes no mention of AI, it offers important insights into the potential use of AI in ensuring an accurate and reproducible method of cephalometric landmark detection.

2.4 The Modern Solution: Artificial Intelligence in Radiology and Cephalometry

Currently, in clinical practice, cephalometric landmarks are identified manually or semi-automatically which can be tedious, time-consuming and prone to subjectivity within and across orthodontists and radiologists. *Inter*-examiner variations may be impacted by the levels of orthodontic training and experience, whilst *intra*-examiner consistency can be affected by time constraints and other commitments (Lindner *et al.*, 2016)

.

Two-dimensional (2D) imaging techniques, whilst familiar to all, are not without limitations such as magnification, distortion and superimposition (Kapila *et al.*, 2011; Agrawal *et al.*, 2013; Machado, 2015). Yet traditional 2D cephalometry is still considered the “gold standard” diagnostic tool for evaluating craniofacial growth and skeletal deformities. CBCT-generated cephalograms have also grown in popularity and provided orthodontists with an anatomically representative visualization of the craniofacial complex. This is especially beneficial in complex cases involving clefts, craniofacial disorders, or orthognathic cases, where traditional 2D cephalograms are no longer considered optimum (Cattaneo *et al.*, 2008; Kapila *et al.*, 2011; Iannucci and Howerton, 2012). However, it is important to reiterate that whilst the radiation exposure of a CBCT scan has an advantageous lower dose than that of a multi-slice CT scan; it is not recommended for the common and ordinary orthodontic case (Cattaneo *et al.*, 2008; Kapila *et al.*, 2011; Iannucci and Howerton, 2012). As such, monitoring the progress of orthodontic treatment with CBCT would not be realistic or ethical as a large FOV (field of view) would be required and a greater amount of radiation exposure in comparison to the 2D counterpart (Scarfe and Farman, 2008; Gribel *et al*., 2011). In addition, de Oliveira *et al.* (2009) states that using 3D landmark identification is more time-consuming than using conventional 2D cephalometry (cited in Shahidi *et al*., 2014).

CBCT was considered the key to accurate landmark detection, but it is not viable for everyday orthodontic cases. Furthermore, despite the current use of CBCT scans, the ability to accurately identify cephalometric landmarks has not markedly improved over conventional lateral cephalograms (Miloro *et al.*, 2014). What is trending now is something that orthodontists and radiologists could never have imagined. Today, the solution to determine accurate cephalometric landmark detection lies in artificial intelligence (Durão *et al*., 2015; Lindner *et al*., 2016; Tridandapani, 2018; Park *et al.*, 2019; Hwang *et al.*, 2019; Tadinada, 2019).

2.5 An introduction to Artificial Intelligence (AI)

AI is the “simulation of human *intelligence* processes by machines, especially computer systems”. It has been defined as the branch of computer science involving the development of computer algorithms to achieve tasks conventionally associated with human intelligence (Deshmukh, 2018; Tang *et al.*, 2018; Yaji *et al*., 2019).

As in any field, there is specialized vocabulary related to it. In radiology, special terms are used to describe findings and communicate them to others in the field. As AI technology continues to grow, it is now predicted to become part of clinical workflow and radiologists will be expected to become familiar with the terminology and underlying concepts. The hierarchy of AI fields is demonstrated in Figure 2.2.

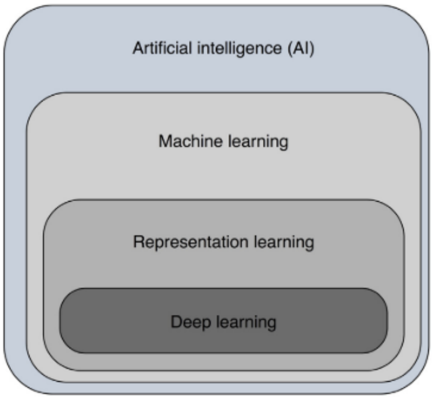


Figure 2.2: Diagram illustrating the hierarchy of AI fields
(Tang *et al.*, 2018)

AI is a broad term that designates a variety of fields and techniques. By using large groups of data and a numerous amount of pre-populated clinical scenarios processed by AI, machines can provide diagnostic suggestions, sometimes more accurately than humans. This is possible because of a remarkable branch of AI called "machine learning" (Tadinada, 2019). Machine learning (ML) refers to the "part of research on AI that seeks to provide knowledge to computers through data and observations without being explicitly programmed" (Tang *et al.*, 2018). Machine learning utilizes algorithms that can learn and make predictions on data and in doing so teaches computers to do what comes naturally to humans and animals—learn from experience and retain memories to construct a response or reaction. This is what most radiologists do by creating large memory banks of cases they have seen and applying their knowledge of pathophysiology to arrive at a differential diagnosis. Information is directly learnt from data using algorithms and computational methods. Just as a radiologist's knowledge improves over time and with experience, the algorithms in ML adapt and improve their performance as the samples available for learning increases. The diagnostic output improves significantly and is therefore directly proportional to the number of interactions, experiences, and patient scenarios (Tadinada, 2019).

Representation learning refers to a subtype of ML in which the computer algorithm learns the features required to classify the provided data. Machine learning has also expanded to "deep learning," inspired largely by the way the human nervous system works, albeit on a simpler scale (Tadinada, 2019). Deep learning refers to "a subfield of representation learning which relies on multiple processing layers (hence, deep) to learn representations of data with multiple layers of abstraction" (Tang *et al.*, 2018; Yaji *et al.*, 2019). It refers to learning through the use of a cascade of layers that can teach the machine to make decisions based on data representations, as opposed to task-specific algorithms (Tadinada, 2019).

2.6 Radiology and AI

Radiology is a discipline that has always been at the forefront of technology (Dreyer and Allen, 2018; Sana, 2018; Tadinada, 2019), yet is highly dependent on visual skills, retaining knowledge and images and consequently the formation of an individual's memory banks. Radiologists are primarily known for their image interpretation skills (Tang *et al.*, 2018), but a major concern is that identification and reporting has been subjective (Leonardi *et al.*, 2008; Lindner *et al.*, 2016; Rozylo-Kalinowska, 2018). The famous quote "the eye sees only what the mind is prepared to comprehend" is the best description of radiology (Rozylo-Kalinowska, 2018). Without knowledge and experience, a radiologist is limited in what they can see, interpret and diagnose. Over time, as knowledge increases, radiologists gain experience in identifying structures and reporting becomes more refined and adept. Yet along with time passing by, routine also takes over accompanied by fatigue, ageing, and even professional burnout. This may negatively influence the quality of reporting. A question is now being raised as to whether AI can be the solution (Tang *et al.*, 2018; Tridandapani, 2018).

Whilst artificial intelligence was thought to be restricted to science-fiction, it has now become a reality, and radiology has welcomed this new novice. Many have considered it to be a threat to humans, but to others, it can be seen as an enhancement of skills. The purpose of AI in radiology is not to replace radiologists but rather to aid radiologists and orthodontists in their daily routines (Tridandapani, 2018). This technology can improve the accuracy and efficiency of diagnostics resulting in a better patient outcome. AI has the potential to change the landscape of clinical practice and scientific research. Even more advantageous, is that it can assist in creating faster turnaround times in radiology practices (Rozylo-Kalinowska, 2018). Implementation of AI in radiology is expected to considerably transform clinical workflows and patient care. Thus, a radiologist must be aware of AI and its applications in their field (Tang *et al.*, 2018; Yaji *et al.*, 2019).

2.7 Automated Cephalometric Landmark Detection

As mentioned previously, cephalometric tracing is the standard analysis tool for orthodontic diagnosis and treatment planning. Having discussed the importance of cephalometry and the introduction of AI into radiology, the final section of this review addresses the use of automated cephalometric landmark detection.

Automated cephalometric landmark identification would greatly assist in overcoming time constraints and inconsistencies within and across examiners (Ongkosuwito *et al*., 2002; Shahidi *et al*., 2014; Durão *et al.*, 2015; Lindner *et al*., 2016). Lindner *et al.* (2016) proposed that if a computerized system was able to accurately locate cephalometric landmarks then this would have the potential to significantly improve the clinical workflow in orthodontic treatment. In an investigation into this topic, Lindner *et al.* (2016) set out to develop and validate a fully automatic landmark annotation (FALA) system (BoneFinder®) for identifying cephalometric landmarks in lateral cephalograms. The IEEE (Institute of Electrical and Electronics Engineers) International Symposium on Biomedical Imaging (ISBI) Grand Challenges organized a challenge on automated landmark detection in cephalograms. Preliminary results of the approach by Linder *et al.* (2016) were presented at the 2015 ISBI Grand Challenge in Dental X-ray Image Analysis. Their system was awarded the first prize. Their methodology achieved an average point-to-point error of 1.66 mm compared to errors ranging from 1.85 mm to 2.85 mm for all other techniques, demonstrating that their method performed significantly better than any of the other six techniques ($p < 0.0001$).

The FALA system follows a machine learning approach. Digital cephalograms of 400 subjects (age range: 7–76 years) were used and all cephalograms had been manually traced by two experienced orthodontists with 19 cephalometric landmarks, and eight clinical parameters had been calculated for each subject. The system was evaluated via comparison to the manual tracings. The system achieved an average point-to-point error of 1.2 mm, and 84.7% of landmarks were located within the clinically accepted precision range of 2.0 mm. It is important to note that some researchers have suggested that landmark detection errors of less than 1 mm are clinically acceptable. It has also been proposed that errors of less than 2° or 2 mm would most likely not affect treatment (Miloro *et al.*, 2014; Durão *et al.*, 2015).

The automatic landmark localization performance was within the inter-examiner variability between two clinical experts. The automatic classification achieved an average classification accuracy of 83.4% which was comparable to an experienced orthodontist. It was concluded that the FALA system accurately identifies cephalometric landmarks in lateral cephalograms, and has the potential to significantly improve the clinical workflow in orthodontic treatment (Lindner *et al.*, 2016).

The newest deep learning technique based on the You-Only-Look-Once version 3 algorithm (YOLOv3) recently recognized 80 landmarks in the field of automated cephalometric landmark recognition, resulting in not only more accurate but also faster detecting performance (Park *et al.,* 2019). The You-Only-Look-Once version 3 (YOLOv3) and Single Shot Multibox Detector (SSD) techniques were used to evaluate the accuracy and computational efficiency of two of the most recent deep-learning algorithms for automatic detection of cephalometric landmarks.

Following the results, of this study, Hwang *et al.* (2019) set out to compare detection patterns of 80 cephalometric landmarks identified by an automated identification system (AI) based on a recently proposed deep-learning method, the You-Only-Look-Once version 3 (YOLOv3) with those identified by human examiners. With custom modifications, the YOLOv3 algorithm was executed and trained on 1028 cephalograms. A total of 80 landmarks, consisting of two vertical reference points and 46 hard tissue and 32 soft tissue landmarks, were identified. On the 283 test images, the same 80 landmarks were detected by human examiners and AI twice. Statistical analyses were performed to detect whether any significant differences between AI and human examiners existed. AI consistently recognized identical positions on each landmark in repeated testing, but human intra-examiner variability of repeated manual detections revealed a detection error of 0.97 -1.03 mm. Between AI and humans, the mean detection error was 1.46 - 2.97 mm. Human examiners had a mean difference of 1.50 - 1.48 mm. The detection errors of AI and human examiners were often less than 0.9 mm, which did not appear to be clinically significant. It was concluded that AI was comparably accurate in the identification of cephalometric landmarks. The AI system always detected identical positions, upon repeated trials. This holds the promise that AI might be a more reliable option for repeatedly identifying multiple cephalometric landmarks (Hwang *et al.,* 2019).

Without question, AI appears to have a bright future ahead as a potentially “game-changing” tool in healthcare. Whilst some fear still exists regarding this overwhelming, the trajectory at which AI is changing the field of radiology now warrants more research and insight. This has led to the Radiology Society of North America (RSNA) Congress in the USA to introduce a new journal called “Radiology: Artificial Intelligence”. It appears almost inevitable that AI will be introduced not only to the diagnostic side of radiology but also to assist in triaging radiological investigations. It is likely in the near future, AI will be introduced into radiology practice and included in radiology training curricula (Pakdemirli, 2019).

What these studies provide is an exciting opportunity to advance our knowledge in automated cephalometric detection. The findings of the proposed research could assist in more precise location of landmarks and making sure that automated cephalometric systems make their way into orthodontic and radiology practices very soon.

2.8 Conclusion

Radiology has always been at the forefront of technology (Dreyer and Allen, 2018; Sana, 2018; Tadinada, 2019): not only has this discipline mastered the digitization of medical imaging and picture archiving and communications systems (PACS), it has also made what may have seemed impossible 30 years ago a reality. Many radiologists have not seen film radiographs in over a decade, and there are those being trained who may wonder what that is. Today, studies with hundreds of images can be easily transmitted across a hospital, a city, a country, or across the world within seconds. According to Tridandapani (2018), if PACS was our end goal, then we have arrived, and there are no more evolutionary hurdles to cross in radiology. However, Tridandapani (2018) reminds us that we should always be asking ourselves: "What do we do? How do we do it? Why do we do it? And how can we do it better?" The answer to these questions lies within AI. And it seems radiologists have opened our arms to this exciting tool (Sana, 2018; Tridandapani, 2018). It is time for orthodontists to do the same. It is hoped that the profession will take an interest in and embrace the potential of AI (Sana, 2018). The literature on artificial intelligence in healthcare and particularly radiology has only just begun and the future of AI in this field looks promising. Whether it is for locating landmarks or detecting lesions, AI has the potential to detect what the human grayscale cannot discern (Tadinada, 2019).

In the current section, the use of AI in cephalometric landmark detection was reviewed. The scope of the proposed research intends to provide a means for easy and precise detection of cephalometric landmarks within a South African context. This is to substantiate the benefit of implementing fully automated cephalometric landmark detection programmes in orthodontic practices that will ultimately assist with workflow and improve treatment planning with increased precision. The next chapter will discuss the aims and objectives of the study.

CHAPTER 3 : AIMS AND OBJECTIVES

In the previous section, studies were compared showing cephalometric landmark detection methods and were found to be limited. This next section provides an overview of the aims and objectives of the current study.

3.1 Aim

This study aimed to determine the precision of two cephalometric landmark identification methods, namely an artificial intelligence programme (BoneFinder®) and a computer-assisted human examination software (Dolphin Imaging™).

3.2 Objectives

1. To calibrate the main researcher by obtaining cephalometric landmark consensus from two experienced observers using Dolphin Imaging™ software
2. To determine *x*, *y* coordinates for 19 cephalometric landmarks for the entire sample using computer-assisted human examination approach with Dolphin Imaging™
3. To determine *x*, *y* coordinates for the 19 cephalometric landmarks for the entire sample using artificial intelligence software (BoneFinder®)
4. To calculate and compare the Euclidean distance between the computer-assisted/human plot and artificial intelligence plot thereby determining the **precision**
5. To suggest an opinion on the use of AI in cephalometric analysis.

3.3 Research Question

What is the difference in precision between cephalometric landmark detection in artificial intelligence and computer-assisted human examination?

3.4 Null Hypothesis

There is no significant difference between the artificial intelligence programme (BoneFinder®) and the computer-assisted human examination (Dolphin Imaging™) regarding the precision of landmark detection**.**

CHAPTER 4 : MATERIALS AND METHODS

This chapter encompasses the methodology that was utilized in the current study. The programmes used for landmark detection were selected based on availability and cost.

4.1 Research Design

This was a retrospective quantitative cross- sectional analytical study.

4.2 Study Population

The study population consisted of retrospective cephalograms of patients who required orthodontic treatment and presented at the Diagnostics and Radiology Department of the Faculty of Dentistry, Tygerberg Oral Health Centre, University of the Western Cape, Cape Town, South Africa.

4.3 Sample selection process and size

Cephalograms were retrieved from current records (the study starting date – 03 April 2020) and were backdated until the study sample was reached. The sample size was confirmed by the statistician after a sample determination test was carried out. An initial search within the database resulted in a convenient sample frame total of 1818 cephalograms, obtained from January 2016 to March 2020. Only a single time point cephalogram (i.e., pre-operative) was selected from each patient, resulting in the exclusion of repeat cephalograms. The preliminary refinement resulted in the exclusion of 517 cephalograms. A second refinement led to a final sample size of 409 cephalograms, after strict adherence to the inclusion and exclusion criteria set below. Since this was a retrospective study, no new cephalograms were specifically taken for the study and no patient was exposed to unnecessary radiation to fulfil the sample size requirements relating to the study.

Utilizing systematic random sampling techniques, cephalograms were selected for the inter- and intra- reliability tests. With a final sample size of 409 (N), the required cohort for the inter-rater reliability tests was 10 (n). The interval size was calculated as N/n = 409/ 10 = 40. Therefore, every 40th cephalogram was selected to obtain the required 10 cephalograms for the inter-rater reliability tests. The inter-observer agreement was carried out by the primary researcher, an experienced chief radiologist and an experienced orthodontist. The chief radiologist was a dentist with a MSc degree in Oral and Maxillofacial Radiology and the Head of the Oral and Maxillofacial Radiology Department. An MSc degree in Oral and

Maxillofacial Radiology is the highest qualification in South Africa. Both the chief radiologist and orthodontist had over 10 years of experience at the time of this study.

The intra-reliability test was carried out on 40 randomly selected cephalograms. This was calculated by using 10% of the overall sample. In this case, every 10th cephalogram was selected (N/n = 409/10 = 40).

4.4 Inclusion and Exclusion Criteria

The criteria for selecting the cephalograms are summarized in Figure 4.1 and described below.

4.4.1 Inclusion Criteria

1. Cephalograms of patients requiring orthodontic treatment, but no evidence of current orthodontic treatment
2. Cephalograms of patients with no missing permanent incisors or first molars
3. Cephalograms of patients in occlusion
4. High-resolution cephalograms of adequate diagnostic quality (sharp and free of distortion).
5. Cephalograms of patients with no unerupted or supernumerary teeth overlying areas of interest
6. Cephalograms with correct cephalostat placement

4.4.2 Exclusion Criteria

1. Cephalograms of patients with gross skeletal asymmetries and genetic anomalies in the area of study
2. Cephalograms with distortion, artefacts and superimposed metal objects in the area of study
3. Cephalograms without a cephalostat

IDENTIFICATION

Records identified through database searching:

Cephalograms of patients between January 2016 to March 2020

Assessed for eligibility N = 1818

SCREENING

Excluded/Ineligible: n = 1409

- Repeat cephalograms (only single time points cephalograms were included, i.e., pre-operative)
- Patients with fixed orthodontic appliances
- Magnification error
- Patient positioning errors
- Impacted teeth overlying ANS
- Displaced central incisors
- Missing mandibular 6's
- Patient not in occlusion
- Surgical screws and plates overlying landmarks
- Absence of cephalostat
- Broken cephalostat

SELECTION FOR RELIABILTY

Intra-rater Reliability Sample (10% of sample)

10% of 409 = 40.9 ≈ 40

N/n = 409 /10 = 40

**Every 10th cephalogram was selected.*

Inter-rater Reliability Sample: 10 cephalograms

N/n = 409 /40 = 10

**Every 40th cephalogram was selected.*

ELIGIBILTY & INCLUSION

Total eligible cephalograms:

Final sample: N = 409

Figure 4.1: Flow Diagram depicting sample selection

4.5 Instruments and machines

All cephalograms were acquired in DICOM format with the Orthophos XG 5 machine (Dentsply Sirona, Germany) using Sidexis software (version 4.3). The image resolution was 1280 x 1024 pixels. The retrospective cephalograms were acquired by positioning the patients in a cephalostat in a natural head position. The cephalometric modality has a left-sided arm with a digital line sensor with CCD technology. The active sensor area is 2.30 – 6.48 mm. The pixel size is 0.027 mm and the focus-sensor distance is 1 714 mm. The retrospective cephalograms were taken by either experienced and trained radiographers or undergraduate students at UWC Dental Faculty, Tygerberg, with compliance to the manufacturer's instructions. The radiographs were taken under routine daily conditions and the head positions were standardized with conventional cephalogram techniques using a cephalostat.

The software used to conduct the computer-assisted human examination of cephalometric landmarks was Dolphin Imaging™ 11.95 Service Pack 2 (Patterson Dental Supply, Chatsworth, California, USA) (Appendix A).

The artificial intelligence software, BoneFinder® (University of Manchester, England) is freely available online for research purposes and was downloaded from: https://www.click2go.umip.com/i/s_w/Biomedical_Software/Bonefinder.html. The license was activated on 4 October 2019. It was provided under licence no: MAN_002-3494548-v2-UMIP Annual Research Licence C2G final 1.00. The licence was valid for 12 months and expired on 4 October 2020. Each cephalogram was uploaded to the programme, after which the landmarks were automatically determined.

The landmark detection and inter- and intra- examiner reliability tests were accomplished using a Dell® Inspiron 3580 8th Generation laptop comprising of an Intel Core i7-8565U CPU @ 1.80GHhz, 16GB RAM, 256GB Ultra-Fast SSD and 1TB Hard Drive, with a 15.6" FHD 1920 x 1080 anti-glare display monitor, 64-bit operating system, x64-based processor, running Windows® 10 Home, © 2019 Microsoft Corporation.

4.6 Data Collection

The radiology database was reviewed and cephalograms meeting the inclusion criteria were selected for the analysis. Patient demographic information was recorded as part of the data collecting process with retrieval of cephalograms in DICOM and JPEG format.

This study evaluated 19 landmarks chosen to represent common structures in cephalometric analyses like the Steiner Analysis and Wits Appraisal (Appendix B, C, D) (Lindner *et al.*, 2016; Meric and Naoumova, 2020). Instructions to examiners were also provided (Appendix E).

Cephalometric landmarks were identified on the conventional 2D digital cephalogram using human examination on computer-aided cephalometric analysis software, Dolphin Imaging™, and the artificial intelligence programme, BoneFinder®.

To prevent operator bias, the 19 landmarks were first identified on the conventional 2D digital cephalograms using the computer-aided cephalometric analysis software Dolphin Imaging™. Technical support and training were provided to the researcher by a Dolphin Imaging™ technician.

4.7 Landmark Detection

This section explains how landmarks were detected using Dolphin Imaging™ and BoneFinder®. The methods used in Dolphin Imaging™ (Appendix F) is presented first, followed by BoneFinder® (Appendix G).

4.7.1 Landmark Detection using Dolphin Imaging™

The primary researcher uploaded the entire sample of cephalograms to Dolphin Imaging™ prior to landmark detection. A customized cephalometric analysis (named "19 Landmarks") was created by the primary researcher to include the study's intended landmarks (Appendix F). The ruler length was set at 30mm, to represent the real distance length of the fixed corner points of the nasion-guiding rod. This was done as there was no ruler used during the acquisition of the cephalograms.

The mouse-driven cursor was used to detect landmarks. Its location was indicated by a red dot displayed on the monitor. The placement of the landmark could be adjusted until the operator was satisfied. To better visualize structures of interest, the researcher and inter-examiners could utilize any of the software's image-enhancing capabilities (e.g. magnifying glass). The definitions described in this study were used and not those that automatically appear in Dolphin™. When bilateral structures were involved, landmarks on the patients' right side was only identified. The right side was chosen because the right and left sides would be a repetitive estimate of a single landmark. All landmark identification sessions were conducted in a darkly lit room, with no interruptions, for as long as each examiner required. To ensure standardization, the same operator (i.e. the primary researcher) detected all landmarks for the human approach. After being calibrated, the recordings made by the primary researcher were taken as the manual ground truth.

4.7.2 Landmark Detection using BoneFinder®

The above process, described in section 4.7.1, was repeated with the BoneFinder® software (Appendix G). The cephalogram was imported into the programme and the search button was selected to automatically detect the landmark points. Since automatic detection systems are deterministic i.e., the same image will yield the same result every time, the landmarks attained by BoneFinder® were then compared to the manual ground truth

The *x* and *y* coordinates were extracted from each cephalogram from each programme (Dolphin Imaging™ and BoneFinder® and saved into an Excel sheet (Microsoft, Seattle, WA) (Appendix H and I). The coordinates were saved in millimetres (mm).

4.8 Data Analysis

The next section describes how the data was analysed and includes a description of validation and describes the statistical analysis.

4.8.1 Criteria for Validation

According to Hwang *et al.* (2019) "when it comes to a reliability measure when identifying a certain cephalometric landmark, there is no firm 'ground truth' or gold standard that can provide validation as to where the true location of the landmark is".

The landmarks were calibrated after inter- and intra-reliability tests were conducted, to reach a consensus landmark for each point. Using the Dolphin Imaging™ software for 2D cephalometric images, the same three examiners digitally identified the same landmarks. This was conducted by the primary researcher (1st examiner), the chief radiologist (2nd examiner) and an experienced orthodontist (3rd examiner).

The inter-examiner reliability tests were conducted using 10 random cephalograms. The intra-reliability test was carried out on 40 randomly selected cephalograms. Both inter- and intra-reliability tests were done at two intervals, 2 weeks apart.

4.8.2 Statistical Analysis

Data analysis was discussed with a statistician. Statistical tests were performed per the study by Katkar *et al.* (2013).

The Euclidean distance is "the square root of the sum of squared coordinate differences between the two selected landmark positions." The Euclidean distance was calculated for each pair of observations (either the duplicate measures made by a single observer or the measures of the same landmark by the three different observers). The Euclidean distance is defined by

$$Euclidean\ Distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \qquad 4.1$$

where (x_1, y_1) and (x_2, y_2) are the coordinates of the two selected landmarks from each program.

Descriptive statistics were determined for these Euclidean differences, and the differences in the distribution of Euclidean differences between the two software programs were evaluated using the Wilcoxon rank-sum test. R Core Team (2013) was used to compare the two methods (BoneFinder® and computer-assisted human examination with Dolphin Imaging™) for each measurement. For the inter and intra-class correlation tests, the ICC reliability Calculator was used (Mangold International Germany, LabSuite version 2015, Program version 1.5.) All measurements were recorded onto a Microsoft Excel spreadsheet (Appendix H and I).

The mean, standard deviation, minimum and maximum of the detection errors from AI and differences between the human examiners was determined. The mean difference between AI and human examination was also determined. Differences were considered significant at $P < 0.05$.

4.9 Ethical Considerations

Permission was obtained from the Dean of the Faculty of Dentistry, and the Head of the Department of Diagnostics and Radiology to analyze and use the cephalograms taken at the faculty. Permission to access these records was requested via a letter to the Dean's office and Head of the Department of Diagnostics and Radiology (Appendix J - L).

The anonymity of all patients was ensured by allocating record numbers to the cephalograms and data files were deidentified by use of specialty software. Demographic data and file numbers were captured on a Microsoft Excel spreadsheet (Appendix M). This data will be kept by the primary researcher in a secured location via a password protected PC. The electronic data will be stored for 5 years and thereafter deleted. The results obtained from the study will be used for educational and research purposes only. No conflict of interest has been reported.

A record number was assigned to the file number as well as other personally identifiable patient information (names, dates of birth, gender, etc.). On the data collection form, only the former was noted. The data identifying the record number of a patient remained anonymous. This number was used for record purposes only and was only kept for the duration of the study. Patient records were stored on a password-protected computer and printed information was stored in a locked office. Cephalograms investigated in this study were de-identified and did not jeopardise patient identity. Backup of the data was conducted periodically using the primary researcher's student account on Google Drive and a portable external hard drive, WD Elements Portable 4TB USB 3.0.

This mini-thesis proposal was presented to the Faculty of Dentistry of the University of the Western Cape Research Committee and was approved by the Senate Research Ethics Committee (approval number: BM19/10/3) of the University of the Western Cape (Appendix N).

4.10 Budget

This was a self-funded research project.

4.11 Research Deliverables

The proposed research intended to provide a means for easy and precise detection of cephalometric landmarks within a South African context. This was to substantiate the benefit of implementing fully automated cephalometric landmark detection programmes in orthodontic practices that will ultimately assist with workflow and improve treatment planning with increased precision.

4.12 Summary

This chapter summarized the methodology that was utilized in the current study. The computer-assisted human examination approach was carried out using Dolphin Imaging^{TM} software. The AI program used was BoneFinder®. The next section displays the results attained during the study.

CHAPTER 5 : RESULTS

Having described the methodology of the study, the results are now presented in the sections below.

5.1 Demographic Data

Of the final cohort, 57.94% (n = 237) were female and 42.05% (n = 172) were male. The mean age of the patients was 15.78 years; with the minimum age being 7 years and the maximum age being 40 years. The median age was 14 years (Table 5.1).

The race of the patients was as follows: 0.49% were Asians, 9.78% were Black, 59.66% were Coloured/ Mixed race; 2.68% were Indians, 16.14% were Caucasian. 11.24% of cases did not have the race specified on the medical record data (Table 5.2).

Table 5.1: Study Population Characteristics

Maximum Age	40
Minimum Age	7
Mean Age	15.78
Median Age	14
Gender Ratio (Male: Female)	1: 1.38

*Age in years

Table 5.2: Study Demographics

	No of records	Percentage (%)	Race	No of Females	No of Males
	2	0.49	Asians	1	1
	40	9.78	Black	25	15
	244	59.66	Coloured	124	120
	11	2.68	Indians	10	1
	66	16.14	Caucasian	49	17
	46	11.24	Not specified	28	18
Total	**409**	**100**	**-**	**237**	**172**

5.2 Intra-examiner Assessment

To ensure the reliability of the measurements, the primary researcher carried out intra-reliability tests twice with a two-week interval. No more than 20 cephalograms were examined in a single session to minimize errors due to examiner fatigue. 10 cephalograms were viewed in the morning and 10 in the afternoon resulting in two sessions a day. Therefore, landmarking using the computer-assisted human examination approach with Dolphin was carried out in twenty-one sessions. These results were assessed with Pearson's product-moment correlation r two-sided, true correlation ≠0 (non-zero) with their p-values to test for association between the paired samples for each landmark from interval 1 versus interval 2. The results are summarized in Table 5.3.

Table 5.3: Intra-examiner tests - Interval 1 versus Interval 2

Landmark	X co-ordinate	Y co-ordinate
1	r= 0.973360 p-value: 2.88	r= 0.944142 p-value: 2.00
2	r= 0.931418 p-value: 7.99	r= 0.985348 p-value: 5.01
3	r= 0.877407 p-value: 2.32	r= 0.955702 p-value: 3.02
4	r= 0.914413 p-value: 4.15	r= 0.800422 p-value: 9.59
5	r= 0.899096 p-value: 7.65	r= 0.951787 p-value: 1.40
6	r= 0.960221 p-value: 4.29	r= 0.934161 p-value: 3.84
7	r= 0.969004 p-value: 4.57	r= 0.914998 p-value: 3.68
8	r= 0.970427 p-value: 1.94	r= 0.921161 p-value: 9.64
9	r= 0.971155 p-value: 1.23	r= 0.923846 p-value: 5.19
10	r= 0.965452 p-value: 3.30	r= 0.839113 p-value: 2.5
11	r= 0.954903 p-value: 4.18	r= 0.947352 p-value: 6.87
12	r= 0.954973 p-value: 4.06	r= 0.952481 p-value: 1.07
13	r= 0.925082 p-value: 3.88	r= 0.931686 p-value: 7.45
14	r= 0.936034 p-value: 2.29	r= 0.942169 p-value: 3.73
15	r= 0.924609 p-value: 4.34	r= 0.963968 p-value: 7.10
16	r= 0.966329 p-value: 2.06	r= 0.930671 p-value: 9.70
17	r= 0.939087 p-value: 9.51	r= 0.884146 p-value: 8.65
18	r= 0.875362 p-value: 3.09	r= 0.962968 p-value: 1.16
19	r= 0.975206 p-value: 7.75	r= 0.873159 p-value: 4.20

5.3 Inter-examiner Assessment

To control for bias and adequate calibration of the primary researcher (examiner 1), landmark detection was carried out by two other individuals: (1) the chief radiologist (examiner 2) and (2) an experienced orthodontist (examiner 3) (Figure 5.1 and 5.2).

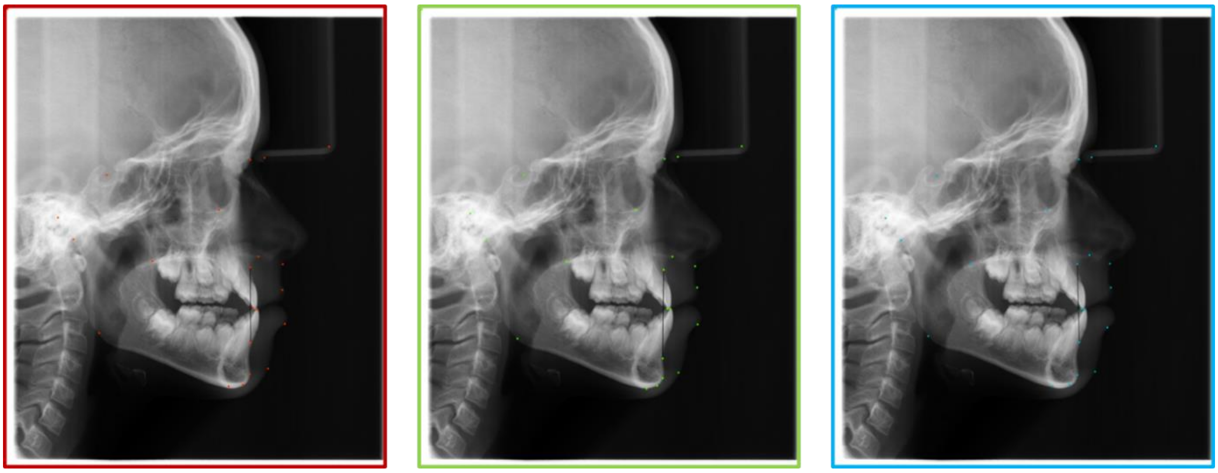


Figure 5.1: Example of identical cephalogram used for the inter-examiner reliabilty test.

