INSTITUTO UNIVERSITÁRIO EGAS MONIZ

MESTRADO INTEGRADO EM MEDICINA DENTÁRIA

A NARRATIVE EXPLORATION OF ARTIFICIAL INTELLIGENCE FOR ORTHODONTIC DIAGNOSIS AND DECISION-MAKING IN TREATMENT PLANNING.

Trabalho submetido por
Charlotte Cachau-Herreillat
para a obtenção do grau de Mestre em Medicina Dentária

outubro de 2024

INSTITUTO UNIVERSITÁRIO EGAS MONIZ

MESTRADO INTEGRADO EM MEDICINA DENTÁRIA

A NARRATIVE EXPLORATION OF ARTIFICIAL INTELLIGENCE FOR ORTHODONTIC DIAGNOSIS AND DECISION-MAKING IN TREATMENT PLANNING.

Trabalho submetido por
Charlotte Cachau-Herreillat
para a obtenção do grau de **Mestre** em Medicina Dentária

Trabalho orientado por
Mestre François Durand Pereira

outubro de 2024

Agradecimentos

Em primeiro lugar, gostaria de expressar os meus mais profundos agradecimentos ao Mestre François Durand Pinto Pereira, meu orientador de tese, pelo seu apoio constante, pelos seus bons conselhos e pela sua disponibilidade durante a realização deste trabalho. A sua competência e rigor acompanharam-me ao longo de todo o processo e reforçaram a qualidade deste trabalho.

Foi um prazer colaborar e trabalhar convosco, aceitem a minha gratidão através destas palavras.

Gostaria também de agradecer à instituição Egas Moniz, que contribuiu para a minha formação académica e me forneceu os conhecimentos necessários para a realização deste projeto.

À mes parents et au soutient constant qu'ils m'ont apporté au cours de ses longues années. Merci de m'avoir inculqué les valeurs de persévérance et de détermination, je réalise la chance que nous avons eu que vous soyez aussi présents et impliqués dans nos éducations. Ce projet n'aurait jamais vu le jour sans vous.

À mon frère, pour sa patience avec moi au quotidien et pour m'avoir expliqué les produits de matrices de dizaines de fois.

À mes amis de France qui ont toujours présents pour moi malgré la distance, je vous aime.

À toutes les personnes merveilleuses que j'ai rencontrées ici, grâce à vous j'ai vécu 5 années inoubliables, merci d'avoir su adoucir les moments difficiles et d'avoir multiplié les moments de joie.

À ma binôme, le travail n'aura pas la même saveur sans toi.

Resumo

O autor pretende avaliar a utilização da inteligência artificial (AI) no campo da ortodontia, e em particular no diagnóstico e na tomada de decisão para o plano de tratamento.

Antes de mais será realizada uma explanação da inteligência artificial atual assim como das suas principais áreas, em particular a aprendizagem automática (*machine learning*) e a aprendizagem profunda (*deep learning*), no contexto dos diagnósticos ortodônticos (Leonardi et al., 2021). Explorando os conceitos fundamentais da AI nas aplicações atuais na área da medicina, esta revisão pretende realçar as várias metodologias, técnicas e modelos relevantes para a ortodontia.

É também nosso objetivo examinar o impacto dos modelos de AI no planeamento personalizado do tratamento ortodôntico, avaliando a qualidade das bases de dados onde a AI se alimenta, fornecendo uma avaliação abrangente das vantagens e desafios associados à sua integração no processo de tomada de decisão (Etemad et al., 2021).

Abordando o impacto clínico e as perspetivas futuras, o autor propõe-se analisar a eficácia da AI na tomada de decisões ortodônticas, explorando suas implicações para a eficiência, qualidade dos resultados, tendências e desafios futuros na área da ortodontia (Jung & Kim, 2016).

O autor pretende também realçar as considerações éticas em torno da aplicação de AI na ortodontia, assim como as limitações das práticas atuais e recomendações propostas para utilização ética e responsável da AI na área da saúde e em particular no paciente ortodontico (Mörch et al., 2021).

Concluindo a revisão narrativa, o autor propõe-se destacar as eventuais áreas onde a aplicação da AI em ortodontia necessitará de maior orientação e investigação para aumentar a eficiência no diagnóstico e planeamento do tratamento ortodôntico.

Palavras-chave: Artificial Intelligence; Orthodontics; Decision-Making; Machine Learning

Abstract

The author aims to evaluate the use of artificial intelligence (AI) in the field of orthodontics, and in particular in diagnosis and decision-making for the treatment planning.

First of all, an explanation will be given of current artificial intelligence and its main areas, in particular machine learning and deep learning, in the context of orthodontic diagnosis (Leonardi et al., 2021b). By exploring the fundamental concepts of AI in current medical applications, this review aims to highlight the various methodologies, techniques and models relevant to orthodontics.

We also aim to examine the impact of AI models on personalised orthodontic treatment planning by evaluating the quality of the databases that AI is fed into, providing a comprehensive assessment of the advantages and challenges associated with integrating them into the decision-making process (Etemad et al., 2021).

Addressing the clinical impact and future prospects, the author sets out to analyse the effectiveness of AI in orthodontic decision-making, exploring its implications for efficiency, quality of results, trends and future challenges in the field of orthodontics (Jung & Kim, 2016).

The author also aims to highlight the ethical considerations surrounding the application of AI in orthodontics, as well as the limitations of current practices and proposed recommendations for ethical and responsible use of AI in healthcare and in particular in the orthodontic patient (Mörch et al., 2021).

In concluding the narrative review, the author proposes to highlight possible areas where the application of IA in orthodontics will require further guidance and research in order to increase efficiency in orthodontic diagnosis and treatment planning.

Key-Words: Artificial Intelligence; Orthodontics; Decision-Making; Machine Learning

SUMMARY

FIGURES INDEX

TABLES INDEX

LIST OF ACRONYMS

AGI - Artificial General Intelligence

AI - Artificial Intelligence

ANI - Artificial Narrow Intelligence

ANN - Artificial Neural Network

ASI - Artificial Super Intelligence

AUC - Area Under the Curve

BN - Bayesian Network

CBCT - Cone-Beam Computed Tomography

CDSS - Clinical Decision Support System

CNN - Convolutional Neural Network

CT - Computed Tomography

DL - Deep Learning

EMR - Electronic Medical Records

ICC - Intraclass Correlation Coefficient

kNN - k-Nearest Neighbors

LSTM - Long Short-Term Memory

MAE - Mean Absolute Error

ML - Machine Learning

MLP - Multilayer Perceptron

MRI - Magnetic Resonance Imaging

NN - Neural Network

ReLU - Rectified Linear Unit

RF - Random Forest

ROI - Region of Interest

RNN - Recurrent Neural Network

SSD - Single Shot Multibox Detector

TMD - Temporomandibular Disorders

TMJ - Temporomandibular Joint

YOLO - You Only Look Once

I- INTRODUCTION

Artificial intelligence (AI) is increasingly integrated into our daily lives, transforming many sectors through its advanced data processing and analysis capabilities. From voice assistants like Siri and Alexa to autonomous cars, AI is improving the efficiency and accuracy of tasks once reserved for humans. This omnipresence of AI offers opportunities for technological breakthroughs and process improvements in a number of fields, including medicine and more specifically orthodontics.

In the medical field, AI is already being used to improve diagnosis, treatment planning and patient care management. For example, AI systems can analyse large amounts of medical data to identify patterns and tendencies that might be missed by human clinicians. This can improve the accuracy of diagnoses and personalise treatments based on patients' specific needs (E. Topol, 2019).

In radiology, AI is helping to detect medical image abnormalities, such as tumours or fractures, with greater accuracy and speed (S. Jha & Topol, 2016). In genomics, AI enables entire genomes to be sequenced and analysed more quickly, making it easier to discover new genetic associations with diseases and identify potential therapeutic targets (Libbrecht & Noble, 2015).

The use of AI in medicine is not limited to diagnostics. It is also being employed in planning and executing robot-assisted surgeries, where advanced algorithms are guiding surgical instruments with unmatched precision, reducing risk and improving patient outcomes (Yang et al., 2017).

AI is also transforming healthcare management by optimising administrative processes, such as scheduling appointments, handling medical records and coordinating care between different healthcare professionals. These improvements allow healthcare professionals to focus more on patient care rather than administrative tasks (Davenport & Kalakota, 2019).

With all these advances, AI's help is not negligible in the medical field, so it's interesting to wonder what role AI plays in orthodontics. We can hope that it also has the potential to completely revolutionise diagnosis and treatment planning.

As we are going to see in this narrative review, several studies have already demonstrated the effectiveness of AI in the field of orthodontics. Whether through performance comparisons between AI and human experts in the identification of cephalometric landmarks, or concerning the results of AI accuracy and reproducibility (Hwang et al., 2020).

Systematic reviews have also been conducted to analyse these studies, concluding that AI offers clinically acceptable diagnostic performance (Junaid et al., 2022)

In order to explore this topical subject, we will divide this work into several parts. Firstly, we will present the different types of AI as well as the basic models and those used in orthodontics. Next, we will examine their clinical use, and the performance of the models used. Finally, we will discuss the challenges that AI still faces, including ethical issues, and futurs prospects for overcoming these challenges.

II- DEVELOPMENT

1. ARTIFICIAL INTELLIGENCE

The Massachusetts Institute of Thechnology (MIT) defines the Artificial Intelligence (AI) as the ability for computers to imitate cognitive human functions such as learning and problem-solving, using math and logic to simulate the human reasoning (jeferson.zambrano, 2022). AI is usually classified as either weak or strong.

- Weak AI, also known as Artificial Narrow Intelligence (ANI), are designed and developed to perform particular task. It has access to data and human assistance, and must be capable of simulating human reasoning, without understanding the phenomena involved.
- Strong AI comprises both Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI).
  AGI refers to AI systems designed to attain human-level intelligence, aspiring to showcase human-like adaptability, learning capabilities, and reasoning abilities. On the other hand, ASI aims to achieve machine intelligence that surpasses human capabilities, possessing superior cognitive functions. Both forms of Strong AI remain theoretical technologies with no practical examples currently in use. However, achieving AGI represents an ongoing challenge in the field of AI research.

Weak AI powers the majority of today's AI applications and has demonstrated several advantages, including increasing the efficiency for repetitive tasks, or facilitates informed decision-making by rapidly processing vast amounts of data inputs, including big data (Matteis, 2022; Nancholas, 2023; *Stryker*, 2024).

AI includes different sub-section such as Machine Learning (ML), Neural Network (NN) and Deep Learning (DL), which we will describe in more details below. ML consists of learning implicit rules to respond to a given problem, either from experience or from a database. This field is particularly focused on the statistical analysis of training data. In the context of ML, we will see that there are many algorithms using a variety of mathematical models.

An algorithm is a well-defined sequence of steps or instructions designed to solve a specific problem or perform a calculation. It takes an input, processes it through a series of operations and produces an output. They are fundamental to computing and are used to automate tasks, optimise processes and solve problems efficiently (Cormen et al., 2009).

The NN is one of these models, certainly the most widespread and the one used in the most diverse fields. DL is a set of techniques that use neural networks to solve complex problems (Matteis, 2022).

1.1. MACHINE LEARNING

As we explain earlier, Machine Learning is the technique that improves system performance by learning from experience, to resolve problem that have never been encountered (Zhou, 2021). This is known as learning by generalisation, in opposition of the learning by rote, which is characterised by the explicit memorisation of all possible examples. Historically, this branch of AI is defined as the development of machines capable of learning without being specifically programmed to learn a task.

In ML, several methods are used to handle data-related issues, however, if all possible problems are averaged over, the "No Free Lunch Theorem" asserts that all optimization algorithms perform equally well. That implies there is no single one-size-fits-all type of algorithm that is best to solve a problem, the choice of algorithm depends on the specific problem you aim to solve, the number of variables involved, the most suitable model for the task, and other relevant factors (Mahesh, 2018; Matteis, 2022).

We will now take a quick look at the three basics machine learning paradigms.

1.1.1. Supervised learning

In supervised learning, we work with labelled training data, consisting in a training example where each input is paired with a known output. The objective is to train the model to predict outputs for unseen data, referred to as the test dataset (Mahesh, 2018).

Noting the N inputs xi and the associated target outputs yi, we have the dataset $D = \{xi, yi\}_{i \in [1, N]}$. These type of algorithms require external guidance (Matteis, 2022).

Supervised learning is mostly use for classification with categorical output values or for regression where the output values are numeric (Shen et al., 2017).

There are differents types of supervised learning, such as:

- <u>Linear regression</u>, where the output y is express based on an input x, in a linear way, meaning that y= $ax + b,$ during the training data the algorithm is looking for the optimum a and b (Matteis, 2022). This is a type of parametric model.

- k-Nearst Neighbors (kNN), where the outputs correspond to differents classes and not to continuous values, which make it a non-parametric classification model. The learning consists in storing the training data, then finding the k nearests points to the new data by measuring the distance between them. The value of k is chosen before the training and affects how the process works, it's called an hyperparameter (Nyuytiymbiy, 2022).

  The class of the new data is determined by the classes of the k nearest neighbours, it corresponds to the most frequent class among them. An upgrade of this approach would be to weight the importance of each neighbour by their distance.

  Although this type of algorithm is very effective, it is becoming significantly slow as the size of the data grows (Badillo et al., 2020; Bansal et al., 2022; Mahesh, 2018; Matteis, s. d.).

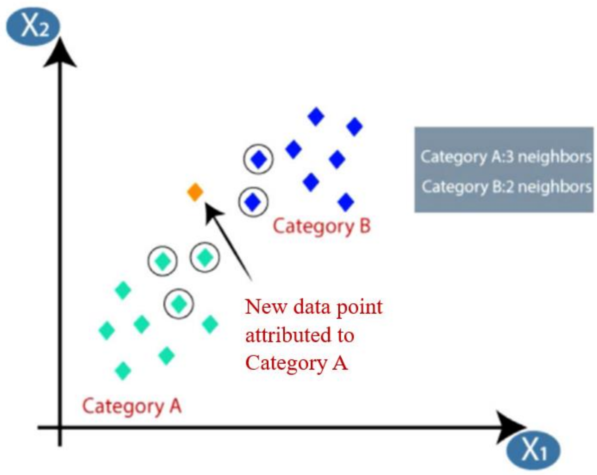


Figure 1 : Classification of new data point based on k-Nearest Neighbours (adapted from Bansal and al., 2022)

1.1.2. Unsupervised learning

Unlike the supervised learning, unsupervised learning is generated without any supervision or external guidance. We provided an unsegregated and unlabelled dataset, meaning we only have the input and the output remain unknown, the dataset will therefore be noted as $D = \{xi\}\ i_{\in[1,\,N]}$. The goal will be to identify common characteristics in the training data. The previously learned features are used to recognize the class of the data when a new data is introduced. (Bansal et al., 2022; Matteis, s. d.)

