BULLETIN OF STOMATOLOGY AND MAXILLOFACIAL SURGERY

Volume 21, Issue 12

DOI:10.58240/1829006X-2025.21.12-429

ORIGINAL RESEARCH

THE ACCURACY OF COMPUTER AIDED DETECTION OF PERIAPICAL RADIOLUCENCIES ON CONE BEAM COMPUTED TOMOGRAPHY IMAGES USING ARTIFICIAL INTELLIGENCE: DIAGNOSTIC ACCURACY STUDY.

Yasmin A. Youssef[1] , Hossam A. Kandil[2], Ahmed M. Elsheikh[3], Noha O. Issa [4]

[1] Assistant lecturer of Oral and Maxillofacial Radiology, Faculty of Dentistry, Cairo University, Egypt.
Email: Yasmin.aboulmaaty@dentistry.cu.edu.eg

[2] Professor of Oral and Maxillofacial Radiology, Faculty of Dentistry, Cairo University, Egypt.
Email: hossam.kandil@dentistry.cu.edu.eg

[3] Associate professor of Engineering Mathematics Faculty of Engineering Cairo University, Egypt.
Email: ahmed.elsheikh@eng.cu.edu.eg

[4]Professor of Oral and Maxillofacial Radiology Faculty of Dentistry, Cairo University, Egypt.
Email: noha.issa@dentistry.cu.edu.eg

***Corresponding author:** Yasmin A. Youssef Assistant lecturer of Oral and Maxillofacial Radiology, Faculty of Dentistry, Cairo University, Egypt. Email: Yasmin.aboulmaaty@dentistry.cu.edu.eg

Received: *Nov 25. 2025;* ***Accepted:*** *Dec 25, 2025;* ***Published:*** Jan. *21, 2026*

ABSTRACT

Objectives: To assess the accuracy of a deep learning model in the automatic detection of periapical radiolucent lesions of upper and lower jaws by comparing it with experienced radiologists' opinion, which represents the ground truth.

Material and methods: CBCT scans of 90 patients were imported into **Blue sky bio** software to be cropped and annotated for detection of periapical Radiolucencies. The annotated data were sent to computer science expert to use the data in training, validation and testing to evaluate the performance of a Bayesian Convolutional Neural Network (Bayesian CNN) model for the automatic detection of periapical radiolucent lesions within maxillary and mandibular cone-beam computed tomography (CBCT) images.

Results: The cross-validated diagnostic performance of the Bayesian ensemble architectures. deep learning model, with ROC-AU values of 0.9839± 0.0139, The PR-AUC values further confirm strong performance, achieving (0.9933 ± 0.0052), achieved high accuracy (0.9441± 0.0283), balanced accuracy (0.9140 ± 0.0458), and F1-scores (0.9616 ± 0.0201), indicating excellent balance between sensitivity and specificity despite class imbalance.

Conclusion: From a clinical perspective, these results suggest that Bayesian ResNet-18 Architecture deep learning model function as balanced diagnostic tools with high sensitivity while maintaining acceptable specificity in computer aided detection of periapical Radiolucencies on CBCT images. ResNet-18 provides also showed an optimal balance between diagnostic performance and computational efficiency, this makes it highly suitable for most clinical deployment scenarios.

Keywords: Artificial intelligence, CBCT, Computer aided detection, Deep learning.

INTRODUCTION

Periapical pathosis can be seen radiographically as periapical Radiolucencies. CBCT creates high-resolution (3D) images without the distortion and superimposition of bone and dental structures seen in conventional radiographs. Cone beam computed tomography scans provide higher diagnostic accuracy compared to digital periapical radiographs in detection of periapical radiolucency[1].In the last few years, there was a revolution in application of artificial intelligence (AI) in the dental field.

Studies show that these AI-powered automated systems performed extremely well in different dental fields[2]. Computer aided diagnosis (CAD) is a non-invasive tool for diagnosis, providing medical professionals with a valuable second opinion and nowadays, many studies are investigating its role in dental application.

The development of computer-aided detection/diagnosis (CAD) systems for dental imaging

is progressing. CAD is also useful in the detection and evaluation of dental and maxillofacial lesions[3].

