Artificial intelligence for detection of periapical lesions on intraoral radiographs: Comparison between convolutional neural networks and human observers

Ruben Pauwels, MS, PhD,[a,b] Danieli Moura Brasil, DDS, MS, PhD,[c] Mayra Cristina Yamasaki, DDS, MS, PhD,[c] Reinhilde Jacobs, DDS, MS, PhD,[b,d] Hilde Bosmans, MS, PhD,[b] Deborah Queiroz Freitas, DDS, MS, PhD,[c] and Francisco Haiter-Neto, DDS, MS, PhD[c]

Objective. The aim of this study was to compare the diagnostic performance of convolutional neural networks (CNNs) with the performance of human observers for the detection of simulated periapical lesions on periapical radiographs.

Study Design. Ten sockets were prepared in bovine ribs. Periapical defects of 3 sizes were sequentially created. Periapical radiographs were acquired of each socket with no lesion and with each lesion size with a photostimulable storage phosphor system. Radiographs were evaluated with no filter and with 6 image filter settings. A CNN architecture was set up using Keras-TensorFlow. Separate CNNs were evaluated for randomly sampled training/validation data and for data split up by socket (5-fold cross-validation) and filter (7-fold cross-validation). CNN performance on validation data was compared with that of 3 oral radiologists for sensitivity, specificity, and area under the receiver operating characteristic curve (ROC-AUC).

Results. Using random sampling, the CNN showed perfect accuracy for the validation data. When data were split up by socket, the mean sensitivity, specificity, and ROC-AUC values were 0.79, 0.88, and 0.86, respectively; when split up by filter, they were 0.87, 0.98, and 0.93, respectively. For radiologists, the values were 0.58, 0.83, and 0.75, respectively.

Conclusions. CNNs show promise in periapical lesion detection. The pretrained CNN model yielded in this study can be used for further training on larger samples and/or clinical radiographs. (Oral Surg Oral Med Oral Pathol Oral Radiol 2021;131:610−616)

Radiographic examination is essential in the diagnosis of periapical lesions. Though intraoral radiography (IOR) is commonly used for this purpose, it has shown limited diagnostic efficacy for small, periapical bony lesions because of anatomic overlap.[1-4] Cone beam computed tomography (CBCT) allows for an accurate, 3-dimensional assessment of periapical lesions,[1-4] but because of its higher radiation dose it is not justified for routine diagnosis and follow-up.[5] Furthermore, the added dimension of CBCT may not improve the diagnosis of inflammation,[6] and the diagnostic value of CBCT can be severely affected by metal artefacts (e.g., from root canal fillings or adjacent implants).[7] Thus, IOR is expected to remain a frequently used imaging modality for routine periapical evaluation before, during, and after treatment.[5]

Because of the limited diagnostic efficacy of IOR for periapical lesions, several previous studies have evaluated techniques to optimize lesion detection, including through image postprocessing.[8,9] Image filtering is commonly applied to optimize the diagnostic efficacy of a radiograph for a given diagnostic task. The most commonly used filters either (1) increase sharpness by enhancing edges (e.g., unsharp masking) or (2) reduce noise, often at the cost of blurring the image (e.g., a Gaussian filter). Furthermore, filters may increase contrast for certain tissues and lesions by stretching part of the histogram.

Recently, the potential use of artificial intelligence (AI) as a diagnostic tool has gained attention in radiology.[10] Owing to innovations in deep learning, particularly convolutional neural networks (CNNs),[11] it is now feasible to train and evaluate AI systems for specific diagnostic tasks. Recent improvements in computing hardware, especially in the graphics processing unit (GPU), allow CNN models to be trained on very large and/or augmented data sets. The main potential benefits of AI in radiology are improved diagnostic efficacy and

Statement of Clinical Relevance

Convolutional neural networks are used to develop artificial intelligence systems for diagnostic tasks in oral and maxillofacial radiology. This study indicates that convolutional neural networks can yield diagnostic performance comparable to or better than that of human observers for detection of periapical lesions.

