REVIEW ARTICLE

Werner Schupp, Asad Abu-Tarif, Julia Haubrich, Fayez Elkholy, James Mah, Karl-Friedrich Krey

Artificial intelligence in orthodontics: Part 1

Werner Schupp

KEY WORDS *aligner orthodontics, artificial intelligence, convolutional neural networks, deep learning, machine learning, neural networks*

Artificial intelligence is now firmly established in society. Whether we are searching on Google (Mountain View, CA, USA), being guided by recommendation algorithms or using facial recognition or smart home software, artificial intelligence is a daily presence in our lives, and in almost all areas: business, science, medicine, and increasingly in orthodontics. What is artificial intelligence, what does it encompass, what can it already do in orthodontics and where is it taking us in the discipline? What opportunities does it offer and what negative effects does it have? In this article, the first in a three-part series, we will discuss these questions and strive to answer them.

Werner Schupp, Dr med dent
Private practice, Cologne, Germany

Asad Abu-Tarif, PhD, MBA, PMP, MCSD, CSM
Vice President, Research and Development, Spark and Digital Orthodontics, Envista Holdings Corporation, Brea, CA, USA

Julia Haubrich, Dr med dent
Private practice, Cologne, Germany

Fayez Elkholy, BSc, Dr med dent
Clinic of Orthodontics and Dentofacial, Ulm University (UULM), Ulm, Germany

James Mah, DDS, MSc, DMSc
Professor and Orthodontic Programme Director, University of Nevada, Las Vegas, NV, USA

Karl-Friedrich Krey, DMD, PhD, MME
Head of Department of Orthodontics and Dentofacial Orthopedics, Department for Orthodontics, University of Greifswald, Greifswald, Germany

Correspondence to: Dr Werner Schupp, Hauptstrasse 50, 50996 Cologne, Germany. Email: schupp@schupp-ortho.de

Introduction

When the term artificial intelligence (AI) was coined at a conference held in the USA in 1956, it was initially only of interest to computer scientists. Now, 65 years later, AI has found its way into all our lives. In medicine, AI has become indispensable in certain areas; one of the most frequently cited examples is the analysis of radiographs and MRI scans using AI-supported software. The first part in this series "AI in Orthodontics" addresses the presentation of AI in general. It examines how AI is divided into subfields (machine learning [ML], neural networks [NNs], deep learning [DL] and convolutional neural networks [CNNs]), and how Big Data is used in AI software so it can make decisions independently.

What is AI?

AI (or Assistant Intelligence as some like to call it) is a subfield of computer science that involves incorporating human thought and decision-making processes into computer-based procedures. An exact definition of AI is still pending. According to Oxford Reference, AI is "the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages"[1].

There are many ways of classifying AI algorithms, which encompass three main tasks:

1. Classification: The objective here is to classify data into one or more classes; for example, if given a credit card transaction, classifying that transaction as fraudulent or not, or, if given a 3D intraoral scan of a tooth, classifying that tooth by giving it the correct tooth number.
2. Regression: Instead of classifying things into discrete classes, regression or estimation algorithms try to predict a number/value. For example, an AI algorithm can use various parameters to forecast sales numbers. Can an AI algorithm predict the length of orthodontic treatment for a patient?
3. Generation: This class of AI algorithms essentially model how the data were generated and thus can generate synthetic data. Can an algorithm generate synthetic (fake) images of a horse once it has learnt from images of real horses? Can an AI algorithm generate a missing tooth to be used in dental implants based on how teeth look and given a scan of a patient's mouth?

AI can also be considered to consist of several subgroups (Fig 1): ML, as part of AI, in turn has two subgroups, NNs and DL, of which CNNs are a further subgroup.

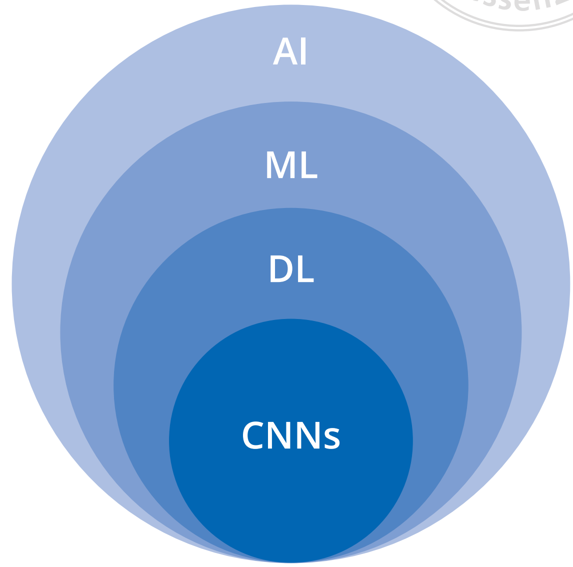

Fig 1 AI, ML, NNs/DL and CNNs.

