Université de Montréal

The reliability of cephalometric tracing using AI

By

Emmanuel Suissa

« Département de santé buccale – Section d'orthodontie »

« Faculté de Médecine Dentaire »

Thesis presented to the « Faculté des études supérieures » to obtain a Master's degree (M.Sc.)
in dental medicine, orthodontic option

February 2023

Université de Montréal

Department of Oral Health, Section of Orthodontics, Faculty of Dentistry

---

This memoire entitled

The reliability of cephalometric tracing using AI

Presented by

Emmanuel Suissa

Was evaluated by the following jury.

Jack Turkewicz

Chair-Rapporteur

Normand Bach

Research Director

Jeremie Abikhzer

Jury Member

Résumé

Introduction : L'objectif de cette étude était de comparer la différence entre l'analyse céphalométrique manuelle et l'analyse automatisée par l'intelligence artificielle afin de confirmer la fiabilité de cette dernière. Notre hypothèse de recherche était que la technique manuelle est la plus fiable des deux méthodes.

Méthode : Un total de 100 radiographies céphalométriques latérales étaient recueillies. Des tracés par technique manuelle (MT) et par localisation automatisée par intelligence artificielle (AI) étaient réalisés pour toutes les radiographies. La localisation de 29 points céphalométriques couramment utilisés était comparée entre les deux groupes. L'erreur radiale moyenne (MRE) et un taux de détection réussie (SDR) de 2 mm étaient utilisés pour comparer les deux groupes. Le logiciel AudaxCeph version 6.2.57.4225 était utilisé pour l'analyse manuelle et l'analyse AI.

Résultats : Un des radiographies a été éliminé de l'étude , parce que l'échelle millimétrique manquait, laissant 99 radiographies dans l'étude. Le MRE et SDR pour le test de fiabilité inter-examinateur étaient respectivement de 0,87 ± 0,61mm et 95%. Pour la comparaison entre la technique manuelle MT et le repérage par intelligence artificielle AI, le MRE et SDR pour tous les repères étaient respectivement de 1,48 ± 1,42 mm et 78 %. Lorsque les repères dentaires étaient exclus, le MRE a diminué à 1,33 ± 1,39 mm et le SDR a augmenté à 84 %. Lorsque seuls les repères des tissus durs étaient inclus (excluant les points des tissus mous et dentaires), le MRE a diminué encore à 1,25 ± 1,09 mm et le SDR a augmenté à 85 %. Lorsque seuls les points de repère des tissus mous étaient inclus, le MRE a augmenté à 1,68 ± 1,89 mm et le SDR diminue à 78 %.

Conclusion: La performance du logiciel était similaire à celles précédemment rapportée dans la littérature pour des logiciels utilisant un cadre de modélisation similaire. Nos résultats ont révélé que le repérage manuel a donné lieu à une plus grande précision. Le logiciel a obtenu de très bons résultats pour les points de tissus durs, mais sa précision a diminué pour les tissus mous et dentaires. Nous avons conclu que cette technologie est très prometteuse pour une application en milieu clinique sous la supervision du docteur.

Mots-clés : Identification automatique ; analyse céphalométrique; points céphalométriques ; Intelligence artificielle ; Apprentissage automatique ; Apprentissage profond

Abstract

Introduction: The objective of this study was to compare the difference between manual cephalometric analysis and automatic analysis by artificial intelligence, to confirm the reliability of the latter. Our research hypothesis was that the manual technique was the most reliable of the methods and is still considered the gold standard.

Method: A total of 100 lateral cephalometric radiographs were collected in this study. Manual technique (MT) and automatic localization by artificial intelligence (AI) tracings were performed for all radiographs. The localization of 29 commonly used landmarks were compared between both groups. Mean radial error (MRE) and a successful detection rate (SDR) of 2mm were used to compare both groups. AudaxCeph software version 6.2.57.4225 (Audax d.o.o., Ljubljana, Slovenia) was used for both manual and AI analysis.

Results: One of the radiographs was eliminated from the study since it was lacking the millimetric scale, leaving 99 radiographs in the study. The MRE and SDR for the inter-examiner reliability test were 0.87 ± 0.61mm and 95% respectively. For the comparison between the manual technique MT and landmarking with artificial intelligence AI, the MRE and SDR for all landmarks were 1.48 ± 1.42mm and 78% respectively. When dental landmarks were excluded, the MRE decreased to 1.33 ± 1.39mm and the SDR increased to 84%. When only hard tissue landmarks were included (excluding soft tissue and dental points) the MRE decreased further to 1.25 ± 1.09mm and the SDR increased to 85%. When only soft tissue landmarks were included the MRE increased to 1.68 ± 1.89mm and the SDR decreased to 78%.

Conclusion: The software performed similarly to what was previously reported in the literature for software that use analogous modeling framework. Comparing the software's landmarking to manual landmarking our results revealed that the manual landmarking resulted in higher accuracy. The software operated very well for hard tissue points, but its accuracy diminished for soft and dental tissues. Our conclusion was that this technology shows great promise for future application in clinical settings under the doctor's supervision.

Keywords: Automated identification; Cephalometric analysis; Cephalometric landmarks; Artificial intelligence; Machine learning; Deep learning

Table of Contents

List of tables

List of Figures

List of acronyms and abbreviations

AI: Artificial Intelligence

VTO: Visual Treatment Objective

ALARA: As Low As Reasonably Achievable

ML: Machine Learning

DL: Deep Learning

NN: Neural Network

ANN: Artificial Neural Network

CNN: Convolutional Neural Network

CBCT: Cone-beam Computed Tomography

AJO-DO: American Journal of Orthodontics and Dentofacial Orthopedics

RF: Random Forest

RFRV: Random Forest Regression-voting

SDR: Successful detection Rate

MRE: Mean Radial Error

« Ne jamais oublier que peu importe notre origine,

nous sommes tous des humains et que

la vie sur terre est belle quand on ne

fait pas de différence entre les personnes »

-Dr Claude Remise

Acknowledgements

I would like to thank my research director Dr. Normand Bach for his help and support throughout every step of this thesis. Thank you for your patience and tolerance of my procrastination. You've become a friend throughout this experience. I would also like to thank M. Pierre Rompré for his statistical expertise and support throughout the project. Thanks to the Graduate and Postdoctoral Fellowship in Artificial Intelligence and the Paul Geoffrion funds for their financial support.

Thank you to my classmates, Etienne, Hortense, Marie-Hélène and Eliyahou and to all my other co-residents. My incredible experience throughout these past few years is all thanks to you.

I would also like to thank the members of my jury, Dr. Jack Turkewicz and Dr. Jeremie Abikhzer, for your time and valuable input.

I would like to thank all my professors and clinicians in our section for your dedication to the residents and for offering such high quality of teaching. I aspire to be as skilled and sharp as you all. A special thanks to Dr Jack Turkewicz who wore many hats throughout the past year. Your efforts, positive attitude, and your true presence with the residents were crucial and well acknowledged during the difficult times.

Thank you to the love of my life, Magalie, for putting up with me during these past 9 years when I decided to go back to school. Your love and support gave me the energy to thrive and the environment to succeed. Thank you for giving me the best gift I could ever ask for, our daughter Adelle, that is now the light of our lives.

Thank you to my parents and my sisters for your support and encouragement. You guys are my rocks. Thank you for teaching me that there is nothing in life more important than family. I know I can always count on you.

A very special thought goes to Dr Claude Remise. Your knowledge, both of orthodontics and of life in general, will remain with us for the rest of our lives. Your love for all humans and animals was inspiring. You were a guru in the field and your teachings will guide me through my entire career. I will make sure to think of you when I visit the Magic Kingdom in Orlando with my daughter. May you rest in peace.

1 Introduction

Since the beginning of cephalometric radiography in the 1930s, cephalometric analysis has always been an important tool in diagnosis, treatment planning and growth analysis, as well as a method to quantify the effects of orthodontic and surgical treatments. Manual cephalometric tracing is a painstaking and time-consuming task. Although digital tracing software can automatically calculate cephalometric measurements and angles, time is required for manual localization and positioning of cephalometric landmarks on the monitor. In addition, errors in cephalometric point location, operator experience, and the subjective nature of tracking also pose problems with this approach. To overcome these shortcomings, a fully automated approach based on the use of artificial intelligence is now available to automatically position cephalometric points.

The purpose of this study was to compare the difference between manual analysis and automatic analysis by artificial intelligence, to confirm the reliability of the latter. Our research hypothesis was that the manual technique is the most reliable of the methods.

2 Literature Review

2.1 Cephalometry

2.1.1 History of Cephalometry

The first efforts in craniofacial studies began with anthropologists and artists from the 13th to 15th centuries. Anthropologists used direct measurements on dry skulls to obtain data on facial shape. Leonardo da Vinci, in the 1400s, provided the first applicable form of facial characterization using a multi-line system that allowed him to reliably reproduce the position of the head and assess aspects of facial shape (1). Two important developments in the second half of the 19th century paved the way for cephalometric radiography. The first was the need to standardize the position of the skull, which led to the first craniostat by Pierre Broca, a French anthropologist, and the second was the use of X-rays, for which Wilhelm C. Roentgen received the Nobel Prize in 1917. (1)

In 1924, orthodontist Dr. Holly Broadbent modified Dr. Wingate Todd's craniostat by adding a millimeter scale that allowed direct measurements on dry skulls (Figure 1. –). In 1925, Drs. Broadbent and Todd modified the craniostat a second time by adding an X-ray sensor (the "roentgenographic craniometer"). This allowed for accurate standardization of skull radiographs using dry skulls. In 1926, Broadbent adapted the X-ray craniometer to hold the head of a living subject, while taking lateral and posterior-anterior radiographs(2). A few years later, in 1931, the first commercial cephalometer was introduced: "The Broadbent-Bolton cephalometer" (Figure 2). (1)

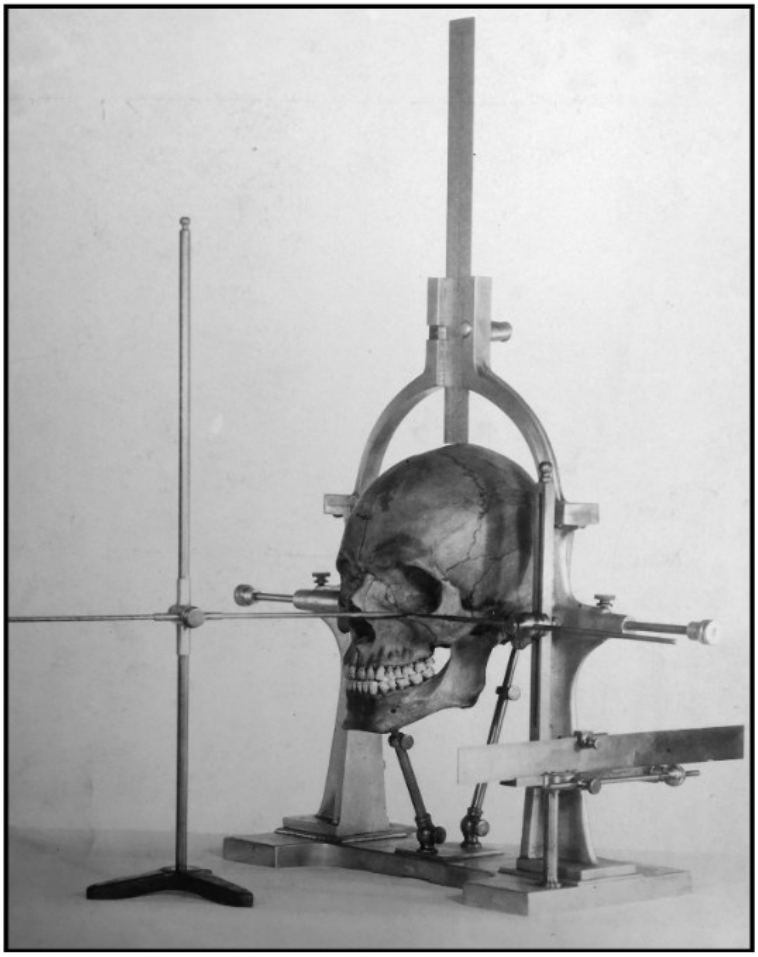


Figure 1. – Craniostat developed by Broadbent and Todd(1)

Broadbent and Todd realized that the true value of cephalometric radiography was its ability to study changes in the anatomy of the normal human skeleton over time. Longitudinal growth studies thus became popular. In Cleveland, Todd initiated the Brush survey to study normal skeletal development from birth to adulthood, and Broadbent directed the Bolton study that focused on lateral and frontal cephalometric radiographs. The Bolton Study began in 1929 and examined the craniofacial and dentofacial development of 4309 children, twice a year from infancy to age 20. In 1937 Broadbent presented the results of the Bolton Study to the Angle School of Orthodontia. He demonstrated the growth in the vertical dimension, the impact of tooth eruption on the vertical dimension, the downward and forward growth of the face, the position of the hard palate during growth, and the cessation of growth around 9 years of age of the spheno-occipital suture. (3)

Figure 2. – The first cephalometer on display on the third floor of the dental school in the Bolton Brush Growth Study Center, Cleveland, Ohio (1)

About twenty years after its introduction to the market, the cephalometer became a popular modality in clinical and pedagogical orthodontics, as a result of W.B. Downs publishing in 1948, the first cephalometric analysis of dental and skeletal patterns. (4) In 1953, Steiner introduced the notion of using cephalometry in the establishment of a treatment plan, by considering the skeletal elements, the angulation of the incisors, the amount of overlap and the profile of the patient. (5) Another important advancement was the 1975 publication of the "Bolton Standards" by Dr. Broadbent Jr., based on the Bolton Study subjects. In 1979, Ricketts popularized the "Visual Treatment Objective (V.T.O.)", to better establish treatment plans and also proposed the superimposition of pre-treatment and post-treatment cephalograms to understand the effects of orthodontic treatment. In the same study, Ricketts proposed methods for predicting growth using the cephalogram. (6)

Cephalometry occupied a significant presence in the orthodontic literature in the 20th century and gave the specialty two important tools. Firstly, imaging allowed the production of serial radiographs, which led to the development of superimposition techniques that allowed us to isolate changes in skeletal and tooth movement over time and to evaluate the effects of our treatments. Finally, cephalometry has given us a diagnostic tool to confirm our clinical assessment of a patient's craniofacial morphology. (1)

2.1.2 Clinical importance of cephalometry

In the early 20th century, Angle taught his students how to develop a treatment plan based on the patient's profile and dental malocclusion (7). Since the popularization of cephalometric radiography after World War II, orthodontists have been able to measure the changes in the teeth and jaws produced by both growth and treatment. These radiographs have clearly demonstrated that a large majority of Class-II and Class-III malocclusions have skeletal components that are not necessarily associated with the dental relationships.(8) In fact, the majority of malocclusions have a combination of both components. An example often reported in the literature is the excessive overjet demonstrated by Bjork in 1961. Figure 3. – schematically demonstrates how excessive horizontal overjet can be a result of alveolar protrusion/retrusion, maxillary/mandibular incisor tilt, or maxillary protrusion/mandibular retrognathism giving 243 possible combinations that may contribute to the presence of an increased overjet in our patients(7). Therefore, when developing a treatment plan, most orthodontists will obtain diagnostic models, intra-oral and extra-oral photographs, as well as panoramic and cephalometric radiographs to fully understand and analyze the source of the malocclusion(9).

