Applications of deep learning in dentistry

Stefano Corbella, PhD, DDS,[a,b,c] Shanmukh Srinivas, BDS, MDS,[d] and Federico Cabitza, PhD, BSc, MEng[e]

Over the last few years, translational applications of so-called artificial intelligence in the field of medicine have garnered a significant amount of interest. The present article aims to review existing dental literature that has examined deep learning, a subset of machine learning that has demonstrated the highest performance when applied to image processing and that has been tested as a formidable diagnostic support tool through its automated analysis of radiographic/photographic images. Furthermore, the article will critically evaluate the literature to describe potential methodological weaknesses of the studies and the need for further development.
This review includes 28 studies that have described the applications of deep learning in various fields of dentistry. Research into the applications of deep learning in dentistry contains claims of its high accuracy. Nonetheless, many of these studies have substantial limitations and methodological issues (e.g., examiner reliability, the number of images used for training/testing, the methods used for validation) that have significantly limited the external validity of their results. Therefore, future studies that acknowledge the methodological limitations of existing literature will help to establish a better understanding of the usefulness of applying deep learning in dentistry. (Oral Surg Oral Med Oral Pathol Oral Radiol 2021;132:225−238)

This article will focus on specific discriminative computational models that have produced highly accurate performance in tasks of medical and dental image processing. These kinds of computational models, which are also less accurately referred to as algorithms, are often put under the much more evocative (and suggestive) rubric of artificial intelligence (AI).

Interest in the medical application of AI has increased recently due to the impact of this technology on the outcome and quality of clinical practice in and after the 1980s, when applications known as "expert systems" resulted in several expectations that were realized in the following decades.[1] This likely occurred for 2 reasons: (1) the wide diffusion and use of diagnostic information in medicine derived from medical imaging from various sources, such as magnetic resonance imaging, computed tomography (CT), positron emission tomography, cone beam computed tomography (CBCT), and standard radiographs, and (2) the growing visibility in medical literature of works dealing with the performance of AI systems. Both the selected studies and recently published meta-analyses and literature reviews have acknowledged the performances of these systems, noting that they are on par with and even superior to the performance of human specialists, especially in image-based diagnostic tasks. Regardless, the validity of the presented results has been weakened by the paucity of high-quality studies (just a few of the studies in the 2 cited papers fulfilled review requirements) and their methodological limitations.[2,3]

These results are the final achievement of algorithms that were first devised in the 1980s and only materialized after they could run on more efficient computers equipped with faster processing power backed by much larger data sets. In addition, the expert systems mentioned previously were based on a rigid framework of rules and methods that aimed to make logical decisions that mimic the ways in which clinicians and experts think. Though purely computational in nature, the results reported in recent literature have regarded this technological advancement as a shift in the paradigm, from the rule-based algorithms of expert systems, which have rarely achieved optimal discriminative performance, to data-driven models, which are capable of achieving superhuman performance and are developed using machine learning (ML).[4] In general, ML was developed in the quest to build an intelligent machine that could construct accurate statistical models to classify cases and/or predict continuous outcomes through the development of numerous methods (e.g., classification trees, regression trees, bootstrap forest, boosted tree, *k*-nearest neighbors, naïve Bayes, multiple logistic regression, and neural networks).

[a]Department of Biomedical, Surgical and Dental Sciences, Università degli Studi di Milano, Milan, Italy.
[b]IRCCS Istituto Ortopedico Galeazzi, Milan, Italy.
[c]Department of Oral Surgery, Institute of Dentistry, I. M. Sechenov First Moscow State Medical University, Moscow, Russia.
[d]Indian Institute of Science, Bangalore, India.
[e]Department of Informatics, Systemics and Communication, University of Milano-Bicocca, Milan, Italy.
Received for publication Jun 3, 2020; returned for revision Oct 9, 2020; accepted for publication Nov 8, 2020.

2212-4403/$-see front matter
https://doi.org/10.1016/j.oooo.2020.11.003

Statement of Clinical Relevance

Deep learning machines could be a viable and extremely useful aid for dental diagnosis and, in general, for the management of images in any field of dentistry. The accuracy of such methods should be improved in order to be considered for everyday practice.

In other words, ML represents a newer set of statistical techniques and computational methods through which specific programs (suggestively called "learners") produce mathematical models that can classify input data based on a larger data set of available data and feedback data. These models are mathematical functions built to accurately interpolate a set of data points (called training data) in the belief that this function can be used to approximate new data that yield reliable predictions or produce new metadata (i.e., data describing other data) and that is accurate in the classification process. Methods of ML typically iteratively and automatically build accurate interpolating functions by optimizing a given function (often related to prediction errors). In other words, the above-mentioned learners tune the parameters of the functions through incremental trial and error by exploring the space of possible parameter values to minimize the loss function, which is usually represented by the difference between predicted values and actual values. To clarify, neural nets often use activation functions that result in equations that are similar to regression equations where the intercept is replaced with a bias term and the coefficients are typically referred to as weights.

For image classification tasks, a particular subset of ML can achieve higher performance. This specific class of ML technique and architecture is called deep learning (DL). This particular expression is used among data scientists for a multilayer artificial neural network (ANN) with a set of methods applied to optimize the interpolating functions represented by their means. DL can be viewed as an evolution of ML that aims to automate data preprocessing to create features that can optimize classification tasks (e.g., data cleansing, missing imputation, feature selection, data normalization, and standardization). If the expression "learning" in ML is misleading, then through its very designation DL acts as a source of further misunderstanding. The word "deep" in DL does not represent these techniques' obscurity or power. Rather, it functions as a primary reference to their architecture in interpolating models, such as an ANN. These structures partially resemble neuronal networks in the human brain. They are computational structures where simple computing nodes—which usually compute a simple summation and nonlinear transformation functions (the latter are functions that produce nonlinear transformations in the inputs, such as the hyperbolic tangent function, which is a sigmoid function)—are connected by oriented and weighted links. Therefore, deep in DL refers to the fact that these structures encompass many layers of nodes between the input and output nodes. As a result, the architecture of DL machines can be very complex and include multiple layers and millions of parameters.

