DETECTING APICAL RADIOLUCENCIES USING DEEP LEARNING TECHNOLOGY

Manal Hamdan

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the School of Dentistry (Oral and Maxillofacial Radiology).

Chapel Hill
2021

Approved by:

Donald Tyndall

André Mol

Peter Tawil

ABSTRACT

Manal Hamdan: Detecting Apical Radiolucencies Using Deep Learning Technology
(Under the direction of Donald Tyndall)

Objectives: To assess the effectiveness of the Denti.AI CAD system in assisting dentists with detecting apical radiolucencies on intraoral periapical radiographs.

Methods: Using CBCT as the ground truth reference, (n=68) IO radiographs were randomly selected to serve as the testing subset. Twelve readers were asked to view the subset and to record their confidence about the presence of apical radiolucencies. Readers analyzed the same images under two conditions: with and without AI predictions shown. The readers' performance for both conditions was compared. AFROC was chosen as the main metric of performance measurement.

Results: The AFROC-AUC metric showed a statistically significant improvement by 4.9 or 5.8% compared to the unaided reading session. Subgroup performance analysis showed a statistically significant improvement in the detection of mandibular molar lesions, small lesions, and in endodontically-treated teeth.

Conclusion: Using a limited testing dataset, Denti.AI improved localization of apical radiolucencies. Further AI training is necessary to increase the sensitivity and specificity of apical radiolucency detection.

To my pillars: my husband, my Dad and my Mom. Thank you for your endless love and encouragement! Without your support, this wouldn't be possible.
To my baby boy, you are the reason why!

ACKNOWLEDGEMENTS

A special thank you to my advisor, Dr. Donald Tyndall, and committee members Dr. André Mol and Dr. Peter Tawil for generously offering their support, time and expertise. I'm grateful for Lyudmila Tuzova and Denti.Ai for their statistical support.

I would also like to acknowledge all the wonderful people that helped make this journey unforgettable, Dr. Broome, Prof. Johnson, Prof. Chen, Sherri, Paulette and my brilliant co-residents!

TABLE OF CONTENTS

LIST OF FIGURES

Literature Review Figures:

Manuscript Figures:

LIST OF TABLES

LIST OF ABBREVIATIONS

AFROC	Alternative Free-Response Receiver Operating Characteristic
AI	Artificial Intelligence
AR	Apical Radiolucency
AUC	Area under the ROC curve
CAD	Computer-Aided Detection
CBCT	Cone Beam Computed Tomography
CCD	Charge-Coupled Device
CNN	Convolutional Neural Network
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
FPF	False Positive Fraction
IO	Intraoral Radiography
IRB	Institutional Review Board
LLF	Lesion Localization Fraction
MH	Manal Hamdan (Primary Investigator)
ML	Machine learning
PSP	Photostimulable Phosphor
ROC	Receiver Operating Characteristic (Curve)
TPF	True Positive Fraction
UNC	University of North Carolina at Chapel Hill

REVIEW OF THE LITERATURE

Introduction:

Technological advances in Oral and Maxillofacial Radiology are no longer limited to imaging modalities and hardware. The age of deep learning is upon us, with several companies marketing their online CAD tools for various diagnostic problems and procedural planning. This advancement necessitates that proper research be conducted to provide the scientific community with evidence-based knowledge. To be able to understand the current study, it is imperative to understand the main concepts that artificial intelligence encompasses.

Artificial intelligence, Machine Learning and Deep Learning:

AI is classically defined as the ability of computer systems to perform tasks conventionally done using human intelligence [(1)]. The term "artificial intelligence" was coined in 1956, at a workshop that took place in Dartmouth College. A 17- page typescript was authored in August 31, 1955 by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon proposing a summer research project on AI. Duplicates of the typescript are housed in the archives at Dartmouth College and Stanford University. The Dartmouth Project team proposed a simplified description of an otherwise complex problem. AI was explained as "making a machine behave in ways that would be called intelligent if a human were so behaving". Those scientists hypothesized that a machine can be trained to learn through experimental "trial and error" much like humans do [(2)]. Hence, this science is focused on engineering intelligent machines and computer systems that would primarily reduce time and effort spent on daily tasks.

The rapid advancement in AI in the past few decades was attainable largely due to the availability of "big data" or huge digital data sets as well as the jump in computational power, which doubles every two years per Moore's law [(3)]. The great eagerness for the development of AI systems in radiology is mirrored by the increase in publications on this topic. Pesapane et al. concluded in his review that publications on AI have drastically increased from about 100–150 per year in 2007–2008 to 700–800 per year in 2016–2017 [(4)]. Recent radiologic meetings, including the 2017 Annual Meeting of the Radiological Society of North America (RSNA), the 2018 European Congress, and the 2020 Annual Meeting of the American Academy of Oral and Maxillofacial Radiology, have also proven the interest in AI applications with many AI-related talks. With the growing interest in AI applications in radiology, the Radiological Society of North America started a journal dedicated to AI research called Radiology: Artificial Intelligence [(5)]. The aim of this journal is to highlight the emerging applications of AI in the field of imaging across multiple disciplines.

Conventional machine learning and deep learning are subsets of AI (Figure 1.1). Machine learning is relatively limited compared to deep learning in that it requires engineering and domain expertise in order to design a "feature extractor" (Figure 1.3). A feature extractor transforms raw data into a suitable internal representation or feature vector. Features can represent pixels of an image for example. This enables the learning subsystem, which is often a classifier, to detect and/or classify patterns seen in the input.

A key concept is that deep learning is a form of representation learning (Figure 1.3). Representation learning is defined as: "a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification". In contrast to ML, layers of features are not designed by human engineers. DL methods are

representation learning methods with multiple levels of representation. Those levels are obtained by composing simple but non-linear modules. Each module transforms the representation at one level, starting with the raw input, into a representation at a higher and slightly more abstract level. With multiple non-linear layers, very complex functions can be learned.

In tasks such as classification, higher layers of representation amplify aspects of the input important for discrimination, whereas irrelevant variations are suppressed. For instance, images come in the form of an array of pixel values; the learned features in the first layer of representation can represent the presence/absence of edges at particular orientations and locations in the image. In the second layer, motifs or patterns are detected by spotting particular arrangements of edges irrespective of small variations in the edge positions. Furthermore, a third layer may assemble these motifs into larger combinations that correspond to parts of an object. Succeeding layers would detect objects as combinations of these parts. In summary, the key aspect of deep learning lies in that the layers of features are not designed by human engineers but are rather learned from data using a general-purpose learning procedure. Hence, the important features are extracted automatically from the input data for the purpose of interpreting previously unseen samples [(6)].

Neural Networks:

Artificial neural networks consist of connected nodes, inspired by biological nervous systems. An artificial neural network is composed of interconnected artificial neurons. Deep learning systems encode features by using an architecture of ANN. Each artificial neuron implements a simple classifier model, which outputs a decision signal based on a weighted sum of evidences. The weights of the network are adjusted via a learning algorithm where pairs of input signals and desired output decisions are presented. In addition to the weighted sum of

evidences, an activation function integrates signals from previous neurons subsequently outputting a decision signal. Hundreds of these basic computing units make up an artificial neural network computing device [7]. DL algorithms most relevant to radiology are called convolutional neural networks (CNNs), they are used mainly because of their effectiveness in image segmentation and classification [8, 9]. In CNNs, lower level information inputs, similar to cutaneous sensory nerves, form synaptic connections to the next level or "layer" of neurons. Each neuron in this second layer can combine the inputs from lower level neurons to form a newer, more complex output. As the number of intermediate or hidden layers increases, so too does the accuracy of the output from the highest layer. CNNs capture the spatial features from an image. Spatial features are the relationship and arrangement of pixels in an image. They help us in identifying the object accurately, as well as its location and relation with other objects in an image [10].

Applications of AI, ML and DL in Dental Radiology

Today, there are several applications of AI in dental radiography. Studies have investigated the accuracy and efficiency of such tools in the various diagnostic dilemmas. Among the current applications is periodontal bone loss detection. In one study, A CNN trained on a limited set of radiographic image segments showed similar discrimination ability to dentists for assessing periodontal bone loss on panoramic radiographs [11]. In another study utilizing a CNN called DeNTNet for detecting periodontal bone loss on panoramic radiographs, DeNTNet achieved an F1 score (harmonic mean of the precision and recall) of 0.75 on the test set, while the average performance of annotating dentists was 0.69 [12]. Moreover, a CNN was developed and used to predict periodontally-compromised teeth and need for extractions based on intraoral periapical radiographs. The CAD tool showed promising results, with the diagnostic accuracy

being 81.0% for premolars and 76.7% for molars. The accuracy of predicting extractions was 82.8% for premolars and 73.4% for molars [13].

In the literature, several CAD tools have emerged in the last decade to assist dentists in caries detection. Logicon Caries Detector (LCD), an automated caries detection program, was found in a study conducted on extracted teeth to be less accurate than human observers in detecting proximal carious lesions [14]. A second study on extracted teeth compared observers' performance aided and unaided by LCD, using Micro CT as the reference standard. The study concluded that there was no statistically significant difference in observers' performance (between Az values) with and without the use of LCD when lesions of all depths were considered. Nonetheless, the study showed an improvement in the detection of proximal lesions that extended into the inner half of the enamel or into the dentine [15]. On the other hand, in a clinical study the authors concluded that LCD enabled dentists found 20% more cases of caries penetrating into dentin than they were able to find without it, without impacting the specificity, i.e. treating healthy teeth [16]. In a study conducted by Araki et al., results were consistent with the previous study's conclusions, as the observers' sensitivity doubled when detecting early caries that needed restoration, while specificity remained relatively constant [17]. Notably, this was the first diagnostic tool for caries detection purposes to be demonstrated efficacious in a clinical study and to be cleared by the FDA[16]. Other commercially unavailable tools were developed mainly for research purposes. The list includes a pre-trained GoogLeNet Inception v3 CNN network which was used for caries detection on periapical radiographs with experts' opinions serving as the ground truth reference. The overall accuracy for caries detection in the premolar and molar sites was found to be nearly 82%, a considerably good result given the associated advantages of fast and accurate diagnosis [18]. Furthermore, Valizadeh S et al.

designed a software application that was able to diagnose 60% of enamel caries and 97% of dentinal caries, when histological sectioning was used as a reference standard. The limitation in detecting enamel caries is somewhat predictable and in line with the literature since lesions are usually not radiographically visible until 30 - 40% of the enamel affected by the lesion has demineralized [19].