Landmark detection by the primary researcher (red); chief radiologist (green) and an orthodontist (blue) using Dolphin™ software.

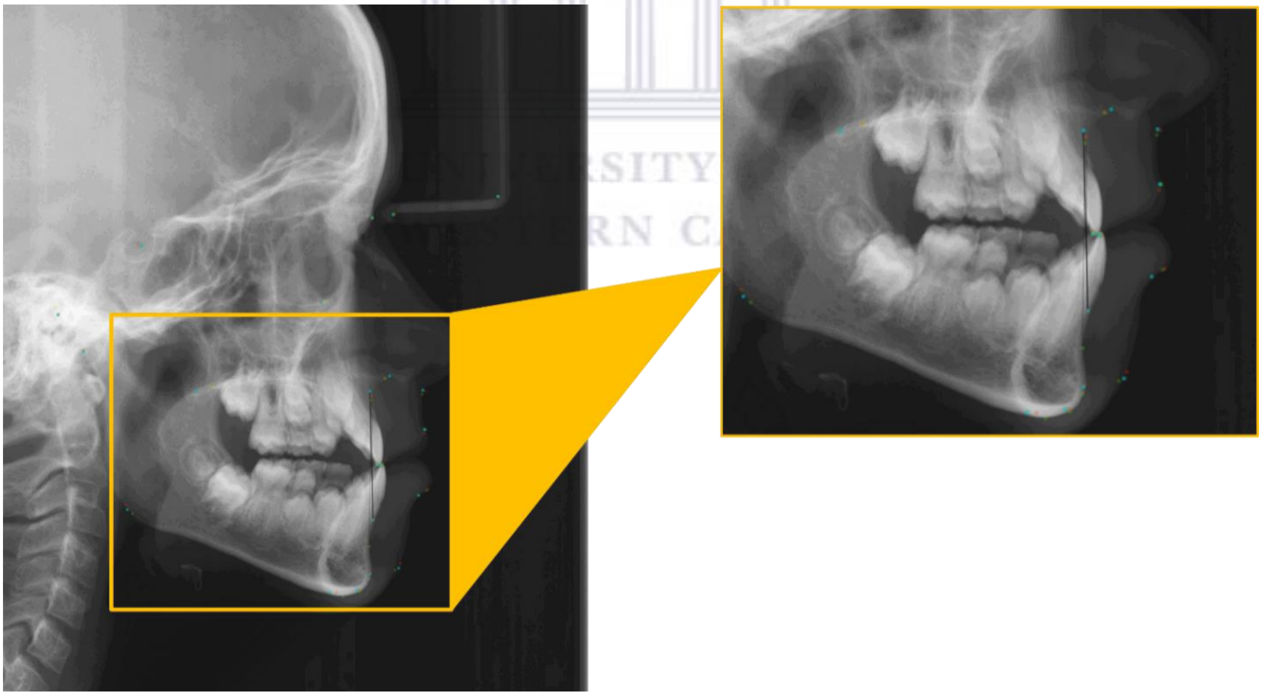


Figure 5.2: Superimposition of 3 cephalograms

Left - Superimposed cephalograms with all 3 observers' landmarks for the same patient. Red- primary observer, green – chief radiologist, blue – orthodontist, yellow -coinciding landmarks. Zoomed in image depicting proximity of landmarks.

The inter-reliability tests were carried out on 10 cephalograms twice with a two-week interval. The primary researcher acted as the control value for the inter-class correlation. The ICC was calculated by Equation 5.1.

$$ICC = \frac{variance\ of\ interest}{variance\ of\ interest + unwanted\ variance} \quad \mathbf{5.}1$$

For each coordinate (*x* and *y*) of each landmark, 2 mm was taken to be acceptable to represent concurrence of examiner 2 (chief radiologist) and 3 (orthodontist) with the primary researcher. When the ICC was determined with 4mm, the ICC had an agreement level of 1 (good) for all x and y coordinates across all landmarks.

The results of the inter-examiner correlation tests are summarized in tables 5.4 and 5.5:

Table 5.4: Inter-examiner correlation at Interval 1

	Interval 1	
Landmark	**X co-ordinate**	**Y-co-ordinate**
1	0.9	0.8
2	0.76	0.93
3	0.66	0.8
4	0.83	0.66
5	0.7	0.7
6	0.8	0.63
7	0.73	0.63
8	0.63	0.6
9	0.66	0.66
10	0.86	0.56
11	0.7	0.63
12	0.66	0.7
13	0.66	0.6
14	0.7	0.63
15	0.7	0.66
16	0.6	0.6
17	0.66	0.56
18	0.66	0.73
19	0.9	0.73
Average	**0.72**	**0.67**

Table 5.5: Inter-examiner correlation at Interval 2

	Interval 2	
Landmark	**X co-ordinate**	**Y-co-ordinate**
1	0.9	0.86
2	0.86	0.9
3	0.76	0.76
4	0.73	0.7
5	0.86	0.53
6	0.7	0.6
7	0.8	0.56
8	0.63	0.7
9	0.76	0.53
10	0.7	0.6
11	0.6	0.66
12	0.66	0.53
13	0.66	0.76
14	0.66	0.6
15	0.63	0.76
16	0.7	0.5
17	0.73	0.73
18	0.7	0.66
19	0.93	0.53
Average	**0.73**	**0.65**

The agreement of the examiners between interval 1 and 2 was determined as well with the ICC and represented in Table 5.6, which indicates that between interval 1 and 2 the mean x and y values for the cephalogram as a whole was essentially the same. This indicates that the examiners 2 and 3 between the two intervals were reliable in their assessment of the landmarks in relation to the primary researcher with a moderate agreement for the Y value and a good agreement for value x.

Table 5.6: Agreement between examiner at Interval 1 and 2

Interval	X value	Y value
1	0.72	0.67
2	0.73	0.65
	Good	Moderate

5.4 Euclidean distance measurements

19 landmarks were identified in each of the 409 cephalograms by the primary researcher using two methods [(409 cephalograms x 19 landmarks) x 2 methods = 15542 landmarks]. Each

landmark generated two coordinate values (*x,y*), thus a total of 31084 values were generated (15 542 x 2).

The mean value of the 409 cephalogram records was determined with the Standard Deviation (SD) for each landmark (L1 - L19) (Table 5.7). The literature states that if a landmark is within a distance of 2-4mm from the "control landmark" then the method is acceptable (Katkar *et al.*, 2013*;* Lindner *et al.*, 2016*;* Wang *et al.*, 2016; Park *et al.*, 2019; Moon *et al*., 2020). However, a surprisingly large discrepancy was noted. This was concerning as *visually* the landmarks' locations between the two systems were in close approximation (Figure 5.3). Overall, the greatest Euclidean distances were observed for L18 (anterior nasal spine) (the highest being 92.43mm). The next largest Euclidean distance was observed for L16 (soft tissue pognion) (87.63) followed by L2 (nasion) (52.36). The smallest Euclidean distance was observed for L15 (subnasale). The minimum range of Euclidean distances was 2.34mm and the maximum range was 76.01mm.

Table 5.7: Mean value of the Euclidean distances for the various Cephalometric landmarks

Landmark	Mean	±SD	±SEM	Min	Max
L1	6.19	2.02	0.0998	1,99	17,9
L2	8.75	3.76	0.1859	0,58	52,36
L3	9.64	3.16	0.1562	1,46	21,39
L4	8.98	3.41	0.1686	0,57	23,37
L5	10.57	3.75	0.1854	1,92	24,24
L6	10.84	4.15	0.2052	1,11	29,92
L7	10.43	4.09	0.2022	2,07	29,23
L8	11.29	4.38	0.2165	1,05	29,08
L9	11.28	4.25	0.2101	1,96	28,87
L10	10.43	5.81	0.2872	0,42	31,22
L11	10.2	3.93	0.1943	1,76	24,93
L12	10.59	4.08	0.2017	1,14	26,92
L13	9.65	4.1	0.2027	1,61	27,28
L14	11.09	4.69	0.2319	0,81	42,3
L15	9.6	4.62	0.2284	0,22	54,47
L16	10.66	5.4	0.2670	2,56	87,63
L17	7.66	3.71	0.1834	0,50	50,06
L18	8.88	5.55	0.2744	1,09	92,43
L19	7.10	2.65	0.1310	0,54	16,42

SD = Standard Deviation; SEM = Standard error of the mean

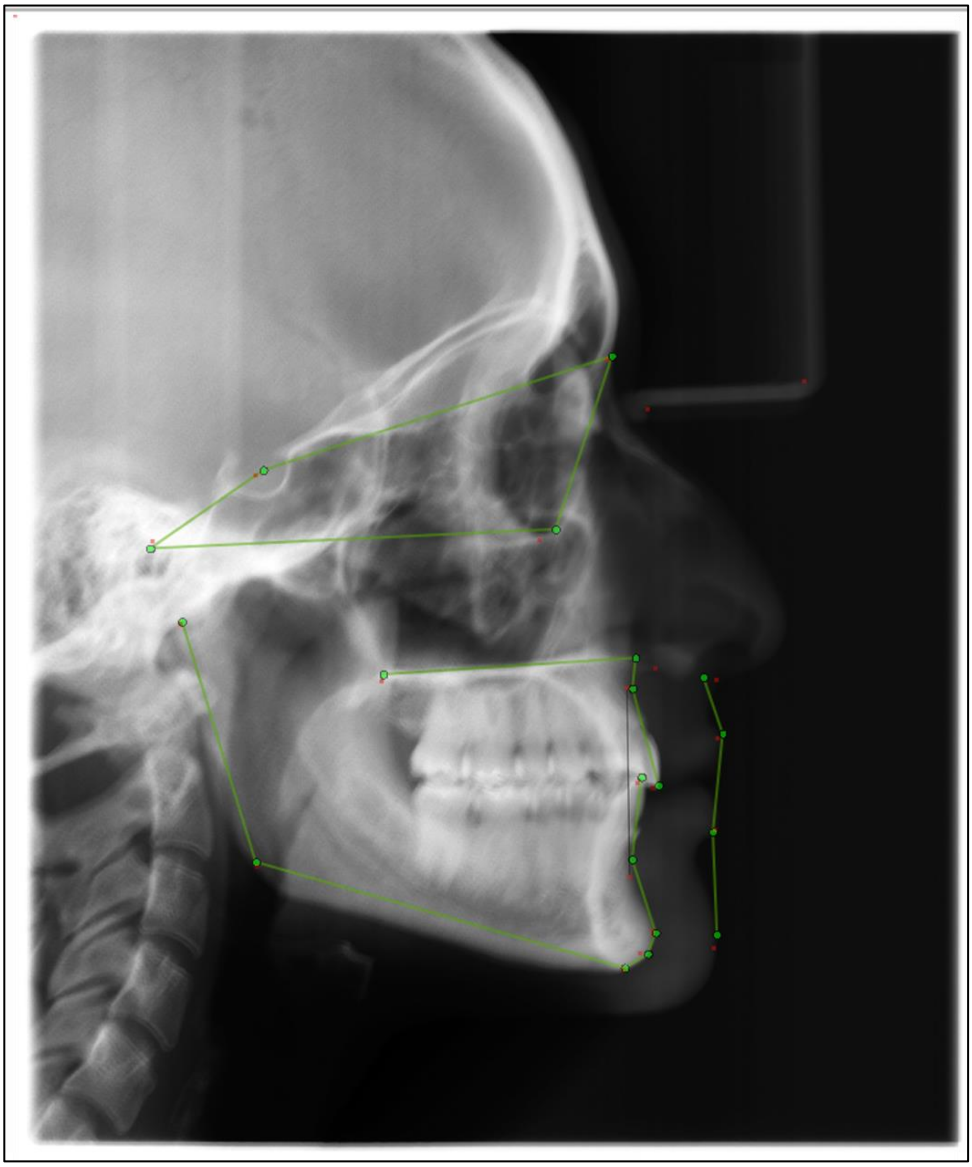


Figure 5.3: Superimposed image comparing the landmarks detected by BoneFinder® (green) and human examination using Dolphin™ (red)

Precision is usually expressed in terms of standard deviation or standard error of a range (difference between the highest and the lowest result) (ISO, 1998; Menditto, Patriarca and Magnusson, 2007). Less precision is reflected by a larger standard deviation. This investigation was carried out using repeatability conditions, where independent test results were obtained with the same method on identical test items in the same location by the same operator using the same within short intervals of time (ISO, 1998). According to Juneja *et al.*, (2021), to quantitatively assess the results of the different landmark identification techniques, two important metrics must be used in the literature; namely, mean error and standard deviation of mean error. The Standard Deviation (SD) represents the difference ±value from the mean

Euclidean value of each landmark of the 409 cephalograms. The standard error of the mean (SEM) values from the Euclidean distances is represented in Table 5.7. The SEM represents the precision of how close the value is to the whole sample size of 409 cephs for each landmark. The small values of the SEM illustrate that the SEM is closely related with a narrow distribution to the SD. A small value of mean error represents acceptable landmark detection results in the case of cephalometric analysis.

Table 5.8: Standard Deviation and Standard Error of the Mean for *xy* coordinates for each method

Landmark	**BoneFinder® *x***		**Dolphin™ *x***		**BoneFinder® *y***		**Dolphin™ *y***	
	SD	**SEM**	**SD**	**SEM**	**SD**	**SEM**	**SD**	**SEM**
1	5.6	0.2769	5.5	0.2720	6.11	0.3021	5.72	0.2828
2	7.86	0.3887	7.35	0.3634	9.67	0.4782	9.2	0.4549
3	5.97	0.2952	6.19	0.3061	8.17	0.4040	7.93	0.3921
4	5.66	0.2799	3.77	0.1864	6.63	0.3278	5.22	0.2581
5	6.77	0.3348	7.16	0.3540	9.08	0.4490	8.9	0.4401
6	9.12	0.4510	8.91	0.4406	9.88	0.4885	10.22	0.5053
7	10.4	0.5142	10.04	0.4964	10.49	0.5187	10.53	0.5207
8	10.83	0.5355	10.17	0.5029	10.29	0.5088	10.37	0.5128
9	10.69	0.5286	10.22	0.5053	10.48	0.5182	10.54	0.5212
10	8.72	0.4312	6.52	0.3224	9.09	0.4495	6.9	0.3412
11	8.25	0.4079	8.45	0.4178	9.54	0.4717	9.68	0.4786
12	8.48	0.4193	8.63	0.4267	9.71	0.4801	9.88	0.4885
13	7.81	0.3862	8.13	0.4020	10.5	0.5192	10.57	0.5227
14	8.78	0.4341	8.9	0.4401	10.46	0.5172	10.53	0.5207
15	7.36	0.3639	7.96	0.3936	9.96	0.4925	10.52	0.5202
16	9.98	0.4935	10.49	0.5187	10.97	0.5424	10.95	0.5414
17	5.68	0.2809	6.12	0.3026	6.31	0.3120	6.2	0.3066
18	6.62	0.3273	8.34	0.4124	9.27	0.4584	9.25	0.4574
19	5.99	0.2962	4.74	0.2344	6.32	0.3125	5.23	0.2586

SD = standard deviation; SEM = standard error of the mean

5.5 Wilcoxon Rank Test and Bland-Altman Plots

Dolphin Imaging™ versus BoneFinder® landmark statistical analysis was calculated at a 95% confidence interval with the Wilcoxon-signed rank test. It was conducted with continuity correction for *x, y* coordinates and the Euclidean distance of the 409 cephalograms.

The Wilcoxon-signed rank test is applied in situations of paired data when the paired data samples come from a population that cannot be assumed to be normally distributed due to the variation of facial profiles and anatomical variation. Since the Wilcoxon sign test is a non-para-continuous-level test, it does not require a special distribution of the dependent variable in the analysis (von Fraunhofer, 2010).

The p-value represents a significant difference that BoneFinder® is different from Dolphin™ if $p < 0.05$. If $p > 0.05$ then there is no significant difference. Indicating the truth of the null hypothesis, there was no significant difference between BoneFinder® and Dolphin™ (Table 5.9).

The *y* coordinate of L2 (Nasion) in BoneFinder® was presented with a significant difference ($p = 0.000031$) concerning the *y*-values obtained in Dolphin™ (Table 5.9). There was no significant difference between the *x* nor *y*-values of BoneFinder® compared to Dolphin™ ($p > 0.05$).

Some cephalometric landmarks are more reliable in either the vertical or horizontal plane (Chen *et al.*, 2000). Large variations of the x-coordinates and y-coordinates occurred. L8 (Menton) and L5 (Point A) were the most reliable landmarks in the horizontal plane (p-value of 9.19 and 8.08 respectively). L9 (Orbitale) was the most reliable in the vertical plane (p-value of 8.66). L2 in the vertical dimension (y-value) presented with the significant difference ($p = 0.000031$).

Table 5.9: Comparison between vertical and horizontal planes for both Dolphin Imaging™ and BoneFinder®

Landmark	X value for Dolphin™ vs BoneFinder®	Y value for Dolphin™ vs BoneFinder®
L1	p-value : 4.64	p-value : 7.21
L2	**p-value : 1.11**	**p-value : 0.000031**
L3	p-value : 1.03	p-value : 8.51
L4	p-value : 1.04	p-value : 3.50
L5	p-value : 8.08	p-value : 7.28
L6	p-value : 4.73	p-value : 1.92
L7	p-value : 4.00	p-value : 2.66
L8	p-value : 9.19	p-value : 6.46
L9	p-value : 4.46	p-value : 8.66
L10	p-value : 2.11	p-value : 5.71
L11	p-value : 5.72	p-value : 3.73
L12	p-value : 4.24	p-value : 2.97
L13	p-value : 4.39	p-value : 1.49
L14	p-value : 7.07	p-value : 2.56
L15	p-value : 1.65	p-value : 2.61
L16	p-value : 1.18	p-value : 2.86
L17	p-value : 8.47	p-value : 4.18
L18	p-value : 1.21	p-value : 3.42
L19	p-value : 1.07	p-value : 5.73

The difficulty in locating the landmarks Orbitale, Porion, and Articulare and Gonion may be the result of a blurred image due to the superimposition of adjacent or bilateral structures (McClure *et al.*, 2005). The landmark Orbitale was more inaccurate in the horizontal plane, most likely the result of the left and right images of the orbits being more closely aligned vertically than anteroposteriorly (p-value of the x-axis = 1.03, p-value for the y-axis = 8.51). Alternatively, Articulare was more imprecise vertically (p-value of the x-axis = 1.07, p-value for the y-axis = 5.73) since this landmark is defined as the most posterior point on the neck of the vertically oriented condyle. The convoluted route of the ear canals creates multiple vertically overlapping radiolucent structures, which was likely a contributory factor in the imprecision of identification of Porion in the vertical direction (p-value of the x-axis = 1.04, p-value for the y-axis = 3.50). The uncertainty in the detection of Gonion may result from the difficulty of establishing this landmark's position along a curved anatomical structure (SD = 5.81).

5.6 Bland Altman

The Bland-Altman analysis was carried out on L2 (Nasion) due to the y-value statistical analysis with the Wilcoxon showing a significant difference between Dolphin™ and BoneFinder®. L16 (Soft tissue pogonion) was also analysed for comparison.

The interpretation of the Bland Altman was limited to Landmarks 1 (Sella), 2 (Nasion) and 16 (Soft tissue pogonion) (Figures 5.4 – 5.6). These landmarks were identified based on the smallest Euclidean distance of Landmark 1 that was the smallest (6.19), and landmark 16 a soft tissue landmark (10.66), where the Wilcoxon statistical significance for the *y*-coordinates of Landmark 2 was determined to be p = 0.000031.

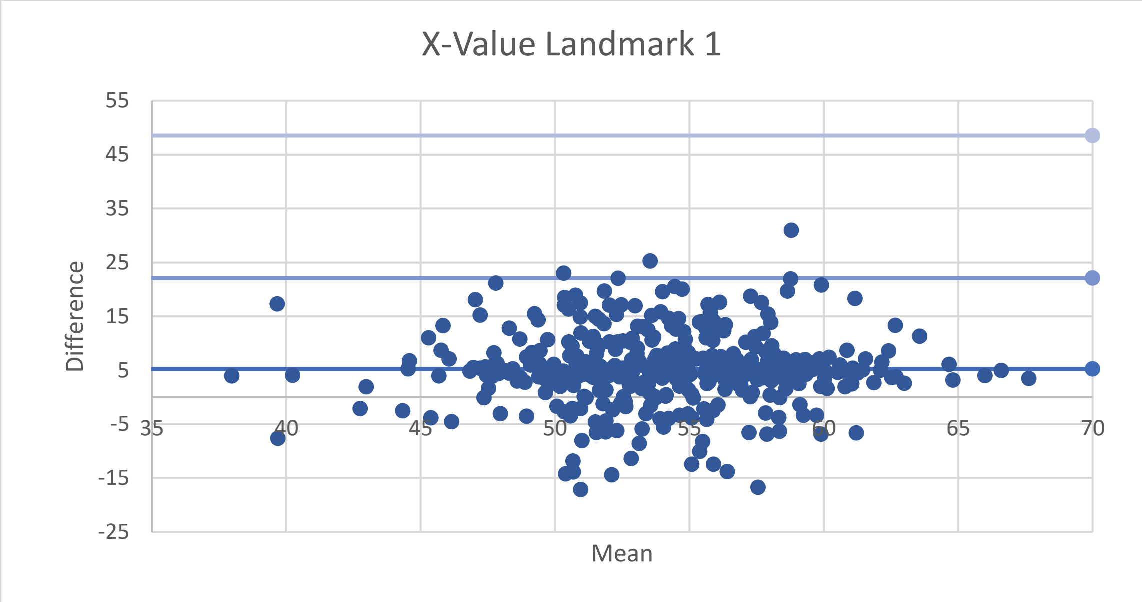


Figure 5.4: Bland-Altman graph for Landmark 1, *x* value

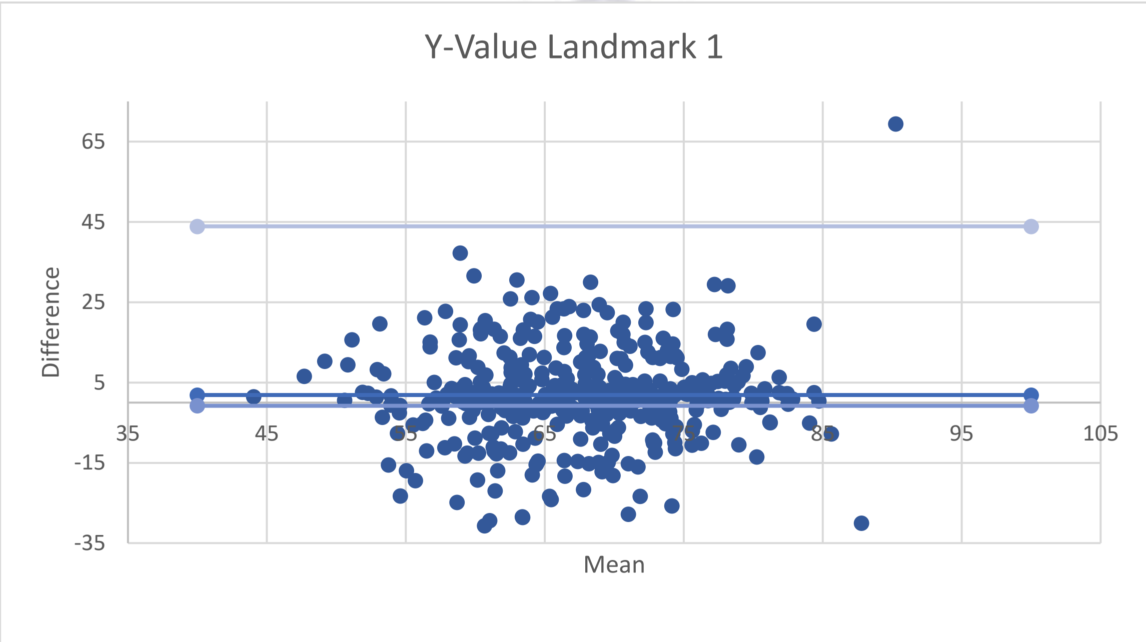


Figure 5.5: Bland-Altman graph for Landmark 1, ***y*** value

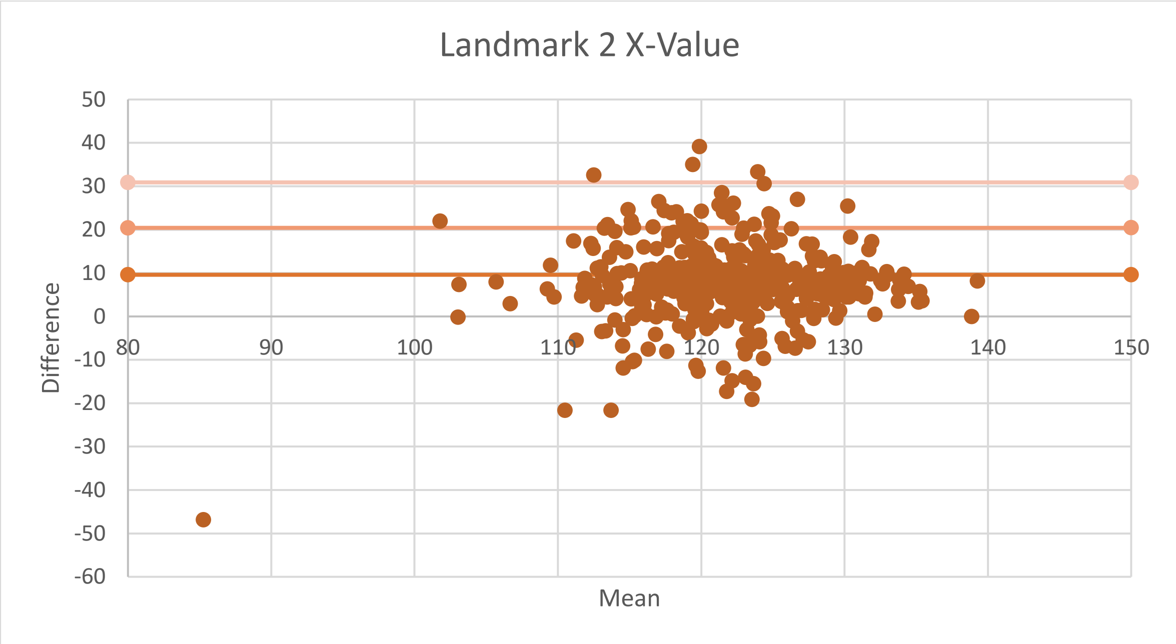


Figure 5.6: Bland-Altman graph for Landmark ***x*** value

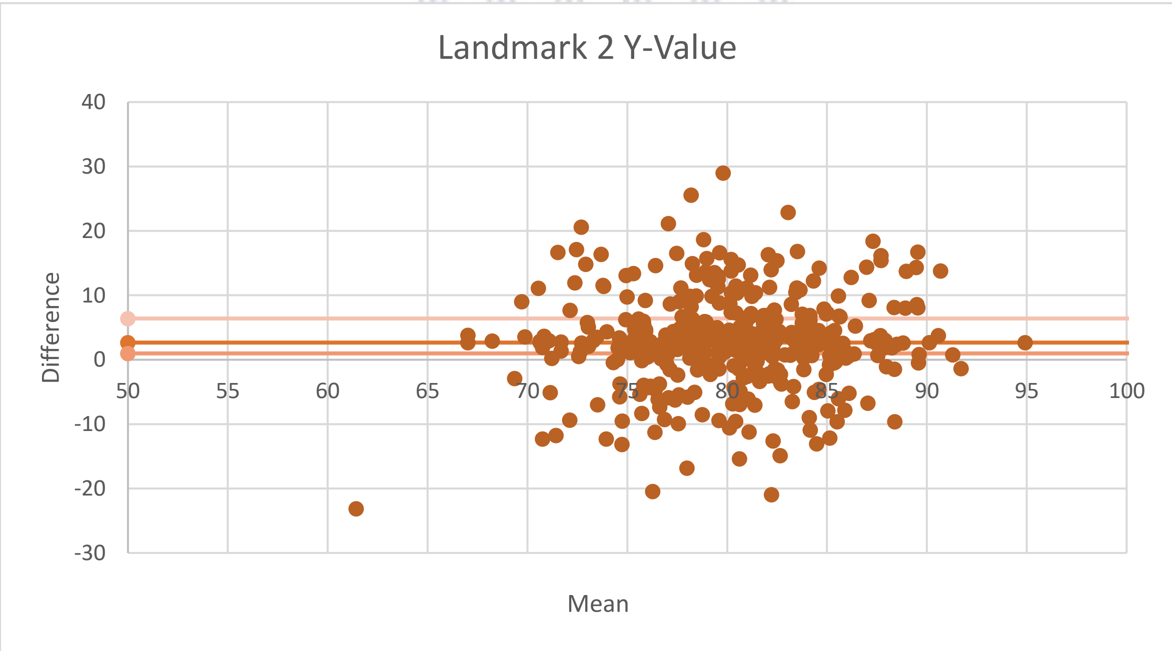


Figure 5.7: Bland-Altman graph for Landmark 2, *y* value

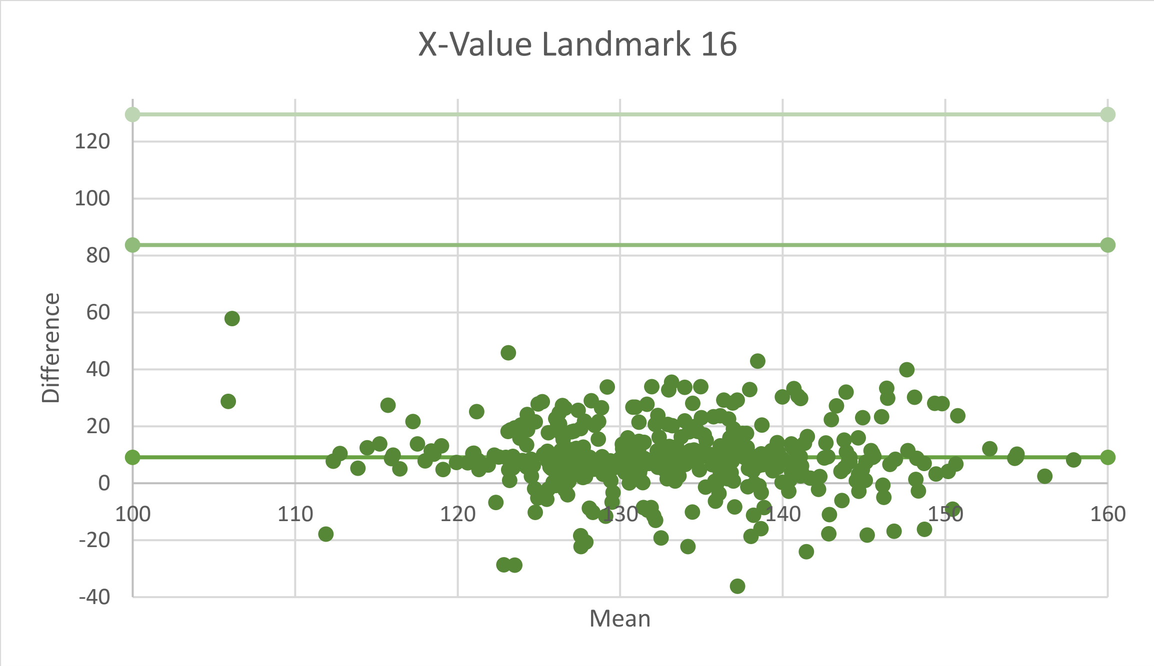


Figure 5.8: Bland-Altman graph for Landmark 1, *x* value

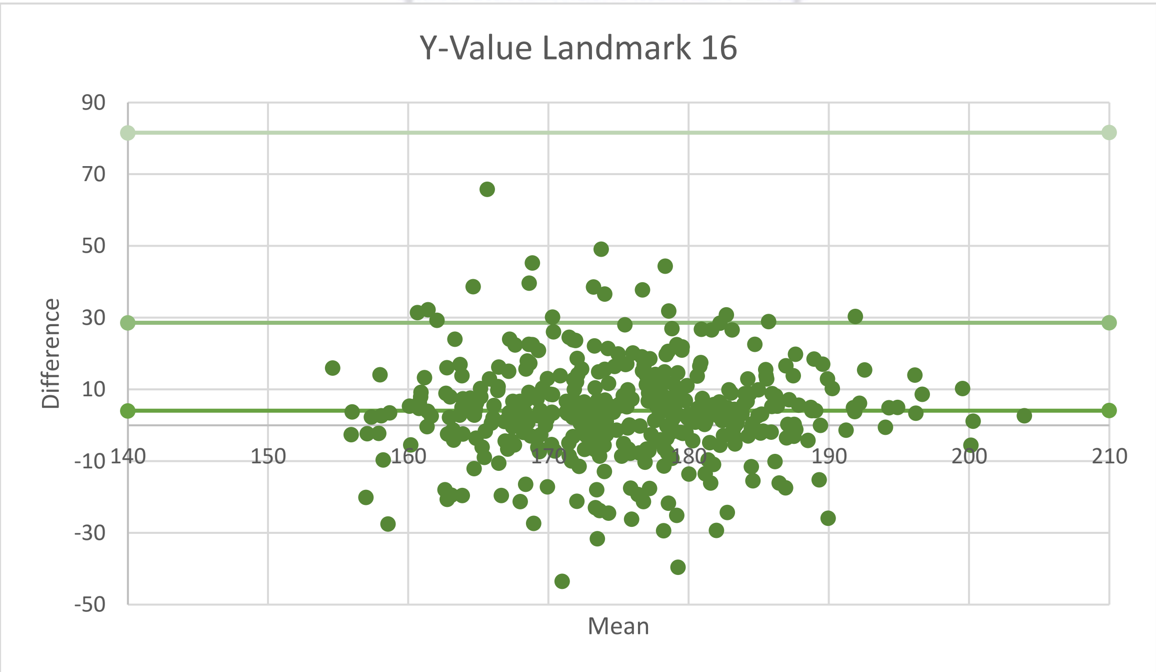


Figure 5.9: Bland-Altman graph for Landmark 1, *y* value

The comparison and interpretation of the Bland Altman graphs provided insight to why Landmark 2; *y*-values presented with a significant difference. As per Table 5.9, the number of *y*-coordinates above the Upper Limit of Agreement is 99. This large number of coordinates resulted in the statistical significance noted for the *y*-coordinate for Landmark 2. The ULOA (upper limit of agreement) was 6.402 and the difference in the coordinates between BoneFinder® and Dolphin™ was up to 28.98. BoneFinder® had much larger y-coordinates for Landmark 2, resulting in the result that there is a significant difference with the Wilcoxon test for the BoneFinder® coordinates in relation to Dolphin™.