As hinted above, convolutional (artificial) neural networks (CNNs) frequently serve as the architecture used to recognize and classify images. A CNN is a type of ANN wherein many of the intermediate nodes apply convolutional functions to their input data and a type of filter functions them into isolate patterns, such as the edges, vertices, and other higher-level elements of an image. The hierarchical structure of a CNN is suitable for processing features from raw images to more abstract ones. The final nodes are classification nodes, where the inputs coming from the preceding nodes are interpreted in terms of generalized concepts, such as dog, cat, gorilla, or, in the case of dental imaging, a radiograph that positively identifies a specific pathologic condition. In DL, many other special, more efficient, and task-oriented CNN architectures are currently being employed, such as recursive and recurrent neural networks.

One can distinguish a basic CNN architecture from others according to the assigned task, the procedure by which the networks' weights are adjusted, the architecture of the neurons, and the functions of the nodes. The most commonly used architectures in image detection include regional CNN, faster regional CNN, and you only look once. Unlike image classification, wherein only the contents of an object inside an image are determined without specifying the location of an object inside the image, the location of the image is determined with these architectures. Object detection architectures can classify, locate, and identify the targeted contents from an image at a comparatively faster rate than conventional CNNs by employing multiple bounding boxes for object edge detection.

Image segmentation is a further variation of image detection that is achieved by applying similar principles where the targeted object in an image is segregated from its background. Image segmentation includes semantic and instance segmentation. In semantic segmentation, multiple objects belonging to the same category (e.g., teeth) are considered as a single entity (e.g., all of the teeth segmented under the same mask). In the case of instance segmentation, multiple objects of the same category are considered to be distinct and individual objects (e.g., every single tooth in an image is segmented under a separate mask; Figure 1). Some frequently used architectures for image segmentation include fully convolutional networks and U-NET (Figure 1).

Recognizing AI as a form of automation enables commonalities and differences to be recognized through regular automation. Applying autonomous technologies to a set of human activities can either replace humans or make them significantly more effective and efficient in their decision-making processes. As recently suggested by the American Medical

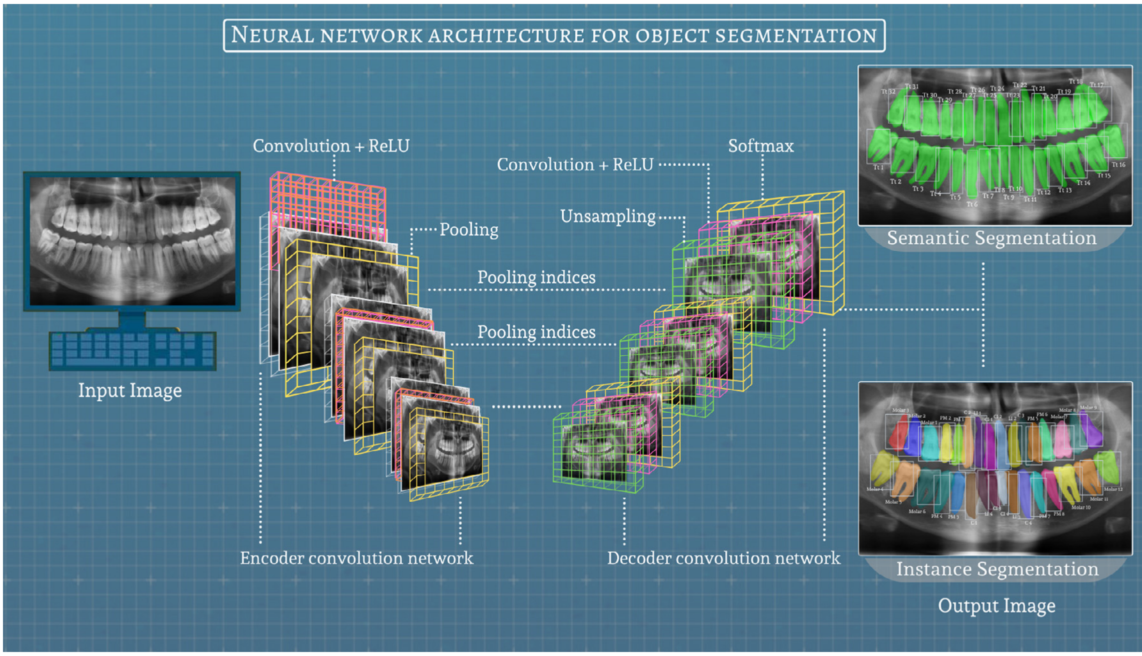

Fig. 1. Image segmentation process.

Association, the same acronym (AI) can be used to denote the augmented intelligence of human beings. Nonetheless, the automation of intellectual tasks and the potential replacement of interpretative and cognitive skills, especially in the case of medical diagnostics, can also induce the phenomena of deskilling[5] and automation bias.[6]

The present article will report on the main studies that have investigated DL and shed light on the level of penetration that DL techniques can have in dental research and clinical practice. Unlike other specialties, such as ophthalmology,[7] dermatology,[8] and image-intensive fields like radiology and pathology,[9] DL has the enormous potential to bring genuinely impactful applications to the field of dentistry. Nonetheless, except for the many studies that this article will review in forthcoming sections, the benefits that this class of computational methodology can bring to dental research and practice have yet to be determined.