In recent years the application of artificial intelligence (AI) in the form of deep convolutional neural networks (CNN) have been extensively employed for developing automated tools to achieve an accurate diagnosis[4].

Therefore, this diagnostic accuracy study was designed to evaluate the performance of a Bayesian Convolutional Neural Network (Bayesian CNN) EfficientNet-B0 model for the automatic detection of periapical radiolucent lesions within maxillary and mandibular cone-beam computed tomography (CBCT) images.

MATERIAL AND METHODS

Sample selection and preparation

Sample selection

90 CBCT scans of anonymized retrospective data will be used for research, without the active involvement of patients. Different CBCT machines scans will be used and preliminarily imported into CBCT viewer software program **Blue Sky Bio** to detect the periapical radiolucent lesions.

Inclusion criteria:

- CBCT scans of maxilla and mandible with good quality free pf periapical radiolucent lesions.
- CBCT scans of maxilla and mandible with good quality showing periapical radiolucent lesions.

Exclusion criteria:

- CBCT images of sub-optimal quality or artifacts / high scatter interfering with proper assessment.

Data annotation:

CBCT scans were imported into **Blue sky bio** (open-source free software version 4.12) for data annotation**.** At first the radiologist **navigates the CBCT scans in all cuts (axial, coronal, sagittal and reformatted panorama).**

The radiologist-dependent detection of the periapical area was conducted on **axial cuts** over the apical portion of the root and subsequently identified as either present or absent, with confirmation and consensus from two Oral and Maxillofacial Radiologists (OMFR) with 8 and 15 years of experience. This classification functions as the ground truth **Figure (1).**

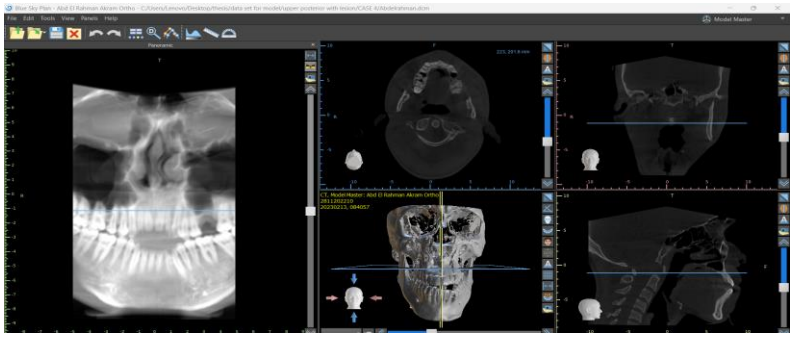

Figure 1. RAW CBCT image in the axial, coronal, and sagittal views on blue sky bio software

The most representative axial cut was selected and 2 cuts above and another two cuts below were selected for a total of 5 cuts in each scan then organized, then scans were cropped in fixed dimensions and organized for easy collaboration with the mathematical engineer for building the model as the following table (1)

Table 1.

Upper arch	**Lower arch**
Upper Anterior Normal **(UAN)**	Lower Anterior Normal **(LAN)**
Upper Anterior with Lesion **(UAL)**	Lower Anterior with Lesion **(LAL)**
Upper Posterior Normal **(UPN)**	Lower posterior Normal **(LPN)**
Upper Posterior with Lesion **(UPL)**	Lower Posterior with Lesion **(LPL)**

Image preprocessing and patch generation

For each category (UAN, UAL, etc.), rectangular axial crops centered on the periapical region were extracted and resized to fixed dimensions. Anterior regions were resized to 275 pixels (height) × 440 pixels (width), whereas posterior regions were resized to 440 pixels (height) × 275 pixels (width), preserving anatomical proportions while standardizing input size across the

Dataset **fig 2**.