Machine learning

With AI, the computer is programmed to make decisions. A traditional AI paradigm is shown in Fig 2 and involves studying the problem, summarising the rules, and then programming the computer to implement those rules. Once the program has been evaluated, the application is launched. Let us take the example of trying to detect fraudulent credit card activities. A financial expert can devise certain rules, for example that if two transactions are initiated from two different states and less than 10 minutes apart, the transaction that took place furthest from the user's home address is fraudulent. Once computers are programmed with such rules, we have AI-based fraudulent activity detection.

So, what is ML and why do we need it? Firstly, as seen in the aforementioned example of detecting fraudulent credit card activities, an expert was needed to summarise the rules. This results in the development of custom AI. Secondly, in many applications, such rules are highly complex or not even agreed upon (is there a book that contains all the rules that would classify a fraudulent activity?). Finally, after we have spent years developing the system and release it, thieves adapt! Once they discover how a computer classifies a fraudulent activity, they will adjust their behaviour to avoid detection. Another example is autonomous vehicles. Just imagine the rules that someone would need to program manually to cover all the different driving scenarios!

The basic principle of ML is to learn and improve autonomously from input data. It is an alternative to conventional programming. Instead of a prescribed program, the computer is given data with known relationships. In this process, the computer learns from known structures to apply them later to unknown contexts (Fig 3). ML has existed for decades in some specialised applications like optical character recognition (OCR).

ML uses algorithms to analyse known data. It detects statistical regularities that are represented in models. The models respond to new, as yet unknown data and sort them into categories. The more learning data ML receives, the better the models become. ML is very precise when structured data are available. From these structures, e.g., data from an Excel (Microsoft, Redmond, WA, USA) file, al-

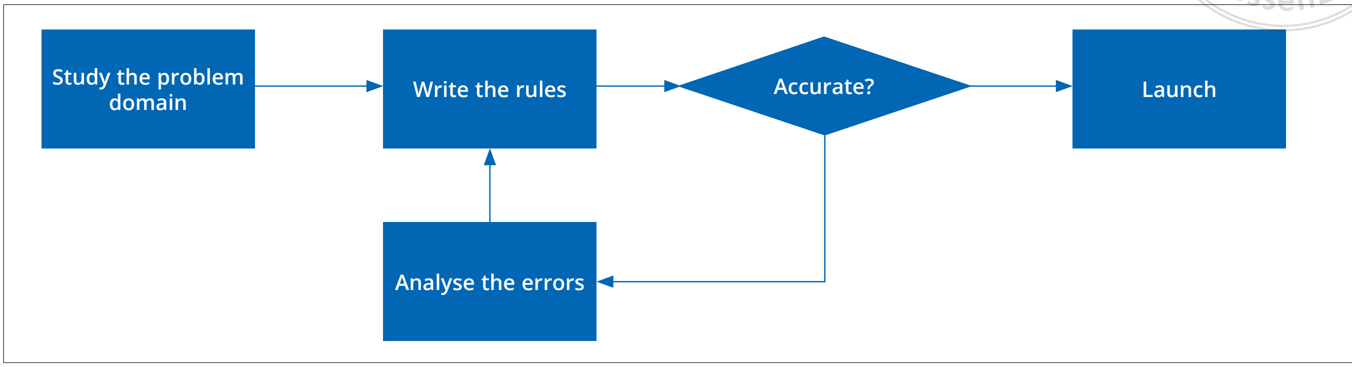

Fig 2 A traditional AI paradigm.