APPLICATIONS OF DEEP LEARNING IN DENTISTRY

General considerations and methodological issues

For the present literature review, the PubMed/Medline, EMBASE, and Cochrane Library electronic databases were searched using a combination of keywords that referred to the topic of DL in dentistry. The reference lists of all pertinent articles were screened for potentially relevant papers. Furthermore, we performed a manual search of all issues from 2010 onward in the following dentistry journals: *Journal of Dental Research, Journal of Dentistry, Oral Diseases, Journal of Clinical Periodontology, Journal of Periodontology, Journal of Endodontics*, and *International Endodontic Journal*. We decided to exclude congress abstracts and nonindexed publications and to use only papers written in English. In the collected CNN research, most studies applied DL to analyze radiographic images (panoramic, periapical, CBCT, etc.), some studies used DL to assess photographs, and a single study used DL to analyze data from the electronic databases of clinical records. The results of this literature review are summarized in Table I. The included papers were all published recently, with 1 paper published in 2017, 7 papers published in 2018, 18 papers published in 2019, and 2 papers published in 2020.

Deep learning as a diagnostic aid

Convolutional networks have been widely applied in the studies listed in the literature review in the recognition and identification of particular regions within digital images.[10] This specific ability has been demonstrated to identify an application in every medical field that involves the management and use of digital images, such as photographs and, specifically, radiographic images.[11]

In the field of dentistry, both photographic and radiographic images are frequently used, and they often represent one of the first steps in the assessment and diagnostic processes of patients. In particular, radiographic images (e.g., periapical, panoramic, or CBCT images) are widely used by both general practitioners and specialists. In dentistry, CNNs are trained to recognize, classify (Figure 1), and segment (Figures 1 and 2)

Table I. Characteristics of the included studies

Study	*Architecture*	*Training/testing modalities*	*Dental specialty*	*Object*	*Imaging modalities*	*Reference standard*	*Results*	*Conclusions*	*Limitations*
Chen et al.[14]	R-CNN (Inception Resnet v2)	Training set: 800 Validation set: 200 Testing set: 250	General/anatomy	Tooth classification	Periapical radiographs (300 to 500 × 300 to 400 pixels)	Expert dentist	Mean IOU (intersection over union) = 0.91	The R-CNN has the same performance as a junior dentist	Sensitivity/specificity not reported; methods of sample selection
Vinayahalingam et al.[18]	CNN	—	Oral surgery, anatomy	Third molars and mandibular nerve identification	Panoramic radiographs	Not reported	Mean Dice coefficient for third molar = 0.947 (0.033); Mean Dice coefficient for mandibular nerve = 0.847 (0.099)	CNN results are encouraging, though further enhancement of the algorithm is advised to improve the accuracy	Reliability of the examiners; methods of sample selection
Miki et al.[12]	CNN (AlexNet)	Training set: 42 Testing set: 10	Prosthodontics	Tooth classification	Cone beam CT images	Not reported	The average classification Ac using the augmented training data was 88.8%	The proposed method is advantageous in obtaining high classification accuracy without the need for precise tooth segmentation	Low number of images used for training/testing; images not standardized; reliability of the examiners
Tuzoff et al.[15]	CNN (VGG-16)	Training set: 1352 Testing set: 222	Prosthodontics	Tooth classification	Panoramic radiographs	Not specified "expert"	For the tooth detection task, CNN achieved a sensitivity of 99.41% and a specificity of 99.45%. For tooth numbering, its sensitivity and specificity were 98.93% and 99.94% respectively	The performance of the proposed computer-aided diagnosis solution is comparable to the level of experts	Reliability of examiners

(*continued on next page*)

Table I. Continued

Study	*Architecture*	*Training/testing modalities*	*Dental specialty*	*Object*	*Imaging modalities*	*Reference standard*	*Results*	*Conclusions*	*Limitations*
Xu et al.[13]	CNN	Training set: 1000 Validation set: 50 Testing set: 150	Prosthodontics	Tooth classification	3D dental models	Not specified (professional orthodontic company)	Accuracy for tooth segmentation and labeling achieved for maxillary dental models was 99.06%; for mandibular dental models it was 98.79%	Optimal discriminatory ability	Limitations related to the quality of the boundaries between 2 teeth; methods of sample selection
Zhang et al.[16]	CNN (label tree with cascade network structure)	Training set: 200 images with 639 teeth	General/anatomy	Tooth classification	Periapical radiographs	Not specified	Compared to the state-of-the-art convolutional neural network the precision was 95.8%	Quite good performance even when dealing with complex cases such as decayed tooth, filled tooth, and tooth loss	Limited training data
Hiraiwa et al.[17]	CNN (AlexNet), (GoogLeNet)	Training set: 608 Testing set: 152	Endodontics	Distal root of mandibular first molars	Panoramic radiographs/CT images	Radiologist	Ac for AlexNet and GoogLeNet was 87.4% and 85.4%, respectively	High accuracy of DL system	Sample selection; reliability of the examiners
Lee et al.[34]	CNN (VGG-19 modified)	Training set 1044 Validation set: 348 Testing set: 348	Periodontology	Tooth prognosis	Periapical radiographs (224×224 pixels)	Three calibrated, board-certified periodontists	Diagnostic accuracy for detecting periodontally compromised teeth was 81.0% for premolars and 76.7% for molars and the accuracy for predicting extraction was 82.8% for premolars and 73.4% for molars	CNN algorithm was useful for assessing the diagnosis of periodontally compromised teeth	The diagnosis of periodontally compromised teeth is not based only on periapical radiographs; low-resolution images; reliability of examiners

(*continued on next page*)