Clustering

This type of algorithms is mainly used for clustering, the principle is to separate the data into different groups in order to automatically labels them. "*A cluster is a subset of the data which are "similar" to each other, whereas points belonging to different clusters are more "different.""* (Badillo et al., 2020). K-means clustering is probably the simplest unsupervised learning algorithm, the number of clusters is defined by a parameter *k*, and for each one you defined an artificial point called *k* center, the smartest approach being to put them as far away as possible from each other's.

The first step is the assignment where data point is assigned to each *k* represented by the closest *k* center. The second step is the center shift, each bary center is re-calculated based on the composition of the cluster in step one. The previous steps are repeated until the centers no longer move between two iterations, that is what we call convergence (Mahesh, 2018; Matteis, 2022).

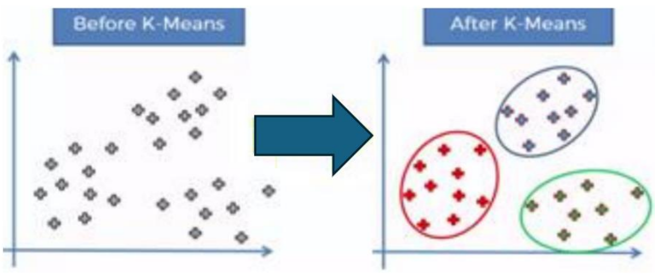


Figure 2 : Classification based on k-Means Clustering (adapted from Mahesh, 2018)

1.1.3. Reinforcement learning

Reinforcement learning is another type of algorithm that contain two fundamental components: the agent and the environment. The agent interacts with the environment, taking actions at each iteration, and it is rewarded for each correct move, and penalized for the incorrect action. The ultimate goal is to teach the agent to interact efficiently with the environment, aiming to achieve the highest score.

An easy example would be a ping pong game, where the game itself would be the environment, and the agent controls the paddle hitting the ball back and forth. The agent interacts with the environment through a predefined action set (ex: moveUp, moveDown) and receives rewards when the player scores.

It is therefore very important to choose carefully the rewards and penalties to be applied, as this determines the results of the final model (Bansal et al., 2022; Dong et al., 2020; Mahesh, 2018; Matteis, 2022).

1.2. NEURAL NETWORK AND DEEP LEARNING

As we already know DL is part of the ML algorithm, but it is based on artificial Neural Network (Goodfellow et al., 2016), that is why we will first focus on NN and then on the explanation of DL.

1.2.1. Neural Network

Artificial NN are a group of models that imitate the structure of the neural system and identify patterns present in data, the most basic being the perceptron. In fact, it consists exclusively of an input layer and an output layer (Shen et al., 2017). The perceptron follows the same pattern as the brain where a neuron receives electrical impulses from its dendrites, once this particular neuron reaches a sufficient level of polarization, it will transmit a spike of action potential through its axon to the other adjacent neurons (Dong et al., 2020). A perceptron, or a modified perceptron with multiple output units, is regarded as a linear model (Shen et al., 2017).

Perceptron

A NN is organised in differents layers, and as we already know the perceptron is only composed by two layers, the input and the output. In this first example shown in Figure 3, the perceptron also only has one output, and here we will set three inputs or neurones. Each of them are related to the output by an edge where differents weights can be applied, representing the different degrees of impact of each neuron on the next, the greater the impact, the greater the value. (Badillo et al., 2020; Dong et al., 2020)

At the end, to fit the date better, we add an extra value to the output called the bias.

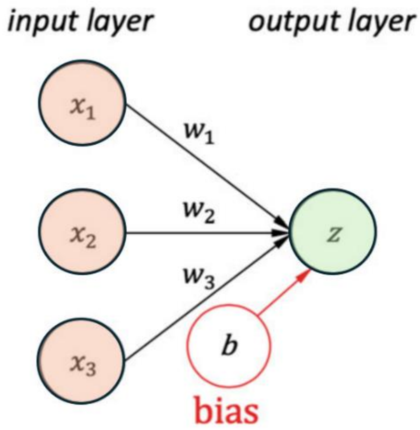


Figure 3 : Single neural network or perceptron with bias (adapted from Dong and al., 2020)

Modified Perceptron

We will now have an example of NN with two outputs, the principle remains exactly the same. If we modified a little bit the Figure 3, we have:

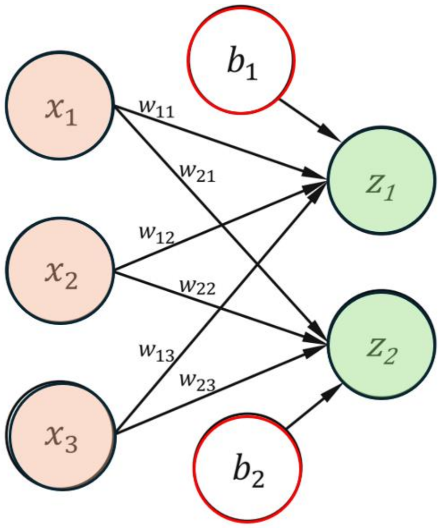


Figure 4 : NN or modified perceptron with three input neurons and two output neurons (adapted from Dong and al., 2020)

A so-called hidden layer can be introduced between the input layer and the output layer in order to overcome the limitations that may be implied by complicated data patterns. The number of hidden layers define whether the system is a shallow learning system (with one or a few hidden layer) or deep learning (with many hidden layers) (Badillo et al., 2020; Shen et al., 2017).

Multilayer Perceptron (MLP)

We refer at a multilayer perceptron when there is one or more hidden layer between the input and the output, this type of model can fit more complex data and has a stronger learning capability. Even if the number of layers is not directly related to the learning capacity, MLP usually have several hidden layers, especially those used for deep learning as we will see below (Dong et al., 2020).

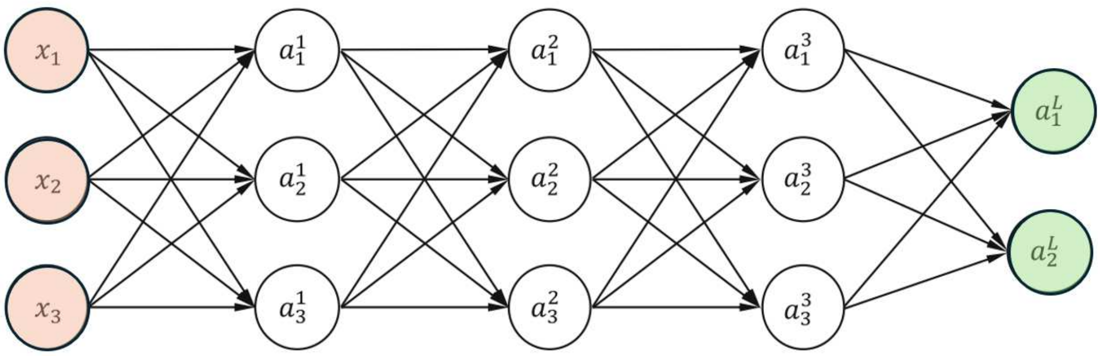


Figure 5 : Multilayer percetron with three hidden layers (adapted from Dong and al., 2020)

In this illustration we can see three hidden layers where the neurons are represented by α, with the upper number representing the layer index (or the column) and the lower number representing the output index (or the line).

The MLP is the quintessential example of a deep learning model, it is also called Feedforward deep network (Goodfellow et al., 2016).

Activation functions

NN are in fact an extrapolation of the linear regression to larger dimensions by rewriting it in its matrix form, and the output layer of a MLP can be represented by a matrix multiplication. In a simple way we can represente : $z = Wx + b$, where W is a matrix representing the weights, z represent the output, x the input, and b the bias.

As a result, matrix addition and multiplication are both linear operators, wich give a limited learning capacity to the model. To respond to more complex problems and to learn a nonlinear mapping, we are introducing a non-linear part to the model, the activation function.

There is the most commonly used:

- Sigmoid: can be use on the output layer, mostly for classification purpose
- Hyperbolic tangent (tanh): on the hidden or output layers.
- Rectified linear unit (ReLU): more promising than the sigmoid because it is easier to implement and to compute, and also easier for a network to optimized.
- Softmax: provides normalization, only used on the output layer to normalize into a probability (Dong et al., 2020; Kelleher, 2019).

1.2.2. Deep Learning

« *The term deep learning describes a family of NN models that have multiple layers of simple information processing programs, known as neurons, in the network* » (Kelleher, 2019).

DL networks are NN that have at least two hidden layers of neurons, they can have a greater data representation capacity in comparison with shallow learning system. They can also be called end to end learning as we only care about the input and the output and less on the feature. Of course, there is an inherent trade-off between the number of layers and time required to train the model. (Badillo et al., 2020; Dong et al., 2020; Kelleher, 2019).

In the following section, we will explore several deep NN models to help in our understanding of the following topics.

Feedforward Neural Network

As we already saw, feedforward NN represent the most fundamental type of NN, operating by transmitting information sequentially from the input layer through hidden layers before reaching the output layer.

The term "feedforward" derives from the absence of feedback loops, all connections point forward to subsequent layers, without any links back from the output of a neuron to the input of an earlier layer. Their goal is to approximate a target function and, despite this simplicity, they can approximate complex functions. Feedforward networks current state is not defined by any past state, making them memory-less systems.

Consequently, feedforward networks are designed for statistical generalization, similar to linear models, necessitating decisions on optimizers, cost functions, and output unit forms during training (Badillo et al., 2020; Goodfellow et al., 2016; Kelleher, 2019).

Convolutional Neural Networks

The Convolutional NN or CNNs are a variant of MLP, designed for image recognition tasks. The main idea was to build a network where early layer neurons would extract local visual characteristics, and later layer neurons would combine to construct higher-order features (Dong et al., 2020; Kelleher, 2019).

One challenge inherent to this network type is ensuring that the feature detection functions work in a translation-invariant manner. This implies that the CNN should effectively identify the presence of local visual features within an image, regardless of where in the image it occurs.

Structurally, CNNs have two kinds of layers, convolutional layer and pooling layer, followed by fully connected layer, the dense layer, that operates in exactly the same way as a standard layer in a feedforward network. In the convolutional layer, the neuron is locally connected rather than being connecting to all the units from the previous layer, which means that each neuron inspects a different portion of the image, those are called the receptive field.

The receptive fields of two neurons can overlap, the amount of that overlap is controlled by an hyperparameter called the stride length.

Another characteristic of the CNNs is weight sharing, and if two neurons use the same set of weights then they both implement the same function, in this case feature detector. Consequently, it makes translation invariant feature detection possible.

The kernel defines the feature detection function, it is the equivalent to passing a local visual feature detector across the image and recording all the locations in the image where the visual feature was present. The output of this convolutional process is a map of all the location in the image where the relevant visual feature occurred, it is also known as a feature map.

A convolution operation does not include a nonlinear activation function; therefore, it is standard to apply a nonlinearity operation to a feature map.

Following the convolutional part a pooling layer is used to downsample the updated feature map, it takes advantage that neighbouring pixels are similar. The most common pooling functions are max-pooling and average-pooling. The first one returns the maximum value of any of its inputs and the second one the average value.

The standard operation sequence for a CNN is applying a convolution, which give us a feature map, followed by a nonlinearity and then a pooling layer to reduce sampling. To generalize beyond one feature, we can have multiple convolutional layers in parallel, also called filters.

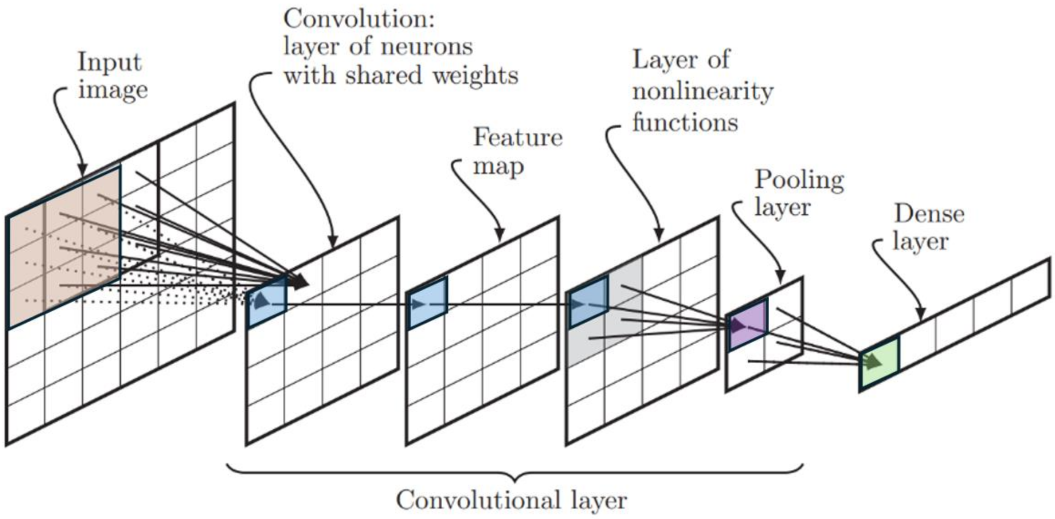


Figure 6 : Different stages of processing in a convolutional layer (adapted from Kelleher, 2019)

Convolutional layers have the advantage of being able to be trained more quickly than dense layers because they have considerably less connections to the previous layer (Dong et al., 2020; Goodfellow et al., 2016; Kelleher, 2019; Shen et al., 2017).

Recurrent Neural Networks

The Recurrent NN, or RNNs, are designed to process sequential data which refers to a sequence of values, unlike CNNs which is specialized for processing a grid of values, with the network processing each element in the sequence one at time.

RNN only has one hidden layer, combined to a memory buffer that stores the output of this hidden layer for one input and feeds it back into the hidden layer along with the next input from the sequence. The network processes each input within the context generated by processing the previous input, creating a recurrent flow of information. This recurrence allows the network to maintain a memory, processing each input within the context of what it has previously processed. At each step, the network receives two elements as an input, the input itself and the elements of the memory. In the Figure 7, we can see the process of a sequence of inputs, the active paths of information are represented by the bold arrows at each time point, and the dashed arrows represent the inactive connections at that time.