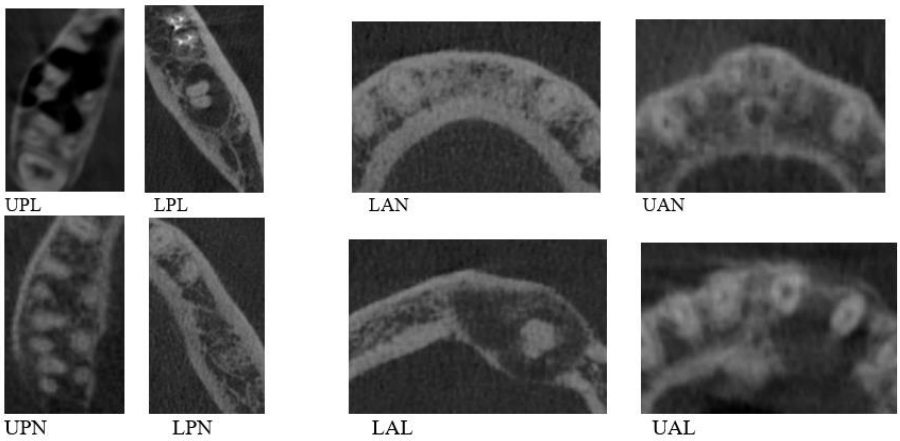

Figure 2. Showing cropped CBCT axial cuts of upper and lower anterior and posterior region

Development of the AI model:
CBCT images were organized into subfolders and preprocessed for uniform cropping and normalization. Model development was conducted in the Python environment (v3.9), utilizing PyTorch and torchvision for diversified data transformations and augmentations (center cropping, random flips, rotation, contrast adjustment) to mitigate overfitting under limited data scenarios [5].
Each ensemble member uses a ResNet-18 backbone with final drop-out layer activated both during training and inference. For each test sample, T stochastic forward passes (typically T = 30) are performed, yielding a predictive distribution rather than a point estimate, as justified by variational Bayesian inference theory[6].
Training, Validation, and Testing A five-fold cross-validation strategy was used to robustly estimate generalization performance.
Training and test splits were maintained at 80:20 per fold and results were aggregated for ensemble prediction. The Bayesian CNN model provides not only a lesion/no-lesion binary label but also a calibrated confidence score, enabling practitioners to assess the reliability of each automated decision[7]. Fig (3)

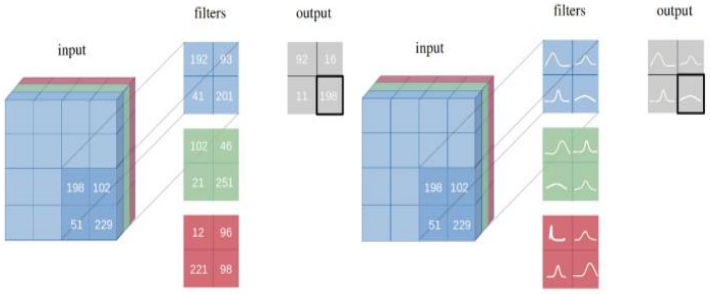

Figure 3. Comparison of conventional and Bayesian CNN architectures
Bayesian ResNet-18 Architecture was included as a mid-complexity baseline. This architecture is based on the residual learning framework, which addresses the degradation problem in deep networks through identity skip connections. The network consists of an initial stem comprising a 7 × 7 convolution with stride 2 andmax pooling, followed by four stages of residual BasicBlock modules with
channel dimensions {64, 128, 256, 512} and down sampling strides. fig (4)

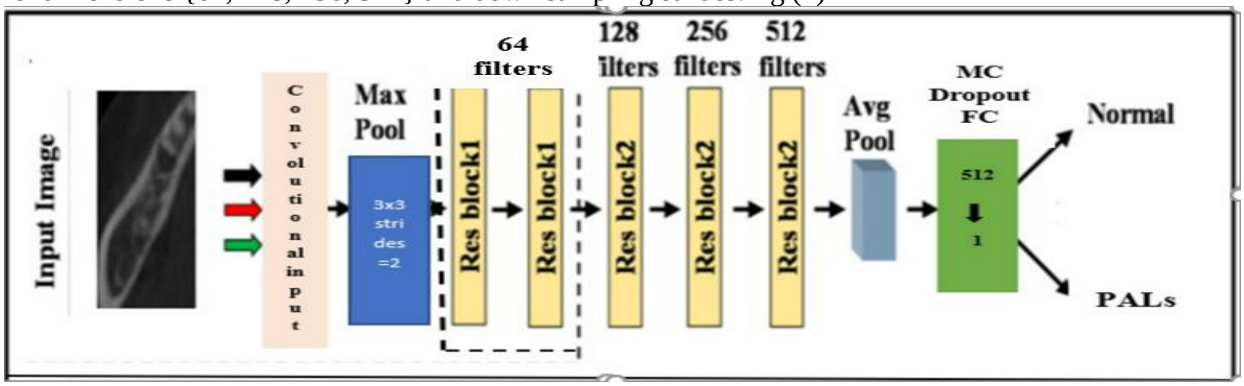

Figure 4. Bayesian ResNet-18 Architecture of Cropped CBCT image in mandibular posterior region.