Table I. Continued

Study	*Architecture*	*Training/testing modalities*	*Dental specialty*	*Object*	*Imaging modalities*	*Reference standard*	*Results*	*Conclusions*	*Limitations*
Casalegno et al.[19]	CNN	Trained on 185 samples and validated on 32 samples	Restorative dentistry	Caries diagnosis	NILT pictures of the occlusal surface (256 × 320 pixels) in grayscale	Expert identification	AUC = 85.6% for proximal lesions; AUC = 83.6% for occlusal lesions	DL approach for the analysis of dental images could increase speed and accuracy of caries detection	Characteristics of the pictures could influence the accuracy (overexposed, underexposed); trained sample is limited; reliability of examination
Lee et al.[21]	CNN (GoogLeNet Inception v3)	Training set 2400 Testing set: 600	Restorative dentistry	Caries diagnosis	Periapical radiographs (299 × 299 pixels)	Four calibrated board-certified dentists	Diagnostic Ac for premolar 89%, molar: 88%, premolar and molar models: 82%, AUC achieved by CNN, premolar: 0.917, molar: 0.890, both molar and premolar: 0.845	CNN algorithms demonstrated to be the most effective method to detect caries from periapical radiographs	Limited sample size; low-resolution images
Schwendicke et al.[20]	CNN (Resnet18, Resnext50)	Trained on online database. Test on 226 images	Restorative dentistry	Caries	NILT images (224 × 224)	Two experienced dentists	Resnet18: AUC: 0.73, Ac: 0.69, Se: 0.46, Sp: 0.85, PPV: 0.71, NPV: 0.69. Resnext50: AUC: 0.74, Ac: 0.68, Se: 0.59, Sp: 0.76, PPV: 0.63, NPV: 0.73	CNNs may be useful to assist NILT-based caries detection	Reliability of examiners was limited; larger data sets for training are needed

(*continued on next page*)

Table I. Continued

Study	*Architecture*	*Training/testing modalities*	*Dental specialty*	*Object*	*Imaging modalities*	*Reference standard*	*Results*	*Conclusions*	*Limitations*
Ekert et al.[22]	CNN	2001 cropped image segments, each representing a tooth, from 85 randomly chosen digital panoramic dental radiographs	Endodontics	Apical lesions	Panoramic radiographs (cropped focusing on one particular tooth)	Majority vote of 6 independent and experienced dentists	AUC for CNN was acceptable at 0.85. Apical lesion detection sensitivity was significantly higher for molars than in other tooth types, whereas specificity was lower	Satisfying discriminatory ability of CNN in detecting apical lesions on panoramic radiographs	Doubtful reliability of examiners; broad inclusion criteria
Kim et al.[28]	CNN (Residual Net)	Training set: 8000 Validation set: 1000 Testing set: 540	Oral surgery, ENT	Maxillary sinusitis	Waters' view radiographs	Five radiologists	AUC = 0.93 and 0.88 for the temporal and geographic external test set	The CNN algorithm could diagnose sinusitis with higher AUC and sensitivity and specificity comparable to those of radiologists	Use of CT to confirm sinusitis; only maxillary sinuses
Murata et al.[29]	CNN (AlexNet)	Training set 400 healthy, 400 inflamed Testing set: 60 healthy, 60 inflamed	Oral pathology	Maxillary sinusitis	Panoramic radiographs	Premade diagnosis (methods not known)	Ac DL: 87.5%; Se DL: 86.7%; Sp DL: 88.3%; PPV DL: 88.1%; NPV DL: 86.9% Comparable to the results obtained by specialists AUC DL: 0.875	The diagnostic performance of the DL system for maxillary sinusitis on panoramic radiographs was sufficiently high	Only single sinuses were considered; limited sample size

(*continued on next page*)

Table I. Continued

Study	*Architecture*	*Training/testing modalities*	*Dental specialty*	*Object*	*Imaging modalities*	*Reference standard*	*Results*	*Conclusions*	*Limitations*
Nakano et al.[35]	DL	Testing set: 45 weak/no oral malodor, 45 marked oral malodor	Periodontology	Salivary microorganisms	Salivary samples were collected and 16 S rRNA genes were amplified followed by gene sequence analysis to categorize operational taxonomic units	Organoleptic test and gas chromatography	Predictive Ac of DL was 97%	High accuracy of the algorithm	Limited sample size
Ariji et al.[26]	CNN (DetectNet)	Training set 210 Testing set: 50, 25	Oral medicine and radiology	Mandibular radiolucent lesions	Panoramic radiographs (900 × 900 pixels)	Histopathologic verification of the diagnosis	The detection and classification Ac achieved were 71% and 60% for ameloblastomas, 100% and 13% for odontogenic keratocysts, 88% and 82% for dentigerous cysts, and 81% and 77% for radicular cysts, respectively	CNN demonstrated high sensitivity in detecting radiolucent lesions	Relatively limited sample size
Lee et al.[38]	CNN (GoogLeNet Inception v3)	Training set: 684 panoramic images; 789 CBCT images Validation set: 228 panoramic images; 197 CBCT images Testing set: 228 panoramic images; 197 CBCT images	Oral pathology	Cystic lesion (odontogenic keratocysts, dentigerous cyst, periapical cyst)	Panoramic images; CBCT images (299 × 299 pixels)	Histopathologic examination	Panoramic images: AUC: 0.847, Se: 88.2%, Sp: 77.0%. CBCT images: AUC: 0.914, Se: 96.1%, Sp: 77.1%	Using CBCT images the CNN demonstrates higher diagnostic performance than using panoramic images	The diagnostic accuracy of OCLs using radiologic assessment alone is less than that using histologic examination, and accurate diagnosis with radiologic images only is still challenging

(*continued on next page*)