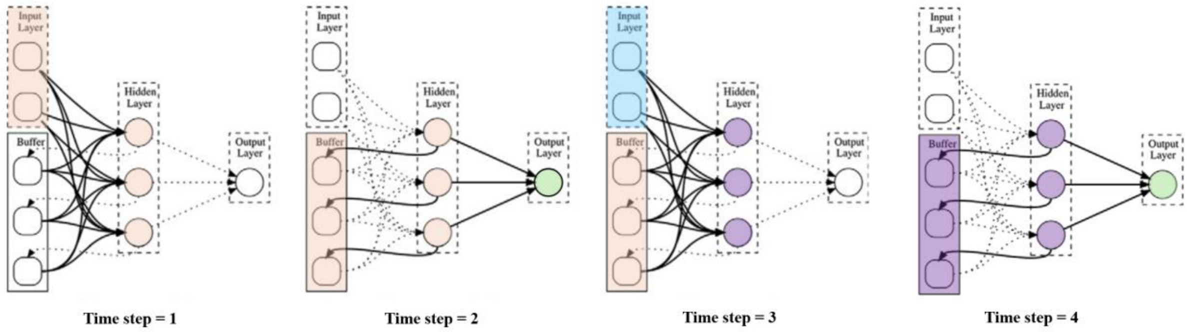


Figure 7 : Different steps in a recurrent neural network (adapted from Kelleher, 2019)

Like CNNs, they can also adopt parameter sharing, allowing to repeatedly use the same weight across multiple locations in the input sequential data, by using a very deep computational graph.

One of the problems of RNNs is the vanishing gradient, which describes the fact that as more layers are added to a network, it takes longer to train it. In order to reduce this effect, we used Long Short-term Memory Networks (LSTMs), they are more sophisticated to handle the long-term dependencies in long sequences (Dong et al., 2020; Goodfellow et al., 2016; Kelleher, 2019).

In addition to these different types of machine learning, other tools can be used to improve the algorithms, such as Bayesian networks, which are probabilistic graphical models used to represent the conditional dependency relationships between different variables in a complex system, or in the case of a causal BN, direct cause and effect relationships. They offer a compact and powerful way of modelling non-deterministic systems, meaning systems where events do not occur with certainty but with a certain probability (Kitson et al., 2023).

2. AI IN ORTHODONTIC DIAGNOSTIS

AI now plays a central part in medical science, revolutionising both diagnostic methods and treatment strategies. Its ability to process massive amounts of data, extracting relevant information quickly and accurately, makes it a particularly valuable tool in many fields such as oncology, AI-assisted surgery or radiology.

In oncology, for example, AI systems can process large genomic datasets to identify specific biomarkers and therefore guide personalised treatments adapted to each patient's molecular characteristics. In surgery, AI and surgical robots have made it possible to perform procedures with extreme precision, reducing post-operative complications and improving clinical outcomes. Particularly effective in image analysis, AI is used in radiology to detect anomalies such as tumours or fractures with an accuracy often superior to the human eye (Esteva et al., 2017; E. J. Topol, 2019).

Orthodontics is no exception to this technological transformation, in particular in cephalometric analysis.

The use of AI in orthodontic diagnosis represents a significant advance in terms of accuracy, speed and efficiency of cephalometric analyses. AI technologies, particularly CNNs, have demonstrated a remarkable ability to automate the identification of cephalometric landmarks and the analysis of radiographic images. These advances not only offer significant improvements for clinical diagnosis, but also opportunities for clinician learning and development (Jiang et al., 2023; Menezes et al., 2023; Park et al., 2019). Therefore, in this section we will explore the different applications of AI to identify cephalometric landmarks and analyse cephalometric images, while highlighting the advanced techniques that have achieved high levels of accuracy and efficiency. We will also look at a case study illustrating the success of these methods, comparing the diagnostic effectiveness of AI with human experts. Finally, we will discuss current challenges and future prospects, looking at how expected technological developments could further improve the performance of AI in orthodontic diagnosis (Bao et al., 2023; Hwang et al., 2020; Panesar et al., 2023).

2.1. AI-ENHANCED DIAGNOSTICS

2.1.1. AI applications for identifying landmarks and analysing radiographic images.

In this part we are going to explore how AI is transforming orthodontic diagnosis, focusing particularly on the identification of cephalometric landmarks and the analysis of radiographic images.

The first introduction of an automated landmark identification method has been suggested since the mid-1980s, but the lack of accuracy at the time made it useless for clinical purposes. However, recent algorithms have shown significant improvements in accuracy (Menezes et al., 2023; Park et al., 2019).

One advanced technique involves the use of two-layer cascaded CNNs. Initially, an orthodontist manually annotated 1,063 cephalograms with 30 cephalometric landmarks, including skeletal, dental and soft tissue landmarks. This manual annotation forms the training part of the CNN model, establishing a reference data set to improve its accuracy and efficiency.

The CNN itself consists of two parts, the RegionNet and the LocationNet. The RegionNet is designed to detect 10 regions of interest, operating like the convolutional layer by analysing each of these regions separately to identify important local characteristics. The resulting sub-images from this first stage are then used as input by the LocationNet, which uses a similar architecture to provide a precise location for each landmark.

After testing on the full data set, this system showed an average landmark prediction error of 0.94 ± 0.74 mm, with an average classification accuracy of 89.33%. Using a large set of training samples and an innovative algorithm, they achieved high accuracy and applicability (Jiang et al., 2023).

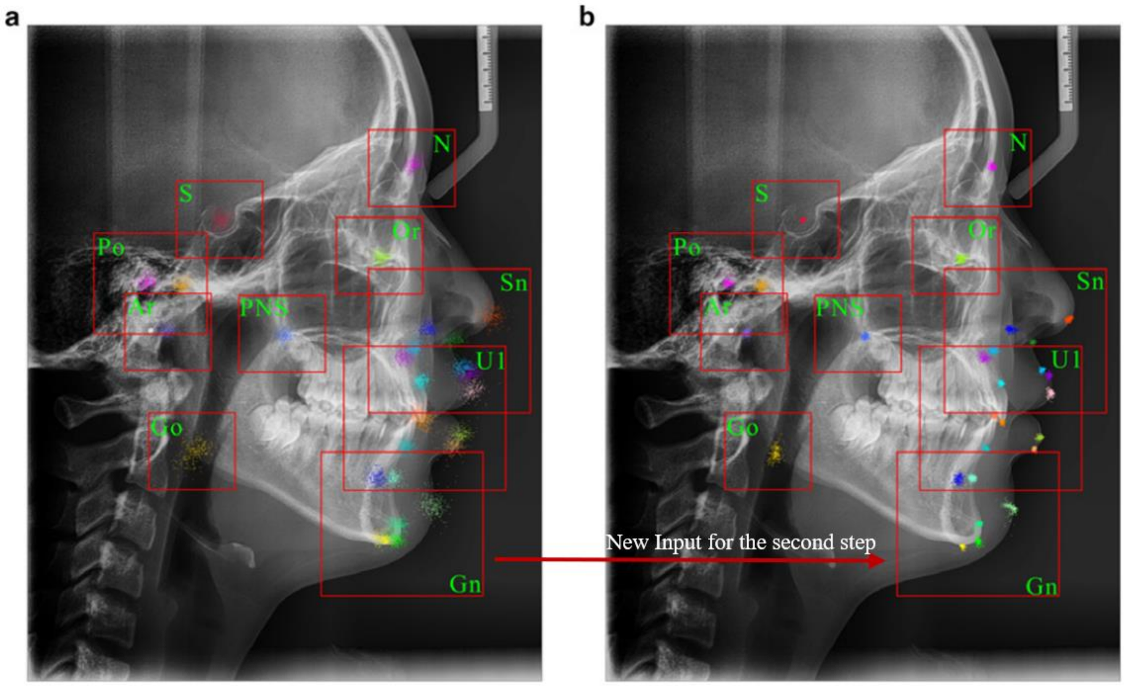


Figure 8 : a) ROIs boxes drawn in cephalogram with the prediction error of RegionNet, b) ROIs boxes drawn in cephalogram with the prediction error of LocationNet (adapted from Jiang and al., 2023)

Another example of a cascade NN for automated and accurate cephalometric analysis, consisting in two phases like the first one. The first phase involves identifying the Regions of Interest (ROI), a task performed here by RetinaNet, and the second consists of precisely detecting the landmarks, using U-Net. From 1,000 lateral cephalometric images, an experienced orthodontist identified 42 landmarks. The model for the cascade CNN recorded an average error of 0.79±0.91 mm, leading to the conclusion that this type of model is applicable in current clinical practice, which is similar to the one of the first algorithm (I.-H. Kim et al., 2021).

The diagnostic value of a cephalometric analysis relies on reproducibility and accuracy, both skeletal and soft tissue landmarks are essential for evaluation, treatment planning, and outcome prediction. In order to improves the accuracy of treatment outcome predictions it is important to use an increased number of landmarks. The major source of error in is known to be the inconsistency in landmark identification. This inconsistency is mainly a result of the subjectivity and variability inherent in human examiners (Menezes et al., 2023).

Although the use of AI for automatic cephalometric marking can be considered as excellent, another source of error can be the uncertainty in visual identification. This variability can be influenced by different brightness and contrast settings. The literature suggests that these variations in setting result in statistically significant differences in reproducibility, especially at high contrast and low brightness, highlighting the need for consistent imaging conditions (Menezes et al., 2023).

Moreover, while AI performs well in identifying landmarks, it faces similar challenges as human examiners. Some factors such as overlapping anatomical structures, image sharpness limitations, and individual anatomical variability can influence the accuracy of AI-based landmark identification. In contrast, the results of multiple linear regression analysis indicate that image variables such as gender, skeletal classification, image quality, and the presence of metal artifacts do not significantly affect the accuracy of AI-based landmark identification. In fact, AI behaves in the same way as a human examiner, where the observer has difficulty identifying a landmark, the AI will encounter the same difficulties. These issues underline the importance of high-quality images and the potential need for manual corrections by a clinician to ensure clinical reliability (Hwang et al., 2020; Menezes et al., 2023).

To illustrate a successful case, it is worth mentioning the latest deep learning method based on an algorithm called 'You Only Look Once version 3' (YOLOv3). This innovative system predicts several boundary zones and class probabilities in a single evaluation from complete images, using a single CNN. This approach has considerably improved the accuracy and speed of cephalometric point identification, demonstrating the growing efficiency of artificial intelligence in orthodontic diagnosis (Redmon et al., 2016).

A comparative study was conducted in which an orthodontist manually identified 80 landmarks in 1311 images. Of these images, 1028 were used as training data using supervised learning techniques, and 283 images were used as test data. The tests were performed using two CNNs, the latest You Only Look Once version 3 (YOLOv3) and the Single Shot Multibox Detector (SSD), while SSD is based on a feed-forward convolutional network (W. Liu et al., 2016).

The results showed YOLOv3 outperformed SSD in terms of accuracy for 38 of the 80 landmarks identified, with the others showing no statistically significant differences. YOLOv3 took only 0.05 seconds to process each image, which is 2.84 seconds faster than SSD. In terms of reproducibility, YOLOv3 systematically detected identical positions for each landmark, clearly outperforming human analyses. Clearly, there have been significant improvements in recent deep learning methods. Furthermore, as we will see in the following sections, identification by this deep-learning model now achieve an accuracy comparable to an experienced orthodontist (Hwang et al., 2020; Park et al., 2019).

Further supporting the efficacy of YOLOv3, a study showed that YOLOv3 had smaller error ranges (Figure 9) and more isotropic error distributions compared to SSD, which had more elongated error ellipses (Figure 10), indicating less consistent accuracy. To clarify the difference between these two types of error, it is important to understand that the smaller error range means that the errors in localising the landmarks are generally smaller, in other words that the algorithm's predictions are closer to the actual positions of the landmarks. The smaller the error range, the more accurate the algorithm is considered to be.

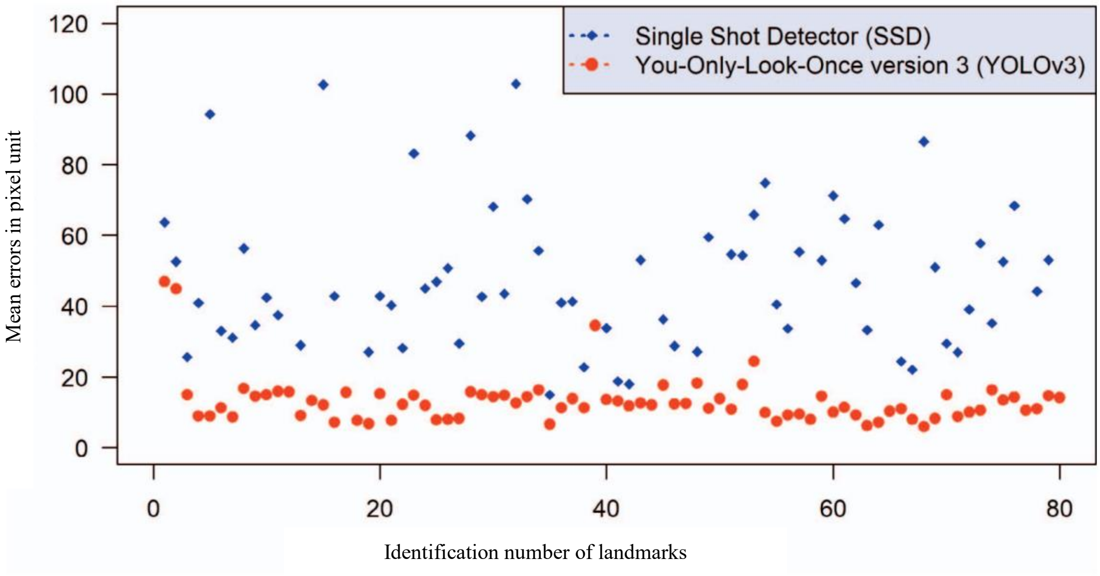


Figure 9 : Comparison of the mean errors between the two methods, SSD in blue and YOLOv3 in red (adapted from Park and al., 2019)

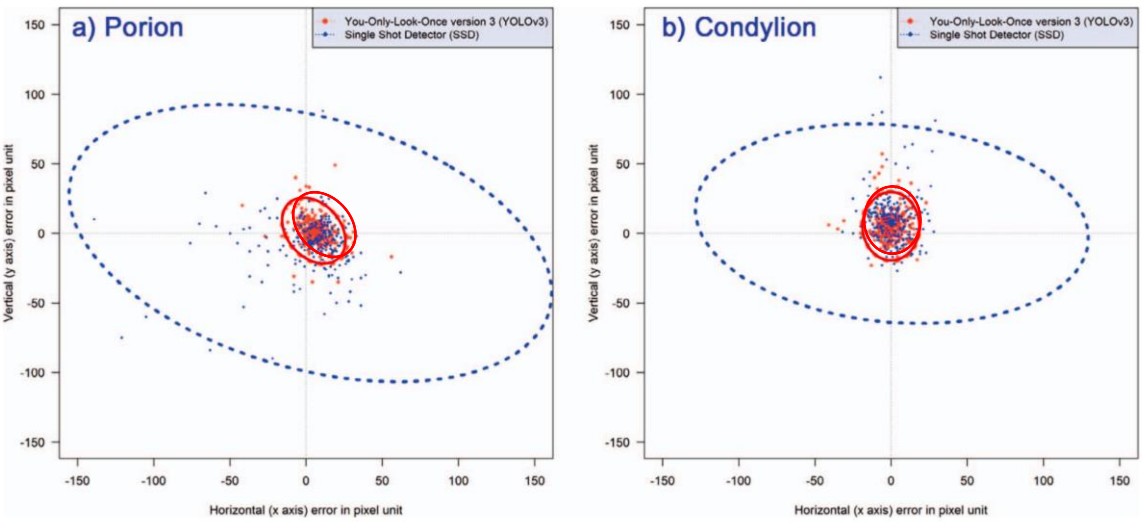


Figure 10 : Error scattergrams and 95% confidence ellipses after detecting porion and condylion from the SSD in blue and the YOLOv3 in red (adapted from Park and al., 2019)

On the other hand, the isotropic error distributions refer to the way in which the errors are distributed around the actual position of the landmarks. If the error distribution is isotropic, it means that the errors are uniformly distributed in all directions around the actual position of the point, indicating that the algorithm has no directional bias. This is an additional sign of strength and accuracy in landmark identification.

