Deep Learning Application in the Detection of Pathological Exposure of Pulp Using Periapical Radiographs: A Retrospective Multicentric Study

Abdulbadea Altukroni (research-manager@swc-dent.com)
Ministry of Health

Abdullah Alsaeedi
Taibah University

Cristina Gonzalez-Losada
Complutense University of Madrid

Jae-Hong Lee
Jeonbuk National University

Mohamed Alabudh
Ministry of Health

Mahir Mirah
Taibah University

Sakina El-Amri
Ministry of Health

Omar Ezz El-Deen
Ministry of Health and Population

Research Article

Keywords: Artificial Intelligence, Caries, Deep Learning, Pulp Exposure, Periapical Radiograph, Yolov5x

Posted Date: March 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2629211/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Introducing artificial intelligence into the medical field proved to be beneficial in automating tasks and streamlining the practitioners’ lives. Hence, this study was conducted to design and evaluate an Artificial Intelligence (AI) tool called Make Sure Caries Detector and Classifier (MSc) for detecting pathological exposure of pulp on digital periapical radiographs and to compare its performance with dentists.

Methods

This study was a retrospective multi-centric study, with over 3461 digital periapical radiographs from different countries’ and centers. MSc was built using Yolov5-x model, which was used for exposed and unexposed pulp caries detection. The dataset was split into train, validate, and test datasets; the ratio was 8-1-1 to prevent overfitting. 345 images with 752 labels were randomly allocated to test MSc. The performance metrics used to test MSc performance included mean average precision (mAP), precision, F1 score, recall, and area under receiver operating characteristic Curve (AUC). The metrics used to compare the performance with that of 10 certified dentists were: right detection exposed, right detection not exposed, false detection exposed, false detection not exposed, missed diagnosis, and over diagnosis.

Results

MSc achieved a performance of more than 90% in all metrics examined: an average precision of 0.928, recall of 0.918, F1-score of 0.922, and an mAP@ .5 (AUC) of 0.956 (P < .05). The results showed a higher mean for all right diagnosis parameters in MSc group while a higher mean for all wrong diagnosis parameters in the dentists group (P < .05).

Conclusions

The designed MSc tool proved itself reliable in the detection and differentiating between Exposed and Unexposed pulp. It also showed a better performance when compared to the 10 dentists.

1 Background

Dental caries is considered a multifactorial disease, resulting from the demineralization of the tooth's hard tissues and the product of bacterial activity. (1) As a consequence of this process, a caries lesion develops, which may lead to pulpal exposure.

It is important to prevent pulpal exposure as the pulp's vitality is critical to the tooth's ability to function physiologically in the mouth. Pulpal inflammation generally occurs due to microbial challenges. The
severity of the inflammatory response increases as the carious lesion approaches the pulp. \(^{(2)}\)

The magnitude and severity of pulp-related problems should not be underestimated since they may lead to oral sepsis. Hence, correct diagnosis and management are essential. Practitioners need precise knowledge of the status of exposure to determine the appropriate type of restoration and treatment planning. \(^{(2)}\)

Furthermore, the clinical and radiographic examinations are the primary criteria for assessing, describing, and diagnosing the depth and the risk of dental caries to prevent pulpal exposure. Also, the thickness of residual dentine cannot be evaluated clinically; this is why a radiographic examination should be used. \(^{(2)}\)

However, it is challenging to diagnose the exposure of the pulp using radiographs alone, especially for the dentists who do not have enough specialized training or time devoted to detailed diagnosis. Numerous studies have reported large variations in the reliability and accuracy of detecting dental caries, depending on the clinician’s level of experience. \(^{(3)}\)

This is why it’s important to find ways to diagnose the pulp status of teeth with deep caries. One of these ways is the use of Artificial Intelligence (AI).

AI and deep learning (DL) methods can 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. \(^{(2)}\)

Today, AI began emerging in the health care fields, which reduces diagnostics errors in daily practice. Dentist’s workflow became more efficient with automated suggestions for complex cases, better treatment planning, and prediction of diseases and outcomes. Many people, including doctors and scientists, are not yet familiar with the concepts and true potential of AI and the impact that it can have on our personal and professional lives. \(^{(4)}\)

This study was conducted to design and evaluate an AI tool called Make Sure Caries Detector and Classifier (MSc) for detecting pathological exposure of pulp on digital periapical radiographs and to compare how correct is the diagnosis between MSc and Dentists.

