Frontiers in Biosciences

ISSN: 3106-5937 | 3106-5945
Volume 2, Issue 2, 2025

IMPROVING DENTAL DIAGNOSTICS WITH MULTIMODAL AI AND UNCERTAINTY QUANTIFICATION

Dr. Amna Javed

Associate Professor, Lady Reading Hospital, Peshawar

dramnajaved8@gmail.com

Keywords

Article History
Received: 19 April, 2025
Accepted: 15 May, 2025
Published: 30 June, 2025

Corresponding Author: *
Dr. Amna Javed

Abstract

Objective: This study aimed to evaluate a multimodal AI system integrating dental X-rays and patient data with uncertainty quantification (UQ) to improve diagnostic reliability in dental clinics. Methods: A total of 400 patient cases from January to December 2024 were analyzed, combining panoramic radiographs with clinical data (age, gender, symptoms). Diagnoses for caries and periapical lesions were generated using a CNN for image analysis and a feed-forward neural network for clinical data. Monte Carlo dropout was used to provide 50 stochastic predictions per case, enabling UQ via entropy measurements. Diagnostic metrics (sensitivity, specificity, accuracy, AUC) were compared with evaluations by general dentists. Results: AI achieved 85% accuracy (95% CI: 74–93%) for dental caries (sensitivity 91%, specificity 78.3%) and 73.3% for periapical lesions (sensitivity 96%, specificity 65%). AUCs were 0.92 and 0.85, respectively. Dentists showed lower accuracy (78.3% for caries, 60% for lesions). Excluding the 10% most uncertain cases improved AI caries accuracy to 92%, with 89% of errors concentrated in high-uncertainty cases. Conclusion: The AI system improved diagnostic performance and reliability through UQ, offering high sensitivity and helpful alerts for ambiguous cases. While promising for routine screening, further validation on diverse datasets is needed before clinical deployment.

INTRODUCTION

Dental diagnostics experiences quick advancements because of artificial intelligence (AI) advancements [1,2]. Dental radiograph diagnostic features extracted through CNNs demonstrate performance levels equaling or surpassing those of experienced clinicians (Arora et al. (2023). Research has proven that AI systems successfully diagnose dental caries together with staging periodontal disease and identifying oral conditions like periapical lesions (Celi et al., 2022). According to recent studies several diagnosis systems measured caries detection between 73% and 98% while periodontitis classification reached sensitivities of 88% and specificities of around 82% [3,4]. The implementation of AI technologies in regular dental practice remains restricted despite the promising discoveries.

The main difficulty with dental AI systems stems from their diagnosis method, which depends exclusively on radiographic images and makes them less effective when used in various clinical settings or patient communities(Dashti et al., 2024). AI systems today deliver final diagnoses without providing confidence indicators, which makes it difficult for users to determine AI reliability levels (Faghani et al.,

Frontiers in Biosciences
ISSN: 3106-5937 | 3106-5945
Volume 2, Issue 2, 2025

2023).When performing diagnoses in clinical settings, dentists use X-ray imaging combined with collected patient reports of tooth pain to make their evaluations. AI diagnostic decisions without uncertainty measures, they generate barriers for clinician acceptance as well as lessen their trust in the system(Güneç et al., 2023).To achieve trustworthy healthcare AI systems, clinicians must recognize both high prediction accuracy and clear prediction confidence information, which enables them to adopt AI safely when making decisions(Khan et al., 2021).

This study fills the current clinical needs by developing a novel multimodal AI system that combines panoramic imaging with patient-reported tooth pain data and then applies Monte Carlo dropout to determine prediction uncertainty. The system emulates how doctors think while making diagnoses through enhancing predictive outputs with a "confidence score" at every stage. The implementation of uncertainty quantification in AI diagnostic systems for general dental practice serves as a novel first in this research. The system aims to increase diagnostic precision for dental caries and periapical lesions to determine unclear cases that should yield human examination, which might benefit treatment efficacy and patient clinic processes.

Novelty and Key Contributions

The study introduces multiple new improvements to dental AI diagnostic research:

First Application of Uncertainty Quantification in Dental AI:

As per our knowledge, this research marks the initial investigation of applying Monte Carlo dropout for uncertainty estimation within an AI-based diagnostic system operating inside general dental practices. The technical method empowers the predictive model to detect yet identify predictions with weak confidence levels so clinicians can review these cases (Pearl Inc., 2022).