In addition, the computational efficiency of YOLOv3 makes it highly suitable for clinical practice and demonstrates its superiority in terms of speed and efficiency (Park et al., 2019).

The YOLOv3 algorithm highlights the potential of deep learning methods in clinical applications, setting a new standard for future advances in this field. This example demonstrates remarkable efficiency and accuracy in identifying cephalometric landmarks, making orthodontic diagnosis faster and more reliable.

2.1.2. AI combined with 3D imaging

Traditionally, in orthodontic diagnosis, the focus has been on hard tissues. This was originally believed to be reasonable because an orthodontist moves the teeth, which are included in hard tissues, and cephalograms, often used in diagnosis, show these hard tissues better than soft tissues. However, soft tissues play a crucial role in orthodontic diagnosis, as they directly influence the visual appearance of the face, a factor that is key to evaluating the success of treatment. Soft tissue analysis is therefore essential for understanding the overall impact on the face. With advances in 3D technology, it is now possible to analyse soft tissues more accurately, allowing improved assessment and monitoring of orthodontic results. The integration of these 3D soft tissue analyses helps to understand the aesthetic changes post-treatment and to meet the expectations of patients and their relatives. In this case, some points were premarked before scanning to provide a better result, the fronto-temporal point, soft-tissue gonion, and zygomatic points (Baik et al., 2007).

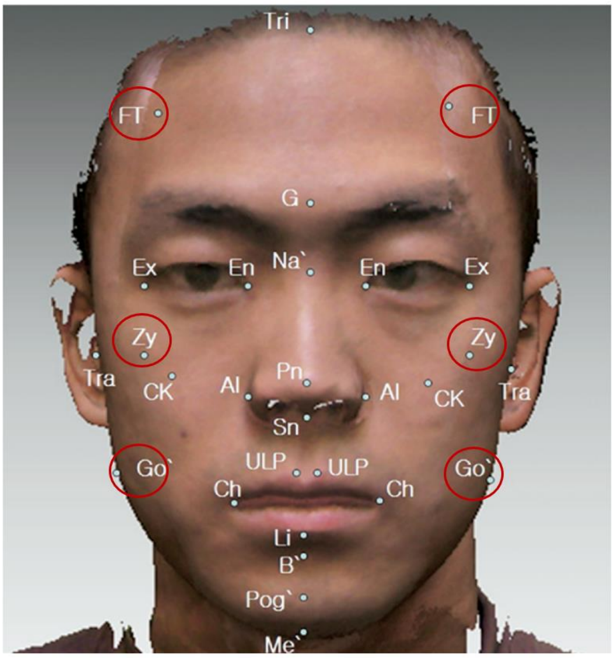


Figure 11 : Soft tissue landmarks (adapted from Baik and al., 2007)

In the assessment of facial aesthetics for example, 3D facial soft tissue analysis is essential for accurate orthodontic diagnosis (Figure 11). One study showed that using a 3D scanner on Korean adults identified significant differences, such as nasal prominence in Figure 12, making it easier to personalise treatments to improve facial aesthetics (Baik et al., 2007).

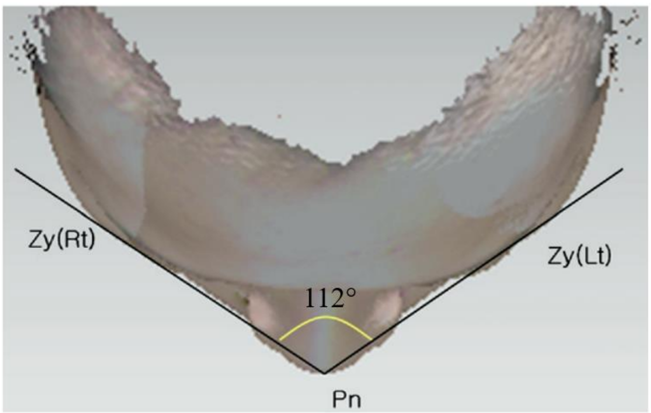


Figure 12 : Transverse nasal prominence (adapted from Baik and al., 2007)

The reproducibility of these facial landmarks is also a crucial aspect in orthodontics, as it determines the reliability of measurements and diagnoses.

In 2006, a study by Gwilliam et al. evaluated the reproducibility of 24 soft tissue landmarks on 3D facial images using both intra-operator and inter-operator assessments. The results showed that, when a single operator repeatedly placed the landmarks, 12 of the 24 landmarks had high reproducibility, with a standard deviation of less than 1 mm in all three planes of space (x, y, and z axes). However, reproducibility between different operators was significantly lower, with only two landmarks showing a standard deviation of less than 1 mm in all three axes, revealing significant variability in how different orthodontists place landmarks.

This inter-operator variability, due to factors such as the subjective perception of landmarks, variations in the angle of view, and the difficulty in identifying complex anatomical structures, can compromise the quality of diagnoses. This is where AI provides a valuable solution. By automating the identification of facial landmarks, AI eliminates human subjectivity and applies standardised criteria for locating landmarks, achieving accuracy with an average error of less than 1 mm, which is essential for ensuring reliable and reproducible diagnoses (Gwilliam et al., 2006).

By integrating AI into the diagnostic process, orthodontists can not only improve the accuracy of measurements, but also ensure consistency in diagnoses across different practitioners. This technological advance reduces human error and offers increased reliability, contributing to high quality orthodontic treatment (Gwilliam et al., 2006).

Dental aesthetic analysis is also essential for the evaluation of the impact of planned dental movements on the appearance of the smile. AI allows dental movements to be simulated and predicted using 3D imagery, giving orthodontists and patients an accurate overview of possible aesthetic outcomes before treatment even begins. This ability to predict the appearance of the smile allows clinicians to design treatment plans that maximise both function and aesthetics.

It is important to know that in orthodontics, achieving optimised aesthetic goals is as important as correcting functional malocclusions, as it is intrinsically linked to patient satisfaction and the overall success of orthodontic treatment (Hung et al., 2020).

The prediction of dental movements can also be applied to the impact on periodontal tissues. A detailed analysis of these tissues is essential to minimise the risk of periodontal complications during orthodontic treatment. This simulation ability is once again useful for anticipating potential adverse effects, such as bone loss or gingival recession, and for adjusting treatment plans as necessary (Hung et al., 2020).

AI can also play a role in the analysis of occlusal forces. Certain models, like CNNs for example, can accurately simulate occlusal forces even before orthodontic treatment begins, considering the interactions between soft tissues and bony structures. This accurate prediction of occlusal forces is a key factor in improving masticatory function and reducing the risk of recurrence or post-treatment malocclusion (Zhu et al., 2023).

For the analysis of temporomandibular joints (TMJs), AI combined with 3D imaging offers a detailed, fine-tuned evaluation of these complex structures. This technological advance allows early detection of temporomandibular disorders (TMD), which can lead to chronic pain, functional limitations and other severe complications.

A study evaluated the performance of AI algorithms in diagnosing TMDs. The results show that AI models can achieve a diagnostic accuracy of 91%, often outperforming traditional manual diagnostic methods. This increased accuracy is essential for detecting TMJ abnormalities at an early stage, allowing those patients to be identified and informed before treatment begins (Baik et al., 2007; N. Jha et al., 2022).

2.2. BENEFITS AND PERFORMANCE OF AI

2.2.1. Comparison of diagnostic efficiency between AI and human experts.

We can now ask whether the cephalometric analysis performed by artificial intelligence can be compared with the analysis performed by a human, and its performance in terms of clinical acceptance. The YOLOv3 model, which we have already examined, was compared with the human analysis. For this purpose, 80 landmarks, including two vertical references, as well as 46 landmarks for hard tissue and 32 for soft tissue, were evaluated. These landmarks were initially identified manually by an experienced examiner with 28 years of experience. The model was then trained on a test set of 283 images. Three months later, another examiner manually identified the same 80 landmarks on these same images.

The results showed that for 14 of the 46 skeletal landmarks, AI was more accurate, the human examiner was more accurate for another 14, while for the remaining 18, there was no significant difference. For the 32 soft tissue landmarks, AI demonstrated greater accuracy for 5 landmarks, compared with 7 for the human examiner. The mean error of detection between the AI and the human examiner was 1.46 ± 2.97 mm, compared with 1.50 ± 1.48 mm between the human examiners.

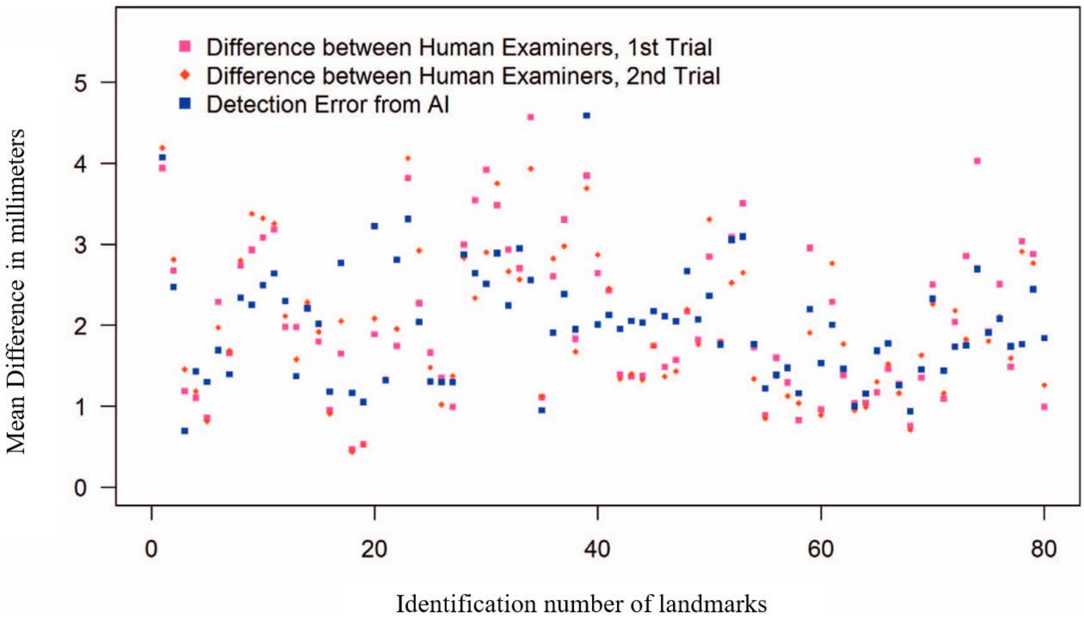


Figure 13 : Point plots summarizing the mean differences between human examiners and between AI vs humans (adapted from Hwang and al., 2020)

The AI demonstrated an ability to identify points with an accuracy comparable to experienced orthodontists, but it outperformed human examiners in terms of reproducibility, as the positions identified were systematically identical.

We can conclude that, although AI is not intended to completely replace trained specialists, it can nevertheless be used in a complementary way in the current clinical context especially in cephalometric analysis, under the supervision of expert orthodontists (Hwang et al., 2020).

In 2021, Y. H. Kim et al. conducted a study to compare the performance of AI with the variability between two experienced orthodontists, who annotated 13 specific landmarks. The distance detection error was used to establish a clinically acceptable error range for each point.

In this study, for 6 of the 13 landmarks, the successful detection rate based on expert variability was greater than 50%. This means that for these points, the AI location error is less or equal to the largest difference observed between the two orthodontists in at least 50% of cases.

In addition, 11 of the 13 landmarks had a general successful detection rate greater than 50% within a 2mm margin of error, illustrating the overall robust performance of the CNN model in this study. However, it is crucial to point out that for two points, the orbital and the porion, although the identified landmarks show high success detection rate values based on general accuracy ranges, their very low expert variability success detection rate could mean that they are still clinically unacceptable.

This demonstrates once again that, although AI can be a valuable tool and can sometimes surpass human capabilities, it nevertheless requires tight supervision (Y. H. Kim et al., 2021).

To support these observations, a systematic review by Junaid et al (2022) analysed 17 studies using deep learning-based artificial intelligence for cephalometric analysis. Most of the models used in these studies were CNNs, and two were based on the YOLOv3 system already mentioned.

Of these studies, 14 concluded that AI offered clinically acceptable diagnostic performance, while three observed no significant difference compared with the manual gold standard technique. These results highlight, once again, the potential of AI systems to provide a level of accuracy and precision that is equal to or better than those of trained orthodontists in cephalometric analysis (Junaid et al., 2022).

2.2.2. Improving the accuracy and speed of diagnostics

Although we recognise that AI cannot fully replace experienced clinicians, it is pertinent to discuss whether it can provide effective assistance to orthodontists. Panesar et al. (2023) conducted a comprehensive study to analyse the impact of human augmentation on the precision and accuracy of cephalometric analyses performed by deep learning. The study was conducted in several stages, initially four examiners of different levels of experience manually identified 31 landmarks on 30 cephalometric radiographs and they repeated this task a week later. The same radiographs were then retraced with the help of AI.

The software used incorporates four sub-systems, three of them being CNNs and the last one being a sub-system acting as a quality control. This method, called AI/human augmentation, begins with an initial localisation of the landmarks by the AI, followed by potential adjustments of the locations by the examiners.

The gold standard for this study was set by an orthodontist and a radiologist, both with 25 years' experience in their field. The results showed that with the assistance of AI, the accuracy of the examiners increased by 10.47%. In addition, accuracy was improved by 27.27%, in both cases, this improvement was especially marked in the less experienced examiners.

This study clearly illustrates that AI is a valuable tool for improving clinicians' accuracy and precision. It can also be used as an effective teaching tool for less experienced clinicians, providing them with accurate guidance and increasing their skill in identifying cephalometric points. These results highlight the potential of AI not only as a technical aid but also as a means of training and professional development in the orthodontic field (Panesar et al., 2023).

