Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v29i03.5607

PERIAPICAL LESIONS DETECTION USING AN ARTIFICIAL INTELLIGENCE TOOL: A RETROSPECTIVE MULTICENTRIC STUDY OVER PERIAPICAL RADIOGRAPHS

Abdulbadea Altukroni[1*], Abdullah Alsaeedi[1] Sakina El-Amri[1], Omar Ezz El-Deen[1] , Mahmoud Khatiri[1], Mohammed Alabduh[1], Turki Jari1, Mwaed Ashri[1], Ahmed Jamaan[1], Omar Alshaikh[1]

[1*]Ministry of Health, Medina, Saudi Arabia, aaltukroni@moh.gov.sa
[1]Taibah University, Medina, Saudi Arabia, aasaeedi@taibahu.edu.sa
[1]Ministry of Health, Ankara, Turkey, sakina.ela.se@gmail.com
[1]Ministry of Health, Cairo, Egypt, research-manager@swc-dent.com
[1]Ministry of Health, Medina, Saudi Arabia, mkhatiri@moh.gov.sa
[1]Ministry of Health, Medina, Saudi Arabia, Malabduh@moh.gov.sa
[1]Ministry of Health, Medina, Saudi Arabia, tjari@moh.gov.sa
[1]Ministry of Health, Medina, Saudi Arabia, maashri@moh.gov.sa
[1]Ministry of Health, Medina, Saudi Arabia, dr-de3o@hotmail.com
[1]Ministry of Health, Medina, Saudi Arabia, Okalshaikh@moh.gov.sa

***Corresponding Author:** Abdulbadea Altukroni
*Ministry of Health, Saudi Arabia, Email: aaltukroni@moh.gov.sa
Address: Alfrhini 3841, Medina, SA, 42363

Abstract

This study was conducted to design and evaluate an AI tool called Make Sure periapical explorer (MSp) to detect periapical lesions on digital periapical radiographs and to compare its performance with the dentists'. This study was a diagnostic, retrospective, and multi-centric study, with a sample size of 2,200 digital periapical radiographs (with 3,680 periapical lesions). The dataset was split into train, validate, and test datasets; the ratio was 8-1-1. 220 images were randomly allocated to test MSp AI model, and the same images were allocated to test 10 certified dentists. The performance metrics used to test and compare MSp performance and the dentist's performance included precision, F1 score, recall, and mean average precision (mAP). Kolmogorov-Smirnov test was used to test the normalization of the distributions. The Kruskal-Wallis test was used to determine the significant difference between the mAP, precision, recall, and F1 scores. The statistical significance was set at $p < 0.05$. MSp achieved a higher performance in all metrics in comparison to the dentists group. There was no statistical difference in the precision metric and recall metric, while there was a statistically significant difference in F1-score and mAP between the two groups. The designed MSp tool proved itself reliable in the detection of periapical lesions in digital periapical radiographs. It also showed a higher performance metrics in detecting periapical lesions when compared to the dentists' group consensus.

1. Introduction

Periapical lesions are one of the most common periodontal pathologies in humans [1]. They develop as sequelae to pulp disease. They often occur without any episode of acute pain and are considered incidental findings [2]. They can cause local inflammation, hard tissue resorption, and the destruction of other periapical tissues [3].

The clinical and radiographic examinations should be the primary criteria for assessing the presence of the periapical lesions. Since clinical methods cannot precisely evaluate their presence, a radiographic examination should be used. However, radiographs provide two-dimensional representation of three-dimensional structures [4]. In addition, some dentists have difficulty detecting and diagnosing periapical lesions on dental radiographs. When there are a few hundred cases to diagnose, an experienced dentist who only needs a few seconds to diagnose and decide whether there is a periapical lesion or not on a single periapical X-ray image may get confused, making errors inevitable [5].

To avoid probable misinterpretation of apical periodontitis associated with certain radiolucencies, it is critical to accurately diagnose periapical lesions. There are several radiolucent lesions that may be similarly relevant to endodontic practice. Some examples are periapical cemental dysplasia, early stages of fibrous dysplasia, ossifying fibroma, odontogenic keratocyst, median maxillary or mandibular cyst, and traumatic bone cyst [6]. This is why a new intervention is needed to support the diagnosis of periapical lesions.