\section*{2 Methods}

\subsection*{2.1 Study Design}

This study was established as a retrospective, multi-centric study with the goal of proposing and evaluating an AI tool that was used for detection of pathological pulp exposure in digital periapical radiographs as well as to compare how correct is the diagnosis between MSc and dentists.
The study was testing the hypothesis to evaluate if the designed AI tool was able to detect Exposed/Unexposed pulp caries correctly as compared to dentists.

Based on our updated knowledge, this is the first study evaluating the ability of AI algorithms in differentiating between Exposed and Unexposed pulp caries.

The study protocol was approved by the IRB in the local committee for ethics of health and scientific research in health affairs in Medina region (IRB 25/2021), Daejeon Dental Hospital, Wonkwang University College of Dentistry (Dr. Lee), Complutense University of Madrid (Dr. Cristina), and Taibah University (Dr. Maher). Informed consent was waived by the IRB in the local committee for ethics of health and scientific research in health affairs in Medina region (IRB 25/2021) due to retrospective nature of the study. All methods were performed in accordance with the Declaration of Helsinki.

2.2 Study Population

3461 Anonymized labeled digital periapical radiographs, with 3106 exposed pulp caries and 4612 unexposed pulp caries, were selected between April 2021 and November 2021 from different centers in different countries, including Saudi Arabia (Specialized Dental Center, Aohd Dental Center, and Alhijra Dental Center), (Faculty of Dentistry, Taibah University), Spain (Faculty of Dentistry, Complutense University of Madrid), and Korea (Faculty of Dentistry Daejeon Dental Hospital). The periapical radiographs for this research were retrospectively selected from 18000 collected periapical radiographs.

The research included all sizes of digital radiograph and all carious teeth, including those with periapical or periodontal issues. Additionally, radiographs with noticeable caries by human eyes were included, whether permanent or deciduous, anterior or posterior teeth.

However, digital periapical radiographs with root caries and orthodontic brackets were excluded. Also, unclear radiographs, those with more than half of the film missing, or ones that are difficult to discern due to extreme distortion, were omitted.

2.3 Data Cleaning and Labeling

Each hospital's data was cleaned and labeled internally in each hospital through collaborated qualified dentists before being submitted to the Principal Investigator (PI). The PI then revised the cleaning and labeling processes by randomly distributing all collected labeled data to two collaborated dental practitioners, who then checked and confirmed that all data included dental caries and met our eligibility criteria, and then excluded any data that did not. Each digital periapical radiograph was labeled when the conclusion was affirmed.

QuestionPro testing platform was used for distributing all data in each institution, as well as for the reviewer dentists. QuestionPro was also used for randomizing the database on sample randomization before distributing the data for every dentist in each hospital.
All data were interpreted and labeled using LabelImg (windows_v1.8.0). Lastly, all categorized and labeled datasets were fully anonymized and fed into an electronic cloud (Google Drive) as radiographs images in JPG format and TXT files as YOLO format.

### 2.4 Data Processing

Using CLAHE, pre-processing techniques were used to obtain more contrasted black-and-white images (Fig. 1). Then, dataset was broken down into three parts: Train, validate, and test, with an (8-1-1) ratio. The test dataset was used for testing the MSc and comparing MSc results with those of the dentists. On the other hand, the validation set was used for validation only.

Yolo (You Only Look Once) 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 only one forward pass through the neural network. Then, it delivers the detected objects. Its architecture consists mainly of three parts.

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

### 2.5 Study Groups

The randomized test dataset contained 345 images with 752 labels, and was used for evaluating the performance of MSc. Also, the data was used for comparing the diagnostics differences between 10 certified dentists and the MSc tool. For the dentist groups, the data was divided randomly into 10 blocks, and every block had between 34 to 35 radiographs.

### 2.6 Outcome measures

#### 2.6.1 MSc metrics outcome

##### 2.6.1.1 Primary outcome

**Mean Average Precision 0.5 (mAP@0.5)**

Mean average precision, calculated by taking the mean AP (accuracy of our AI tool) over all Exposed and Unexposed pulp caries and/or overall 0.5 (IoU) thresholds

##### 2.6.1.2 Secondary outcome

- **Precision**: The ratio of correctly predicted positive Exposed/Unexposed pulp caries to the total predicted Exposed/Unexposed pulp caries: \( \text{Prec.} = \frac{TP}{TP + FP} \)
• **Recall**: Calculates how many actual Exposed/Unexposed pulp caries true positives the model has captured, labeling them as positives. Recall = TP / TP + 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

• **AUC**: Area under the Receiver Operating Characteristic (ROC) Curve (AUC). AUC integrated from (0, 0) to (1, 1) gave the aggregate measure of all possible Exposed and Unexposed pulp caries detection and classification thresholds.