The goals of cephalometric analysis can be summarized into three categories:

1. To evaluate the vertical and horizontal relationships of the five major functional components of the face: the cranium or cranial base, the skeletal maxilla, the skeletal mandible, the maxillary dentition and its alveolar process, the mandibular dentition and its alveolar process (Figure 4. – (8)
2. To analyze growth and treatment effects
3. To make a prediction or simulation of growth or treatment (7)

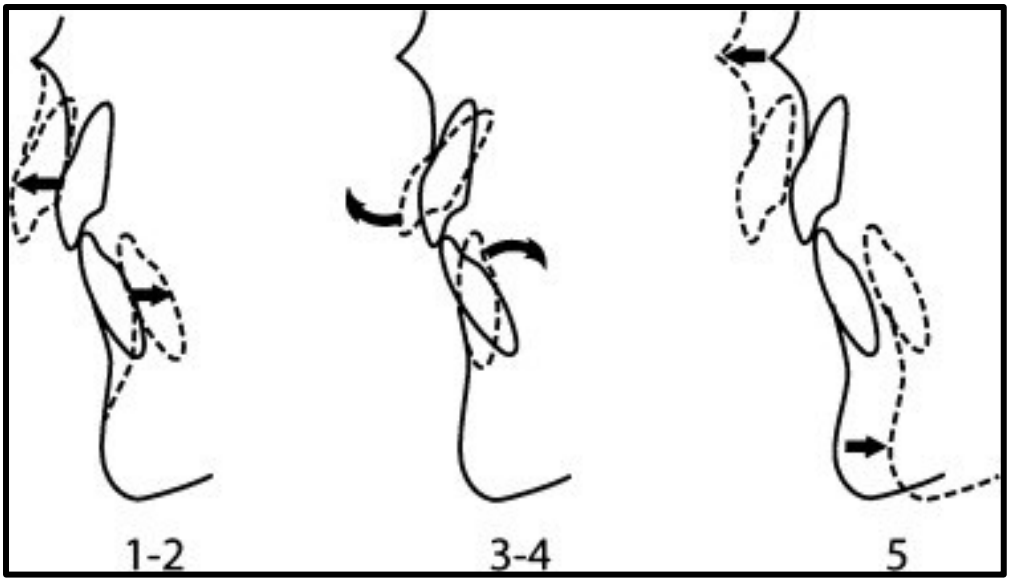


Figure 3. – Possible combinations of dentoalveolar and skeletal relationships in cases with excessive overjet (7)

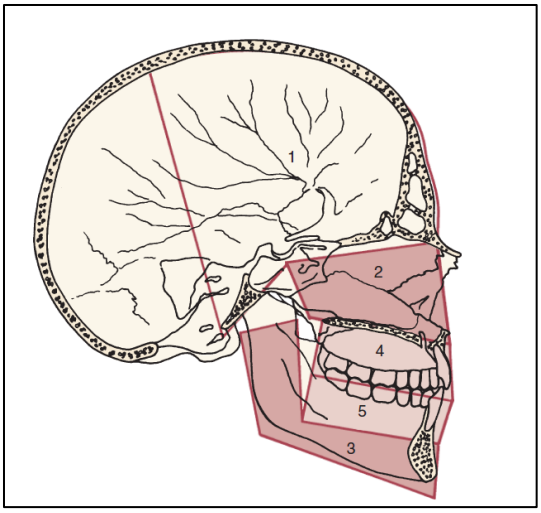


Figure 4. – The five structural components of the face as seen on a cephalometric x-ray (8)

2.1.3 Cephalometric Analysis

Cephalometric analysis begins with the acquisition of a radiological image of the lateral or frontal surface (posterior-anterior cephalometry). This requires an x-ray source, an adjustable cephalostat, a film cassette or digital image device (Figure 5). The cephalostat is used to maintain the patient's head position through bilateral ear rods placed in the external auditory canals. The orthodontist must then perform the analysis by defining skeletal and dental structures named cephalometric landmarks, connecting these points to create a cephalometric tracing, and finally making measurements of angles, distances, and ratios on these tracings.

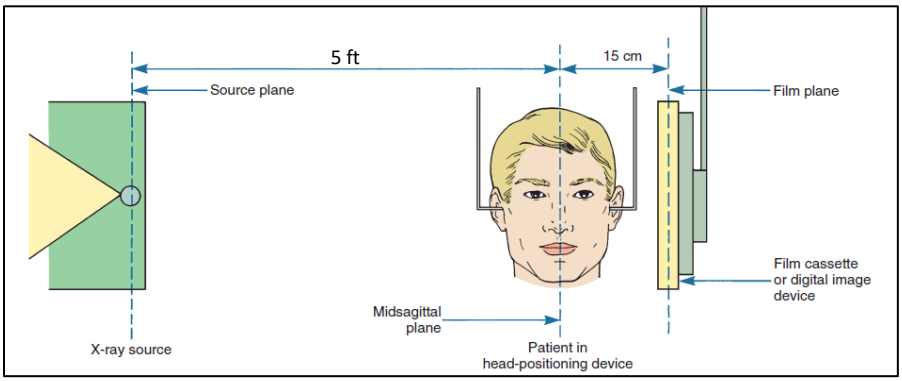


Figure 5. – Diagrammatic representation of the American standard for lateral cephalometric image capture. Figure modified from Fig 6.42 in Proffit W.R.'s *Contemporary Orthodontics* 6th edition (8)

Cephalometric landmarks are a series of points which can either be defined as exact positions on a physical model, an extreme point (e.g. most anterior point of the chin), or as a constructed point such as the intersection of two planes(8). Cephalometric landmarks can be divided into hard tissue, soft tissue and dentoalveolar landmarks. The most commonly used landmarks can be found in Figure 6 and their definitions can be found in Table 1. - Common cephalometric landmarks and their definitions (13).

Errors in cephalometric analysis may occur for many reasons. The most important type of errors involves the inconsistent and imprecise landmark identification, and may sometimes lead to erroneous diagnoses and treatment plans (10). Since cephalometric x-rays are two-dimensional images of many bilateral structures, there will inevitably be different degrees of superimposition. The localizations of certain landmarks such as porion (Po), orbitale (Or), condylion (Co), anterior and posterior nasal spine (ANS & PNS), may be more difficult to locate and thus more prone to error due to the overlapping structures superimposed on the landmarks. Another source of error is the quality of the radiographic image which can interfere with the identification of landmarks such as Po, Co, Or, ANS, gonion (Go), and glabella (G). Additionally, some authors have argued that the level of an observer's experience plays an important role in landmark identification. Studies have shown that the landmarks with the largest localization variability among orthodontists were ANS, Or, Po, Co and Me (10) (11). Some cephalometric landmarks seem to be more reliable in the horizontal (x) or in the vertical (y) planes, which indicates that the distribution of error is asymmetric on lateral cephalograms. Studies have shown that differences along the x-axis tend to be greater than those on the y-axis(10, 11) . Some authors have argued that landmark identification errors of less than 1 mm are clinically acceptable. The consensus in the literature is that that errors of less than 2 degrees or 2 mm would likely not make a significant difference in the diagnosis and treatment (10-12).

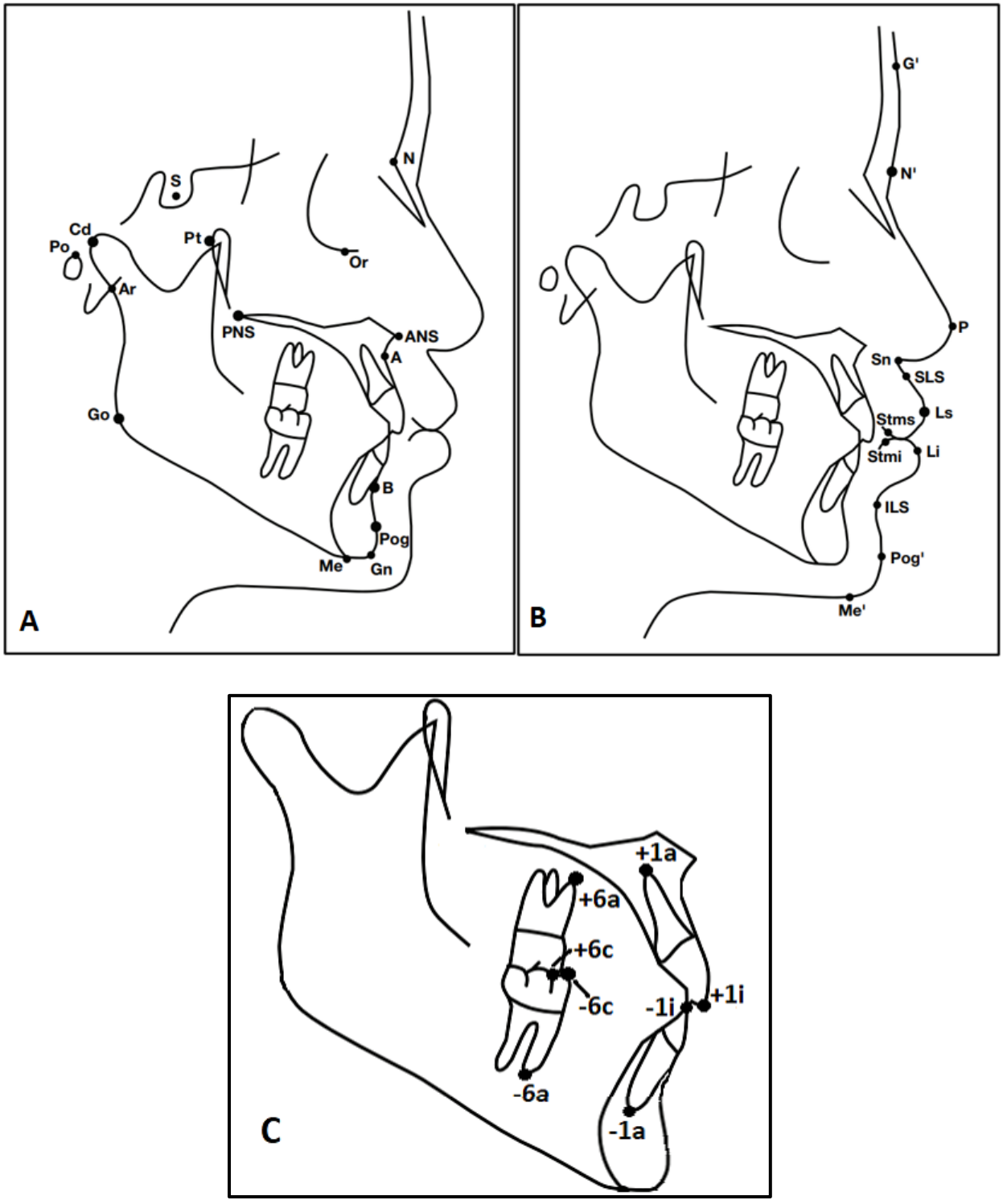


Figure 6. – (A) Cephalometric hard tissue landmarks (13), (B) Cephalometric soft tissue landmarks (13), (C) Cephalometric dentoalveolar landmarks.

Table 1. - Common cephalometric landmarks and their definitions (13)

Landmark	Abbreviation	Definition
Hard Tissue Points		
Subspinale	A	Deepest point on the maxilla below ANS
Anterior nasal spine	ANS	Anterior point of maxilla.
Articulare	Ar	Point on the posterior border of the ramus at the intersection with the basilar portion of the occipital bone
Supramentale	B	Most posterior point on the bony curve of the mandible above pogonion
Condylion	Cd	Most superior and posterior point on the head of the condyle
Gonion	Go	Most posterior and inferior point on the outline of the angle of the mandible
Menton	Me	Lowest point on the symphysis of the mandible
Nasion	N	Junction of frontonasal suture
Orbitale	Or	Inferior border of orbit
Pogonion	Pg or Pog	Most anterior point of bony chin
Posterior Nasal Spine	PNS	Posterior point of bony hard palate.
Porion	Po	Top of external auditory meatus.
Pterygomaxillary Fissure	PTM	Most posterior and superior point on the outline of the pterygomaxillary fissure
Sella	S	Mid-point of sella turcica
Soft Tissue Points		
Soft tissue A point	A' or SLS	Deepest midline point on outline of the Superior labial sulcus.
Soft tissue B point	B' or ILS	Deepest midline point on outline of the Inferior labial sulcus
Soft tissue Glabelle	G’	Most prominent point in the mid sagittal plane of the forehead.
Labius Inferius	LL or Li	Most anterior point on outline of lower lip (vermillion border)
Soft tissue Menton	Me'	Lowest point on outline of soft tissue chin.
Soft tissue Nasion	N’	Deepest part of the soft tissue outline in front of Nasion.
Soft tissue Pogonion	Pg’ of Pog’	Most anterior point on outline of soft tissue chin
Pronasale	Pn or P	Anterior tip of the nose
Subnasale	Sn	Junction of nasal columella and upper lip in mid-sagittal plane.
Stomium Superior	STM or Stms	Lowest midline point on outline of upper lip.
Stomium Inferior	St- or Stmi	Highest midline point on outline of lower lip.
Labius Superius	UL or Ls	Most anterior point on outline of upper lip (vermillion border)
Dental Points		
Apex of upper incisor	+1a	Tip of the apex of the upper incisor
Apex of lower incisor	-1a	Tip of the apex of the lower incisor
Incisal edge of upper incisor	+1i	Tip of the incisal edge of the upper incisor
Incisal edge of lower incisor	-1i	Tip of the incisal edge of the lower incisor
Upper molar mesial Apex	+6a	Mesial Apex of the first upper molar
Lower molar mesial apex	-6a	Mesial apex of the first lower molar
Mesial Buccal Cusp of upper 1st Molar	+6c	Tip of the mesiobuccal cusp of the first upper molar
Mesial buccal Cusp of Lower 1st Molar	-6c	Tip of the mesiobuccal cusp of the first lower molar

The traditional method for analyzing cephalometric radiographs was by manual tracing, where the orthodontist manually identified landmarks, reference planes and angles on a matte acetate tracing paper viewed through a light source. Angular and linear measurements were done using a protractor and millimeter scale (14). Nowadays, computer assisted cephalometric analysis is more commonly used with the advent of digital x-ray films. The points are identified directly on a monitor, the computer software then completes the analysis by automatically measuring angles and distances (14). Many cephalometric analyses have been developed, and the common names are the Down, Steiner, Sassouni, McNamara, Harvold & Wits, and Enlow's analyses (8). The analysis is completed when the patient's measurements are compared with average values from the literature, which are averages of patients with "ideal" dental and skeletal relationships. A major database for contemporary analysis is the Michigan growth study. Other major sources are the Burlington growth study and the Bolton study in Cleveland (8). Most orthodontists understand that these average values should not be considered as absolute values since most of the databases are based on Caucasians of European ancestry. These averages provide only an indication to help the orthodontist characterize the patient's facial morphology, and the standard deviations give an idea of the severity of the deviation. These values are sometimes viewed as treatment goals with the idea that if the patient does not conform to these averages, there must be an underlying problem. This is an erroneous concept, as this type of practice will futilely guide the clinician's treatment plan to follow cephalometric numbers, as opposed to evaluating and treating patients' faces. It is important to remember that there are many racial and ethnic groups among our patients, and also that the aesthetic goals of one group may be different from another(7).

2.1.4 Controversy over the use of cephalometry

In 1979, Dr. George Silling asked the question, "Is a cephalometric radiograph always necessary to establish a treatment plan?" He explains that some clinicians find this radiograph indispensable in all cases, while others find it useful only in specific situations or for certain

malocclusions such as Class II Division 1 malocclusions.(15) In 2002, 90% of orthodontists in the United States routinely asked for cephalometric radiographs in their diagnostic records (16). The arguments that some authors suggest against their use are that the clinical examination could give us enough information to establish the treatment plan, and that a cephalometric analysis based on standardized normative values can mislead the orthodontist, due to the great variation in craniofacial morphology of our patients. In 2011, Devereux et al published a study in the American Journal of Orthodontics and Dentofacial Orthopedics where six patients were presented to 199 orthodontists. The availability of a lateral cephalometric radiograph and its tracing did not make a significant difference in treatment planning decisions, except for one patient who was Class I dental but Class II skeletal(17). A second more recent study done in Portugal with 43 patients and 10 orthodontists concluded that while all 10 orthodontists felt that it was important to have the cephalometric radiograph to establish a treatment plan, the results of the study showed that its presence did not have an impact on it (18).

According to the ALARA principle, it is necessary to reduce radiation exposure and eliminate unnecessary radiography. As with any form of radiography, there is an associated dose of ionizing radiation to which the patient will be exposed. Although the radiation dose from lateral cephalometric radiographs is relatively low, lateral cephalometric radiography still emits radiation to several organs that are considered radiosensitive, such as the brain, bone marrow, thyroid gland and salivary glands(19).

The effective dose of lateral cephalometric radiographs with photostimulable phosphor is 5.6 microsieverts, which represents a 51% increase in effective dose compared to the calculated effective dose in 1990. This trend of increasing effective dose is an important indicator to study the diagnostic value of radiographs and whether there are acceptable methods to limit patient radiation exposure. The ionizing radiation dose delivered by lateral cephalometry is an important concern in the field of orthodontics because orthodontic patients are often children and adolescents who are still in the growth and development phase. Ionizing radiation has the potential to damage DNA and thus increase the overall risk of cancer. (9)

2.2 Artificial Intelligence (AI)

2.2.1 General Concepts

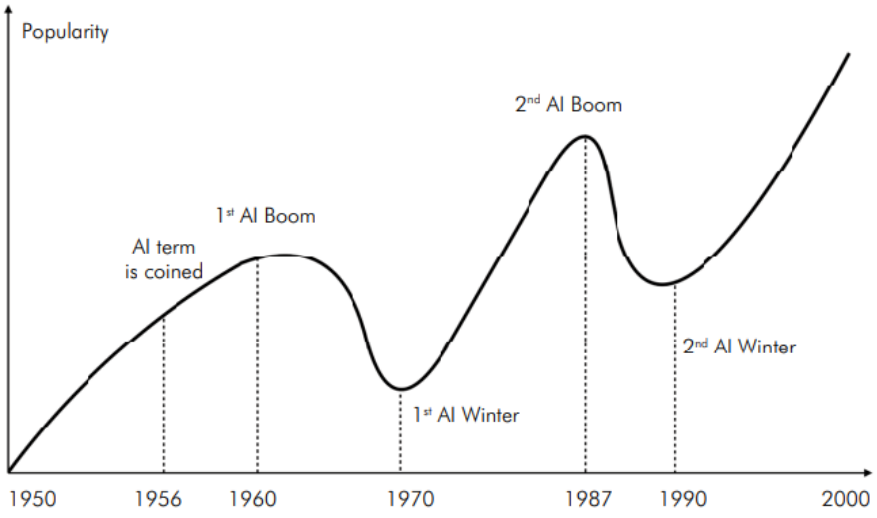
Artificial intelligence (AI) is the general term used to describe technologies that allow machines to perform tasks that would normally require human cognition(20). This technology is a branch of computer science and it's ultimate goal is to build intelligent machines that are often under the form of software programs(21). These programs are composed of a sequence of operations that are designed to perform a specific task. Historically, the task of "teaching" a machine to perform an intelligent task required knowledge of the specific domain and manual fine tuning of a software. This type of program required what we call supervised learning, meaning that the programmer needed to teach the software the algorithms and patterns that it wanted it to detect. The term AI is not new, as it was first used in 1956. Since the 1950's, AI has gone through several phases of popularity and disappointments. These periodic downfalls of AI throughout history were called "AI Winters" and are shown in Figure 7. – along with a brief history of its development.