Artificial intelligence (AI) and Deep learning (DL) methods proved to mimic humans' cognitive functions to perform tasks of problem-solving and learning. DL allows computational models composed of multiple processing layers to learn representations of data with various abstraction levels. In this way, images can be used as an input for the neural networks to achieve several different outputs [7].

Radiology is deemed the front door for AI into dentistry; dental literature focuses on three main AI technology applications, including automated diagnosis of dental diseases, localization of anatomical landmarks, and general improvement of image quality [8]. Because of the contrast in radioopacity between the standard tooth apex structure and bone structure and the radiolucent appearance of periapical lesions, their presence or absence may be easily determined. AI prediction capability relies on these distinction features, which allows AI technologies to "learn" to analyze dental radiographs.

This study aims to design and evaluate an AI tool called Make Sure periapical explorer (MSp) to detect periapical lesions on digital periapical radiographs and to compare its performance with the dentists'. The study was testing the hypothesis to evaluate if the designed AI tool was able to detect periapical lesions correctly as compared to dentists. MSp was a trial model created by Smile with Confidence (SWC) Company to test the ability of AI to detect periapical lesions.

2. Materials and Methods

2.1 Study Design

This study was established as a diagnostic, retrospective, and multi-centric study in many hospitals of Saudi Ministry of Health (MOH) in Makkah and Tabuk Regions, Saudi Arabia. The study protocol was approved by the Institutional Review Board (IRB) in the local committee for ethics of health and scientific research in health affairs in Makkah region (H-02-K-076-0821-544) and in Tabuk region (H-07-TU-077). Informed consent was not required for this study as per these committees. All methods were performed in accordance with the Declaration of Helsinki.

2.2 Study Population

2,200 anonymized labeled digital periapical radiographs with periapical lesions were selected between October 2021 and May 2022 from different dental centers and hospitals, including Alnoor Specialist Hospital Makkah; King Faisal Hospital, Makkah; King Abdulaziz Hospital, Makkah; Hera General Hospital, Makkah; North Jeddah Specialized, Jeddah; King Fahd Hospital, Jeddah; King Abdulaziz Hospital, Jeddah; Al Thaghr Hospital, Jeddah; East Jeddah General Hospital, Jeddah; and

Tabuk Specialized Dental Center, Tabuk. The periapical radiographs for this research were retrospectively selected from 26,000 collected periapical radiographs.
The research included all sizes (size 0, 1, and 2) of digital periapical radiographs (taken using the parallel technique), and all teeth (anterior and posterior, upper and lower) with periapical radiolucency, whether they are endodontically treated or not. Radiographs with any number of periapical radiolucencies were included (after interrater agreement between two collaborated dentists). (Figure 1)