In addition to accuracy and precision, AI aims to improve the speed of analysis. We know that cephalometric analysis is essential for orthodontic diagnosis, however it is a laborious and time-consuming task for the clinician, which can lead to fatigue and errors. According to research, a single manual cephalometric tracing takes an average of 20 minutes for an experienced orthodontist. (Jiang et al., 2023; I.-H. Kim et al., 2021).

By comparison, as we saw earlier, the YOLOv3 algorithm can analyse 80 landmarks in 0.05 s. So even if the clinician may need to make adjustments, a completed cephalometric analysis can be done in around 1 min (Menezes et al., 2023; Park et al., 2019).

This is an excellent indicator of the remarkable time-saving potential of AI in this type of diagnostic analysis.

3. AI IN ORTHODONTIC DECISION MAKING AND TREATMENT PLANNING

In this new section, we will examine the application of AI in treatment planning and clinical decision-making in orthodontics. The application of AI in this field has opened new prospects for predicting the need for extractions and planning interventions. In particular, Artificial Neural Networks (ANNs) have demonstrated their ability to provide accurate and reliable predictions, optimising clinical decision-making and improving overall patient outcomes.

3.1. TRADITIONAL TO AI METHOD

3.1.1. Traditional method and its limits

The traditional orthodontic decision-making process is detailed and rigorous, based on clinical expertise, accurate radiographic analysis and careful planning (Subramanian et al., 2022).

As discussed in previous parts, cephalometric analysis is a crucial step in diagnosis. This analysis helps to determine the nature and extent of the anomalies and to plan the necessary dental movements (Leonardi et al., 2021).

Planning orthodontic treatment obviously involves determining specific objectives, such as correcting malocclusions, improving tooth alignment and improving smile aesthetics. A detailed treatment plan is then drawn up, specifying the type of orthodontic appliance to be used, the expected duration of the treatment and the specific steps to be followed, such as whether extractions are required, the extraction pattern or the anchorages needed, and is then discussed with the patient.

Treatment progress is monitored by regular visits, follow-up radiographs and ongoing evaluations to ensure goals are being met and to make adjustments if necessary (Leonardi et al., 2021; Subramanian et al., 2022).

However, traditional decision-making methods in orthodontics present several limitations and challenges. First, subjectivity and interpretative variability are important factors. Clinical decisions are highly dependent on the experience and expertise of the practitioner, which can lead to variations in diagnoses and treatment plans. Additionally, as discussed earlier, interpretations of radiographs and cephalometric analyses may vary from one orthodontist to another, resulting in different treatment recommendations for the same patient (Faber et al., 2019).

The process of data collection, analysis and planning is time-consuming and requires a lot of manual work. Periodic adjustments and regular monitoring also require a significant time commitment from practitioners and patients. Human error is another significant limitation (Subramanian et al., 2022).

Integrating different data sources, such as radiographs, models and medical history, into a coherent and comprehensive view can be challenging. Practitioners have to juggle large amounts of data to arrive at an informed decision, which again can be a source of error and cognitive overload. The lack of standardisation in diagnostic and treatment protocols between different practices and geographical regions leads to variability in clinical outcomes (Faber et al., 2019).

The challenges of traditional methods show the importance of the continuous evolution of practices and the adoption of modern technologies to improve the precision, efficiency and standardisation of orthodontic treatment. This is why AI is increasingly present in this field.

3.1.2. Improvements introduced by AI

In order to overcome the limitations of the so-called traditional methods, AI-based decision support systems (DSS) have been created and represent a significant advance in the field of orthodontics, offering valuable assistance to clinicians in their complex decision-making processes. DSS date back to the 1970s, but at the time they were poorly integrated, time-consuming and often restricted to academic institutions. Ethical and legal concerns relating to the use of computers in medicine, physician autonomy, and liability in the event of the use of a recommendation from a system of imperfect explicability were also major obstacles.

Since their first use, clinical DSS have evolved rapidly and are now commonly integrated into electronic medical records (EMRs) and other computerised clinical workflows (Sutton et al., 2020).

AI can be use in many different way, for example, aligner company claims that AI helps to optimize orthodontics planning, therefore those algorithms are industry secret so it is really hard to know where the truth is and where marketing begins (Faber et al., 2019).

3.2. AI MODELS USED IN DECISION-MAKING AND TREATMENT PLANNING

3.2.1. AI models used for extraction

We will now briefly review the evolution of algorithms over the last few years, in order to present their chronological development, their applications, and their improvements in terms of accuracy and efficiency.

We will start in 2009 with simple statistical models and decision trees. The use of decision trees has been explored to determine optimal sites for tooth extraction based on patient-specific dentoskeletal characteristics. The decision tree provided a structured and interpretable approach to orthodontic decision-making, providing the foundation for more complex models.

The model uses algorithms to evaluate different extraction scenarios and their potential impact on dental arch alignment. It incorporated patient-specific variables such as incisor protrusion, overjet and severity of dental crowding. This data guides the model's decision-making process by providing a detailed picture of the patient's dental structure.

This type of model achieved a coincidence rate of 86.0%, meaning that in 86% of cases, the model's recommendations matched the orthodontists' decisions. This high rate of agreement already demonstrates the reliability of the model and its potential use in clinical settings (Takada et al., 2009; Yagi et al., 2009).

One of the first ANN for orthodontic decision-making was described in 2010, with the aim of determining whether tooth extraction is necessary prior to orthodontic treatment, based on cast measurement, cephalometry and growth. This type of model is beneficial because it can deal with complex problems of uncertainty, non-configuration, non-linearity and multifactorial interactions.

The ANN model works by Back Propagation of errors, adjusting the weights of the connections to minimise the global error. This BP ANN had a 23-13-1 construction, meaning 23 inputs, for the 23 indices used, 13 hidden layers, and 1 output for the outcome of extraction or non-extraction.

Order	Input Index	Type of indice
1	Anterior teeth uncovered by incompetent lips	Nonquantification indice
2	Incisor Mandibular Plane Angle (LI-MP)	Hard tissue cephalometric indice
3	Overjet	Cast measurement indice
4	Crowding in the upper dental arch (mm)	Cast measurement indice
5	Space for correcting Spee's curve (mm)	Cast measurement indice
6	Overbite	Cast measurement indice
7	ANB angle	Hard tissue cephalometric indice
8	LI-NB	Hard tissue cephalometric indice
9	LI-E plane (mm)	Soft tissue cephalometric indice
10	Soft tissue convexity (angle Ns-Sn-Pos)	Soft tissue cephalometric indice
11	Interincisal angle (UI-LI)	Hard tissue cephalometric indice
12	UI-NA (mm)	Hard tissue cephalometric indice
13	Z angle	Soft tissue cephalometric indice
14	Wits (mm)	Hard tissue cephalometric indice
15	NLA (angle Cm-Sn-UL)	Soft tissue cephalometric indice
16	UI-SN	Hard tissue cephalometric indice
17	UL- E plane (mm)	Soft tissue cephalometric indice
18	UI-NA angle	Hard tissue cephalometric indice
19	Heredity	Nonquantification indice
20	Crowding in the lower dental arch (mm)	Cast measurement indice indice
21	Frankfort-Mandibular Incisor Angle (L1-FH)	Hard tissue cephalometric indice
22	LI-NB (mm)	Hard tissue cephalometric indice
23	Frankfort-Mandibular Angle (FH-MP)	Hard tissue cephalometric indice

Table 1 : Analysis of Input Index (adapted from Xie et al., 2010)

The results implied that 'anterior teeth uncovered by incompetent lips' and IMPA (L1-MP) were the two indices to be considered first. In terms of accuracy, the ANN was able to achieve a success rate of 80% (Xie et al., 2010).

In the mid-2010s, ANNs had gained more prominence, it is a neural network commonly used for classification and regression tasks. As we already know, they are made up of interconnected layers of nodes, or 'neurons', where each connection between nodes is associated with a weight. In 2016, a neural network model was presented to diagnose extraction needs in orthodontics (Jung & Kim, 2016).

This model uses supervised learning algorithms to predict extraction decisions based on 12 cephalometric measurements (Figure 14) and 6 additional indices (Table 2) which included the maxillary arch length index, mandibular arch length index, molar key index, excessive overbite index, protrusion index, and chief complaint index for protrusion.

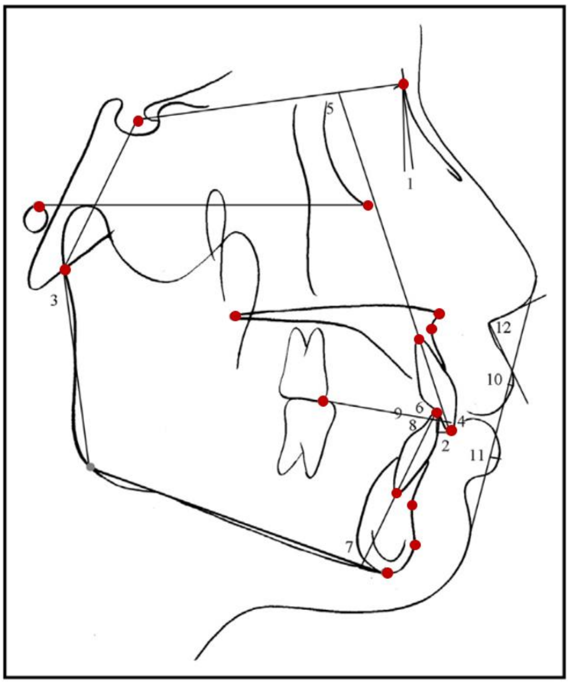


Figure 14 : 12 cephalometric measurements used in this study (adapted from Jung & Kim, 2016)

1	ANB angle
2	Overjet
3	Bjork sum
4	Overbite
5	Maxillary central incisor to SN angle
6	Maxillary central incisor to occlusal plane angle
7	Incisor mandibular plane angle
8	Mandibular central incisor to oclusal plane angle
9	Interincisal angle
10	Upper lip to E-line
11	Lower lip to E-line
12	Nasolabial angle

Index	*Weighting*	*Criterion (mm)*
Arch length discrepancy *ALD*		
Spacing	0	ALD > 0
Normal	0.25	$-1 < \text{ALD} \leq 0$
Mild crowding	0.5	$-3 < \text{ALD} \leq -1$
Moderate crowding	0.75	$-5 < \text{ALD} \leq -3$
Severe crowding	1	$\text{ALD} \leq -5$
Molar key		
Class III key	0	
Super Class I key	0.25	
Class I key	0.5	
End-on key	0.75	
Class II key	1	
Large overjet		
Not severe	0	Overjet ≤ 5
Severe	1	Overjet > 5
Protrusion		
Concave profile	0	
Normal profile	0.25	
Mild protrusion	0.5	
Moderate protrusion	0.75	
Severe protrusion	1	
Chief complaint for protrusion		
No protrusion in chief complaint	0	
Protrusion in chief complaint	1	

Table 2 : 6 additional indexes used in this study (adapted from Jung & Kim, 2016)

The ANN used here is a two-layer network (Figure 15), comprising a hidden layer with four hidden nodes, and a backpropagation algorithm was applied to adjust these weighted values according to the errors made during predictions. This process enables the accuracy of the model to be gradually improved.

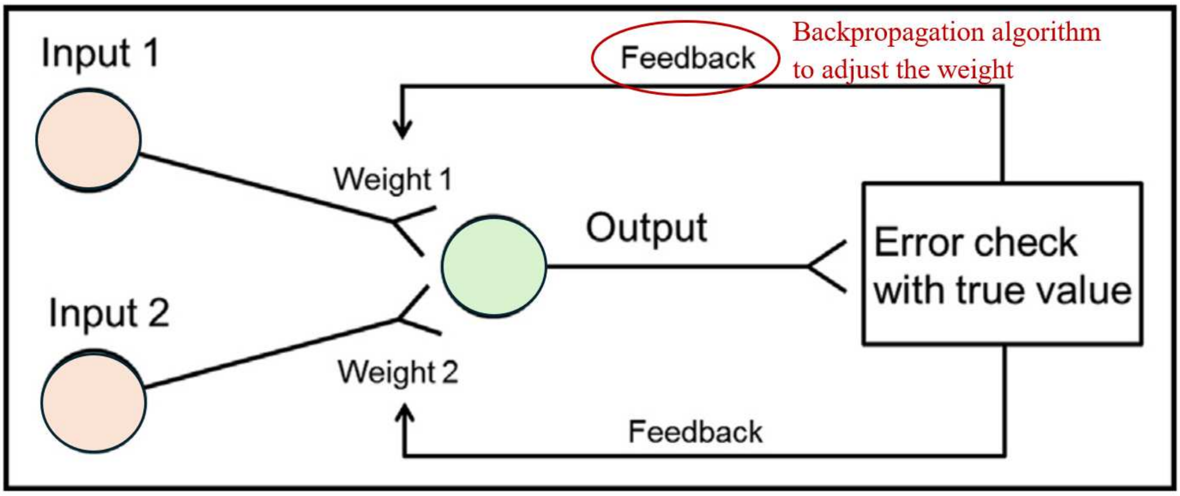


Figure 15 : Schematic diagram of the two-layer ANN (adapted from Jung & Kim, 2016)

The results showed that the artificial neural network model achieved a 93% success rate for extraction versus non-extraction diagnosis and an 84% success rate for detailed extraction pattern. Intraclass correlation coefficients (ICC) for assessing test-retest reliability of the plots ranged from 0.97 to 0.99, demonstrating excellent reliability.

Extraction decisions were classified according to three indices: Dx_ext (indicating the need for extractions), Dx_diff (indicating differential extractions between the maxillary and mandibular arches) and Dx_more (indicating the need for further retraction, which involves moving the teeth further backward). Success rates for these indices ranged from 84% to 96%.

This artificial neural network-based decision support system is very efficient and could be a valuable tool to help orthodontists make informed decisions about tooth extractions, as well as its potential clinical application in orthodontic decision-making (Jung & Kim, 2016).

Another type of model for decision support systems has been widely used in medical organisations, for example to help diagnose breast cancer, known as the Bayesian network (Choi et al., 2009). These models are very interesting because they can make inferences in many cases where data is missing. They also make it possible to understand the causal relationship between several factors.

In 2018, an evaluation was conducted to evaluate the system in analysing the need or not for orthodontic treatment, it has a much simpler output consisting just in a Need or No Need response. 20 intra- and extra-oral patient data points were selected to compare the results of two orthodontists and those generated by the decision support system. The BN model generated really promising results since the lowest degree of accuracy had a kappa value of 0.894, which can be characterised as almost perfect agreement between the system and the expert (Viera & GaSrrett, 2005).

For non-orthodontist dentists, analysing orthodontic treatment needs can sometimes be difficult, this model can therefore be seen as the first step in developing a decision support system for these practitioners (Thanathornwong, 2018).