2.2.2 Machine Learning (ML)

Machine learning (ML) is a more recent subfield of AI where the system learns rules from data, rather than having humans provide these rules. A basic analogous example of this method is one of an adult showing a series of photos of cats to a child, and the child eventually learns the patterns necessary in recognizing a cat(21). The data that is used for ML can be simple or complex, and when complex data is used, neural networks (NNs) are usually employed(20). The main constituent of any NN is the artificial neuron that was inspired by the human brain (22). The artificial neural network (ANN) is a structure composed of many communicating neurons that are organized in layers. The basic composition of the neural network is to have an input layer, an output layer, and at a minimum, one hidden layer (Figure 9. – The term "hidden" is used to describe the layers in between the input and output because their values are not pre-determined

nor visible, and their goal is to build progressively from information from the visible input layer, to then calculate an output which is then taken as input by the next layer, and ultimately computes the correct value of the visible output layer. (21)

A



B

1938 - 1946: Golden age of science fiction

1950 - 1960: First AI Boom: Can machines think? Turing test, method to determine the intelligence of a computer by A. Turing; Logic Theorists, the fist AI programme is created; Dartmouth summer conference on AI, the term AI is used for the first time;

1965: ELIZA, a natural language programme is created, which simulates a conversation with a psychotherapist;

1970s: First AI winter;

1980s - 1990s: Second AI Boom: E. Feigenbraum introduces expert systems, which emulates decisions of human experts;

1990s: Second AI winter;

1997: Deep Blue, a chess-playing computer beats the world champion Gary Kasparov;

2006: Geoffrey Hinton from the University of Toronto develops Deep Learning and publishes "A Fast Learning Algorithm for Deep Belief Nets";

2011 - 2014: Watson (IBM) wins jeopardy; Apple integrates Siri, an assistant with voice; Alexa, virtual assistant from Amazon;

2018: European Union establishes guidelines for dealing with ethics in AI;

Figure 7. – Timeline illustration of AI development (20)

2.2.3 Deep Learning

When an ML process requires the application of multilayered neural networks it would be defined as Deep Learning (DL). The hierarchy from AI to deep learning is shown in Figure 8. Deep learning is a more recent sub-branch of ML that uses a hierarchy of composable patterns that build on each other (21), and the term "deep learning" is a reference to the "deep" / multilayered NN architectures. DL is especially suitable for complex data structures such as imagery, where they can represent an image and its hierarchical features such as edges, corners and macroscopic patterns.(22) If we use the same analogy as in ML for the DL example, the child will not recognize a cat from a single pattern matching step (layer), but will rather recognize edges and simple shapes that will help outline groups such as eyes or ears, and these will then lead in the recognition of larger groups such as legs, bodies and heads. A particular grouping of these defines the whole cat. (21)

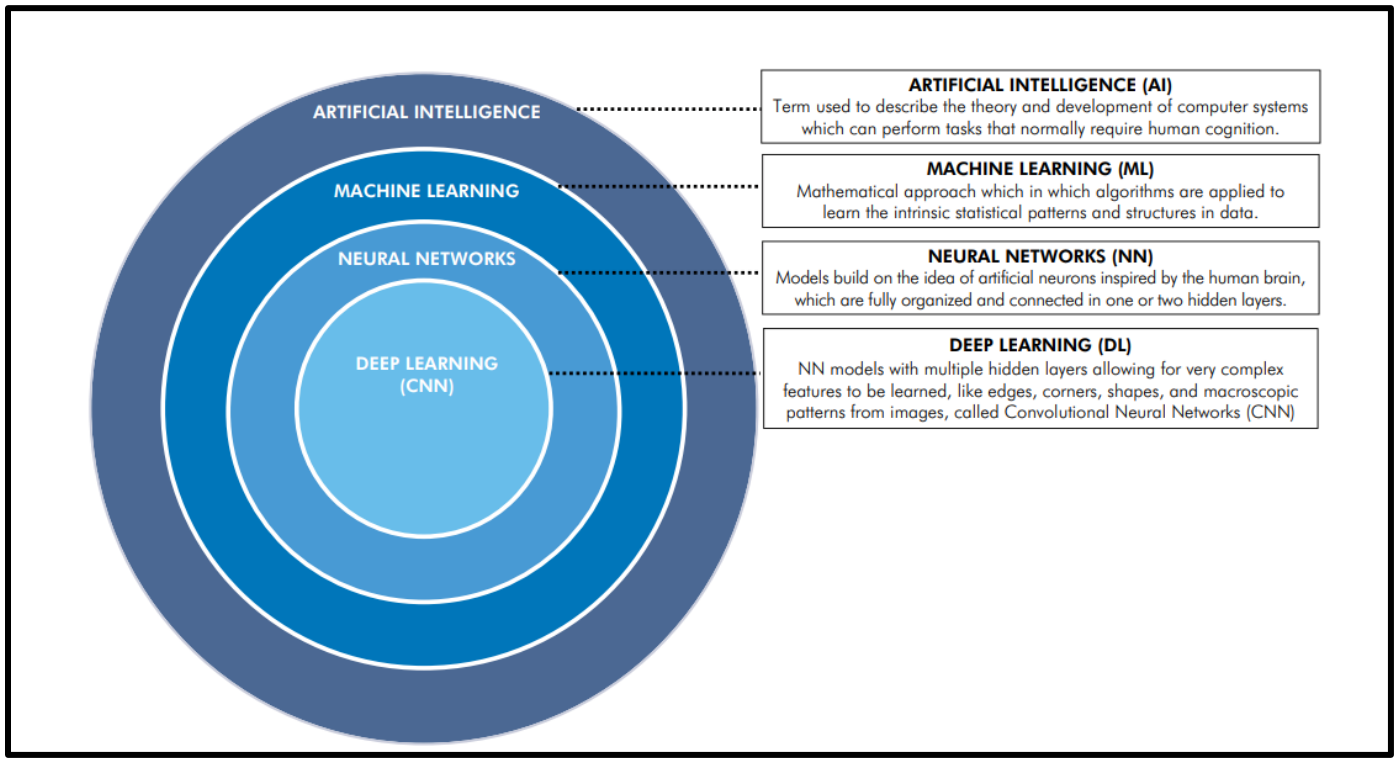


Figure 8. – Hierarchy of AI (20)

2.2.4 Convolutional neural networks

The most used deep learning algorithm for image recognition and processing in medicine and dentistry is a subclass of ANN called the convolutional neural network (CNN) (Figure 9. – The CNN algorithm uses a sliding filter to scan a small neighborhood of inputs in order to analyze a larger image(21). Its strong impact on computer vision is thanks to an architecture based on the mathematical operation called "convolution", that is applied as a matrix multiplication between the filter and the data. CNNs were first introduced in the 1980's but only became popular once more powerful computers were developed with access to larger quantities of data. As previously mentioned, what makes CNNs so popular for computer vision is its ability to extract features from data. Previously, these features had to be extracted by hand for later processing, and was considered one of the toughest and most complex tasks in computer vision.

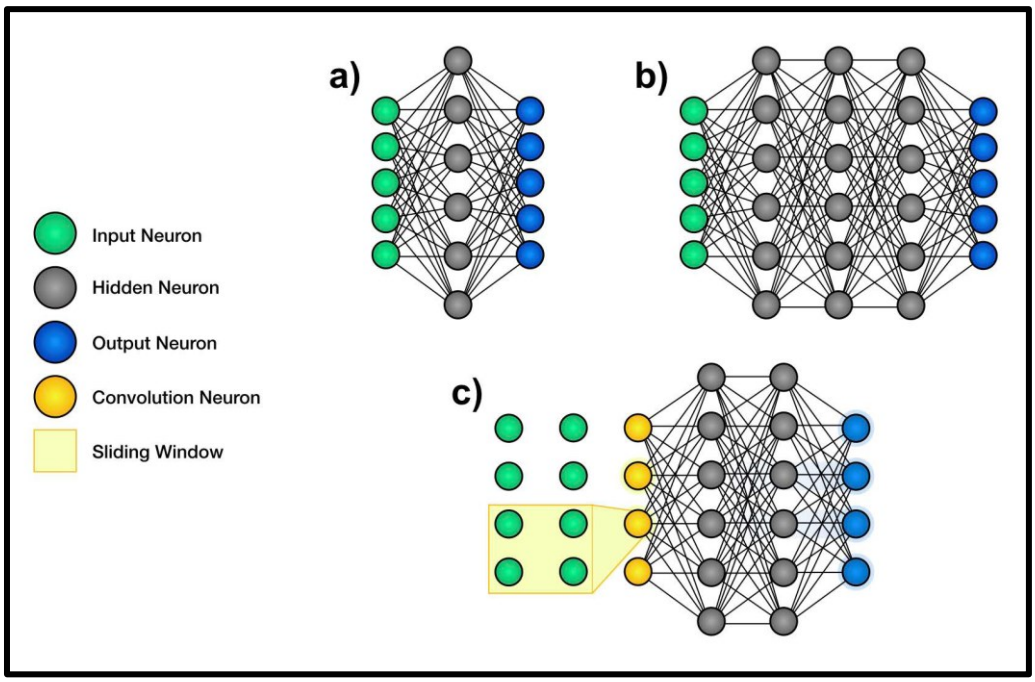


Figure 9. – Schematic representation of ANNs. **a)** ANN with a single hidden layer, typically referred to as ML. **b)** ANN with multiple layers of hidden neurons (DL). **c)** Convolutional neural networks use filters to scan a local zone of inputs(21)

2.2.5 The scope of AI in healthcare

The introduction of artificial intelligence in healthcare is revolutionizing and pushing the industry towards advancements in many clinical specialties and hospital processes. Even the most modern of healthcare organizations face many challenges in collecting, organizing and applying structured and unstructured data to diagnose and treat diseases(23). The data mining and recognition abilities of AI can provide effective methods for patient care and can lead the clinician to provide unprecedented diagnosis, treatment and care to the patient at the correct time(24).

The three major steps of medical diagnosis are (Figure 10):

1. Collecting patient medical history, signs and symptoms, observation and examination, and interpretation of the data obtained from the patient.
2. Formulation of a diagnosis based on the clinician's knowledge and experience
3. Establishment of a therapeutic plan

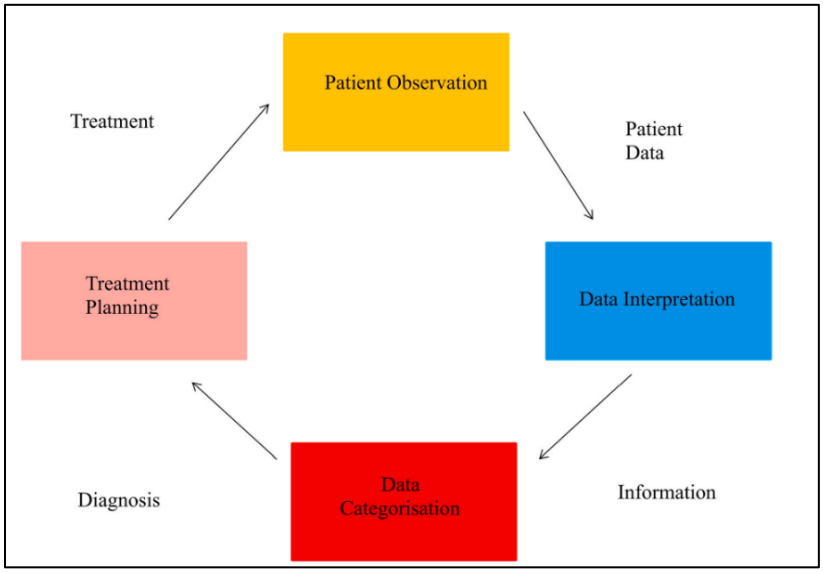


Figure 10. – Conventional medical diagnostic cycle (23)

In this conventional diagnostic cycle, the clinician represents the intelligent agent, the patient data represents the input, and the diagnosis represents the output. There are various advantages to implementing AI into this cycle:

1. AI allows for more curated and structured collection of patient data
2. AI will reduce the chance of human error, and reducing the bias associated with clinician knowledge and experience
3. Diagnostic and treatment costs are reduced.
4. Lowering costs of treatment by reducing routine tasks
5. Reducing the traditional tedious tasks associated to collecting patient details and thus increases more time for face-to-face discussion between patient and clinician. (25)

AI is currently mostly present in research labs and tech firms rather than in clinical practice. Barely a week goes by without a claim that a firm has developed an approach to use AI to diagnose and treat a disease with greater accuracy than a human clinician. Many of these findings are based on radiological image analysis, retinal scanning or genome-based precision medicine(24). Tech firms and startups are also involved in developing AI for the healthcare industry. For example, Google is building prediction models from big data to alert clinicians of high-risk conditions such as sepsis and heart failure(26). Jvion, a healthcare & clinical AI platform, helps identify which patients are more at risk, as well as those that are more likely to respond to therapeutic protocols. The immensely complex genetic nature of certain cancers makes it inevitably very difficult for clinicians to understand all the cancer genetic variants and their response to protocols. Firms like Foundation Medicine and Flatiron Health build AI that specifically focus on diagnosis and treatment recommendations for certain cancers based on their genetic profiles (24). Machine learning models are also being developed to assess population health, such as predicting populations at risk for specific diseases, accidents or hospital readmissions(27).

2.2.6 Clinical use of AI in dentistry

Thanks to the fact that most of our medical data is now stored digitally, deep artificial neural networks can be applied to many medical fields. In the field of dentistry, CNNs have shown very promising results in diagnosis and prediction in both radiology and pathology through the use of disease identification and image segmentation (20). Furthermore, deep learning algorithms are

being implemented to assist in the clinical decision-making process and even treatment planning in the fields of orthodontics(28) (29).

2.2.6.1 Disease identification and radiology

Dental caries is a chronic infectious disease and is experienced by more than 90 percent of all adults in the United States. Although dental caries represent the main dental disease that general dentists treat on a regular basis, there still seem to be difficulties with their detection using traditional methods. Studies have shown that caries detection by clinicians using visual, tactile, radiographic or a combination of these, have overall mean sensitivity of 19%-94%, meaning that sometimes up to 80% of lesions can be missed (30). A more accurate detection and diagnosis of dental caries would reduce the cost of oral health management and increase the likelihood of natural tooth preservation. ANNs have been used successfully in the detection of dental caries from periapical and bitewing radiographs. In 2018, a group used a pre-existing CNN network (GoogleNet inception v3) and trained it with 3000 periapical radiographs to detect and diagnose dental caries in premolars and molars. The detection sensitivity was 84.0 % for premolars and 92.3 % for molars (31). Another study also used another pre-existing CNN network (U-net) and trained it with 3686 bitewings radiographs and found that the neural network had a sensitivity of 75% while the dentists in the study were at 36%(32). Another impressive study showed that through a deep ANN, they were able to distinguish subjects with an absence of caries from those subjects with caries or restorations with a high degree of accuracy using only their demographic and dietary factors as input(33).

The detection and diagnosis of oral pathologies by dentists is of utmost importance during routine examinations, as their early detection can have a significant impact on their prognosis, especially for lesions that may be cancerous or precancerous in nature. Unfortunately, the overall prognosis of oral cancer remains poor because over half of patients are diagnosed at advanced stages (34). To aid in the detection and diagnosis of oral cancer and to improve the prognosis, hyperspectral images were used as input in CNNs and yielded an accuracy of 91.4% for the

classification of cancerous versus benign tumors, and an accuracy of 94.5% for the classification of cancerous versus normal tissues(35). For the detection of oral cancer, a CNN was trained with 44,409 clinical images of biopsy proven oral cavity squamous cell carcinoma and reached a detection accuracy of 92.3% (34). One study trained a CNN to distinguish radiographically between ameloblastomas and keratocystic odontogenic tumors, which can have similar radiologic presentations. The accuracy of the CNN was similar to that of clinical specialists (83.3% versus 83.2%, respectfully), the main difference being that the specialists took an average of 23.1 minutes to distinguish the differences on all images, while the total calculation time taken by the CNN to analyze all the images was 38 seconds (36).

Osteoporosis is a systemic disease that is characterized by low bone mineral density and deterioration of bone architecture. According to the International Osteoporosis Foundation, one in three women and one in five men above the age of 50 will experience an osteoporotic bone fracture. The gold standard in diagnosing osteoporosis is by evaluating bone mineral density using dual-energy X-ray absorptiometry, but this technique is known to be complex and expensive. (37) Recently, digital images of dental panoramic radiographs have been deemed a cost-effective and available method for screening osteoporosis thanks to the widespread use of panoramic radiographs in dentistry. These methods utilized manual categorization of feature indexes for screening osteoporosis from panoramic images (38). Four different CNN models were proposed and trained for screening of osteoporosis with panoramic radiographs. The best model yielded an accuracy of 84% in detecting osteoporosis from panoramic images (39). These results are very promising since the earlier detection of this disease can help prevent fractures in older populations.