Table 5.10: The distribution of the difference values for the various landmarks, based on the *x* and *y* coordinates

Landmark and co-ordinate	Below LLOA	Between LLOA and ULOA	Above ULOA	Around the bias line not crossing any limits – Above Bias line	Around the bias line not crossing any limits – Below Bias line
L1 X-Value	409	4	0	188	218
L2 X-Value	376	29	4	119	257
L16 X-Value	409	0	0	185	224
L1 Y-Value	135	273	1	193	80
L2 Y-Value	**128**	**182**	**99**	**116**	**66**
L16 Y-Value	392	17	0	178	214

**ULOA = Upper limit of agreement, LLOA = Lower limit of agreement*

When looking at the race of a sample of 10 cephalograms that were around the bias, and the large values above the ULOA respectively; the majority was of mixed ethnicity. L2 (Nasion) is usually a reliable landmark as it is situated at a well-defined anatomic point at the intersection of frontal and nasal bones. This region was dark and on our sample of radiographs. Furthermore, patient tilting also resulted in the landmark requiring interpretation.

5.7 Incidental Findings

Attempts to explain the large discrepancy in the Euclidean distances were actively sought but were not the primary reason for the study. These attempts included:

5.7.1 Comparison of different file inputs (i.e. DICOM versus DICOM, DICOM versus JPEG)

With BoneFinder®, DICOM files export coordinates in millimetres automatically, and JPEG files exported coordinates in pixels. As a result, DICOM files were used to compare Dolphin™ and BoneFinder® and this resulted in large discrepancies in the Euclidean distance.

Using Dolphin™, the same patient's cephalogram was used in a DICOM format, followed by JPEG format. This comparison of JPEG files in Dolphin™ with DICOM in BoneFinder® revealed more 'agreeable' Euclidean distance results. When comparing Dolphin ™ coordinates from the JPEG file to the DICOM file, it differed significantly (Table 5.11).

In the table below, there are important things to note. It can be seen that BoneFinder® provides deterministic results, i.e. the coordinates output would be the same, no matter how many times the same image is imported into the software. When a JPEG image was imported into Dolphin™, it yielded *different* coordinates than that of the DICOM file. The mean Euclidean distance of Dolphin™ JPEG and BoneFinder DICOM was within the accepted range of 2.15mm, whereas the mean Euclidean distance of Dolphin™ DICOM and BoneFinder® DICOM was 8.49mm.

Table 5.11: A comparison of DICOM and JPEG files with Dolphin Imaging™ and DICOM files with BoneFinder® and their respective Euclidean Distances

BoneFinder® DICOM		Dolphin™ JPEG		Euclidean Distance	Dolphin™ DICOM		BoneFinder® DICOM		Euclidean Distance
X	Y	X	Y		X	Y	X	Y	
56.0692	84.7893	53.8	-86.0	2,57	51.4	-83.1	56.0692	84.7893	2,57
124.513	57.8378	124.0	-58.7	1	116.8	-57.1	124.513	57.8378	7,75
116.972	92.6589	114.4	-92.7	2,57	103.9	-92.4	116.972	92.6589	13,07
39.4534	107.933	32.8	-102.5	8,59	33.2	-103.4	39.4534	107.933	7,72
138.273	121.236	137.1	-120.5	1,38	130.9	-117.5	138.273	121.236	8,27
141.43	159.992	139.8	-159.7	1,66	134.0	-153.5	141.43	159.992	9,87
147.175	176.163	145.7	-175.5	1,62	139.5	-169.3	147.175	176.163	10,3
142.089	183.857	141.3	-183.1	1,09	135.2	-177.2	142.089	183.857	9,58
146.032	180.966	145.0	-180.2	1,29	138.8	-173.8	146.032	180.966	10,18
67.9345	169.588	65.8	-171.6	2,93	60.9	-165.0	67.9345	169.588	8,4
143.968	143.858	142.8	-144.7	1,44	136.4	-139.7	143.968	143.858	8,64
146.584	144.692	145.0	-144.4	1,61	139.2	-139.2	146.584	144.692	9,2
160.239	128.198	160.8	-130.6	2,47	154.1	-127.5	160.239	128.198	6,18
158.483	148.165	157.1	-147.9	1,41	151.7	-140.9	158.483	148.165	9,94
155.459	116.939	155.1	-116.1	0,91	148.1	-112.7	155.459	116.939	8,49
162.005	172.31	160.3	-175.5	3,62	153.8	-172.6	162.005	172.31	8,21
84.5971	126.213	86.6	-126.9	2,12	79.5	-122.0	84.5971	126.213	6,61
142.189	114.68	140.8	-114.1	1,51	134.7	-110.8	142.189	114.68	8,43
48.8331	123.398	47.8	-123.7	1,08	43.0	-118.0	48.8331	123.398	7,95

5.7.2. Comparison: Adjusting the Dolphin Imaging™ ruler calibration

As described in the methodology (Appendix F; Figure 9.9), a ruler length of 30mm was used. This is in accordance with the real length of the corner points of the nasion-positioning rod. This was needed for image calibration since there was no calibration ruler was included during the acquisition of the image.

When re-evaluating the Dolphin™ parameters, it was apparent that changes to the calibration ruler significantly changed the results. The possibility of inaccurate measurement of the nasion-positioning rod was explored. Table 5.12 shows that when this distance was changed to 31mm, it considerably altered the Euclidean distance.

Table 5.12: Changes of ruler calibration resulting in changes of cartesian coordinates

Landmarks	Dolphin™				BoneFinder®		E1	E2
	30mm calibration		31mm calibration					
	X	Y	X	Y	X	Y		
1	57.8	75.4	58,5	75,8	59.3796	79.0643	3,99	3,38
2	131.9	70.5	133,5	71,2	132.361	70.1415	0,58	1,55
3	112.8	99.3	115,4	100,5	120.138	100.923	7,52	4,76
4	28.0	93.5	27,1	92,4	33.0467	99.824	8,09	9,51
5	129.3	132.1	131,3	133,5	131.298	137.655	5,9	4,16
6	117.9	174.7	119,8	176,5	126.251	182.461	11,4	8,78
7	118.4	192.3	120,1	195,4	125.675	198.117	9,31	6,2
8	110.7	198.8	112,3	201,7	117.724	204.58	9,1	6,14
9	116.0	197.6	116,4	200,7	123.112	202.761	8,79	7,02
10	39.4	161.9	42,3	166	48.1837	165.141	9,36	5,95
11	123.7	157.0	126	159,7	135.644	160.908	12,57	9,72
12	129.3	159.9	130,8	161,9	138.028	162.627	9,14	7,26
13	146.9	145.9	147,7	146,5	151.633	148.325	5,32	4,34
14	138.0	168.1	138,7	169,9	147.621	177.573	13,5	11,77
15	150.3	131.2	154,6	131,1	145.338	135.312	6,44	10,17
16	131.9	191.8	133,8	191,9	139.775	196.754	9,3	7,7
17	74.4	125.4	77	125,9	81.2463	128.275	7,43	4,87
18	136.8	128.5	138,7	130,1	134.155	130.224	3,16	4,55
19	35.8	113.3	36,2	116,1	42.5032	115.531	7,06	6,33

*E1 = Euclidean Distance (Dolphin™ Original calibration of 30mm and BoneFinder®), *E2 = Euclidean Distance (Dolphin™ New calibration of 31mm and BoneFinder®), Record number 10 was used.

5.7.3 Comparison: Image alignment on Dolphin Imaging™

As described in the methodology (Appendix F, Figure 9.4 and 9.5), the cephalogram was dragged into the image box. The automatic alignment of the image was used for this study (Figure 5.10) (scenario 1). To assess whether the alignment changed the coordinate outputs, the image was then re-aligned so that the cephalogram image border and the boundary box corresponded (Figure 5.11) (scenario 2). The coordinates from both the first scenario and second scenario were exported and as anticipated the coordinates did indeed differ. Furthermore, the Euclidean distance discrepancy still existed (Table 5.13).

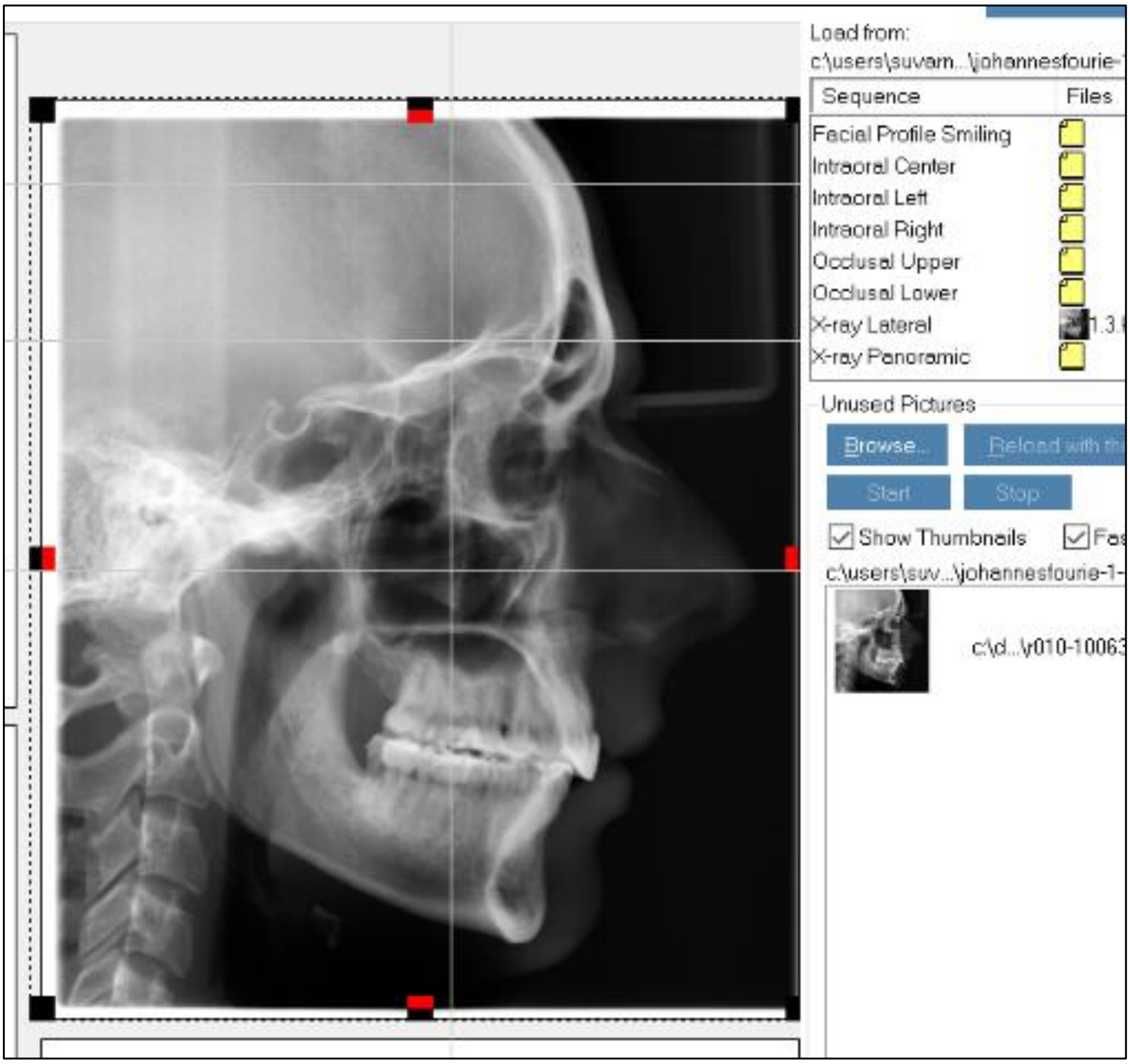


Figure 5.10: Scenario 1

The automatic alignment places the cephalogram image border *OUTSIDE* the black boundary (dotted border lies outside the black solid line)

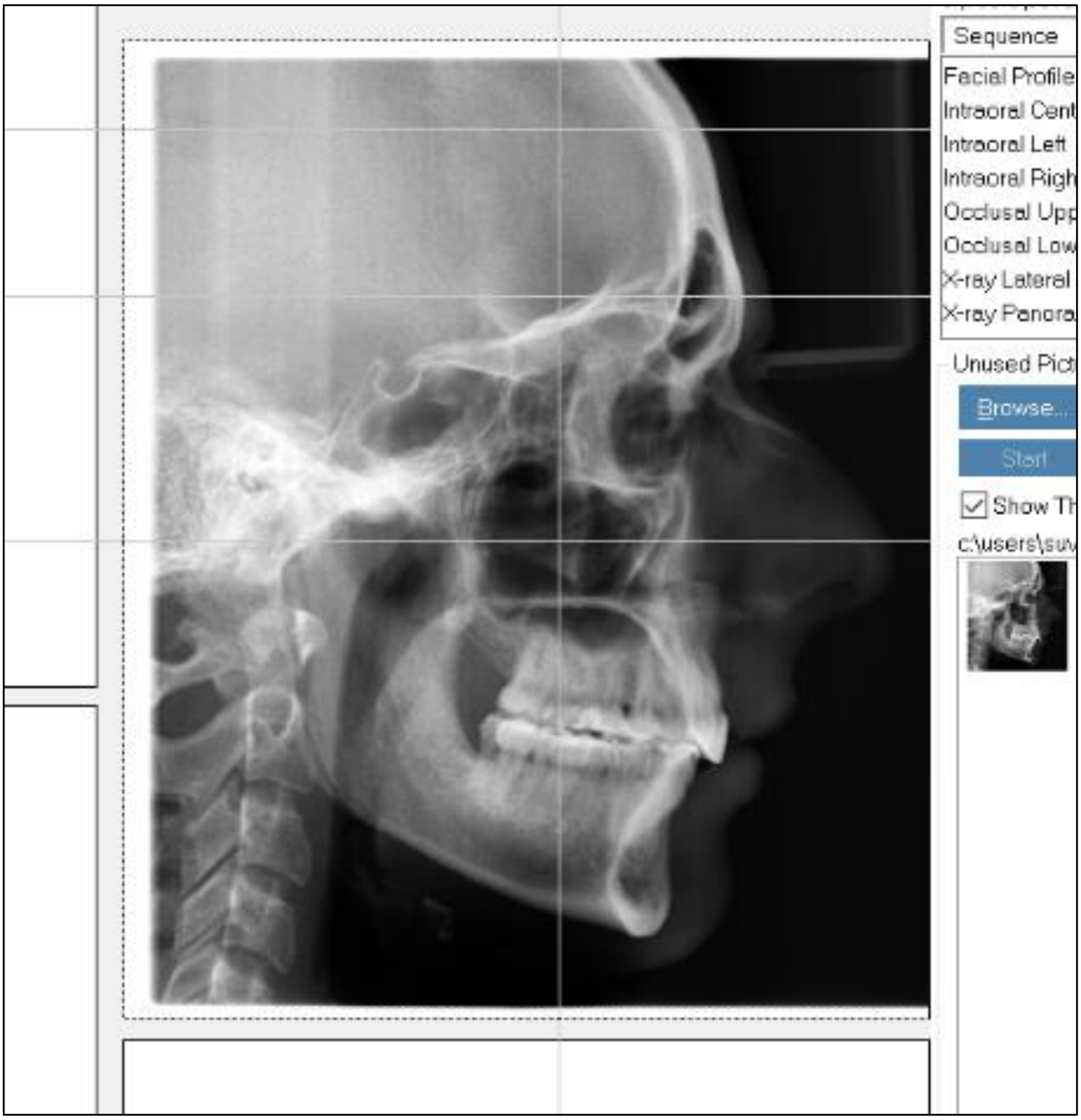


Figure 5.11: Scenario 2

The image was aligned so that the cephalogram image border and the boundary box corresponded (dotted line).

Table 5.13: Comparison of the Euclidean Distance with changes of the alignment

Landmarks	Dolphin™ Auto-Aligned		Dolphin™ Manually aligned		BoneFinder® Original		E1	E2
	X	Y	X	Y	X	Y		
1	57.8	-75.4	57.8	-73.2	59.3796	79.0643	3.99	6.07
2	131.9	-70.5	129.3	-68.0	132.361	70.1415	0.58	3.74
3	112.8	-99.3	112.0	-96.6	120.138	100.923	7.52	9.21
4	28.0	-93.5	28.3	-86.0	33.0467	99.824	8.09	14.62
5	129.3	-132.1	128.4	-127.4	131.298	137.655	5.9	10.66
6	117.9	-174.7	116.7	-169.1	126.251	182.461	11.4	16.42
7	118.4	-192.3	116.9	-186.4	125.675	198.117	9.31	14.64
8	110.7	-198.8	109.9	-193.4	117.724	204.58	9.1	13.65
9	116.0	-197.6	114.6	-191.5	123.112	202.761	8.79	14.12
10	39.4	-161.9	40.7	-156.9	48.1837	165.141	9.36	11.13
11	123.7	-157.0	122.1	-152.9	135.644	160.908	12.57	15.73
12	129.3	-159.9	127.5	-155.7	138.028	162.627	9.14	12.6
13	146.9	-145.9	143.8	-139.8	151.633	148.325	5.32	11.58
14	138.0	-168.1	134.7	-163.7	147.621	177.573	13.5	18.96
15	150.3	-131.2	150.9	-124.9	145.338	135.312	6.44	11.8
16	131.9	-191.8	130.5	-184.3	139.775	196.754	9.3	15.53
17	74.4	-125.4	75.3	-120.4	81.2463	128.275	7.43	9.87
18	136.8	-128.5	135.0	-124.4	134.155	130.224	3.16	5.88
19	35.8	-113.3	36.3	-111.3	42.5032	115.531	7.06	7.51

*E1 = Euclidean Distance (Dolphin™ Original calibration of 30mm and BoneFinder®), *E2 = Euclidean Distance (Dolphin™ New calibration of 31mm and BoneFinder®), Record number 10 was used.

CHAPTER 6 : DISCUSSION

This study aimed to determine the precision of two cephalometric landmark identification methods, namely an artificial intelligence programme (BoneFinder®) and a computer-assisted human examination software (Dolphin Imaging™). To preserve clarity an overview of each topic is discussed. This section is organized as a sequential review and this study will now be compared to the findings of previous work.

6.1 Demographic Data

Overall, this study did not set out to compare differences between males and females and the different racial groups. However, the primary researcher notes that patients of mixed ancestry can present with different skeletal patterns contributing to different landmark norms.

A large component of this study sample comprised of Coloured/Mixed race individuals (n = 244, 59.66%) and Blacks (n = 40, 9.78%). Seedat (1983) noted that patients of mixed ancestry may have skeletal differences and as a result landmark detection in these individuals may fluctuate. The literature on South African cephalometric norms appears sparse. Seedat's 1983 study, which aimed to determine clinical cephalometric values applicable to the Cape Coloured community, is the only comparative study performed to date. According to Tobias (1953) (cited in Seedat, 1983), the Cape Coloureds "are a community resident in the Western Cape, South Africa whose origin stems from an admixture of Caucasoids, Negroids and Mongoloid races." No comparative cephalometric studies have been performed on the South African Cape Coloured community (which comprised the majority of the current study).

Seedat's conclusions found that the cephalometric and dental norms of this "Coloured group" showed significant differences in the majority of parameters when compared to Steiner values. While Seedat's study did not directly investigate landmark detection, the values obtained in cephalometry first arise from accurate landmark detection and therefore need to be studied concertedly.

Furthermore, an evaluation of the mean cephalometric values for black South African adults (mean age of 24.5 years) in the Western Cape region of South Africa was last conducted in 1997 by Naidoo and Miles (1997). Their studies indicated that both the hard and soft tissue profiles for black South Africans differ from the North American Caucasians and African-American people. As such treatment planning for these individuals would be more accurate if based on a diagnosis that includes measurements for specific populations groups. This should replace the standard Caucasian norms.

Barter *et al.*, (1995) also conducted a cephalometric study on 50 male and 54 females of Sotho-Tswana children 11- 16 years of age in South Africa. Various analyses such as Steiner, Wits and McNamara were used. This present study agrees with the observations made by Barter *et al*. that Black and Coloured patients are now presenting in increasing numbers for orthodontic treatment. Cotton *et al*. (1951), Altemus (1960) and Jacobsen *et al.* (1977) (cited in Barter *et al.*, 1995) also stated that the norms and standards of one group cannot be used, without modification, in orthodontic treatment planning for another racial group. Kula and Ghoneima (2018) also have compiled literature from Cotton *et al.* (1951) and Altemus (1960) highlighting the central theme that each continent or country will have differences in cephalometric values among various ethnic groups. In South Africa, particularly, years of integration of ethnic groups has taken place leading to difficulties in characterizing those groups based on the norms.

It can be seen that literature on the South African population groups is outdated. The findings of this current study motivate the development of cephalometric norms to provide a closer approximation of the profiles of the South African population.

It is also important to note that the AI program, BoneFinder® was trained on an unknown population group. Marked differences may have occurred due to this. This will be discussed in more detail in Chapter 7.

6.2 Intra-examiner and Inter-examiner Reliability

Agreement is defined as "how well an observation produces the same value on repeated measurements in the same patient". The intraclass correlation coefficient was used to assess reliability for continuous data. According to the literature, an intraclass correlation coefficient greater than 0.70 is considered acceptable (Anvari, *et al.*, 2015). An important consideration with regards to the evaluation of a method of measurement is the agreement over a range of patients and a range of observers. Since there is no "true" value in cephalometry (Hwang *et al.*, 2019), precision was measured – i.e. how close the observers' measurements were to each other (Jones, *et al.*, 2011).

6.2.1 Intra-examiner reliability

All the data from interval 1 versus interval 2 with regard to the intra-reliability of examiner 1 had a positive correlation coefficient and the p-values were above 1. An *r* value above +0.70 indicated a strong uphill (positive) linear relationship. All the *r* values were greater than +0.70. A p-value of 1 implied that a linear equation describes the relationship between the data of interval 1 and interval 2 perfectly, with all data points lying on a line for which Interval 2 increases as interval 1 increases. The primary researcher remained consistent.

6.2.2 Inter-examiner reliability

According to Hwang *et al.* (2019) “when it comes to a reliability measure when identifying a certain cephalometric landmark, there is no firm ‘ground truth’ or gold standard that can provide validation as to where the true location of the landmark is”.

Therefore, to achieve the first objective of this study, the primary researcher was calibrated by comparing her to two experienced observers using Dolphin Imaging™ software.

Chen *et al*. (2000) found that inter-observer error presents greater values than intra-observer error. This also accords with the results found by Durão *et al.*, (2014), which confirmed that there was a higher rate of inter-observer error. Meric and Naoumova (2020) also recently reported that inter-operator error is greater than intra-operator error. Therefore, to prevent such errors, the landmark detection in this study was carried out by the primary researcher. The primary researcher remained consistent and was therefore calibrated.

Some landmarks also show a wider variation in localization than others. Landmarks such as Gonion, Porion, Orbitale, and the lower incisor apex may be difficult to identify due to its superimposition between bilateral anatomical structures (Durão *et al.*, 2014). Gonion, porion and orbitale also showed variation in this study’s inter-examiner tests.

The discrepancies that occur amongst observers are well-documented in the literature (Ongkosuwito *et al.*, 2002; Durão *et al.*, 2014). Variation can occur due to inconsistent landmark detection relating to several random errors such as superimposition of the structures, image resolution, radiograph quality, anatomical complexity, examiner experience and observer subjectivity (Anuwongnukroh *et al.*, 2018). The primary researcher had 3 years of experience within the Oral and Maxillofacial Radiology department. Other factors that could have affected intra-examiner agreement included orthodontic experience and day-to-day activities that led to time constraints.

Errors can also be induced by anatomical variations. We can explain this by using interval 2, landmark 17 (PNS) y-co-ordinate as an example (Figure 6.1), which had an interclass correlation of 0.56. Due to the maxillary third molars that are commonly unerupted, they can obscure the detection of PNS. As a result, the location of this landmark moves from "identifiable" or clearly recognizable to requiring interpretation – subjectivity and experience play a big role here.

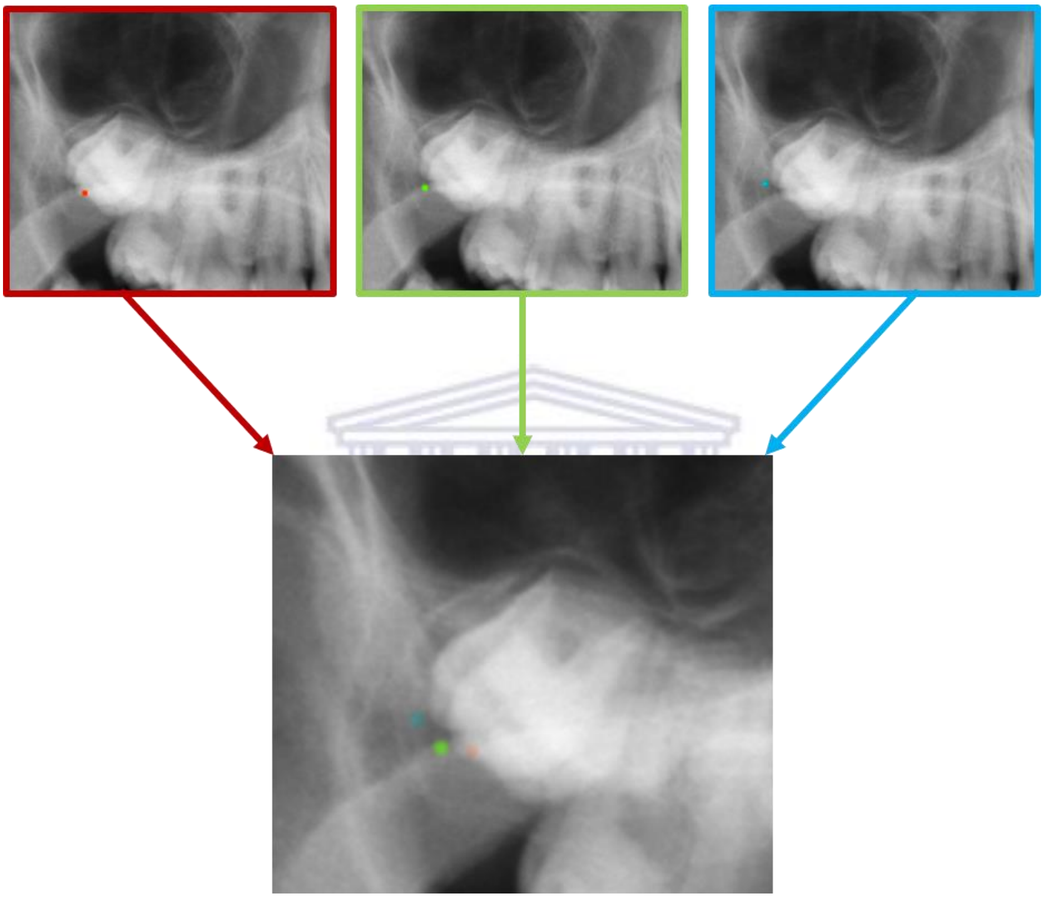


Figure 6.1: Comparison of PNS landmark detection.

The unerupted third molar superimposed on the landmark region renders the PNS indifferentiable. Top: All examiners detected PNS at varying points - the primary researcher (red); chief radiologist (green) and an orthodontist (blue) using Dolphin™ software. Below: Superimposed and zoomed in image depicting proximity of landmarks cephalograms with all 3 observers' landmarks for the same patient.

It is apparent from this, that variations can still exist within a single examiner and between different examiners. As mentioned in the literature review, this substantiates the need for an objective AI programme.

6.3 Accuracy versus Precision

Many studies regarding cephalometric landmark detection with AI can be found in the literature. However, to the author's current knowledge, no studies have been performed using the aforementioned programs nor using a South African population. With the underlying aim of determining the precision of two cephalometric landmark identification methods, a comparison between an artificial intelligence programme (BoneFinder®) and a computer-assisted human examination software (Dolphin Imaging™) was done.

Accuracy and *precision* are commonly used synonymously (Wilson and Hernandez-Hall, 2015), but there is a fine semantic distinction between the term *accuracy* and *precision* (Stallings and Gillmore, 1971). Both terms denote a reference to correctness. However, accuracy refers to how close a sample estimate is to a *gold standard or a true (or accepted) value*, i.e. how nearly correct it is. The term accuracy refers to the “trueness" or the closeness of the analytical result to an accepted reference value. An accurate determination, therefore, produces a "true" quantitative value, i.e. it is free of bias (ISO, 1998; Menditto, *et al.*, 2007; Dunn, 2019). However, with cephalometric landmark detection, the true location of the landmark cannot be validated (Hwang *et al.* 2019). In fact, Schwendicke *et al*. (2021) stated that a major difficulty in AI studies is the construction of the reference test in the absence of a hard “gold standard”. Researchers in dentistry, are sometimes compelled to use multiple human examiners to independently label data, thereby creating a “fuzzy” gold standard (e.g. in this study, a pool of cephalograms were labelled by two experienced examiners, who may not always agree on their verdict).

Precision refers to how close the sample estimates from *different samples are likely to be to each other*, i.e. the “spread” of the measurements or how close they are together (Wilson and Hernandez-Hall, 2015; Dunn, 2019). Like reliability, precision refers to consistency. Simply put, this study aimed to compare if BoneFinder® and Dolphin™ were comparable at locating landmarks.

A useful analogy is that of the dartboard. The more precise a group of measurements, the closer together they are. However, a large degree of precision does not necessarily imply accuracy, as illustrated in Figure 6.2. Based on the small SEM values, both Dolphin™ and BoneFinder® by themselves as a tool for the determination of the 19 landmarks were consistent in their respective findings. This is similar to a dart player that never hits the bull’s eye but is consistently hitting the exact same spot with the dart board; round after round.

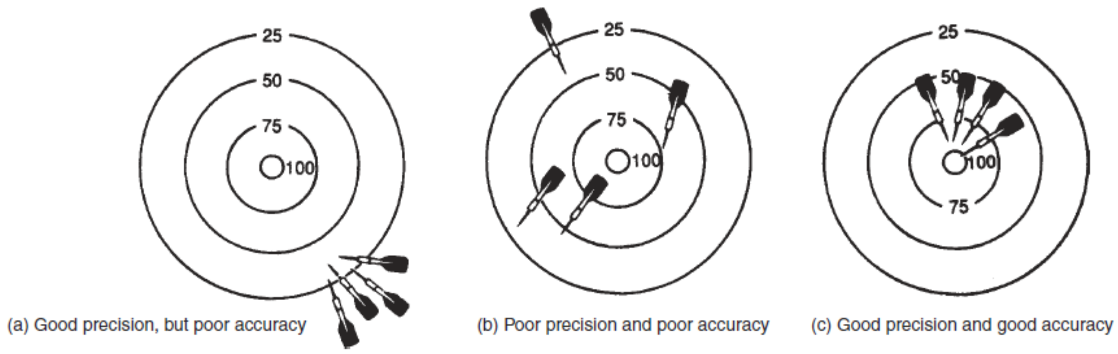


Figure 6.2: Accuracy and precision

The true value in this analogy is the bull's eye. The degree of scattering is an indication of precision—the closer together a dart grouping, the greater the precision. A group (or symmetric grouping with an average) lose to the true value represents accuracy (Wilson and Hernandez-Hall, 2015)

6.4. Errors

It is important to note that obtaining *greater accuracy* for an experimental value is dependent on *minimizing systematic errors*. To obtain *greater precision random errors* must be minimized (Wilson and Hernandez-Hall, 2015). There is always an inherent discrepancy in *any* measurement taken. This is referred to as "experimental error". This does not infer incompetence of the researcher, but rather infer that the measuring tools are merely approximated. A measurement, whether digital or physical, is never exact as the tool used or the skill of the observer always bears limitations (Coan, 2006).