In 2019 the use of an artificial neural network (ANN) model was detailed for orthodontic treatment planning. The model used here is also an ANN, more specifically a multilayer perceptron (MLP). Inputs to the model include cephalometric variables, demographic data and clinical data. The model also incorporates radiographic images, including cephalometric and panoramic radiographs, as well as 3D images such as cone-beam computed tomography (CBCT). In the case of missing data, a k-NN is used to estimate the missing values.

The outputs of this model provide predictions of extraction requirements, extraction patterns and anchorage patterns, each associated with their probability indicating the accuracy of the model.

Training is always carried out in the same way using supervised learning. Back-propagation was used to adjust the connection weights between neurons to minimise prediction error.

The ANN demonstrated an accuracy of 94.0% in extractions versus non-extraction decision-making, with an area under the curve (AUC) of 0.982, a sensitivity of 94.6% and a specificity of 93.8%. These results indicate excellent performance in identifying cases requiring extraction from those that do not. The model also showed an accuracy of 83.3% for extraction patterns and 92.8% for anchorage patterns, demonstrating its ability to provide detailed and personalised treatment recommendations.

The fact that the accuracy of the extraction patterns is lower can be explained by the fact that this is the most complicated part of the treatment, and each practitioner uses a specific pattern. To make it more applicable and to take this subjectivity into consideration, the model offers several alternatives for the doctor to choose from (Li et al., 2019).

In 2021, the performance of different machine learning models was compared to predict the need for tooth extraction in orthodontic patients.

- A MLP: with a backpropagation algorithm used to adjust the weights of the connections between neurons in order to minimise the prediction error.
  22 clinical and cephalometric variables are used as input features and the model achieved an accuracy of 79%.
- Random Forest (RF): this is an overall method that combines several decision trees to improve prediction accuracy. The RF was also trained with the same 22 features and showed slightly lower performance with an accuracy of 75%.

To improve model performance, a technique was introduced to identify and process incongruent data. Using this technique, the data samples were divided into two groups, G1 whose samples followed the general tendencies of the model, and G2 with samples that were more difficult to classify, representing a minority of the data. The MLP and RF models were re-trained on these separate groups, and the results showed a significant improvement in performance, particularly for the G1 group where the MLP achieved 96% accuracy.

These models show great promise for clinical application, although further improvements are required. This approach allows clinicians to recognise that some cases (G2) may require special attention and further analysis, while the majority of cases (G1) can be handled with more standardised machine learning models (Etemad et al., 2021).

In 2022, a machine learning model was developed using Auto-WEKA to predict the need for orthodontic extractions. Auto-WEKA is a machine learning software that selects and adjusts machine learning algorithms. The variables used to develop the predictive models include gender, dental model variables (such as overjet, overbite, molar class, etc.) and cephalometric data (vertical, sagittal, dental, as well as soft tissue) from 214 patients. These data were obtained from an anonymized database containing 314 cases treated by two experienced orthodontists, using different algorithms such as support vector machines, random forests and gradient boosting machines.

The tool performed a comprehensive search of possible configurations and selected the best model on the basis of cross-validation performance. Three settings were studied here, the first with clinical and Rx data, the second with clinical data only, and the last one with Rx data only.

Unsurprisingly, the most accurate model was based on the configuration that included both model variables and cephalometric data and achieved an accuracy of 93.9% using an MLP. However, it is also interesting to compare the other two models : performance was 15.4% better for the second setting, suggesting that model variables may be more relevant for extraction prediction than cephalometric data alone (Real et al., 2022).

In 2023, AI marked significant progress in decision making for tooth extraction and dental overlap categorisation, illustrated by two new key studies.

The first, by Kapoor et al., aimed to develop an AI model based on ANNs using feed-forward backpropagation method to reduce errors, to assist in tooth extraction decisions in borderline orthodontic cases. In this type of cases, the decisions to extract or not is quite difficult, and a wrong decision can lead to subsequent problems.

Using this back-propagation algorithm, the model was trained on data from 40 cases and tested on a further 20, showing a remarkable accuracy of 97.97%. The 30 diagnostic parameters included 13 cephalometric measurements, 12 cast measurements and 5 soft tissue parameters. These parameters were chosen according to the most consistent extraction vs non-extraction decision-making diagnostic criteria (Shukla et al., 2019).

The results showed high performance scores with an accuracy of 0.90 for extraction decisions and 0.80 for non-extractions. This study pointed out the importance of variables such as the angles of the lower incisor to the mandibular plane (IMPA), the upper incisor to the nasion-point A line (U1-NA degree), the esthetic line (E-line) and the nasolabial angle (NL-angle), which would be the most influential for extraction decisions (Kapoor et al., 2023).

The second study focused on the use of CNNs for dental landmark detection, overlap categorisation and extraction diagnosis, this time using intra-oral photography. Diagnosis is carried out as we saw previously with extraction of ROI windows from the input images, after these models were used for landmark detection, identifying the central point of the basal arch form and the mesial and distal points of each tooth.

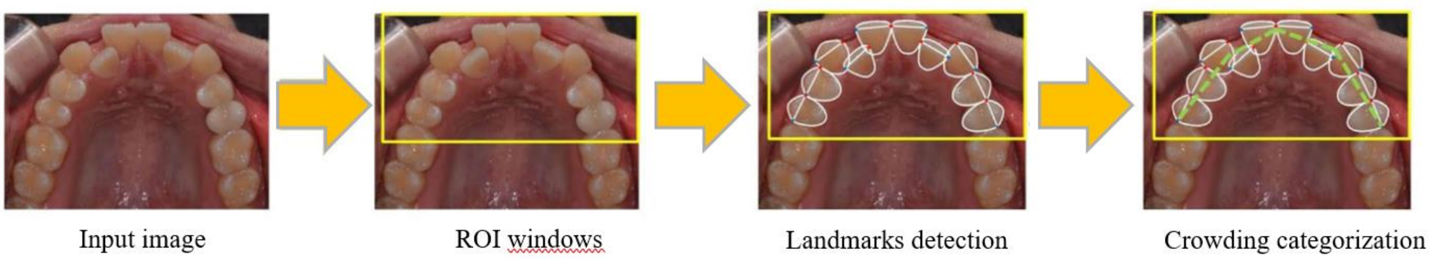


Figure 16 : Flow diagram of the processes (Ryu et al., 2023)

The models were then adapted for tooth extraction diagnosis. Transfer learning and fine-tuning techniques were applied to modify the fully connected classifier part to conform to the extraction diagnosis output.

Four CNNs models were evaluated: ResNet50, ResNet101, VGG16 and VGG19. These models were chosen for their efficiency in image processing tasks and their previous success in various medical and dental applications.

The first two are designed to overcome the degradation problem seen in deep neural networks by using jump connections, which allow the network to learn residual functions with reference to the layer inputs. This technique improves the model's ability to train deeper networks without significant loss of accuracy.

The last two are known for their simple architecture, which consists of very deep convolutional networks using small convolution filters. The depth of the network is increased to achieve better performance. VGG16 has 16 weight layers, while VGG19 has 19.

The results showed that the VGG19 model demonstrated the best performance with an accuracy of 0.922 and an AUC of 0.961 for the diagnosis of maxillary extraction. Detection of dental landmarks by VGG19 also showed the lowest mean errors.

These studies demonstrate that AI can not only improve the accuracy of orthodontic decisions but also reduce clinicians' subjectivity. Deep learning technologies make it possible to process large amounts of diagnostic data with high accuracy, facilitating more standardised and efficient decision-making (Ryu et al., 2023).

The evolution of AI algorithms in orthodontics illustrates the significant progress made over the last few decades. From early decision trees and Bayesian networks to sophisticated neural networks and deep learning models, AI has changed the approach to orthodontic treatment planning and decision-making. These advances have improved the accuracy, efficiency and reliability of clinical practice, providing orthodontists with valuable tools to improve patient outcomes.

3.2.2. Other example of software used in treatment planning

Orthodontic aligners, such as those used in the Invisalign® system, have become a major part of modern orthodontics thanks to the integration of digital technologies and AI. The aligners are custom-made, removable devices designed to move teeth gradually based on the patient's digital data. AI plays a key role in the planning and execution of these treatments.

Treatment with aligners begins with the collection of accurate clinical data from the patient, such as intra-oral scans, digital casts and cephalometric radiographs. This information is essential for creating a detailed 3D model of the patient's dentition. As the treatment plan is decided, the ClinCheck® software, in the case of Invisalign®, is designed to visualise the dental movements required to achieve the desired results, generating a virtual treatment plan, where each stage of tooth movement is decided until the final positions (Joffe, 2003).

The role of the ClinCheck® algorithm is crucial in this process. It uses machine learning algorithms to simulate and predict optimal movements, adjusting parameters such as movement speed and tooth biomechanics. The data is compared with thousands of similar cases to improve accuracy.

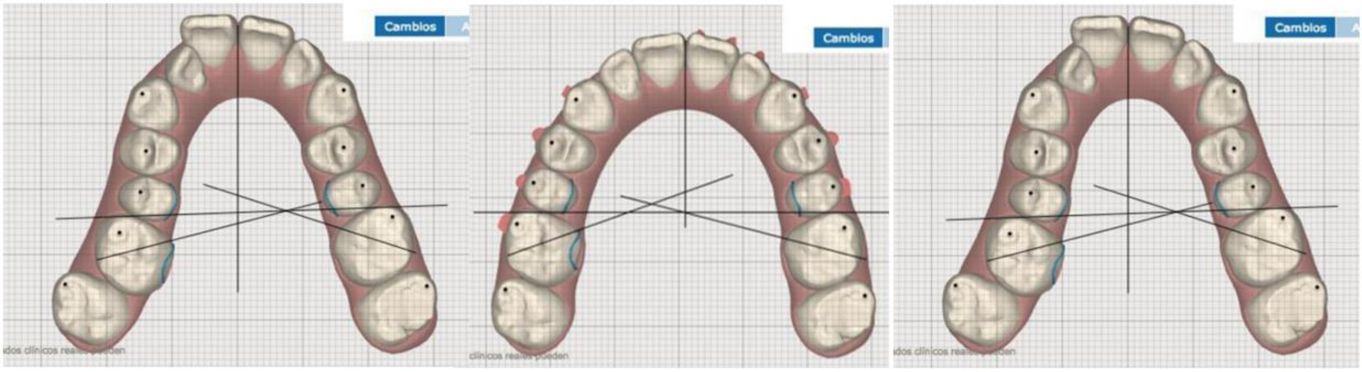
To explore the accuracy and effectiveness of dental movements performed with Invisalign® aligners on patients, a 2014 study examines three specific types of movement: incisor torque, premolar shifting and molar distalization. The authors also evaluated the influence of accessories, such as attachments which are temporarily bonded composite buttons and Power Ridges which are pressure lines close to the gingival margin, as well as the amount of movement by aligner on treatment efficacy.

Clinical results were obtained by scanning plaster models before and after treatment, then compared with predictions made by the software.

The study concludes that efficiency is highly dependent on the type of movement. While the average overall effectiveness of the movements was 59%, molar distalization, upper incisor torque and premolar shifting showed effectiveness of 87%, 42% and approximately 40% respectively. These results suggest that certain dental movements, particularly those involving torque and rotation, are less predictable with the use of the Invisalign® system. The results also show greater accuracy of movements with accessories in complex treatment plans (Simon et al., 2014).

Regarding transversal malocclusions and the movements predicted by the ClinCheck® software compared with the actual results after treatment, a 2020 study was carried out on a sample of 114 patients treated with the Invisalign® system. Arch expansion was measured at the canine, premolar and molar levels for both rotations and inclinations. Pre-treatment measurements (T1) were compared with ClinCheck® predicted results (T2) and with actual measurements at the end of the first phase of treatment (T3) (Figure 17).

The study shows that arch widths increased significantly after treatment. However, the actual results were generally lower than predicted by ClinCheck®. For example, an average expansion of 0.63 mm less was observed in the canines, 0.77 mm less in the first premolars, and 0.69 mm less in the first molars, compared with predictions.



T1 T2 T3

Figure 17 : Measurements taken at T1, T2, T3 (adapted from Morales-Burruezo et al.,2020)

Although the results showed that expansion in the premolar area was more accurate than in the canine and molar areas, ClinCheck® treatment plans tend to overestimate results.

In order to obtain results in line with expectations, it is essential that orthodontists adjust the plan according to the clinical results and the specific characteristics of each patient. In fact, overcorrection, which means predicting greater movements than necessary, during virtual treatment planning can compensate for the discrepancies between predictions and actual clinical results (Morales-Burruezo et al., 2020).

More recently, a study conducted in 2023 by Alswajy et al. evaluated the accuracy of this software in predicting dental movements before and after Invisalign® treatment. The main objective was once again to compare ClinCheck®'s predictions of tooth movement with the actual clinical results after treatment. The study focused on several dimensions: sagittal, vertical, transverse, and the length of the dental arch.

The study involved 206 patients, who all received full treatment with Invisalign® aligners. Data was collected using digital scanners, and the digital models obtained were compared with ClinCheck® predictions. The orthodontists taking part in the study had different levels of experience, allowing the robustness of the software's results to be tested in a variety of clinical settings. ClinCheck® numerical models were used for initial (pre-treatment), intermediate (predictions) and final (post-treatment clinical results) measurements.

The results show that ClinCheck® predictions were overall reliable, with an average accuracy of 76.85% for all dental movements. The highest reliability rates were for interincisal angle (96.23%) and intercanine width (97.97% for the upper arch and 97.67% for the lower arch). However, predictions of arch length were less reliable, with an accuracy of only 38.79% for the upper arch and 30.02% for the lower.

While widely used and effective for planning treatment with Invisalign® aligners, there are limitations to its predictions, particularly for more complex movements such as changes in arch length. This means that orthodontists must rely on their clinical expertise to adjust ClinCheck® predictions and ensure optimal results (Alswajy et al., 2023).

3.2.3. Dental Monitoring

Other software, such as Dental monitoring, allows patients to be monitored online. The system works by analysing images periodically taken by patients themselves using their smartphones. These images are then sent to the clinician for further evaluation, but before that, Dental Monitoring's AI performs a pre-analysis. The software compares the images with the pre-established treatment plan, monitors progress and detects any deviations. If a problem occurs, an alert is sent to the orthodontist, who can adjust the treatment accordingly. This automated monitoring capability improves treatment accuracy and reduces the need for office visits, optimising both patient comfort and clinical efficiency (Strunga et al., 2023).

The integration of artificial intelligence into orthodontic treatment with various software, such as ClinCheck® or Dental Monitoring, has undeniably transformed the profession.