### 2.6.2 Diagnostics outcome

#### 2.6.2.1 Primary outcome

- **Correct Diagnosis**: The sum of Right Diagnosis Exposed (RDE), which is all exposed pulp caries that were diagnosed correctly based on the ground truth, and Right Diagnosis Not Exposed (RDNE), which is all not exposed pulp caries that were diagnosed correctly based on the ground truth

#### 2.6.2.2 Secondary outcome

- **Wrong diagnosis**: The sum of False Diagnosis Exposed (FDE), which is all exposed pulp caries that were diagnosed as unexposed pulp caries based on the ground truth, False Diagnosis Not Exposed (FDNE), which is all unexposed pulp caries that were diagnosed as exposed pulp caries based on the ground truth, Over Diagnosis (OV, Exposed\Unexposed), which is all non-carious objects that were diagnosed as exposed pulp caries or unexposed pulp caries based on the ground truth, and Missed Diagnosis (MD, Exposed\Unexposed), which is all exposed\unexposed pulp caries that were undiagnosed based on the ground truth.

### 2.7 Sample size calculation

Based on the results of the study, (3) which aimed to estimate optimal deep CNN algorithm weight factors for training and validation dataset of both carious and non-carious molars and premolars teeth, at diagnostic accuracy 82.0%, sensitivity 81.0%, specificity 83.0%, PPV 82.7%, and NPV 81.4%, and with an alfa error of 5% and a confidence interval of 95%, a sample size of 3000 periapical radiographs in total were chosen. To achieve higher diagnostic performance metrics, the teeth were not classified based on tooth position, and 3445 digital periapical radiographs were selected in this study.

### 2.8 Statistical Analysis

The descriptive metrics of MSc model performance presented as the percentage of Mean Average Precision (mAP @0.5), Precision, Recall, F1 score, and area under the curve (AUC) of the test dataset were calculated using the Keras library on top of TensorFlow "Yolo v5" in Python. On the other hand, the mean, frequency distribution, median, and range of the diagnostics measures (right detection exposed, false detection exposed, right detection not exposed, false detection not exposed as well as over detection,
missed diagnosis, correct diagnosis, and wrong diagnosis) were calculated to compare both the performance of the MSc model and 10 certified dentists.

Shapiro-Wilk test was used to test the normalization of the distributions. Then, the Wilcoxon Signed Ranks Test was used to determine the significant difference between the values (right detection exposed, false detection exposed, right detection not exposed, false detection not exposed, as well as, over detection, missed diagnosis, correct diagnosis and wrong diagnosis) for both the dentists over test dataset and MSc over the test dataset. The statistical significance was set at p < 0.05.

3 Results

The total data assessed for eligibility consisted of 18000 images; however, 14539 images were excluded for not meeting the inclusion criteria and for technical reasons. Most images were excluded because they were caries free; therefore, only 3461 images were randomized. In addition, further 16 images were excluded because of inter-rater disagreement; hence, 3445 images were labeled and annotated (7718 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 2,755 images (6,171 labels), the validation set consisted of 345 images (795 labels) and the test set comprised of 345 images (752 labels). Lastly, the test dataset was used for analyzing MSc group and dentists group (Fig. 2).

3.1. MSc performance metrics

The number of true positive (TP) exposed and unexposed pulp caries that were detected by MSc over the test dataset is 691 labels, the number of false positive (FP) exposed and unexposed pulp caries that were detected by MSc over the test dataset is 56 labels, and the number of false negative (FN) exposed and unexposed pulp caries that were detected by MSc over the test dataset is 61 labels. Therefore, the obtained results of our model showed a mAP < 0.5 of 95.6%, precision of 92.8%, a recall of 91.8%, and an F1 score of 92.2% (P > .05) (Table 1). The Area Under Curve (AUC) value was 0.956. (p > .05) (Fig. 3).