Several CAD tools aimed at tooth detection and labeling have emerged lately. The inspiration behind the development of those tools relates mainly to the prospects of saving time and improving workflow by automatic filling of dental charts and in cases of large-scale disasters. Miki et al, investigated the accuracy of a tooth detection and labeling CNN method on cone-beam CT volumes. Results of the tooth detection and labeling accuracy were 77.4% and 77.1%, respectively. However, the previous study used only 10 volumes for testing purposes, a very limited set [20]. In a study by Tuzoff et al. a CNN model was used for teeth identification, and numbering. This CAD technique had a mean sensitivity of 0.987 and precision of 0.9945, a result that matched dentists' performance [21]. Zhang et al., described the use of fast Regions with CNN features (fast R-CNN) in teeth recognition by relying on the label tree along with cascade network structure. The label tree was used to give each tooth several labels to address the low number of data. Whereas the use of cascade network structure was utilized to do automatic identification for the 32 teeth positions, using several CNNs as its basic module. The author claims that his method can address many complex cases such as radiographic images with tooth loss, decayed and restored teeth. Their results showed a precision and recall of 95.8% and 96.1%, respectively [22]. In a study by Chen et al., faster regions with convolutional neural network features (faster R-CNN) was applied to detect and label teeth on 1,250 digitized dental periapical films. The model demonstrated a detection precision of 90% when compared with human

experts, and 71.5% precision in tooth numbering [23]. Those studies demonstrate the possibilities of DL technologies, making it a convenient and efficient automatic aid for dentists in filling out their patients' dental charts.

In the field of Orthodontics and Orthopedics, AI advances have been applied for deciding if extractions are necessary prior to the orthodontic treatment as well as determining if a surgical or non-surgical approach is needed using the lateral cephalometric radiographs, automatic identification of cephalometric landmarks and in determining growth and development by cervical vertebrae stages [24]. Using an ANN, Xie et al. reported 80 % accuracy in determining whether extraction or non-extraction treatment was best for malocclusion patients [25]. In accordance with those results, Jung et al. reported a high accuracy (93%) for deciding on tooth extraction using cephalometric radiographs [26]. These studies suggest that AI models can be used as aids in making decisions in clinical practice by predicting the need for extraction ahead of treatment initiation. Studies on identification of cephalometric landmarks showed promising results [27-30]. Kunz et al. concluded that there were no statistically significant differences between human experts' as the gold standard and the AI's predictions in a study using a specialized artificial intelligence (AI) algorithm [29]. Similarly, Huang et al. found that AI was as accurate in the identification of 80 cephalometric landmarks as were trained orthodontists, using the latest deep learning method based on the You-Only-Look-Once version 3 algorithm (YOLOv3) [30]. Realizing that treatment planning is a crucial step in orthodontics and orthognathic surgeries, Choi et al. developed and tested an AI model deciding on surgical versus non-surgical treatment and for need of extractions determination. The model showed a success rate of 91% and 96% for the detailed diagnosis of surgery type and the extraction along with the need for surgery/non-

surgery decisions [31]. Those result suggest that machine learning could be applied for the diagnosis and planning of orthodontic and orthognathic surgery cases.

In Endodontics, deep convolutional neural networks (CNNs) were applied to detect apical lesions on panoramic radiographs. In one study, a moderately deep CNN trained on a limited set of panoramic images, showed satisfactory ability to detect apical lesions on panoramic radiographs. With the consensus of six examiners serving as the reference standard, the AUC of the CNN was 85%. Sensitivity and specificity were 65% and 87%, respectively. Subgroup analysis for tooth type was also performed showing a significantly higher sensitivity in molars than in other tooth types, whereas specificity was lower [32]. Thus, the application of neural networks may assist dentists in reliably and accurately detecting apical lesions. A second study, evaluated based on clinically validated ground truth, investigated the detection of periapical lucencies on panoramic radiographs. The periapical lucencies had a differential diagnosis that included infections, granuloma, cysts and tumors. Results demonstrated that the deep learning algorithm achieved a better performance than 14 out of 24 OMF surgeons within the cohort, exhibiting an average precision of 0.60, and an F1 score of 0.58, a PPV of 0.67 and TPR of 0.5. While not exceptionally high, the results of this study showed that the algorithm has a potential in aiding oral surgeons in detecting periapical lucencies on panoramic radiographs [33].

A deep learning object detection technique was utilized in a study by Ariji et al. to automatically detect and classify radiolucent lesions in the mandible on panoramic radiographs. In this study, histologically-verified mandibular radiolucent lesions of 10 mm or greater were incorporated. The five types of lesions included ameloblastomas, odontogenic keratocysts, dentigerous cysts, radicular cysts and simple bone cysts. The detection sensitivity was 88% using two testing data sets, with 50 images in the first testing data set and 25 images in the second data set. The false-

positive rate per image was 0.00 for the testing 1 data set, and 0.04 for the testing 2 data set, indicating that the learning model incorrectly predicted the presence of lesions in areas without lesion in 1 of 25 images from the testing 2 data set [34]. Despite the limited validation data set size in group 2, this study demonstrated that radiolucent lesions of the mandible can be detected with high sensitivity using deep learning. In a study by Lee et al. three cystic lesions (odontogenic keratocyst, dentigerous cysts and periapical cysts) were evaluated using panoramic radiographs and cone beam computed tomographic images. The pre-trained model using CBCT images showed good diagnostic performance (AUC = 0.914, sensitivity = 96.1%, specificity = 77.1%). Those results were significantly greater than those achieved by other models using panoramic images (AUC = 0.847, sensitivity = 88.2%, specificity = 77.0%) (p=0.14) [35].

Setzer et al. demonstrated excellent results in another study that aimed to use a DL algorithm for the automated segmentation of CBCT images and the detection of periapical lesions. Lesion detection accuracy was 0.93, specificity was 0.88, with a positive predictive value of 0.87 and a negative predictive value of 0.93[36]. At an attempt to evaluate a CNN method for detecting apical pathosis on CBCT, Orhan et al. included 153 periapical lesions obtained from 109 patients. The AI system was able to detect 142 of a total of 153 periapical lesions, which represents a sensitivity of 92.8 %[37].

REFERENCES

1. Mupparapu, M., Wu, C. W., & Chen, Y. C. (2018). Artificial intelligence, machine learning, neural networks, and deep learning: Futuristic concepts for new dental diagnosis. Quintessence international (Berlin, Germany: 1985), 49(9), 687.

2. McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine, 27(4), 12-12.

3. Mayo, R. C., & Leung, J. (2018). Artificial intelligence and deep learning–Radiology's next frontier?. Clinical imaging, 49, 87-88.

4. Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. European radiology experimental, 2(1), 35.

5. Kulkarni, S., Seneviratne, N., Baig, M. S., & Khan, A. H. A. (2020). Artificial intelligence in medicine: where are we now?. Academic radiology, 27(1), 62-70.

6. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

7. Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdzal, M., Turcotte, S., Pal, C. J. & Tang, A. (2017). Deep learning: a primer for radiologists. Radiographics, 37(7), 2113-2131.

8. Moeskops, P., Viergever, M. A., Mendrik, A. M., De Vries, L. S., Benders, M. J., & Išgum, I. (2016). Automatic segmentation of MR brain images with a convolutional neural network. IEEE transactions on medical imaging, 35(5), 1252-1261.

9. McBee, M. P., Awan, O. A., Colucci, A. T., Ghobadi, C. W., Kadom, N., Kansagra, A. P., ... & Auffermann, W. F. (2018). Deep learning in radiology. Academic radiology, 25(11), 1472-1480.

10. Pai, A. (2020, October 19). CNN vs. RNN vs. ANN – Analyzing 3 Types of Neural Networks in Deep Learning. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/

11. Krois, J., Ekert, T., Meinhold, L., Golla, T., Kharbot, B., Wittemeier, A., & Schwendicke, F. (2019). Deep learning for the radiographic detection of periodontal bone loss. Scientific reports, 9(1), 1-6.

12. Kim, J., Lee, H. S., Song, I. S., & Jung, K. H. (2019). DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. Scientific reports, 9(1), 1-9.

13. Lee, J. H., Kim, D. H., Jeong, S. N., & Choi, S. H. (2018). Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. Journal of periodontal & implant science, 48(2), 114-123.

14. Wenzel, A., Hintze, H., Kold, L. M., & Kold, S. (2002). Accuracy of computer-automated caries detection in digital radiographs compared with human observers. European journal of oral sciences, 110(3), 199-203.

15. Araki, K., Matsuda, Y., Seki, K., & Okano, T. (2010). Effect of computer assistance on observer performance of approximal caries diagnosis using intraoral digital radiography. Clinical oral investigations, 14(3), 319-325.

16. Gakenheimer, D. C. (2002). The efficacy of a computerized caries detector in intraoral digital radiography. The Journal of the American Dental Association, 133(7), 883-890.

17. Tracy, K. D., Dykstra, B. A., Gakenheimer, D. C., Scheetz, J. P., Lacina, S., Scarfe, W. C., & Farman, A. G. (2011). Utility and effectiveness of computer-aided diagnosis of dental caries. Gen Dent, 59(2), 136-144.

18. Lee, J. H., Kim, D. H., Jeong, S. N., & Choi, S. H. (2018). Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. Journal of dentistry, 77, 106-111.

19. Valizadeh, S., Goodini, M., Ehsani, S., Mohseni, H., Azimi, F., & Bakhshandeh, H. (2015). Designing of a computer software for detection of approximal caries in posterior teeth. Iranian Journal of Radiology, 12(4).