While AI-driven tools offer accurate simulations and predictions, actual clinical results can sometimes deviate from these projections, particularly for more complex movements. Orthodontists must remain vigilant and use their clinical judgement to adjust treatment plans if necessary. In addition, a few challenges remain unresolved.

4. CHALLENGES AND PROSPECTS

4.1. CURRENT LIMITATIONS OF AI

The progress of AI in orthodontics is evident, but a number of challenges remain that could affect its clinical application. This section explores the current limitations and the futures prospects of AI in orthodontic diagnosis as well as in decision making and treatment planning.

4.1.1. In orthodontic diagnosis.

As we have seen, AI has shown promise in automated cephalometric analysis, but despite its progress, several limitations persist that could impact its clinical application.

First of all, the accuracy of AI in identifying cephalometric landmarks can be influenced by the quality of radiographic images. Poor image quality, the presence of metallic artefacts and variable resolutions can affect the performance of AI. The difficulties encountered by AI with poor quality images have also been observed with human examiners, which may explain why some studies have not shown a significant difference in AI accuracy as a consequence of these factors, however poor quality images can still be challenging in practical use (Y. H. Kim et al., 2021).

Another major challenge is the differences in imaging parameters, such as exposure, angle and resolution, which can affect the quality of 3D images and, consequently, the performance of AI models. The lack of standardisation in data gathering can introduce bias into diagnoses (Hung et al., 2020).

Likewise, traditional AI methods often analyse a limited number of landmarks, restricting their application to full cephalometric analyses. Although recent developments, such as YOLOv3, have expanded this to include up to 80 landmarks, there is still a need to extend their application to cover the full range of skeletal and soft tissue landmarks required for comprehensive diagnosis and effective treatment planning (Park et al., 2019).

Another limitation is the acceptable margin of error in the identification of landmarks. Automated systems have shown an average error of around 2 mm, however this may still be too high for some clinical applications. Human examiners can sometimes achieve higher accuracy, particularly in complex anatomical regions. The complexity of some regions may be caused by inter-patient variability (such as the gonion, protuberance menti and condylion) or by osseous complexity, proximity to other structures, and location on anatomical curves according to some researchers (like the A-point) (I.-H. Kim et al., 2021; Y. H. Kim et al., 2021).

On a more algorithmic level, in both radiographic and 3D analysis, AI systems are dependent on high-quality, well-annotated sets of data for model training, to achieve accuracy comparable to human experts, that can include cone-beam computed tomography (CBCT) and intra-oral and facial scans for 3D imaging for example. This requirement can be an obstacle, as acquiring and annotating large datasets is time-consuming and requires considerable resources, which limits the availability of robust AI models. In addition, as some experts have pointed out, the majority of AI research are conducted on specific and often geographically limited datasets, which can lead to problems of generalisation. It is necessary to have diverse and representative training data sets in order to guarantee the reliability of AI performance (Hung et al., 2020; Hwang et al., 2020; Willemink et al., 2020).

Also, some diagnostic subtleties still remain beyond the reach of AI algorithms, particularly in complex clinical cases or when image interpretation requires an advanced understanding of the clinical context. Which makes it a complementary tool to support orthodontists, rather than a complete substitute for human expertise (Hung et al., 2020).

Even with the significant advances in AI for cephalometric analysis and 3D imaging, limitations remain, and efforts must be made to improve image quality, increase the number of landmarks analysed, reduce margins of error and increase the diversity, volume and quality of training data. An additional clinical expertise is always necessary. Addressing these challenges is essential to developing the clinical adoption of AI in orthodontics.

4.1.2. In decision making and treatment planning

One of the main challenges is the quality and quantity of data available for training AI models. Data sets are often limited and heterogeneous, making it difficult to standardise models and reducing the reliability of predictions. AI algorithms require large amounts of diverse, high-quality data to produce reliable and generalisable results. Acquiring high-quality annotated data requires a lot of time and resources, which is a significant obstacle (J. Liu et al., 2021, 2023).

The lack of reproducibility of results is also a critical limitation. Model performance can vary depending on hyper-parameters such as learning rate, number of iterations and initialization strategies (Ryu et al., 2023). As a result, it is difficult to reproduce the results obtained by different research teams, which reduces confidence in the models. To improve reproducibility, rigorous sensitivity tests and greater transparency of the models' internal processes are needed. Also, data cleaning method could be an effective alternative to prevent manual errors in labelling measurement from affecting the reproducibility (J. Liu et al., 2021).

In the same spirit, the interpretability of AI models remains a major challenge. Clinicians need to understand how models arrive at their decisions in order to use them with confidence. AI models are often perceived as 'black boxes', which creates mistrust among potential users. Improving the visualisation of models' inner processes and conducting standardised clinical trials can help build user confidence (J. Liu et al., 2021, 2023; Ryu et al., 2023).

Overfitting is a common problem meaning that AI models perform well on training data but fail to generalise to unknow samples. This often results from insufficient diversity of clinical data in the training and test datasets. To avoid overfitting, techniques such as early stopping, dropout, and regularisation penalties can be used. Multimodal learning, which integrates various sources of information, is also proposed as a solution for dealing with imbalanced data (Kapoor et al., 2023; J. Liu et al., 2021, 2023).

Finally, the adoption of AI in orthodontics raises many ethical and legal issues. It is crucial to inform patients about the principles, functions and limitations of AI models. Transparency of AI systems is essential to maintain public trust, and decisions made by these systems must be explicable. In addition, responsibility for AI-influenced decisions must be clearly defined and supervised by healthcare professionals to ensure safe and effective care.

As mentioned previously, algorithms are often industry secrets, making it difficult to separate truth from marketing. In addition, AI could select treatment plans of patients who require less refinement, appearing to achieve orthodontic goals more efficiently. This selective optimisation is part of the ethical concerns, as treatment plans become the property of aligner companies, which could prioritise profit over patient-centred care. The problem lies not only in the transparency of AI algorithms, but also in the potential moral implications, as AI could favour cases that deliver faster and more marketable outcomes, rather than universally accessible treatment solutions (Faber et al., 2019; Naeem et al., 2023; Strunga et al., 2023).

The increasing industrialisation of orthodontic treatment, particularly with the use of standardised technologies such as Invisalign® aligners and software such as ClinCheck®, presents an increased risk of iatrogenic effects. Although these technologies make it possible to improve the efficiency and speed of treatment, they can also lead to complications when treatment plans are not sufficiently personalised to each patient's specific needs. Among the most common iatrogenic effects are root resorption, occlusal imbalances and facial asymmetries, often caused by the uniform application of orthodontic forces. These complications underline the need for orthodontists not to rely exclusively on technological tools, but to combine clinical expertise with these technologies to ensure optimal results and minimise iatrogenic risks (Barreto et al., 2016).

Health data protection is another major concern, and it is important to be able to ensure the confidentiality of patient data. The combination of blockchain technology and federated learning is expected to facilitate data sharing without compromising data privacy. Ethical implications also include the need to guarantee equity and minimise bias in AI models, in order to avoid discrimination and provide equal access to healthcare technologies. A validated and ethical approach is needed to ensure that AI improves healthcare safely, effectively and equitably. (Kapoor et al., 2023; J. Liu et al., 2023; Mörch et al., 2021).

4.2. FUTURES PROSPECTS AND EXPECTED DEVELOPMENTS

Despite the current challenges, the future prospects for AI in orthodontics are vast and promising. Recent advances point the way to new applications and clinical improvements. Using deep learning and computer vision algorithms, AI facilitates complex tasks such as automatic cephalometric analysis, segmentation of anatomical structures from CBCT images, and prediction of dental extraction needs (Faber et al., 2019; Leonardi et al., 2021).

To begin with, as the effectiveness of AI systems is directly linked to the quality and quantity of training data, increasing the diversity and volume of this data could improve the strength and accuracy of AI models in the future. For example the use of over 250,000 images from various sources of X-rays can enhance the AI's ability to generalize across different patient populations and conditions (Subramanian et al., 2022).

Multimodal data integration represents a promising advance in the use of AI in orthodontics. Future AI systems could integrate data from a various imaging modality, such as computed tomography (CT), magnetic resonance imaging (MRI) and standard radiography. This multimodal approach could improve the accuracy of landmark identification and provide a more complete view of anatomical structures, which is essential for orthodontic treatment planning. One study demonstrated the usefulness of combining CBCT images to improve the accuracy of 3D cephalometric landmarks (Bao et al., 2023; J. Liu et al., 2021).

In terms of clinical improvement, AI has significantly improved the accuracy and precision of cephalometric landmark identifications by itself alone, reaching precision with an average intraclass correlation coefficient (ICC) of 0.97, as we saw earlier. Deep learning algorithms, such as CNNs, can automatically identify these points with an accuracy comparable to, or even better than that of human experts, with also reducing errors due to clinician subjectivity.

Furthermore, the combination of AI and human expertise, known as AI/human augmentation, improves this precision and accuracy even further for dental professionals. This collaborative approach significantly reduces the mean absolute error (MAE), particularly for less experienced clinicians, allowing them to benefit from AI recommendations while making adjustments based on their expertise (J. Liu et al., 2023; Panesar et al., 2023).

AI also plays a crucial role in decision support for tooth extraction. ANN network models help to decide whether tooth extractions are necessary prior to orthodontic treatment. These systems analyse cephalometric variables to provide accurate recommendations, thereby reducing the variability of decisions between orthodontists (Bichu et al., 2021).

Similarly, AI-based clinical decision support systems can provide detailed treatment protocols, improving consistency and quality of care, particularly for less experienced orthodontists, and help improve patients' understanding of their treatments (Faber et al., 2019; Jung & Kim, 2016).

For example, these systems can guide the treatment of severe malocclusions such as deep overbites, with a high accuracy (Al Turkestani et al., 2021).

AI offers substantial gains in accuracy, speed and reliability, making it possible to transform cephalometric analysis. With continued advances in deep learning algorithms, multimodal integration and open collaboration, combined with the increasing quality and diversity of training data, AI promises to play a central role in future orthodontic diagnosis and treatment. The full integration of facial analysis, medical history and functional considerations into AI models will also provide more comprehensive and effective treatment plans. These innovations will not only improve clinical outcomes, but also increase patient confidence in AI-assisted orthodontic treatment (Al Turkestani et al., 2021; Bichu et al., 2021; Evangelista et al., 2022. J. Liu et al., 2021, 2023).

III- CONCLUSION

The integration of AI into the field of orthodontics marks a significant advance in diagnosis and treatment planning. Throughout this work, we explored various applications of AI, highlighting its ability to surpass, or at least equal, human performance in certain critical tasks, such as identifying cephalometric landmarks and predicting the need for tooth extractions.

ANN, particularly CNNs, have demonstrated a remarkable ability to automate the identification of cephalometric landmarks and the analysis of images, making diagnoses both faster and more reliable.

AI technologies combined with 3D scans have also shown notable advances in orthodontic analysis, providing a more complete view of facial and dental anatomy, once again improving the accuracy of diagnoses.

These technologies not only improve the speed of diagnosis, but also standardise processes, reducing variation between practitioners and increasing consistency of treatment. It also offers learning and development opportunities for young orthodontists.

Advanced treatment planning and simulation give clinicians the ability to anticipate outcomes and design more personalised plans, not only reducing the risk of error, but also optimising aesthetic and functional outcomes for patients.

Similarly, traditional decision-making processes in orthodontics are frequently limited by the subjectivity and variability of interpretations, requiring practitioners to make a significant investment in terms of time and effort. Integrating AI into these processes overcomes these limitations by providing more accurate, rapid and standardised analysis. AI models, including ANNs, can analyse complex cephalometric variables to provide detailed and personalised treatment recommendations, reducing the variability of decisions between orthodontists.

However, despite these advances, the integration of AI into orthodontic clinical practice is not without its challenges. The quality of radiographic images, the diversity and quantity of training data, the reproducibility and interpretability of AI models, along with the complexity of some clinical interpretations where subtleties still remain beyond the reach of current algorithms, are crucial aspects that require improvement to ensure widespread adoption of these technologies in orthodontics. Continued improvements in deep learning algorithms and multimodal data integration are expected to further reinforce the abilities of AI.

These developments will offer more accurate and personalised treatment plans, while increasing patient and practitioner confidence in AI systems. In addition, the ethical and legal concerns associated with the use of AI in medicine, such as patient data confidentiality and algorithm transparency, require particular consideration.

In conclusion, AI is proving to be a powerful tool in the transformation of orthodontic practice. This process will require close collaboration between researchers, practitioners and regulators to ensure that AI truly improves the quality of orthodontic care, while preserving the fundamental principles of medicine, namely autonomy, beneficence, non-maleficence and justice.

Bibliography

Al Turkestani, N., Bianchi, J., Deleat-Besson, R., Le, C., Tengfei, L., Prieto, J. C., Gurgel, M., Ruellas, A. C. O., Massaro, C., Aliaga Del Castillo, A., Evangelista, K., Yatabe, M., Benavides, E., Soki, F., Zhang, W., Najarian, K., Gryak, J., Styner, M., Fillion-Robin, J., … Cevidanes, L. H. S. (2021). Clinical decision support systems in orthodontics : A narrative review of data science approaches. *Orthodontics & Craniofacial Research*, *24*(S2), 26-36. https://doi.org/10.1111/ocr.12492

Alswajy, W., Baeshen, H., Al-Turki, G., & Alsulaimani, F. (2023). The Reliability of ClinCheck® Accuracy before and after Invisalign® Treatment—A Multicenter Retrospective Study. *Applied Sciences*, *13*(8), 4670. https://doi.org/10.3390/app13084670

Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An Introduction to Machine Learning. *Clinical Pharmacology & Therapeutics*, *107*(4), 871-885. https://doi.org/10.1002/cpt.1796

Baik, H.-S., Jeon, J.-M., & Lee, H.-J. (2007). Facial soft-tissue analysis of Korean adults with normal occlusion using a 3-dimensional laser scanner. *American Journal of Orthodontics and Dentofacial Orthopedics*, *131*(6), 759-766. https://doi.org/10.1016/j.ajodo.2005.08.038

Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, *3*, 100071. https://doi.org/10.1016/j.dajour.2022.100071

Bao, H., Zhang, K., Yu, C., Li, H., Cao, D., Shu, H., Liu, L., & Yan, B. (2023). Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence. *BMC Oral Health*, *23*(1), 191. https://doi.org/10.1186/s12903-023-02881-8

Barreto, G. M., Feitosa, H. O., Universidade de São Paulo, Brazil, & Universidade Cidade de São Paulo, Brazil. (2016). Iatrogenics in Orthodontics and its challenges. *Dental Press Journal of Orthodontics*, *21*(5), 114-125. https://doi.org/10.1590/2177-6709.21.5.114-125.sar

Bichu, Y. M., Hansa, I., Bichu, A. Y., Premjani, P., Flores-Mir, C., & Vaid, N. R. (2021). Applications of artificial intelligence and machine learning in orthodontics : A scoping review. *Progress in Orthodontics*, *22*(1), 18. https://doi.org/10.1186/s40510-021-00361-9

Choi, J. P., Han, T. H., & Park, R. W. (2009). A Hybrid Bayesian Network Model for Predicting Breast Cancer Prognosis. *Journal of Korean Society of Medical Informatics*, *15*(1), 49. https://doi.org/10.4258/jksmi.2009.15.1.49

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms* (Third edition). MIT Press.