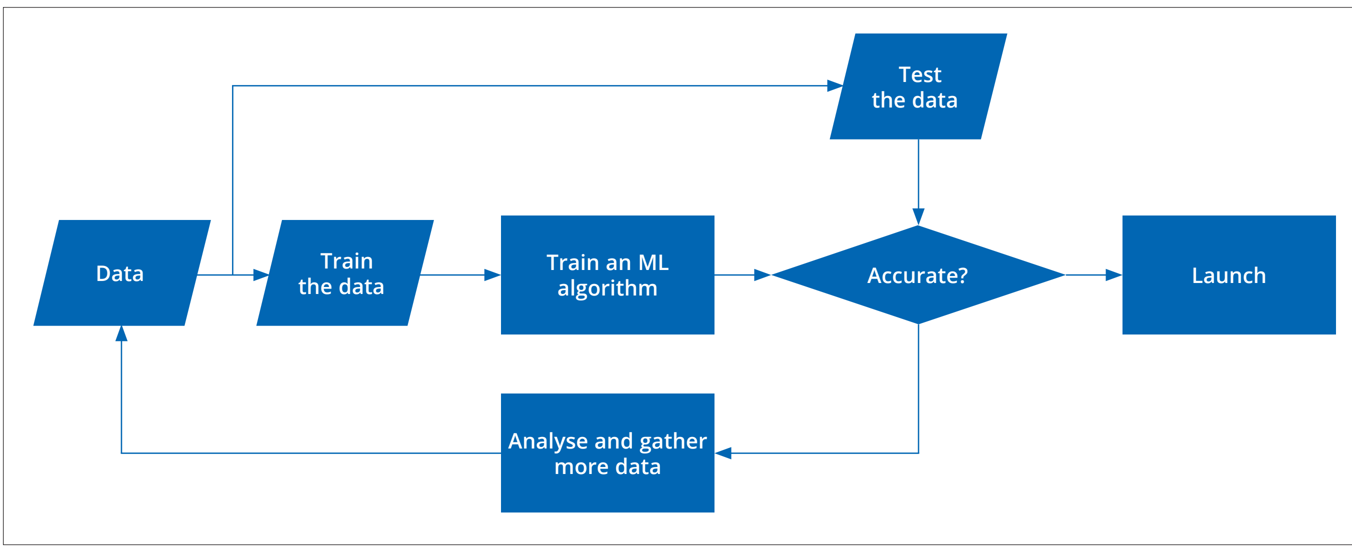

Fig 3 ML approach.

gorithms are built that search these structures for new, unknown information.

The main applications of ML in medical diagnostics are the evaluation of computed tomography (CT) scans and electrocardiograms (ECGs) and detection of skin lesions. Robotics and ML are becoming indispensable in surgery, or at least increasingly important. The da Vinci Surgical System (Intuitive Surgical, Sunnyvale, CA, USA), although not fully automated, is an impressive example.

There are many classifications of ML algorithms based on different criteria. For example, algorithms can be classified based on whether they are trained with human supervision, or whether they can learn incrementally on the fly. According to their classification, they can be described as supervised, unsupervised, semi-supervised and reinforcement learning algorithms.

In supervised learning, the algorithm is fed the training data that include the labels (the true classes) in the case of classification algorithms and the desired value in the case of predictor algorithms. Examples of supervised learning algorithms include k-nearest neighbours, logistic regression, linear regression, support vector machines, random forests, decision trees and NNs (DL).

With unsupervised learning, the training data are not labelled. For example, there are various pictures that may be known to include a certain number of object classes (e.g., horses, humans), but the class that any particular image contains is unknown. Unsupervised learning algorithms try to group the training data together automatically and define the classes. Once new data are encountered, they can be classified as belonging to one of the defined classes. Some of the algorithms that can be used in unsupervised learning