Table I. Continued

Study	*Architecture*	*Training/testing modalities*	*Dental specialty*	*Object*	*Imaging modalities*	*Reference standard*	*Results*	*Conclusions*	*Limitations*
Poedjiastoeti and Suebnukarn[27]	CNN (VGG-16)	Training set 400 Testing set: 100	Oral pathology	Ameloblastomas/ keratocystic odontogenic tumors	Panoramic radiographs	Histopathologic examination	Se: 81.5%; Sp: 83.3%; Ac: 83.0% (38 s for response)	Ac of CNN comparable to that of manual diagnosis by oral and maxillofacial specialists	Only frontal radiographs; no use of medical histories; no information about the characteristics of the images used
Ariji et al.[31]	CNN (AlexNet)	Testing set: 127 prediagnosed CT images of positive cervical lymph nodes and 314 CT images of prediagnosed negative lymph nodes from 45 patients with oral squamous cell carcinoma	Oral medicine and radiology	Cervical lymph nodes	CT images	Premade histologic analysis	Diagnostic Ac of 78.2%,	CNN may be valuable for diagnostic support	Limited sample size
Uthoff et al.[24]	CNN	—	Oral pathology	Oral cancer detection	Pictures taken with smartphone	Oral oncology specialist	Se: 0.850; Sp: 0.887; PPV: 0.877; NPV = 0.855	Initial feedback is positive	No biopsies; data set size
Yu et al.[25]	CNN	12 samples with tumorous tissues; 12 samples with non-tumorous tissue; 80% used for training, 20% for testing	Oral pathology	Tongue squamous cell carcinoma	Raman spectroscopy	Premade diagnosis	Se: 99.31%; Sp: 94.40%; Precision: 94.70%; Ac: 96.90%	The high sensitivity and specificity of the CNN would be helpful in obtaining adequate resection margins	Small sample size
Kats et al.[32]	CNN (Region based)	65 prediagnosed panoramic images with atherosclerotic carotid plaques (ACPs).	Oral medicine and radiology	Carotid artery	Panoramic radiographs	Not specified	Se = 75%; Sp = 80%; Ac = 83%.	Further improvements are needed to apply CNN	Relatively limited sample size

(*continued on next page*)

Table I. Continued

Study	*Architecture*	*Training/testing modalities*	*Dental specialty*	*Object*	*Imaging modalities*	*Reference standard*	*Results*	*Conclusions*	*Limitations*
Kim et al.[33]	DL (DeepSurv)	Training set: 179 patients Testing set: 76 patients	Oral pathology	Survival to oral cancer	Data from one database (sex, age, site, histologic grade, TNM stage, T stage, N stage, others)	Clinical data	C-index (training set) = 0.810; C-index (testing set) = 0.781	Survival prediction may be improved by using DL	Small data set of one single center
Krois et al.[23]	CNN	Training set: 1456 Testing set: 200	Periodontology	Periodontal bone loss	Panoramic radiographs	Three examiners	Ac = 0.81 (0.02); Se = 0.81 (0.04); Sp = 0.81 (0.05)	A CNN showed discrimination ability at least similar to that of dentists	Small data set; manually cropped images
Lee et al.[30]	CNN (AlexNet)	Training set: 535 healthy; 533 with osteoporosis Testing set: 200	Oral medicine and radiology	Osteoporosis	Panoramic radiographs	The diagnosis was made when 2 observers agreed, and a diagnosis of osteoporosis was made when cortical erosion was observed	AUC for different approaches ranged from 0.9763 to 0.9991	High agreement with experienced oral and maxillofacial radiologists in detecting osteoporosis	Reliability of examiners
Park et al.[36]	CNN (YOLO; SSD)	Training set: 1028 Testing set: 283	Orthodontics	Cephalometric landmarks	Lateral cephalograms (608 × 608 pixels)	One single expert examiner	5% higher accuracy than other methods	YOLO presented higher accuracy and processing speed compared to SSD	Reliability of examiners
Patcas et al.[37]	CNN	Trained on >0.5 million images	Orthodontics, oral surgery, maxillofacial surgery	Facial attractiveness, estimated age	Pictures (256 × 256 pixels)	Not reported	—	CNN might be considered to score facial attractiveness and apparent age in orthognathic patients	Reliability of the standard
Shoukri et al.[39]	DL	Trained on 259 condyles (105 controls and 154 with osteoarthritis); tested on 34 condyles	Oral medicine	TMJ osteoarthritis	High-resolution CBCT scans	Dental specialist through clinical examination	Ac = 73.5% DL	High degree of conformity in classifying and categorizing condyles	Reliability of examiners

R-CNN, regional convolutional neural network; *CNN*, convolutional neural network; *CT*, computed tomography; *Ac*, accuracy; *DL*, deep learning; *NILT*, near-infrared light transillumination; *AUC*, area under the curve; *Se*, sensitivity; *Sp*, specificity; *PPV*, positive predictive value; *NPV*, negative predictive value; *ENT*, ear-nose-throat specialist; *CBCT*, cone beam computed tomography; *OCL*, Oral cystic lesion; *TNM*, tumor-node-metastasis; *YOLO*, you only look once; *SSD*, Single shot detector; *TMJ*, temporomandibular joint.