For example, multiple measurements may have variations in answers. This variation is called random error. As previously mentioned in the literature review, the progression of manual cephalometry to computer assisted-cephalometric analysis and automatic landmark detection is directed at improving the diagnostic value of cephalometric analysis by reducing any systematic or random errors and saving time (Ongkosuwito *et al.*, 2002).

Inconsistency in landmark identification is the most central source of random errors both in computer-aided digital cephalometry and in manual cephalometric analysis (Leonardi *et al.*, 2008). Both methods are time-consuming, thus resulting in efforts to automate cephalometric analysis, improving the accuracy of landmark identification and reducing the errors due to clinicians' subjectivity (Ongkosuwito *et al.*, 2002; Leonardi *et al.*, 2008).

Many factors can induce random error, including the quality of the radiographic image, the precision of landmark definition, the reproducibility of the landmark location, the operator experience, and the recording procedure (Anuwongnukroh *et al.*, 2018; Kula and Ghoneima, 2018).

Efforts to minimize random error were made by ensuring there was minimal subjectivity, therefore only one examiner performed the landmark detections. The main researcher was also calibrated to ensure consistency and reliability. Aksakalli *et al.* (2017) also ensured that their measurements were performed by the same single investigator. Aksakalli *et al*. also echo the observation found generally throughout the literature on this topic, that uncertainty in the landmark identification is the main source of error – which is highly dependent on the examiner's experience. To minimize any sources of errors, Dolphin™ software enables the operator to zoom in, zoom out, move the pointer and reposition the image to choose the ideal location of the landmark.

It has been reiterated that the main source of error in cephalometry is landmark identification, therefore it is vital to ascertain whether the use of automated detection is reliable (Meric and Naoumova, 2020; Juneja *et al.*, 2021).

Measurements may display good precision but yet be very inaccurate. The most surprising aspect to emerge from the data was the large Euclidean distances, where it was greater than 4mm. These measurements were *consistently* too high. The largest value was calculated to be 92.43 for L18 (ANS) and the minimum value was 0.22 for L15 (subnasale) (Table 5.6). Contrary to the Euclidean distance data, this study did not find a statistically significant difference between BoneFinder® and Dolphin™.

These results are both revealing and unrevealing in several ways. First, the large Euclidean distance was concerning. However, it wasn't apparent as to why. There are several possible explanations for these results. Systematic errors are often difficult to detect and usually requires an understanding of the measurement tools or programs. This leads us to the technical parameters of BoneFinder® and Dolphin™.

6.5 Dolphin Imaging™

In both clinical and research contexts, Dolphin Imaging™ 11.0 has been used to reliably diagnose, plan, monitor and evaluate orthodontic treatment. Several studies have been cited in testing the Dolphin's reliability and reproducibility (Paixão *et al.*, 2010; Mosleh *et al.*, 2016; Aksakalli *et al.*, 2017; Wanjau *et al.*, 2019).

When taking measurements, the goal should always be to reduce as many sources of error as possible and to keep track of those errors that cannot be eliminated. The types of error that could have been incurred during this study are briefly mentioned here: (1) unclear landmark definitions, landmark definitions subject to bias/interpretation/observer experience, (2) environmental factors such as distractions, lighting, (3) personal errors such as operator fatigue and (4) calibration errors.

Calibration is an important source of systematic error (Albarakati *et al.*, 2012). Before gathering data, an instrument's calibration should be verified. If a calibration standard is not available, the instrument's accuracy should be verified by comparing it to another instrument of similar precision or reviewing the manufacturer's technical data. One of the major drawbacks in this study was sensitivity to ruler calibration. Calibration was carried out according to the manufacturer's instructions. Since no ruler was used during the acquisition of the images, the calibration of the actual size of each image in millimetres was based on the measurement of the known distance (30 mm) between the two fixed corner points of the nasion-guiding rod on the screen. This calibration standardized all images. However, this also could have introduced random error as the placement of the mouse-driven cursor was highly sensitive. The actual process of the calibration was also not perfectly repeatable; therefore uncertainty was introduced through the calibration process (Muelaner, 2018).

In this study, we saw large differences in the Euclidean distance. The literature has stated a 2-4 mm difference is diagnostically acceptable (Katkar *et al.*, 2013; Lindner *et al.*, 2016; Wang *et al.*, 2016; Park *et al.*, 2019; Moon *et al.*, 2020). However, our results yielded far greater differences. This was concerning as *visually* the landmarks' locations between the two systems were in close approximation (Figure 5.3). Unfortunately, changing the measuring technique (i.e. the software parameters) was not possible and therefore we cannot be sure as to what systematic error could be present. A suggested contributing factor could be the reference frame used in Dolphin™. A reference frame refers to the coordinate system used whereby the origin, orientation and scale are defined by a set of reference points (Lindner *et al.*, 2016). Very little was found in the literature on the question of reference frames concerning cephalometric studies. Protection of this proprietary information resulted in this factor being unknown. Condylion was not identified at its true anatomical position but was selected arbitrarily to be used as the centre of origin to determine the *x, y* coordinates of the other landmarks (Appendix F, Figure 9.10). This also may have introduced a calibration error. Again, this process was not perfectly repeatable and was highly subjective, and this may have also contributed to the uncertainty.

It is important to note that Dolphin™ is used more often than other digital cephalometric software programs (Aksakalli *et al.*, 2017), and its many references in the literature account for its validity. The discrepancies that have been identified do not reflect uncertainties of functionality but rather variations in comparison. Using a ruler analogy (shown in Figures 6.3 and 6.4) we can describe this observation simply. Both methods of landmark detection can be likened to comparing two rulers. No one has questioned testing two rulers, as we assume the measurements will be the same. In the same vein, when we compared the two landmark detection methods, we assumed they would be in agreement. When one uses a ruler, it can produce random errors as the user may not consistently get the ruler's zero line aligned exactly with the measured object's border. Furthermore using a tool with a different resolution can also impact the precision of readings. In addition, the alignment of rulers can also significantly impact measurements (Figure 6.3). Another inconsistency that may have occurred is the scenario in which inherent scaling of Dolphin™ software was applied, which may have added to the existing discrepancies.

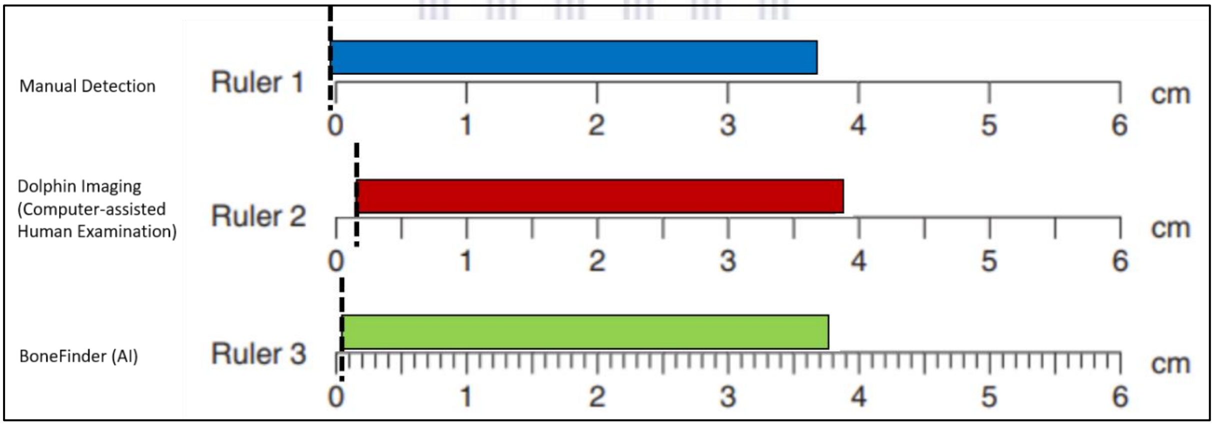


Figure 6.3: Effect of ruler alignment

If a ruler is not accurately aligned with the object you are measuring, differences may occur (adapted from Wilson and Hernandez-Hall, 2015).

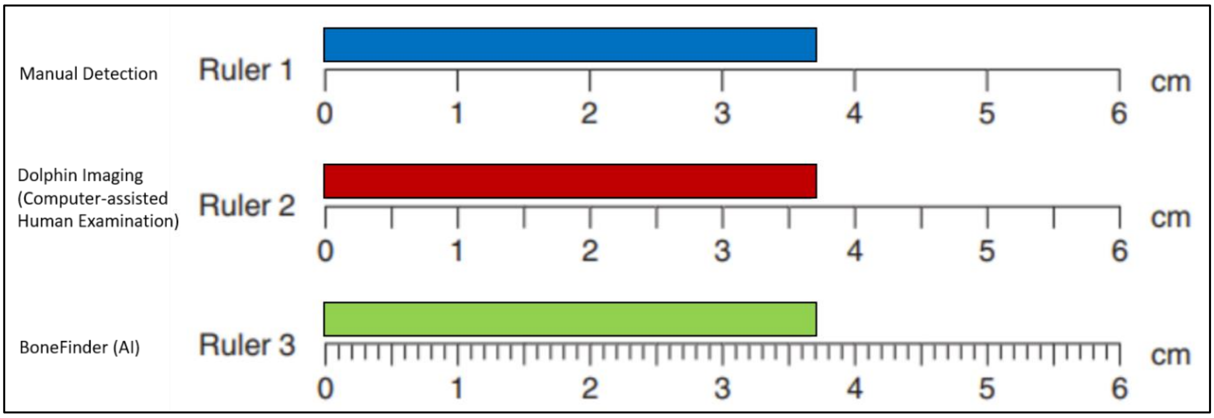


Figure 6.4: Effect of ruler resolution

By using an instrument with a higher resolution and reading it to the smallest reading possible will reduce the uncertainty in results. Ruler 2 will give a more precise reading, than ruler 1. Ruler 3 will give the most precise reading. BoneFinder® provided millimeters to up to 4 decimal places, whereas Dolphin™ provide measurement with 2 decimal places (adapted from Wilson and Hernandez-Hall, 2015).

According to the author's knowledge, no study has compared Dolphin™ with an AI program and no studies exist in relation to a South African demographic. The discrepancies found do not warrant cessation of the use of Dolphin™, but rather advocate the need for further research into this field. The aim of this study was not to verify and validate the use of Dolphin™, but to investigate the application of AI cephalometric landmark detection.

6.6 BoneFinder®

Automatic cephalometric analysis has been a topic of interest during the past years. Due to the increasing number of computer-assisted cephalometric programs and apps, there is a need for comparative studies to enable informed decision making amongst practitioners and researchers (Meric and Naoumova, 2020). The enticement is that AI can be a replacement for sufficient landmark detection. As previously stated random error can occur due to observer bias and experience (Durão *et al.*, 2015). Using Dolphin™, landmark detection and extraction was carried out to the best of the primary researcher's ability, ensuring consistent parameters were applied. However, an automated approach removes all subjectivity. Furthermore, BoneFinder did not require any calibration.

BoneFinder® was developed at the University of Manchester and is currently being used for research purposes. In contrast, to the computer-assisted human examination approach, AI offers objectivity. Like most AI programs (Meric and Naoumova, 2020), BoneFinder® is deterministic, i.e. the same image will give the same result every time (Lindner *et al.*, 2016). BoneFinder® is a full-automated landmark annotation (FALA) system based on the machine learning approach. Mitchell (1997) provided a comprehensive definition of machine learning: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*." With reference to this study, BoneFinder® learnt to detect landmarks and will improve its performance (precision) as measured by its ability to detect landmarks through experiences obtained by the training data set. Simply put, the landmark detection (*T*), as measured with precision (*P*), would improve with BoneFinder® training set (*E*). However, the training data set was limited. BoneFinder® was trained on only 400 digital cephalograms. AI is only as smart as the dataset it was trained on. A study by Moon *et al.* (2020) set out to determine the optimal quantity of learning data needed to develop artificial intelligence (AI) that can automatically identify cephalometric landmarks. As expected, the greater the quantity of learning data, the better the accuracy of AI. It was approximated that at least 2300 learning data sets, would be required to develop accurate and clinically applicable AI in orthodontics (Moon *et al.*, 2020).

Since the number of studies involving AI is rapidly growing, and suffer from many limitations (Juneja *et al.*, 2021; Schwendicke *et al.*, 2021); Schwendicke *et al.*, (2021) developed a checklist on planning, conduction and reporting of AI studies for researchers. When reviewing this; it was found that the study by Lindner *et al.* (2016) and their development of BoneFinder® had one downfall: they had not provided details of the source of the data (inclusion and exclusion criteria, the sampling framework and the target population). However, they did briefly discuss their limitations namely that the performance of their system was dependent on (1) the quality of the training data; (2) the size of the training dataset; and (3) the shape and appearance variation exhibited in the training data (e.g. age, type and degree of malformations).

Schwendicke *et al.*, (2021) also utilized the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines in their checklist. Reporting of observational research is often not detailed and clear enough to assess the strengths and weaknesses of the investigation. To improve the reporting of observational research, Vandenbroucke *et al.* (2007) developed a checklist of items that should be addressed: the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement (Appendix O). Research should be reported transparently. This enables readers to have a full understanding of the methodology, results and conclusion (von Elm *et al.*, 2014). This current study followed these guidelines.

BoneFinder®'s training set consisted of cephalograms of patients between the ages of 7-76. No inclusion and exclusion criteria were specified in the study by Wang *et al.*, (2016) and (Lindner *et al.*, 2016). Furthermore, the demographics (race) was not specified. If a European or Asian dataset was utilized, it may have influenced the training data set. As noted earlier, craniofacial patterns differ with patients of mixed ethnicities (Kula and Ghoneima, 2018). This was the first time BoneFinder® was used on a South African population, and this may have also been attributed to the discrepancies observed.

Furthermore, the BoneFinder® system may have also inherited some of the inaccuracies from the manual training data. Similar remarks were made by the creators of BoneFinder® (Lindner *et al.*, 2016). AI aims to be objective, but to identify landmarks, training needs to be *annotated by humans*. This in turn can be subjective, depending on the level of experience and knowledge of the examiners.

AI also offers other advantages apart from objectivity. The faster results can assist with workflow (Lindner *et al.*, 2016; Tang *et al.*, 2018; Yaji *et al.*, 2019). However, when using BoneFinder, a minor latency was experienced. A comparison between the time taken to conduct the detection and exportation of landmarks was not an objective in this study. This minor latency experienced is still negligible in comparison to the time taken for manual or computer-assisted landmark detection.

6.7 Landmarks and Case Examples

In this dataset, there were two kinds of landmarks: (a) *anatomic or identifiable landmarks* and (b) *derived or interpreted landmarks* (Figures 6.5-6.13). The former landmarks refer to anatomic structures that were clearly recognized and the latter are derived from neighbouring anatomic structures and require interpretation which may be subjective (Perillo *et al.*, 2000; Kwon *et al.*, 2021). Automated methods also suffered from inaccurate localization. The difficulty in the localization of L10 (Gonion) (one of the bilateral landmarks of the mandible) has been reported by Lindner *et al.* (2016) and Wang *et al.* (2016). This is usually caused by the asymmetry of the mandible (Kwon *et al.*, 2021). In this study, Gonion also showed the greatest standard deviation (SD=5.81).

Factors influencing landmark detection include image resolution, anatomical complexity, superimposition of the structures, the experience of the observers when locating a landmark, and manual measurement errors. As a rule for bilateral structures, when overlapping of the right and left anatomical structure such as the inferior border of the mandible, condyle, porion, orbitale, and teeth occurred, the observer "traced" the average part of bilateral structures before locating the landmark on the tracing line (Anuwongnukroh *et al.*, 2018).

According to McClure *et al.* (2005), the reliability of cephalometric landmarks is dependent on whether their dimensions are horizontal or vertical. The causes of these discrepancies in the distribution of landmark identification errors are usually linked to the definitions of the landmarks. This is frequently due to the landmark's anatomical variation. A landmark location linked with a more gradual curve, such as Gonion, Gnathion, Pogonion, and Menton, for example, would have less detection error than a sharp incisal edge. The vertical or horizontal orientation of the curve might impact errors in the former. Errors involving Gonion, Gnathion, Pogonion, and Menton can be influenced by the vertical or horizontal orientation of the curve. For example, landmarks such as Me, Go, ANS, and PNS are likely to have more x-axis error than y-axis error.

This study would be incomplete without addressing and relating the clinical implications of the differences at each landmark with the importance of the landmark identification error. Since cephalometric landmarks are used in conjunction with others to assess linear and angular measurements, error at each landmark site is of great clinical significance. Both the magnitude as well as the distribution of the landmark identification error is of importance when selecting a cephalometric analysis to arrive at diagnostic conclusions and treatment planning decisions (McClure *et al.*, 2005). The importance of the distribution of error for a given landmark is determined by the use of that landmark in various cephalometric analyses. If a landmark is to be used to determine the magnitude of a horizontal discrepancy of the jaws relative to one another in an angular measure such as SNA, SNB, and ANB, then the error of the landmarks A point and B point along the horizontal axis would be of greater significance than the error of these landmarks along the vertical axis (McClure *et al.*, 2005; Durão *et al.*, 2014; Tam and Lee, 2015).

Errors can also be induced by anatomical variations. As previously mentioned; unerupted maxillary third molars can obscure PNS (Figure 6.1). In Figure 6.5, the discrepancy between the soft tissue pogonion determined by the researcher (shown in red) and the automatically detected pogonion (green) locations are displayed. The disparity is most commonly produced by abnormal lip tension in individuals with forcefully corrected lip incompetence, which distorts the chin profile and causes it to deviate from the template, resulting in the detection discrepancy (Tam and Lee, 2015).

Some landmarks suffer from a combination of poor definition due to superimposition and uncertain interpretation. For example, Articulare (Figure 43) is composed of three independent bones: the inferior surface of the cranial base, and the posterior outlines of the mandibular rami and condyles. As per the bilateral rule; the observer "traced" the average part of bilateral structures before locating the landmark on the midpoint of the tracing line.

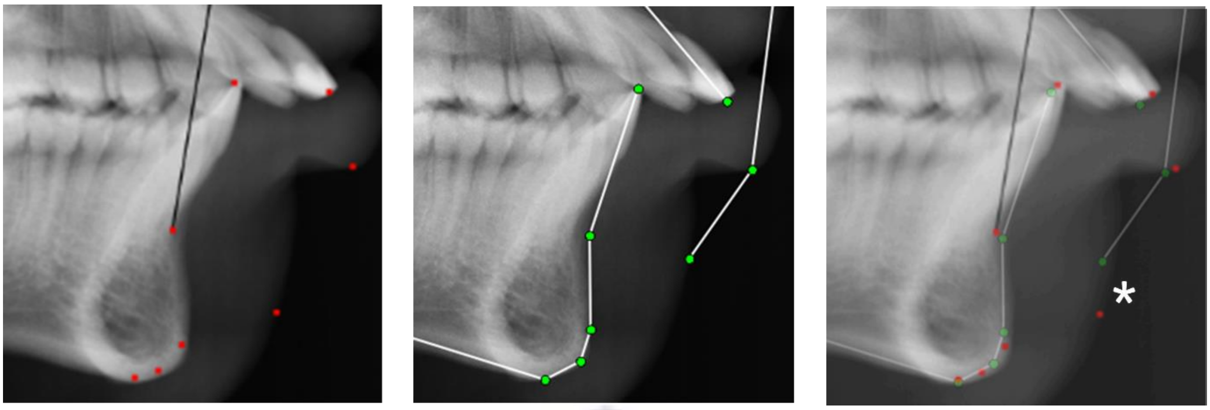


Figure 6.5: Effect of improper lip tension

Improper lip tension shifts landmark soft tissue pogonion to a position midway (green) between the human-detected pogonion (red) and the lower lip.

Right – landmark detection with Dolphin™; Middle – landmark detection with BoneFinder®; Left – Superimposed images.

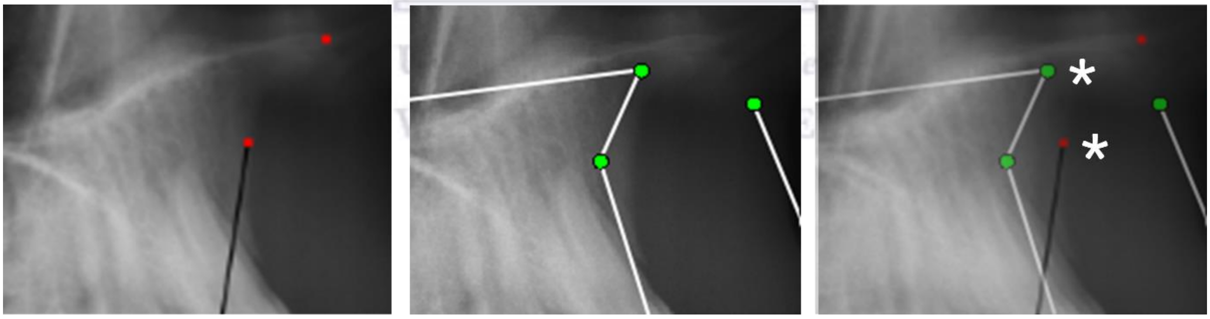


Figure 6.6: Differences of the automatically detected ANS and A point.

Right – landmark detection with Dolphin™; Middle – landmark detection with BoneFinder®; Left – Superimposed images.

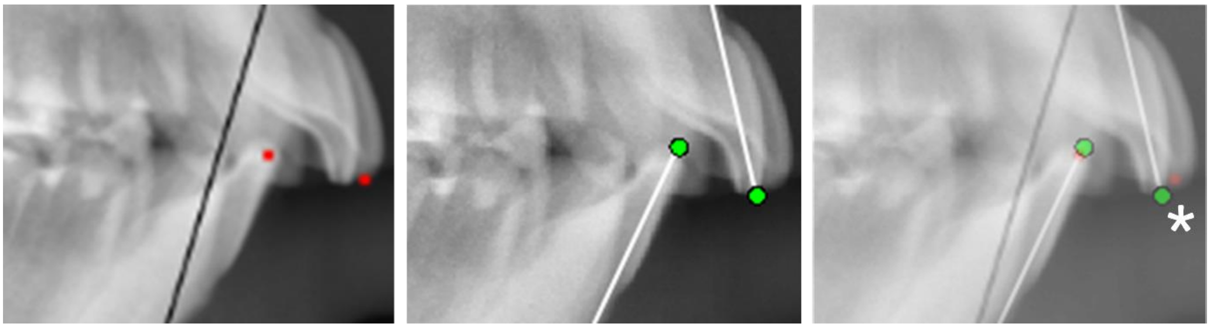


Figure 6.7: Effect of crowded anterior teeth

BoneFinder® did not select the most inferior aspect of the incisal tip of the most labially places maxillary incisor. Right – landmark detection with Dolphin™; Middle – landmark detection with BoneFinder®; Left – Superimposed images.

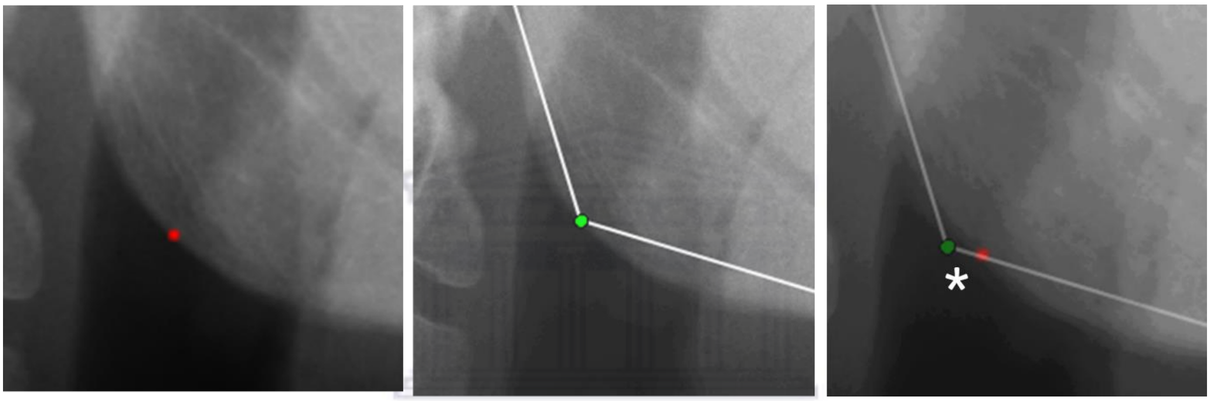


Figure 6.8: Gonion landmark

A discrepancy was noted between the human-detected landmark and the automatically detected landmark. Right – landmark detection with Dolphin™; Middle – landmark detection with BoneFinder®; Left – Superimposed images

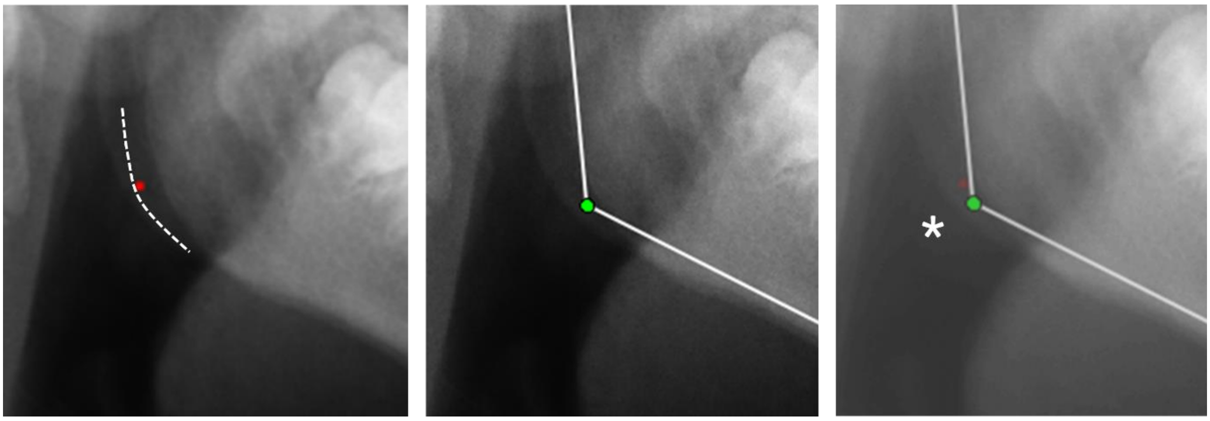


Figure 6.9: Bilateral rule for detecting gonion

Right – landmark detection using Dolphin™; Middle – landmark detetcion with BoneFinder®, Left- Superimposed images. Using the bilateral rule, Gonion is detected on the most inferior and posterior border (white dashed line) of the mandible.

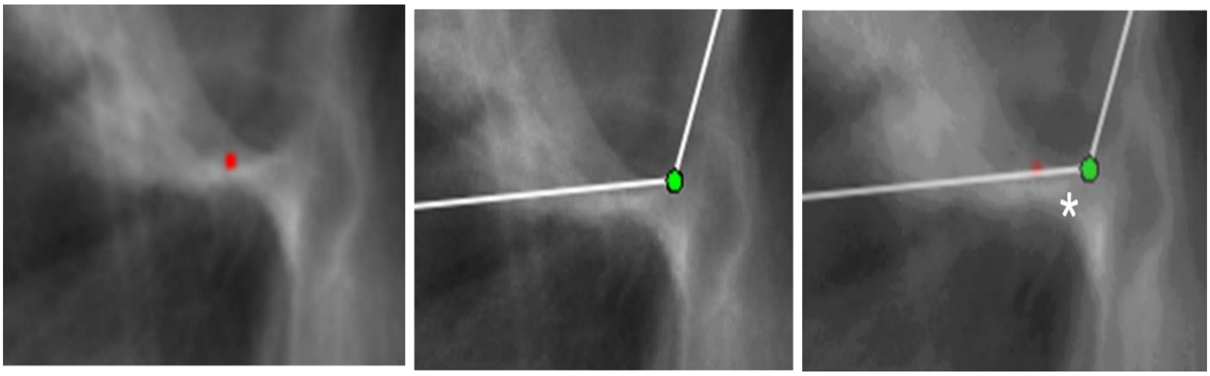


Figure 6.10: Detection of Orbitale

Right – landmark detection with Dolphin^{TM}; Middle – landmark detection with BoneFinder®; Left – Superimposed images. Discrepancy noted in detecting orbitale.

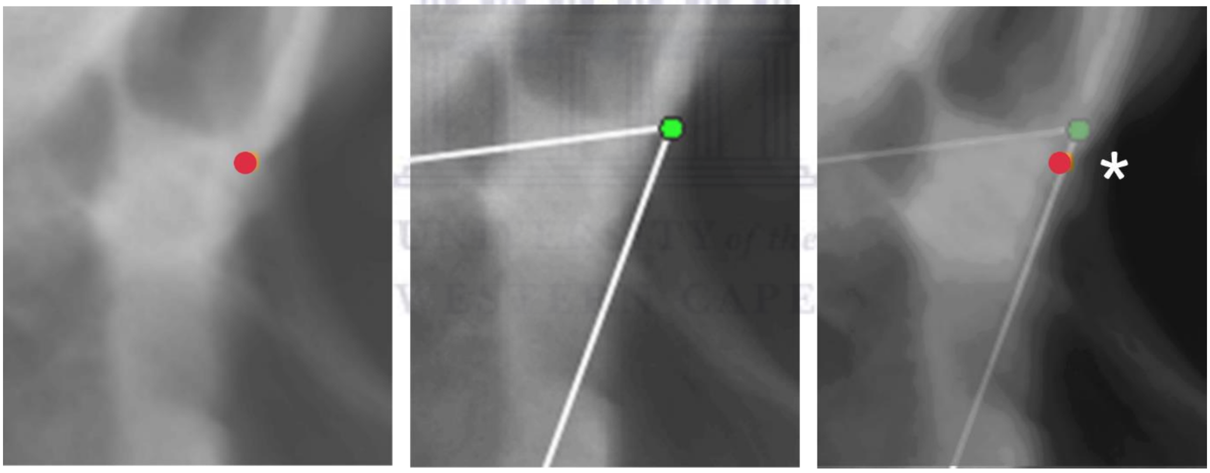


Figure 6.11: Detection of nasion

Right – landmark detection with Dolphin^{TM}; Middle – landmark detection with BoneFinder®; Left – Superimposed images. Fronto-nasal suture is not easily detectable in this case. Discrepancy of human detected landmark for nasion, and automatically detected landmark.

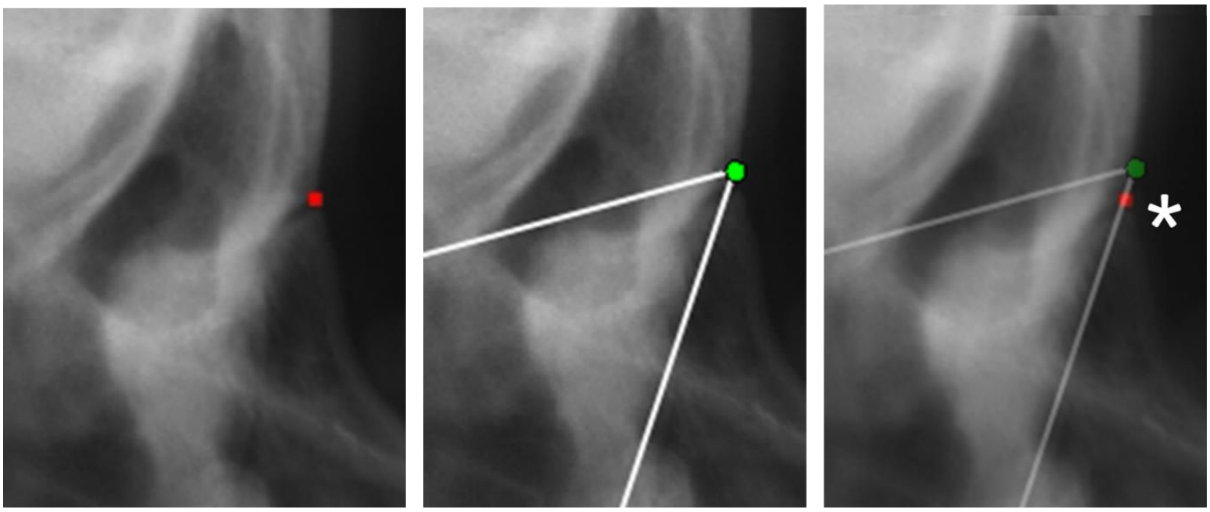


Figure 6.12: Detecction of nasion

Right – landmark detection with Dolphin™; Middle – landmark detection with BoneFinder®; Left – Superimposed images. Fronto-nasal suture is easily detectable in this case. Close proximity of human detected landmark for nasion, and automatically detected landmark.