[a]Aarhus Institute of Advanced Studies, Aarhus University, Aarhus, Denmark.
[b]Department of Imaging & Pathology, Biomedical Sciences Group, Catholic University of Leuven, Leuven, Belgium.
[c]Department of Oral Diagnosis, Division of Oral Radiology, Piracicaba Dental School, University of Campinas, Piracicaba, Brazil.
[d]Department of Dental Medicine, Karolinska Institutet, Huddinge, Sweden.

Received for publication Oct 13, 2020; returned for revision Dec 22, 2020; accepted for publication Jan 18, 2021.

2212-4403/$-see front matter
https://doi.org/10.1016/j.oooo.2021.01.018

reduced image interpretation time.[10] In dentistry, several pilot studies have investigated the use of CNNs for various diagnostic tasks,[11] such as periodontal bone loss,[12-15] dental caries,[14,15] periapical/endodontic evaluation,[14,16] root fractures,[17] maxillary sinusitis,[18] and osteoporosis.[19] A previous study that involved CNN-based periapical lesion detection on IOR showed high performance on test images using transfer learning.[14] However, the data were labeled by clinicians (dentists and radiologists), making them prone to error, which may lead to suboptimal CNN training. To overcome this issue, simulated lesions can be used, allowing for perfect annotation of training and validation data. Furthermore, the aforementioned study[14] did not compare the performance of the trained CNN with that of human observers, which limits the conclusions that can be drawn in terms of CNN performance.

The aim of this study was to explore the use of CNNs for the detection of simulated periapical lesions on intraoral radiographs and to compare the performance of trained CNNs with that of human observers. The null hypothesis stated that there would be no differences in diagnostic performance in the detection of simulated lesions between the CNNs and human observers.

MATERIALS AND METHODS

Sample preparation

Sample preparation and human observation conditions have been described in Brasil et al.[9] In summary, 10 sockets were prepared in bovine ribs, and bone defects of incremental size (1.6 mm, 1.8 mm, and 2.1 mm) were created with a round carbide bur (KG Sorensen, Cotia, SP, Brazil) to simulate periapical lesions (Figure 1). A bovine tooth with a single, fully formed, straight root was used for all subsequent radiographs.

Periapical radiographs were acquired using size 2 VistaScan photostimulable storage phosphor plates (Dürr Dental AG, Bietigheim-Bissingen, Germany) and a FOCUS radiography unit (Instrumentarium Dental, Tuusula, Finland) at 70 kV, 7 mA, 0.08 second exposure time, and 40 cm source-to-object distance. For each socket, a radiograph was acquired before inducing the bone defect. Next, a radiograph was taken after each progressive enlargement of the defect. The aforementioned bovine tooth was placed in the socket