2.2.6.2 Periodontics

Globally, there exist over 4000 different types of dental implant systems (40). If periodontists are unable to identify and classify what implant system is present when mechanical or biological complications occur, they would be more likely to treat with a more invasive modality. Deep CNNs have been found useful in tackling the issue of identifying the implant system on

radiographs, and were found to have an accuracy of 97.1% compared to 92.5% from board certified periodontists (41).

The American Academy of Periodontology clinically classified the two forms of periodontitis as either aggressive or chronic , and since the pathogenesis of this disease is so complex, there is no clinical, histopathological, microbiological or genetic test that can discriminate between the two forms(21). An ANN was developed and used to distinguish between the two forms of periodontitis and used specific immunologic parameters as inputs such as leukocytes, interleukins and IgG antibodies, and had an accuracy of 90-98% in classifying patients as having either the aggressive or chronic forms(42).

Another challenge in the field of periodontics is establishing the proper diagnosis and prognosis, such as the need for extraction of periodontally compromised teeth. A CNN algorithm was used to evaluate if artificial intelligence would be a useful tool to assist in establishing the diagnosis and prognosis of periodontally compromised teeth. The algorithm yielded an accuracy of 76.7-81% in evaluating the diagnosis of periodontally compromised teeth, while it yielded an accuracy of 73.4-82.8% in detecting the prognosis of these teeth(43). According to the authors, this range in accuracy was due to differences in the complexity of root anatomy between molars and premolars, where the CNN exhibited more difficulty with multirooted teeth.

2.2.6.3 Endodontics

Although mandibular molars tend to have straightforward root canal systems, several atypical differences can complicate their morphology. Cone-beam computed tomography (CBCT) has become the gold standard method of imaging in endodontics, as it helps guide the endodontist through the root canal anatomy and minimize treatment failures. The issue with CBCT is its higher dose of radiation, and for this reason it is not yet used systematically in all cases (44), especially not where tooth morphologies tend to be straightforward. A deep learning method has been proposed using a CNN to help identify the presence of additional root canals in the distal root of mandibular first molars using only dental panoramic radiographs. Once the CBCT confirmed the presence of an additional canal, the corresponding panoramic radiograph was used to train the

neural network. The CNN yielded a high accuracy of 86.9% in detecting additional root canals (45). The implementation of such technology can lead to lower doses of radiation for patients undergoing endodontic therapy.

Vertical root fractures are dental diseases that are difficult to diagnose and treat. Endodontically treated teeth are more likely to suffer from vertical root fractures and is predominantly seen in mandibular posterior teeth(46). Although the typical treatment for a tooth with vertical root fracture is extraction, an early diagnosis may allow for treatment by hemisection or root separation. This approach can have relatively high survival rates, ranging from 94% and 64% for 5 and 10 years, respectively (47). While the gold standard for detection of a vertical root fracture is CBCT, as mentioned previously, they provide a higher dose of radiation, and it should be noted that endodontically treated teeth don't always show symptoms when root fractures are present. For these reasons, panoramic radiographs would be useful in screening for vertical root fractures during routine examinations. A CNN based deep learning model for the detection of vertical root fractures from panoramic radiographs was trained. Of the 330 vertical root fractures, 267 were detected, which shows it to be a promising tool for the early diagnosis of vertical root fractures, with the goal of saving natural teeth (48).

2.2.7 Clinical use of AI in orthodontics

The term malocclusion refers to a common dental condition that can have many adverse repercussions (8):

1) Psychosocial problems: Often the main reason for consultation, a severe malocclusion can represent a social handicap and can severely affect one's self esteem.
2) Occlusal function: Even though oral function can adapt relatively well to a malocclusion, for some it can add difficulties and necessitate additional effort during function.
3) Dental trauma: Malocclusions, more specifically in the presence of protrusive maxillary incisors in Class II malocclusions, can increase the risk of dental trauma in children.
4) Oral Health: Malocclusions with severely crowded teeth can increase the incidence of caries and periodontal disease due to the added difficulty in dental hygiene (8).

An American epidemiologic survey showed that 57%-59% of each racial group had some degree of orthodontic treatment need(49). In 2015 the American Dental Association reported the following statistics on oral health and well-being(50):

1) One in four adults, and one in three young adults avoid smiling because of the condition of their mouth and teeth.
2) 22% of young adults reduce participation in social activities due to the condition of their mouth and teeth.
3) 23% of adults, and 35% of young adults feel embarrassment due to the condition of their mouth and teeth.
4) 82% of adults believe that straight, bright teeth help you get ahead in life.
5) One in five adults experience anxiety due to the condition of their mouth and teeth (50).

In order to achieve acceptable orthodontic treatment outcomes, treatment planning must be meticulously performed before the therapy begins(8). The practice of orthodontics is considered by some to be partly an art and partly a science, that is based upon the experience and bias of each clinician (51). The fact that all malocclusions are unique makes it impossible for the human brain to predictably correlate the different patterns expressed by the entire stomatognathic system(52). Clinically speaking, the fact that orthodontists treat one patient at a time makes it difficult to gather and share large datasets of their treatment outcomes to make connections from multiple clinical observations. Instead, they usually rely on their experience to provide the treatment plan and treatment modality that provides them with their perceived "maximum" efficiency. Thorough and careful evaluation of numerous factors can make treatment planning a complex process without any objective patterns, and greatly depends on the subjective judgment of the clinician (53). This treatment approach, that is based purely on experience, may lead to non-optimal outcomes and prolonged treatment times(54). Based on how AI has helped the numerous dentistry fields previously described, one can only imagine that the introduction of artificial intelligence in orthodontics will offer a new method to achieve sharper predictions from data by concurrently analyzing the different variables present in a malocclusion. This will facilitate the clinician in obtaining the best outcomes when treating malocclusions, or even assist him in determining the need for orthodontic treatment(55).

The traditional diagnostic method for evaluating and diagnosing a malocclusion possesses many difficulties which can bring uncertainty with treatment outcomes due to the numerous variables present in the analysis. When an orthodontist is faced with the variety of variables in a malocclusion, he or she must mentally evaluate and compute the best approach to solve the problem often based on previous experience. In order to help make this process simpler, many orthodontists adopt a process called the “feedforward approach” shown in Figure 11. This workflow shows the traditional forward build up based on information received during the initial examination but does not necessarily incorporate any feedback mechanisms to help improve on previous diagnostics and outcomes analyses. This lack of a feedback mechanism makes it difficult to re-evaluate treatments and learn from previous positive or negative outcomes(56). Although this traditional method is accepted and advocated by clinicians, its drawbacks are now evident and thus can be deemed inefficient to treat patients(57).

The ideal scenario would be one where the orthodontist can find potential negative outcomes and go backwards in the workflow to prevent creating errors. This process would add value in the diagnostic procedure and immensely improve the traditional unidirectional thought process. Deep learning is starting to be used to input large amounts of malocclusions and letting the algorithm predict the “ideal” treatment options. In order to predict a suitable treatment plan to a given malocclusion despite the numerous variables presented, this requires that the software accumulates and trains on a large quantity of data, and this data must be correctly labeled and weighted.

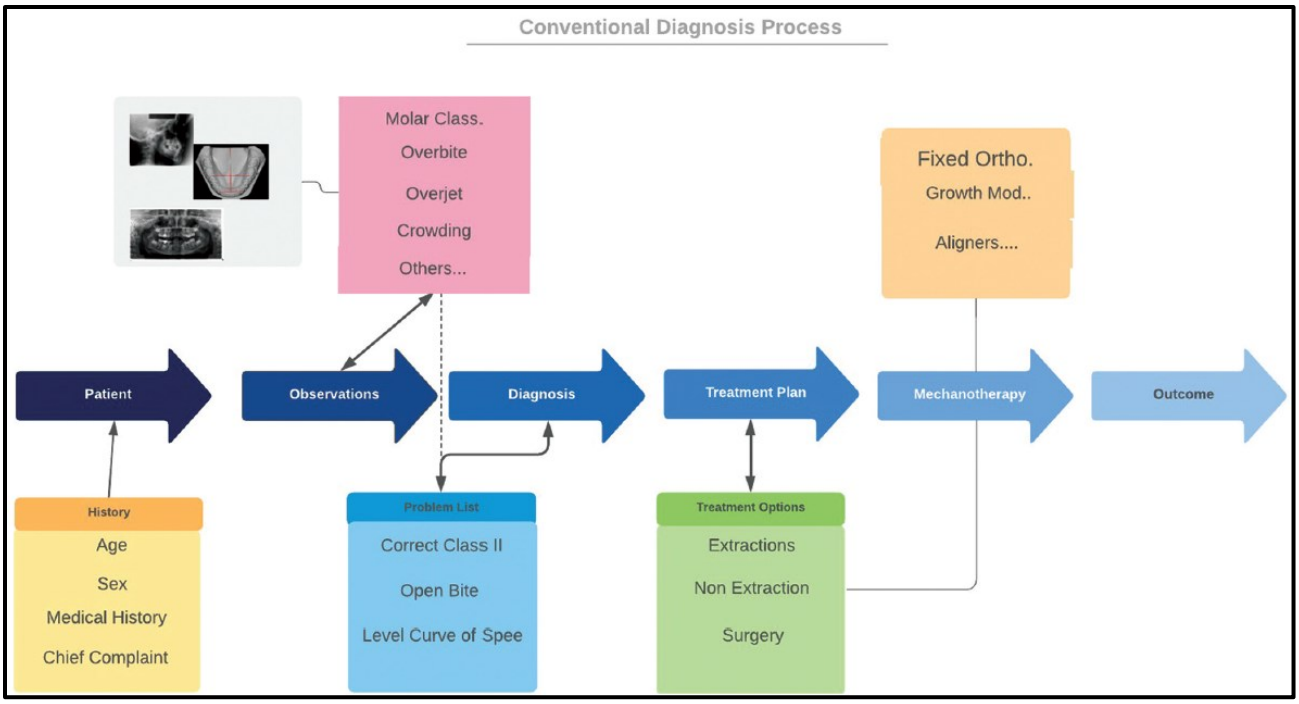


Figure 11. – Conventional orthodontic treatment planning workflow (57)

As previously mentioned, an artificial neural network (ANN) has the power of mining features from massive medical data, and in recent years it has been used to help in orthodontic treatment planning(28, 53, 58).

A 2010 study published in Angle Orthodontist attempted to use ANN modeling to help clinicians in deciding whether extractions were necessary prior to orthodontic treatment. By training a Back Propagation ANN model with 120 extraction cases and 80 non extraction cases, they used a data set of 20 patients to test the algorithm (Figure 12). They established 23 quantifiable indexes: 5 indexes derived from cast measurements, 13 from hard tissue cephalometrics and 5 from soft tissue cephalometrics. The untrained data from the 20 patients in the testing set were 80% accurate in determining whether extraction or nonextraction treatment was best for the malocclusion. Another interesting result that the algorithm yielded was that the two most important index inputs for determining extraction/nonextraction were “anterior teeth uncovered by incompetent lips” and the Incisor -mandibular plane angle -IMPA (L1-MP) , while the least important index was the Frankfort-mandibular plane angle- FMA (FH-MP)(28).

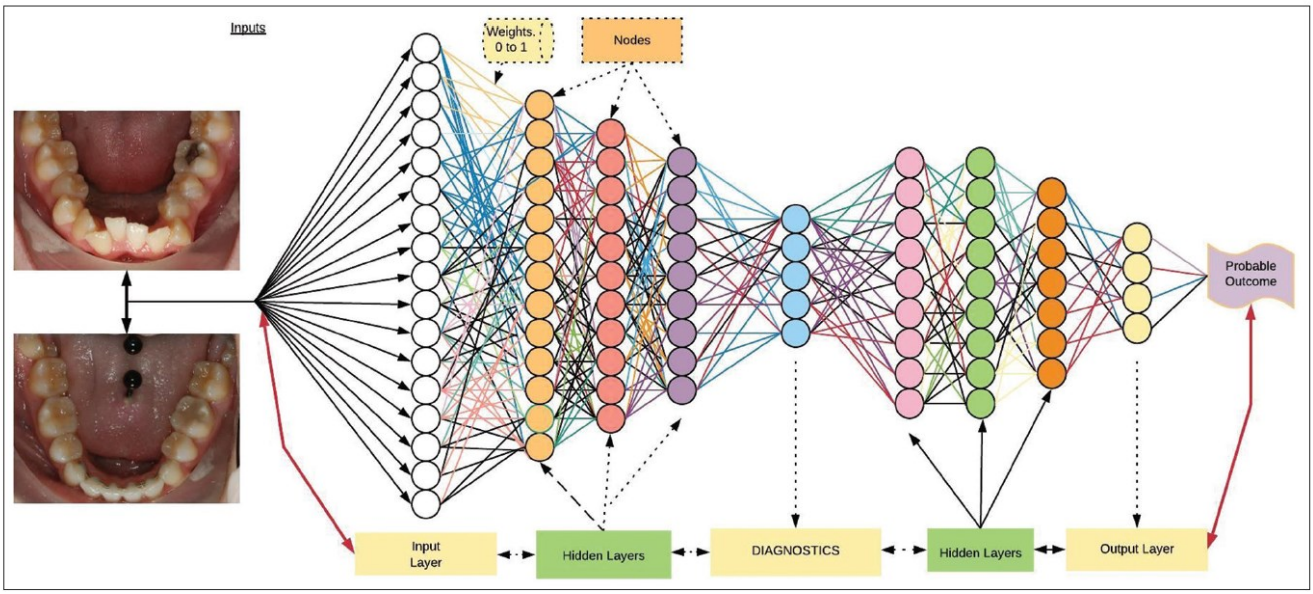


Figure 12. – Deep neural network designed to allow for learning through back propagation(57)

A Korean study published in the AJO-DO in 2016 performed a similar study to the one previously mentioned but took it one step further. They not only attempted to determine between extraction/nonextraction treatment plans, but also trained the ANN to output what the best extraction pattern would be for the given malocclusion. The output data was split into 5 groups:

1. Nonextraction
2. Extraction of upper and lower second premolars
3. Extraction of upper and lower first premolars
4. Extraction of upper first and lower second premolars
5. Extraction of upper first premolars

In the diagnosis of extraction vs nonextraction, the accuracy was 93% and 84% for the detailed diagnosis of extraction patterns. Although in a real orthodontic setting there are many more possibilities for extraction patterns, the results were quite promising considering that these are the most common patterns encountered in daily practice (58).

Another study, published in 2019, also focused on predicting the necessity of extractions and on extraction patterns. An important additional factor when determining a treatment plan is the anchorage requirement in the case of space closure, especially in cases requiring maximum anchorage where appropriate means must be taken into account early in the treatment process (8). This newer study introduced the prediction of anchorage types for closing extraction spaces using additional factors such as the patient's nasolabial angle, the relationship between upper lip and lower lip to the esthetic plane, and lip incompetence (53) (Figure 13). The predictive accuracy for the extraction/nonextraction decision reached 94%, the extraction pattern prediction reached 84.2%, and the accuracy for prediction of anchorage type was 92.8%. These results suggest that ANN's could potentially be useful tools in assisting orthodontists in making more detailed treatment plans.