20. Miki, Y., Muramatsu, C., Hayashi, T., Zhou, X., Hara, T., Katsumata, A., & Fujita, H. (2017, March). Tooth labeling in cone-beam CT using deep convolutional neural network for forensic identification. In Medical Imaging 2017: Computer-Aided Diagnosis (Vol. 10134, p. 101343E). International Society for Optics and Photonics.

21. Tuzoff, D. V., Tuzova, L. N., Bornstein, M. M., Krasnov, A. S., Kharchenko, M. A., Nikolenko, S. I., ... & Bednenko, G. B. (2019). Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofacial Radiology, 48(4), 20180051.

22. Zhang, K., Wu, J., Chen, H., & Lyu, P. (2018). An effective teeth recognition method using label tree with cascade network structure. Computerized Medical Imaging and Graphics, 68, 61-70.

23. Chen, H., Zhang, K., Lyu, P., Li, H., Zhang, L., Wu, J., & Lee, C. H. (2019). A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Scientific reports, 9(1), 1-11.

24. Khanagar, S. B., Al-Ehaideb, A., Maganur, P. C., Vishwanathaiah, S., Patil, S., Baeshen, H. A., & Bhandi, S. (2020). Developments, application, and performance of artificial intelligence in dentistry–A systematic review. Journal of dental sciences.

25. Xie, X., Wang, L., & Wang, A. (2010). Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. The Angle Orthodontist, 80(2), 262-266.

26. Jung, S. K., & Kim, T. W. (2016). New approach for the diagnosis of extractions with neural network machine learning. American Journal of Orthodontics and Dentofacial Orthopedics, 149(1), 127-133.

27. Park, J. H., Hwang, H. W., Moon, J. H., Yu, Y., Kim, H., Her, S. B., ... & Lee, S. J. (2019). Automated identification of cephalometric landmarks: Part 1—Comparisons between the latest deep-learning methods YOLOV3 and SSD. The Angle Orthodontist, 89(6), 903-909.

28. Lindner, C., Wang, C. W., Huang, C. T., Li, C. H., Chang, S. W., & Cootes, T. F. (2016). Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms. Scientific reports, 6, 33581.

29. Kunz, F., Stellzig-Eisenhauer, A., Zeman, F., & Boldt, J. (2020). Artificial intelligence in orthodontics. Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie, 81(1), 52-68.

30. Hwang, H. W., Park, J. H., Moon, J. H., Yu, Y., Kim, H., Her, S. B., ... & Lee, S. J. (2020). Automated identification of cephalometric landmarks: Part 2-Might it be better than human?. The Angle Orthodontist, 90(1), 69-76.

31. Choi, H. I., Jung, S. K., Baek, S. H., Lim, W. H., Ahn, S. J., Yang, I. H., & Kim, T. W. (2019). Artificial intelligent model with neural network machine learning for the diagnosis of orthognathic surgery. Journal of Craniofacial Surgery, 30(7), 1986-1989.

32. Ekert, T., Krois, J., Meinhold, L., Elhennawy, K., Emara, R., Golla, T., & Schwendicke, F. (2019). Deep learning for the radiographic detection of apical lesions. Journal of endodontics, 45(7), 917-922.

33. Endres, M. G., Hillen, F., Salloumis, M., Sedaghat, A. R., Niehues, S. M., Quatela, O., & Gaudin, R. A. (2020). Development of a deep learning algorithm for periapical disease detection in dental radiographs. *Diagnostics*, *10*(6), 430.

34. Ariji, Y., Yanashita, Y., Kutsuna, S., Muramatsu, C., Fukuda, M., Kise, Y., ... & Ariji, E. (2019). Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. Oral surgery, oral medicine, oral pathology and oral radiology, 128(4), 424-430.

35. Lee, J. H., Kim, D. H., & Jeong, S. N. (2020). Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral diseases*, *26*(1), 152-158.

36. Setzer, F. C., Shi, K. J., Zhang, Z., Yan, H., Yoon, H., Mupparapu, M., & Li, J. (2020). Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. *Journal of endodontics*, *46*(7), 987-993.

37. Orhan, K., Bayrakdar, I. S., Ezhov, M., Kravtsov, A., & Özyürek, T. A. H. A. (2020). Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *International endodontic journal*, *53*(5), 680-689.

MANUSCRIPT

Introduction

The term “artificial intelligence” was coined in 1956, at a workshop that took place at Dartmouth College [1]. The Dartmouth Research Project team proposed a simplified description of a complex problem. AI was explained as “making a machine behave in ways that would be called intelligent if a human were so behaving”. Scientists hypothesized that a machine can be trained to learn through experimental “trial and error” much like humans do [2].Since then, AI has been progressing very rapidly in the medical field where it serves as an aid to physicians in diagnostic and treatment decisions. AI shows the most promise in providing non-specialists with easily accessible, expert-level predictions [3].

In dentistry, multiple computer-aided softwares (CAD) and deep learning (DL) tools have emerged recently for the assessment of dental caries [4, 5, 6, 7]. Additionally, neural networks and DL algorithms have been utilized in applications that include predicting dental pain, [8] teeth numbering and classification [9],in deciding if extractions are necessary prior to orthodontic treatment [10] and for detecting periodontal bone loss on panoramic radiographs. [11,12] In the field of Endodontics, deep convolutional neural networks (CNNs) were applied to detect apical lesions on panoramic radiographs [13, 14] and CBCT [15, 16, 17], root morphology assessment of the mandibular first molar [18] and vertical root fracture evaluation on panoramic radiographs [19].

The need for AI and automation is largely linked to an anticipated need to increase accuracy and speed in order to improve workflow and efficient use of resources. While computer-assisted detection of apical radiolucencies might benefit both experienced radiologists

and general practitioners, we anticipate that this would be more relevant for the latter because of their greater number and need for efficient screening tools. In the current study, we aim to investigate the effectiveness of the (Denti.AI) CAD system for detecting apical radiolucencies on intraoral periapical radiographs.

The apical radiolucency detection module is based on DL techniques, specifically deep convolutional neural networks (CNN). Deep learning is a class of artificial intelligence (AI) algorithms that allows a computer program to learn from input data for further interpretation of previously unseen samples. CNN architectures are commonly used for image recognition tasks. The CNNs exploit specific characteristics of an image data input to effectively represent and learn hierarchical image features using multiple levels of abstraction. The key aspect of deep learning is that these features are not designed by humans, but automatically extracted and learned from the raw data (such as pixels of images) [21,22].

Aims and testable hypotheses

The aims of the study are to:

1. Assess the effectiveness of the Denti.AI CAD system in assisting dentists with detecting apical radiolucencies on intraoral periapical radiographs.

2. Assess the effectiveness of DL in subgroups by location, extent, and treatment status of the tooth.

3. Assess the effectiveness of DL by reader specialty training.

The null hypotheses are:

1. There is no mean change in the performance metrics of readers under the two reading scenarios (aided and unaided by the Denti.AI CAD system).

2. There is no mean change in the sensitivity performance metric under the two reading scenarios (aided and unaided by the Denti.AI CAD system) in different lesion characteristics (location, extent, treatment status of the tooth).

3. There is no mean change in the performance metrics of readers under the two reading scenarios (aided and unaided by the Denti.AI CAD system) between radiologists and non-radiologists.

Materials and Methods

Case Selection and Ground Truth

Ethical approval was granted by the University Biomedical Institutional Review Board (#19-1430). The dental school's OMR-CBCT referral database was searched for all CBCT volumes acquired for endodontic purposes between August-2014 and March-2019. The case selection was retrospective and sequential in nature. The endodontic CBCT referrals (n=367) were assessed for the presence of apical radiolucencies. The finalized radiology reports were analyzed for findings related to apical radiolucencies including: "apical rarefying osteitis" or "apical radiolucent lesions" or "apical widening of the PDL space". All radiology reports were written by a board-certified oral and maxillofacial radiologist with at least 10 years of experience. If a report was found to contain any of the above findings, the accompanying CBCT volumes were downloaded and re-examined by the investigator (MH) to confirm the presence of a lesion at the specified location mentioned in the report. Measurements of the lesions were then recorded in all three dimensions; mesiodistal, buccolingual or buccopalatal and apico-coronal. Apical radiolucencies measuring less than 2 mm at their widest dimension were excluded to reduce the imperfect reference standard bias [(22)] and beam hardening artifacts resulting from

endodontic treatment fill-material. In addition, the exclusion criteria encompassed radiographs of patients under 18 years old.

Once a case was considered eligible to be included in the sample, the patients' records were searched for a corresponding, same-site IO periapical radiograph. The inclusion criteria for IO periapical radiographs comprised diagnostically acceptable IO radiographs acquired within a six-month window of the accompanying CBCT. If available, that periapical radiograph was included in this study. For control cases, periapical radiographs of teeth with CBCT-proven sound periodontium were uploaded in the negative subgroup. The intraoral images were all verified to have no evidence of apical radiolucencies on CBCT volumes and associated reports.

A total of 184 positive IO radiographs were collected and divided into a training subset (n=54) and a testing subset (n=130). A final set of 132 IO radiographs with sound apical periodontium was collected to serve as controls.

The positive training, positive testing and control sets were de-identified and uploaded to Denti.AI (Toronto, CA). The IO radiographs were annotated using the [Denti.AI](http://denti.AI) labeling tool. The annotation process included drawing a box around the borders of the apical radiolucencies on IO radiographs by utilizing the patient's reference CBCT. The AI model was pre-trained ahead of this study. However, the training subset (n=54) was utilized to further adapt the model to images acquired at the dental school's clinics. The margins of the box or rectangle had to conform to the size of the radiolucency and not exceed 2 mm from any side.

Apical lesions were then divided based on their anatomic location and size into the following groups:

1) By location:

- Incisors-canines (anterior teeth)

- Maxillary

- Mandibular

- Premolars (posterior teeth)

- Maxillary

- Mandibular

- Molars (posterior teeth)

- Maxillary

- Mandibular

2) By extent:

- Small radiolucencies (2-5 mm)

- Large radiolucencies (≥ 5mm)

3) By treatment status:

- Endodontically-treated teeth

- Untreated teeth

Correspondent tags were added for each annotated lesion showing the location and extent of a lesion. The detailed annotation guide was provided to MH by Denti.AI with the description of the task and labeling interface features. For the purpose of this clinical study, 68 images were randomly selected with an equal distribution from the positive testing and control sets to serve as a testing subset.