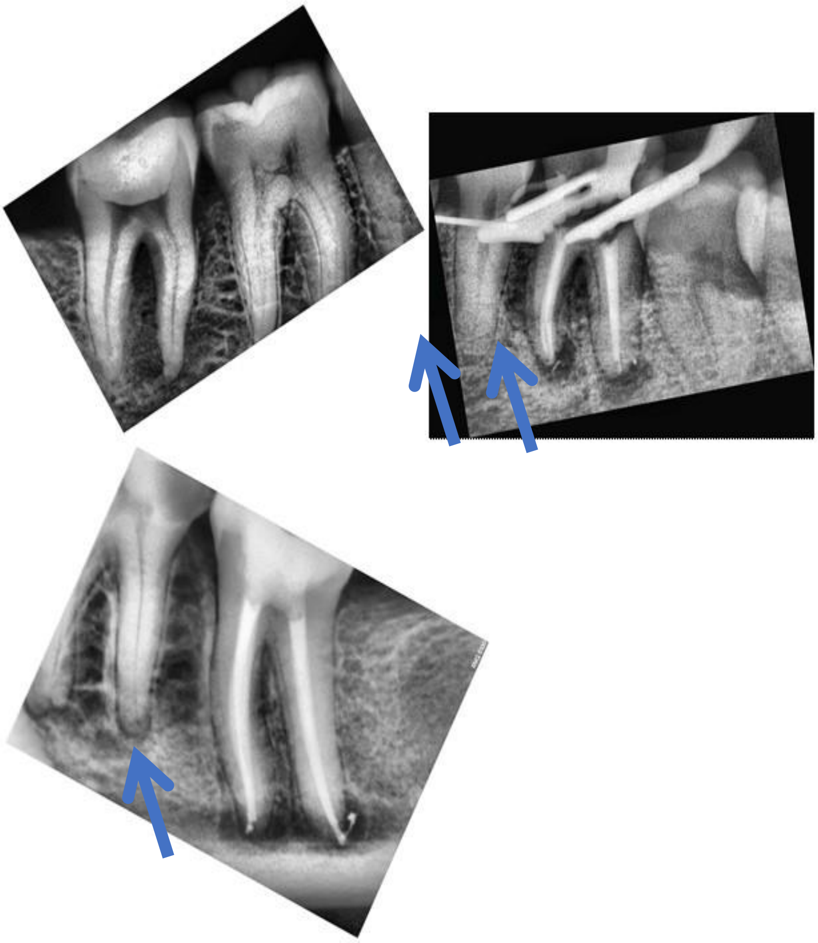

Figure (1): Eligible radiograph samples showing periapical radiolucencies

However, radiographs with more than 50% of the radiograph image missing or unclear and poor contrast between the alveolar bone and dental root apex were excluded from the study. Also, unclear radiographs that were difficult to distinguish because of the severe distortion, artificial noise (scattered radiation from X-ray machine), blur, and low quality were omitted.

2.3 Data Cleaning and Labeling

Each hospital's data was cleaned and labeled internally through two collaborated, qualified dentists with more than two years of experience before being submitted to the Principal Investigator (PI) (Specialist of Restorative Dentistry) through an electronic cloud (Google Drive). The PI then revised the cleaning and labeling process by randomly distributing all collected labeled data to another two collaborated dental practitioners with more than two years of experience, who then checked and confirmed that all data included periapical radiolucency and met our eligibility criteria, and then excluded any data that did not; any radiograph on which the interrater disagreed was excluded. At last, the qualified information was sent to the electronic cloud (Google Drive) of the PI.

All confirmed radiograph images were sent to 10 clinicians (endodontists and general practitioners with more than two years of experience) who manually classified the digital periapical radiographs based on whether there were periapical lesions or not using "Labellmg (Windows_v1.8.0, tzutalin)". This was used to establish the ground truth. A different digital file was sent to each clinician. The annotation was made using boxes. Labeling was done on any radiolucency detected on each radiograph.

The PI received the clinicians' responses as radiographs images in Yolo format and TXT files without knowing their names or contacts (only their titles were known). Clinicians were not given access to each other's data, so they were unfamiliar with one another.

All radiographs were coded by sequential numbers. Later, during data processing, the labeled dataset was randomly divided into train, validate (internal validation), and test datasets using Python's random package. All categorized and labeled datasets was fully anonymized and fed into Google Drive. The programmers were able to have full access to all labeled data.

All data were randomized (sample randomization) using (True Random Generator, Version 2.0.3) before distributing the data for 10 clinicians. This happened when establishing ground truth and in the test dataset, too, for comparison.

The total data assessed for eligibility consisted of 26,000 images; however, 23,660 images were excluded for not meeting the inclusion criteria and for technical reasons. Most images were excluded because they were without periapical lesions; therefore, only 2340 images were randomized. In addition, further 140 images were excluded because of inter-rater disagreement; hence, 2,200 images were labeled and annotated (3680 labels in total).

2.4 Data Processing

Using CLAHE, pre-processing techniques were used to obtain more contrasted black-and-white images with parameter clipLimit = 5.0. Then, dataset was broken down into three parts: train, validate, and test, with an (8-1-1) ratio to prevent overfitting and inaccurate results. The test dataset was used for testing the MSp group and the dentists group (human raters) as well as comparing MSp results with those of the dentists. On the other hand, the validation set was used for validation only. The test dataset was randomized again and distributed to the same 10 human raters.