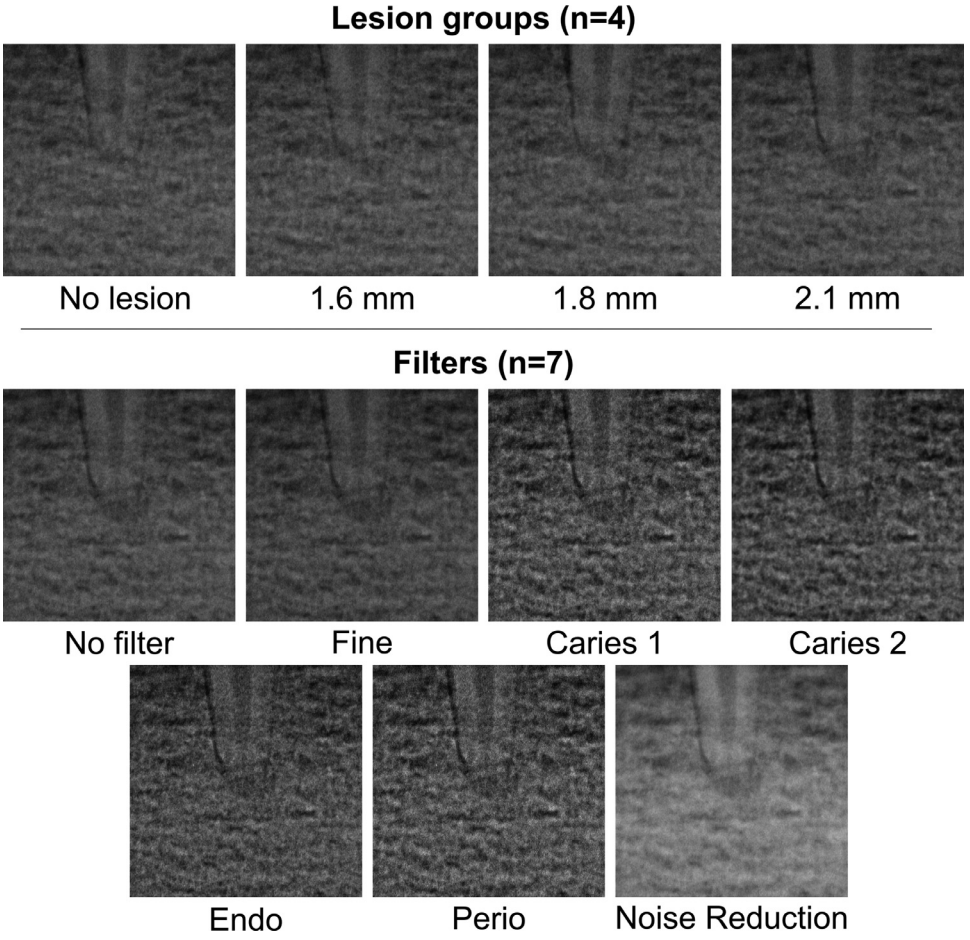

Fig. 1. Lesion groups and filters.

Table I. Mean gray value and noise measured in dentin and description of the filter according to the DBSWIN instruction manual

Filter	*Mean gray value*	*Noise (SD)*	*Description*
No filter	92	5.9	
Fine	86	6.0	General sharpening filter that enhances structures of 2 line pairs per millimeter
Caries 1	98	11.8	Enhances structures of 2 line pairs per millimeter with strong attenuation of low frequencies
Caries 2	94	7.8	Enhances structures of 2 line pairs per millimeter with strong attenuation of low and high frequencies
Endo	93	11.4	Enhances structures of 3 line pairs per millimeter with strong attenuation of low frequencies
Perio	103	16.9	Enhances structures of 2.5 line pairs per millimeter with strong attenuation of low frequencies
Noise reduction	138	6.4	Matches outlier pixels to their neighboring pixels

SD, standard deviation.

for each radiograph. An acrylic resin device was used to fix the exposure geometry, allowing for reproducible radiographs. To simulate soft tissue, a 2.5-cm−thick acrylic resin block was placed between the x-ray tube and the bovine rib.[9]

For each radiograph, along with an unfiltered image, 6 image filters (DBSWIN, Dürr Dental AG) were applied (Figure 1). Table I provides an indication of the differences between filters based on the mean gray value and noise (calculated as standard deviation) measured using a small region of interest inside the dentin. The total sample of 280 radiographs (10 sockets × 4 lesion conditions × 7 filter settings) was exported in 8-bit TIF (tagged image file) format.

Data preparation and sampling for CNN

The radiographs were cropped to regions of interest of 256 × 256 pixels, centered on the root apex, by a researcher who was not involved in the human observation study. Data augmentation was performed through smoothening, adding noise, and affine transformations, resulting in 5600 unique images (1400 per each of the 4 lesion groups). Three different sampling methods were used to split up the data into training and validation subsets (Figure 2):

- Random sampling: The entire data set was randomized and subsequently split into training (60%) and validation (40%) data. This approach assessed the overall ability of the CNN to extract the expected output information from the input data; it primarily served as a validation of the CNN architecture and hyperparameters rather than an assessment of CNN performance vs human observers.
- Sampling by socket (5-fold cross-validation): The data were split up by socket. Because there were 10 sockets in the sample, the training data consisted of 80% of the data set (i.e., images from 8 sockets) and the validation data comprised images from the 2 remaining sockets. It was ensured that all 10 sockets were included in the validation data once; that is, sockets 1 and 2 for the first fold, sockets 3 and 4 for the second fold, and so on for a total of 5-fold cross-validation. CNNs were trained to detect a lesion but not for classification of lesion size. This approach evaluated the robustness of a CNN model to identify lesions in anatomic locations it has not encountered before (such as new teeth or new patients).
- Sampling by image filter (7-fold cross-validation): The data were split up by image filter. In a first evaluation, the training data consisted of 85.7% of the data set (i.e., 6 of the 7 filters). This approach evaluated the robustness of the CNN model to variations in image quality (brightness, contrast, sharpness, and noise).