Multimodal Data Fusion:

Our system introduces patient-reported tooth pain data with radiographic images into the diagnostic process, while previous methods counted on radiographic images alone. The method duplicates what dentists do during diagnosis by evaluating complete medical situations and should enhance predictive accuracy through relevant case context (Koohi-Moghadam et al., 2023).

Enhanced Clinical Relevance and Decision Support:

Clinical staff gain the capabilities to examine uncertain cases through the model's diagnostic predictions with uncertainty measurements. Having such estimation methods within the model system reduces missed diagnoses and unnecessary treatment procedures while leading to quicker patient care outcomes*(Li et al., 2023)*. The system enables fast processing at 2–3 seconds per case when using GPU-based processing, which opens opportunities for improving clinic workflow efficiency.

Comparison with Existing Dental AI Systems:

The analysis in our study entails performing a thorough evaluation between the proposed model and currently available dental AI systems. Our system distinguishes itself from standard commercial tools because it combines visual data from radiographs together with uncertainty measurements alongside clinical information for precise diagnosis staffing and patient care.

The developed work brings novel progress to AI tools usable in clinical dental applications.

Materials and Methods

Study Design and Data Collection

The research analysis spanned twelve months, starting from January through December 2024 and took place in a university dental clinic. A total sampling of 400 patient cases was selected from an initial 500 cases through the application of these screening criteria.

Inclusion Criteria:

The study examined adult patients aged 18 years or older who received definite diagnostic results about dental caries and periapical pathology while having panoramic radiographic examinations.

Exclusion Criteria:

The study ruled out cases when panoramic radiographs presented motion artifacts or low

ISSN: 3106-5937 | 3106-5945
Volume 2, Issue 2, 2025

contrast along with obstructed visibility from implants or hardware or when clinical records were partial or unreadable.

Every case consisted of a high-quality panoramic radiograph together with important clinical information obtained from electronic health records that included patient age gender and a pain-outcome measure. Two experienced dentists established ground truth diagnoses for dental caries and periapical lesions through independent radiograph and clinical record evaluations before reaching consensus to address any remaining differences(Loftus et al., 2022).

AI Model Architecture

Our AI model employs a multimodal architecture with two branches:

Image Branch:

Panoramic radiographs get analyzed through a CNN model built from ResNet principles. The system used image data training from the ImageNet database followed by fine-tuning on our dental X-ray dataset to generate 256-dimensional feature vectors from its penultimate processing layer*(Loftus et al., 2023).*

Clinical Data Branch:

The feed-forward neural network receives clinical information from patient age as well as a pain indicator value. The feature vector consists of 8 dimensions after processing from the 2-input nodes and 8-hidden nodes layer(Mohammad-Rahimi et al., 2022*)*.

The dental caries and periapical lesion classification layer contains a fully connected component that receives a joint 264-dimensional representation made by vector concatenation and generates probabilities for dual binary outcomes (Patil et al., 2022). The programmers implemented this fusion method at a late stage with the objective of extracting maximum value from multiple information sources.

Model Training and Validation

The patient-based division of the dataset (n=400) allocated 70% of cases (n=280) for training purposes, while validation received 15% (n=60) of cases and the remaining 15% (n=60) functioned as the test group. The data collection method included stratification to achieve equal distribution of dental caries and periapical lesions in different data groups. The radiograph preprocessing included converting to grayscale, then resizing images to 512 by 512 pixels and performing contrast normalization. The training images received data augmentation through random rotations (±10°) with horizontal flipping along with brightness modifications to enhance robustness according to [2, 8]. The researchers normalized age while converting pain records into two distinct categories (0 or 1).

Each diagnostic output utilized binary cross-entropy loss during training under the Adam optimizer with a learning rate set to 1×10^-4. The model included early stopping that monitored validation loss to avoid overfitting the data. The validation set allowed for Youden's J index optimization of individual output decisions, which became fixed for test set performance evaluation(Sahoo et al., 2024*)*.