Devices and Imaging Instruments

The 3D volumes were all acquired in the main Radiology clinic using either the Orthophos XG 3D, the Orthophos SL 3D (Dentsply Sirona, Charlotte, NC), the CS 9000 or the

CS 9300 (Carestream Dental, Atlanta, GA). Of the IO radiographs, 42 were obtained using photostimulable phosphor (PSP) plates scanned using the ScanX (Air Techniques, Hicksville, NY), 19 with PSP plates scanned using the Soredex Digora Optime (Kavo Dental, Charlotte, NC), three with the Sirona Schick33 Direct Digital Sensor (Dentsply Sirona, Charlotte, NC), one with the XDR Anatomic Sensor (Cyber Medical Imaging, Los Angeles, CA), and three where the sensor was unknown. The inclusion of multiple units and sensor types allowed for additional generalizability of the results.

Cases Distribution

Each periapical radiograph showed multiple teeth (mean = 3.9), and each positive case displayed 1-3 apical radiolucencies (mean = 1.5). All positive cases were obtained from unique patients. As for the negative cases, 1-3 images were taken for a patient (Table 2.1). Table 2.2 and Figure 2.1 show the distribution of cases by age groups. Table 2.3 shows the distribution of cases by gender.

The negative radiographs that were obtained from the same patients were taken from different sextants. Each periapical radiograph of the resulting testing collection was read independently and readers were not aware whether the case was taken from the same patient or not. All cases included in this study (both negative and positive) were selected randomly from a collection of approximately 260 images, so there was no exact order that might bias the reading results.

AR Distribution Figures

Figure 2.2 shows the distribution of AR by location, Figure 2.3 shows the distribution of AR by extent, and Figure 2.4 displays the distribution of cases by treatment status. Table 2.4 provides the figures for each lesion cohort.

Readers

Twelve readers took part in this clinical study, each reading the IO testing subset independently and under two scenarios. The qualifications of the readers comprised one or a combination of the following: general dentists, oral and maxillofacial radiologist, endodontist.

The readers included residents of operative dentistry and biomaterials program (n=6) and a general dentist (n=1) with experience ranging from 3 to 30 years. In addition, two oral and maxillofacial radiology residents, two board certified radiologists and one board-certified endodontist, having experience ranging from 16 to 40 years.

Reader Study Execution

The 12 readers performed a cross-over reading scenario. Each reader analyzed the same testing subset collection of 68 images (full-crossed design) under two conditions; without and with the aid of AI predictions. The reading sessions were separated by a washout period of more than one month (~120 - 340 days, with an average of ~200 days).

Prior to each of the two reading sessions, readers underwent a training conducted by the investigator (MH). The training included sample images and practice annotations. The sample images were not included in the main case collection. Additionally, a written interface manual containing the annotation guide was provided along with a detailed task description. For the conventional reading session, without the assistance of AI predictions, the readers were instructed to draw bounding boxes using the labeling tool around sites with suspected AR (annotate apical radiolucencies). Readers were then asked to ensure that each bounding box annotation covered the finding with the margin not exceeding 2 mm beyond any side of the AR. In addition, they were asked to add confidence score tags that reflected their confidence regarding the presence of their decision; a (1-5) Likert scale was used for this step. The tags were

as follows: “C1: not confident”, “C2: slightly confident”, “C3: somewhat confident”, “C4: moderately confident”, “C5: very confident”. It was emphasized that the tags only reflected confidence scores and did not have to match the severity or extent of the AR.

For the second reading session the task was to review the same IO radiographs (n=68) and either confirm or modify the bounding box, add a new finding or delete the prediction generated by the CAD system. With each predicted annotation, the AI system provided a confidence score. Nonetheless, the readers were instructed to provide their own confidence tags for each of the final annotations using the abovementioned (1-5) confidence scale.

Statistical Analysis

RJafroc R-library (version 1.2.0) was used to evaluate the performance. Dorfman-Berbaum-Metz with Hillis' improvements (DBMH) method of analysis was applied: a method for multi-reader multi-case (MRMC) analysis that uses the jackknife technique and conventional analysis of variance (ANOVA). The "Random-Reader Random Case" option of analysis was evaluated. With this analysis option, both cases and readers are considered as random factors. Thus, the results of the analysis can be applied to the general population of readers and cases.

The following common definitions were applied:

1) Lesion: an apical radiolucency that is shown on the image in the form of a bounding box; each lesion annotated by the reader is supported by a confidence score.

2) Case: an image that is interpreted by the reader. The inferred ROC rating paradigm was applied to define a confidence rating for the case annotated by the reader: the highest rating was used in the case of multiple lesions shown. Alternative Free-Response Receiver Operating Characteristic (AFROC) AUC metric was evaluated as the primary endpoint for comparing the performance of the readers for the two reading scenarios. AFROC AUC measures the area under

the AFROC curve. This curve shows a tradeoff between lesion localization fraction (LLF) and the false positive fraction for the range of decision thresholds. The LLF shows how many ground truth lesions were correctly detected by the reader, i.e. sensitivity on the by-lesion basis.

FPF shows how many actual negative cases were mistakenly classified as positive, i.e. 1 - specificity on a by-case basis. AFROC was chosen as the main metric of performance as it both provides aggregated measurement over different thresholds and accounts for localization accuracy.

Secondary endpoint analysis included the following metrics: ROC AUC, sensitivity (by case), specificity (by case), sensitivity by lesion.

Subgroup performance analysis was conducted to measure the effect of the CAD system for different characteristics of lesions. The analysis was stratified based on AR location, extent, and treatment status of the tooth (whether an affected tooth was endodontically treated or not).

It is worth noting that the sensitivity metric was used for “by lesion” analysis, whereas the specificity metric was not calculated. This is mainly because this metric depends on true negatives and there is no meaningful way to calculate true negative lesions, as there is an undefined number of locations where the lesions might be shown on an image. Additionally, the AFROC AUC, ROC AUC, sensitivity by case, specificity by case, and sensitivity by lesion metrics were calculated for the different groups of readers stratified by readers' specialties.

Results

Primary Endpoint Analysis

A statistically significant difference in AFROC ($p = 0.023$), with the AI-aided session showing a 5.8% improvement over the unaided reading session was found. Table 2.5 shows the

resulting metrics. Furthermore, Figure 2.5 shows the AFROC curves: the average performance for all 12 readers under both aided and unaided modalities.

Secondary Endpoint Analysis

Secondary endpoints results showed that there was no statistically significant difference in the performance between the two sessions in relation to the following metrics: ROC AUC, sensitivity and specificity on a "by case" basis or for sensitivity on a "by lesion" detection basis. The following p-values were obtained for the above metrics:

- ROC AUC (p-value 0.440)
- Sensitivity by case (p-value 0.780)
- Specificity by case (p-value 0.180)
- Sensitivity by lesion (p-value 0.065)

Table 2.6 and Figure 2.6 provide the resulting figures and the plot of the ROC curve; respectively.

Subgroup Analysis Stratified by Lesion Characteristics

There was a statistically significant difference in the sensitivity by lesion between the two sessions for small ARs ($p = 0.022$), as well as for the detection of ARs in the mandibular molar region ($p = 0.046$) and for ARs associated with endodontically treated teeth ($p = 0.009$). The sensitivity for small lesions increased by 10.5%. The mandibular molar region showed a sensitivity increase of 9.7% and endodontically treated status increased the sensitivity by 10.8%. Table 2.7 shows the results by lesion extent. Table 2.8 shows the results for different tooth treatment status and Table 2.9 displays the results for different lesion locations.

Subgroup Analysis Stratified by Readers' Specialty

Subgroup performance analysis was conducted to measure the effect of the CAD system stratified by readers' specialization. Two groups were evaluated: radiologists and non-radiologists, specifically general dentists, operative dentists, and endodontists. The same metrics were assessed that were included in the primary and secondary endpoints of the study: AFROC AUC, ROC AUC, sensitivity, specificity, and sensitivity on a by-lesion basis.

For the non-radiologists group, the AFROC, specificity, and sensitivity-by lesion metrics showed a statistically significant difference between the first and second reads with p-values of 0.005, 0.047 and 0.049, respectively (Table 2.10.) All three metrics showed an increase which ranged from 7.1 – 13.8 %.ROC AUC and sensitivity metrics were not statistically significant.

For the radiologists group, there was no statistically significant difference found between the two reads in any of the metrics (AFROC AUC, ROC AUC, sensitivity, specificity, and sensitivity on a by-lesion basis), see Table 2.11.

Discussion:

Dentists assume the responsibility of selecting the appropriate radiographic modality, acquiring the radiographs and interpreting the results, and making decisions based on the interpretation. If radiographs are inattentively read, this could potentially lead to over or under-treating patients. By serving as adjuncts to dentists, diagnostic software and tools have the potential to reduce fatigue and diagnostic errors. [23]

In endodontics, deep convolutional neural networks (CNNs) have been applied to tasks that include the detection of apical lesions on panoramic radiographs and CBCT. In one study, a moderately deep CNN trained on a limited set of panoramic images showed satisfactory ability to detect apical lesions. Based on the consensus of six examiners the AUC of the CNN was 85%.