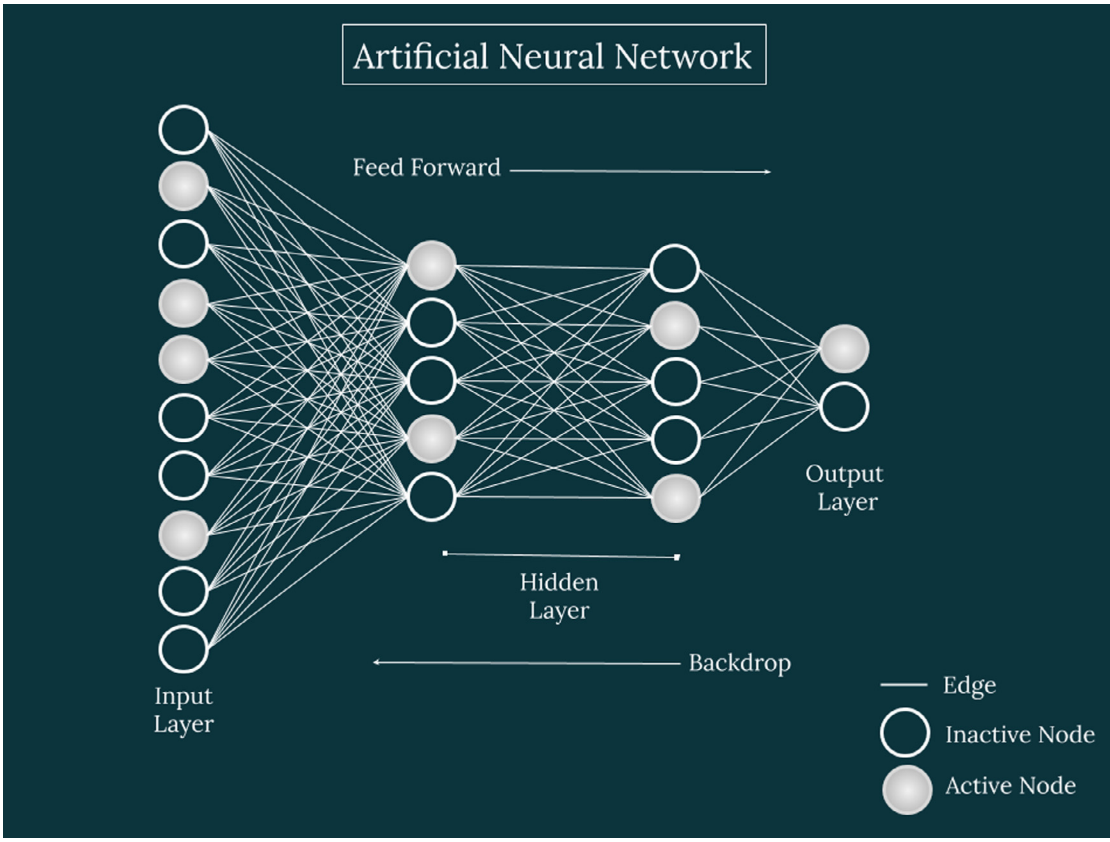

Fig. 2. Architecture of artificial neural networks (ANNs). The feed-forward mechanism represents the flow of information from input nodes toward output nodes through which an ANN is trained. The backdrop mechanism represents the flow of information from output nodes toward the input nodes; in this regard, the ANN learns through feedback of the results. Designing an ANN architecture demands consideration of various parameters such as the number of nodes in each layer, the value assigned to every individual node, the corresponding weight on each edge (connection between 2 nodes), and the bias assigned to each node in subsequent layers. To put the complexity of an ANN in perspective, the presented simple illustration of an ANN with an input layer with 10 nodes, 2 hidden layers with 5 nodes each, and an output layer with 2 nodes has 85 weighted edges $\{(10 \times 5) + (5 \times 5) + (5 \times 2)\}$ and 12 biases $\{5 + 5 + 2\}$ for a total of 97 parameters. In designing radiographic image recognition software on a standard 1024×768 pixel resolution image, where each pixel acts as an input node, the numbers of parameters in the architecture will increase to 4,024,355.

the anatomic structures or pathologic conditions of different points on images requiring analysis.

By examining radiographic images, it is relatively easy for clinicians to recognize the positions and characteristics of teeth as part of their preliminary examinations. In general, CNNs have performed well in recognizing and classifying teeth from both tridimensional images (i.e., CBCT)[12,13] and 2D panoramic and periapical radiographs.[14-16] In the same way, CNNs have been applied in the field of endodontics to interpret the anatomy of first molar roots and identify the presence of abnormalities in panoramic radiographs.[17] Two studies used CNNs to detect distal root abnormalities from panoramic radiographs in mandibular first molars. Although the accuracy of the tested systems (GoogLeNet and AlexNet) was 85.4% and 87.4%, respectively, both studies demonstrated a number of methodological limitations including sample selection and examiner reliability.

Identifying anatomic structures with the aid of CNNs is important when imaging cardinal surgical structures during surgical interventions. The potential to detect the course of the inferior alveolar nerve and establish its spatial relationship with the roots of third molars is fundamental to the preplanning process for surgical extraction of third molars.[18] Though most pathologic conditions that affect the oral cavity are initially identified by clinicians through visual assessment, certain cases that require further confirmation can be assessed with clinical radiographs and histopathologic examinations.

In general, CNNs and DL have been studied as adjuncts in the identification of dental caries, periapical lesions, and periodontal bone resorption from digital radiographs and other digital images.

In clinical practice, the diagnosis of caries is performed through visual observation and radiographs to identify alterations in the appearance of teeth after the loss of enamel and dentin. Near-infrared transillumination images are produced to help clinicians diagnose the presence of dental caries without the aid of radiographs. CNNs have been demonstrated to correctly detect the presence of caries in approximately 4 out of 5 cases from these types of imaging data.[19,20] When evaluating the potential presence of caries using CNNs and periapical radiographs, research has identified them as the most effective methods for detecting caries in experimental settings, with a diagnostic accuracy of 89% in premolars and 88% in molars.[21]

In the case of dental pathoses, evidence of the presence of a periapical radiolucency may be important in the detection of chronic inflammatory processes involving teeth. Periapical radiolucencies can be detected through the presence and interpretation of specific radiographic signs. One study used images cropped from panoramic radiographs to test CNNs for the detection of apical lesions, concluding that the discriminatory ability of CNNs was satisfactory and highly sensitive in the imaging of molars, likely because of the reduced distortion of the images in the posterior areas of the jaw.[22]

Periodontal bone loss can be evaluated with dental radiographs and is one of the parameters for classifying periodontal health and diseases. In a study comparing the ability to detect periodontal bone loss in panoramic radiographs, no differences were determined between the discriminatory ability of CNNs and specialists.[23] In other examples, ML has been applied to identify cancerous or precancerous lesions of the tongue and oral mucosa.[24,25] Though the results of studies in this field are preliminary and based on a few cases, the authors have stressed the potential advantages of using diagnostic support to make timely diagnoses of malignant lesions in regions with scarce medical coverage. Potentially malignant lesions can also be detected as radiolucent areas in panoramic radiographs by using CNNs.[26,27]