CNN setup, training, and evaluation

A CNN architecture was set up using the Keras (v2.1.6, François Chollet and contributors) and TensorFlow (v1.12.0, Google Brain Team) application programming interface (Figure 3). The CNNs were trained using the binary cross-entropy loss function, a batch size of 16, and the AMSGrad optimizer.[20] Rectified Linear Unit (ReLU) activation was used for each layer, and sigmoid activation was used for the final classification. For the first sampling method, separate CNNs were trained for a binary evaluation of the presence of a lesion (yes/no) as well as a categorical evaluation of lesion size (none/small/medium/large). For the sampling methods involving cross-validation, because of the increased computational workload, CNNs were trained for lesion detection but not for classification of lesion size. Training was performed on a GeForce 1070 GTX GPU using CUDA v9.0 and cuDNN v7.4.1 (Nvidia, Santa Clara, CA, USA) for 100 epochs (lesion detection) and 500 epochs (lesion size classification); additional epochs did not improve performance for training or validation data. For the lesion detection task, the data were stratified to ensure an equal distribution of images with and without a lesion.

Human observation

As previously reported by Brasil et al.,[9] human observation was performed on the abovementioned sample

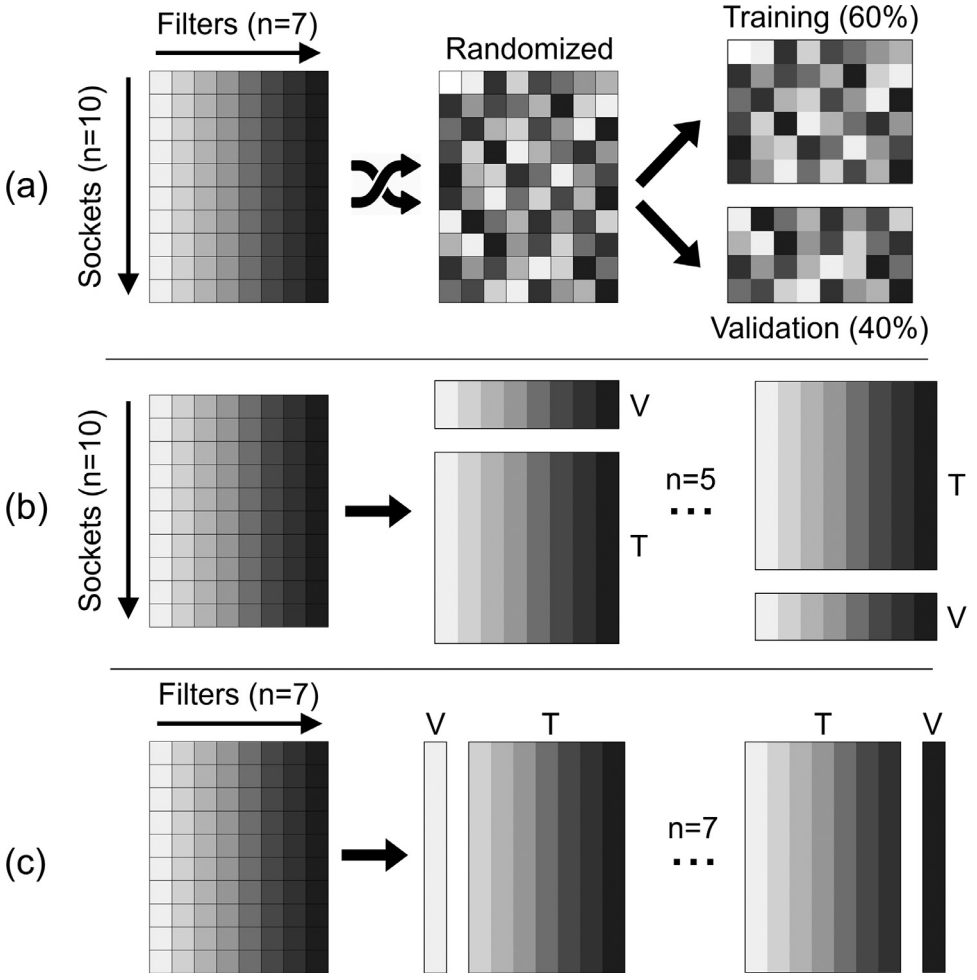

Fig. 2. Three complementary approaches for data sampling used in this study. **(A)** Random sampling. **(B)** Sampling by socket using 5-fold cross-validation. **(C)** Sampling by image filter using 7-fold cross-validation. *T*, training; *V*, validation.

of 280 radiographs by 3 oral radiologists, each with 3 years of experience in radiographic examination. A 5-point confidence scale regarding the presence of a periapical lesion was used: (1) *definitely absent*; (2) *probably absent*; (3) *unsure*; (4) *probably present*; (5) *definitely present*. All observations were performed in