The MSp model employs multiple CNN algorithm optimization tactics, such as auto learning bounding box anchors, mosaic data augmentation, and the cross-stage partial network. It uses Yolo (You Only Look Once), which is an object detection algorithm. It divided images into cells. Each cell is responsible for detecting objects within it. Yolo uses a single neural network to process the entire picture and then separates it into parts and predicts the bounding boxes for each part. This algorithm only looks once at the image in a way that it makes predictions after one forward pass through the neural network. Then, it delivers the detected objects. Its architecture consists mainly of three parts [9] (Figure 2):

- Backbone (CSPDarknet): It is used to extract key features (rich in useful characteristics) from the input image
- Neck (PANet): A series of layers to mix and combine image features to pass them forward to prediction
- Head (Output): It is responsible for the final detection step

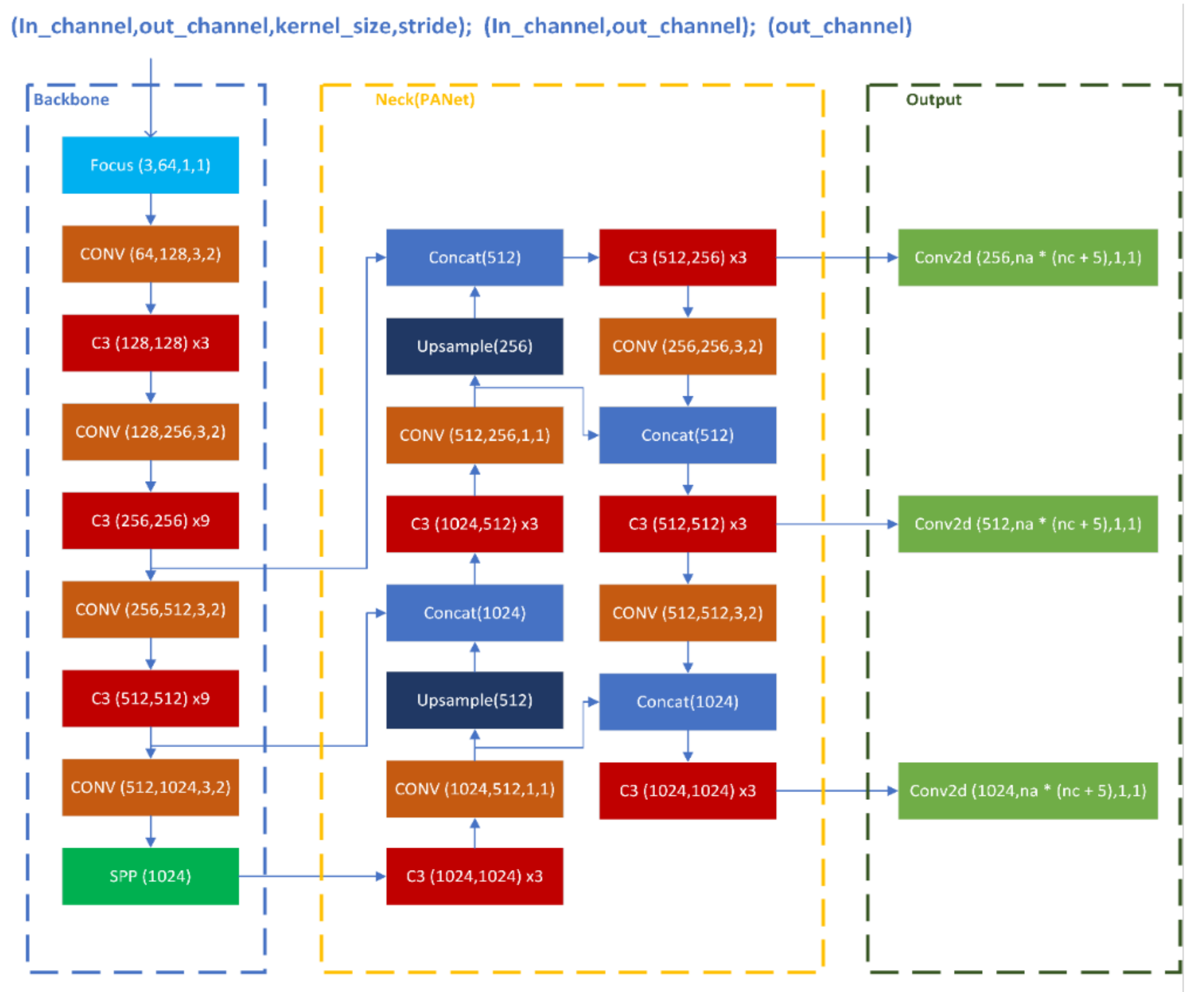

Figure (2). Yolov5 network architecture

2.5 Study Groups

The randomized test dataset contained 220 images with 358 labels and was used for evaluating the performance of MSp. Also, the data was used for comparing the diagnostics differences between 10 certified dentists and the MSp tool. For the dentist groups, the data was divided randomly into 10 blocks, and every block had 22 radiographs.

2.6 Outcome measures

- Mean Average Precision 0.5 (mAP@0.5): Mean average precision, calculated by taking the mean AP (accuracy of our AI tool) over all periapical lesions detected by MSp or Dentists, and/or overall 0.5 (IoU) thresholds.
- Precision: The ratio of correctly predicted positive periapical lesions to the total predicted periapical lesions by MSp or dentists: Precision = True Positive (TP)/ TP + (False Positive) FP.
- Recall: Calculates how many actual periapical lesions the model or the dentists has captured. Recall = TP/TP + (False Negative) FN.
- F1 Score: Defined as the function of precision and recall. It is calculated when a balance between precision and recall is needed. F1 = 2 × Precision × Recall/ Precision + Recall.

2.7 Sample size calculation

Based on Pauwels R et al., [10] 5,600 periapical radiographs were taken from 10 prepared bovine ribs sockets for the detection of simulated periapical lesions on periapical radiographs. When data were split up by socket, the mean sensitivity, specificity, and ROC-AUC values were 0.79, 0.88, and 0.86, respectively. To achieve higher generalizability in our study, 2,200 (training, validation, and testing sets) retrospective digital periapical radiographs for real patients were selected for periapical lesions detection.