Statistical analysis
Assessment metrics:
Evaluation
Assessment metrics:
Evaluation

The accuracy of the deep learning model (DLM) versus the Ground truth (GT) was evaluated by the following 2 methods:

1. periapical radiolucent lesion detection accuracy using the classifier DLM:

A dichotomized outcome of "presence of PLAs" and "absence of PLAs" of DLM was compared with GT, where the GT was determined and agreed by two maxillofacial radiologists with eight and 15 years of experience. The results of the test group were put into a confusion matrix of true positive (TP), false positive (FP) and false negative (FN) parameters. Based on this confusion matrix, *Precision*, *Recall (Sensitivity)* and *F1* score were calculated and graded according to the ranking for diagnostic tests by Leonardi Dutra et alwith scores .80% considered excellent outcomes, between 70% and 80% good, between 60% and 69% fair, and below 60% as poor (8)

- **Accuracy** evaluates the correct predictions generated by the model throughout the complete dataset. The calculation is the ratio of true positives (TP) and true negatives (TN) to the total number of samples.
- **Precision,** or positive predictive value, evaluates just true positives among all positive predictions generated by the model. It is determined by the ratio of true positives (TP) to the total of true positives and false positives (FP).
- **Recall**, also referred to as sensitivity or true positive rate, quantifies the ratio of true positive forecasts to all real positive cases, thereby accounting for missed positives. It is determined by the ratio of true positives (TP) to the total of true positives and false negatives (FN).
- **F1 Score** — The F1 Score is a statistic that equilibrates precision and recall. It is computed as the harmonic mean of precision and recall.

Accuracy assesses the overall correctness of the model's predictions, whereas precision and recall concentrate on the quality of positive and negative predictions, respectively. The F1 Score offers an appropriate ratio between precision and recall, rendering it a more balanced indicator for assessing classification models. The area under the curve (AUC) of the receiver-operating characteristic (ROC) curve was calculated as well.

RESULTS

Overall Diagnostic Performance

The cross-validated diagnostic performance of the Bayesian ensemble architectures. Bayesian ResNet-18 Architecture deep learning model, with ROC-AU values of 0.9694±0.0175, The PR-AUC values further confirm strong performance, achieving 0.9828± 0.0125, achieved high accuracy 0.9412± 0.0345 , balanced accuracy 0.9081±0.0561, and F1-scores 0.9597±0.0235, indicating excellent balance between sensitivity and specificity despite class imbalance.

Table 2. Cross-validated diagnostic performance (mean -standard deviation) of the Bayesian ensembles across 5 folds with ensemble and Monte Carlo averaging.

Metric	Bayesian ResNet-18 Architecture
ROC-AU	0.9694±0.0175
PR-AUC	0.9828±0.0125
Accuracy	0.9412±0.0345
Balanced accuracy	0.9081±0.0561
F1-score	0.9597±0.0235
MCC	0.8499±0.0963
Sensitivity	0.9836±0.0163
Specificity	0.8326±0.1011
PPV	0.9373±0.0355
NPV	0.9484±0.0630
Brier score	0.0433 ±0.0236

Confusion Matrix Analysis

EfficientNet-B0 achieving approximately 16.8 true negatives and 47.2–47.4 true positives per fold. False negatives are minimal (0.6 patches), confirming the very high sensitivity. positives are moderate (3.2 patches), resulting in acceptable specificity suitable for clinical deployment where minimizing missed lesions takes priority while avoiding excessive false alarms.

Table 3. Mean confusion matrices (average patch counts per test fold) for model.

TN	FP	FN	TP
16.8	**3.2**	**0.6**	**47.4**

TN = true negatives, FP = false positives, FN = false negatives, TP = true Positives.