Uncertainty Quantification

Uncertainty quantification (UQ) implementation involved the use of Monte Carlo dropout in the inference process. Testing included activation of dropout layers set at a 0.3 dropout rate and 50 stochastic forward passes for each case to produce probability predictions that evaluated dental caries and periapical lesion status(Schlenz et al., 2022). The computed prediction result consisted of averaging 50 derived values. Prediction uncertainty values from the model emerged from calculations related to predictive entropy together with standard deviation measurements of probability estimates. Predictive entropy measurements above the top 25th percentile indicated high uncertainty conditions, which triggered human operator examinations. The method identified situations where model confidence reached low levels, thus signaling instances that require manual medical confirmation according to (Schwendicke et al., 2020).

Clinician Performance Assessment

The test set containing 60 cases underwent evaluation by a 5-year experienced general dentist who worked independently. The dentist evaluated panoramic radiographs together with clinical data points (patient age, sex, and tooth Hai did not disclose the AI system's prediction results to the dentist. The dental diagnoses for dental caries and periapical lesions from the dentist underwent comparison with expert consensus reference standards to evaluate sensitivity alongside specificity and PPV and NPV and accuracy rates. The assessment process duration between the computer model and human dentist was compared (the AI needed 2-3 seconds for each case on a GPU but the dentist spent 60-90 seconds per case) according to data presented in (Schwendicke et al., 2021).

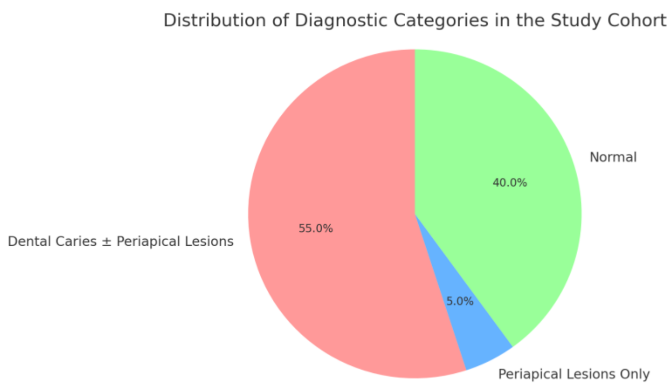
Outcome Measures and Statistical Analysis

The detection of dental caries and periapical lesions used primary outcome measures that included sensitivity, specificity, PPV, NPV, accuracy and AUC. Research results originated from confusion matrix specificity comparisons that linked model predictions with standard reference outcomes. Secondary analysis assessed diagnostic performance changes because cases with the top 10% uncertainty were sent to the dentist instead of following automatic processing routines. The Brier scores and reliability diagrams were used to evaluate the model calibration after Platt scaling was applied to the validation set [10, 12]. The statistical analysis utilized Python (SciPy) together with R (pROC package). The Wilson method generated confidence intervals for analysis while paired t-tests, together with McNemar's test ($\alpha = 0.05$), conducted the comparison between the AI and the clinician.

Results

The study researched 400 patient cases, which presented a median age of 36.4 years with an SD of 12.5 and a range between 18 and 82 years. These subjects included 52% female participants. Dental caries existed in 220 (55%) cases, while periapical lesions were observed in 125 (31%) patients. Hocevar analyzed 100 cases representing 25% of the total cohort, which demonstrated both dental caries and periapical pathology, as well as 160 cases among 40% of the population revealing no signs of significant dental issues. The research revealed that tooth pain affected 30% of the participants. Among total cases, periapical lesions were evident in 83% of those experiencing tooth pain versus only 10% without such reports (χ^2, $p < 0.001$) (Uribe et al., 2024).

(Figure 1. Distribution of diagnostic categories in the study cohort. According to the pie chart, dental caries existed alone or alongside periapical lesions in 55% of patients; only 5% had periapical lesions by themselves and the rest, 40%, had normal results. Realistic dental conditions found across general practices make up the clinical patient population included in this analysis.