Sensitivity and specificity were 65% and 87%, respectively. Subgroup analysis for tooth type was also performed showing a significantly higher sensitivity in molars than in other tooth types, whereas specificity was lower [13].Thus, the application of neural networks shows promise in assisting dentists in detecting apical lesions on panoramic radiographs with a reasonable accuracy. A second study, evaluated based on a clinically validated ground truth, investigated the detection of periapical radiolucencies on panoramic radiographs. The periapical radiolucencies included infections, granulomas, cysts and tumors. Results demonstrated that the deep learning algorithm achieved a better performance than 14 out of 24 participating OMF surgeons within the cohort, exhibiting an average precision of 0.60, and an F1 score of 0.58, a PPV of 0.67 and TPR of 0.5. While not exceptionally high, the results of this study showed that the algorithm has the potential to aid oral surgeons in detecting periapical lucencies on panoramic radiographs [14]. Setzer et al. demonstrated excellent results in another study that aimed at using a DL algorithm for the automated segmentation of CBCT images and the detection of periapical lesions with lesion detection accuracy of 0.93 [16]. At another attempt to evaluate a CNN method at detecting apical pathosis on CBCT, Orhan et al. included 153 periapical lesions obtained from 109 patients. The AI system was able to detect 142 of a total of 153 periapical lesions with a reliability of correctly detecting a periapical lesion of 92.8 % [17].

These studies were designed to assess the diagnostic efficacy of DL systems in detecting periapical lesions as a standalone system and compared to clinicians. The authors of this study posit that assessing the ability of the DL software to assist the clinician is of equal value to the readership. This was one of the chief aims of this investigation.

For the average performance for all 12 readers, the results of this study showed a statistically significant difference in AFROC (p-value 0.023), with the AI-aided session showing

a 5.8% improvement over the unaided reading session. Those results reflect that the detection of AR lesions by clinicians can be improved using this tool. However, the results of the secondary endpoints results showed that there was no statistically significant difference in the performance between the two sessions in relation to ROC AUC, sensitivity and specificity on a by-case basis as well as sensitivity on a by-lesion detection basis. A possible explanation lies in the versatility and inhomogeneity in observers' experiences and specialties. This is a plausible explanation since the results of the non-radiologists group showed a statistically significant difference between the first and second reads in the following metrics: AFROC, specificity, and sensitivity-by lesion metrics. All three metrics showed an increase which ranged from 7.1 – 13.8%. Whereas in the radiologists group, there was no statistically significant difference found between the two reads in any of the metrics. This observation suggests that non-radiologists may benefit the most from CAD tools.

Additionally, the results of the subgroup analysis by lesion characteristics showed a significant difference in the sensitivity by-lesion metric between the two sessions in the following subgroups: small extent lesions, mandibular molar sites and in AR lesions associated with endodontically-treated teeth. The sensitivity increased for small lesions by 10.5%. The mandibular molar region showed an increase of 9.7% and endodontically-treated teeth showed an increase in sensitivity of 10.8%.

Hence, all three null-hypotheses are rejected. The results of the subgroup analysis by-lesion characteristics follows the distribution characteristics within each subgroup with a larger sample of smaller lesions, mandibular molar lesions and endodontically-treated status. However, the sample size is too small to determine whether this finding is coincidental or not.

To the best of the authors' knowledge, no study investigated the accuracy of detecting apical radiolucencies on periapical radiographs using CBCT as a gold standard. The sample was selected retrospectively with due diligence given to preventing selection bias by performing consecutive collection and randomization of the pilot sample. However, the sample was limited in number. Limitations also included the low number of observers in each subgroup and the variability of experiences in the observers as a whole body. Furthermore, another limitation in our study is that we did not attempt to correlate radiographic findings with clinical findings or patient symptoms relating to the apical radiolucencies detected. Hence, it is not known to what extent periapical lesions detected in our CBCT volumes reflect the histological status of the periapical tissues.

Various observer studies have been conducted to evaluate the accuracy of periapical radiographs, panoramic radiographs, and CBCT imaging in the diagnosis of apical radiolucent lesions. While studies showed varying results, we decided to utilize CBCT volumes as a reference standard since it has been proven to have significantly higher diagnostic accuracy compared to 2D imaging modalities. In a study by Patel et. al, The ROC Az values were 0.79 and 1.00 for IO and CBCT, respectively [(24)] .It's also reported that (20%–39%) of AP radiolucencies are diagnosed with CBCT and missed with IO radiography [(25)]. In a study by low et al., CBCT showed significantly more lesions (34%, $p < 0.001$) than periapical radiographs [(26)]. The accuracy of detecting apical periodontitis with panoramic radiographs was evaluated by Nardi et al. who reported a low sensitivity of 34.2%, a diagnostic accuracy of 65% and a high specificity 95.8% [(27)]. A major advantage of using CBCT as the reference is the ability to detect apical radiolucencies in anatomically challenging areas such as the posterior maxilla, where anatomic overlap takes place in 2D images [(28)]. A study by Uraba et al. showed that the overall periapical

lesion detection rates of periapical radiographs and CBCT imaging were 31.5% and 52.2%, respectively. Notably, the ability of CBCT imaging to identify periapical lesions that were not detected by PR was statistically significant for the maxillary incisors/canines and maxillary molars [29]. CBCT can display the details of the lesions and adjacent structures and provide correct clinical diagnosis as it shows destruction of cortical bone that couldn't detect by periapical radiography [30]. Thus, CBCT could be considered too high of a standard and the impact of AI on a clinician's diagnostic performance may be confounded by the choice of the gold standard.

Furthermore, the prevalence of apical pathology undetected on periapical radiographs is considerably high as 30-50% of mineral loss is needed to visualize the lesions. Thus, the limitations of periapical radiographs as a diagnostic tool should not be disregarded, mainly to reduce false-negative results [31, 32].

Conclusions:

Using a limited testing dataset, Denti.Ai improved localization of apical radiolucencies. Further AI training is necessary to increase the sensitivity and specificity of apical radiolucencies detection. Statistically significant improvement in the performance of non-radiologists, detection of mandibular molar apical radiolucent lesions, small lesions, and in endodontically-treated teeth was observed in the current study.

REFERENCES

1. Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62(1), 15-25.

2. McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine, 27(4), 12-12.

3. Park, W. J., & Park, J. B. (2018). History and application of artificial neural networks in dentistry. European journal of dentistry, 12(4), 594.

4. Valizadeh, S., Goodini, M., Ehsani, S., Mohseni, H., Azimi, F., & Bakhshandeh, H. (2015). Designing of a computer software for detection of approximal caries in posterior teeth. Iranian Journal of Radiology, 12(4).

5. Behere, R. R., & Lele, S. M. (2011). Reliability of Logicon caries detector in the detection and depth assessment of dental caries: an in-vitro study. Indian Journal of Dental Research : Official Publication of Indian Society for Dental Research, 22(2), 362. doi:10.4103/0970-9290.84277

6. Wenzel, A., Hintze, H., Kold, L. M., & Kold, S. (2002). Accuracy of computer-automated caries detection in digital radiographs compared with human observers. European journal of oral sciences, 110(3), 199-203.

7. Araki, K., Matsuda, Y., Seki, K., & Okano, T. (2010). Effect of computer assistance on observer performance of approximal caries diagnosis using intraoral digital radiography. Clinical oral investigations, 14(3), 319-325.

8. Kim, E. Y., Lim, K. O., & Rhee, H. S. (2009). Predictive modeling of dental pain using neural network. Studies in health technology and informatics, 146, 745-746.

9. Tuzoff, D. V., Tuzova, L. N., Bornstein, M. M., Krasnov, A. S., Kharchenko, M. A., Nikolenko, S. I., ... & Bednenko, G. B. (2019). Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofacial Radiology, 48(4), 20180051.

10. Xie, X., Wang, L., & Wang, A. (2010). Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. The Angle orthodontist, 80(2), 262-266.

11. Krois, J., Ekert, T., Meinhold, L., Golla, T., Kharbot, B., Wittemeier, A., & Schwendicke, F. (2019). Deep learning for the radiographic detection of periodontal bone loss. Scientific reports, 9(1), 1-6.

12. Kim, J., Lee, H. S., Song, I. S., & Jung, K. H. (2019). DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. Scientific reports, 9(1), 1-9.

13. Ekert, T., Krois, J., Meinhold, L., Elhennawy, K., Emara, R., Golla, T., & Schwendicke, F. (2019). Deep learning for the radiographic detection of apical lesions. Journal of endodontics, 45(7), 917-922.

14. Endres, M. G., Hillen, F., Salloumis, M., Sedaghat, A. R., Niehues, S. M., Quatela, O., & Gaudin, R. A. (2020). Development of a deep learning algorithm for periapical disease detection in dental radiographs. Diagnostics, 10(6), 430.

15. Lee, J. H., Kim, D. H., & Jeong, S. N. (2020). Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. Oral diseases, 26(1), 152-158.

16. Setzer, F. C., Shi, K. J., Zhang, Z., Yan, H., Yoon, H., Mupparapu, M., & Li, J. (2020). Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. Journal of endodontics, 46(7), 987-993.

17. Orhan, K., Bayrakdar, I. S., Ezhov, M., Kravtsov, A., & Özyürek, T. A. H. A. (2020). Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. International endodontic journal, 53(5), 680-689.

18. Hiraiwa, T., Ariji, Y., Fukuda, M., Kise, Y., Nakata, K., Katsumata, A., ... & Ariji, E. (2019). A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. Dentomaxillofacial Radiology, 48(3), 20180218.

19. Fukuda, M., Inamoto, K., Shibata, N., Ariji, Y., Yanashita, Y., Kutsuna, S., ... & Ariji, E. (2019). Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. Oral radiology, 1-7.

20. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

21. Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdzal, M., Turcotte, S., Pal, C. J. & Tang, A. (2017). Deep learning: a primer for radiologists. Radiographics, 37(7), 2113-2131.

22. Trikalinos, T. A., & Balion, C. M. (2012). options for summarizing medical test performance in the absence of a “gold standard”. Journal of general internal medicine, 27(1), 67-75.

23. Kahn Jr, C. E. (2017). From images to actions: opportunities for artificial intelligence in radiology.

24. Patel, S., Dawood, A., Mannocci, F., Wilson, R., & Pitt Ford, T. (2009). Detection of periapical bone defects in human jaws using cone beam computed tomography and intraoral radiography. International endodontic journal, 42(6), 507–515.

25. Abella, F., Patel, S., Duran-Sindreu, F., Mercadé, M., Bueno, R., & Roig, M. (2012). Evaluating the periapical status of teeth with irreversible pulpitis by using cone-beam computed tomography scanning and periapical radiographs. Journal of endodontics, 38(12), 1588-1591.