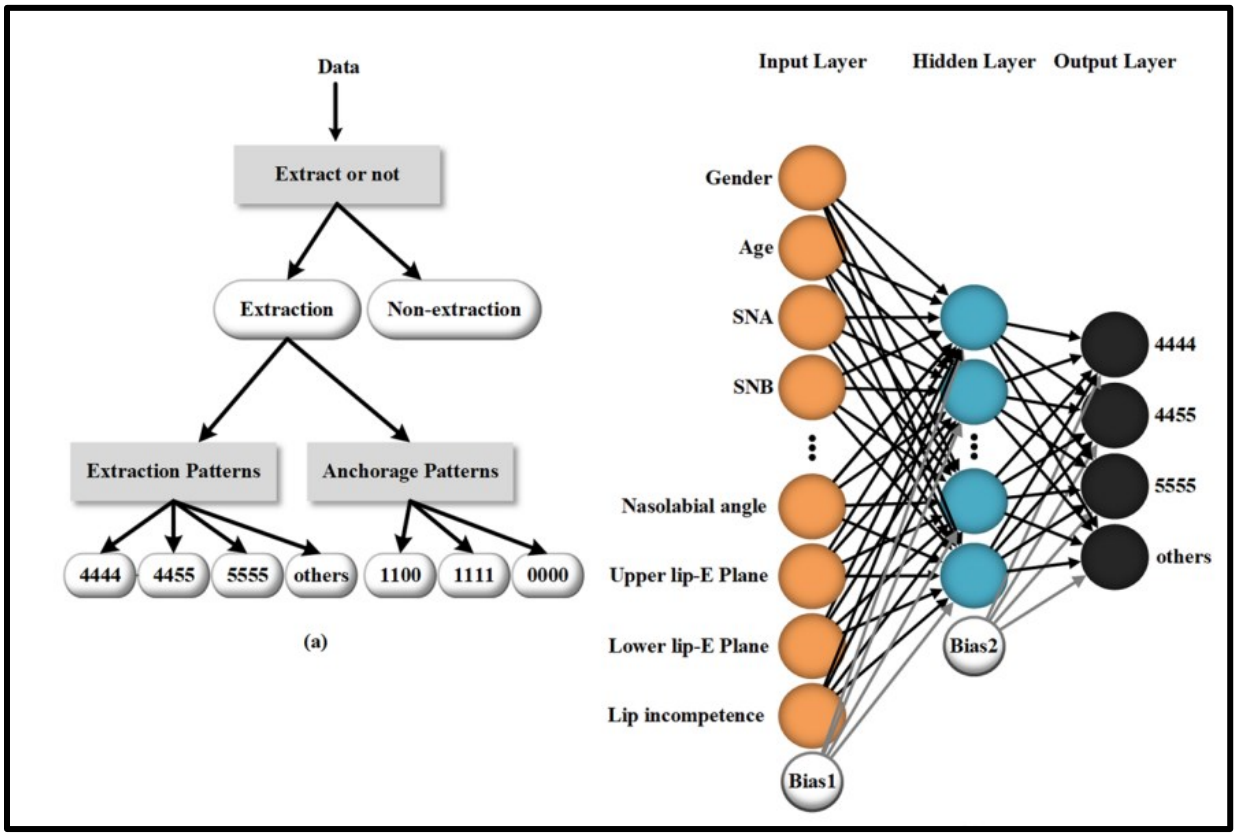


Figure 13. – Structure of the neural network to predict the need for extraction, extraction patterns and anchorage patterns. (53)

2.2.7.1 AI for Cephalometric Landmark Detection

Another interesting use of AI in the field of orthodontics is in the automatic localization of cephalometric landmarks. The diagnostic value of a given cephalometric analysis will highly rely on the clinician's accuracy and reproducibility in landmark identification (59). Errors in cephalometric analysis can be either systematic or random. A systematic error can arise when no compensation is made for the variable geometry of a cephalogram, while random error involves tracing, landmark identification and measurement errors. Studies have shown that variability in landmark identification was five times greater than measurement variability. With the use of computer aided software, measurement errors have greatly been reduced, but landmark identification remains a necessary time-consuming human task and will still involve variability and error. For these reasons there have been efforts to automate the analysis with the following goals:

1. Reducing the time required to complete the analysis
2. Improve the accuracy and reproducibility of landmark identification
3. Reduce human error and subjectivity related to landmark identification

The first attempt at developing an automated landmarking software for cephalograms was done by Cohen et al. in 1984 (60). Since then, many authors have attempted to design software using different ways that involve computer vision and artificial intelligence, and these can be classified into four categories(61):

1. knowledge-based approaches (edge detection + image-processing techniques),
2. model-based approaches,
3. soft-computing approaches (neural networks and support vector machines)
4. hybrid systems (a combination of the three previous methods). (61)

Throughout the various techniques and studies that have been presented, many did not yield results that were accurate enough for use in clinical practice (14). Different success levels in landmarking detection were reported depending on the approach that was used and the number of landmarks that the software was asked to detect The accepted accuracy standard for

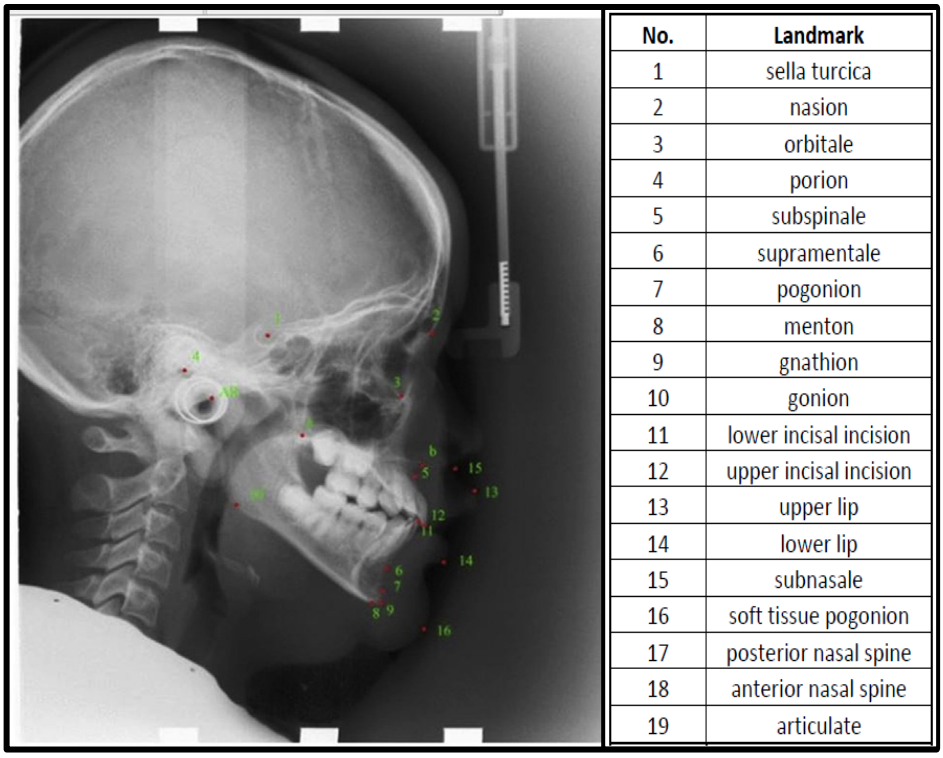
automized cephalometric landmark detection in the literature is that if the distance between the position of the manual localization ("gold standard") and the position of the automatic localization is less than 2mm (Euclidean distance) , the detection is considered accurate or "correct" (61-63).This measure is known as the successful detection rate (**SDR**). If the distance is less than 4mm, the detection is considered acceptable. Table 2 shows a summary, in chronological order, of the studies in the literature published on the subject from 1989 to today, along with their successful detection rates (SDR) of 2mm or less, mean radial error (MRE), and the number of landmarks that they were designed to detect.

The issue with basing ourselves on the success rates of each study is that only a certain number of landmarks were selected to be studied, and this is especially true for the older studies where a specific approach was only good for a certain set of landmarks. Furthermore, one cannot quickly judge the performance of a study's approach/software based only on the published MRE and SDR, since each test set of radiographs can be very different. Thanks to the 2015 International Symposium on Biomedical Imaging (ISBI) that launched challenges on cephalometry landmark detection, the best studies that were published were where all teams had to test their proposed algorithms on the same cephalometric x-ray data sets, thus eliminating the bias associated to testing on different sets. (64) The nineteen landmarks that were part of the 2015 ISBI challenge are shown in Figure 14.Can you make Figure 14 less blurry? The studies in Table 2 that used the 2015 ISBI are labeled in the "No. of x-rays tested" column. Concerning the 2-mm SDR, Song et al (64), Oh et al (65), Zhong et al (66) and Gilmour et al (67) showed the highest performance on the 2015 grand challenge in dental x-ray image dataset by reaching more than 86% of landmarks with less than 2-mm SDR. Gilmour et al (67) reported the lowest MRE of 1.01 mm.

Table 2.- Summary of the articles published on landmark detection in lateral cephalometry

Research Group	Year	No. of Landmarks	No. of X-rays tested	MRE (mm)	SDR < 2mm	Architecture/modelling framework
Parthasarathy et al (68)	1989	9	5	2.06	58%	Image filtering plus knowledge-based
Tong et al (69)	1989	17	5	0.33	76%	Image filtering plus knowledge-based
Cardillo et al (70)	1994	20	40	NR	75%	Model-based
Forsyth et al (71)	1996	19	10	NR	79%	Image filtering plus knowledge-based
Rudolph et al (72)	1998	15	14	3.07	13%	Model-based
Hutton et al (73)	2000	16	63	4.08	35%	Model-based
Liu et al (74)	2000	13	38	2.86	23%	Hybrid
Grau et al (75)	2001	17	20	1.03	88.6%	Hybrid
Yang et al (76)	2001	16	11	NR	80%	Hybrid
Innes et al (77)	2002	3	109	NR	72%	Artificial Neural Network (Soft-Computing)
Chakrabartty et al (78)	2003	8	40	NR	93%	Machine Learning (Soft Computing)
Ciesielski et al (14)	2003	4	36	NR	85%	Soft-Computing
El-Feghi et al (79)	2004	20	600	NR	90%	Artificial Neural Network (Soft Computing)
Giordano et al (14)	2005	8	26	1.07	85%	Hybrid
Yue et al (80)	2006	12	86	NR	71%	Hybrid
Ibragimov et al (81)	2014	19	250 (ISBI 2015)	1.82-1.92	71.70%	Hybrid : **Random Forests (RF)** and Game Theory
Vandaele et al (82)	2014	19	100	1.83	75.37%	Hybrid: **Random Forests** & simple pixel-based multiresolution features
Kaur and Singh (83)	2015	18	85	1.84	89.50%	Soft-Computing (Zernike Moments)
Lindner and Cootes (84)	2015	19	250 (ISBI 2015)	1.67	74.84%	Machine Learning -**Random Forests** (RF)
Ibragimov et al (85)	2015	19	150	1.84	75.40%	Hybrid : **Random Forests (RF)** and Game Theory
Lindner et al (86)	2016	19	250 (ISBI 2015)	1.2	84.70%	Machine Learning -**Random Forests** (RF)
Arik et al. (87)	2017	19	250 (ISBI 2015)	NR	75.58%	Deep Learning / Convolutional Neural Networks
Hwang et al. (59)	2019	80	283	1.46	NR	Deep Learning / Convolutional Neural Networks
Qian et al(88)	2019	19	250 (ISBI 2015)	NR	82.50%	Deep Learning / Convolutional Neural Networks
Chen et al (89)	2019	19	250 (ISBI 2015)	1.17	86.21%	Deep Learning / Convolutional Neural Networks
Zhong et al (66)	2019	19	250 (ISBI 2015)	1.14	86.74%	Deep Learning / Convolutional Neural Networks
Lee JH et al. (41)	2020	19	250 (ISBI 2015)	1.53	82.11%	Deep Learning / Convolutional Neural Networks
Song et al (64)	2020	19	250 (ISBI 2015)	1.3095	86.40%	Deep Learning / Convolutional Neural Networks
Gilmour et al (67)	2020	19	250 (ISBI 2015)	1.01	88.32%	Deep Learning / Convolutional Neural Networks
Oh et al (65)	2020	19	250 (ISBI 2015)	NR	86.20%	Deep Learning / Convolutional Neural Networks
Kim et al (90)	2020	19	250 (ISBI 2015)	1.16	83.13%	Deep Learning / Convolutional Neural Networks
Yao et al (62)	2021	19 37	250 (ISBI 2015) 100	1.14 1.04	86.84% 97.30%	Deep Learning / Convolutional Neural Networks

***NR = not reported**



No.	Landmark
1	sella turcica
2	nasion
3	orbitale
4	porion
5	subspinale
6	supramentale
7	pogonion
8	menton
9	gnathion
10	gonion
11	lower incisal incision
12	upper incisal incision
13	upper lip
14	lower lip
15	subnasale
16	soft tissue pogonion
17	posterior nasal spine
18	anterior nasal spine
19	articulate

Figure 14. – The nineteen landmarks of the ISBI 2015 Grand Challenges in Dental X-ray Image Analysis (62)

With evolving computational power and newer advanced algorithms, recent approaches have shown significant improvement in accuracy, increasing their interest for clinical use (87) (86) (62) (91). As previously mentioned, deep learning (DL) is a newer branch of machine learning, and it is only recently since 2017 that some authors have used this method in automatic cephalometric analysis (87). Its increasing popularity has led to increased research for DL in cephalometric analysis (59) (62, 63, 91).

A study in 2019 conducted an experiment with the purpose of comparing two of the latest deep learning algorithms (YOLOv3 and SSD) for automatic identification of landmarks on lateral cephalograms in order to compare their accuracies in detecting 80 landmarks (Figure 15) (59). Both algorithms yielded promising results, with YOLOv3 yielding more accurate detections at higher computational speeds (0.05 seconds versus 2.89 seconds per cephalogram).

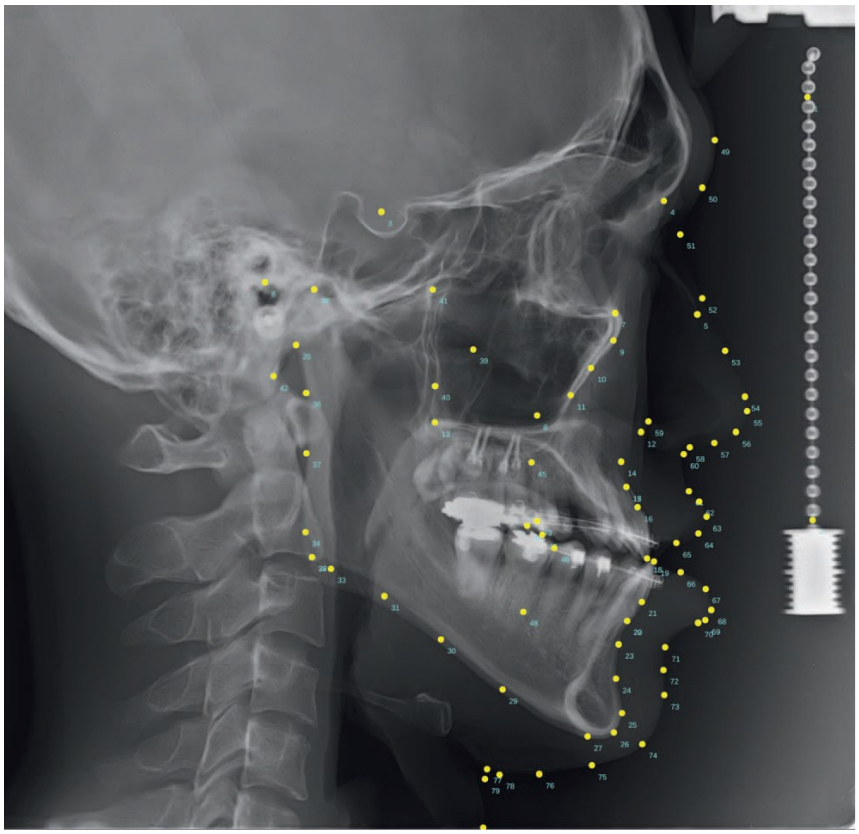


Figure 15. – Cephalogram indicating the 80 cephalometric landmarks detected in the study (59)

In this study, 80.4% of the landmarks were less than 2mm (SDR < 2mm) away from the manual reference, while 96.2% of the landmarks were less than 4mm away from the gold standard. These results showed a 5% increase in detection since the first time DL algorithms were used for detection of cephalometric landmarks only two years prior (87).

After realizing that the first study showing that DL algorithms yielded excellent results, the same team conducted a second study with one question in mind: “Might it be better than human?”. In order to answer this question, they compared the AI to a human examiner with 28 years of clinical orthodontic practice experience, who was considered the “gold standard”. (63) They then compared the human examiner to a second human examiner (3rd year orthodontic resident) to determine whether the differences between AI and human examiners would be smaller than those between both human examiners. They also wanted to test the reproducibility of the AI versus the reproducibility of the human examiners. Their results showed that out of 46 skeletal landmarks, the AI had better accuracy for 14/46 landmarks, the human had better accuracy for

14/46, and the remaining 18 did not show statistically significant differences, and the results were very similar for soft tissues. The mean detection error between AI and human was 1.46 ± 2.97mm, while the difference between human examiners was 1.50 ± 1.48mm. They noticed that the DL software behaved similarly to human examiners in the sense that when the human had difficulties in identifying landmarks on poor quality images, so did the AI. AI always detected identical landmark positions which can imply that AI might be a more reliable option for repeatedly identifying multiple cephalometric landmarks.

3 – Research Article

3.1 Abstract

Introduction: The objective of this study was to compare the difference between manual cephalometric analysis and automatic analysis by artificial intelligence to confirm the reliability of the latter. Our research hypothesis is that the manual technique is the most reliable of the methods and is still considered the gold standard.

Method: A total of 100 lateral cephalometric radiographs were collected in this study. Manual technique (MT) and automatic localization by artificial intelligence (AI) tracings were performed for all radiographs. The localization of 29 commonly used landmarks were compared between both groups. Mean radial error (MRE) and a successful detection rate (SDR) of 2mm were used to compare both groups. AudaxCeph software version 6.2.57.4225 (Audax d.o.o., Ljubljana, Slovenia) was used for both manual and AI analysis.

Results: The MRE and SDR for the inter-examiner reliability test were 0.87 ± 0.61mm and 95% respectively. For the comparison between the manual technique MT and landmarking with artificial intelligence AI, the MRE and SDR for all landmarks were 1.48 ± 1.42mm and 78% respectively. When dental landmarks are excluded, the MRE decreases to 1.33 ± 1.39mm and the SDR increases to 84%. When only hard tissue landmarks are included (excluding soft tissue and dental points) the MRE decreases further to 1.25 ± 1.09mm and the SDR increases to 85%. When only soft tissue landmarks are included the MRE increases to 1.68 ± 1.89mm and the SDR decreases to 78%.

Conclusion: The software performed similarly to what was previously reported in the literature for software that use analogous modeling framework. Comparing the software's landmarking to manual landmarking our results reveal that the manual landmarking resulted in higher accuracy. The software operated very well for hard tissue points, but its accuracy went down for soft and dental tissue. Our conclusion is that this technology shows great promise for application in clinical settings under the doctor's supervision.