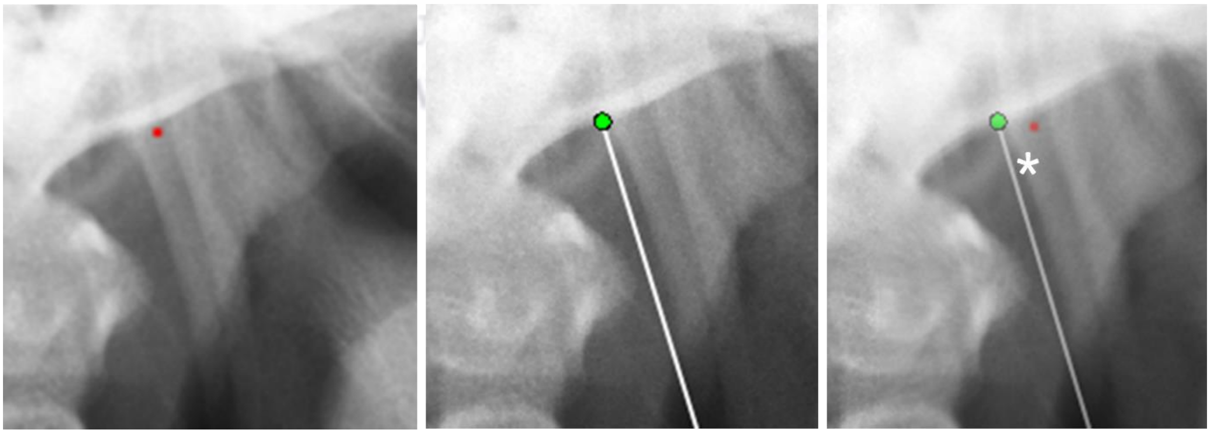


Figure 6.13: Detection of articulare

Right – landmark detection with Dolphin™; Middle – landmark detection with BoneFinder®; Left – Superimposed images. Interpretation required to determine location of articulare. BoneFinder® did not utilize the bilateral rule

6.8 Incidental Findings

According to Ells and Thombs (2014), incidental findings are becoming directly proportional to the advancement of medical technologies used in clinical treatment and research. These potentially important discoveries fall outside the scope of conducting a study or clinical test. In this study, unanticipated findings were detected. These were findings that were unknown to the researcher to be associated with the investigation, leading to heuristic results. As mentioned above, the explanations for the large discrepancy in the Euclidean distances were actively sought but were not the primary motive for the study.

BoneFinder® provided coordinates in millimetres if the file input was in DICOM format. JPEG files produced coordinates in pixels. In view of this and the consensus that DICOM files have the highest image quality (Graham, *et al.*, 2005; Faccioli *et al.*, 2009; Varma, 2012; Burgess, 2015), DICOM files were used in both methods.

An incidental finding revealed that if one compared the same cephalogram using BoneFinder® coordinates (using a DICOM file) with the Dolphin™ coordinates (JPEG file), the results were more comparable to those in the literature, i.e. within the accepted 2mm range (Wang *et al.*, 2016; Juneja *et al.*, 2021). The Euclidean distances also appeared more comparable to studies in the literature (the mean Euclidean distance was 2.15mm). Interestingly, when comparing Dolphin™ with itself, the coordinates extracted from the DICOM and the JPEG files differed. This was contradictory to the results of Saez *et al.*, (2016) and Saghaie and Ghaffari (2014) where they evaluated the influence of DICOM and JPEG formats on cephalometric landmarks detection and found JPEG file formats to be reliable.

When re-evaluating Dolphin™'s parameters, it was apparent that changes to the calibration ruler significantly changed the results. The possibility of inaccurate measurement of the nasion-positioning rod was explored. Table 5.12 shows that when this distance was changed to 31mm, it considerably altered the output. As mentioned previously, calibration is important. In this study, the lack of a ruler during the acquisition stage meant that calibration needed to be performed using two fixed points. This measurement of a known distance (30 mm) between the two fixed corner points of the nasion-guiding rod on the screen was chosen. However, this also could have introduced random error as the placement of the mouse-driven cursor was highly sensitive. The actual process of the calibration was also not perfectly repeatable; therefore, uncertainty was introduced through the calibration process. In Table 12, a change of 1mm increased the Euclidean distance, further showing the sensitivity of Dolphin™'s parameters.

Overall, it was difficult to draw a robust comparison, as Dolphin™ was very sensitive to parameters leading to fluctuating results. Together these results provide important insights as well as raises more questions into AI and computer-assisted approaches.

6.9 Opinion of AI in Cephalometric Analysis:

The first attempt at automated landmarking of cephalograms was made by Cohen in 1984. Computer-assisted cephalometric analysis eliminates the mechanical errors that occur during manual tracing i.e. identifying landmarks, drawing lines between landmarks, reading measurements off a protractor. However, inconsistency in landmark identification is still an important source of random errors both in computer-assisted digital cephalometry and in manual cephalometric analysis (Leonardi *et al.*, 2008). According to Leonardi *et al.* (2008), variability in landmark detection has been determined to be five times greater than measurement variability, with both methods open to considerable subjectivity. Although to a different extent, both manual and computer-assisted cephalometry methods are time-consuming. Computer-assisted cephalometry can be likened to a manual digital system, i.e. although the landmarks are detected digitally on-screen, they still must be manually pinpointed using a mouse-driven cursor.

Entering the arena, to alleviate such problems is AI which offers objective, quick and deterministic values. Overall, these random errors resulting from uncertainty may be eliminated by automated programs (Tam and Lee, 2015).

Some researchers have been infatuated with the thought that all the conclusions drawn from AI studies are much more optimistic than reality, allowing readers to think that automatic cephalometric analysis will be available very shortly (Leonardi *et al.*, 2008). This is not wrong, but a cautious approach needs to be taken. In fact, Anuwongnukroh *et al.* (2018) used Carestream Dental, Version 6.14, which is a fully automatic cephalometric analysis program, to determine and compare the reliability of a fully automatic cephalometric analysis software with manual cephalometric tracing. They concluded that automated programs should only be used to *support a diagnosis* and not as a diagnostic tool. Recommendations included that the operator must review, check and change all landmarks that are inaccurately identified by the software before completion of cephalometric analysis. Pauwels (2020) agrees with Anuwongnukroh *et al.* (2018) and suggests that AI could even be an option as a 'first pass' analysis to save time, by highlighting potential outliers that warrant further evaluation from the orthodontist.

The use of automated software may reduce the errors that occur during computer-assisted approaches. Some landmarks, particularly those involving crowded maxillary and mandibular incisors, are difficult to identify; hence, such structures have been shown to have low reliability in digital tracings, despite the possibility of using filtering and zooming (Meric and Naoumova, 2020).

Inconsistency in landmark detection is specific to each landmark and can be affected by the experience and training of the observers, individual anatomical variations and image quality such as radiographic film magnification (Chen *et al.*, 2000; Tam and Lee, 2015). Many efforts have been developed to automate computerized identification of cephalometric landmarks. One needs to remember that two cephalometric points are needed to trace a reference plane or line, the resulting special position of the line will be affected by the errors of two points, not a single one, and thus the error will be increased (Leonardi *et al.*, 2008). This emphasizes the importance of landmark detection. By creating an objective AI program, all uncertainty is removed.

Skeletal classification is only possible by assessing the vertical and anteroposterior locations of the jaws in the cephalograms. This is a crucial aspect of orthodontic diagnosis and treatment planning. As a result, successful treatment is directly related to proper landmark identification and analysis which in turn contributes to an accurate diagnosis. It can thus be seen that accurate landmark detection is the most critical and sensitive procedure in cephalometric analysis, but is highly time-consuming with the potential for errors and variability. By eliminating the landmark identification process as a whole, the diagnostic process is expected to be expedited with better accuracy (Yu *et al.*, 2020).

It is worth noting that with machine learning, algorithms are only as good as the training sets used to train the system (Ahuja, 2019). As shown in Hwang’s (2019) study, an AI model could reach or even surpass the performance of human observers (Pauwels, 2020). Hwang *et al.* (2019) also set out to compare landmark detection patterns of 80 cephalometric landmarks identified by an automated identification system. Their system was trained on 1028 cephalograms. Their study was more relevant and clinically applicable to clinicians as it tested whether this new AI method was better and more reliable than clinically experienced human experts. In general, comparisons in the detection errors between AI and human examiners were less than 0.9 mm, which did not seem to be clinically significant. It was found that the AI system always detected identical positions, upon repeated trials.

Despite the flaws that still exist within AI, the benefits far outweigh these faults. AI holds the promise that it may be a more reliable option for repeatedly identifying multiple cephalometric landmarks.

CHAPTER 7 : CONCLUDING REMARKS

Several interesting findings were discovered during this study. Having discussed the differences between Dolphin Imaging^{TM} and BoneFinder$^{®}$, the final section of this thesis addresses the limitations, recommendations and conclusion of this study.

7.1 Limitations

Several limitations need to be considered. First, as previously mentioned, "when it comes to a reliability measure when identifying a certain cephalometric landmark, there is no firm 'ground truth' or gold standard that can provide validation as to where the true location of the landmark is" (Hwang *et al.* 2019). As a result, no "gold standard" was used; however, the primary researcher was calibrated to conform to manual ground truth.

The sample size was also relatively small due to the lack of records that complied with the inclusion and exclusion criteria. It is also worth noting that 147 cephalograms were excluded due to errors related to ruler placement or the absence thereof. Furthermore, 128 cephalograms were also excluded due to incorrect head positioning and movement. It would aid future studies utilizing cephalograms if the cephalograms taken could be standardized. The majority of cephalograms at the institution were performed by undergraduate dental and oral hygiene students; although images were diagnostically acceptable, they may not always be optimal.

As recommended by Durão *et al.* (2015), precise landmark location requires a sufficiently high-quality digital cephalometric image for landmark identification, with the ruler visible on the film, allowing image calibration in the cephalometric analysis software program. Since these cephalograms did not have a ruler, calibration may have been affected. The re-exposure of patients for the sole purpose of obtaining data for a study would be unethical, as a result, the cephalograms lacking the ruler were utilized and efforts to calibrate the image were made.

According to the literature, landmark identification is the main source of errors. The factors contributing to the identification error are examiner experience (the primary researcher had 3 years of experience in the Oral and Maxillofacial Radiology Department), landmark definition, and the density and sharpness of the image. Furthermore, operator fatigue and subjectivity may also have contributed to discrepancies in landmark detection.

BoneFinder$^{®}$ is freely available for non-commercial research purposes, and this software also exhibited several limitations, namely: the quality of the data set that this programme was trained on, (2) the size that it was trained on (400) and (3) the shape and appearance variation exhibited in the training data (e.g. age, type and extent of malformations). This was the first time that BoneFinder® was used on a sample of a South African population, and changes in bone structure and may have contributed to outliers.

Such limitations mean that these findings need to be interpreted with caution. The current study only examined the DICOM images and did not compare the two methods with JPEG images. Furthermore, the study was not designed to evaluate factors related to reference frames used by both modalities.

The study was also limited by the lack of literature available on the topic. According to the primary researcher's current knowledge and the literature reviewed; no similar study was found. The nearest latest and relevant references were based on findings of studies on automatic landmark detection using both deep learning and machine learning approaches.

7.2 Recommendations

The proposed research intended to provide a means for precise detection of cephalometric landmarks within a South African context. This was to substantiate the benefit of implementing fully automated cephalometric landmark detection programmes in orthodontic practices that will ultimately assist with workflow and improve treatment planning with increased precision. The results in this study were very sensitive to several variables. Therefore, several questions remain unanswered at present. However, there is abundant room for further progress in determining whether AI can replace computer-assisted landmark detection approaches.

A number of possible future studies using the same experimental setup are apparent: (1) a study similar to this one should be carried out with correctly calibrated cephalometric images to explore BoneFinder®'s reliability within a South African context, (2) the dataset used by Lindner *et al.* (2016) in their development of BoneFinder® could be used to compare the precision to Dolphin™. It is also suggested that the association of reference frames is investigated studies.

There is also a lack of robustness of available training datasets and this is influenced by the inaccessibility of standard datasets (Juneja *et al.*, 2021). It would be beneficial to create an open standard South African dataset with the ground truth marked and validated by experienced clinicians for future research into automated landmark detection. Due to the significant variation in anatomical features among different ethnic groups, the datasets would also need to be representative of each ethnic group.

It would be interesting to create an automated landmark detection system that is trained on an African dataset. Data sets are usually trained according to inclusion and exclusion criteria, however, distortion to the skull caused by diseases etc need to also be included. This will enable an AI system to detect landmarks on anomalous skulls. A comprehensive medical history should also be included during the acquisition of the training set.

Before AI can be fully adopted in a clinical setting in South Africa, further studies determining South African cephalometric norms should be carried out. A better understanding of this would also contribute to a South African data set of cephalograms.

Two-dimensional (2D) cephalograms are questionable due to several limitations, such as the superimposition, magnification and distortion, and the influence of head position during image acquisition (De Oliveira Lisboa *et al.*, 2014). The Faculty of Dentistry, Tygerberg, is a training institution for dental and oral hygiene students, and as a result, special attention should be paid to ensure the students are taking their radiographs correctly and that is not only of acceptable diagnostic quality but also so that standardization is ensured.

According to Tam and Lee (2015) to improve image quality**,** it is recommended that in addition to using ear position rods for head stabilization, additional stabilizing points are required to position the head with optimal symmetry. Patients must also be instructed to raise their heads when cephalograms are captured so that the mandibular rami appear distinct from the vertebrae. Due to the indistinct demarcation between neighbouring landmarks, to improve the accuracy of results, high-resolution images in the data sets must also be included.

7.3 Conclusion

In conclusion, and within the limitations outlined above, the null hypothesis was accepted – there was no significant difference between the artificial intelligence programme (BoneFinder®) and the computer-assisted human examination (Dolphin Imaging™) regarding the precision of landmark detection. Whilst this study did not confirm that AI is superior to computer-assisted human examination, it did partially substantiate the use of AI in radiology and orthodontics. AI should not be seen as a replacement, but rather an aid. Taken together, these findings suggest a promising role for the future of AI in cephalometry.

CHAPTER 8 : BIBLIOGRAPHY

Abdelkarim, A. (2012) ‘Myths and facts of cone-beam computed tomography in orthodontics’, *Journal of the World Federation of Orthodontists*, 1(1), pp. e3–e8. doi: 10.1016/j.ejwf.2012.04.002.

Agrawal, J. M. Agrawal, M. S., Nanjannawar, L. G. and Parushetti, A. D. (2013) ‘CBCT in orthodontics: The wave of future’, *Journal of Contemporary Dental Practice*, 14(1), pp. 153–157. doi: 10.5005/jp-journals-10024-1291.

Ahuja, A. S. (2019) ‘The impact of artificial intelligence in medicine on the future role of the physician’, *PeerJ*, 2019(10). doi: 10.7717/peerj.7702.

Aksakalli, S., Yilanci, H., Gorukmez, E. and Ramoglu, S. I. (2017) ‘Reliability Assessment of Orthodontic Apps for Cephalometrics’, *Turkish Journal of Orthodontics*, 29(4), pp. 98–102. doi: 10.5152/turkjorthod.2016.1618.

Albarakati, S. F., Kula, K. and Ghoneima, A. A. (2012) ‘The reliability and reproducibility of cephalometric measurements: a comparison of conventional and digital methods’, *Dentomaxillofacial Radiology*, 41(1), pp. 11–17. doi: 10.1259/dmfr/37010910.

American Academy of Oral and Maxillofacial Radiology (2013) ‘Clinical recommendations regarding use of cone-beam computed tomography in orthodontic treatment. Position statement by the American Academy of Oral and Maxillofacial Radiology’, *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 116(2), pp. 238–257. doi: 10.1016/j.oooo.2013.06.002.

Anuwongnukroh, N., Dechkunakorn, S., Damrongsri, S., Nilwarat, C., Pudpong, N., Radomsutthisarn, W. and Kangern, S. (2018) ‘Assessment of the reliability of automatic cephalometric analysis software’, *International Journal of Mechanical Engineering and Robotics Research*, 7(1), pp. 61–65. doi: 10.18178/ijmerr.7.1.61-65.

Anvari, A., Halpern, E. F. and Samir, A. E. (2015) ‘Statistics 101 for radiologists’, *Radiographics*, 35(6), pp. 1789–1801. doi: 10.1148/rg.2015150112.

Barter, M. A., Evans, W. G., Smit, G. L. and Becker, P. J. (1995) ‘Cephalometric Analysis of a Sotho-Tswana Group’, *Journal of the Dental Association of South Africa*, 50(11), pp. 539–544.

Burgess, J. (2015) ‘Digital DICOM in Dentistry’, *The Open Dentistry Journal*, 9(1), pp. 330–336. doi: 10.2174/1874210601509010330.

Cattaneo, P. M., Bloch, C. B., Calmar, D., Hjortshøj, M. and Melsen, B. (2008) ‘Comparison between conventional and cone-beam computed tomography-generated cephalograms’, *American Journal of Orthodontics and Dentofacial Orthopedics*, 134(6), pp. 798–802. doi: 10.1016/j.ajodo.2008.07.008.

Chen, Y. J., Chen, S. K., Chang, H. F. and Chen, K. C. (2000) ‘Comparison of Landmark Identification in Traditional Versus Computer-Aided Digital Cephalometry’, *Angle Orthodontist*, 70(5), pp. 387–392. doi: 10.1043/0003-3219(2000)070<0387:COLIIT>2.0.CO;2.

Coan, T. E. (2006) *Measurement and Error - SMU Physics*. *Measurement and Error - SMU.*

Available at: http://www.physics.smu.edu/~coan/1313/labs/errors.doc (Accessed: 23 July 2021)

da Silveira, H. L. D. and Silveira, H. E. D. (2006) ‘Reproducibility of Cephalometric Measurements Made by Three Radiology Clinics’, *Angle Orthodontist*, 76(3), pp. 394–399.

de Lima Navarro, R. Oltramari-Navarro, P.V.P., Fernandes, T. M. F., de Oliveira, G. F., Conti, A. C. D. F., de Almeida, M. R. and de Almeida, R. (2013) ‘Comparison of manual, digital and lateral CBCT cephalometric analyses’, *Journal of Applied Oral Sciences*, 21(2), pp. 167–176. doi: http://dx.doi.org/10.1590/1678-7757201302326.

De Oliveira Lisboa, C., Masterson, D., da Motta, A. F. J. and Motta, A. T. (2014) ‘Reliability and reproducibility of three-dimensional cephalometric landmarks using CBCT: A systematic review’, *Journal of Applied Oral Science*, 23(2), pp. 112–119. doi: 10.1590/1678-775720140336.

Deshmukh, S. V. (2018) ‘Artificial Intelligence in Dentistry’, *Journal of International Clinical Dental Research Organization*, 10(2), pp. 47–48. doi: 10.4103/jicdro.jicdro.

Dreyer, K. and Allen, B. (2018) ‘Artificial Intelligence in Health Care: Brave New World or Golden Opportunity?’, *Journal of the American College of Radiology*, 15(4), pp. 655–657. doi: 10.1016/j.jacr.2018.01.010.

Dunn, P. K. (2019) *Scientific Research Methods: An introduction to quantitative research in science and health*. Available at: https://srm-course.netlify.com (Accessed: 15 January 2021).

Durão, A. P. R., Morosolli, A., Pittayapat, P., Bolstad, N., Ferreira, A. P. and Jacobs, R. (2015) ‘Cephalometric landmark variability among orthodontists and dentomaxillofacial radiologists: A comparative study’, *Imaging Science in Dentistry*, 45(4), pp. 213–220. doi: 10.5624/isd.2015.45.4.213.

Durão, A. R., Bolstad, N., Pittayapat, P., Lambrichts, I., Ferreira, A. P. and Jacobs, R. (2014) ‘Accuracy and reliability of 2D cephalometric analysis in orthodontics’, *Revista Portuguesa de Estomatologia, Medicina Dentaria e Cirurgia Maxilofacial*, 55(3), pp. 135–141. doi: 10.1016/j.rpemd.2014.05.003.

Ells, C. and Thombs, B. D. (2014) ‘The ethics of how to manage incidental findings’, *Canadian Medical Association Journal*, 186(9), pp. 655–656. doi: 10.1503/cmaj.140136.

Faccioli, N. Perandini, S., Comai, A., D'Onofrio, M. and Pozzi M. R. (2009) ‘Proper use of common image file formats in handling radiological images’, *Radiologia Medica*, 114(3), pp. 484–495. doi: 10.1007/s11547-009-0378-6.

von Fraunhofer, J. A. (2010) *Research Writing in Dentistry*. First. Iowa: Wiley-Blackwell.

Graham, R. N. J., Perriss, R. W. and Scarsbrook, A. F. (2005) ‘DICOM demystified: A review of digital file formats and their use in radiological practice’, *Clinical Radiology*, 60(11), pp. 1133–1140. doi: 10.1016/j.crad.2005.07.003.

Gribel, B. F., Gribel, M. N., Frazão, D. G., McNamara, J. A. and Manzi, F. R. (2011) ‘Accuracy and reliability of craniometric measurements on lateral cephalometry and 3D measurements on CBCT scans’, *Angle Orthodontist*, 81(1), pp. 28–37. doi: 10.2319/032210-166.1.

Halazonetis, D. J. (1994) ‘Computer-assisted cephalometric analysis’, *American Journal of Orthodontics and Dentofacial Orthopedics*, 105(5), pp. 517–521.

Hwang, H., Park, J., Moon, J., Yu, Y., Kim, H., Her, S., Srinivasan, G., Aljanabi, M. N. A., Donatelli, R. E. and Lee, S. (2019) ‘ Automated identification of cephalometric landmarks: Part 2-Might it be better than human? ’, *The Angle Orthodontist*, 00(00), pp. 1–8. doi: 10.2319/022019-129.1.

ISO (1998) *ISO 5725-1 (1998) Accuracy (trueness and precision) of measurement methods and results–Part 1: General principles and definitions.* Geneva. Available at: https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en. (Accessed: 23 July 2021)

Jones, M., Dobson, A. and O’Brian, S. (2011) ‘A graphical method for assessing agreement with the mean between multiple observers using continuous measures’, *International Journal of Epidemiology*, 40(5), pp. 1308–1313. doi: 10.1093/ije/dyr109.

Juneja, M., Garg, P., Kaur, R., Manocha, P., Prateek, N. V. and Batra, S. (2021) ‘A review on cephalometric landmark detection techniques’, *Biomedical Signal Processing and Control*, 66(November 2019), p. 102486. doi: 10.1016/j.bspc.2021.102486.

Kapila, S., Conley, R. S. and Harrell, W. E. (2011) ‘The current status of cone-beam computed tomography imaging in orthodontics’, *Dentomaxillofacial Radiology*, 40(1), pp. 24–34. doi: 10.1259/dmfr/12615645.

Katkar, R. A., Kummet, C., Dawson, D., Moreno U. L., Allareddy, V., Finkelstein, M. and Ruprecht, A. (2013) ‘Comparison of observer reliability of three-dimensional cephalometric landmark identification on subject images from Galileos and i-CAT cone-beam CT’, *Dentomaxillofacial Radiology*, 42(9), pp. 1–11. doi: 10.1259/dmfr.20130059.

Kula, K. and Ghoneima, A. (2018) ‘A Perspective on Norms and Standards’, in Kula, K. and Ghoneima, A. (eds) *Cephalometry in orthodontics: 2D and 3D*. 1st edn. Quintessence Publishing Co, Inc, pp. 83–86. doi: 10.1016/j.ajodo.2019.04.021.

Kwon, H. J., Koo, H. I., Park, J. and Cho, N. I. (2021) ‘Multistage Probabilistic Approach for the Localization of Cephalometric Landmarks’, *IEEE Access*, 9, pp. 21306–21314. doi: 10.1109/ACCESS.2021.3052460.

Leonardi, R. Giordano, D., Maiorana, F. and Spampinato, C. (2008) ‘Automatic cephalometric analysis: A systematic review’, *Angle Orthodontist*, 78(1), pp. 145–151. doi: 10.2319/120506-491.1.

Lindner, C., Wang, C.W., Huang, C. T, Li, C. H., Chang, S. W. and Cootes, T. F. (2016) ‘Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms’, *Scientific Reports*, 6(September), pp. 1–10. doi: 10.1038/srep33581.

Machado, G. L. (2015) ‘CBCT imaging - A boon to orthodontics’, *Saudi Dental Journal*, 27(1), pp. 12–21. doi: 10.1016/j.sdentj.2014.08.004.

Manosudprasit, A. Haghi, A., Allareddy, V. and Masoud, M.I. (2017) ‘Diagnosis and treatment planning of orthodontic patients with 3-dimensional dentofacial records’, *American*

Journal of Orthodontics and Dentofacial Orthopedics, 151(6), pp. 1083–1091. doi: 10.1016/j.ajodo.2016.10.037.

McClure, S. R. Sadowsky, P. L., Ferreira, A. and Jacobson, A. (2005) 'Reliability of digital versus conventional cephalometric radiology: A comparative evaluation of landmark identification error', *Seminars in Orthodontics*, 11(2), pp. 98–110. doi: 10.1053/j.sodo.2005.04.002.

Menditto, A., Patriarca, M. and Magnusson, B. (2007) 'Understanding the meaning of accuracy, trueness and precision', *Accreditation and Quality Assurance*, 12(1), pp. 45–47. doi: 10.1007/s00769-006-0191-z.

Meric, P. and Naoumova, J. (2020) 'Web-based Fully Automated Cephalometric Analysis: Comparisons between App-aided, Computerized, and Manual Tracings', *Turkish Journal of Orthodontics*, 33(3), pp. 142–149. doi: 10.5152/turkjorthod.2020.20062.

Miloro, M., Borba, A. M., Ribeiro-Junior, O., Naclério-Homem, M. G. and Jungner, M. (2014) 'Is there consistency in cephalometric landmark identification amongst oral and maxillofacial surgeons ?', *International Journal of Oral and Maxillofacial Surgery*, 43(4), pp. 445–453. doi: 10.1016/j.ijom.2013.08.007.

Mitchell, L., Mitchell, D. A. and McCaul, L. (2011) 'Orthodontics', in *Oxford Handbook of Clinical Dentistry*. 5th edn. New York: Oxford University Press, p. 120.

Mitchell, T. M. (1997) 'Machine Learning'. McGraw-Hill, p. 2.

Moon, J. H., Hwang, H. W., Yu, Y., Kim, M. B., Donatelli, R. E. and Lee, S. J. (2020) 'How much deep learning is enough for automatic identification to be reliable? A cephalometric example', *Angle Orthodontist*, 90(6), pp. 823–830. doi: 10.2319/021920-116.1.

Mosleh, M. A. A., Baba, M. S, Malek, S. and Almaktari, R. A. (2016) 'Ceph-X: Development and evaluation of 2D cephalometric system', *BMC Bioinformatics*, 17(Suppl 19). doi: 10.1186/s12859-016-1370-5.

Muelaner, J. (2018) *Evaluating Uncertainty of Measurement*. Available at: https://www.muelaner.com/uncertainty-of-measurement/ (Accessed: 23 July 2021).

Naidoo, L. C. D. and Miles, L. P. (1997) 'An evaluation of the mean cephalometric values for orthognathic surgery for black South African adults. Part I: Hard Tissue', *Journal of the Dental Association of South Africa*, 52(9), pp. 495–502.

Nervina, J. (2012) 'Cone beam computed tomography use in orthodontics', *Australian Dental Journal*, 57, pp. 95–102. doi: 10.1111/j.1834-7819.2011.01662.x.

Noar, J. H. and Pabari, S. (2013) 'Cone beam computed tomography – current understanding and evidence for its orthodontic applications?', *Journal of Orthodontics*, 40(1), pp. 5–13. doi: 10.1179/1465313312Y.0000000040.

Ongkosuwito, E. M., Katsaros, C., van't Hof, M. A., Bodegom, J. C. and Kuijpers-Jagtman, A. M. (2002) 'The reproducibility of cephalometric measurements : a comparison of analogue and digital methods', *European Journal of Orthodontics 24*, 24(6), pp. 655–665.

Paixão, M. B., Sobral, M. C., Vogel, C. J. and de Araujo, T. M. (2010) 'Comparative study between manual and digital cephalometric tracing using Dolphin Imaging software with lateral radiographs', *Dental Press Journal of Orthodontics*, 15(6), pp. 123–130. doi: 10.1590/S2176-94512010000600016.

Pakdemirli, E. (2019) 'Artificial intelligence in radiology: friend or foe? Where are we now and where are we heading?', *Acta Radiologica Open*, 8(2), p. 205846011983022. doi: 10.1177/2058460119830222.

Park, J., Hwang, H., Moon, J., Yu, Y., Kim, H., Her, S., Srinivasan, G., Aljanabi, M. N A., Donatelli, R. E. and Lee, S. (2019) 'Automated identification of cephalometric landmarks: *Part 1—Comparisons between the latest deep-learning methods YOLOV3 and SSD*', *The Angle Orthodontist*, 00(00), pp. 022019–127.1. doi: 10.2319/022019-127.1.

Pauwels, R. (2020) 'A brief introduction to concepts and applications of artificial intelligence in dental imaging', *Oral Radiology*, (0123456789). doi: 10.1007/s11282-020-00468-5.

Pereira, U. L., Kalia, D., Batham, P. R., Pujari, P., Chitalia, A., Prasad, S. and Dilliwal, S. (2015) 'Cone Beam Computed Tomography: A must in orthodontic diagnosis and treatment planning?', *International Journal of Stomatology & Occlusion Medicine*, 8(2), pp. 1–7. doi: 10.1016/j.cden.2018.03.002.

Perillo, M. A., Beideman, R. W., Shofer, F. S., Jacobsson-Hunt, U., Higgins-Barber, K., Laster, L. L. and Ghafari, J. G. (2000) 'Effect of landmark identification on cephalometric measurements: Guidelines for cephalometric analyses', *Orthodontics and Craniofacial Research*, 3(1), pp. 29–36. doi: 10.1034/j.1600-0544.2000.030106.x.

Phulari, B. S. (2013). *An Atlas on Cephalometric Landmarks* (First). Jaypee Brothers Medical Publishers.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rozylo-Kalinowska, I. (2018) 'Artificial intelligence in dentomaxillofacial radiology: Hype or future?', *Journal of Oral and Maxillofacial Radiology*, 6(1), pp. 1–2. doi: 10.4103/jomr.jomr.

Sadowsky, P. L. (2006) 'The Geometry of Cephalometry', in Jacobsen, A. and Jacobsen, R. L. (eds) *Radiographic Cephalometry: From Basics to 3D Imaging*. 2nd edn. Chicago: Quintessencces Publishing Co, Inc, pp. 137–151.

Saez, D. M., Bommarito, S. and Sannomiya, E. K. (2016) 'Reproducibility of cephalometric landmarks on posteroanterior digital radiographs using DICOM and JPEG formats', *Oral Radiology*, 32(2), pp. 79–86. doi: 10.1007/s11282-015-0217-5.

Saghaie, S. and Ghaffari, R. (2014) 'Effect of image compression of digital lateral cephalograms on the reproducibility of cephalometric points', *Dental Research Journal*, 11(1), pp. 27–31. doi: 10.1259/dmfr/40996636.

Sana, M. (2018) 'Machine Learning and Artificial Intelligence in Radiology', *Journal of the American College of Radiology*, 15(8), pp. 1139–1142. doi: 10.1016/j.jacr.2017.11.015.

Scarfe, W. C. and Farman, A. G. (2008) 'What is Cone-Beam CT and How Does it Work?', *Dental Clinics of North America*, 52(4), pp. 707–730. doi: 10.1016/j.cden.2008.05.005.

Schwendicke, F., Schwendicke, F., Singh, T., Lee, J. H., Gaudin, R., Chaurasia, A., Wiegand, T., Uribe, S. and Krois, J. (2021) 'Artificial intelligence in dental research: Checklist for authors, reviewers, readers', *Journal of Dentistry*, 107(October 2020). doi: 10.1016/j.jdent.2021.103610.

Seedat, A. K. (1983) 'Cephalometric Analysis of a Group of Cape Coloureds', *Journal of the Dental Association of South Africa*, 38(11), pp. 673–675.

Shahidi, S., Bahrampour, E., Soltanimehr, E., Zamani, A., Oshagh, M., Moattari, M. and Mehdizadeh, A. (2014) 'The accuracy of a designed software for automated localization of craniofacial landmarks on CBCT images', *BMC Medical Imaging*, 14(1), pp. 1–8. doi: 10.1186/1471-2342-14-32.

Stallings, W. M. and Gillmore, G. M. (1971) 'A Note on "Accuracy" and "Precision"', *Journal of Educational Measurement*, 8(2), pp. 127–129.

Tadinada, A. (2019) 'Artificial intelligence, machine learning, and the human interface in medicine: Is there a sweet spot for oral and maxillofacial radiology?', *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 127(4), pp. 265–266. doi: 10.1016/j.oooo.2018.12.024.

Tam, W. K. and Lee, H. J. (2015) 'Improving point correspondence in cephalograms by using a two-stage rectified point transform', *Computers in Biology and Medicine*, 65, pp. 114–123. doi: 10.1016/j.compbiomed.2015.07.022.

Tang, A., Tam, R., Cadrin-Chenevert, A., Guest, W., Chong, J., Barfett, J., Chepelev, L. Mitchell, J. R., Cicero, M. D., Poudrette, M. G., Jaremko, J. L., Reinhold, C., Gray, B. and Geis, R. (2018) 'Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology', *Canadian Association of Radiologists Journal*, 69, pp. 120–135. doi: 10.1016/j.carj.2018.02.002.

Tridandapani, S. (2018) 'Radiology "Hits Refresh" with Artificial Intelligence', *Academic Radiology*, 25(8), pp. 965–966. doi: 10.1016/j.acra.2018.02.006.

Varma, D. R. (2012) 'Managing DICOM images: Tips and tricks for the radiologist', *Indian Journal of Radiology and Imaging*, 22(1), pp. 4–13. doi: 10.4103/0971-3026.95396.

Vandenbroucke, J. P., Von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., Poole, C., Schlesselman, J. J. and Egger, M. (2007) 'Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration', *PLoS Medicine*, 4(10), pp. 1628–1654. doi: 10.1371/journal.pmed.0040297.