2.8 Statistical Analysis

The descriptive metrics (mean, median, and range) of MSp model and the dentists performance presented as the percentage of mAP @0.5, precision, recall, and F1 score were calculated to compare both the performance of the MSp model and dentists, using the Keras library on top of TensorFlow "Yolo v5" in Python and SPSS 26 for windows.

Kolmogorov-Smirnov test was used to test the normalization of the distributions. Then, the Kruskal-Wallis test was used to determine the significant difference between the mAP @0.5, precision, recall, and F1 scores for both the dentists over test dataset and MSp over the same test dataset. The statistical significance was set at $p < 0.05$.

3. Results

The total data assessed for eligibility consisted of 26,000 images; however, 2,200 images were labeled and annotated (3680 labels in total). The data were split in a ratio of (8:1:1) for train, test, and validate respectively. The train set consisted of 1,760 images (2972 labels), the validation set consisted of 220 images (350 labels) and the test set comprised of 220 images (358 labels) (Table 1).

Table 1. Data and labelled distribution

Observed Frequency	Train Set	Val Set	Test Set	Total
Images	2200	1760	220	220
Periapical_Radiolucency	3680	2972	350	358

The number of true positive (TP) periapical lesions detected by MSp was 351 labels, while the dentists group had 336. The number of false positive (FP) periapical lesions detected by MSp was 47 labels, while the dentists group had 63. The number of false negative (FN) periapical lesions detected by MSp was 56 labels, while the dentists group had 71.

Therefore, the obtained results for MSp model showed an mAP<0.5 of 0.91, while 0.89 for the dentists group ($P<0.05$). The obtained results for MSp model showed a precision of 0.88, while 0.84 for the dentists group ($P= 0.302$). The obtained results for MSp model showed a recall of 0.86, while 0.82 for the dentists group ($P= 0.068$). The obtained results for MSp model showed an F1 score of 0.87, while 0.83 for the dentists group ($P= 0.029$) (Table 2). (Figure 3)

Table 2. MSp and dentists performance metrics

	MSp	Dentists	P
Precision	0.88	0.84	0.302
Recall	0.86	0.82	0.068
F1 Score	0.87	0.83	0.029