Table 4 reports per-image inference time for the three ensemble architectures, measured in milliseconds per image during evaluation. These timing measurements include the complete inference pipeline: loading the image, performing T = 15 Monte Carlo dropout passes for each of Nens = 5 ensemble members, and computing the final averaged prediction.

ResNet-18 achieves a favorable performance-efficiency balance with mean inference time of 17.700. ±34 ms per image (range: 17.36–19.02 ms), maintaining comparable diagnostic performance. This efficiency advantage makes ResNet-18 particularly attractive for clinical deployment scenarios where computational resources are limited or where high-throughput batch processing is required (e.g., retrospective analysis of large CBCT databases).

Table 4. Per-image inference time statistics for ResNet-18 architectures.

Model	Mean (ms/image)	Std (ms/image)	Range (ms/image)
ResNet-18	**17.70**	**0.34**	**17.36–19.02**

DISCUSSION

In detection of periapical pathosis AI have gained much importance, specifically that used CNN and U-Net architecture. Most of the endodontic-related AI studies which are based on CBCT images, dealt with periapical pathosis detection and segmentation.

In a study done by abdolali et al image retrieval (CBIR) was used to assist clinicians in decision making by retrieving the most similar cases to a given query image from a large database. Quantitative results show that the proposed approach is more effective than previous methods in the literature, and it can facilitate the introduction of CBIR in clinical CBCT applications[5].

Complicated assignments like the classification of 2D or 3Dradiographic images lack the ability to elucidate how and why the model operates. Furthermore, in the event of an incorrect diagnosis, it becomes impossible to identify the issue and to thoroughly examine the model's functionality in a step-by-step manner[9].

For detection of periapical radiolucent lesion in dental radiographic images. Despite the growing success of deep learning models, accurate classification remains a significant challenge as medical images same as dental images often exhibit characteristics such as limited spatial resolution, subtle visual differences between disease categories (low inter-class variance), and substantial variation within the same class (high intraclass variability). This will hinder the ability of standard vision models to generalize effectively, frequently resulting in misclassification[10].

These challenges highlight the need for efficient architectures that can focus on critical spatial and contextual features for reliable performance. Therefore, in our study we used Bayesian ResNet-18, providing amid-complexity baseline with proven performance in medical imaging that achieved high accuracy (0.9441± 0.0283), balanced accuracy (0.9140 ±0.0458), sensitivity (0.9836±0.0163), specificity (0.8326±0.1011), positive predictive value (0.9373±0.0355), and negative predictive value (0.9484±0.0630). in comparison to the results of detection of periapical pathosis using CNN and U-Net architecture. CBCT images to assess lesion detection accuracy, achieving excellent ratings for sensitivity (0.93), specificity (0.88), positive predictive value, and negative predictive value (0.93)[11].

On the contrary, Orhan et al. employed a U-Net-like architecture for the detection of periapical lesions in CBCT images, utilizing binary voxel-based intersection and prediction metrics based on the true mask combination (IoU) to assess the model's efficacy in predicting lesion localization. This deep learning model identified 142 out of 153 periapical lesions, yielding a reliability of 92.8%[12].

In our study Bayesian ResNet-18 Architecture model was designed as a supervised binary classifier, categorizing cropped axial CBCT patches as either “lesion” or “normal.” To address the inherent uncertainty in medical image classification, The model incorporated Bayesian inference through MonteCarlo (MC) dropout, which provides quantitative estimates of prediction confidence alongside the classification decision[7].

Another study performed the detection of periapical lesions using the Spatial Configuration-Net based on heatmap Regression and the sensitivity and specificity values of lesion detection were 97.1% and 88.0%, respectively[13].

In other study they found that Relying on the precision metric for comparison may be misleading since, w their datasets have a well-balanced distribution that does not reflect real-world clinical scenarios. The precision metric is sensitive to class distribution, making it less suitable in this context. In terms of sensitivity and specificity, they report a higher number of false negatives and false positives. While their sensitivity and specificity results are lower than those of the closely related and best-performing work by Setzer et al. In our study ResNet-18 demonstrate confusion matrices achieving approximately 16.8 true negatives and 47.2–47.4 true positives per fold. False negatives are minimal 0.8 patches, confirming the very high sensitivity reported in. False positives are moderate (3.2 patches) resulting in acceptable specificity suitable for clinical deployment where

minimizing missed lesions takes priority while avoiding excessive false alarms. From a clinical perspective, these results suggest ResNet-18 function as balanced diagnostic tools with high sensitivity (missing few true lesions) while maintaining acceptable specificity (avoiding excessive false alarms).