a quiet, darkened room, using a 24.1-in LCD monitor (MDRC-2124, Barco, Kortrijk, Belgium) with a screen resolution of 1920 × 1200 pixels. The observers were allowed to use a zoom tool but could not adjust brightness or contrast. Observers were instructed to assess a maximum of 20 images on a given day and to have an interval of at least 2 days between sessions to avoid visual fatigue and memorization. After 1 month, 30%

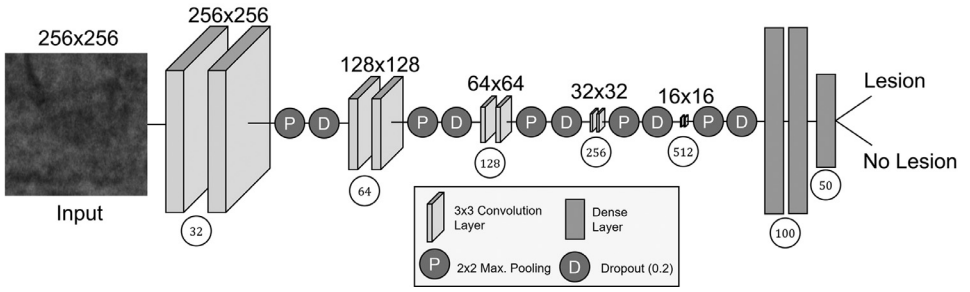

Fig. 3. Architecture of the convolutional neural network used in this study for binary evaluation of the presence of a lesion. The number of trainable parameters was 5,546,200. For the second convolutional neural network, used for categorical evaluation of lesion size, the last dense layer was altered to have 4 output categories; the other network components were identical.

of the periapical radiographs from each group were reassessed under the same conditions to determine intraobserver agreement. Interobserver agreement between the 3 radiologists was also calculated.

Statistical analysis

The performance of the CNN in terms of lesion detection on validation data was compared with the performance of the human observers reported in the study by Brasil et al.[9] Sensitivity (Sn), specificity (Sp), and receiver operating characteristic (ROC) area under the curve (AUC) values were determined for each CNN and each observer. For Sn and Sp assessment, confidence level scores for the presence of a lesion were dichotomized. Scores of 1, 2, and 3 were equated to 'no lesion present' and scores of 4 and 5 to 'lesion present'. Intra- and interobserver agreement was determined by using the quadratic weighted kappa (κ). The statistical analyses were performed using Excel (Microsoft, Redmond, WA, USA) and SPSS Statistics 22.0 (IBM, Armonk, NY, USA).

RESULTS

The computational time for CNN training was approximately 3 minutes and 1 minute 45 seconds per epoch for lesion detection and size classification, respectively.