26. Low, K. M., Dula, K., Bürgin, W., & von Arx, T. (2008). Comparison of periapical radiography and limited cone-beam tomography in posterior maxillary teeth referred for apical surgery. Journal of endodontics, 34(5), 557-562.

27. Nardi, C., Calistri, L., Pradella, S., Desideri, I., Lorini, C., & Colagrande, S. (2017). Accuracy of orthopantomography for apical periodontitis without endodontic treatment. Journal of endodontics, 43(10), 1640-1646.

28. Saidi, A., Naaman, A., & Zogheib, C. (2015). Accuracy of cone-beam computed tomography and periapical radiography in endodontically treated teeth evaluation: A five-year retrospective study. Journal of international oral health: JIOH, 7(3), 15.

29. Uraba, S., Ebihara, A., Komatsu, K., Ohbayashi, N., & Okiji, T. (2016). Ability of Cone-beam Computed Tomography to Detect Periapical Lesions That Were Not Detected by Periapical Radiography: A Retrospective Assessment According to Tooth Group. Journal of Endodontics, 42(8), 1186–1190. doi:10.1016/j.joen.2016.04.026

30. Lan, M. A., Fu-liang, Z. H. A. N., Li-hong, Q. I. U., & Ming, X. U. E. (2012). The application of cone-beam computed tomography in diagnosing the lesions of apical periodontitis of posterior teeth. Shanghai Journal of Stomatology, 21(4).

31. Huumonen, S., & Ørstavik, D. (2002). Radiological aspects of apical periodontitis. Endodontic Topics, 1(1), 3-25.

32. Estrela, C., Bueno, M. R., Leles, C. R., Azevedo, B., & Azevedo, J. R. (2008). Accuracy of cone beam computed tomography and panoramic and periapical radiography for detection of apical periodontitis. Journal of endodontics, 34(3), 273-279.

APPENDIX I – FIGURES

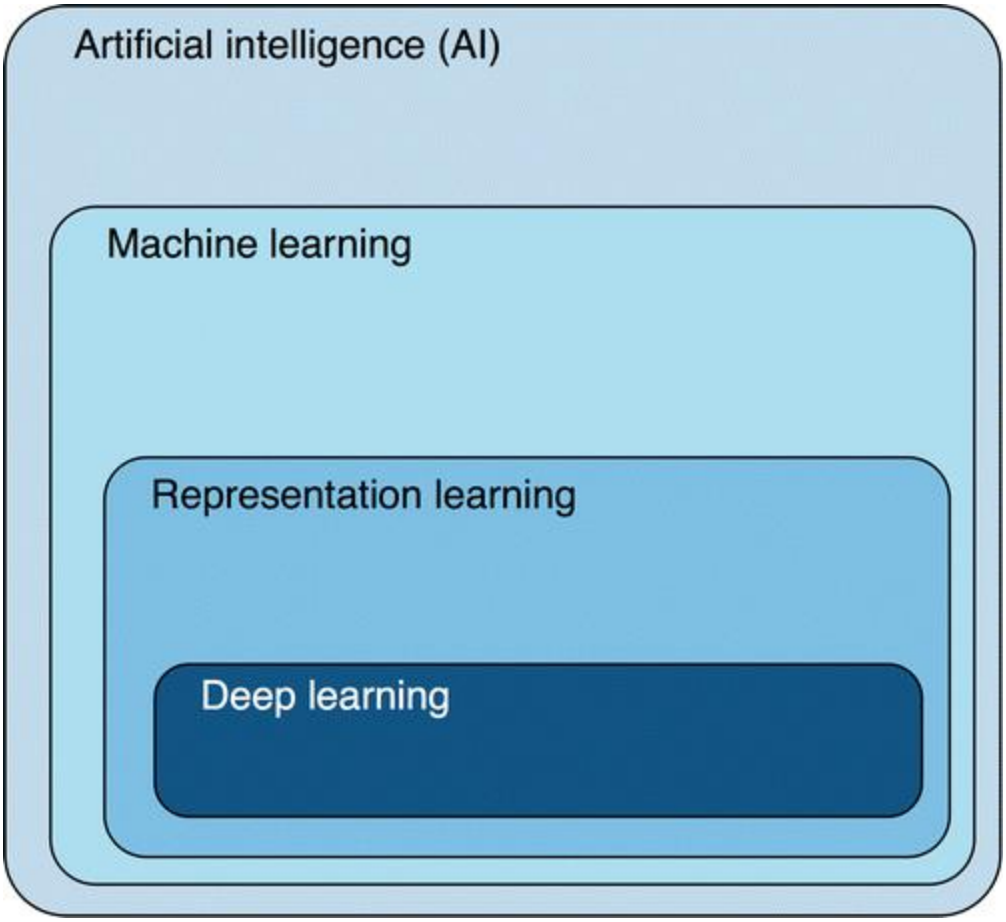

Figure 1.1. Venn diagram. AI is a subfield of computer science devoted to creating systems to perform tasks that ordinarily require human intelligence. ML is a subfield of AI where humans engineer features by which a computer can learn to differentiate patterns of data. Representation learning is a type of ML where no feature engineering is used; instead, the computer learns the features by which to classify the provided data. DL is a type of representation learning where the learned features are hierarchical [7].

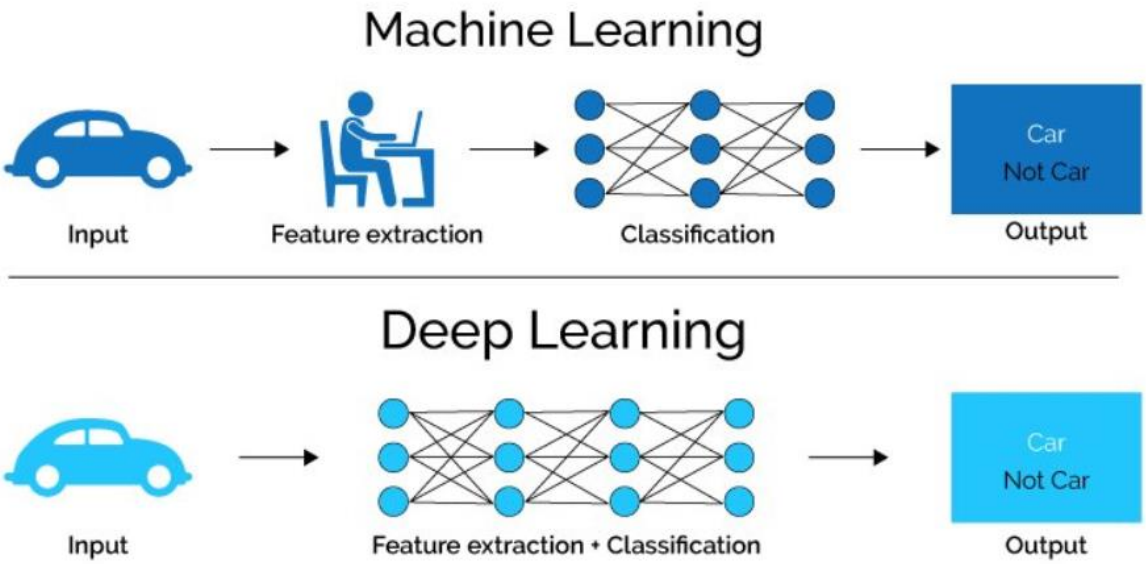

Figure 1.2. Comparison between Machine Learning and Deep Learning.

(https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/)

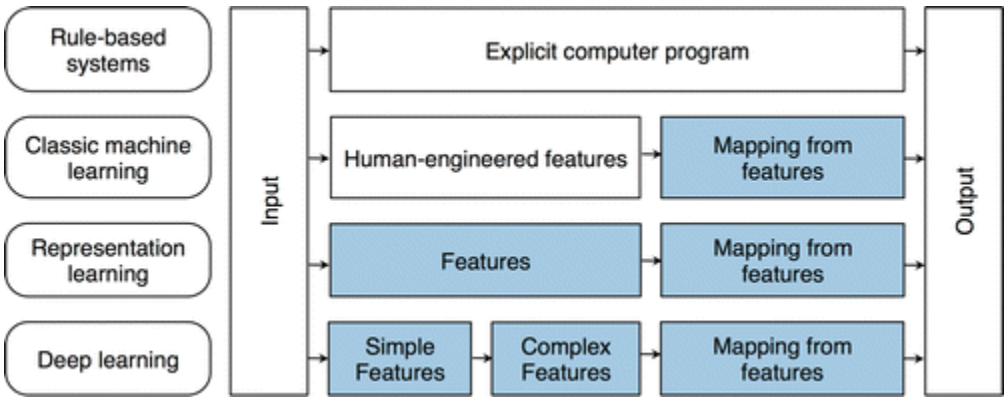

Figure 1.3. Types of Learning. Classic machine learning depends on carefully designed features, requiring human expertise and complicated task-specific optimization. Deep learning bypasses

feature engineering by taking advantage of large quantities of data and flexible hierarchical models. Blue boxes represent components learned by fitting a model to example data; deep learning allows learning an end-to-end mapping from the input to the output [7].