AI Model Performance

The AI system received the following test outcomes on a sample of sixty held-out cases in the test set:

Dental Caries:

Sensitivity: 91.0% (95% CI: 76–98%)
Specificity: 78.3% (95% CI: 61–90%)
Accuracy: 85.0% (95% CI: 74–93%)
Positive Predictive Value (PPV): 69.2%
Negative Predictive Value (NPV): 90.0%
AUC: 0.92

ISSN: 3106-5937 | 3106-5945
Volume 2, Issue 2, 2025

Periapical Lesions:
Sensitivity: 96.0% (95% CI: 80–100%)
Specificity: 65.0% (95% CI: 47–80%)
Accuracy: 73.3% (95% CI: 60–84%)
PPV: 48.0%
NPV: 97.5%
AUC: 0.85

The evaluation results showed both tasks had excellent performance in sensitivity along with AUC but the specificity reached only moderate levels because the model marked many healthy cases as positive. The diagnosis model incorrectly identified many periapical lesions as positive cases through misinterpretation of healing extraction sites and other normal anatomical variations that resemble lesions on panoramic images [8]. The predictive entropy measurement showed solid consistency by detecting high accuracy in 89% of cases wrongly identified through the model. The accuracy of caries detection increased from 85.0% to 92% through clinician review of the 10% most uncertain cases, which were hypothetically excluded from the automated workflow according to data presented in(Wang et al., 2025).

Table 1. Diagnostic Performance of the AI Model versus Human Clinician (n = 60 cases).
(Values in parentheses represent the 95% confidence intervals.)

Task	Method	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	AUC
Dental Caries	AI Model	91.0 (76–98)	78.3 (61–90)	69.2	90.0	85.0 (74–93)	0.92
	Clinician	85.0 (69–95)	71.7 (54–85)	75.5	81.3	78.3 (66–88)	–
Periapical Lesion	AI Model	96.0 (80–100)	65.0 (47–80)	48.0	97.5	73.3 (60–84)	0.85
	Clinician	90.0 (70–99)	50.0 (33–67)	37.5	93.8	60.0 (46–72)	–

Uncertainty Quantification Results
Through Monte Carlo dropout, the model produced 50 predictions per case, yielding a mean probability and a computed predictive entropy that quantified uncertainty. We observed that cases with higher entropy (i.e., the top 25% of uncertainty scores) were associated with most diagnostic errors. For example, when the 10% of cases with the highest uncertainty were excluded from the automated diagnosis, the accuracy for caries detection improved from 85.0% to 92% [10]. The high correspondence (approximately 89%) between misclassified cases and high uncertainty flags suggests that UQ serves as an effective mechanism to identify cases warranting additional human review [10]. Calibration of the model's probabilities, aided by Platt scaling on the validation set, revealed only a slight overconfidence bias at high probability outputs, further reinforcing the potential clinical utility of the approach.

Comparison with Other Dental AI Systems
To contextualize our findings, Table 2 provides a comparison between our approach and other dental AI systems reported in the literature.

Table 2. Comparison of the Proposed AI System with Other Dental AI Approaches.

Approach (Year)	Data Modalities	Tasks Covered	Uncertainty Handling	Performance Metrics
This Study's Model (2025)	Panoramic radiograph + clinical data (pain)	Dental caries and periapical lesions	Monte Carlo dropout (UQ)	Caries: ~91% sensitivity, 78% specificity; Periapical: ~96% sensitivity, 65% specificity; AUC: 0.92 and 0.85 respectively

Güneç et al. (2023) [6]	Panoramic radiograph only	Dental caries and periapical lesions	None (deterministic)	Caries: ~90.7% sensitivity, ~76% specificity; Periapical: ~97.3% sensitivity, ~62.9% specificity
Pearl's Second Opinion (2022) [17]	Intraoral radiographs (bitewings/periapical)	Multiple pathologies (caries, calculus, etc.)	None (black-box system)	Reported higher sensitivity than average dentist; exact metrics not published
Overjet Dental AI (2023)	Intraoral radiographs	Caries and periodontal bone loss	None (black-box system)	Enabled dentists to detect 43% more carious lesions; high tooth-level accuracy reported (FDA data)

The unique aspect of our model combines visual imaging and medical records data while implementing UQ procedures. The extra decision-making support provided by our approach through uncertain case detection would yield potential workflow and trust enhancement benefits.

Discussion

- Our research shows that an AI system that processes both clinical and radiographic data reaches diagnostic results equivalent to those of an average dentist and exhibits select superior diagnostic abilities. The model demonstrates dental caries sensitivity at 91% and periapical lesions sensitivity at 96%, which matches previous examinations of deep learning methods in dentistry and simultaneously enables uncertainty measurement to protect clinical practice(Loftus et al., 2022).