Keywords: Automated identification; Cephalometric analysis; Cephalometric landmarks; Artificial intelligence; Machine learning; Deep learning

3.2 Introduction

Since the beginning of cephalometric radiography in the 1930s, cephalometric analysis has always been an important tool in diagnosis, treatment planning and growth evaluation, as well as a method to quantify the effects of orthodontic and surgical treatments. Manual cephalometric tracing is a meticulous and time-consuming task. Although digital tracing software can automatically calculate cephalometric measurements and angles, time is required for manual localization and positioning of cephalometric landmarks. In addition, errors in cephalometric point location, operator experience, and the subjective nature of landmarking also pose problems with this approach. To overcome these shortcomings, many fully automated approaches based on the use of artificial intelligence are now commercially available to automatically position cephalometric points and perform the analysis.

The introduction of artificial intelligence in healthcare is revolutionizing and pushing the industry towards advancements in many clinical specialties(23). In the field of dentistry, Convolutional Neural Networks (CNNs) have shown very promising results in diagnosis and prediction in both radiology and pathology through the use of disease identification and image segmentation (20). Furthermore, deep learning algorithms are being implemented to assist in the clinical decision-making process and even in treatment planning in the field of orthodontics (28) (29). Random Forest regression-voting (RFRV) is one of the most popular and commonly used machine learning algorithms across real-life data science projects as well as data science competitions (92). Several studies trained and tested the application of RFRV to automatically detect cephalometric landmarks (82) (83) (85) (86) (87), and their results can be found in Table 3.

Table 3- Studies that used RF regression-voting for cephalometric landmarking

Research Group	Year	No. of Landmarks	No. of X-rays tested	MRE (mm)	SDR < 2mm	Architecture/modelling framework
Ibragimov et al (81)	2014	19	250 (ISBI 2015)	1.82-1.92	71.70%	Hybrid : **Random Forests (RF)** and Game Theory
Vandaele et al (82)	2014	19	100	1.83	75.37%	Hybrid: **Random Forests** & simple pixel-based multiresolution features
Lindner and Cootes (84)	2015	19	250 (ISBI 2015)	1.67	74.84%	Machine Learning -**Random Forests** (RF)
Ibragimov et al (85)	2015	19	150	1.84	75.40%	Hybrid : **Random Forests (RF)** and Game Theory
Lindner et al (86)	2016	19	250 (ISBI 2015)	1.2	84.70%	Machine Learning -**Random Forests** (RF)

The latest study using RFRV by Lindner et al (87) obtained an MRE of 1.2mm and 2mm SDR of 84.70%, which was a significant improvement from the previously seen SDR's. Recent studies introduced deep learning algorithms to improve the landmark detection(62, 64, 87-90). Yao et al in their 2021 study tested their DL / CNN algorithm on 37 landmarks and obtained an MRE of 1.04mm and 97.30% SDR which are numbers that come close to human error (62). The purpose of this study is to compare manual analysis and automatic analysis by artificial intelligence to confirm the reliability of a commercially available AI driven software that uses RFRV. We will then compare our results with newer studies that use more advanced deep learning algorithms. Our research hypothesis is that the manual technique is the most reliable of the methods.

3.3 Methodology

The Ethics Committee for Clinical Research of the Université de Montréal assessed and approved this study (CERC Projet # 2022-1334 , Appendix 1).

This is a reliability study conducted at the University of Montreal's Faculty of Dentistry's post-graduate orthodontic clinic. One hundred (100) digital lateral cephalometric radiographs belonging to patients treated at the University clinic were randomly selected and anonymized to remove all patient related information. The sample was drawn from the pre-treatment cephalograms of patients taken at the orthodontic clinic between the years 2000 and 2020. Gender, age, racial group, dental occlusion, skeletal class, or stage of dentition were not

considered. Exclusion criteria were the presence of a syndrome that could alter craniofacial development, rigid post-surgical fixations, and obvious malposition of the head in the cephalostat. All radiographs were taken with the Instrumentarium Orthoceph OC200D (KaVo, Biberach, Germany). A maximum of 85 kVp was used to produce digital images (computed radiography) of the cephalometric radiograph. The size of the cephalograms was 2304 by 2832 pixels. The images were grayscale with 8 bits per pixel and the resolution was 302 DPI. In this study, a total of twenty-nine (29) commonly used cephalometric landmarks were chosen to be analyzed; thirteen (13) hard tissue points, eight (8) dental points, and 8 soft tissue points. The landmarks along with their abbreviations and definitions are found in Table 4.

Table 4. - Definition of landmarks

Landmark		Definition
Hard Tissue Points		
1	A, Subspinale	Deepest point on the maxilla below ANS
2	ANS, Anterior nasal spine	Anterior point of maxilla.
3	Ar, Articulare	Point on the posterior border of the ramus at the intersection with the basilar portion of the occipital bone
4	B, Supramentale	Most posterior point on the bony curve of the mandible above pogonion
5	Go, Gonion	Most posterior and inferior point on the outline of the angle of the mandible
6	Me, Menton	Lowest point on the symphysis of the mandible
7	N, Nasion	Junction of frontonasal suture
8	Or, Orbitale	Inferior border of orbit
9	Pg, Pogonion	Most anterior point of bony chin
10	PNS, Posterior Nasal Spine	Posterior point of bony hard palate.
11	Po, Porion	Top of external auditory meatus.
12	PTM, Pterygomaxillary Fissure	Most posterior and superior point on the outline of the pterygomaxillary fissure
13	S, Sella	Mid-point of sella turcica
Soft Tissue Points		
14	G', Soft tissue Glabelle	Most prominent point in the mid sagittal plane of the forehead
15	LL, Labius Inferius	Most anterior point on outline of lower lip
16	Pg', Soft tissue Pogonion	Most anterior point on outline of soft tissue chin
17	Pn, Pronasale	Anterior tip of the nose
18	Sn, Subnasale	Junction of nasal septum and upper lip in mid-sagittal plane.
19	STM, Stomium Superior	Lowest midline point on outline of upper lip.
20	St-, Stomium Inferior	Highest midline point on outline of lower lip.
21	UL, Labius Superius	Most anterior point on outline of upper lip
Dental Points		
22	+1a, Apex of upper incisor	Tip of the apex of the upper incisor
23	-1a, Apex of lower incisor	Tip of the apex of the lower incisor
24	+1i, Incisal edge of upper incisor	Tip of the incisal edge of the upper incisor
25	-1i, Incisal edge of lower incisor	Tip of the incisal edge of the lower incisor
26	+6a, Upper molar mesial apex	Mesial Apex of the first upper molar
27	-6a, Lower molar mesial apex	Mesial apex of the first lower molar
28	+6c, Upper 1st molar cusp	Tip of the mesiobuccal cusp of the first upper molar
29	-6c, Lower 1st molar cusp	Tip of the mesiobuccal cusp of the first lower molar

The manual tracing technique (MT) used in this study was a computer-assisted technique where the operator manually selects the cephalometric points on the monitor and allows the software to generate all measurements. In this method, the test set of 100 radiographs were analyzed by the same operator by positioning the cephalometric points manually using AudaxCeph version 6.2.57.4225 software (Audax d.o.o., Ljubljana, Slovenia).

To assess the intra-examiner reliability of the MT technique, 25 digital radiographs randomly selected were retraced one month later by the main author. These same 25 radiographs were traced by a second experienced examiner to determine the inter-examiner reliability.

The artificial intelligence (AI) technique consisted of fully automatic software generated positioning and tracing. With the click of a button, the algorithm automatically positioned the cephalometric landmarks and generated the cephalometric measurements. The same AudaxCeph software offers this feature and was used to analyze the test set of 100 radiographs. AudaxCeph uses Hough Forest approach to detect the structure of interest in the image, and then applies Random Forest Regression-Voting in the Constrained Local Model framework to locally refine all point positions.

Using this software, the landmarks of interest were selected and converted into pixel coordinates (x, y) that were subsequently recorded on spreadsheets. These coordinates were then converted to millimeters on the spreadsheet using the millimetric reference scale contained on each digital x-ray. The same radiograph was analyzed five times by Image J to assess the operator's accuracy in selecting cephalometric points.

Statistical analyses: Two indexes were used to compare both methods and for the intra-examiner and inter-examiner reliability. The first one was the mean radial error (MRE). The radial error R was calculated as follows:

$$R = \sqrt{\Delta x^2 + \Delta y^2}$$

Where Δx and Δy are the absolute distances in the x-direction and y-direction between the predicted and referenced landmarks, respectively. MRE and standard deviation were calculated as follows:

$$MRE = \frac{\sum_{i=1}^{N} R_i}{N}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{N} (R_i - MRE)^2}{N-1}}$$

For each landmark, if the distance between the predicted and standard position was higher than a certain value d, the automatic localization was successful. The second index used was the successful detection rate (SDR) within the range of 2mm.

$$2-mm\ SDR = \frac{\#\{j : \| L_d(j) - L_r(j) \| < 2\}}{\#\Omega} \times 100\%$$

where L_d and L_r are the location of the predicted and referenced landmark, respectively. Ω is the number of detections made and j ∈ Ω (where j is an element of the set Ω). Additionally, the Dahlberg and the Bland-Altman methods were used.

3.4 Results

The intra-examiner and inter-examiner results for 25 randomly selected x-rays (725 total landmarks) are shown in Table and Table 6. The MRE and SDR for the intra-examinator reliability test were 0.59 ± 0.44mm and 99% respectively. The MRE and SDR for the inter-examinator reliability test were 0.87 ± 0.61mm and 95% respectively.

Table 5. - Intra-examiner results for 25 x-rays retraced by the same examiner

	MRE (mm)	SD	SDR (<2mm)	nb
ΔX (mm)	0.02	0.47	99%	725
ΔX (pixel)	0.11	3.18		725
ΔY (mm)	-0.03	0.56	100%	725
ΔY (pixel)	-0.17	3.79		725
D (mm)	0.59	0.44	99%	725
D (pixel)	3.96	2.97		725

Table 6.- Inter-examiner results for 25 x-rays traced by a second examiner

	MRE (mm)	SD	SDR (<2mm)	nb
ΔX (mm)	0.00	0.66	99%	725
ΔX (pixel)	-0.02	4.42		725
ΔY (mm)	-0.15	0.82	97%	725
ΔY (pixel)	-0.98	5.49		725
D (mm)	0.87	0.61	95%	725
D (pixel)	5.84	4.06		725

One of the x-rays had to be removed from the test set because it did not contain a millimetric scale, leaving ninety-nine total x-rays in the test set, for a total of 2871 landmarks tested. For the comparison between the manual technique MT and landmarking with artificial intelligence AI, the MRE and SDR were 1.48 ± 1.42mm and 78% respectively. The results, including broken down measurements on the horizontal (x) and (y) axis are shown in Table 7

Table 7 - Results for all x-rays in test set (99) – comparison between MT and AI

	MRE (mm)	SD	SDR (<2mm)	nb
ΔX (mm)	0.57	1.33	87%	2871
ΔX (pixel)	3.86	8.94		2871
ΔY (mm)	-0.02	1.46	91%	2871
ΔY (pixel)	-0.12	9.83		2871
D (mm)	1.48	1.42	78%	2871
D (pixel)	10.00	9.56		2871

0represent column charts for all landmarks where the y-axis corresponds to the mean error between manual and automatic techniques. Figures 16 A, B and C correspond to mean errors in the horizontal axis, vertical axis, and mean radial error (Euclidean distance), respectively, and Table 6 shows the detailed values for MRE and SDR. To better visualize the strengths (and weaknesses) of the AI algorithm, certain groups of landmarks were excluded and isolated from the analysis. The first subset analysis was done by excluding all molar related landmarks: 6c, +6a, -6c, -6a. The second and third subset analysis were done by including only the hard tissue points

and lastly only the soft tissue points from Table 8. - The MRE and SDR of landmarks in our test set These results can be found in Table 7, Table 8 and Table 9.

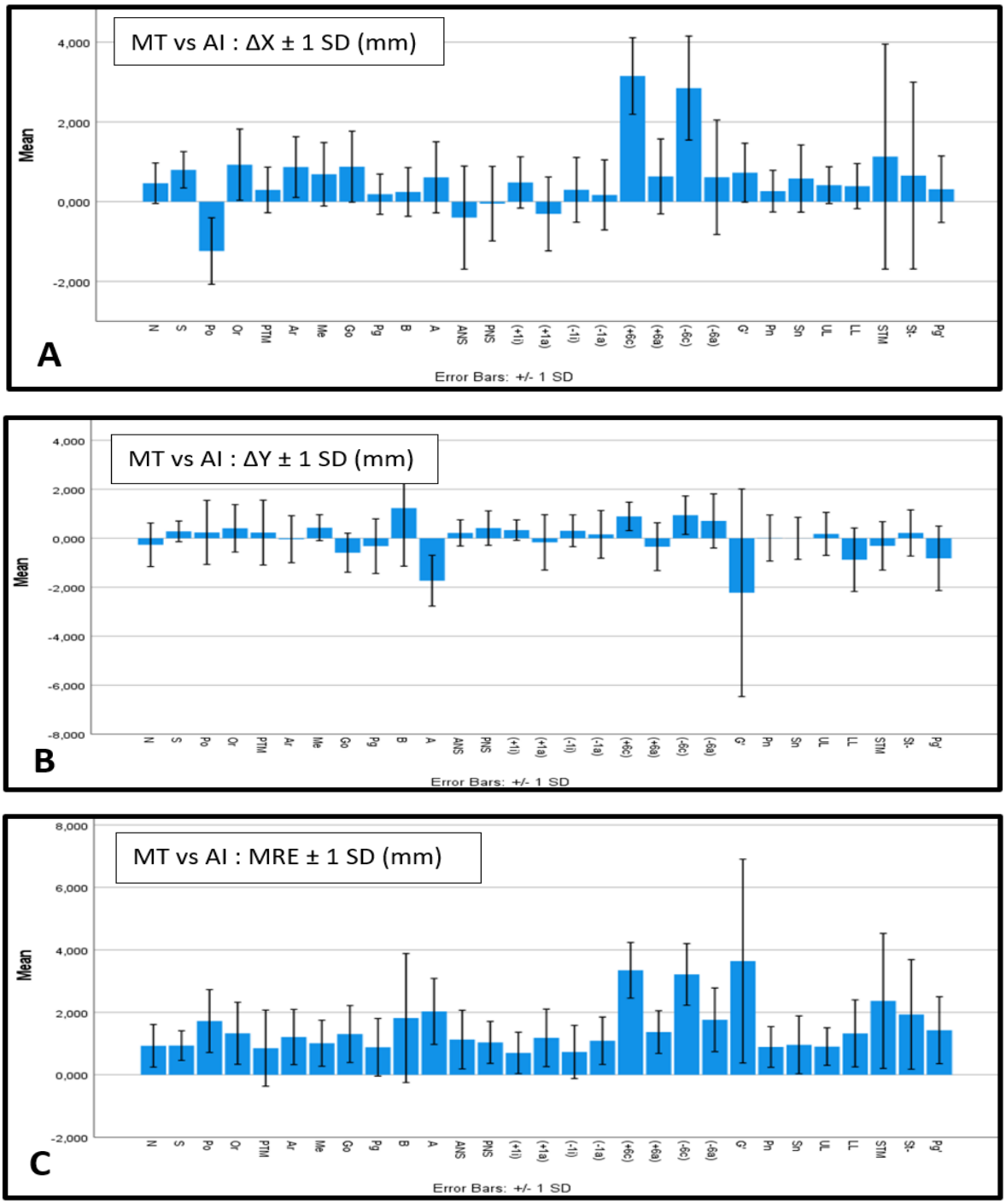


Figure 16. Column charts with A) mean x-axis error measurements for all landmarks, B) mean y-axis error for all landmarks C) MRE for all landmarks. Error bars represent ±1 SD.