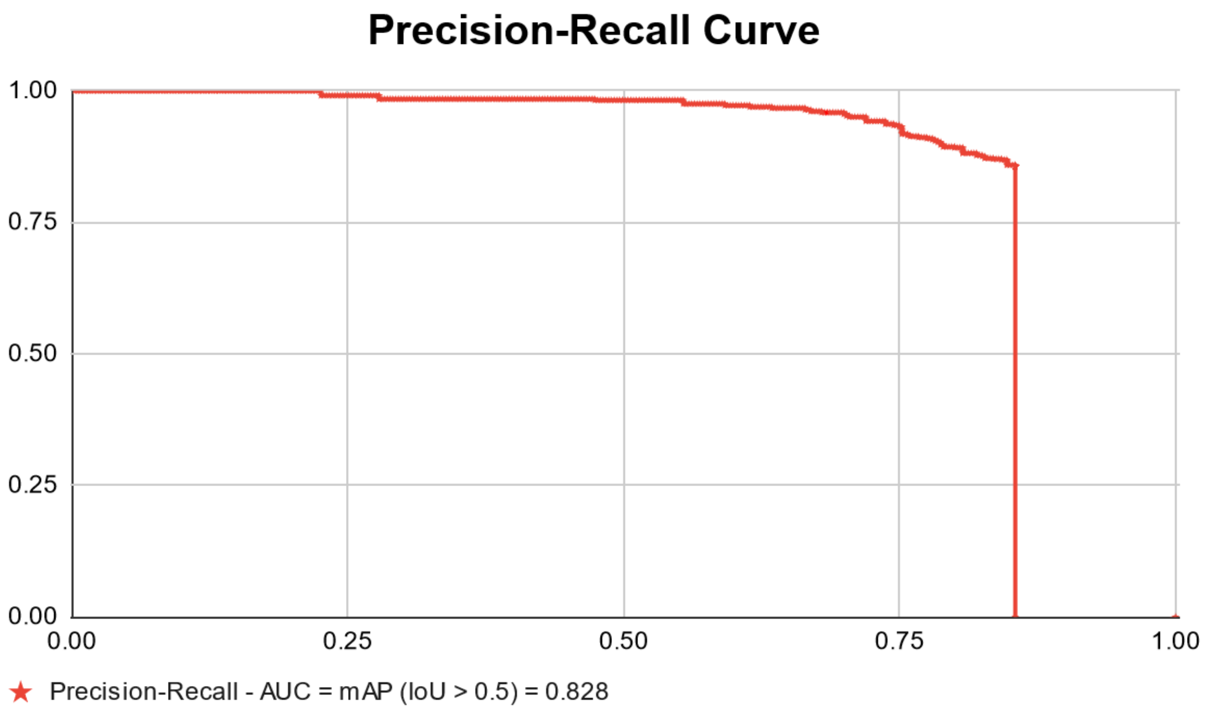

Figure (3): Precision-Recall curve

4. Discussion

Radiographs are the most significant reference tool to help identify various tooth and jaw pathologies and disorders [11]. The most frequently used radiographs in dentistry are periapical radiographs, which may be used to diagnose periodontitis, dental caries, and periapical lesions, as they allow us a close-up look of each tooth [12].

Periapical lesions can be identified by analyzing the different gray scales that the radiographical images create [13]. The differences in gray scales between the normal tooth apex structure, bone structure, and the periapical lesions make it possible to identify the presence or absence of periapical diseases (radiopacity to radiolucent) [14].

However, the intraoral periapical radiographs have some disadvantages because they are a two-dimensional representation of a three-dimensional object. It might be challenging for a new dentist to accurately detect or confirm a diagnosis of periapical pathology on a periapical radiograph, particularly in cases of initial lesions [15].

AI is an emerging boon for overcoming such diagnostic challenges by identifying minute changes in radiographs. As a result, AI may be utilized in conjunction with other imaging modalities to help make appropriate diagnoses and treatment plans [16].

The present study designed and evaluated an AI tool called MSp to detect periapical lesions on digital periapical radiographs and to compare its performance with the dentists'.

10 experienced dentists from different hospitals were used in our investigation. The ground truth was established using their findings, and then the same dentists participated in the comparison. Even though the ground truth was based on the dentists' finding, the inclusion of several dentists from various hospitals and geographical locations ensured that there were diverse experiences in establishing the ground truth, which will produce a more comprehensive understanding of the various outcomes of MSp.