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2014) 'The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies', *International Journal of Surgery*, 12(12), pp. 1495–1499. doi: 10.1016/j.ijsu.2014.07.013.

Wang, C. W., Huang, C. T., Lee, J. H., Li, C. H., Chang, S. W., Siao, M. J, Lai, T. M., Ibragimov, B., Vrtovec, T., Ronneberger, O., Fischer, P., Cootes, T. F. and Lindner, C. (2016) 'A benchmark for comparison of dental radiography analysis algorithms', *Medical Image Analysis*, 31, pp. 63–76. doi: 10.1016/j.media.2016.02.004.

Wanjau, J., Khan, M. and Sethusa, M. (2019) ‘Applicability of the McNamara analysis in a sample of adult Black South Africans’, *South African Dental Journal*, 74(2). doi: 10.17159/2519-0105/2019/v74no2a6.

Weems, R. A. (2006). Radiographic Cephalometry Technique. In A. Jacobsen & R. L. Jacobsen (Eds.),*Radiographic Cephalometry: From Basics to 3D Imaging* (2nd ed., pp. 33–43). Quintessencces Publishing Co, Inc.

Wilson, J. D. and Hernandez-Hall, C. A. (2015) *Physics Laboratory Experiments*. 8th edn. Stamford: Cengage Learning.

Yaji, A., Prasad, S. and Pai, A. (2019) ‘Artificial Intelligence in Dento-Maxillofacial Radiology’, *Acta Scientific Dental Sciences*, 3(1), pp. 2581–4893.

Yu, H. J., Cho, S. R., Kim, M. J., Kim, W. H., Kim, J. W. and Choi, J. (2020) ‘Automated Skeletal Classification with Lateral Cephalometry Based on Artificial Intelligence’, *Journal of Dental Research*, 99(3), pp. 249–256. doi: 10.1177/0022034520901715.

CHAPTER 9 : APPENDICES

Appendix A: Memorandum of Understanding (Dolphin Imaging^{TM})

Dolphin Africa
Imaging & Management Solutions

40 4th Avenue
Newton Park
Port Elizabeth
Tel: (041) 364 0884/0885
Fax: (041) 364 0887
email:royston@dolphinimaging.co.za

Memorandum of Understanding

Between

Dolphin Imaging and Management Solutions (Dolphin Africa)

and

Suvarna Indermun – Post-graduate Student Department of Radiology, University of the Western Cape

This Memorandum of Understanding (MOU) sets for the terms and understanding between

Dolphin Imaging and Management Solutions (Dolphin Africa)

and **Suvarna Indermun – Post-graduate Student Department of Radiology, University of the Western Cape**

to use the software for research purposes.

Background
Dolphin Imaging and Management Solutions developed software over the years to improve and develop research in the field of dentistry as a whole and its main focus in particular orthodontics and radiology.

Purpose
This MOU will create the understanding between the parties that the partnership and collaboration is only for the purpose of research work.

The above goals will be accomplished by undertaking the following activities:

1. The software will be used for the purpose of research
2. The research will be made available to Dolphin Imaging and Management Solutions once completed to be added to their research archive
3. The research can be used for reference for any future or further research

Dolphin Africa
Imaging & Management Solutions

40 4th Avenue
Newton Park
Port Elizabeth
Tel: (041) 364 0884/0885
Fax: (041) 364 0887
email:royston@dolphinimaging.co.za

Duration
This MOU is at-will and may be modified by mutual consent of Dolphin Imaging and Management Solutions (Dolphin Africa) and Suvarna Indermun – Post-graduate Student Department of Radiology, University of the Western Cape

This MOU shall become effective upon signature by the authorized officials and will remain in effect until modified or terminated by any one of the partners by mutual consent. The MOU shall end once the research has been concluded and finalized.

Contact Information
Partner name: Dolphin Imaging and Management Solutions (Dolphin Africa)

Partner representative: Royston Johannes

Position: Director – Dolphin Africa
Address: 40 4th Avenue, Newton Park, Port Elizabeth,6055
Telephone: +27-041 3640884
Fax: (086) 562-2163
E-mail: royston@dolphinimaging.co.za

Partner name: Department of Radiology, University of the Western Cape
Partner representative: Suvarna Indermun
Position: Post-graduate Student
Address: Department of Radiology, University of the Western Cape, Tygerberg Hospital, Bellville

________________________ Study starting date – 03 April 2020

Signature
Royston Johannes

________________________ Study starting date – 03 April 2020
Signature
Suvarna Indermun

Appendix B: Table showing cephalometric landmarks and their description (Lindner *et al.*, 2016)

ID	Landmark name	Landmark description
L1	Sella	The geometric centre of the pituitary fossa (sella turcica), determined by inspection of a constructed point in the midsagittal plane.
L2	Nasion	The intersection of the internasal and frontonasal sutures, in the midsagittal plane.
L3	Orbitale	A point midway between the lowest points on the inferior margin of the two orbits (eye sockets).
L4	Porion	The central point on the upper margin of the external auditory meatus.
L5	Subspinale (A-point)	The deepest (most posterior) midline point on the curvature between the ANS and prosthion.
L6	Supramentale (B-point)	The deepest (most posterior) midline point on the bony curvature of the anterior mandible, between infradentale and pogonion.
L7	Pogonion	The most anterior point on the contour of the bony chin.
L8	Menton	The most inferior point of the mandibular symphysis.
L9	Gnathion	The most anterior inferior point on the bony chin.
L10	Gonion	The most posterior inferior point on the outline of the angle of the mandible.
L11	Incision inferius	The incisal tip of the most labially placed mandibular incisor.
L12	Incision superius	The incisal tip of the most labially placed maxillary central incisor.
L13	Upper lip	Labrale superior (Ls) The point denoting the vermilion border of the upper lip, in the midsagittal plane.
L14	Lower lip	Labrale inferior (Li) The point denoting the vermilion border of the lower lip, in the midsagittal plane.
L15	Subnasale	The point where the base of the columella of the nose meets the upper lip.
L16	Soft tissue pogonion	The most prominent point on the soft tissue contour of the chin.
L17	Posterior nasal spine	The most posterior point on the bony hard palate (nasal floor).
L18	Anterior nasal spine (ANS)	The tip of the bony anterior nasal spine at the inferior margin of the piriform aperture.
L19	Articulare	A constructed point representing the intersection of three radiographic images: the inferior surface of the cranial base and the posterior outlines of the ascending rami or mandibular condyles.

Appendix C: Table with Landmarks Key

Landmarks Key:	
Label:	**Landmark:**
L1	Sella
L2	Nasion
L3	Orbitale
L4	Porion
L5	Subspinale (Point A)
L6	Supramentale (Point B)
L7	Pogonion
L8	Menton
L9	Gnathion
L10	Gonion
L11	Incision inferius
L12	Incision superious
L13	Upper lip
L14	Lower Lip
L15	Subnasale
L16	Soft tissue pogonion
L17	Posterior Nasal spine
L18	Anterior Nasal Spine
L19	Articulare

Appendix D: Cephalometric Landmarks

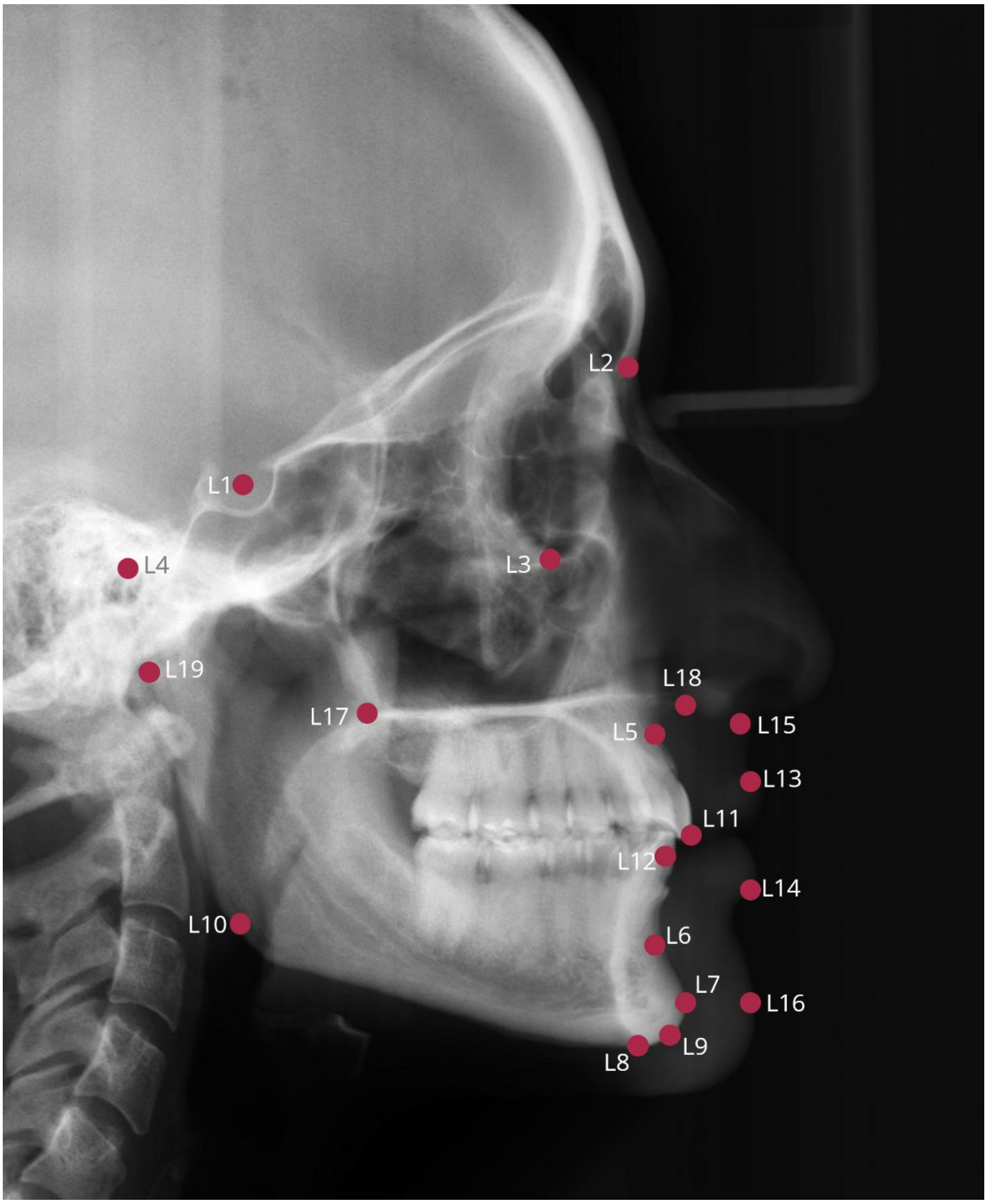


Figure 9.1: Cephalometric landmarks used in this study.

L1 = Sella, L2 = Nasion, L3 = Orbitale, L4 = Porion, L5 = Subspinale (Point A), L6 = Supramentale (Point B), L7 = Pogonion, L8 = Menton, L9 = Gnathion, L10 = Gonion, L11 = Incision inferius, L12 = Incision superious, L13 = Upper lip, L14 = Lower Lip, L15 = Subnasale, L16 = Soft tissue pogonion, L17 = Posterior Nasal spine, L18 = Anterior Nasal Spine, L19 = Articulare.

Appendix E: Instructions for Examiners

<u>INSTRUCTIONS FOR INTER-RELIABILTY TEST:</u>

EXAMINER 1: RESEARCHER, DR INDERMUN

EXAMINER 2: RADIOLOGIST, DR SHAIK

EXAMINER 3: ORTHODONTIST, DR JOHANNES

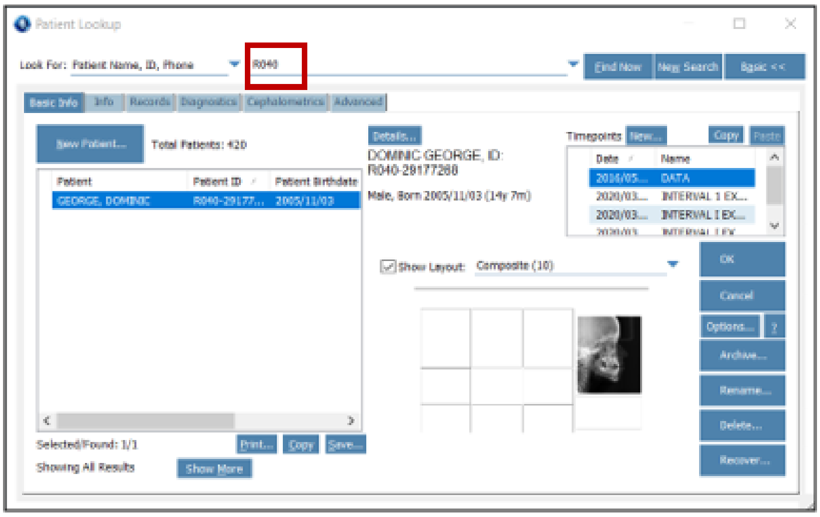
10 CEPHS SELECTED FOR INTER-RELIABILTY TESTS:

Ceph IDs:

1. R040
2. R080
3. R120
4. R160
5. R200
6. R240
7. R280
8. R320
9. R360
10. R400

<u>Step 1:</u>

Select patient according to ceph ID (e.g. R040)

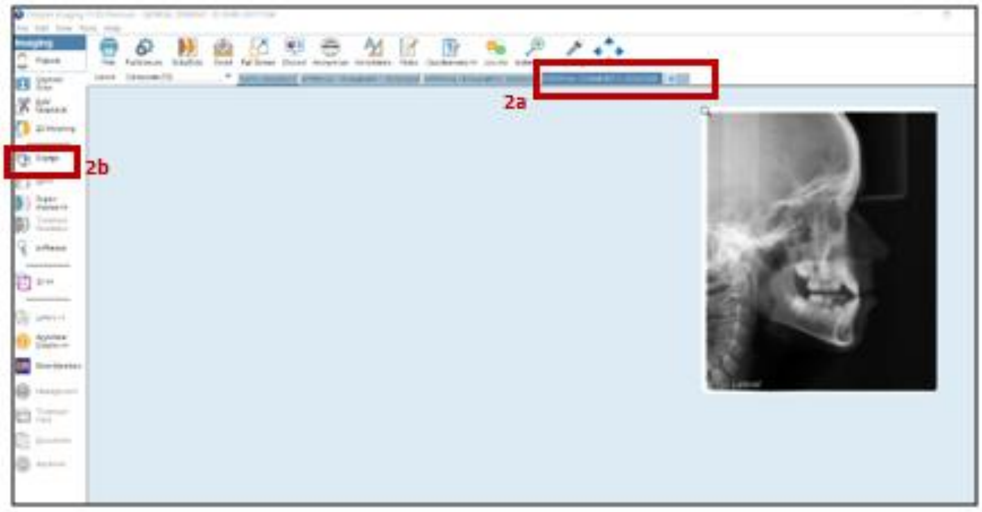


1

<u>Step 2:</u>

2a. Select tab according to interval and examiner (e.g. interval 1, examiner 3)

2b. Select Digitize tab



<u>Step 3:</u>

Use the following settings for the Digitize setup:

2

Step 4:

There are 19 landmarks that will be used in the study, as listed below:

However, a total of 22 points must be located. The extra points include ruler point 1 and ruler point 2. Condylion is will not be detected at its true anatomical position but was selected arbitrarily to be used as the centre of origin to determine the x, y co-ordinates.

Landmarks Key:		
No:	**Landmark:**	
1.	Sella	LANDMARKS FOR STUDY
2.	Nasion	
3.	Orbitale	
4.	Porion	
5.	Subspinale (Point A)	
6.	Supramentale (Point B)	
7.	Pogonion	
8.	Menton	
9.	Gnathion	
10.	Gonion	
11.	Incision inferius	
12.	Incision superious	
13.	Upper lip	
14.	Lower Lip	
15.	Subnasale	
16.	Soft tissue pogonion	
17.	Posterior Nasal spine	
18.	Anterior Nasal Spine	
19.	Articulare	
20.	**Ruler point 1**	**ADDITIONAL POINTS**
21.	**Ruler point 2**	
22.	**Condylion (point of origin)**	

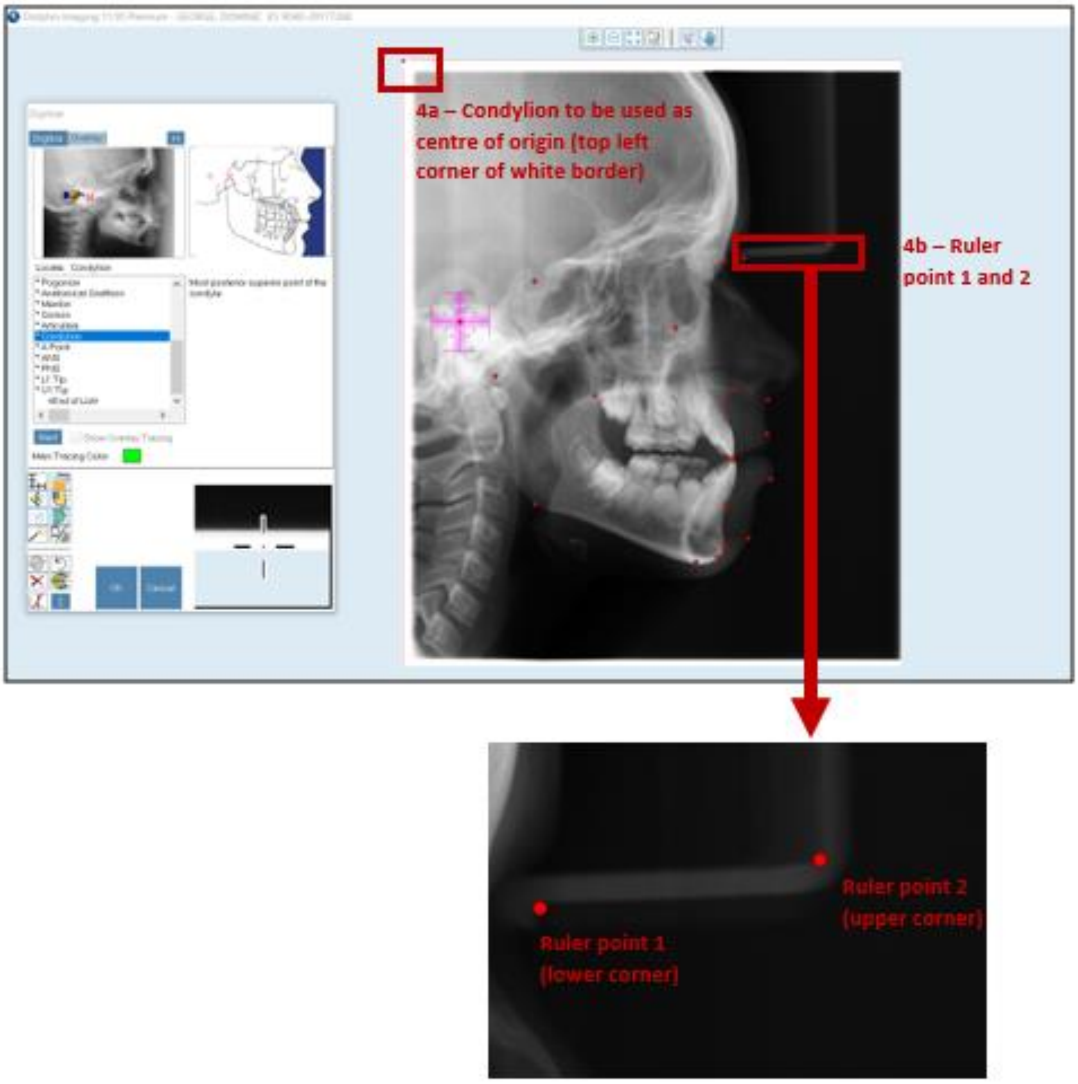
3

Table showing cephalometric landmarks and their description (Lindner et al., 2016)

ID	Landmark name	Landmark description
L1	Sella	The geometric centre of the pituitary fossa (sella turcica), determined by inspection of a constructed point in the midsagittal plane.
L2	Nasion	The intersection of the internasal and frontonasal sutures, in the midsagittal plane.
L3	Orbitale	A point midway between the lowest points on the inferior margin of the two orbits (eye sockets).
L4	Porion	The central point on the upper margin of the external auditory meatus.
L5	Subspinale (A-point)	The deepest (most posterior) midline point on the curvature between the ANS and prosthion.
L6	Supramentale (B-point)	The deepest (most posterior) midline point on the bony curvature of the anterior mandible, between infradentale and pogonion.
L7	Pogonion	The most anterior point on the contour of the bony chin.
L8	Menton	The most inferior point of the mandibular symphysis.
L9	Gnathion	The most anterior inferior point on the bony chin.
L10	Gonion	The most posterior inferior point on the outline of the angle of the mandible.
L11	Incision inferius	The incisal tip of the most labially placed mandibular incisor.
L12	Incision superius	The incisal tip of the most labially placed maxillary central incisor.
L13	Upper lip	Labrale superior (Ls) The point denoting the vermilion border of the upper lip, in the midsagittal plane.
L14	Lower lip	Labrale inferior (Li) The point denoting the vermilion border of the lower lip, in the midsagittal plane.
L15	Subnasale	The point where the base of the columella of the nose meets the upper lip.
L16	Soft tissue pogonion	The most prominent point on the soft tissue contour of the chin.
L17	Posterior nasal spine	The most posterior point on the bony hard palate (nasal floor).
L18	Anterior nasal spine (ANS)	The tip of the bony anterior nasal spine at the inferior margin of the piriform aperture.
L19	Articulare	A constructed point representing the intersection of three radiographic images: the inferior surface of the cranial base and the posterior outlines of the ascending rami or mandibular condyles.

4

Locate each landmark. Locations for Ruler point 1, Ruler point 2, and condylion are indicated in the screenshot below.

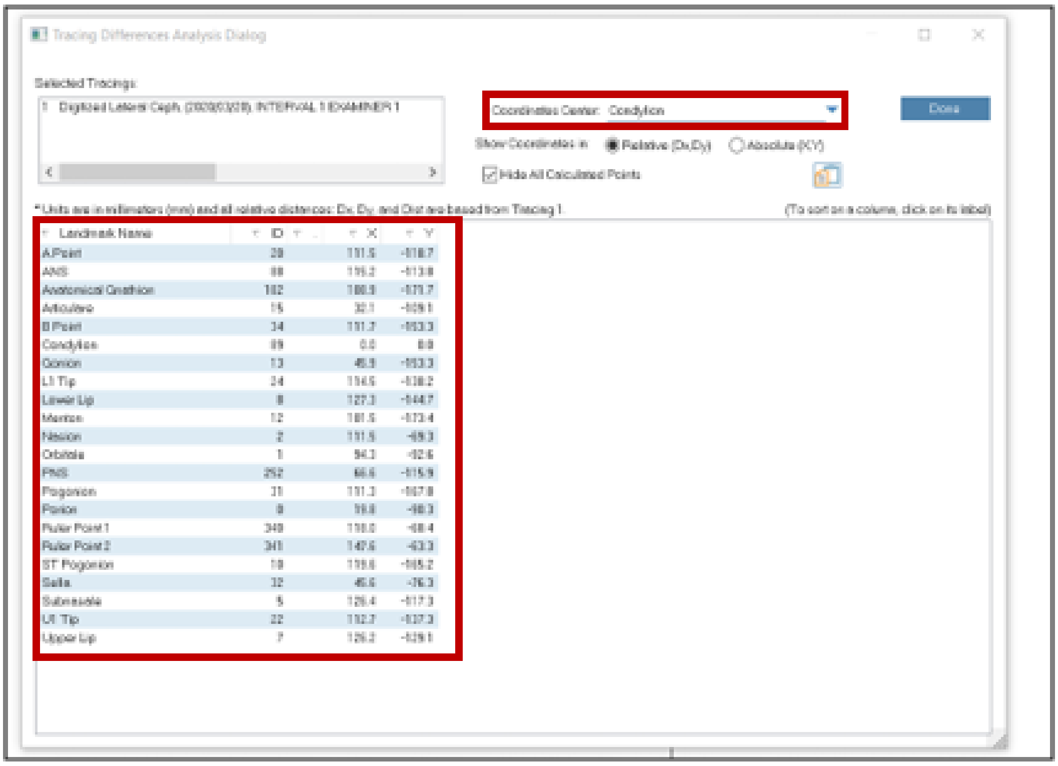


5

Step 5:

Once all landmarks are detected, click ok and continue to next ceph record (R080, R120, R160.... R400) – THEN YOU'RE DONE! 😊 This process will be repeated after 2 weeks (i.e. interval 2)

NOTE: THE RESEARCHER WILL COPY AND PASTE THE X&Y COORDINATES TO AN EXCEL SHEET



6

Appendix F: Demonstration of Landmark Detection using Dolphin Imaging™

The following example is a demonstration of how the coordinates were captured from Dolphin Imaging™ (Figures 9.2-9.14).

1. Dolphin Imaging™: Capture of Cephalograms

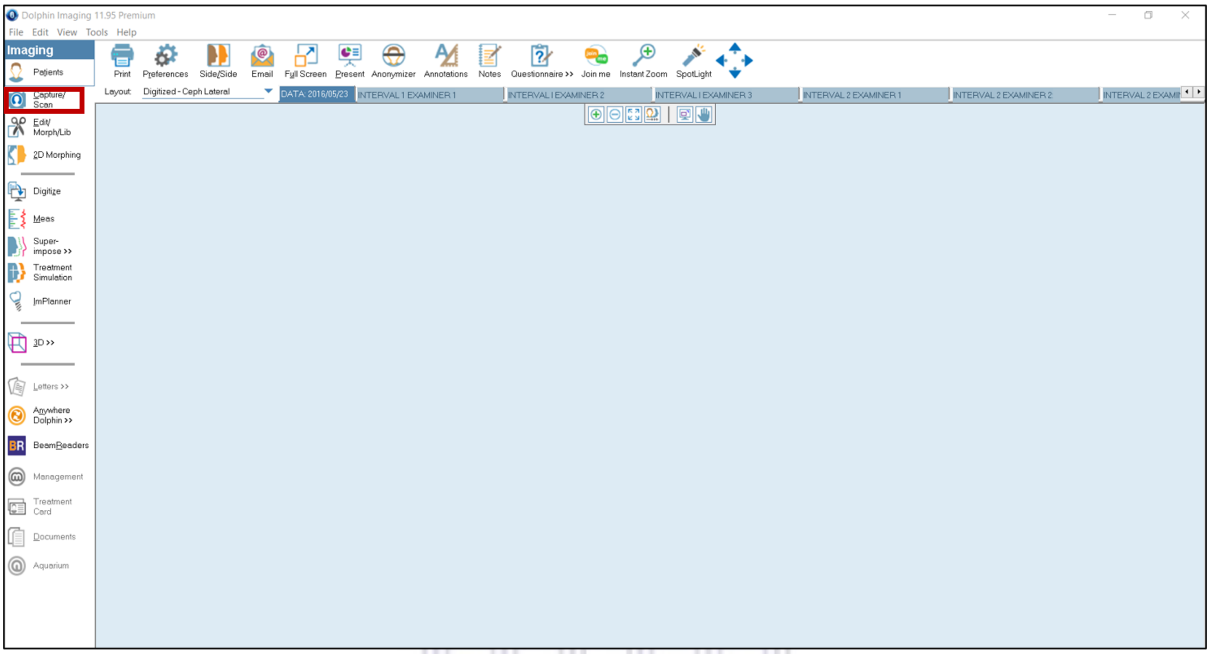


Figure 9.2: Step 1

Tthe primary researcher captured the entire sample of cephalograms prior to landmark detection. This was done by clicking capture (red box).

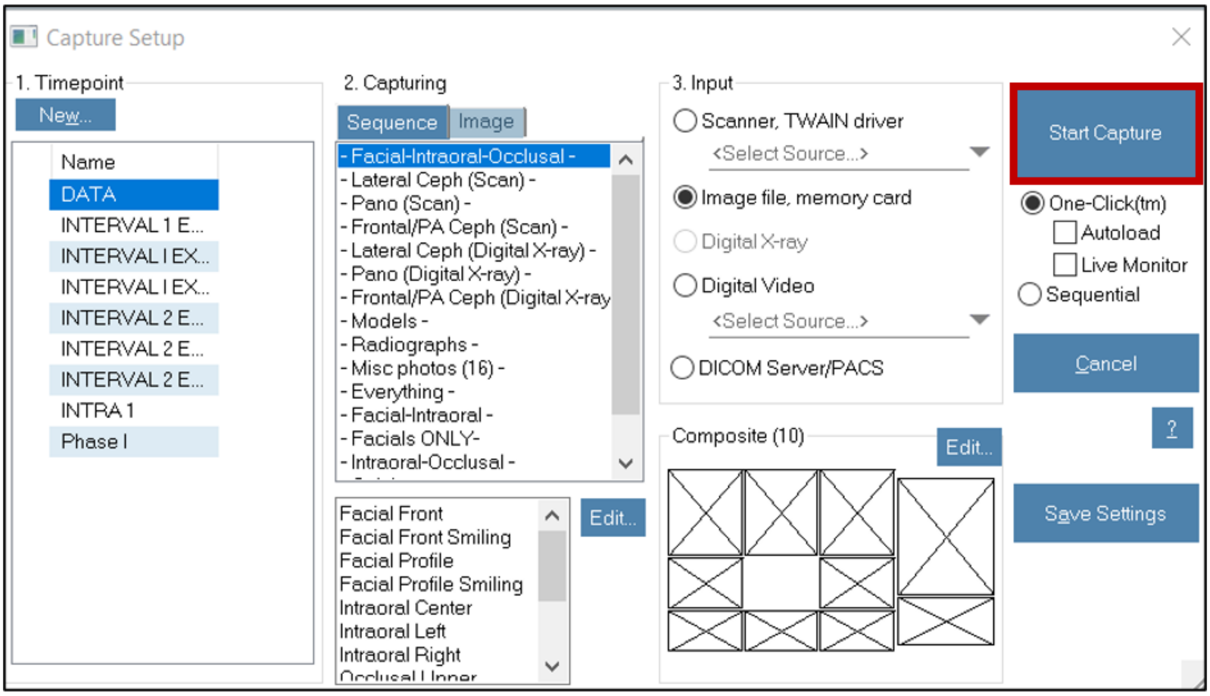


Figure 9.3: Step 2

The Capture setup was enabled by first selecting the timepoint, the images to be captured, and the input. The input was the device storage on the primary researcher's desktop. "Start capture" (red box) was selected to complete the setup.

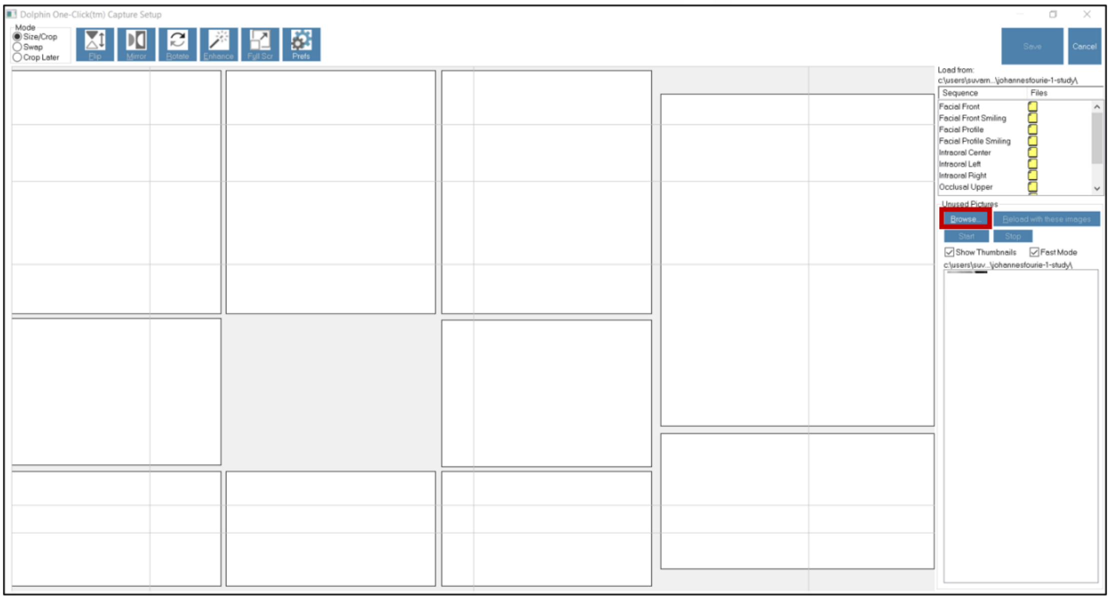


Figure 9.4: Step 3

The DICOM file was selected by browsing through the folder for the intended record number.