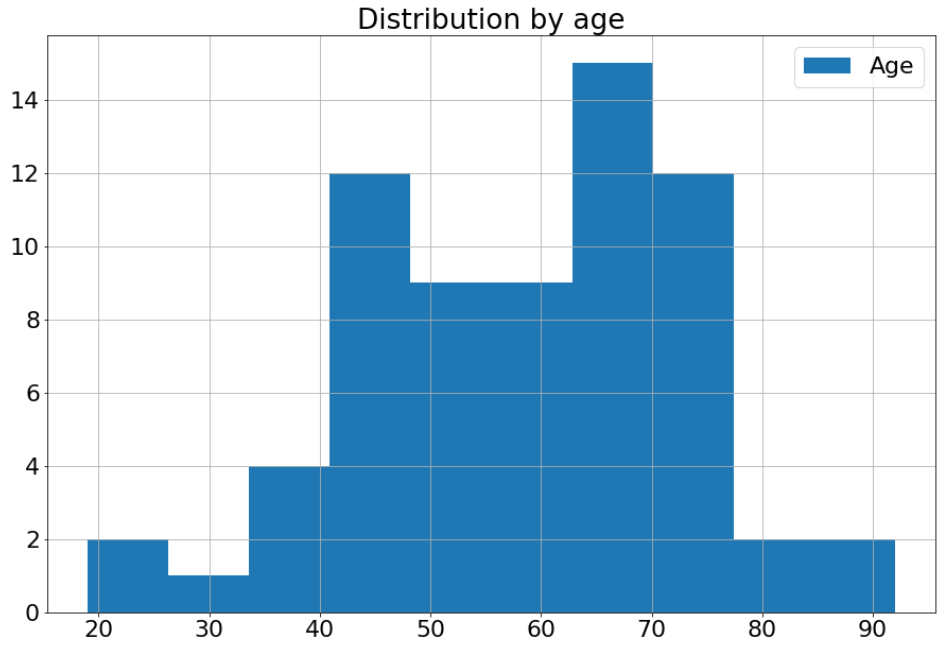

Figure 2.1. Distribution of Cases by Age.

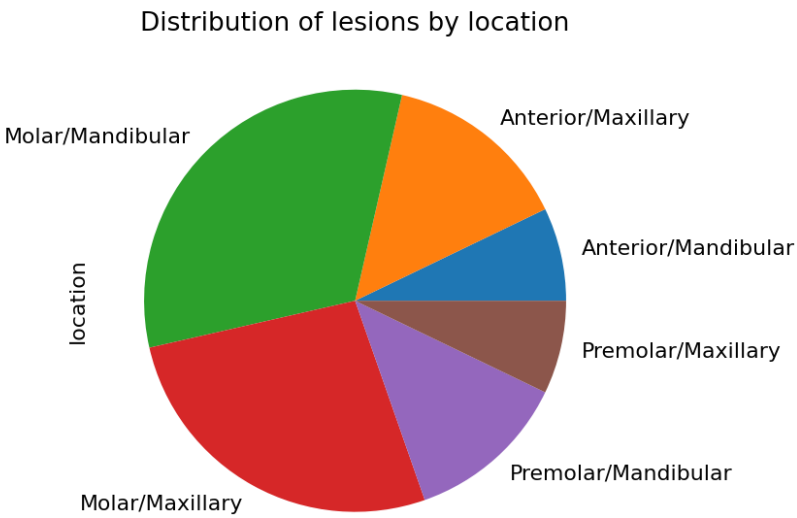

Figure 2.2. Distribution of Lesions by Location.

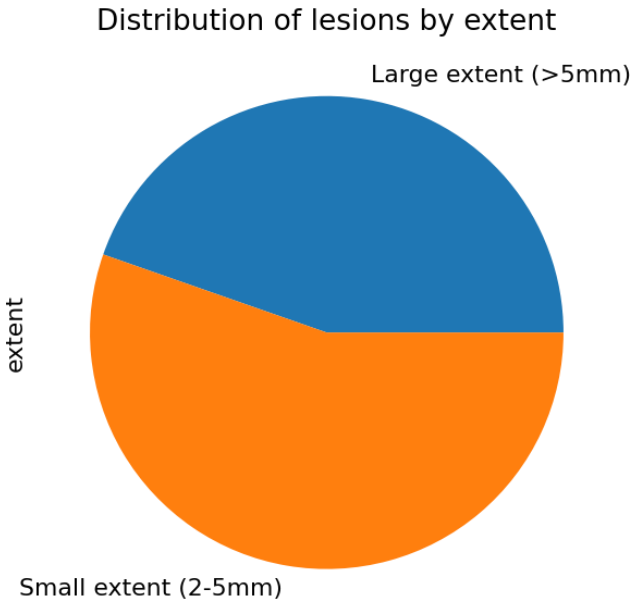

Figure 2.3. Distribution of Lesions by Extent.

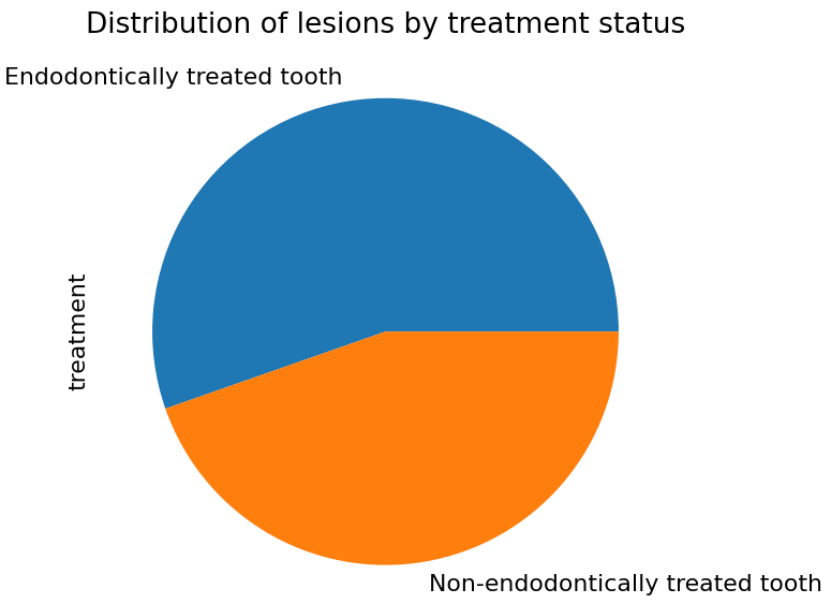

Figure 2.4. Distribution of Lesions by Treatment Status.

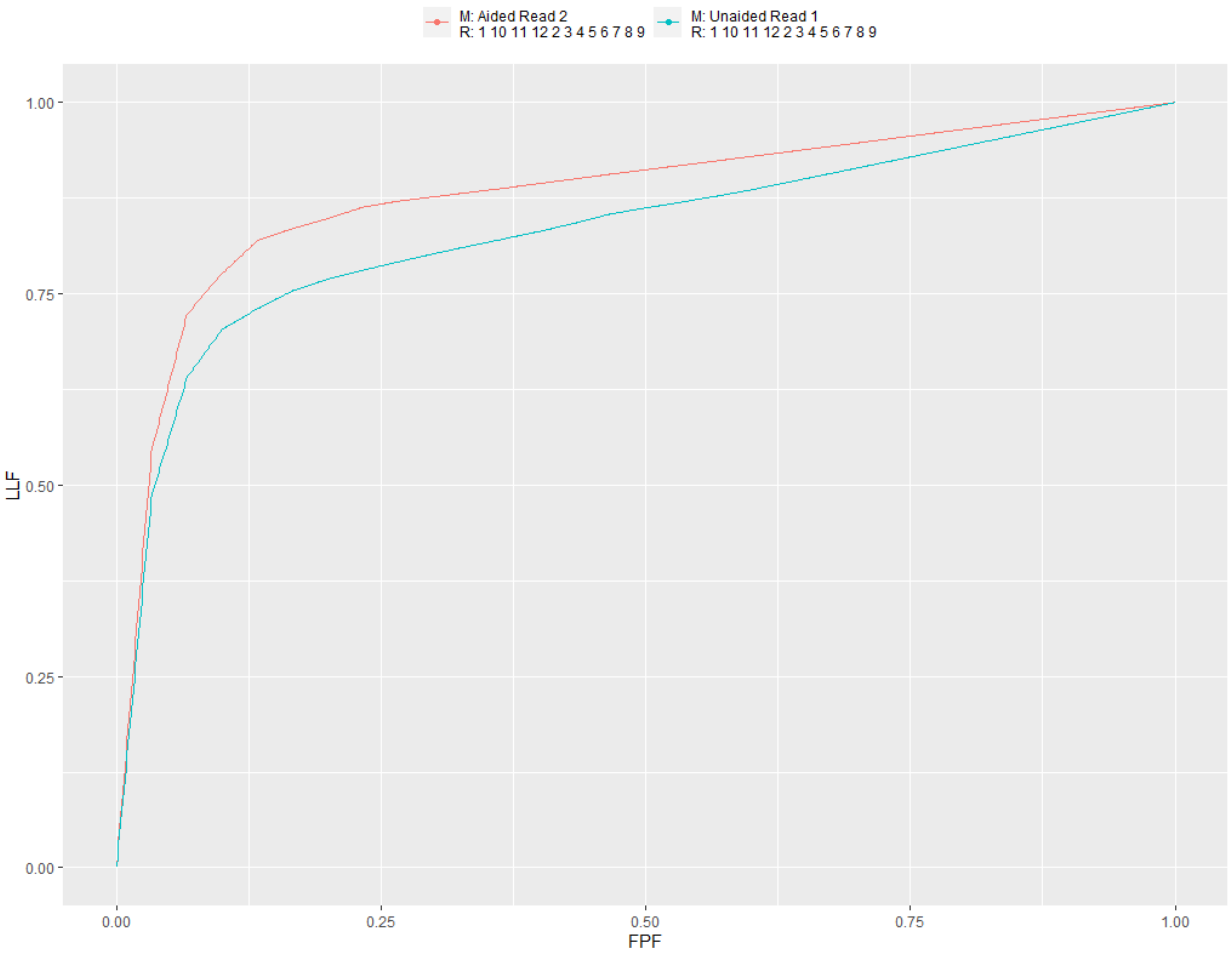

Figure 2.5. AFROC Plot: the average operating characteristics over 12 readers and 2 modalities (aided and unaided).