The number of images (2,200) used in our study was also comparable with that of Endres MG et al., [17], who used 2,902 panoramic radiographs to detect periapical lesions. When these data were compared to the technique and findings of the current investigation, the number of images employed and the verified diagnostic performance were nearly similar.

The results of this study showed that MSp has a higher performance than the dentists in all metrics for detecting periapical lesions. The MSp model showed an mAP<0.5 of 0.91, a precision of 0.88, a recall of 0.86, and an F1 score of 0.87. There was no statistical difference in the precision metric (0.8819 for MSp group, - 0.8421 for the dentists group), recall metric (0.8624 for MSp group, - 0.8256 for the dentists group), F1-score (0.8720 for MSp group, - 0.8337 for the dentists group), and mAP@ .5 metric (0.9154 for MSp group, - 0.8988 for the dentists group) between the two groups.

This was in line with a systematic review that concluded that the neural networks have even outperformed the specialists. Also, these models can be of greater assistance as an expert opinion for less experienced and nonspecialists [18].

A study by Chun-wei Li et.al found that the neural networks showed the possibility of automatically identifying the periapical lesions with a success rate of 92.75%, which was similar to our obtained results for MSp model [13].

The study by Endres MG et al., [17] also demonstrated that a DL model trained on radiographic images can match the mean diagnostic performance of oral and maxillofacial (OMF) surgeons in detecting periapical radiolucent alterations. They later found out that the ability of oral and maxillofacial surgeons to identify periapical radiolucencies on panoramic radiographs may be limited and that radiolucent periapical alterations may be missed. However, the DL model performed better than half of experienced OMF surgeons and may serve as a complementary tool in diagnosing periapical radiolucencies.

Another study that used cone-beam computed tomography images (CBCT) to verify the diagnostic performance of an AI system based on the deep convolutional neural network method in detecting periapical pathosis showed that only one tooth was misidentified. A periapical lesion was appropriately detected 92.8% of the time, and volume measurements taken by humans and AI systems were equivalent. According to the author, deep learning-based AI systems may be useful for detecting periapical pathosis on CBCT images for clinical applications [19]. A recent study by Hamdan et al., [20] also showed that DL technology enhances dental professionals' abilities to detect apical radiolucencies on intraoral radiographs.

However, there were some limitations in our study. Firstly, the clinical parameters were not included, which is an aspect that should be taken into account to have a more accurate diagnosis. Also, neural networks, in general, including our tool MSp, are black boxes that cannot explain machine learning characteristics and the grounds for making decisions based on that learning. The limitations of the digital periapical radiographs, such as image magnification and distortion and the lack of three-dimensional information, may lower the MSp tool's diagnostic accuracy.

Moreover, the dataset was not divergent regarding the age and sex because it was collected without prior knowledge of the patients' details. We did not categorize the teeth according to their types, which may have affected the accuracy of the results according to the teeth type. Finally, further research needs to be conducted with a larger dataset and different experienced dentists for more reliable results.

Future initiatives for improving AI-based periapical lesions diagnosis on intraoral pictures should involve image segmentation as an alternate option, which should be carried out by well-trained and calibrated dental practitioners under the supervision of senior specialists. To accomplish this, periapical lesions must be marked pixel by pixel on each accessible image and the diagnosis accuracy must be reassessed. In comparison to the currently utilized classification methodology, this more precise but otherwise time- and resource-intensive approach provides thorough periapical lesions localization.

5. Conclusions

From the above results and discussion, it is concluded that the designed MSp tool proved itself reliable in the detection of periapical lesions in digital periapical radiographs. It also showed a higher performance metrics in detecting periapical lesions when compared to the dentists group.

6. Author Contributions

All authors have made substantial contributions to the conception and design of the work, approved the submitted version, and have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

7. Conflict of Interest

The authors in this study declare no conflict of interest.

8. Data Availability

Data is available upon request from the corresponding author.

9. Funders

This study was fully funded by Smile with Confidence (SWC) Company. The funding source had no role in any part of the research process.