Random sampling

The diagnostic performance in lesion detection as well as size classification of the CNN was 100% for both the training and validation data, implying 1.00 Sn and Sp and an ROC-AUC of 1.00. In comparison, human observers showed an Sn of 0.58 (ranging among the observers from a minimum of 0.26 to a maximum of 0.86), an Sp of 0.83 (minimum 0.30, maximum 1.00), and an AUC of 0.75 (minimum 0.63, maximum 0.88) for lesion detection (Table II).[9]

Sampling by socket

The mean values of Sn, Sp, and ROC-AUC per fold were higher than the corresponding mean values of the radiologists (Table II). However, variable performance was seen between the 5 folds. This is particularly apparent for Sn, for which one of the folds resulted in a value of 0.50, which would be equivalent to random guessing.

Sampling by image filter

The mean values of Sn, Sp, and ROC-AUC per fold were also greater than the corresponding values calculated from the radiologists' interpretations. Whereas 6 folds showed perfect performance on the validation data (Sn and Sp of 1.00), 1 of the folds showed very poor performance for Sn (0.12) in which almost every validation image was classified as 'no lesion', resulting in a somewhat low overall Sn but a high Sp (Table II). It was found that the fold with poor Sn comprised validation data with the noise reduction filter. As seen in Figure 1 and Table I, this filter resulted in a considerably brighter image (indicated by the high mean gray value of the pixels) compared with all other filters.

For the trained CNNs, because of their inherent deterministic nature, agreement between repeated predictions on validation data (i.e., equivalent to intraobserver agreement) was κ=1.00. As previously reported by Brasil et al.,[9] intra- and interobserver agreement for the 3 radiologists ranged from κ=0.69 to 0.80 and κ=0.61 to 0.67, respectively. These values represent substantial agreement.[21]

DISCUSSION

The present study explored the potential of a CNN-based assessment of periapical lesions. Although overall performance of trained CNN models surpassed that of oral radiologists, the performance was shown to be sensitive to the manner in which training and validation data were split up.

An interesting finding was the poor performance of the CNN on images using 1 out of 7 filters (noise reduction). Though CNNs are believed to be highly robust to variations in image quality, brightness, and contrast, the variation between this filter and the other 6 greatly affected CNN performance when the filter was not included in the training data. This is likely due to the overall difference in brightness between the noise reduction filter and all other filters (Table I). Whereas adjustments of brightness and contrast are

Table II. CNN performance for lesion detection vs human observers.

	Sensitivity (min-max)	*Specificity (min-max)*	*ROC-AUC (min-max)*
CNN-Random sampling	1.00	1.00	1.00
CNN-Sampling by socket	0.79 (0.50-1.00)	0.88 (0.77-1.00)	0.86 (0.64-1.00)
CNN-Sampling by image filter	0.87 (0.12-1.00)	0.98 (0.88-1.00)	0.93 (0.48-1.00)
Human observers[9]	0.58 (0.26-0.86)	0.83 (0.30-1.00)	0.75 (0.63-0.88)

For the human observers, "min-max" refers to values from the worst and best observers. For the CNN, "min-max" refers to the worst and best fold during cross-validation.
CNN, convolutional neural network; *ROC*, receiver operating characteristic; *AUC*, area under the curve

often included during data augmentation, this was not the case in this study due to the possibility of pixel value saturation (i.e., gray values reaching the lower or upper limit of the grayscale). On the other hand, the 3-fold range in noise levels between filters shown in Table I did not affect the CNN outcome. This can be explained by the fact that data augmentation involved both smoothening and addition of noise. Although the robustness of CNNs to image filters warrants further investigation, it can be recommended that training data should always comprise different filters and/or representative data augmentation. Furthermore, this study involved 8-bit images; it remains to be seen whether 16-bit images are of benefit for CNN performance.

The design of the CNN in this study was relatively straightforward, using a stepwise increase in layer depth along with a commensurate decrease in resolution through max-pooling layers, ending with 3 fully connected (dense) layers. Similar CNN architectures, such as AlexNet[22] and VGG, were used in previous diagnostic AI studies.[12,18,23,24] Other studies have used more complex architectures, such as Inception networks.[14,25] Though the inherent black box nature of a CNN makes it difficult to judge the optimal CNN architecture, an overly complex or deep CNN may lead to overfitting and, as a result, poor performance on test data. Future research will focus on adjusting CNN architectures and hyperparameters on the fly during training. A post hoc analysis of CNN models with complex architectures may also reveal connections or pathways that serve no meaningful purpose, allowing a simplified model to be derived, but this is not yet common practice in radiologic AI research.