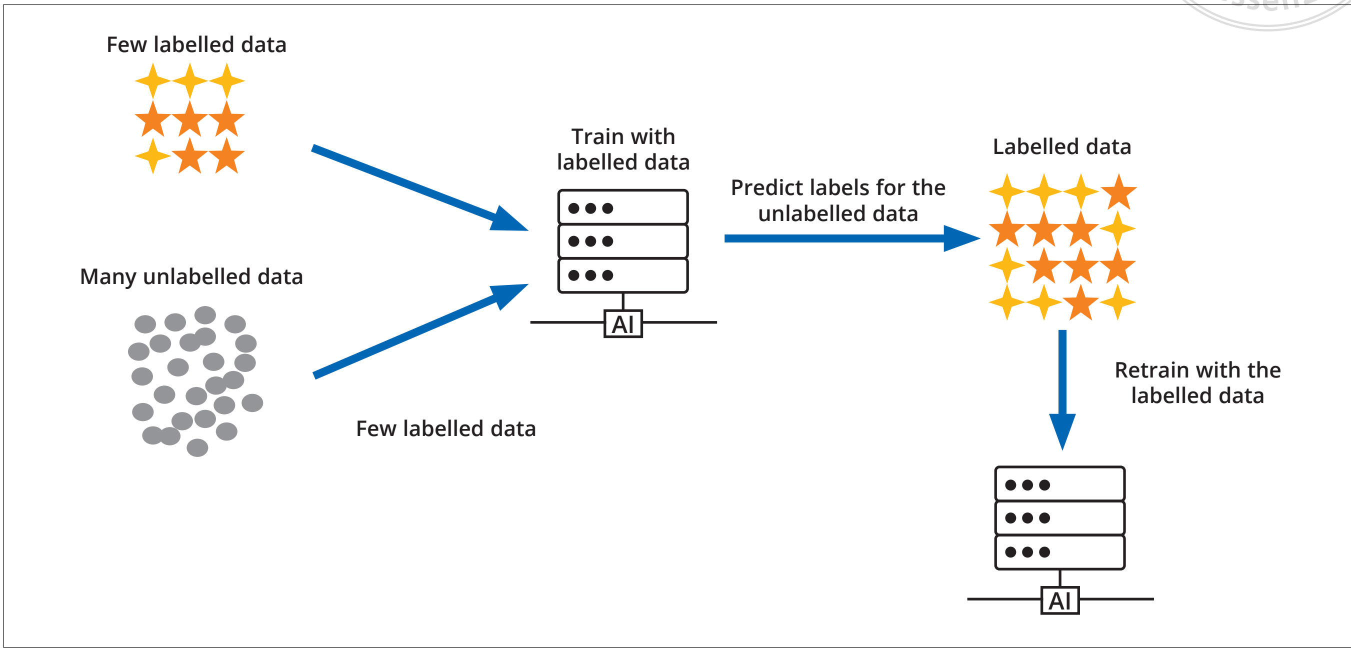

Fig 4 In ML, most models are supervised. They rely on labelled training data. When a model is trained to classify data, labels are used that indicate the class to which each data sample belongs. This allows typical patterns to be learnt for each class.

include K-means, one-class support vector machines, isolation forest and principal component analysis.

With semi-supervised learning, a portion of the training data is labelled but a large amount is not. This is because it is typically time-consuming and difficult to label data (e.g., manually classifying tumours in hundreds of thousands of images). In such instances, labels would be generated for some of the data and then a combination of supervised and unsupervised learning would be used to generate the model.

Reinforcement learning is somewhat different in that ML would have an agent that observes the environment and performs actions. It can then be rewarded or penalised based on a certain regulation to learn how to perform the correct actions. For example, robots use reinforcement learning to learn how to walk. When the robot falls, its actions are penalised, and as it takes steps successfully, its actions are rewarded, such that it ultimately learns how to walk.

Deep learning

Artificial NNs (ANNs) were inspired by the networks of biological neurons found in the human brain and are at the core of DL. DL is a subfield of ML and consists of massive multilayer networks of artificial neurons that can automatically discover useful features, that is, representation of input data needed for tasks such as detection and classification, given large amounts of unlabelled or labelled data (Fig 4). Labelled data are raw data to which meaningful labels such as 'add tags' or 'assign classes' have been allocated.

DL can automatically learn useful representations of data, thereby eliminating the need for handcrafted features. The representations learnt from one dataset can be useful even when applied to another. This property, referred to as transfer learning, is not unique to DL, but the large training data requirements of DL make it particularly useful in cases where the relevant data for a particular task are scarce. In medical imaging, a DL system can be trained using a large number of natural images or those in a different modality to learn proper feature representations that allow it to 'see'. The trained system can subsequently use these representations to encode a medical image that is employed for classification[2].

The success of DL compared to traditional ML is primarily due to two interrelated factors: depth and compositionality. A further advantage of deep architectures relates to how successive layers of the network can utilise the representations from previous layers to compose more complex representations that better capture critical characteristics of the input data and supress the irrelevant variations. In image recognition, deep networks have been shown to capture simple information such as the presence or absence of edges at different locations and orientations in the first layer. Successive layers of the network assemble the edges into compound edges and corners of shapes, and then into increasingly complex shapes that resemble object parts. Hierarchical representation learning is very useful in complicated tasks such as computer vision where adjacent pixels and object parts are correlated with each other and their relative locations provide clues about each class of object[2].