Several studies have evaluated the use of CNNs in detecting pathologic conditions outside the oral cavity by analyzing radiographs that are typically used and prescribed in dental settings. Existing research has tested CNNs for the detection of radiographic signs of maxillary sinusitis in panoramic radiographs and other types of radiographs that are commonly used or prescribed in dental diagnosis and treatment.[28,29]

Dental radiographs (both bidimensional and tridimensional) can provide information about pathologic processes and anatomic landmarks that are not strictly related to dentistry. In this regard, CNNs have been applied to aid in the diagnosis of osteoporosis from panoramic radiographs,[30] detect the presence of reactive lymph nodes in patients with oral squamous cell carcinomas in sections of CT,[31] and determine the presence of atherosclerotic plaques in carotid arteries in panoramic radiographs.[32] Although most authors have admitted that improvements are needed to increase the efficacy of CNNs in the field of dentistry, with time, DL that is applied to dental diagnostic imagery analysis may also provide support for other medical specialists in the diagnoses of lesions and diseases.

Other applications of deep learning in dentistry

In addition to investigations of DL as a diagnostic aid, CNNs have been studied for their benefits in assessing prognosis, diagnostic imaging segmentation, and other applications. As an example, to test the potential to estimate the prognoses of individuals with oral cancer, a single cohort was submitted to CNNs with information derived from numerous large subject databases.[33] Similarly, CNNs were studied to evaluate whether the prognoses of periodontally compromised teeth could be predicted through analyses of periapical radiographs.[34] Forthcoming research may use the scenario of large patient databases to enhance the potential application of CNNs.

In periodontology, DL has been applied in the profiling of periodontal microbial patterns related to the presence of oral malodor.[35] Moreover, CNNs have been applied to help in the detection of landmarks on lateral cephalometric radiographs in the field of orthodontics.[36] CNNs have also been used to determine increases in attractiveness after orthodontic treatment.[37]

CONCLUSIONS AND FUTURE APPLICATIONS

The present review explored the existing literature on the applications of DL in dentistry. The main application included the use of CNN techniques (CNNs are a type of DL) to analyze images predominantly collected with radiographs (periapical, panoramic, and CT scans) that were cropped and processed using these methods. In most of the studies, the results of the CNNs were compared with expert opinions during specific clinical scenarios, from tooth classification and numbering to periodontal diagnosis, and included evaluations for the presence of radiolucent lesions and other anatomic signs and features.[15,22,23,26-30,38] Although the reliability—or ability to provide consistent and correct measurements and diagnoses—of human (reference) examiners was questionable or potentially inaccurate,[22,26,36] the CNN methods appeared to demonstrate comparable accuracy, specificity, or sensitivity under most of the experimental conditions.[23,27,28,30]

One diagnostic limitation of CNNs was represented by the size and characteristics of the images used as input data. In most of the studies,[14,21,34,37,38] radiographs and photographs were cropped and analyzed in low resolution because the calculations needed to be made less complex, reducing the time needed to obtain a response through analysis. Although many of the authors reported significantly high reliability in CNNs when they were tested as supportive tools for diagnosis, it is likely that an increase in image resolution would result in improved diagnostic accuracy. The use of CNNs can also lead to the overfitting of data, thus limiting the external validity of results. This issue is strictly related to how the training sets were selected or inadequately considered within the reviewed studies.

We can hypothesize that AI may provide adjunct support in the diagnoses of lesions and conditions from radiographs in dental settings and may prove particularly useful for dental students, junior dentists, general practitioners, and specialists. This is because the direct adoption of ML methods may lead to an ethical debate on the diagnostic credibility of the techniques, which ought to be thoroughly investigated. Such a hypothesis found significant support in the papers included in the present review. Increasing the diagnostic accuracy of CNNs could also prove vital to the early diagnosis of oral cancer with intraoral photographs.[24] Although these methods have yet to be validated, research in this field should be encouraged because early diagnosis of oral ailments, such as cancer, can drastically affect the prognosis and global outcome of a disease.

According to the existing literature, CNNs have proven useful when applied to the discrimination and classification of teeth and anatomic structures.[12-15,17] Future developments may indicate that integrating discriminating and classifying tools with commonly used software for surgical preplanning may benefit in the discrimination of liable anatomic structures that require assessment during technique-sensitive procedures. However, more research in this field is required so that the current processing prowess of CNNs results in accurate analysis or interpretation.

One model demonstrated limitations in using DL to predict tooth prognoses with periapical radiographs, because the periodontal evaluation consisted of parameters that are more commonly derived from clinical procedures than radiographs.[34] By comparison, another study demonstrated that DL can be used to improve the predictability of survival in a subject with oral cancer after analyzing data about the subject and pathology from a singular source database.[33]

In brief, future studies should apply sound methodology by using rigorous internal and external validation to account for the biases that may arise from their sample selections. Specifically, randomly selected samples in diagnostic studies should be representative of the cases and controls and should be of sufficient size to account for differences in population. Measurements of the outcomes should be standardized and accurate, presenting both sensitivity and specificity in diagnostic studies, and the statistical analyses should be clearly described. Finally, even when specialists are involved, cases and controls should be defined according to the parameters of the studies' objectives rather than to the human examiners' opinions. The examiners' selections should only be adopted when more objective methods are infeasible.

In conclusion, there exists no sufficient evidence to support the routine use of CNNs as diagnostic support tools within clinical practice, because existing literature on the subject is sparse. In addition, some criticism can be raised about examiner reliability and the number of images that CNNs can process. Nonetheless, because the results from existing research are encouraging, and the number of papers pertaining to the topic are growing year by year, it is likely that scientific evidence on the use of CNNs in dentistry will gradually increase with time. Therefore, more studies are needed to confirm the results presented in this narrative review.