In other study group of limitations that was corrected in our study First, the study population was relatively small, which may limit the generalizability of the findings. In our study CBCT scans from different machines were used to account for heterogeneous acquisition parameters, scans were rescaled to a fixed intensity range and normalized per image. Second, the AI tool used in the study was commercially available, and diagnostic performance may vary depending on the specific AI algorithm and training dataset used. Third, the study focused only on PL detection, and other dental pathologies were not considered[14].

Four AI models were evaluated in a study and compared: CBCT-SAM, CBCT-SAM without progressive Prediction Refinement Module (PPR), and two previously developed models: Modified U-Net and PAL-Net. CBCT-SAM achieved an average diagnostic accuracy of 98.92% ± 010.37% and an average[15].

Bayesian ResNet-18 Architecture deep learning model, showed higher results with ROC-AU values of 0.9694±0.0175, The PR-AUC values further confirm strong performance, achieving 0.9828± 0.0125, achieved high accuracy 0.9412± 0.0345, balanced accuracy 0.9081±0.0561, and F1-scores 0.9597±0.0235, indicating excellent balance between sensitivity and specificity despite class imbalance.

When compared to A retrospective study of CBCT scans of 134 molars. PARLs detected by Diagnocat.

Diagnostic performance was assessed at both tooth and root levels showing

demonstrated high sensitivity (teeth: 93.9 %, roots: 86.2 %), moderate specificity (teeth: 65.2 %, roots: 79.9 %), accuracy (teeth: 79.1 %, roots: 82.6 %), PPV (teeth: 71.8 %, roots: 75.8 %), NPV (teeth: 91.8 %, roots: 88.8 %), and F1 score (teeth: 81.3 %, roots: 80.7 %) for PARL detection. The AUC was 0.76 at the tooth level and 0.79 at the root level. Postoperative scans showed significantly lower PPV (teeth: 54.2 %; roots: 46.9 %) and F1 scores (teeth: 67.2 %; roots: 59.2 %)[16].

In other study images under optimal viewing conditions, achieving perfect agreement for axial images of lesion margins (0.98, $P < .001$) and lesion dimensions (length and breadth, 0.98, $P < .001$). Similarly, there was perfect agreement for coronal images of lesion height (1, $P < .001$) and sagittal images of lesion height (0.98, $P < 0.001$)[17].

Our model did not deal with artifacts from in CBCT scans. were excluded from our study; instead, the models were trained solely on cropped images to minimize computational complexity, as advised by the computer scientist, which inevitably impacted the accuracy and generalizability of the model's performance. Furthermore, the normal anatomical periapical radiolucent areas as mental foramen and incisive foramen not considered during evaluation.

CONCLUSION

From a clinical perspective, these results suggest that EfficientNet-B0 deep learning model architecture and function as balanced diagnostic tools with high sensitivity while maintaining acceptable specificity in computer aided detection of periapical Radiolucencies on CBCT images. ResNet-18 provide not only accurate classifications but also reliable confidence estimates through their well calibrated

probabilities and quantifiable uncertainties. This enables sophisticated clinical workflows such as: automatic approval of high-confidence predictions, flagging of high-uncertainty cases for expert review.

DECLARATIONS

FUNDING

This research did not funding

Competing Interests

The authors have no competing interests to declare.

Informed Consent

Not applicable.