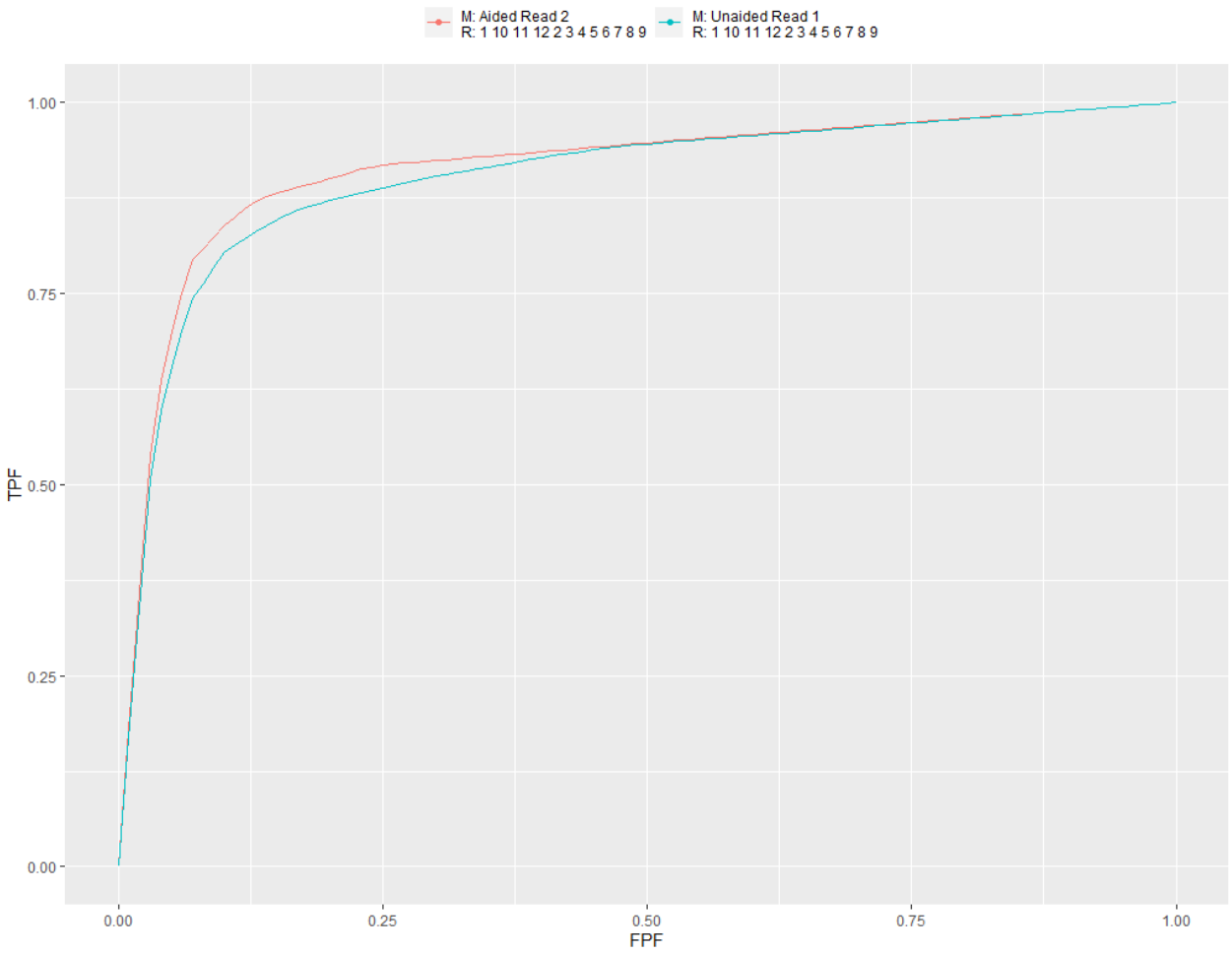

Figure 2.6. ROC Plot: the average operating characteristics over 12 readers and 2 modalities (aided and unaided).

APPENDIX II – TABLES

	Number of periapical radiographs	Number of teeth	Number of lesions	Number of unique patients
Positive cases	38	152	56	38
Control cases	30	116	0	19
Total cases	68	268	56	57

Table 2.1. General Statistics.

Age group	Number of cases
18-30	3
30-40	4
40-50	18
50-60	9
60-70	18
70-80	13
80-90	2

Table 2.2. Distribution of cases by age.

Gender	Number of cases
Male	26
Female	42

Table 2.3. Distribution of cases by gender.

Location	Number of lesions
Anterior/ Mandibular	4
Premolar/ Mandibular	7
Molar/ Mandibular	18
Anterior/ Maxillary	8
Premolar/ Maxillary	4
Molar/ Maxillary	15
Extent	**Number of lesions**
Small (2-5 mm)	31
Large (>5 mm)	25
Treatment status	**Number of lesions**
Endodontically treated	31
Not endodontically treated	25

Table 2.4. Distribution of lesions by location, extent and treatment status.

By-Image Statistics (12 readers, 68 images), RRRC Scenario							
	AFROC	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.896	0.847	0.945	0.049	0.007	0.092	**0.023**
Read 1	0.847	0.784	0.910				

Table 2.5. Primary Endpoint Assessment: AFROC AUC metric.

By-Image Statistics (12 readers, 68 images), RRRC Scenario							
	ROC AUC	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.930	0.888	0.972	0.013	-0.020	0.046	0.440
Read 1	0.917	0.868	0.966				
	Sensitivity	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.930	0.876	0.983	-0.007	-0.053	0.040	0.780
Read 1	0.936	0.888	0.985				
	Specificity	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.781	0.671	0.890	0.075	-0.038	0.188	0.180
Read 1	0.706	0.566	0.845				
By-Lesion Statistics (12 readers, 56 images), RRRC Scenario							
	Sensitivity	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.876	0.809	0.944	0.055	-0.004	0.114	0.065
Read 1	0.821	0.745	0.898				

Table 2.6. Secondary Endpoints assessment results: ROC AUC, Sensitivity, Specificity on the by-image basis, and Sensitivity on the by-lesion basis.

Small Extent (31 lesions)							
	Sensitivity	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.847	0.755	0.939	0.105	0.015	0.194	**0.022**
Read 1	0.742	0.619	0.865				
Large Extent (25 lesions)							
	Sensitivity	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.913	0.835	0.992	-0.007	-0.052	0.039	0.768
Read 1	0.920	0.849	0.991				

Table 2.7. Subgroup analysis stratified by lesion extent.

Endodontically treated (31 lesions)							
	Sensitivity	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.927	0.871	0.983	0.108	0.027	0.188	**0.009**
Read 1	0.820	0.705	0.935				
Non-endodontically treated (25 lesions)							
	Sensitivity	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.813	0.677	0.950	-0.010	-0.086	0.066	0.795
Read 1	0.823	0.718	0.929				

Table 2.8. Subgroup analysis stratified by tooth treatment status.

Molar/Maxillary (15 lesions)							
Read 2 (Aided by AI)	**Sensitivity**	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
	0.694	0.538	0.850	0.039	-0.112	0.190	0.613
Read 1	0.656	0.479	0.832				
Molar/Mandibular (18 lesions)							
Read 2 (Aided by AI)	**Sensitivity**	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
	0.958	0.908	1.000	0.097	0.002	0.193	**0.046**
Read 1	0.861	0.746	0.976				
Premolar/Maxillary (4 lesions)							
Read 2 (Aided by AI)	**Sensitivity**	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
	0.917	0.753	1.000	0.104	-0.101	0.309	0.318
Read 1	0.813	0.445	1.000				
Premolar/Mandibular (7 lesions)							
Read 2 (Aided by AI)	**Sensitivity**	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
	1.000	N/A	N/A	0.012	-0.014	0.038	0.339
Read 1	0.988	0.962	1.000				
Anterior/Maxillary (8 lesions)							
Read 2 (Aided by AI)	**Sensitivity**	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
	0.979	0.937	1.000	0.083	-0.053	0.219	0.229
Read 1	0.896	0.721	1.000				
Anterior/Mandibular (4 lesions)							
Read 2 (Aided by AI)	**Sensitivity**	CI lower	CI upper	Read 2 – Read 1	CI lower	CI upper	P-value
	0.729	0.499	0.959	-0.104	-0.330	0.122	0.365
Read 1	0.833	0.696	0.970				

Table 2.9. Subgroup analysis stratified by lesion location.

General dentists, operative dentists and endodontists (8 readers, 68 images), RRRC Scenario							
Read 2 (Aided by AI)	**AFROC**	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
	0.892	0.833	0.951	0.071	0.022	0.119	**0.005**
Read 1	0.822	0.749	0.894				
Read 2 (Aided by AI)	**ROC AUC**	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
	0.920	0.872	0.969	0.024	-0.016	0.063	0.231
Read 1	0.897	0.841	0.953				
Read 2 (Aided by AI)	**Sensitivity**	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
	0.931	0.869	0.993	-0.007	-0.056	0.043	0.782
Read 1	0.938	0.893	0.982				
Read 2 (Aided by AI)	**Specificity**	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
	0.733	0.599	0.867	0.138	0.002	0.273	**0.047**
Read 1	0.596	0.461	0.731				
By-Lesion Statistics (4 readers, 56 images), RRRC Scenario							
Read 2 (Aided by AI)	**Sensitivity**	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
	0.888	0.812	0.965	0.067	0.0003	0.134	**0.049**
Read 1	0.821	0.738	0.905				

Table 2.10. Subgroup analysis stratified by readers' specialty: general dentists, operative dentists and endodontists.

Radiologists (4 readers, 68 images), RRRC Scenario							
	AFROC	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.904	0.849	0.960	0.007	-0.020	0.035	0.595
Read 1	0.897	0.847	0.948				
	ROC AUC	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.949	0.907	0.990	-0.009	0.907	0.990	0.594
Read 1	0.957	0.916	0.999				
	Sensitivity	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.928	0.843	1.000	-0.007	-0.083	0.070	0.854
Read 1	0.934	0.855	1.000				
	Specificity	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.875	0.717	1.000	-0.050	-0.304	0.204	0.576
Read 1	0.925	0.833	1.000				
By-Lesion Statistics (4 readers, 56 images), RRRC Scenario							
	Sensitivity	CI lower	CI upper	Read 2 –Read 1	CI lower	CI upper	P-value
Read 2 (Aided by AI)	0.853	0.757	0.948	0.031	0.757	0.948	0.314
Read 1	0.821	0.724	0.919				

Table 2.11. Subgroup analysis stratified by readers' specialty: the radiologists group.