REFERENCES

1. Greenhill DAT, Edmunds DBR. A primer of AI in medicine. *Tech Gastrointest Endosc*. 2020;22(2):85-89.
2. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1:e271-e297.
3. Shen J, Zhang CJP, Jiang B, et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. *JMIR Med Inform*. 2019;7(3):e10010.
4. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44-56.
5. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318:517-518.
6. Bahner JE, Hüper A-D, Manzey D. Misuse of automated decision aids: complacency, automation bias and the impact of training experience. *Int J Human Comput Stud*. 2008;66:688-699.
7. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1:39. https://doi.org/10.1038/s41746-018-0040-6.
8. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118.
9. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA*. 2016;316:2353-2354.
10. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.
11. Kulkarni S, Seneviratne N, Baig MS, Khan AHA. Artificial intelligence in medicine: where are we now? *Acad Radiol*. 2020;27:62-70.
12. Miki Y, Muramatsu C, Hayashi T, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput Biol Med*. 2017;80:24-29.
13. Xu X, Liu C, Zheng Y. 3 D tooth segmentation and labeling using deep convolutional neural networks. *IEEE Trans Vis Comput Graph*. 2019;25:2336-2348.

14. Chen H, Zhang K, Lyu P, et al. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci Rep*. 2019;9(1):3840. https://doi.org/10.1038/s41598-019-40414-y.
15. Tuzoff DV, Tuzova LN, Bornstein MM, et al. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofac Radiol*. 2019;48(4):20180051.
16. Zhang K, Wu J, Chen H, Lyu P. An effective teeth recognition method using label tree with cascade network structure. *Comput Med Imaging Graph*. 2018;68:61-70.
17. Hiraiwa T, Ariji Y, Fukuda M, et al. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. *Dentomaxillofac Radiol*. 2019;48(3):20180218.
18. Vinayahalingam S, Xi T, Berge S, Maal T, de Jong G. Automated detection of third molars and mandibular nerve by deep learning. *Sci Rep*. 2019;9(1):9007.
19. Casalegno F, Newton T, Daher R, et al. Caries detection with near-infrared transillumination using deep learning. *J Dent Res*. 2019;98:1227-1233.
20. Schwendicke F, Elhennawy K, Paris S, Friebertshauser P, Krois J. Deep learning for caries lesion detection in near-infrared light transillumination images: a pilot study. *J Dent*. 2020;92:103260.
21. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent*. 2018;77:106-111.
22. Ekert T, Krois J, Meinhold L, et al. Deep learning for the radiographic detection of apical lesions. *J Endod*. 2019;45. 917-922. e915.
23. Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep*. 2019;9 (1):8495.
24. Uthoff RD, Song B, Sunny S, et al. Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities. *PLoS One*. 2018;13(12):e0207493.
25. Yu M, Yan H, Xia J, et al. Deep convolutional neural networks for tongue squamous cell carcinoma classification using Raman spectroscopy. *Photodiagnosis Photodyn Ther*. 2019;26:430-435.
26. Ariji Y, Yanashita Y, Kutsuna S, et al. Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2019;128:424-430.
27. Poedjiastoeti W, Suebnukarn S. Application of convolutional neural network in the diagnosis of jaw tumors. *Healthc Inform Res*. 2018;24:236-241.
28. Kim Y, Lee KJ, Sunwoo L, et al. Deep learning in diagnosis of maxillary sinusitis using conventional radiography. *Invest Radiol*. 2019;54:7-15.
29. Murata M, Ariji Y, Ohashi Y, et al. Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. *Oral Radiol*. 2019;35:301-307.
30. Lee JS, Adhikari S, Liu L, Jeong HG, Kim H, Yoon SJ. Osteoporosis detection in panoramic radiographs using a deep convolutional neural network–based computer-assisted diagnosis system: a preliminary study. *Dentomaxillofac Radiol*. 2019;48 (1):20170344. https://doi.org/10.1259/dmfr.20170344.
31. Ariji Y, Fukuda M, Kise Y, et al. Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2019;127:458-463.
32. Kats L, Vered M, Zlotogorski-Hurvitz A, Harpaz I. Atherosclerotic carotid plaque on panoramic radiographs: neural network detection. *Int J Comput Dent*. 2019;22:163-169.
33. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning−based survival prediction of oral cancer patients. *Sci Rep*. 2019;9(1):6994.
34. Lee JH, Kim DH, Jeong SN, Choi SH. Diagnosis and prediction of periodontally compromised teeth using a deep learning−based convolutional neural network algorithm. *J Periodontal Implant Sci*. 2018;48:114-123.
35. Nakano Y, Suzuki N, Kuwata F. Predicting oral malodour based on the microbiota in saliva samples using a deep learning approach. *BMC Oral Health*. 2018;18(1):128.
36. Park JH, Hwang HW, Moon JH, et al. Automated identification of cephalometric landmarks: part 1—comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod*. 2019;89(6):903-909.
37. Patcas R, Bernini DAJ, Volokitin A, Agustsson E, Rothe R, Timofte R. Applying artificial intelligence to assess the impact of orthognathic treatment on facial attractiveness and estimated age. *Int J Oral Maxillofac Surg*. 2019;48:77-83.
38. Lee JH, Kim DH, Jeong SN. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral Dis*. 2020;26:152-158.
39. Prieto JC, Ruellas A, Yatabe M, Sugai J, Styner M, Zhu H, Huang C, Paniagua B, Aronovich S, Ashman L, Benavides E, de Dumast P, Ribera NT, Mirabel C, Michoud L, Allohaibi Z, Ioshida M, Bittencourt L, Fattori L, Gomes LR, Cevidanes L. Minimally Invasive Approach for Diagnosing TMJ Osteoarthritis. *J Dent Res*. 2019;98 (10):1103-1111. https://doi.org/10.1177/0022034519865187.

Reprint requests:

Dr. Stefano Corbella
IRCCS Istituto Ortopedico Galeazzi
via R. Galeazzi
4, 20161 Milan
Italy
Stefano.corbella@gmail.com