Previous diagnostic studies using CNNs on intraoral radiographs focused on periapical evaluation,[14,16] dental caries,[14,15] and periodontal assessment.[12,13] Generally, high diagnostic accuracy after training was found. A prior investigation involving CNN-based periapical assessment showed an accuracy of 0.88 using transfer learning, but no comparison was made with human observers.[14] Furthermore, data labeling was performed by dentists and radiologists.[14] The overall trend in radiologic AI research is that, when an adequate ground truth is available for data labeling, trained CNNs can reach or surpass the diagnostic performance of experienced clinicians. The performance of CNNs for various tasks has already resulted in concern regarding the potential replacement of radiologists by AI systems.[26,27] Though it is too early to predict the eventual role of AI in diagnostic radiology, it is pivotal to prepare for the clinical implementation of AI tools through the development of dedicated guidelines and ethical frameworks.[10,28]

This study was explorative because of the limited sample size and the use of bovine ribs and simulated lesions. On the other hand, the use of simulated lesions allowed for perfect data labeling, avoiding the need for human data annotation. Though a large sample of clinical IORs could be collected and annotated by clinicians, this approach would imply that the CNN is only able to reach the same diagnostic efficacy as the dentist (s) or radiologist(s) who performed the annotation. Because periapical lesions are often misdiagnosed on IOR, the use of simulated lesions was preferred to ensure that initial training did not include images with erroneous labels. Although it is highly promising that CNNs in our study tended to outperform human observers after training on as few as 8 sockets, the trained CNN model(s) yielded by this study require further validation before considering implementation. Different approaches are possible, such as supervised learning on clinical data with an accurate ground truth or reinforcement learning.[10] Furthermore, this study involved considerable manual cropping of the radiographs to pinpoint a region of interest around the apex. It would be conceivable to automate this process by training a separate CNN for root tip identification. The feasibility of tooth identification and labeling has been demonstrated in several previous studies.[25,29]

CONCLUSION

Taking into account the limited sample size and simulated conditions, this preliminary research shows promising potential for using CNN in periapical lesion detection. The pretrained CNN model yielded in this study can be used for further training on larger samples and/or clinical radiographs.

FUNDING

Ruben Pauwels is supported by the European Union Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant agreement no. 754513 and by Aarhus University Research Foundation (AIAS-COFUND). Financial support was received by the Research Foundation−Flanders (FWO) under grant no. V419619N, the internal research fund of Katholieke Universiteit Leuven (grant no. C24/18/065), and the National Council for Scientific and Technological Development.

PRESENTATION

This study was presented at the 22nd International Congress of DentoMaxilloFacial Radiology and at the 41st Jornada Odontológica de Ribeirão Preto.

REFERENCES

1. Sogur E, Baksi BG, Gröndahl HG, Lomcali G, Sen BH. Detectability of chemically induced periapical lesions by limited cone beam computed tomography, intra-oral digital and conventional film radiography. *Dentomaxillofac Radiol*. 2009;38:458-464.