Davenport, T., & Kalakota, R. (2019). *The potential for artificial intelligence in healthcare*. Future Healthcare Journal 2019 Vol 6, No 2: 94–8

Dong, H., Ding, Z., & Zhang, S. (Éds.). (2020). *Deep Reinforcement Learning : Fundamentals, Research and Applications*. Springer Singapore. https://doi.org/10.1007/978-981-15-4095-0

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118. https://doi.org/10.1038/nature21056

Etemad, L., Wu, T., Heiner, P., Liu, J., Lee, S., Chao, W., Zaytoun, M. L., Guez, C., Lin, F., Jackson, C. B., & Ko, C. (2021). Machine learning from clinical data sets of a contemporary decision for orthodontic tooth extraction. *Orthodontics &* Craniofacial Research, 24(S2), 193-200. https://doi.org/10.1111/ocr.12502

Evangelista, K., Santos de Freitas Silva, B., Yamamoto-Silva, F. P., Valladares-Neto, J. Alves Garcia Silva, M., Soares Cevidanes, L. H., Luca Canto, G., Massignan, C. (2022). Accuracy of artificial intelligence for tooth extraction decision-making in orthodontics: a systematic review and meta-analysis. Clinical Oral Investigations 26, 6893–6905. https://doi.org/10.1007/s00784-022-04742-0

Faber, J., Faber, C., & Faber, P. (2019). Artificial intelligence in orthodontics. *APOS Trends in Orthodontics*, *9*, 201-205. https://doi.org/10.25259/APOS_123_2019

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Gwilliam, J. R., Cunningham, S. J., & Hutton, T. (2006). Reproducibility of soft tissue landmarks on three-dimensional facial scans. *The European Journal of Orthodontics*, *28*(5), 408-415. https://doi.org/10.1093/ejo/cjl024

Hung, K., Yeung, A. W. K., Tanaka, R., & Bornstein, M. M. (2020). Current Applications, Opportunities, and Limitations of AI for 3D Imaging in Dental Research and Practice. *International Journal of Environmental Research and Public Health*, *17*(12), 4424. https://doi.org/10.3390/ijerph17124424

Hwang, H.-W., Park, J.-H., Moon, J.-H., Yu, Y., Kim, H., Her, S.-B., Srinivasan, G., Aljanabi, M. N. A., Donatelli, R. E., & Lee, S.-J. (2020). Automated identification of cephalometric landmarks : *Part 2- Might It Be Better Than human? The Angle Orthodontist*, *90*(1), 69-76. https://doi.org/10.2319/022019-129.1

jeferson.zambrano. (2022). *Machine Learning vs Artificial Intelligence : What's the Difference? | MIT Professional Education*. https://professionalprograms.mit.edu/blog/technology/machine-learning-vs-artificial-intelligence/

Jha, N., Lee, K., & Kim, Y.-J. (2022). Diagnosis of temporomandibular disorders using artificial intelligence technologies : A systematic review and meta-analysis. *PLOS ONE*, *17*(8), e0272715. https://doi.org/10.1371/journal.pone.0272715

Jha, S., & Topol, E. J. (2016). Adapting to artificial intelligence : Radiologists and Pathologists as Information Specialists. *Journal of the American Medical Association*, *316*(22), 2353. https://doi.org/10.1001/jama.2016.17438

Jiang, F., Guo, Y., Yang, C., Zhou, Y., Lin, Y., Cheng, F., Quan, S., Feng, Q., & Li, J. (2023). Artificial intelligence system for automated landmark localization and analysis of cephalometry. *Dentomaxillofacial Radiology*, *52*(1), 20220081. https://doi.org/10.1259/dmfr.20220081

Joffe, L. (2003). Invisalign ® : Early experiences. *Journal of Orthodontics*, *30*(4), 348-352. https://doi.org/10.1093/ortho/30.4.348

Junaid, N., Khan, N., Ahmed, N., Abbasi, M. S., Das, G., Maqsood, A., Ahmed, A. R., Marya, A., Alam, M. K., & Heboyan, A. (2022). Development, application, and performance of artificial intelligence in cephalometric landmark identification and diagnosis : A systematic review. *Healthcare*, *10*(12), 2454. https://doi.org/10.3390/healthcare10122454

Jung, S.-K., & Kim, T.-W. (2016). New approach for the diagnosis of extractions with neural network machine learning. *American Journal of Orthodontics and Dentofacial Orthopedics*, *149*(1), 127-133. https://doi.org/10.1016/j.ajodo.2015.07.030

Kapoor, S., Shyagali, T. R., Kuraria, A., Gupta, A., Tiwari, A., & Goyal, P. (2023). An artificial neural network approach for rational decision-making in borderline orthodontic cases : A preliminary analytical observational in silico study. *Journal of Orthodontics*, *50*(4), 439-448. https://doi.org/10.1177/14653125231172527

Kelleher, J. D. (2019). *Deep learning*. The MIT Press.

Kim, I.-H., Kim, Y.-G., Kim, S., Park, J.-W., & Kim, N. (2021). Comparing intra-observer variation and external variations of a fully automated cephalometric analysis with a cascade convolutional neural net. *Scientific Reports*, *11*(1), 7925. https://doi.org/10.1038/s41598-021-87261-4

Kim, Y. H., Lee, C., Ha, E.-G., Choi, Y. J., & Han, S.-S. (2021). A fully deep learning model for the automatic identification of cephalometric landmarks. *Imaging Science in Dentistry*, *51*(3), 299. https://doi.org/10.5624/isd.20210077

Kitson, N. K., Constantinou, A. C., Guo, Z., Liu, Y., & Chobtham, K. (2023). A survey of Bayesian Network structure learning. *Artificial Intelligence Review*, *56*(8), 8721-8814. https://doi.org/10.1007/s10462-022-10351-w

Leonardi, R., Giudice, A. L., Isola, G., & Spampinato, C. (2021). Deep learning and computer vision : Two promising pillars, powering the future in orthodontics. *Seminars in Orthodontics*, *27*(2), 62-68. https://doi.org/10.1053/j.sodo.2021.05.002

Li, P., Kong, D., Tang, T., Su, D., Yang, P., Wang, H., Zhao, Z., & Liu, Y. (2019). Orthodontic treatment planning based on artificial neural networks. *Scientific Reports*, *9*(1), 2037. https://doi.org/10.1038/s41598-018-38439-w

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, *16*(6), 321-332. https://doi.org/10.1038/nrg3920

Liu, J., Chen, Y., Li, S., Zhao, Z., & Wu, Z. (2021). Machine learning in orthodontics : challenges and perspectives. *Advances in Clinical and Experimental Medicine*, *30*(10), 1065-1074. https://doi.org/10.17219/acem/138702

Liu, J., Zhang, C., & Shan, Z. (2023). Application of artificial intelligence in orthodontics : current state and future perspectives. *Healthcare*, *11*(20), 2760. https://doi.org/10.3390/healthcare11202760

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD : Single Shot MultiBox Detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Éds.), *Computer Vision – ECCV 2016* (Vol. 9905, p. 21-37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2

Mahesh, B. (2018). *Machine Learning Algorithms—A Review*. *9*(1).

Matteis, L. D. (2022). *Introduction à l'apprentissage*.

Menezes, L. D. S., Silva, T. P., Lima Dos Santos, M. A., Hughes, M. M., Mariano Souza, S. D. R., Leite Ribeiro, P. M., Freitas, P. H. L. D., & Takeshita, W. M. (2023). Assessment of landmark detection in cephalometric radiographs with different conditions of brightness and contrast using the an artificial intelligence software. *Dentomaxillofacial Radiology*, *52*(8), 20230065. https://doi.org/10.1259/dmfr.20230065

Morales-Burruezo, I., Gandía-Franco, J.-L., Cobo, J., Vela-Hernández, A., & Bellot-Arcís, C. (2020). Arch expansion with the Invisalign system : efficacy and predictability. *PLOS ONE*, *15*(12), e0242979. https://doi.org/10.1371/journal.pone.0242979

Mörch, C. M., Atsu, S., Cai, W., Li, X., Madathil, S. A., Liu, X., Mai, V., Tamimi, F., Dilhac, M. A., & Ducret, M. (2021). Artificial intelligence and ethics in dentistry : A Scoping Review. *Journal of Dental Research*, *100*(13), 1452-1460. https://doi.org/10.1177/00220345211013808

Naeem, M. M., Sarwar, H., Hassan, M. T., Balouch, N. M., Singh, S. P., Essrani, P. D., & Rajper, P. (2023). Exploring the ethical and privacy implications of artificial intelligence in dentistry. *International Journal of Health Sciences*, *7*(S1), 904-915. https://doi.org/10.53730/ijhs.v7nS1.14294

Nancholas, B. (2023, septembre 1). *Narrow artificial intelligence : Advantages, disadvantages, and the future of AI*. University of Wolverhampton. https://online.wlv.ac.uk/narrow-artificial-intelligence-advantages-disadvantages-and-the-future-of-ai/

Nyuytiymbiy, K. (2022, mars 28). *Parameters and Hyperparameters in Machine Learning and Deep Learning*. Medium. https://towardsdatascience.com/parameters-and-hyperparameters-aa609601a9ac

Panesar, S., Zhao, A., Hollensbe, E., Wong, A., Bhamidipalli, S. S., Eckert, G., Dutra, V., & Turkkahraman, H. (2023). Precision and accuracy assessment of cephalometric analyses performed by deep learning artificial intelligence with and without human augmentation. *Applied Sciences*, *13*(12), 6921. https://doi.org/10.3390/app13126921

Park, J.-H., Hwang, H.-W., Moon, J.-H., Yu, Y., Kim, H., Her, S.-B., Srinivasan, G., Aljanabi, M. N. A., Donatelli, R. E., & Lee, S.-J. (2019). Automated identification of cephalometric landmarks : Part 1—Comparisons between the latest deep-learning methods YOLOV3 and SSD. *The Angle Orthodontist*, *89*(6), 903-909. https://doi.org/10.2319/022019-127.1

Real, A. D., Real, O. D., Sardina, S., & Oyonarte, R. (2022). Use of automated artificial intelligence to predict the need for orthodontic extractions. *Korean Journal of Orthodontics*, *52*(2), 102-111. https://doi.org/10.4041/kjod.2022.52.2.102

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once : Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. https://doi.org/10.1109/CVPR.2016.91

Ryu, J., Kim, Y.-H., Kim, T.-W., & Jung, S.-K. (2023). Evaluation of artificial intelligence model for crowding categorization and extraction diagnosis using intraoral photographs. *Scientific Reports*, *13*(1), 5177. https://doi.org/10.1038/s41598-023-32514-7

Shen, D., Wu, G., & Suk, H.-I. (2017). *Deep Learning in Medical Image Analysis*.

Shukla, P., Thakral, R., Gupta, S., & Nagar, A. (2019). Borderline cases in orthodontics-A review. *IP Indian Journal of Orthodontics and Dentofacial Research*, *5*(3), 78-83. https://doi.org/10.18231/j.ijodr.2019.017

Simon, M., Keilig, L., Schwarze, J., Jung, B. A., & Bourauel, C. (2014). Treatment outcome and efficacy of an aligner technique – regarding incisor torque, premolar derotation and molar distalization. *BMC Oral Health*, *14*(1), 68. https://doi.org/10.1186/1472-6831-14-68

Strunga, M., Urban, R., Surovková, J., & Thurzo, A. (2023). Artificial intelligence systems assisting in the assessment of the course and retention of orthodontic treatment. *Healthcare*, *11*(5), 683. https://doi.org/10.3390/healthcare11050683

Subramanian, A. K., Chen, Y., Almalki, A., Sivamurthy, G., & Kafle, D. (2022). Cephalometric analysis in orthodontics using artificial intelligence—A comprehensive review. *BioMed Research International*, *2022*, 1-9. https://doi.org/10.1155/2022/1880113

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems : Benefits, risks, and strategies for success. *Npj Digital Medicine*, *3*(1), 17. https://doi.org/10.1038/s41746-020-0221-y

Takada, K., Yagi, M., & Horiguchi, E. (2009). Computational formulation of orthodontic tooth-extraction decisions. *The Angle Orthodontist*, *79*(5), 885-891. https://doi.org/10.2319/081908-436.1

Thanathornwong, B. (2018). Bayesian-based decision support system for assessing the needs for orthodontic treatment. *Healthcare Informatics Research*, *24*(1), 22. https://doi.org/10.4258/hir.2018.24.1.22

Topol, E. (2019). *Deep Medicine : How Artificial Intelligence Can Make Healthcare Human Again*. Hachette UK.

Topol, E. J. (2019). High-performance medicine : The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Viera, A. J., & Garrett, J. M. (2005). Understanding Interobserver Agreement : The Kappa Statistic. *Family Medicine*.

What is Artificial Intelligence (AI)? | IBM. (s. d.). Consulté 26 février 2024, à l'adresse https://www.ibm.com/topics/artificial-intelligence

Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L. R., Summers, R. M., Rubin, D. L., & Lungren, M. P. (2020). Preparing medical imaging data for machine learning. *Radiology*, *295*(1), 4-15. https://doi.org/10.1148/radiol.2020192224

Xie, X., Wang, L., & Wang, A. (2010). Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. *The Angle Orthodontist*, *80*(2), 262-266. https://doi.org/10.2319/111608-588.1

Yagi, M., Ohno, H., & Takada, K. (2009). Computational formulation of orthodontic tooth-extraction decisions. *The Angle Orthodontist*, *79*(5), 892-898. https://doi.org/10.2319/081908-439.1

Yang, G.-Z., Cambias, J., Cleary, K., Daimler, E., Drake, J., Dupont, P. E., Hata, N., Kazanzides, P., Martel, S., Patel, R. V., Santos, V. J., & Taylor, R. H. (2017). Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy. *Science Robotics*, *2*(4), eaam8638. https://doi.org/10.1126/scirobotics.aam8638

Zhou, Z.-H. (2021). *Machine learning* (S. Liu, Trad.). Springer.

Zhu, J., Yang, Y., & Wong, H. M. (2023). Development and accuracy of artificial intelligence-generated prediction of facial changes in orthodontic treatment : A scoping review. *Journal of Zhejiang University-SCIENCE B*, *24*(11), 974-984. https://doi.org/10.1631/jzus.B2300244