As larger datasets have become more commonplace and commercial gaming graphical processing units (GPUs) have become available, it is now possible to explore how larger deeper architectures can be trained faster[2].

The longer a network is trained, the more complex its solution becomes; thus, by regularising on time through early stopping, complexity will be reduced and generalisability improved. Dropout, a term that refers to dropping out units in a neural network[2], is another efficient way to prevent overfilling.

Convolutional neural networks

A key problem with ML is that useful features are difficult to design and often require the collective efforts of many researchers over years or even decades to optimise. In addition, the features are domain- or problem-specific. Traditional ML algorithms were heavily dependent on having access to good feature representations; otherwise, it was extremely difficult to improve the state-of-the-art results for a given dataset[2]. CNNs were developed to solve this problem. CNNs are part of NNs. They are very successful in image processing and recognition or speech recognition, and consist of several layers: the convolution layer, pooling layer and fully connected layer.

The convolutional layer is the actual folding level. It can recognise and extract individual features in the input data. In image processing, these features can be lines, edges or specific shapes. The processing of the input data takes place in the form of a matrix. Matrices of defined size (width × height × channels) are used. By adding the convolutional layer to a traditional NN machine, the computer automatically learns what features in the image are important and hence generalises the solution.

The pooling layer, also known as the subsampling layer, compresses and reduces the resolution of the detected features. To achieve this, the layer uses methods such as maximum pooling and mean pooling. Pooling discards superfluous information and reduces the amount of data. This does not diminish the performance in ML; the reduced amount of data increases the computational speed.

The final layer of the CNN is the fully linked layer, which joins the repeating sequences of the convolutional and pooling layers. All features and elements of the upstream layers are linked to each output feature. The fully connected neurons can be arranged in multiple layers, and the number of neurons depends on the classes or objects that the NN is to distinguish[3].

Compared to conventional non-CNNs, CNNs offer numerous advantages. They are suitable for ML and AI applications with large amounts of input data, such as image recognition. The network is robust and insensitive to distortion or other optical changes. They can process images captured in different lighting conditions and from different perspectives and still recognise the typical features of an image.

Because CNNs are divided into several local, partially meshed layers, they have a much smaller memory footprint than fully meshed NNs. The convolutional layers drastically reduce the memory requirements, and training time is also reduced significantly. Using modern graphics processors, CNNs can be trained very efficiently. In image recognition, CNNs are the state-of-the-art method for ML and classification[2].

Let us explain how CNNs work using image recognition as an example. CNNs detect and extract features of the input images using filters. Recognition of the structures is location independent within the image. Initially, CNNs recognise simple structures such as lines, colour spots and edges in the first layers. In the next levels, CNNs learn combinations of these structures, such as simple shapes or curves. With each level, more complex structures can be identified. The data are repeatedly resampled and filtered in the layers. In the final step, the results are assigned to the classes or objects to be recognised[2]. Unlike ML, CNNs can process restructured data; however, huge amounts of Big Data are needed for this.

Other types of NNs (also part of DL) include recurrent NNs (RNNs) and generative adversarial networks (GANs). With NNs, the output from one layer feeds the input to the next. With RNNs, in addition to having feedforward networks like those in NNs, in some layers, the output from one neuron in the layer is fed as input to the neurons in the same layer. These networks can be used for time series forecasting, for example. GANs, on the other hand, are typically CNNs that consist of two models, a generative model (generates data) and a CNN classification model (classifies whether the generated data are fake or real). The job of the generative CNN is to generate data that look real and thus deceive the classifier, and the classification model must prevent this from happening. Training both models and having them compete results in a generative model that generates very realistic data. GANs are used frequently in art to generate images and videos, but have recently also been employed to generate 3D models of missing crowns for dental implants.

DL in medical imaging

In medical imaging, ML algorithms have been used for decades, starting with algorithms to analyse or help to interpret radiographic images in the mid-1960s. Initially, computer-aided detection/diagnosis algorithms predominantly used a data-driven approach, like most DL algorithms today; however, unlike most DL algorithms, the majority of these early methods were heavily dependent on feature engineering. A typical workflow for developing an algorithm for a new task consisted of understanding what types of imaging and clinical evidence clinicians use for the interpretation task, translating the knowledge into computer code to automatically extract relevant features, and then using ML algorithms to combine these features into a computer score. Data were propagated through the networks via convolutions, the networks learnt filter kernels, and the methods did not require feature engineering; that is, the inputs into the networks were image pixel values[2].