(Table 1) MSc performance metrics

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>mAP(&lt; 0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.928</td>
<td>0.918</td>
<td>0.922</td>
<td>0.956</td>
</tr>
<tr>
<td>Exposed</td>
<td>0.93</td>
<td>0.925</td>
<td>0.925</td>
<td>0.97</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>0.926</td>
<td>0.921</td>
<td>0.921</td>
<td>0.943</td>
</tr>
</tbody>
</table>

*All results were statistically significant P < 0.05

3.2. Diagnostics measures
The mean value of RDE, RDNE, OD, RD for the MSc group was higher than that of the dentists group. Meanwhile, the mean value for the FDE, FDNE, MD, and WD for the MSc group was lower than that of the dentists group. The mean and median values are listed in (Table 2).

**Table 2** Dentists versus MSc’s diagnosis

<table>
<thead>
<tr>
<th>Dentists Over Test Data</th>
<th>MSc Over Test Data</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>RD Exposed</td>
<td>0.64</td>
<td>0</td>
</tr>
<tr>
<td>FD Exposed</td>
<td>0.05</td>
<td>0</td>
</tr>
<tr>
<td>RD Not Exposed</td>
<td>0.72</td>
<td>0</td>
</tr>
<tr>
<td>FD Not Exposed</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>Over Detection</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>Missed Diagnosis</td>
<td>0.51</td>
<td>0</td>
</tr>
<tr>
<td>Right Diagnosis</td>
<td>1.36</td>
<td>0</td>
</tr>
<tr>
<td>Wrong Diagnosis</td>
<td>0.64</td>
<td>0</td>
</tr>
</tbody>
</table>

4 Discussion

The present study designed and evaluated an AI tool called Make Sure Caries Detector and Classifier (MSc) to detect pathological exposure of pulp on digital periapical radiographs, demonstrating that MSc detects exposed/unexposed pulp caries accurately more than dentists. This is considered the first study to test the use of AI in the detection of exposed and unexposed pulps.

Radiography is frequently used as a reference tool to aid in the identification of different dental and jaw disorders. The technique of locating or finding all information on a radiograph by providing a black, white, and grey image is known as radiograph interpretation.

The results of the periapical radiograph examination have limitations that are only conjectural and subjective since they are two-dimensional images with three-dimensional.

However, it is challenging to diagnose dental caries using radiographs alone, especially for the dentists who do not have enough specialized training or time devoted for detailed diagnosis; the reason is the settings of various parameters, including brightness, shadow and contrast.
Nowadays, Artificial intelligence (AI) is becoming essential in radiology due to its ability to detect abnormalities in radiographic images unnoticed by the naked human eye.\(^8\)

AI can help in standardizing dental caries according to caries depth. This is why we created our MSc tool: to analyze its accuracy in detection of exposed and unexposed pulps in comparison to dentists.

In our study, we used ten examiners from different hospitals and parts of the world. The dentists involved for the comparison were all experienced experts, and their results were used to set the “ground truth”. Including different dentists from different hospitals and parts of the world ensured that there will be different experiences in setting the ground truth, which will yield a broader view of the different outcomes of MSc.

However, this was not in line with a systematic review that showed that trials with only one examiner yielded the best results, indicating that the same caries detection criteria were always utilized. Also, another study used four experts to analyze the photos, which was considered the second best. Finally, the trial with two examiners produced the worst accuracy result.\(^9\)

The number of images (3445) used in our study was also comparable with that of Cantu et al, who used 3292 bitewing radiographs to detect carious exposure. 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.\(^10\)

In the current study, MSc showed a high mAP (0.956) at 50% Confidence level in detecting exposed and unexposed pulp, and this is considered outstanding.

In addition, with an AUC of 0.956 for all classes in our trial, MSc's performance is considered exceptional since an AUC value of 0.8 to 0.9 is excellent, and more is exceptional.\(^11\)

The obtained results of our MSc model showed a precision of 92.8%, recall of 91.8%, and F1 score of 92.2% in detecting exposed and unexposed pulp caries. In addition, MSc showed excellent results in detecting exposed and unexposed pulp caries correctly and less wrong detections in comparison with the dentists. This means that the model yielded better scores than the dentists. Moreover, the model seemed to be more effective and reliable than the dentists, since the ten experienced dentists did not show good consistency and stability, and the model was much faster and more accurate in lesion classification. However, further research needs to be conducted with a larger dataset and different experienced dentists.