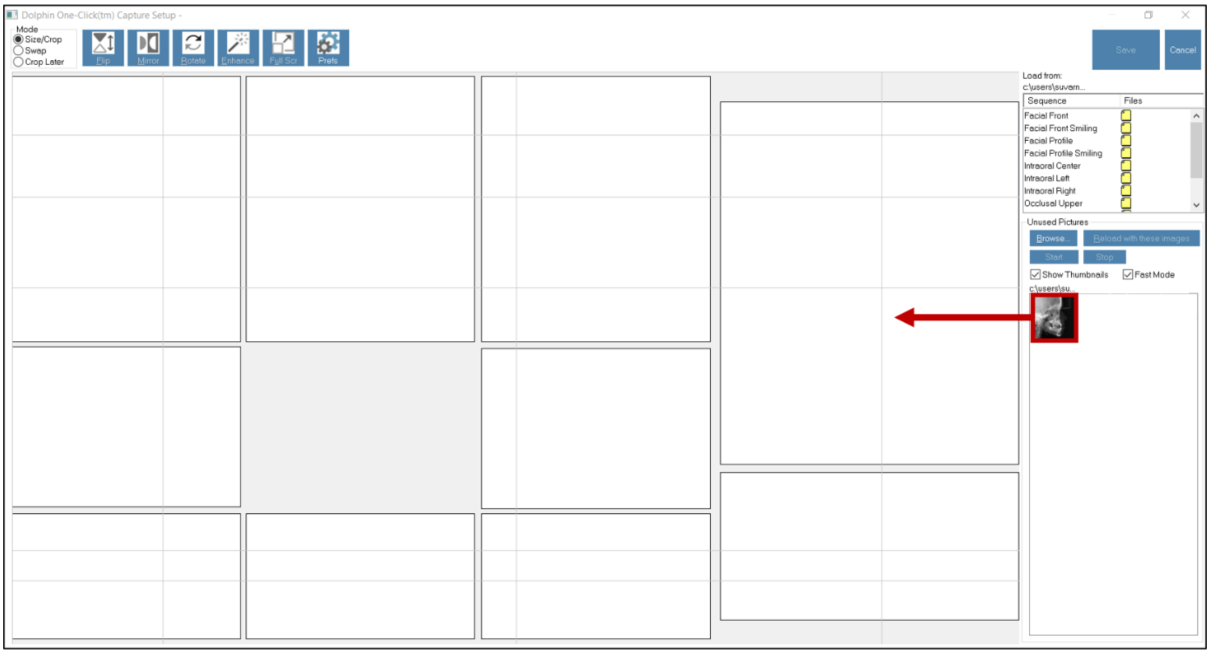


Figure 9.5: Step 4

Once the intended cephalogram was displaced in the display box, it was dragged using the mouse-cursor to the lateral cephalogram box.

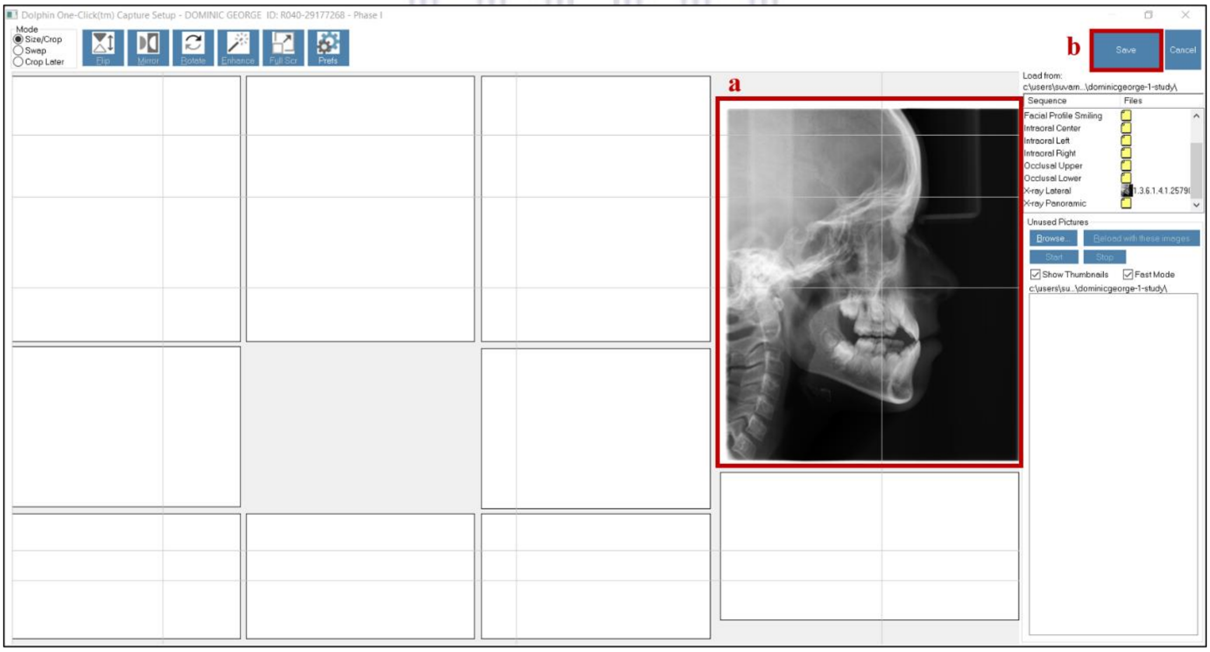


Figure 9.6: Step 5

Once the cephalogram was displayed in the box (a), the image capture was saved (b).

2. Dolphin Imaging™: Landmark Detection

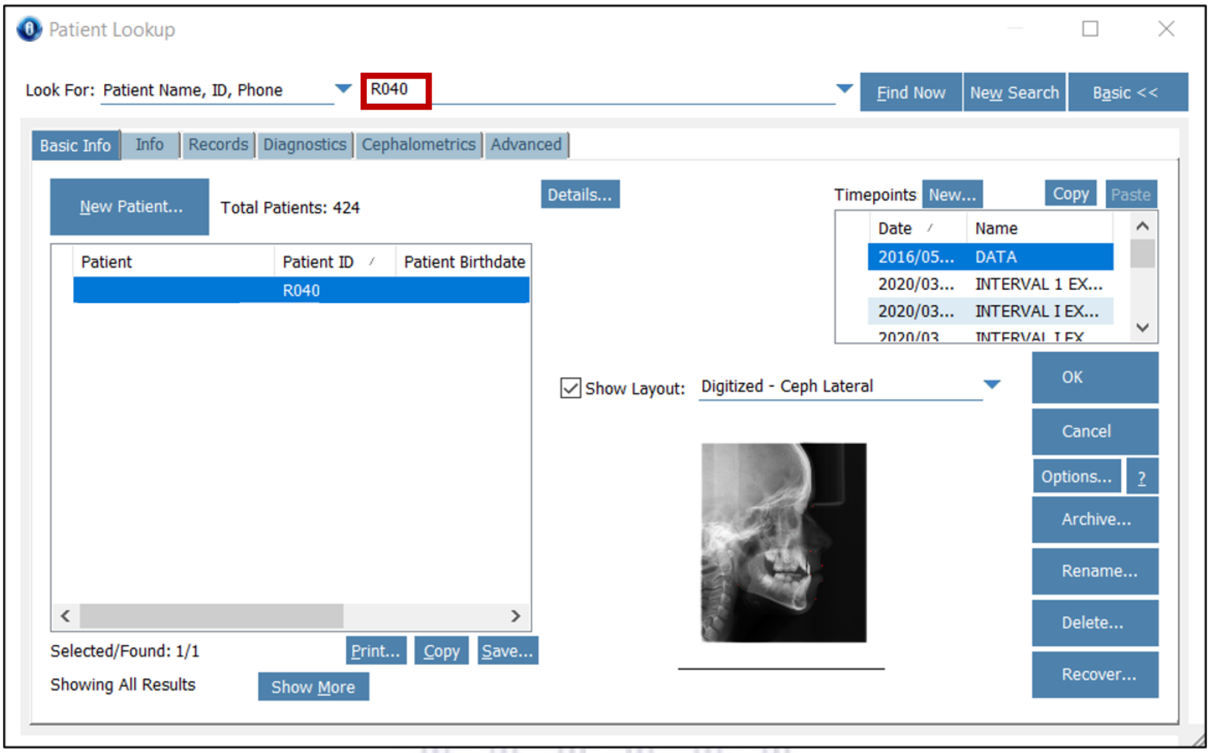


Figure 9.7: Step 6

Patients were selected via the record number, example R040 (red block).

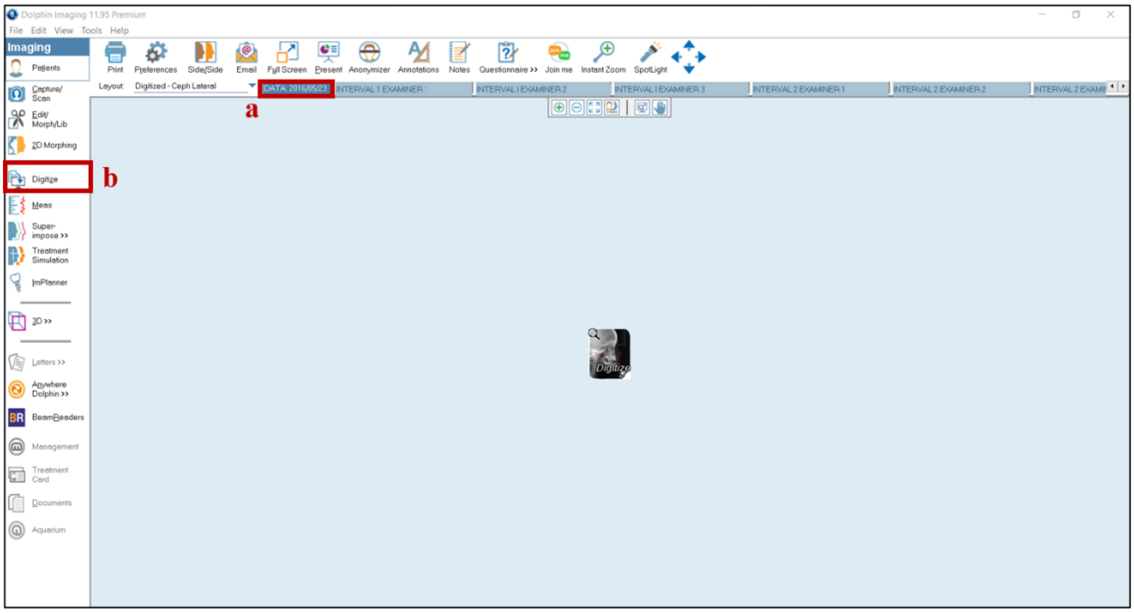


Figure 9.8: Step 7

Timepoints served as tabs to organize data collection; for example: primary data (a). The 'Digitize' tab was selected to open the setup page

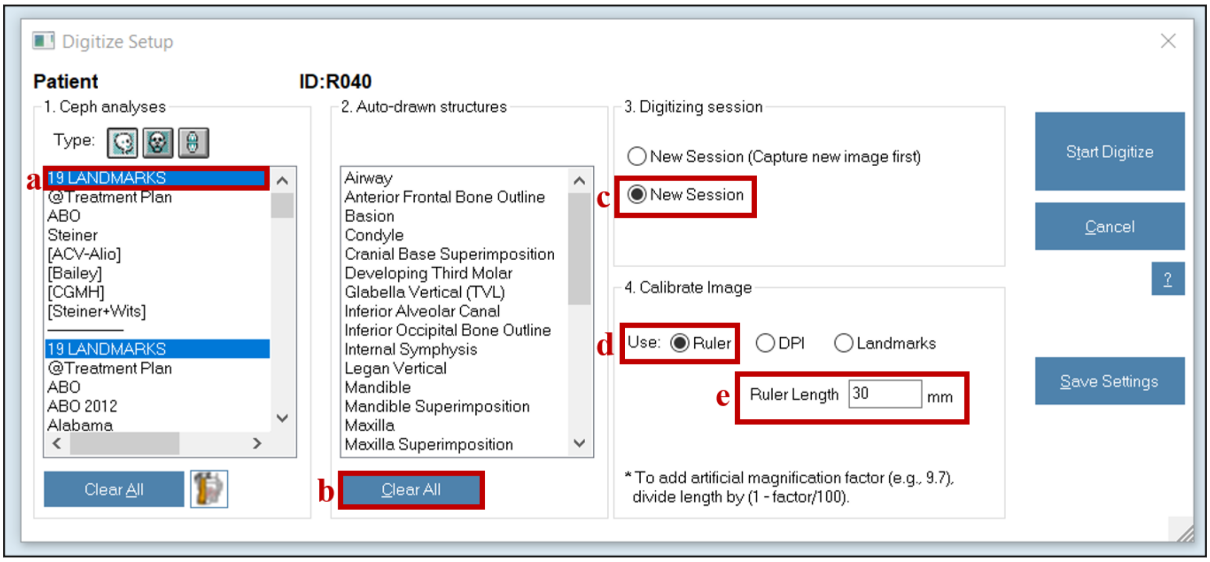


Figure 9.9: Step 8

The settings (a-e) were selected to adhere to the customized landmarks. A customized ceph analysis (named “19 Landmarks”) was created by the primary researcher to include the study’s intended landmarks. The ruler length was set at 30mm, to represent the real distance length of the fixed corner points of the nasion-guiding rod. This was done as there was no ruler used during acquisition of the cephalograms.

Table 9.1: Landmarks that were created in the custom list. Additional points were added for calibration (ruler points 1 and 2). Condylion was not identified at its true anatomical position but was selected arbitrarily to be used as the centre of origin to determine the x, y coordinates of the other landmarks.

Landmarks Key:		
No:	**Landmark:**	
1.	Sella	**LANDMARKS FOR STUDY**
2.	Nasion	
3.	Orbitale	
4.	Porion	
5.	Subspinale (Point A)	
6.	Supramentale (Point B)	
7.	Pogonion	
8.	Menton	
9.	Gnathion	
10.	Gonion	
11.	Incision inferius	
12.	Incision superious	
13.	Upper lip	
14.	Lower Lip	
15.	Subnasale	
16.	Soft tissue pogonion	
17.	Posterior Nasal spine	
18.	Anterior Nasal Spine	
19.	Articulare	
20.	**Ruler point 1**	**ADDITIONAL POINTS**
21.	**Ruler point 2**	
22.	**Condylion (centre of origin)**	

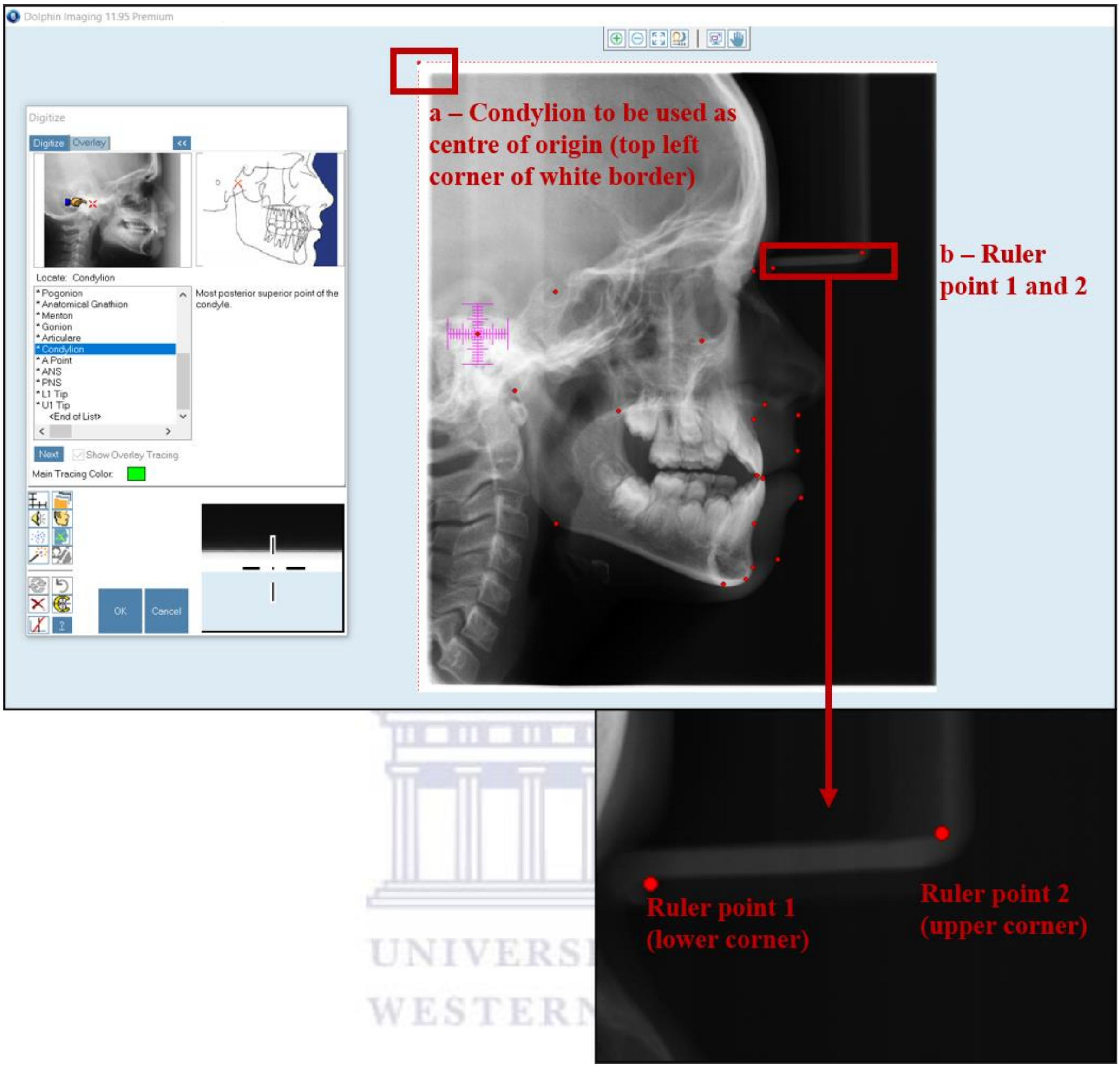


Figure 9.10: Step 9

Identification of landmarks. After recording a landmark with the mouse, a red dot on the monitor-displayed image indicated its position. The landmark location could be corrected until the operator was satisfied. The magnifying glass was also used to locate landmarks. The study's definitions of landmarks were used and not those that automatically appear in Dolphin (a) Locations for condylion, (b) Locations for ruler point 1 and 2 – this was used for calibration by digitizing the two ruler points (30mm).

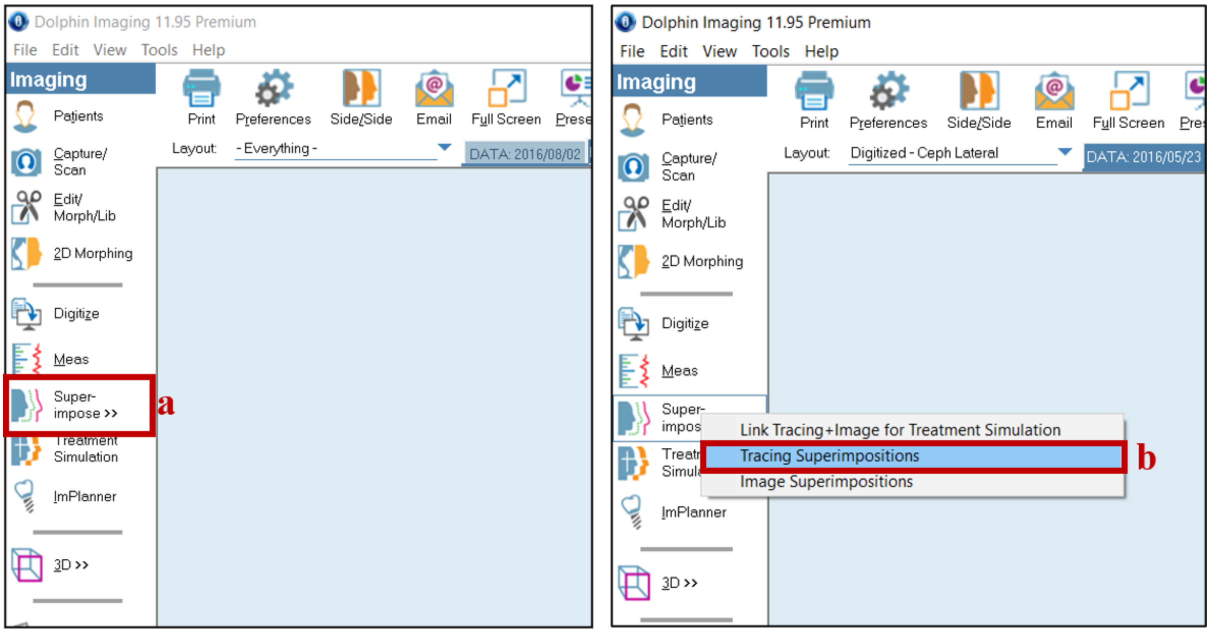


Figure 9.11: Step 10

The "Superimpose" tab was selected (a) to open the tracing superimpositions setup page (b)

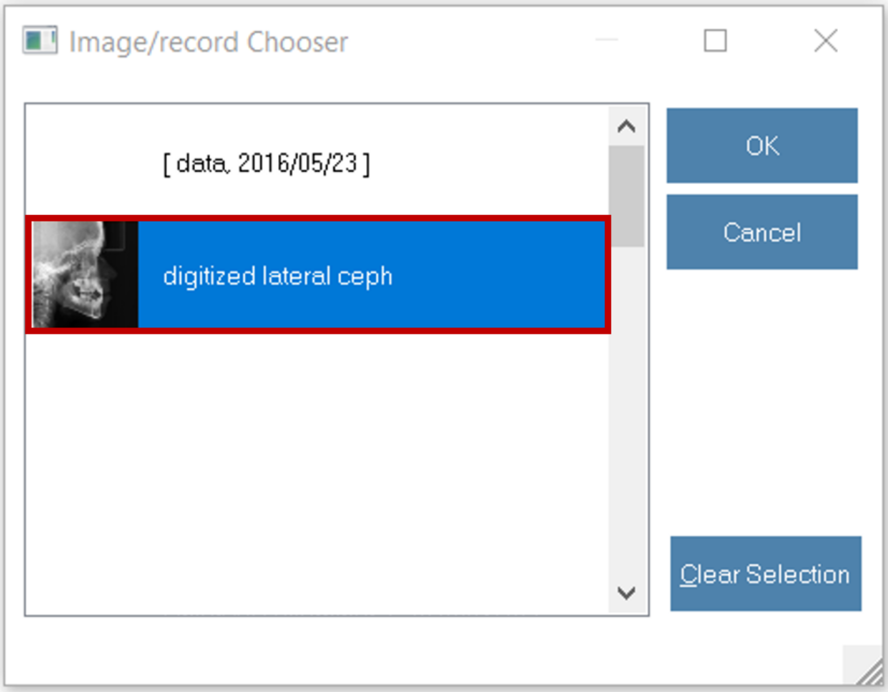


Figure 9.12: Step 11

The intended ceph was selected.

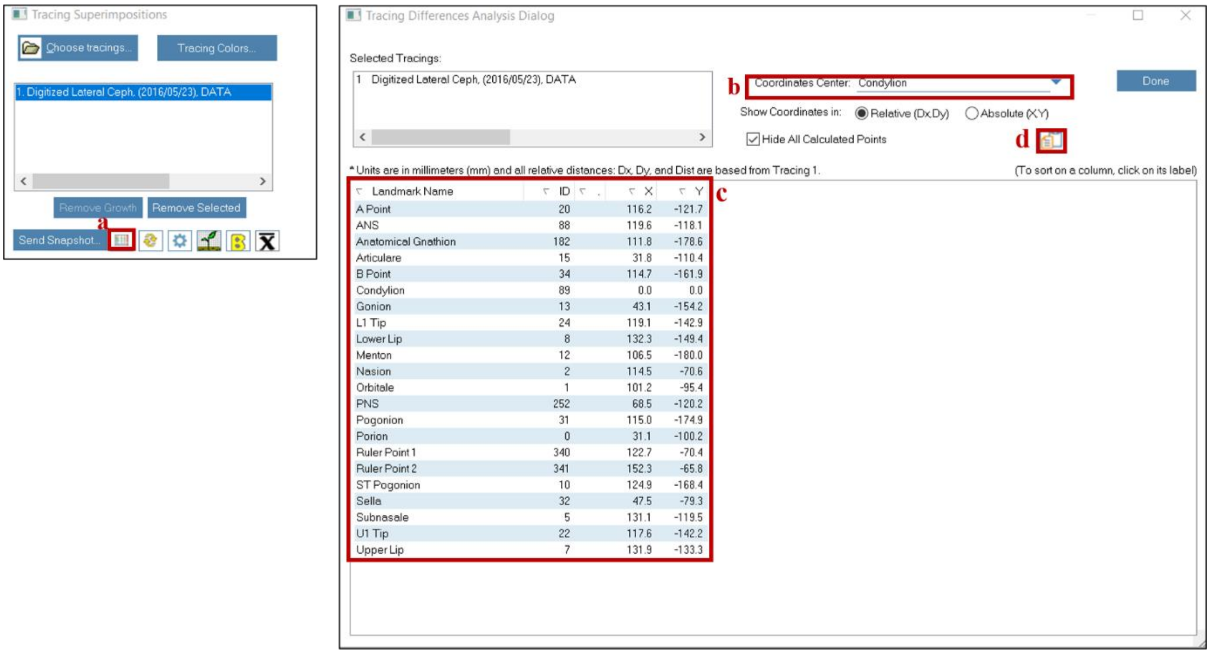


Figure 9.13: Step 12

The icon  was clicked (a) to view the tracing differences analysis dialog. Condylion was selected as the coordinates centre (b). The generated coordinates (c) was exported to the Excel Data Capture sheet (d).

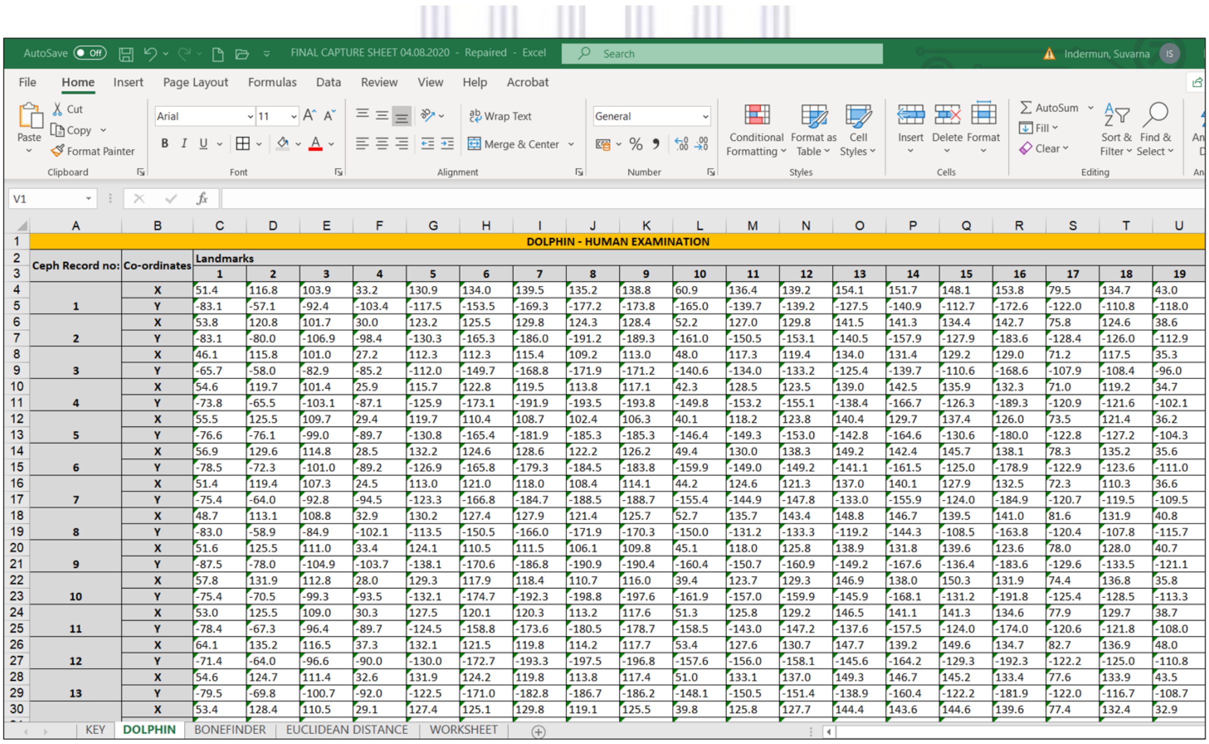


Figure 9.14: Step 13

The generate coordinates were copied to the Excel Capture Sheet.

Appendix G: Demonstration of Landmark Detection using BoneFinder®

The following example is a demonstration of how the coordinates were captured from BoneFinder® (Figures 9.15 – 9.24).

1. BoneFinder® - Landmark Detection

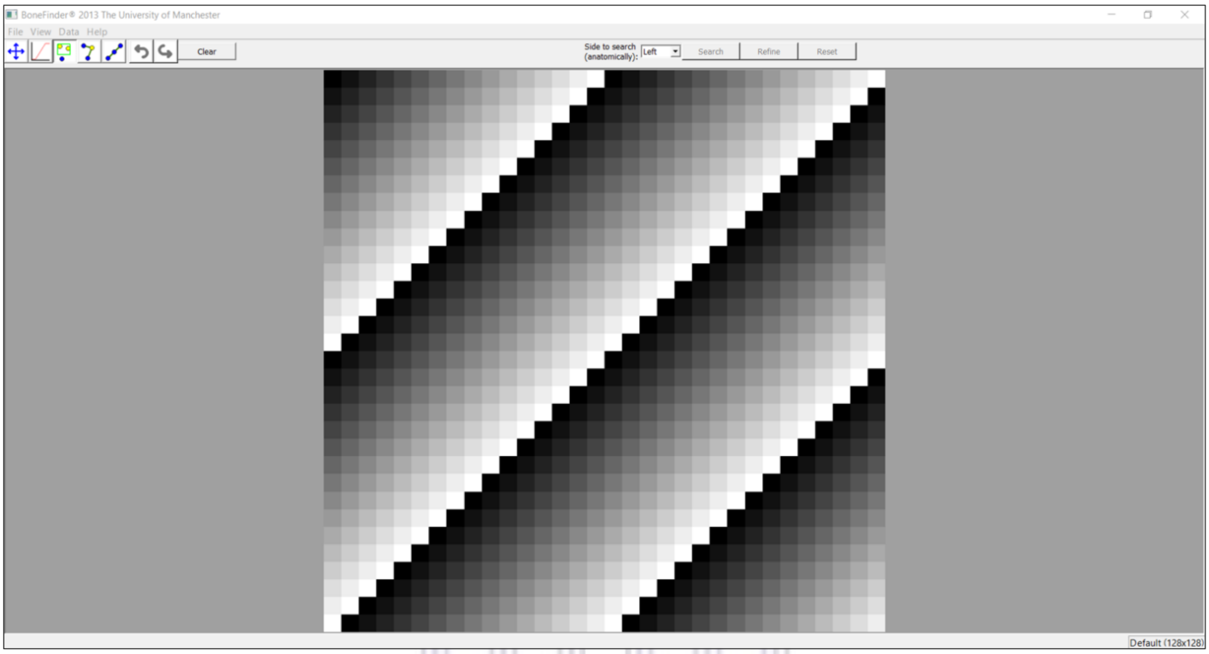


Figure 9.15: Step 1

Open BoneFinder®.

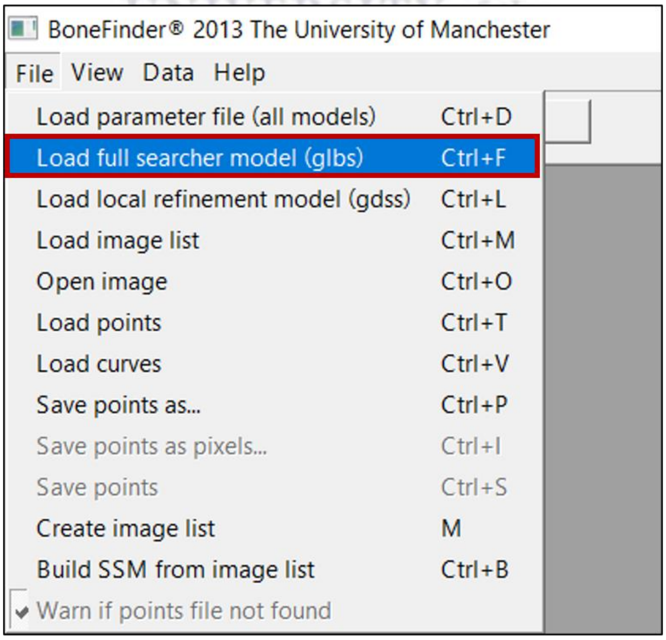


Figure 9.16: Step 2

On start of BoneFinder® a number of models need to be loaded. The “load full searcher model” was selected.