It is impossible to imagine imaging diagnostics without AI. Three- and two-dimensional CBCT images are improving, and AI is responsible for this. Particularly in oncology, DL helps to read images accurately. The results of scientific studies of image evaluation have illustrated that DL sometimes already reads with greater accuracy than trained and experienced physicians[4].

One application area of DL is low-dose image reconstruction. This is important in modalities with ionising radiation such as CT or positron emission tomography (PET). Techniques similar to those described for denoising have been applied to artefact reduction. Metal-affected projections can be eliminated accordingly. A study has shown how to use DL to generate synthetic CT images from MRI free from ionising radiation[2].

Analysis of imaging procedures using DL will not replace physicians but makes it possible to perform analyses more accurately from a second perspective. Fourcade and Khousari[5] call DL in medical image analysis “a third eye for the doctors”; however, they foresee that radiology and pathology will be transformed greatly by it.

Kim et al[6] predict that a fully automated cephalometric analysis algorithm and web-based application can be widely used in various environments to save time and effort for manual marking and diagnosis. They found that the automated algorithm achieved a successful classification rate of 88.43%.

In a pilot study, Schwendicke et al[7] applied CNNs to detect caries lesions in near-infrared light transillumination (NILT) images and concluded that a moderate deep CNN trained with a limited amount of NILT image data showed a satisfying discriminatory ability to detect caries lesions.

DL requires a large amount of data because it learns features directly from the data via an end-to-end process. In an anatomical classification study of LT data, at least 1000 datasets per group were required to achieve 98.0% validation accuracy with DL, and 4092 datasets per group to reach the desired accuracy of 99.5%[8]. It is especially important to emphasise the need to construct a large-scale dataset of public dental information to make the clinical application of DL possible. CBCT images vary widely depending on the machine used and the exposure conditions, and this can impede research into DL. For example, collecting and learning data on a machine-by-machine basis is difficult because models learnt on one machine do not apply to other machines. Although attempts have been made to develop guidelines in Europe and England regarding the image quality of CBCT, no international standard has yet been established[9].

Intraoral scanning is another field that employs DL, as almost all intraoral scanners utilise background algorithms to a certain extent.

Classification and segmentation of teeth from intraoral scans

Accurate segmentation of data derived from intraoral scans (IOSs) is a crucial step in a computer-aided design system (CAD) for many clinical tasks in orthodontics. To obtain the highest possible quality, a segmentation model may process a point cloud derived from an IOS in its highest available spatial resolution, particularly to perform a valid analysis in finely detailed regions such as the curvatures in borders between two teeth[10].

Traditional geometry-based methods tend to achieve undesirable results due to the complex appearance of human teeth. Traditional tooth segmentation methods barely enable labelling of individual teeth. Xu et al[11] developed a generic and robust segmentation model by exploiting CNNs. In their study, the segmentation task was completed by labelling each mesh face, and they extracted a set of geometry features as face feature representations[11]. The accuracy of their mesh labelling methods exceeds that of the state-of-the-art geometry-based methods, reaching 99.06% measured by area which is directly applicable in orthodontic CAD systems, and also offers robust protection against any possible foreign matter on the model surface, such as air bubbles and dental accessories[11].

Tian et al[12] proposed a novel approach based on a sparse voxel octree and 3D dental models for segmenting and classifying tooth types. The tooth classification method capitalised on two-level hierarchical feature learning and was proposed to solve the misclassification problem in highly similar tooth categories[12]. The authors exploited an improved three-level hierarchical segmentation method based on deep convolution features to conduct segmentation of the teeth from the gingiva and of each individual tooth, and the conditional random field model was used to refine the boundary of the gingival margin and the region where tooth fusion occurs[12]. The experimental results showed that the classification accuracy in the level 1 network was 95.96%, the mean classification accuracy in the level 2 network was 88.06% and the accuracy of tooth segmentation was 89.81%[12]. Compared with the existing state-of-the-art methods, the proposed method offers greater accuracy and universality, and there is great potential for its application in computer-assisted orthodontic treatment diagnosis. Tooth segmentation is a core step in many oral medical research processes and forms the basis for computer-aided dental diagnosis and treatment. A high level of accuracy in segmentation is becoming increasingly important, especially in aligner orthodontics.