2. Davies A, Mannocci F, Mitchell P, Andiappan M, Patel S. The detection of periapical pathoses in root filled teeth using single and parallax periapical radiographs versus cone beam computed tomography—a clinical study. *Int Endod J*. 2015;48:582-592.
3. Antony DP, Thomas T, Nivedhitha MS. Two-dimensional periapical, panoramic radiography versus three-dimensional cone-beam computed tomography in the detection of periapical lesion after endodontic treatment: a systematic review. *Cureus*. 2020;12:e7736.
4. Ramis-Alario A, Tarazona-Alvarez B, Cervera-Ballester J, et al. Comparison of diagnostic accuracy between periapical and panoramic radiographs and cone beam computed tomography in measuring the periapical area of teeth scheduled for periapical surgery. A cross-sectional study. *J Clin Exp Dent*. 2019;11:e732-e738.
5. International Atomic Energy Agency. *Radiation Protection in Dental Radiology. Safety Reports Series*. Vienna, Austria: International Atomic Energy Agency; 2021.
6. Kruse C, Spin-Neto R, Reibel J, Wenzel A, Kirkevang LL. Diagnostic validity of periapical radiography and CBCT for assessing periapical lesions that persist after endodontic surgery. *Dentomaxillofac Radiol*. 2017;46:20170210.
7. Mazzi-Chaves JF, de Faria Vasconcelos K, Pauwels R, Jacobs R, Sousa-Neto MD. Cone-beam computed tomographic−based assessment of filled C-shaped canals: artifact expression of cone-beam computed tomography as opposed to micro-computed tomography and nano-computed tomography [e-pub ahead of print]. *J Endod*. 2021. https://doi.org/10.1016/j.joen.2020.07.010. Accessed February 4.
8. Gaêta-Araujo H, Nascimento EHL, Brasil DM, Gomes AF, Freitas DQ, De Oliveira-Santos C. Detection of simulated periapical lesion in intraoral digital radiography with different brightness and contrast. *Eur Endod J*. 2019;4:133-138.
9. Brasil DM, Yamasaki MC, Santaella GM, Guido MCZ, Freitas DQ, Haiter-Neto F. Influence of VistaScan image enhancement filters on diagnosis of simulated periapical lesions on intraoral radiographs. *Dentomaxillofac Radiol*. 2019;48:20180146.
10. Pauwels R. A brief introduction to concepts and applications of artificial intelligence in dental imaging. *Oral Radiol*. 2021;37:153-160.
11. 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.
12. 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.
13. Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep*. 2019;9:8495.
14. Prajapati SA, Nagaraj R, Mitra S. Classification of dental diseases using CNN and transfer learning. In: *Paper presented at: 5th International Symposium on Computational and Business Intelligence (ISCBI)*, Dubai, UAE; 2021. August 11-14, 2017.
15. 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.
16. Yang J, Xie Y, Liu L, Xia B, Cao Z, Guo C. Automated dental image analysis by deep learning on small dataset. In: *Paper presented at: IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Tokyo, Japan; 2021. July 23-27, 2018.
17. Johari M, Esmaeili F, Andalib A, Garjani S, Saberkari H. Detection of vertical root fractures in intact and endodontically treated premolar teeth by designing a probabilistic neural network: an ex vivo study. *Dentomaxillofac Radiol*. 2017;46:20160107.
18. 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*. 2018;35:301-307.
19. Chu P, Bo C, Liang X, et al. Using Octuplet Siamese Network for osteoporosis analysis on dental panoramic radiographs. *Conf Proc IEEE Eng Med Biol Soc*. 2018;2018:2579-2582.
20. Reddi SJ, Kale S, Kumar S. On the convergence of Adam and beyond. arXiv 1904.09237v1 2021. https://arxiv.org/abs/1904.09237.
21. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977;33:363-374.
22. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, San Diego, CA, USA: Neural Information Processing Systems Foundation; 2012:1097-1105.
23. 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.
24. 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;48:20170344.
25. 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;9:3840.
26. Gong B, Nugent JP, Guest W, et al. Influence of artificial intelligence on Canadian medical students' preference for radiology specialty: a national survey study. *Acad Radiol*. 2019;26:566-577.
27. van Hoek J, Huber A, Leichtle A, et al. A survey on the future of radiology among radiologists, medical students and surgeons: students and surgeons tend to be more skeptical about artificial intelligence and radiologists may fear that other disciplines take over. *Eur J Radiol*. 2019;121:108742.
28. Geis JR, Brady A, Wu CC, et al. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Insights Imaging*. 2019;10:101.
29. 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.

Reprint requests:

Ruben Pauwels, MS, PhD
Aarhus Institute of Advanced Studies (AIAS)
Aarhus University, Høegh-Guldbergs Gade 6B
8000 Aarhus
Denmark.
pauwelsruben@hotmail.com