This was in line with a study\(^3\) that used deep CNN-based computer-vision for dental caries detection, and the results were 89.0%, 88.0%, and 82.0% for diagnostic accuracies of premolar, molar, and both premolar and molar models, respectively. The deep CNN algorithm achieved an AUC of 0.917 on premolar, an AUC of 0.890 on molar, and an AUC of 0.845 on both premolar and molar models. In another study conducted by Kühnisch et al, the CNN accurately detected cavities in 92.5% of instances (SE, 89.6; SP, 94.3; AUC, 0.964). This was also similar to the finding in other studies.\(^10\)(\(^12\))
Another study in India evaluated the diagnostic performance of an algorithm where a neural network is developed to detect dental caries in digital radiographs. The system gave an accuracy of 97.1%, false positive (FP) rate of 2.8%, receiver operating characteristic (ROC) area of 0.987 and precision recall curve (PRC) area of 0.987. This is considered high accuracy compared to our study, but the authors only employed 105 photos and performed 10-fold cross-validation; their dataset wasn't enough. (13)

Other accuracy comparisons with studies that used different AI methods were as follows: Singh et al. suggested a caries detection method based on Radon Transformation and DCT employing dental X-ray images. (14) Selected characteristics are retrieved using the PCA approach and used to the Random Forest classifier, yielding an accuracy of 86%.

During another study, (15) a caries detection system utilizing SVM and got 86.15% accuracy for the training dataset and 77.34 percent accuracy for the test dataset.

A study (16) developed Convolutional Neural Network-based categorization of major dental illnesses, which achieved an accuracy of 87.5% in detecting dental caries.

Additionally, a similar study (17) assessed the detection of carious exposure using AI on bitewings, and the neural network had an accuracy of 80%, while dentists had a considerably lower mean accuracy of 71%. The neural network demonstrated significant sensitivities of 70% or higher for both initial and advanced lesions. Dentists' sensitivities for initial lesions were generally low while those for advanced lesions ranged between 40% and 75%. When these data were compared to the technique and findings of the current investigation with MSc, the number of images employed and the verified diagnostic performance were nearly comparable.

Nevertheless, our MSc model had higher mean in over detection variable, which means higher probability for detecting non caries objects as exposed or unexposed pulp caries. That can be considered as limitation for AI, as it can consider some radiographic artificial changes as carious lesions.

There were also 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 MSc, are black boxes that cannot explain machine learning characteristics and the grounds for making decisions based on that learning. Also, the limitations of the digital periapical radiographs, such as image magnification and distortion and the lack of three-dimensional information, may lower the MSc tool's diagnostic accuracy.

Future initiatives for improving AI-based caries 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, caries 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 caries localization.
5 Conclusions

From the above results and discussion, it is concluded that:

1. The designed AI model proved itself reliable in the detection of pathological pulp exposure, and in the differentiation between exposed and unexposed pulp caries on digital periapical radiographs.
2. The designed AI model detected pathological pulp exposure on digital periapical radiographs more correctly and effectively than the 10 dentists.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the IRB in the local committee for ethics of health and scientific research in health affairs in Medina region (IRB 25/2021), Daejeon Dental Hospital, Wonkwang University College of Dentistry (Dr. Lee), Complutense University of Madrid (Dr. Cristina), and Taibah University (Dr. Maher). Informed consent was waived by the IRB in the local committee for ethics of health and scientific research in health affairs in Medina region (IRB 25/2021) due to retrospective nature of the study. All methods were performed in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors in this study declare that they have no competing interests.

Funding

This study was self-funded.

Authors' contributions

All authors contributed to the conception, data acquisition and interpretation, performed the statistical analysis and drafted and critically revised the manuscript.

Acknowledgements

Not applicable
References

Figures

Image 1

a) Before CLAHE

b) After CLAHE

Figure 1

Before and After CLAHE

Assessed for eligibility
(n=18,000)

Excluded (n=14,539)
- Not meeting inclusion criteria (n=14,273)
- Excluded for technical reasons (n=266)

Randomized (n=3,461)

Excluded for intra-rater disagreement
(n=16)

Labelled and annotated

3,445 images
7718 labels

Data Processing

Data Training: Train set
- 2,755 images, 6,171 labels
- Exposed: 2491 labels
- Not exposed: 3680 labels

Data evaluation: Val set
- 345 images, 752 labels
- Exposed: 300 labels
- Not exposed: 452 labels

Data evaluation: Test set
- 345 images, 795 labels
- Exposed: 315 labels
- Not exposed: 480 labels

Processed as a train and validation
dataset for creating the model

Analysis

Projected for 10 Dentists
- 345 images
- 752 labels

Projected for the AI model
- 345 images
- 752 labels

Figure 2

Flow Diagram
Figure 3

AUC for Precision-Recall curve