Table 8. - The MRE and SDR of landmarks in our test set

Landmark	MRE (mm)	SD	SDR (<2mm)
Hard tissue points			
N	0.93	0.68	90.91%
S	0.94	0.48	97.98%
Po	1.72	1.01	71.72%
Or	1.33	0.99	83.84%
PTM	0.85	1.22	86.87%
Ar	1.21	0.88	88.89%
Me	1.01	0.74	92.93%
Go	1.31	0.91	82.83%
Pg	0.88	0.92	94.95%
B	1.82	2.07	71.72%
A	2.03	1.06	57.58%
ANS	1.13	0.94	84.85%
PNS	1.04	0.67	93.94%
Dental Points			
(+1i)	0.70	0.66	95.96%
(+1a)	1.18	0.92	87.88%
(-1i)	0.73	0.85	94.95%
(-1a)	1.09	0.76	90.91%
(+6c)	3.35	0.89	4.04%
(+6a)	1.37	0.68	85.86%
(-6c)	3.21	0.99	6.06%
(-6a)	1.76	1.02	73.74%
Soft Tissue Points			
G’	3.64	3.26	40.40%
Pn	0.89	0.65	95.96%
Sn	0.96	0.93	90.91%
UL	0.90	0.60	96.97%
LL	1.33	1.07	79.80%
STM	2.37	2.16	65.66%
St-	1.93	1.75	71.72%
Pg’	1.43	1.07	78.79%

Table 9. - Results for all x-rays – excluding landmarks associated with molars (+6c, +6a, -6c, -6a)

	MRE (mm)	SD	SDR (<2mm)	nb
ΔX (mm)	0.38	1.15	93%	2475
ΔX (pixel)	2.52	7.70		2475
ΔY (mm)	-0.11	1.50	90%	2475
ΔY (pixel)	-0.73	10.09		2475
D (mm)	1.33	1.39	84%	2475
D (pixel)	8.99	9.34		2475

Table 10.- Results for all x-rays – hard tissue points only

	MRE (mm)	SD	SDR (<2mm)	nb
ΔX (mm)	0.33	0.99	94%	1287
ΔX (pixel)	2.21	6.66		1287
ΔY (mm)	0.04	1.29	92%	1287
ΔY (pixel)	0.26	8.66		1287
D (mm)	1.25	1.09	85%	1287
D (pixel)	8.40	7.34		1287

Table 11. - Results for all x-rays – soft tissue points only

	MRE (mm)	SD	SDR (<2mm)	nb
ΔX (mm)	0.56	1.44	90%	792
ΔX (pixel)	3.76	9.68		792
ΔY (mm)	-0.48	1.94	86%	792
ΔY (pixel)	-3.24	13.05		792
D (mm)	1.68	1.89	78%	792
D (pixel)	11.32	12.66		792

When dental landmarks are excluded from the mean averages, the MRE decreases to 1.33 ± 1.39mm and the SDR increases to 84%. When only hard tissue landmarks are included (excluding soft tissue and dental points), the MRE decreases further to 1.25 ± 1.09mm and the SDR increases to 85%. When only soft tissue landmarks are included (excluding hard tissue and dental points), the MRE increases to 1.68 ± 1.89mm and the SDR decreases to 78%.

3.5 Discussion

The present study was done to evaluate whether AI could be a reliable option to replace the repetitive and time-consuming task of identifying cephalometric landmarks on a two-dimensional lateral x-ray. The software was tested on 99 x-rays and 29 landmarks. The global MRE and 2 mm SDR were 1.48 ± 1.42mm and 78%, which means that the average difference between the manual landmarking and automatic landmarking was 1.48mm and that 78% of landmarks detected by AI were less than 2mm away from the manually placed landmarks. AudaxCeph uses Hough Forest approach to detect the structure of interest in the image, and then applies Random Forest Regression-Voting in the Constrained Local Model framework to locally refine all point positions. Several studies developed and tested similar modeling frameworks (81) (83) (85) (86) (87). The latest study using Random Forest Regression-Voting by Lindner et al (87) obtained an MRE of 1.2mm and 2mm SDR of 84.70% (Table 3- Studies that used RF regression-voting for cephalometric landmarking). When compared to our results, Lindner et al's MRE is lower by 0.28mm and SDR higher by 6.7%. These differences can be explained by the fact that they did not include any molar landmarks in their study, which because of super-impositions on 2D cephalograms, make it difficult for both humans and AI to accurately predict their positions. In order to account for this, we calculated the MRE and SDR of all points while excluding the ones associated with the upper and lower molars (Table 9). In this analysis, we obtained improved MRE and SDR of 1.33 ± 1.39mm and 84%, which are values that compare well to the ones from Lindner et al.

When the results are isolated for hard tissue points and soft tissue points, the software has a net advantage for identifying hard tissue. The MRE and SDR for hard tissue points are 1.25 ± 1.09mm and 85%, while for soft tissue points the performance drops to an MRE of 1.68 ± 1.89mm and SDR decreases to 78%. These differences may be attributed to the contrast differences between the soft and hard tissues on lateral cephalograms, making it easier for the algorithm to identify the hard tissues.

Cephalometric landmarks seem to have slightly different reliability in the horizontal (x) versus the vertical (y) planes, which indicates that the distribution of error is asymmetric on lateral

cephalograms. Studies have shown that differences along the x- axis tend to be greater than those on the y-axis(10, 11). These findings seem to be in accordance with the results that were obtained in this study, as the SDR for ΔX errors was 87% while the ΔY errors was 91%.

For the hard tissue points, those having the largest MRE and SDR were Po (1.72 ± 1.01 & 71.72%), A (2.03 ± 1.06 & 57.58%) and B (1.82 ± 2.07 & 71.72%). The error for Po can be mainly attributed to an x-axis error as seen in 0while the errors for the A and B points can mainly be attributed to errors on the y-axis error as also seen in 0Literature and experience has shown that Po is among the most difficult landmarks to distinguish due to the overlapping structures superimposed on the landmarks (10).

As it was previously reported, landmarks attributed to molars, and more particularly to upper and lower molar cusp tips (+6c, -6c) had the highest errors among the dental landmarks, with MRE's ranging from 3.21-3.35mm and SDR of 4.04%-6.06%. These types of errors may be attributed to increased superimposition between the left and right molars causing the identification of the wrong cusp tips, since the premolars and second molars are within the landmarking area. Landmarks located on molars can be used for quick visualization of the molar's Angle classification on the tracing or for measuring treatment effect on molar position. For cephalometric analysis, the most important use of molar landmarks is in the determination of the occlusal plane, a measure used in many important analyses such as the Wits and Sassouni (8). The original occlusal plane (OOP), as defined by Downs, is a line connecting the point bisecting the first molar cusp height and the point bisecting the incisal overbite (4). The major impact that an error on these points could have would be if the errors were done in the y-axis, affecting the angulation or orientation of the occlusal plane angle. Luckily for us, the major error detected with regards to the molars is seen in the x-axis as seen on Figures 16A and 16B, and the ΔY error remained below the 2mm SDR. The x-axis error is in the positive or right-side direction, meaning that the software had the tendency to landmark cusps associated to premolars rather than molars. Yao et al's recent 2021 study trained their CNN algorithm to identify 37 landmarks including the upper and lower molar cusp tips (62), and obtained an MRE/SDR(2mm) of 1.060mm/90% for the upper molar cusp tip and 1.11mm/88% for the lower molar cusp tip which is a significant difference from what was found in the current study.

As previously mentioned, the software's landmarking was not as accurate for the soft tissue points. When compared to hard tissue points, the MRE for the soft tissue analysis was on average 0.43mm less accurate, and the SDR was 7% smaller. The two highest sources of error were G' and STM. Studies have shown that glabella is also a common source of cephalometric error mainly due to image quality (10). When the G' error is broken down, Figure 16A,B demonstrates that the error is predominantly a Y-axis error, and the direction is in the negative or downward direction, meaning that the software had a tendency to place the G' landmark closer to soft tissue nasion and the supra-orbital rim rather than looking for the most prominent point in the mid sagittal plane of the forehead. The error associated to STM soft tissue point was mainly associated with x-axis error, where the software had more difficulties pinpointing the correct location of the lowest outline of the upper lip. Surprisingly, the software was able to detect well when the patient's lips were in contact or not and demonstrated acceptable accuracy on the y-axis.

The best way to evaluate whether AI could potentially be more accurate than humans to detect landmarks would be to compare human-human error with human-AI error, similarly to what was done in the study conducted by Hwang et al (63). In this study, the main investigator was a third-year orthodontic resident, and the second investigator that examined the x-rays for the inter-examiner reliability was an experienced orthodontist with 15 years of clinical experience. The inter-examiner results in Table show that between the human examiners the MRE was 0.87 ± 0.61mm and SDR was 95%. When comparing these numbers to the global MRE & SDR in Table one can easily argue that manual localization remains the gold standard. Generally speaking, the literature review has shown that since the introduction of deep learning in automatic landmarking in 2017 (87), there has been a trend to improvement in the MREs and SDRs (Table 2.- Summary of the articles published on landmark detection in lateral cephalometry. If we attempt to compare the results obtained by Yao et al in their 2021 study (62) with our results, we see that their software which was tested on their test set that included 37 landmarks, obtained an MRE of 1.04mm and 97.30% SDR, while our human-human MRE and SDR were 0.87 ± 0.61mm and 95% on a smaller number of landmarks, meaning that we performed better for the MRE index but their AI software performed better for the SDR. Since the reproducibility of landmark positions is always constant with the AI software, one can argue that the SDR is a more

accurate measure of accuracy and reliability for AI imaging algorithms (63). From that perspective, the SDR obtained from Yao et al. did yield a better result on a larger number of landmarks and on the larger number of x-rays. As previously mentioned, it is impossible to truly compare two studies that used different test sets as too many factors come into play with regards to data sets. To argue this point further, when Yao et al. tested their software on the 250 x-rays from the 2015 International Symposium on Biomedical Imaging (ISBI), their MRE and SDR dropped to 1.14mm and 86.84%.

From a clinical perspective, our opinion is that currently, AI could never replace a trained orthodontic specialist's analysis. Most orthodontists understand that the values obtained from the analysis should not be considered as absolute values. These averages provide only an indication to help the orthodontist characterize the patient's facial morphology and the standard deviations give an idea of the severity of the deviation (7). When these concepts are understood, supplementing a patient's treatment planning with AI software should not be a concern, especially when the more recent deep learning software will be commercially available. AI software could therefore be compatible with a clinical environment under the constant supervision of an expert, since even the most advanced of software seem to incorporate a certain error.

The initial question regarding the general reliability of AI for automatic landmarking could be answered from different perspectives. When using our data set and results, the conclusion would be that the software AudaxCeph is accurate enough to yield landmarking localization within 2 mm for 78% of the 29 landmarks tested in our study. The way this result can be translated to a clinical standpoint would be to allow the software to generate the tracing, then manually replace the following points: A, B, +6c, -6c, G', Po and STM. This implies a few clinical advantages for a busy orthodontist. Firstly, the time-consuming task of manually tracing or landmarking all the initial, progress or final cephalometric x-rays can be cut down to a task necessitating only a few seconds and minor concentration. Although it was not a measure that was recorded in this study, getting the software to generate the automatic landmarking took between 5-10 seconds. In a clinical setting, we would then add a few seconds to manually relocate a few outlying landmarks. Yao et al reported that their software took 3 seconds to locate 37 landmarks (62). CPU

performance would play an important role in the speed of the analysis, therefore this number will vary depending on the amount of working data that the software necessitates and on the specifications of the computer that it is being run on. Many tasks are typically delegated from the orthodontist to a licensed dental hygienist in North-American orthodontic clinics, and these tasks can vary from taking impressions, fitting or cementing bands and appliances, indirect bonding set-ups, taking photographs and radiographs (93). Manual cephalometric landmarking and analysis is a skill that is taught in orthodontic residency and thus can not easily be delegated. Therefore, a second clinical advantage can be found in the delegation of the automated cephalometric landmarking task. After the hygienist has taken the radiograph, the automatic landmarking can be quickly performed. As with all other delegated tasks, the orthodontist's duty would then be to double check that the x-ray was well taken and fine tune the landmark positions if judged necessary. Finally, the fact that the landmarking is less time-consuming and can now be delegated, the orthodontist could be more likely to take more lateral x-rays for treatment progress analysis or superimposition on final x-rays to understand the true outcome of the treatment.

One major strength with our study was that we third-party tested an AI driven software with no financial benefit, leading to unbiased results of the software's performance. All other studies generally involve training and testing their own data sets, which could inevitably involve some level of performance bias. One obvious limitation of the current study was that only one AI software was used on a single test set coming from the same cephalometric x-ray machine, meaning that most of our generalized conclusions were drawn from a single source, when there is other software readily available on the market. To better understand the validity and reliability of AI for cephalometric landmarking, one should investigate different software on different test sets that come from different cephalometric x-ray machines to understand and test the impact of different image qualities on landmark detection, and to compare the different strengths and weaknesses of various software.

3.6 Conclusions

We tested a commercially available AI based automatic landmark localization software on a data set composed of 99 lateral x-rays with 29 landmarks per image. The software performed similarly to what was previously reported in the literature for software that use analogous modeling framework. Comparing the software's landmarking to manual landmarking, and when considering the accuracy of our intra- and inter-observer evaluations, our results reveal that the manual landmarking resulted in higher accuracy. The software performed very well for hard tissue points, but its accuracy went down for soft tissue and more so for dental landmarks. Nonetheless, this technology shows great promise for application in clinical settings to automatically conduct cephalometric analyses and can be a useful tool for better time management, easy task delegation and to facilitate the acquisition of detailed cephalometric analyses, as long as the orthodontist supervises and fine tunes the results.

3.7 Avenues for further research

As was discussed in the limitations of our current study, further third-party testing of automatic cephalometric x-ray software should be done to test their reliability, using a larger number of images in the test sets along with a higher number of landmarks. Additionally, with 3-dimensional CBCT x-rays gaining popularity (94-96), more research should be conducted on the use of AI for landmarking 3D images and performing 3D growth predictions, surgical simulations, and treatment outcomes.

3.8 Funding

No funding source was required. All x-rays, software and equipment were provided by the orthodontic clinic at the University of Montreal.

Bibliography

1. Hans MG, Palomo JM, Valiathan M. History of imaging in orthodontics from Broadbent to cone-beam computed tomography. Am J Orthod Dentofacial Orthop. 2015;148(6):914-21.
2. Behrents RG, Broadbent BH A Chronological Account of the Bolton-Brush Growth Studies 1984.
3. Hans MG, Broadbent BH, Jr., Nelson SS. The Broadbent-Bolton Growth Study--past, present, and future. Am J Orthod Dentofacial Orthop. 1994;105(6):598-603.
4. Downs WB. Variations in facial relationships; their significance in treatment and prognosis. Am J Orthod. 1948;34(10):812-40.
5. Steiner CC. Cephalometrics for you and me. Am J Orthod. 1953;39(10):729-55.
6. Ricketts RM, Gugino C , Hilgers J, Schulhof R. Visual treatment objective or V.T.O. Bioprogressive therapy. Rocky Mountain Orthodontics. Denver, Colo (1979):pp. 35-54.
7. Nielsen IL. L'analyse morphologique céphalométrique : que peut-elle nous enseigner ? Int Orthod. 2011;9(3):316-24.
8. Proffit WR, Fields HW, Larson B, Sarver DM, Contemporary Orthodontics , 6th Edition: Elsevier; 2019.
9. Dinesh A, Mutalik S, Feldman J, Tadinada A. Value-addition of lateral cephalometric radiographs in orthodontic diagnosis and treatment planning. The Angle Orthodontist. 2020;90.
10. Durão APR, Morosolli A, Pittayapat P, Bolstad N, Ferreira AP, Jacobs R. Cephalometric landmark variability among orthodontists and dentomaxillofacial radiologists: a comparative study. Imaging Sci Dent. 2015;45(4):213-20.
11. Chien PC, Parks ET, Eraso F, Hartsfield JK, Roberts WE, Ofner S. Comparison of reliability in anatomical landmark identification using two-dimensional digital cephalometrics and three-dimensional cone beam computed tomography in vivo. Dentomaxillofac Radiol. 2009;38(5):262-73.
12. Chen YJ, Chen SK, Huang HW, Yao CC, Chang HF. Reliability of landmark identification in cephalometric radiography acquired by a storage phosphor imaging system. Dentomaxillofac Radiol. 2004;33(5):301-6.
13. Hlongwa P. Cephalometric analysis: manual tracing of a lateral cephalogram. South African Dental Journal. 2019;74.
14. Leonardi R, Giordano D, Maiorana F, Spampinato C. Automatic cephalometric analysis. Angle Orthod. 2008;78(1):145-51.
15. Silling G, Rauch MA, Pentel L, Garfinkel L, Halberstadt G. The significance of cephalometrics in treatment planning. Angle Orthod. 1979;49(4):259-62.
16. Durão AR, Pittayapat P, Rockenbach MI, Olszewski R, Ng S, Ferreira AP, et al. Validity of 2D lateral cephalometry in orthodontics: a systematic review. Prog Orthod. 2013;14(1):31.
17. Devereux L, Moles D, Cunningham SJ, McKnight M. How important are lateral cephalometric radiographs in orthodontic treatment planning? Am J Orthod Dentofacial Orthop. 2011;139(2):e175-81.
18. Durão AR, Alqerban A, Ferreira AP, Jacobs R. Influence of lateral cephalometric radiography in orthodontic diagnosis and treatment planning. Angle Orthod. 2015;85(2):206-10.