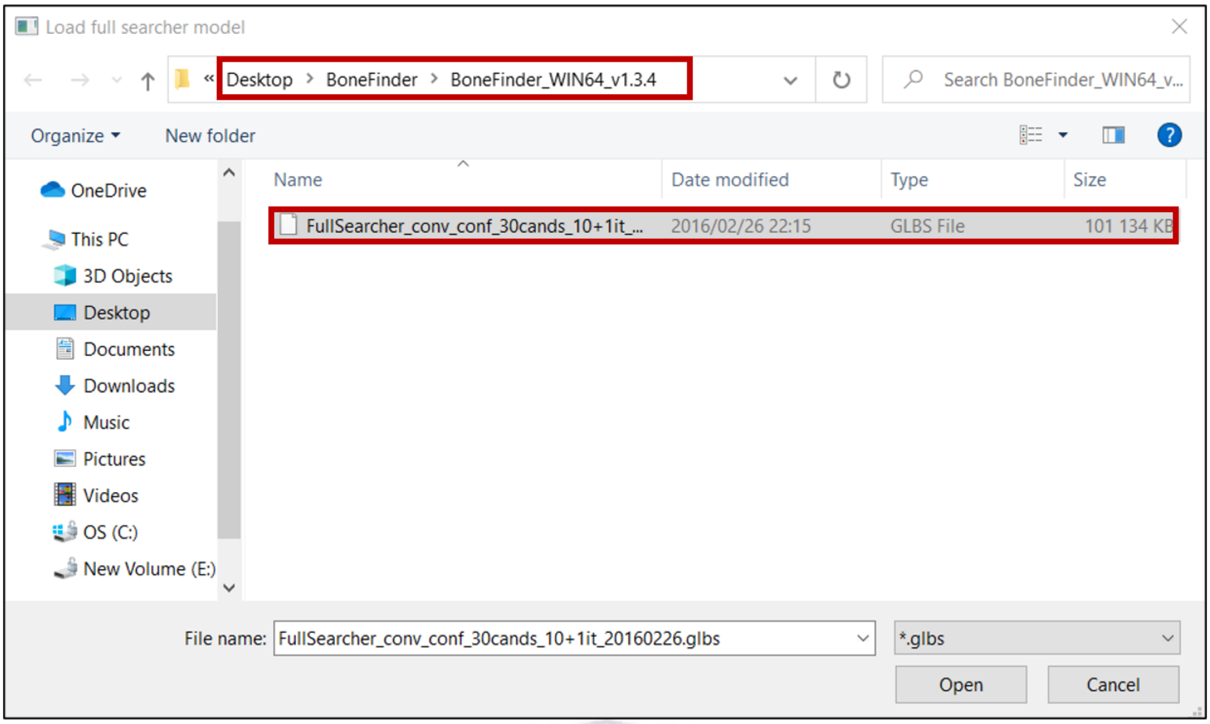


Figure 9.17: Step 3

The Full Searcher GLBS file was opened.

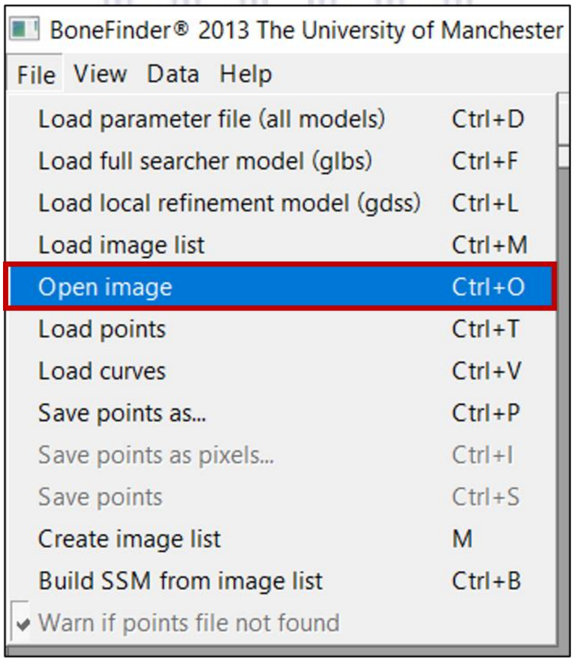


Figure 9.18: Step 4

The next step is to load an image using File/Open image.

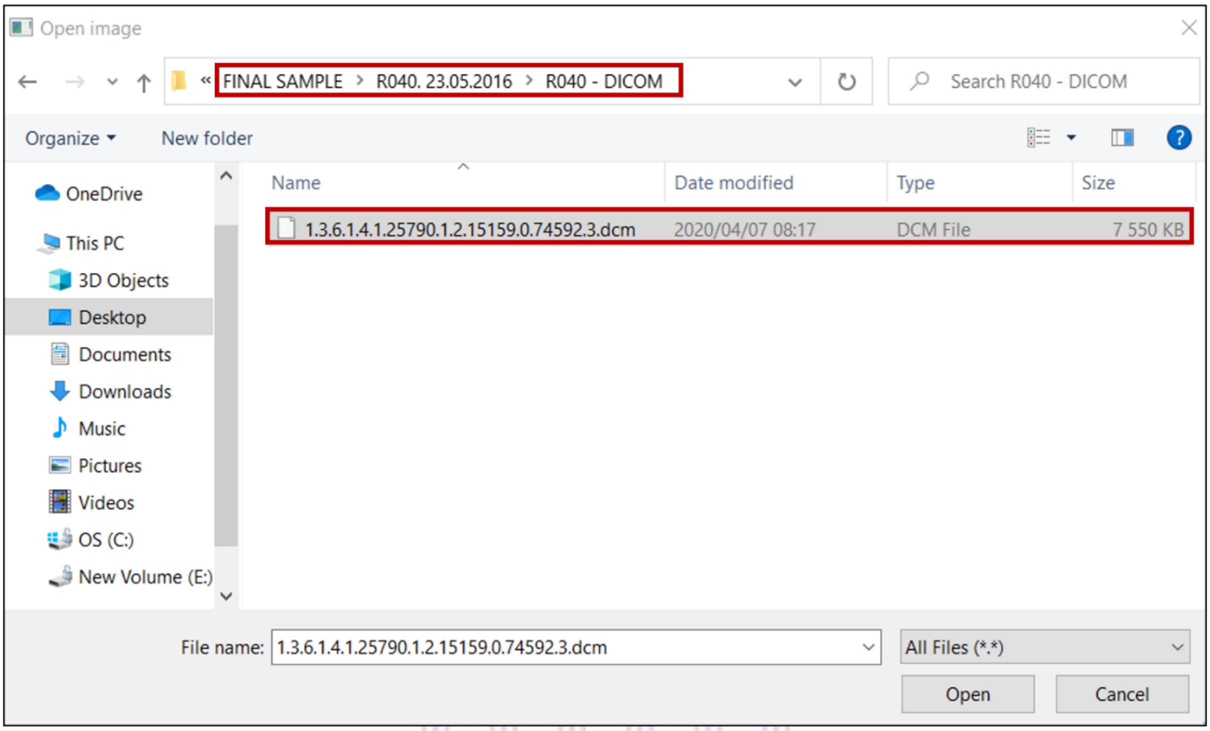


Figure 9.19: Step 5

The folder containing the DICOM files was opened, and the required cephalogram opened via the record number e.g., R040.

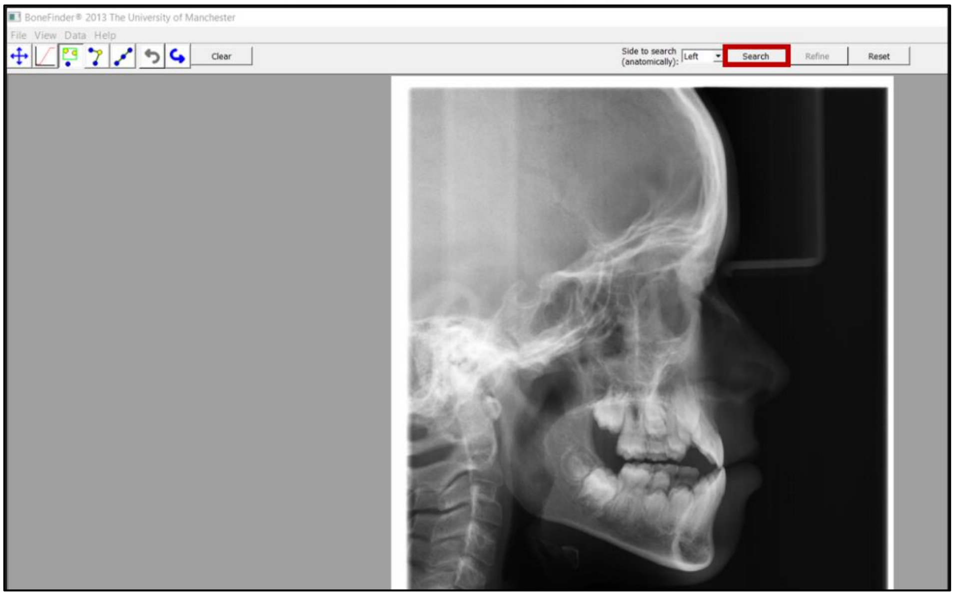


Figure 9.20: Step 6

The fully automatic search can be run on the displayed image using the search button. After a few seconds, the tool will display the annotation result

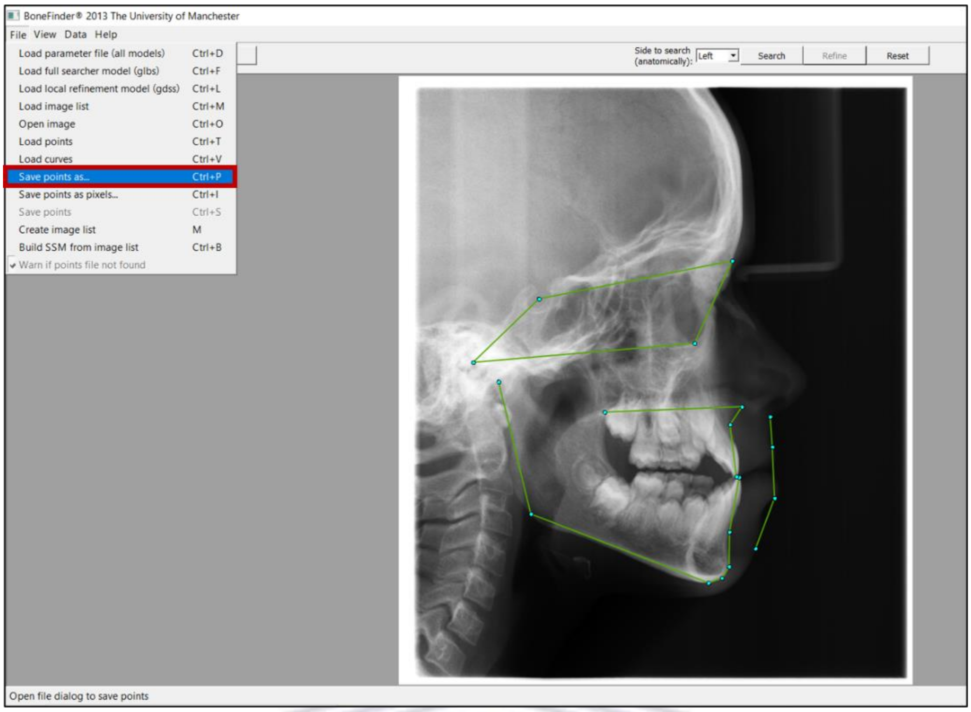


Figure 9.21: Step 7

All landmarks are indicated by light blue markers. The coordinates were saved to a text file using *File/Save points as*. Note that by default every loaded image is histogram equalised to improve the contrast of the image for visualisation and search. This was turned off using "*Data/Always normalise image on opening*" for the purpose of this demonstration.

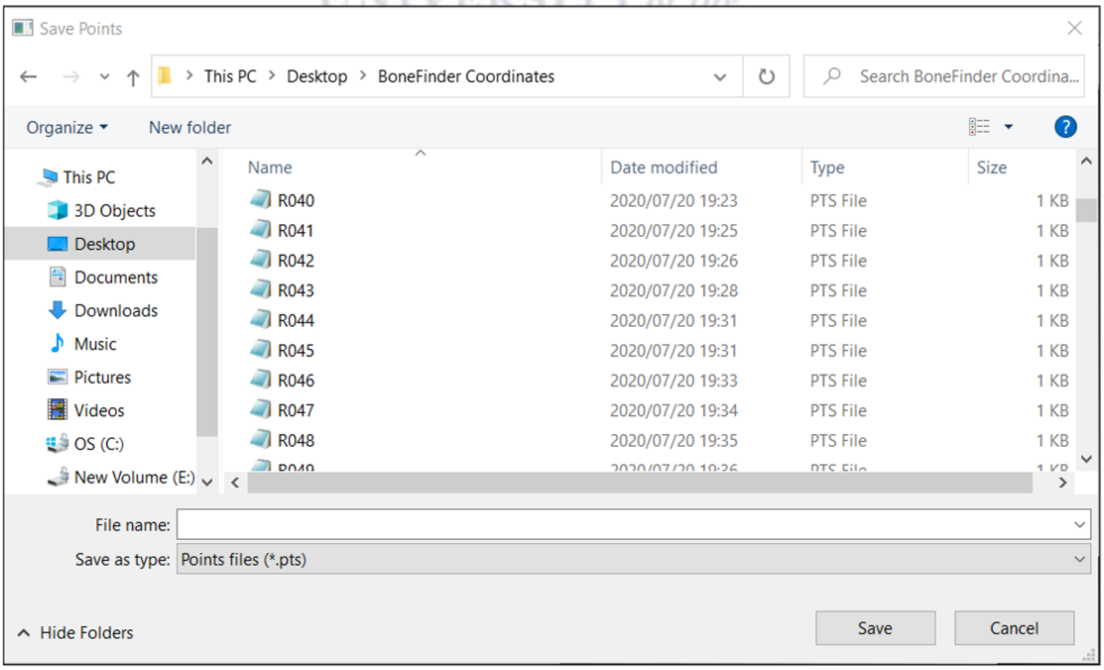
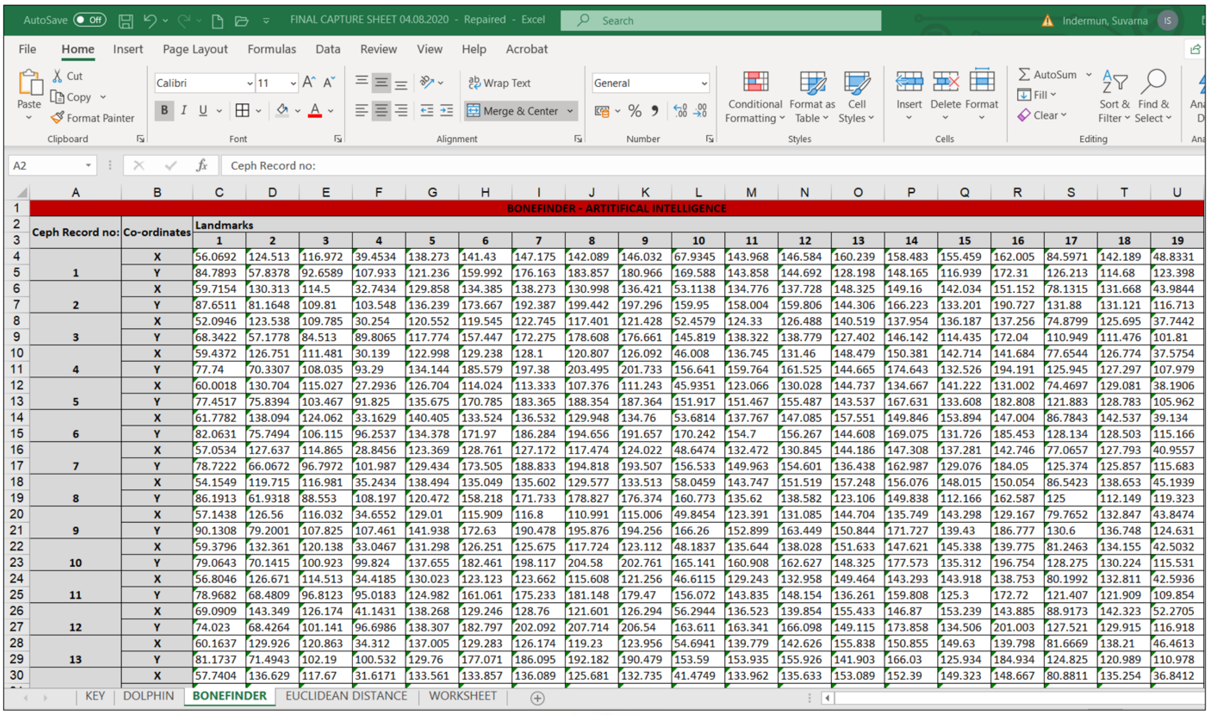


Figure 9.22: Step 8

Coordinates were saved as Points files and exported to the Excel Capture Sheet



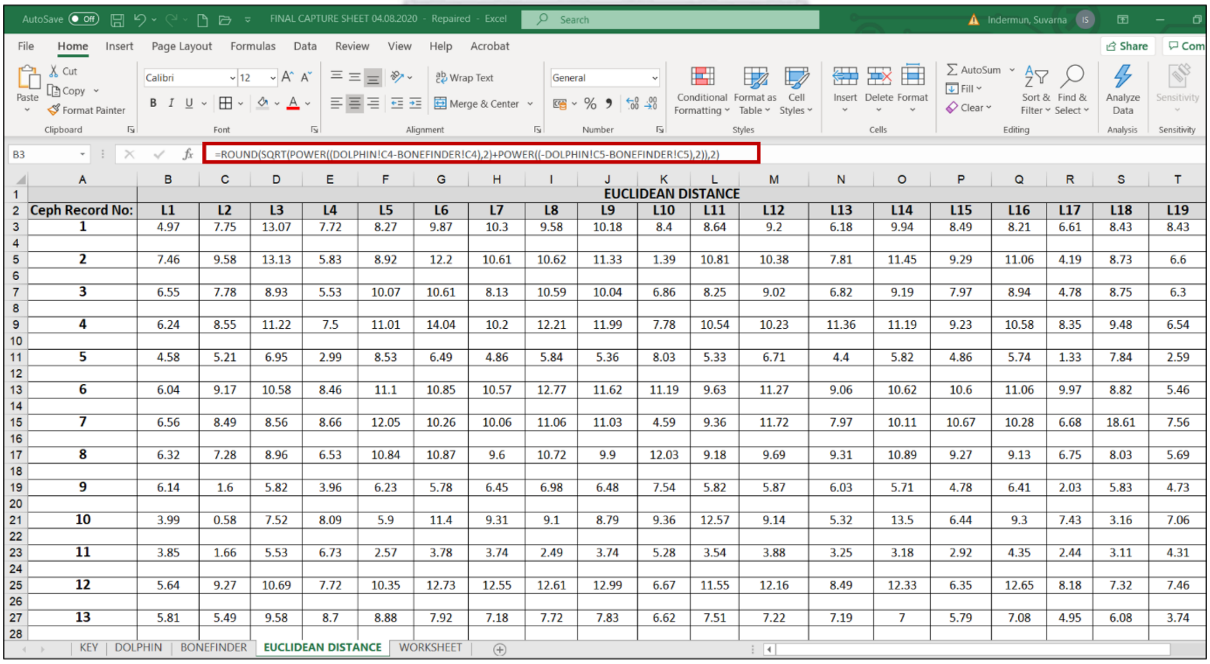
BONEFINDER - ARTITIFICAL INTELLIGENCE

Ceph Record no:	Co-ordinates	Landmarks 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	X	56.0692	124.513	116.972	39.4534	138.273	141.43	147.175	142.089	146.032	67.9345	143.968	146.584	160.239	158.483	155.459	162.005	84.5971	142.189	48.8331
1	Y	84.7893	57.8378	92.6589	107.933	121.236	159.992	176.163	183.857	180.966	169.588	143.858	144.692	128.198	148.165	116.939	172.31	126.213	114.68	123.398
	X	59.7154	130.313	114.5	32.7434	129.858	134.385	138.273	130.998	136.421	53.1138	134.776	137.728	148.325	149.16	142.034	151.152	78.1315	131.668	43.9844
2	Y	87.6511	81.1648	109.81	103.548	136.239	173.667	192.387	199.442	197.296	159.95	158.004	159.806	144.306	166.223	133.201	190.727	131.88	131.121	116.713
	X	52.0946	123.538	109.785	30.254	120.552	119.545	122.745	117.401	121.428	52.4579	124.33	126.488	140.519	137.954	136.187	137.256	74.8799	125.695	37.7442
3	Y	68.3422	57.1778	84.513	89.8065	117.774	157.447	172.275	178.608	176.661	145.819	138.322	138.779	127.402	146.142	114.435	172.04	110.949	111.476	101.81
	X	59.4372	126.751	111.481	30.139	122.998	129.238	128.1	120.807	126.092	46.008	136.745	131.46	148.479	150.381	142.714	141.684	77.6544	126.774	37.5754
4	Y	77.74	70.3307	108.035	93.29	134.144	185.579	197.38	203.495	201.733	156.641	159.764	161.525	144.665	174.643	132.526	194.191	125.945	127.297	107.979
	X	60.0018	130.704	115.027	27.2936	126.704	114.024	113.333	107.376	111.243	45.9351	123.066	130.028	144.737	134.667	141.222	131.002	74.4697	129.081	38.1906
5	Y	77.4517	75.8394	103.467	91.825	135.675	170.785	183.365	188.354	187.364	151.917	151.467	155.487	143.537	167.631	133.608	182.808	121.883	128.783	105.962
	X	61.7782	138.094	124.062	33.1629	140.405	133.524	136.532	129.948	134.76	53.6814	137.767	147.085	157.551	149.846	153.894	147.004	86.7843	142.537	39.134
6	Y	82.0631	75.7494	106.115	96.2537	134.378	171.97	186.284	194.656	191.657	170.242	154.7	156.267	144.608	169.075	131.726	185.453	128.134	128.503	115.166
	X	57.0534	127.637	114.865	28.8456	123.369	128.761	127.172	117.474	124.022	48.6474	132.472	130.845	144.186	147.308	137.281	142.746	77.0657	127.793	40.9557
7	Y	78.7222	66.0672	96.7972	101.987	129.434	173.505	188.833	194.818	193.507	156.533	149.963	154.601	136.438	162.987	129.076	184.05	125.374	125.857	115.683
	X	54.1549	119.715	116.981	35.2434	138.494	135.049	135.602	129.577	133.513	58.0459	143.747	151.519	157.248	156.076	148.015	150.054	86.5423	138.653	45.1939
8	Y	86.1913	61.9318	88.553	108.197	120.472	158.218	171.733	178.827	176.374	160.773	135.62	138.582	123.106	149.838	112.166	162.587	125	112.149	119.323
	X	57.1438	126.56	116.032	34.6552	129.01	115.909	116.8	110.991	115.006	49.8454	123.391	131.085	144.704	135.749	143.298	129.167	79.7652	132.847	43.8474
9	Y	90.1308	79.2001	107.825	107.461	141.938	172.63	190.478	195.876	194.256	166.26	152.899	163.449	150.844	171.727	139.43	186.777	130.6	136.748	124.631
	X	59.3796	132.361	120.138	33.0467	131.298	126.251	125.675	117.724	123.112	48.1837	135.644	138.028	151.633	147.621	145.338	139.775	81.2463	134.155	42.5032
10	Y	79.0643	70.1415	100.923	99.824	137.655	182.461	198.117	204.58	202.761	165.141	160.908	162.627	148.325	177.573	135.312	196.754	128.275	130.224	115.531
	X	56.8046	126.671	114.513	34.4185	130.023	123.123	123.662	115.608	121.256	46.6115	129.243	132.958	149.464	143.293	143.918	138.753	80.1992	132.811	42.5936
11	Y	78.9682	68.4809	96.8123	95.0183	124.982	161.061	175.233	181.148	179.47	156.072	143.835	148.154	136.261	159.808	125.3	172.72	121.407	121.909	109.854
	X	69.0909	143.349	126.174	41.1431	138.268	129.246	128.76	121.601	126.294	56.2944	136.523	139.854	155.433	146.87	153.239	143.885	88.9173	142.323	52.2705
12	Y	74.023	68.4264	101.141	96.6986	138.307	182.797	202.092	207.714	206.54	163.611	163.341	166.098	149.115	173.858	134.506	201.003	127.521	129.915	116.918
	X	60.1637	129.926	120.863	34.312	137.005	129.283	126.174	119.23	123.956	54.6941	139.779	142.626	155.838	150.855	149.63	139.798	81.6669	138.21	46.4613
13	Y	81.1737	71.4943	102.19	100.532	129.76	177.071	186.095	192.182	190.479	153.59	153.935	155.926	141.903	166.03	125.934	184.934	124.825	120.989	110.978
	X	57.7404	136.629	117.67	31.6171	133.561	133.857	136.089	125.681	132.735	41.4749	133.962	135.633	153.089	152.39	149.323	148.667	80.8811	135.254	36.8412

Figure 9.23: Step 9

The generated coordinates were copied to the Excel Capture Sheet.

Calculate the Euclidean Distance



=ROUND(SQRT(POWER((DOLPHIN!C4-BONEFINDER!C4),2)+POWER((-DOLPHIN!C5-BONEFINDER!C5),2)),2)

EUCLIDEAN DISTANCE

Ceph Record No:	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19
1	4.97	7.75	13.07	7.72	8.27	9.87	10.3	9.58	10.18	8.4	8.64	9.2	6.18	9.94	8.49	8.21	6.61	8.43	8.43
2	7.46	9.58	13.13	5.83	8.92	12.2	10.61	10.62	11.33	1.39	10.81	10.38	7.81	11.45	9.29	11.06	4.19	8.73	6.6
3	6.55	7.78	8.93	5.53	10.07	10.61	8.13	10.59	10.04	6.86	8.25	9.02	6.82	9.19	7.97	8.94	4.78	8.75	6.3
4	6.24	8.55	11.22	7.5	11.01	14.04	10.2	12.21	11.99	7.78	10.54	10.23	11.36	11.19	9.23	10.58	8.35	9.48	6.54
5	4.58	5.21	6.95	2.99	8.53	6.49	4.86	5.84	5.36	8.03	5.33	6.71	4.4	5.82	4.86	5.74	1.33	7.84	2.59
6	6.04	9.17	10.58	8.46	11.1	10.85	10.57	12.77	11.62	11.19	9.63	11.27	9.06	10.62	10.6	11.06	9.97	8.82	5.46
7	6.56	8.49	8.56	8.66	12.05	10.26	10.06	11.06	11.03	4.59	9.36	11.72	7.97	10.11	10.67	10.28	6.68	18.61	7.56
8	6.32	7.28	8.96	6.53	10.84	10.87	9.6	10.72	9.9	12.03	9.18	9.69	9.31	10.89	9.27	9.13	6.75	8.03	5.69
9	6.14	1.6	5.82	3.96	6.23	5.78	6.45	6.98	6.48	7.54	5.82	5.87	6.03	5.71	4.78	6.41	2.03	5.83	4.73
10	3.99	0.58	7.52	8.09	5.9	11.4	9.31	9.1	8.79	9.36	12.57	9.14	5.32	13.5	6.44	9.3	7.43	3.16	7.06
11	3.85	1.66	5.53	6.73	2.57	3.78	3.74	2.49	3.74	5.28	3.54	3.88	3.25	3.18	2.92	4.35	2.44	3.11	4.31
12	5.64	9.27	10.69	7.72	10.35	12.73	12.55	12.61	12.99	6.67	11.55	12.16	8.49	12.33	6.35	12.65	8.18	7.32	7.46
13	5.81	5.49	9.58	8.7	8.88	7.92	7.18	7.72	7.83	6.62	7.51	7.22	7.19	7	5.79	7.08	4.95	6.08	3.74

Figure 9.24: Step 1

: The Euclidean distances were calculated for each landmark of each cephalogram. The red box indicates the equation used on Excel.

Appendix H: Data Capture table for BoneFinder® and Dolphin Imaging™ Landmarks

Ceph Landmarks:	L1	L2	L3	L4	L5	L6-L19
Ceph No:						…
1						…
2						…
3						…
4						…
5						…
6						…
7						…
8						…
9						…
10						…
11						…
12						…
13						…
14						…
15						…
…						…

Appendix I: Data Capture table for computer-assisted landmark detection (Dolphin Imaging™)

Ceph Landmarks:	L1			L2			L3-L19
Examiner:	1	2	3	1	2	3	…
Ceph No:							…
1							…
2							…
3							…
4							…
5							…
6							…
7							…
8							…
9							…
10							…
11							…
12							…
13							…
14							…
15							…
…							…

Appendix J: Letter to Dean of Faculty of Dentistry

UNIVERSITY *of the* WESTERN CAPE

FACULTY OF DENTISTRY

Private Bag X 1, Tygerberg, 7505
Cape Town, South Africa
E-mail: suvarna.indermun@gmail.com

25 October 2019

Dear Prof Osman

RE: Request for permission to use cephalometric records from the Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, University of the Western Cape

I am writing this to request permission from the Dean's office to analyze and use the cephalograms of patients who meet the inclusion criteria of my study. The purpose is to complete my mini-thesis that is in partial fulfilment of my requirements for the degree of a Masters in Oral and Maxillofacial Radiology.

All ethical considerations will be adhered to as set out in my protocol presentation (11/10/2019).

Thank you for your consideration.

Kind regards,

Dr Suvarna Indermun
1st Year MSc Student
Department of Oral and Maxillofacial Radiology
University of the Western Cape

FROM HOPE TO ACTION THROUGH KNOWLEDGE.

Appendix K: Letter to Head of Department of Oral and Maxillofacial Radiology

UNIVERSITY *of the* WESTERN CAPE

FACULTY OF DENTISTRY

Private Bag X 1, Tygerberg, 7505
Cape Town, South Africa
E-mail: suvarna.indermun@gmail.com

25 October 2019

Dear Dr Shaik

RE: Request for permission to use cephalometric records from the Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, University of the Western Cape

I am writing this to request your permission to analyze and use the cephalograms of patients who meet the inclusion criteria of my study. The purpose is to complete my mini-thesis that is in partial fulfilment of my requirements for the degree of a Masters in Oral and Maxillofacial Radiology.

All ethical considerations will be adhered to as set out in my protocol presentation (11/10/2019).

Thank you for your consideration.

Kind regards,

Dr Suvarna Indermun
1st Year MSc Student
Department of Oral and Maxillofacial Radiology
University of the Western Cape

FROM HOPE TO ACTION THROUGH KNOWLEDGE.

Appendix L: Letter from Head of Department of Oral and Maxillofacial Radiology

UNIVERSITY *of the* WESTERN CAPE

FACULTY OF DENTISTRY

Private Bag X 1, Tygerberg, 7505
Cape Town, South Africa
Tel: (021) 937 3110
E-mail: sshaik@uwc.ac.za

28 October 2019

To Whom It May Concern

RE: Permission to use cephalometric records from the Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, University of the Western Cape

I hereby grant permission to Dr Suvarna Indermun to analyze and use the cephalograms from the Department of Radiology. The purpose is to complete her mini-thesis that is in partial fulfilment of her requirements for the degree of a Masters in Oral and Maxillofacial Radiology.

All ethical considerations will be adhered to as set out in her protocol presentation (11/10/2019).

Thank you.

Kind regards,

Dr Shoayeb Shaik
HOD
Department of Oral and Maxillofacial Radiology
University of the Western Cape

FROM HOPE TO ACTION THROUGH KNOWLEDGE.

Appendix M: Capture Sheet with patient demographics

DEMOGRAPHICS			
Ceph No:	**OHC No:**	**Age:**	**Gender:**
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
…			

Appendix N: Ethical Clearance Letter

UNIVERSITY of the WESTERN CAPE

OFFICE OF THE DIRECTOR: RESEARCH
RESEARCH AND INNOVATION DIVISION

Private Bag X17, Bellville 7535
South Africa
T: +27 21 959 4111/2948
F: +27 21 959 3170
E: research-ethics@uwc.ac.za
www.uwc.ac.za

19 March 2020

Dr S Indermun
Faculty of Dentistry

Ethics Reference Number: BM19/10/3

Project Title: Cephalometric landmark detection: Artificial intelligence vs human examination.

Approval Period: 29 November 2019 – 29 November 2020

I hereby certify that the Biomedical Science Research Ethics Committee of the University of the Western Cape approved the scientific methodology and ethics of the above mentioned research project.

Any amendments, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.

Please remember to submit a progress report by 30 November for the duration of the project.

The Committee must be informed of any serious adverse event and/or termination of the study.

Ms Patricia Josias
Research Ethics Committee Officer
University of the Western Cape

NHREC REGISTRATION NUMBER -130416-050

FROM HOPE TO ACTION THROUGH KNOWLEDGE.

Appendix O: **STROBE Guidelines**

Checklist of items that should be included in reports of ***cross-sectional studies***

		Item No	Recommendation
Title and abstract	X	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
	X		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
			Introduction
Background/rationale	X	2	Explain the scientific background and rationale for the investigation being reported
Objectives	X	3	State specific objectives, including any prespecified hypotheses
			Methods
Study design	X	4	Present key elements of study design early in the paper
Setting	X	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	X	6	(*a*) Give the eligibility criteria and the sources and methods of selection of participants
Variables	X	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	X	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	X	9	Describe any efforts to address potential sources of bias
Study size	X	10	Explain how the study size was arrived at
Quantitative variables	X	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	X	12	(*a*) Describe all statistical methods, including those used to control for confounding
	X		(*b*) Describe any methods used to examine subgroups and interactions
	X		(*c*) Explain how missing data were addressed
	X		(*d*) If applicable, describe analytical methods taking account of sampling strategy
	X		(*e*) Describe any sensitivity analyses
			Results
Participants	X	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed

Descriptive data	X	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
	X		(b) Indicate the number of participants with missing data for each variable of interest
Outcome data	X	15*	Report numbers of outcome events or summary measures
Main results	X	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
	X		(*b*) Report category boundaries when continuous variables were categorized
	X		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	X	17	Report other analyses done—eg analyses of subgroups and interactions and sensitivity analyses
			Discussion
Key results	X	18	Summarise key results with reference to study objectives
Limitations		19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	X	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	X	21	Discuss the generalisability (external validity) of the study results
			Other information
Funding	X	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for exposed and unexposed groups.