As computer hardware and software technology have improved, many commercial CAD/CAM software programs for orthodontics such as 3Shape (Copenhagen, Denmark), OrthoCAD (Cadent, Carlstadt, NJ, USA) and OnyxCeph (Image Instruments, Chemnitz, Germany) have emerged that can achieve automatic tooth segmentation to a certain extent. Due to the complexity of interactive operation and the high degree of manual intervention in orthodontic CAD systems, their segmentation efficiency appears lower. Orthodontic patients usually have symptoms such as crowding, missing teeth or indistinct boundaries between teeth[12].

The authors report their attempt that used a hierarchical feature learning framework based on 3D CNNs to automatically extract high-dimensional features from 3D tooth models to segment and classify the tooth types. The proposed method of hierarchical segmentation is robust to various complex malformations in patients' teeth that have important application value for virtual tooth arrangement in subsequent orthodontic treatment procedures[12].

The main contributions of the study by Tian et al[12] are as follows:

- A general and robust tooth segmentation and classification framework that achieves 88.06% accuracy for highly similar tooth categories and 89.81% for individual tooth segmentation.
- An optimised design of two-level hierarchical classification network architecture that can solve the misclassification problem in highly similar tooth categories.
- An improved three-level hierarchical segmentation network that is based on conditional random fields to refine the segmentation boundary and that is robust to various complex malformations in teeth.
- A flexible and extendible model that can be retrained to generalise for new dental samples, which allows the framework to improve the intelligence level of orthodontic CAD systems.

As an important step in a computer-aided orthodontic system, the main aim of tooth segmentation is to accurately locate, identify and extract teeth based on patients' digital dental model. The automatic segmentation of individual teeth is not a simple task since the tooth shapes are complex and tooth arrangement varies from person to person; however, the successful application of DL in medical diagnosis has demonstrated its superiority in reducing labour costs and manual intervention. Additionally, the self-learning ability of CNNs can dramatically improve the accuracy and efficiency of the network[12].

Discussion

AI is now present in many areas of orthodontics. We use CBCT, MRI, scanners and virtual treatment software in our practices. Imaging techniques (CBCT and MRI) are becoming increasingly accurate, which gives orthodontists more confidence and results in a more accurate diagnosis and therefore better treatment decisions for patients. In aligner orthodontics, the various software solutions offer the possibility of performing a virtual treatment simulation independently, autonomously and on our own responsibility. High-quality in-office aligner orthodontics has thus become possible. This is where self-learning programs will change the future of orthodontics; however, how much AI we need and want at this stage remains to be discussed.

Conclusion

AI has arrived in our lives and will continue to conquer in the future. This first article in a three-part series explained some of the basics of AI in its current state and aspects of AI used in the medical sector. The following articles will take a more in-depth look at the state of the art and possibilities for AI in the future.

Declaration

The authors declare there are no conflicts of interest relating to this study.

References

1. Oxford Reference. https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960. Accessed 28 October 2021.
2. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. Med Phys 2019;46:e1–e36.
3. Luber S, Litzel N. Was ist ein Convolutional Neural Network? https://www.bigdata-insider.de. Accessed 1 September 2021.
4. Ehtesharni Bejnordi B, Veta M, van Diest PV, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017;318:2199–2210.
5. Fourcade A, Khonsari RH. Deep learning in medical image analysis: A third eye for doctors. J Stomatol Oral Maxillofac Surg 2019;120: 279–288.
6. Kim H, Shim E, Park J, Kim YJ, Lee U, Kim Y. Web-based fully automated cephalometric analysis by deep learning. Comput Methods Programs Biomed 2020;194:105513.
7. Schwendicke F, Elhennawy K, Paris S, Friebertshäuser, Krois J. Deep learning for caries lesion detection in near-infrared light transillumination images: A pilot study. J Dent 2020;92:103260.
8. Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy [preprint]? arXiv 2015:1511.06348.
9. Hwang JJ, Jung YH, Cho BH, Heo MS. An overview of deep learning in the field of dentistry. Imaging Sci Dent 2019;49:1–7.
10. Ghazvinian Zanjani F, Moin DA, Verheij B, et al. Deep learning approach to semantic segmentation in 3D point cloud intra-oral scans of teeth. MIDL 2019;102:557–571.
11. Xu X, Liu C, Zheng Y. 3D tooth segmentation and labeling using deep convolutional neural networks. IEEE Trans Vis Comput Graph 2019;25:2336–2348.
12. Tian S, Dai N, Bei Z, Yuan F, Yu Q, Cheng X. Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. IEEE Access 2019;7:84817–84828.