19. Silva MA, Wolf U, Heinicke F, Bumann A, Visser H, Hirsch E. Cone-beam computed tomography for routine orthodontic treatment planning: a radiation dose evaluation. Am J Orthod Dentofacial Orthop. 2008;133(5):640.e1-5.
20. Rodrigues JK, Schwendicke F, Demystifying artificial intelligence and deep learning in dentistry. Braz oral res. 2021;35.
21. Nguyen T-TD, R. Use of Artificial Intelligence in Dentistry: Current Clinical Trends and Research Advances. J Can Dent Assoc. 2021;87(17).
22. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. J Dent Res. 2020;99(7):769-74.
23. Tandon D, Rajawat J. Present and future of artificial intelligence in dentistry. J Oral Biol Craniofac Res. 2020;10(4):391-6.
24. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94-8.
25. Naylor CD. On the Prospects for a (Deep) Learning Health Care System. JAMA. 2018;320(11):1099-100.
26. Rysavy M. Evidence-based medicine: a science of uncertainty and an art of probability. Virtual Mentor. 2013;15(1):4-8.
27. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18.
28. Xie X, Wang L, Wang A. Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. Angle Orthod. 2010;80(2):262-6.
29. Rousseau M, Retrouvey JM. Machine learning in orthodontics: Automated facial analysis of vertical dimension for increased precision and efficiency. Am J Orthod Dentofacial Orthop. 2022;161(3):445-50.
30. Bader JD, Shugars DA, Bonito AJ. Systematic reviews of selected dental caries diagnostic and management methods. J Dent Educ. 2001;65(10):960-8.
31. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 2018;77:106-11.
32. Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. J Dent. 2020;100:103425.
33. Zanella-Calzada LA, Galván-Tejada CE, Chávez-Lamas NM, Rivas-Gutierrez J, Magallanes-Quintanar R, Celaya-Padilla JM, et al. Deep Artificial Neural Networks for the Diagnostic of Caries Using Socioeconomic and Nutritional Features as Determinants: Data from NHANES 2013–2014. Bioengineering (Basel). 2018;5(2).
34. Fu Q, Chen Y, Li Z, Jing Q, Hu C, Liu H, et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. EClinicalMedicine. 2020;27:100558.
35. Jeyaraj PR, Samuel Nadar ER. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. J Cancer Res Clin Oncol. 2019;145(4):829-37.
36. Poedjiastoeti W, Suebnukarn S. Application of Convolutional Neural Network in the Diagnosis of Jaw Tumors. Healthc Inform Res. 2018;24(3):236-41.

37. Sözen T, Özışık L, Başaran N. An overview and management of osteoporosis. Eur J Rheumatol. 2017;4(1):46-56.
38. Taguchi A, Suei Y, Ohtsuka M, Otani K, Tanimoto K, Ohtaki M. Usefulness of panoramic radiography in the diagnosis of postmenopausal osteoporosis in women. Width and morphology of inferior cortex of the mandible. Dentomaxillofac Radiol. 1996;25(5):263-7.
39. Lee KS, Jung SK, Ryu JJ, Shin SW, Choi J. Evaluation of Transfer Learning with Deep Convolutional Neural Networks for Screening Osteoporosis in Dental Panoramic Radiographs. J Clin Med. 2020;9(2).
40. Esposito M, Ardebili Y, Worthington HV. Interventions for replacing missing teeth: different types of dental implants. Cochrane Database Syst Rev. 2014(7):Cd003815.
41. Lee JH, Jeong SN. Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs: A pilot study. Medicine (Baltimore). 2020;99(26):e20787.
42. Papantonopoulos G, Takahashi K, Bountis T, Loos BG. Artificial neural networks for the diagnosis of aggressive periodontitis trained by immunologic parameters. PLoS One. 2014;9(3):e89757.
43. Lee JH, Kim DH, Jeong SN, Choi SH. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. J Periodontal Implant Sci. 2018;48(2):114-23.
44. Zhang X, Xiong S, Ma Y, Han T, Chen X, Wan F, et al. A Cone-Beam Computed Tomographic Study on Mandibular First Molars in a Chinese Subpopulation. PLoS One. 2015;10(8):e0134919.
45. Hiraiwa T, Ariji Y, Fukuda M, Kise Y, Nakata K, Katsumata A, et al. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. Dentomaxillofac Radiol. 2019;48(3):20180218.
46. Hekmatian E, Karbasi Kheir M, Fathollahzade H, Sheikhi M. Detection of Vertical Root Fractures Using Cone-Beam Computed Tomography in the Presence and Absence of Gutta-Percha. ScientificWorldJournal. 2018;2018:1920946.
47. Prithviraj DR, Bhalla HK, Vashisht R, Regish KM, Suresh P. An overview of management of root fractures. Kathmandu Univ Med J (KUMJ). 2014;12(47):222-30.
48. Fukuda M, Inamoto K, Shibata N, Ariji Y, Yanashita Y, Kutsuna S, et al. Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. Oral Radiol. 2020;36(4):337-43.
49. Proffit WR, Fields HW, Jr., Moray LJ. Prevalence of malocclusion and orthodontic treatment need in the United States: estimates from the NHANES III survey. Int J Adult Orthodon Orthognath Surg. 1998;13(2):97-106.
50. Association AD. Oral Health and Well-Being in the United States 2015 [Available from: https://www.ada.org/en/science-research/health-policy-institute/oral-health-and-well-being.
51. Ackerman JL. Orthodontics: Art, Science, or Trans-science? Angle Orthod. 1974;44(3):243-50.
52. Tarvit DJ, Freer TJ. Assessing malocclusion--the time factor. Br J Orthod. 1998;25(1):31-4.
53. Li P, Kong D, Tang T, Su D, Yang P, Wang H, et al. Orthodontic Treatment Planning based on Artificial Neural Networks. Sci Rep. 2019;9(1):2037.
54. Brightman BB, Hans MG, Wolf GR, Bernard H. Recognition of malocclusion: an education outcomes assessment. Am J Orthod Dentofacial Orthop. 1999;116(4):444-51.

55. Thanathornwong B. Bayesian-Based Decision Support System for Assessing the Needs for Orthodontic Treatment. Healthc Inform Res. 2018;24(1):22-8.
56. Merrifield LL, Klontz HA, Vaden JL. Differential diagnostic analysis system. Am J Orthod Dentofacial Orthop. 1994;106(6):641-8.
57. Retrouvey J-M. The role of AI and machine learning in contemporary orthodontics. APOS Trends in Orthodontics.11.
58. Jung S-K, Kim T-W. New approach for the diagnosis of extractions with neural network machine learning. Am J Orthod Dentofacial Orthop. 2016;149(1):127-33.
59. Park JH, Hwang HW, Moon JH, Yu Y, Kim H, Her SB, et al. Automated identification of cephalometric landmarks: Part 1-Comparisons between the latest deep-learning methods YOLOV3 and SSD. Angle Orthod. 2019;89(6):903-9.
60. Cohen AM, Ip HH, Linney AD. A preliminary study of computer recognition and identification of skeletal landmarks as a new method of cephalometric analysis. Br J Orthod. 1984;11(3):143-54.
61. Shahidi S, Oshagh M, Gozin F, Salehi P, Danaei SM. Accuracy of computerized automatic identification of cephalometric landmarks by a designed software. Dentomaxillofac Radiol. 2013;42(1):20110187.
62. Yao J, Zeng W, He T, Zhou S, Zhang Y, Guo J, et al. Automatic localization of cephalometric landmarks based on convolutional neural network. Am J Orthod Dentofacial Orthop. 2022;161(3):e250-e9.
63. Hwang HW, Park JH, Moon JH, Yu Y, Kim H, Her SB, et al. Automated identification of cephalometric landmarks: Part 2- Might it be better than human? Angle Orthod. 2020;90(1):69-76.
64. Song Y, Qiao X, Iwamoto Y, Chen Y-w. Automatic Cephalometric Landmark Detection on X-ray Images Using a Deep-Learning Method. Applied Sciences. 2020;10(7):2547.
65. Oh K, Oh IS, Le VNT, Lee DW. Deep Anatomical Context Feature Learning for Cephalometric Landmark Detection. IEEE Journal of Biomedical and Health Informatics. 2021;25(3):806-17.
66. Zhong Z, Li J, Zhang Z, Jiao Z, Gao X, editors. An Attention-Guided Deep Regression Model for Landmark Detection in Cephalograms. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019; 2019 2019//; Cham: Springer International Publishing.
67. Gilmour L, Ray N, editors. Locating Cephalometric X-Ray Landmarks with Foveated Pyramid Attention. International Conference on Medical Imaging with Deep Learning; 2020.
68. Parthasarathy S, Nugent ST, Gregson PG, Fay DF. Automatic landmarking of cephalograms. Comput Biomed Res. 1989;22(3):248-69.
69. Tong W, Nugent ST, Jensen GM, Fay DF, editors. An algorithm for locating landmarks on dental X-rays. Images of the Twenty-First Century Proceedings of the Annual International Engineering in Medicine and Biology Society; 1989 9-12 Nov. 1989.
70. Cardillo J, Sid-Ahmed MA. An image processing system for locating craniofacial landmarks. IEEE Trans Med Imaging. 1994;13(2):275-89.
71. Forsyth DB, Davis DN. Assessment of an automated cephalometric analysis system. Eur J Orthod. 1996;18(5):471-8.
72. Rudolph DJ, Sinclair PM, Coggins JM. Automatic computerized radiographic identification of cephalometric landmarks. Am J Orthod Dentofacial Orthop. 1998;113(2):173-9.

73. Hutton TJ, Cunningham S, Hammond P. An evaluation of active shape models for the automatic identification of cephalometric landmarks. Eur J Orthod. 2000;22(5):499-508.
74. Liu JK, Chen YT, Cheng KS. Accuracy of computerized automatic identification of cephalometric landmarks. Am J Orthod Dentofacial Orthop. 2000;118(5):535-40.
75. Grau V, Alcañiz M, Juan MC, Monserrat C, Knoll C. Automatic localization of cephalometric Landmarks. J Biomed Inform. 2001;34(3):146-56.
76. Yang J, Ling X, Lu Y, Wei M, Ding G. Cephalometric image analysis and measurement for orthognathic surgery. Med Biol Eng Comput. 2001;39(3):279-84.
77. Innes A, Ciesielski V, Mamutil J, John S. Landmark Detection for Cephalometric Radiology Images Using Pulse Coupled2002. 511-7 p.
78. Chakrabartty S, Yagi M, Shibata T, Cauwenberghs G, editors. Robust cephalometric landmark identification using support vector machines. 2003 International Conference on Multimedia and Expo ICME '03 Proceedings (Cat No03TH8698); 2003 6-9 July 2003.
79. El-Feghi I, Sid-Ahmed MA, Ahmadi M. Automatic localization of craniofacial landmarks for assisted cephalometry. Pattern Recognition. 2004;37(3):609-21.
80. Yue W, Yin D, Li C, Wang G, Xu T. Automated 2-D Cephalometric Analysis on X-ray Images by a Model-Based Approach. IEEE Trans Biomed Eng. 2006;53(8):1615-23.
81. Ibragimov B, editor Automatic Cephalometric X-Ray Landmark Detection by Applying Game Theory and Random Forests2014.
82. Vandaele R, Marée R, Jodogne S, Geurts P, editors. Automatic Cephalometric X-Ray Landmark Detection Challenge 2014: A tree-based algorithm2014.
83. Kaur A, Singh C. Automatic cephalometric landmark detection using Zernike moments and template matching. Signal, Image and Video Processing. 2015;9(1):117-32.
84. Lindner C, Cootes T. Fully Automatic Cephalometric Evaluation using Random Forest Regression-Voting2015.
85. Ibragimov B, Likar B, Pernus F, Vrtovec T, editors. Computerized Cephalometry by Game Theory with Shape-and Appearance-Based Landmark Refinement2016.
86. Lindner C, Wang CW, Huang CT, Li CH, Chang SW, Cootes TF. Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms. Sci Rep. 2016;6:33581.
87. Arık S, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. J Med Imaging (Bellingham). 2017;4(1):014501.
88. Qian J, Cheng M, Tao Y, Lin J, Lin H. CephaNet: An Improved Faster R-CNN for Cephalometric Landmark Detection. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). 2019:868-71.
89. Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. Quintessence Int. 2020;51(3):248-57.
90. Kim H, Shim E, Park J, Kim Y-J, Lee U, Kim Y. Web-based fully automated cephalometric analysis by deep learning. Comput Methods Programs Biomed. 2020;194:105513.
91. Gil S-M, Kim I, Cho J-H, Hong M, Kim M, Kim S-J, et al. Accuracy of auto-identification of the posteroanterior cephalometric landmarks using cascade convolution neural network algorithm and cephalometric images of different quality from nationwide multiple centers. Am J Orthod Dentofacial Orthop. 2022;161(4):e361-e71.

92. Ravindran SK. Random Forest in Simple English: Why is it so popular? 2021 [Available from: https://towardsdatascience.com/random-forest-in-simple-english-why-is-it-so-popular-3ba04d0374d.
93. Seeholzer H, Adamidis JP, Eaton KA, McDonald JP, Sieminska-Piekarczyk B. A survey of the delegation of orthodontic tasks and the training of chairside support staff in 22 European countries. J Orthod. 2000;27(3):279-82.
94. Knoops PGM, Borghi A, Breakey RWF, Ong J, Jeelani NUO, Bruun R, et al. Three-dimensional soft tissue prediction in orthognathic surgery: a clinical comparison of Dolphin, ProPlan CMF, and probabilistic finite element modelling. Int J Oral Maxillofac Surg. 2019;48(4):511-8.
95. Gupta A, Kharbanda OP, Sardana V, Balachandran R, Sardana HK. A knowledge-based algorithm for automatic detection of cephalometric landmarks on CBCT images. Int J Comput Assist Radiol Surg. 2015;10(11):1737-52.
96. Hung K, Yeung AWK, Tanaka R, Bornstein MM. Current Applications, Opportunities, and Limitations of AI for 3D Imaging in Dental Research and Practice. Int J Environ Res Public Health. 2020;17(12).

Appendix

1. Ethics Approval

Comité d'éthique de la recherche clinique (CERC)

Bureau de la conduite responsable en recherche

Université de Montréal

04 mai 2022

Normand Bach, professeur agrégé
Faculté de médecine dentaire
Université de Montréal

Emmanuel Suissa
Candidat à la maîtrise
Faculté de médecine dentaire
Université de Montréal

OBJET :	Projet # 2022-1334 - Approbation éthique finale Étude de la fiabilité du traçage des radiographies céphalométriques par intelligence artificielle.

M. Bach,

Le Comité d'éthique de la recherche clinique (CERC) de l'Université de Montréal a évalué votre projet de recherche en comité restreint à sa rencontre du 21 février 2022. Suite à cette réunion, une approbation conditionnelle vous a été émise en date du 21 février 2022.

Nous accusons réception des précisions et corrections demandées via le formulaire de conditions F20 ainsi que des documents en vue de l'approbation finale du projet mentionné en rubrique. Suite à la révision de ces documents, le tout ayant été jugé satisfaisant, j'ai le plaisir de vous informer que votre projet de recherche a été approuvé à l'unanimité par le CERC.

Le document que le CERC a approuvé et que vous pouvez utiliser pour la réalisation de votre projet est le suivant :

- 22_05_04_Protocole_v2022-02-22_approuvéCERC

La version approuvée du document est disponible dans la section **Documents approuvés par le CER**, située sous l'onglet "Fichiers" de votre projet.

Cette approbation éthique est valide pour un an, à compter du 04 mai 2022 jusqu'au 04 mai 2023. Il est de votre responsabilité de compléter le formulaire de renouvellement (formulaire F9) que nous vous ferons parvenir annuellement via Nagano 1 mois avant l'échéance de votre approbation, à défaut de quoi l'approbation éthique délivrée par le CERC sera suspendue.

Dans le cadre du suivi éthique continu, le Comité vous demande de vous conformer aux exigences suivantes en utilisant les formulaires Nagano prévus à cet effet :

- Soumettre, pour approbation préalable, toute demande de **modification** au projet de recherche ou à tout autre document approuvé par le Comité pour la réalisation du projet (formulaire F1).
- Soumettre, dès que cela est porté à votre connaissance, toutes **informations supplémentaires,**

nouveau renseignement et/ou correspondances diverses (formulaire F2).

- Soumettre, seulement pour essais cliniques sous la juridiction de Santé Canada et dès que cela e porté à votre connaissance, tout **événement indésirable grave et inattendu** (EIGI) survenu da votre site ou dans un site pour lequel le Comité a juridiction (formulaire F3).
- Soumettre, dès que cela est porté à votre connaissance, tout **incident ou accident** lié à réalisation du projet de recherche (formulaire F5).
- Soumettre, dès que cela est porté à votre connaissance, l'**interruption prématurée** du projet recherche, qu'elle soit temporaire ou permanente (formulaire F6).
- Soumettre, dès que cela est porté à votre connaissance, toute **déviation** au projet de recherc susceptible de remettre en cause le caractère éthique du projet (formulaire F8).
- Soumettre une demande de **renouvellement** un mois avant l'échéance de la date d'approbation a de renouveler l'approbation éthique (formulaire F9).
- Soumettre le rapport de la **fin du projet de recherche** (formulaire F10).

Nous vous rappelons que la présente décision vaut pour une année et peut être suspendue ou révoqu en cas de non-respect de ces exigences.

Le CERC de l'Université de Montréal est désigné par le ministre de la Santé et des Services Sociaux a fins de l'application de l'article 21 du Code civil du Québec. Il exerce ses activités en conformité avec *Politique sur la recherche avec des êtres humains* (60.1) de l'Université de Montréal ainsi que l'*Énoncé politique des trois conseils* et les *Bonnes pratiques cliniques* de la CIH. Il suit également les normes règlements applicables au Québec et au Canada.

Cordialement,

Pour la présidente du CERC, Nathalie Folch,

Camille Assemat
Conseillère en éthique de la recherche
Comité d'éthique de la recherche clinique (CERC)
Bureau de la conduite responsable en recherche
Université de Montréal
3333, chemin Queen-Mary, bureau 220
Montréal (Québec) H3V 1A2
Tél. 514 343-6111, poste 27395
cerc@umontreal.ca

Envoyé par :
Camille Assémat