REFERENCES

1. Venskutonis T, Daugela P, Strazdas M, Juodzbalys G. Accuracy of Digital Radiography and Cone Beam Computed Tomography on Periapical Radiolucency Detection in Endodontically Treated Teeth. J Oral Maxillofac Res [Internet]. 2014 Jul 1 [cited 2025 Dec 12];5(2). Available from: http://www.ejomr.org/JOMR/archives/2014/2/e1/v5n2e1ht.htm
2. Khanagar SB, Al-ehaideb A, Maganur PC, Vishwanathaiah S, Patil S, Baeshen HA, et al. Developments, application, and performance of artificial intelligence in dentistry – A systematic review. J Dent Sci. 2021 Jan;16(1):508–22.
3. Katsumata A, Fujita H. Progress of computer-aided detection/diagnosis (CAD) in dentistryCAD in dentistry. Jpn Dent Sci Rev. 2014 Aug;50(3):63–8.
4. Ayidh Alqahtani K, Jacobs R, Smolders A, Van Gerven A, Willems H, Shujaat S, et al. Deep convolutional neural network-based automated segmentation and classification of teeth with orthodontic brackets on cone-beam computed-tomographic images: a validation study. Eur J Orthod. 2023 Mar 31;45(2):169–74.
5. Abdolali F, Zoroofi RA, Otake Y, Sato Y. A novel image-based retrieval system for characterization of maxillofacial lesions in cone beam CT images. Int J Comput Assist Radiol Surg. 2019 May;14(5):785–96.
6. Milanés-Hermosilla D, Trujillo Codorniú R, López-Baracaldo R, Sagaró-Zamora R, Delisle-Rodriguez D,

Villarejo-Mayor JJ, et al. Monte Carlo Dropout for Uncertainty Estimation and Motor Imagery Classification. Sensors. 2021 Oct 30;21(21):7241.
7.Gal Y, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning [Internet]. arXiv; 2016 [cited 2025 Dec 13]. Available from: http://arxiv.org/abs/1506.02142
8.Leonardi Dutra K, Haas L, Porporatti AL, Flores-Mir C, Nascimento Santos J, Mezzomo LA, et al. Diagnostic Accuracy of Cone-beam Computed Tomography and Conventional Radiography on Apical Periodontitis: A Systematic Review and Meta-analysis. J Endod. 2016 Mar;42(3):356–64.
9.Sadr S, Mohammad-Rahimi H, Motamedian SR, Zahedrozegar S, Motie P, Vinayahalingam S, et al. Deep Learning for Detection of Periapical Radiolucent Lesions: A Systematic Review and Meta-analysis of Diagnostic Test Accuracy. J Endod. 2023 Mar;49(3):248-261.e3.
10. Ferdous Md, Mahmud S, Shimul MdEZ. MedNet: a lightweight attention-augmented CNN for medical image classification. Sci Rep. 2025 Nov 25;15(1):41936.
11. Setzer FC, Shi KJ, Zheng Z, Yan H, Yoon H, Mupparapu M, et al. Artificial Intelligence for the Computer-aided Detection of Periapical Lesions in Cone-beam Computed Tomographic Images. J Endod. 2020 Jul;46(7):987–93.
12. Orhan K, Bayrakdar IS, Ezhov M, Kravtsov A, Özyürek T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. Int Endod J. 2020;53(5):680–9.
13. Kirnbauer B, Hadzic A, Jakse N, Bischof H, Stern D. Automatic Detection of Periapical Osteolytic Lesions on Cone-beam Computed Tomography Using Deep Convolutional Neuronal Networks. J Endod. 2022 Nov;48(11):1434–40.
14. Kazimierczak W, Wajer R, Wajer A, Kiian V, Kloska A, Kazimierczak N, et al. Periapical Lesions in Panoramic Radiography and CBCT Imaging—Assessment of AI's Diagnostic Accuracy. J Clin Med. 2024 May 4;13(9):2709.
15. Chau KK, Zhu M, AlHadidi A, Wang C, Hung K, Wohlgemuth P, et al. A novel AI model for detecting periapical lesion on CBCT: CBCT-SAM. J Dent. 2025 Feb;153:105526.
16. Allihaibi M, Koller G, Mannocci F. Diagnostic accuracy of an artificial intelligence-based platform in detecting periapical radiolucencies on cone-beam computed tomography scans of molars. J Dent. 2025 Sep;160:105854.
17. Naidu G, Makkad RS, Khan F, Sinha P, Shrivastava S, Tripathi NP, et al. Evaluation of Artificial Intelligent Systems Based Analysis in Dental Periapical Lesions – A Radiological Study. J Pharm Bioallied Sci. 2025 Jun;17(2):1475–7.