10. References

1. Schulz, M., von Arx, T., Altermatt, H.J., Bosshardt, D. Histology of Periapical Lesions Obtained During Apical Surgery. *J Endod*. **35**, 634–42 (2009).
2. Fernandes M, Ataide I. Nonsurgical management of periapical lesions. *J Conserv Dent*. **13**, 240 (2010).
3. Croitoru, I.C. *et al.* Clinical, imagistic and histopathological study of chronic apical periodontitis. *Rom J Morphol Embryol*. **57**, 719-728 (2016).
4. Keser, G., & Pekiner, F. Comparative Evaluation of Periapical Lesions Using Periapical Index Adapted for Panoramic Radiography and Cone Beam Computed Tomography. *Clin Exp Health Sci.* **8**, 50-55 (2017).
5. McCaul, L. K., McHugh, S., & Saunders, W. P. The influence of specialty training and experience on decision making in endodontic diagnosis and treatment planning. *Int Endod J*. **34**, 594–606 (2001).
6. Modi, K., Padmapriya, R., Elango, S., Khandelwal, P., Arul, B., & Natanasabapathy, V. Nonmalignant nonendodontic lesions mimicking periapical lesions of endodontic origin: A systematic review. *J Conserv Dent*. **25**, 214–225 (2022).
7. Vranckx, M. *et al.* Artificial Intelligence (AI)-Driven Molar Angulation Measurements to Predict Third Molar Eruption on Panoramic Radiographs. *Int J Environ Res Public Health.* **17**:3716 (2020).
8. Hung, K., Yeung, A. W. K., Tanaka, R., & Bornstein, M. M. Current Applications, Opportunities, and Limitations of AI for 3D Imaging in Dental Research and Practice. *Int J Environ Res Public Health*. **17**, 4424 (2020).
9. Zhu, X., Lyu, S., Wang, X., & Zhao, Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. *Proc IEEE Int Conf Comput Vis*. 2778–2788 (2021).
10. Pauwels, R. et al. Artificial intelligence for detection of periapical lesions on intraoral radiographs: Comparison between convolutional neural networks and human observers. *Oral Surg Oral Med Oral Pathol Oral Radiol*. **131**, 610–616 (2021).
11. Shah, N., Bansal, N., & Logani, A. Recent advances in imaging technologies in dentistry. *World J Radiol*. **6**, 794–807 (2014).
12. Zaki, H. A. M., Hoffmann, K. R., Hausmann, E., & Scannapieco, F. A. Is Radiologic Assessment of Alveolar Crest Height Useful to Monitor Periodontal Disease Activity? *Dent Clin North Am*.**59**:859-72 (2015).
13. Li, C. W. *et al.* Detection of Dental Apical Lesions Using CNNs on Periapical Radiograph.

Sensors (*Basel, Switzerland).* **21**, 7049 (2021).

14. Camps, J., Pommel, L., & Bukiet, F. Evaluation of periapical lesion healing by correction of gray values. *J Endod*. **30**, 762–766 (2004).
15. Fuhrmann, R. A. W., Bücker, A., & Diedrich, P. R. Assessment of alveolar bone loss with high resolution computed tomography. *J Periodontal Res*. **30**, 258–263 (1995).
16. Ahuja, A. S. The impact of artificial intelligence in medicine on the future role of the physician. *Peer J*. **7**, e7702 (2019).
17. Endres, M,G. *et al.* Development of a Deep Learning Algorithm for Periapical Disease Detection in Dental Radiographs. *Diagnostics (Basel).* **10**:430 (2020).
18. Boreak, N. Effectiveness of Artifcial Intelligence Applications Designed for Endodontic Diagnosis, Decision-making, and Prediction of Prognosis: A Systematic Review. *J Contemp Dent Pract.* **21**, 926–934 (2020).
19. Orhan, K., Bayrakdar, I. S., Ezhov, M., Kravtsov, A., & Özyürek, T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *Int Endod J.* **53**, 680–689 (2020).
20. Hamdan, M. H., Tuzova, L., Mol, A., Tawil, P. Z., Tuzoff, D., & Tyndall, D. A. The effect of a deep-learning tool on dentists' performances in detecting apical radiolucencies on periapical radiographs. *Dentomaxillofac Radiol.* **51**, 20220122 (2022).