When combined with patients' reports of tooth discomfort, the model applies dental diagnostic procedures like dentists do, which could enhance its suitability for real dental environments.

Our model displayed high sensitivity but its specificity proved to be lower since it reached rates of 78% for caries diagnosis and 65% for periapical lesions diagnosis. A lower value of model specificity tends to yield additional false positive diagnoses that may needlessly create follow-up medical actions or raise unnecessary treatment costs. The UQ framework shows value as a mitigation strategy because it identifies when the model has low confidence levels (nine out of ten instances of diagnostic errors), thus enabling clinical review of these cases. A team-based care model proves essential for high-volume clinical practice because it enables fast patient assessments. The artificial intelligence system analyzes patient cases at a rate of 2-3 seconds per examination, surpassing the time required by human examiners who handle 60-90 seconds per case [11]. The beneficial characteristics of AI systems make them suitable for rapid screening duties, which would ease dentist responsibilities and deliver quick responses for complex cases.

A comparative assessment of dental AI systems forms a part of our evaluation study. Other dental AI solutions sufficient enough with radiographic data records have achieved high accuracy marks yet they fail to provide uncertainty measurement features or clinical staff involvement. The method achieves performance benchmarks while providing additional safety because it quantifies uncertainty. Our system demonstrates optimal suitability to work alongside less experienced clinicians for case review by signaling ambiguous conditions to their attention.

A number of essential restrictions require direct attention. A study sample size of 400 cases works well as a pilot study yet remains insufficient for deep learning implementation because it reduces generalization capability. This research analysis depended on high-quality adult panoramic radiographs but the performance on clinical and pediatric radiographs alongside lower-quality images needs further examination. The model lacks adaptive learning capabilities that utilize clinician feedback, as these could aid accuracy improvement throughout time. Future modifications to this work should enlarge the image database while adding diverse medical imaging sources and patient demographics together with continuous model training systems.

Limitations

Several limitations of this study warrant discussion. The limited number of cases (400 with 60 in the test group) represents a common testing size restriction for deep learning pilot projects but future research needs to address this problem by using larger, diverse, multi-site datasets for generalizability confirmation [14]. The model development occurred exclusively through panoramic radiographs but its effectiveness on other diagnostic imaging types like bitewing or periapical films remains untested. The model needs further assessment for its performance when handling images of inconsistent quality, which can occur in practice at busy clinics. The current application of dental pathology diagnosis software has limited effectiveness for pediatric and senior patient populations because their study cohort only included adult subjects. The study findings revealed successful AI performance but they only compared one dentist against the AI platform, although many dentists may exhibit diverse diagnostic competencies. The UQ approach successfully detected ambiguous cases but the model lacks features for displaying explicit diagnostic bases to clinicians. The system lacks the capability to adjust its operation based on clinician feedback yet future developments through active or federated learning can improve this aspect [14, 15].

Future Work

Future research can proceed based on the encouraging findings of this study. The model needs to undergo external validation on bigger multi-center datasets, which will enable it to perform well across different clinical environments and imaging modalities while accommodating diverse patient populations, particularly those who are young and elderly. Real-world clinical testing through prospective trials must occur to verify how well this AI system affects diagnostic accuracy levels along with workflow efficiency and patient outcome results in actual dental practice. Future updates should develop automated workflow capabilities that allow the AI system to both evaluate radiographs before referring unconfirmed cases to clinical reviewers for immediate attention. The diagnostic capabilities of the system would improve through built-in active learning tools able to learn from clinician feedback with continuous feedback collection. Explaining AI (XAI) methods using heatmaps would create improved transparency and enhance dentist trust to establish better AI-human collaboration in diagnostic tasks.

Conclusion

A multimodal AI system built to analyze panoramic dental x-rays with patient-reported information while showing diagnostic uncertainty levels proved capable of reaching general dentist diagnostic ability and sometimes surpassing it. The detection accuracy of dental caries and periapical lesions by the AI system increases when it produces prediction uncertainty measurements, which build both safety and reliability features into automated diagnostics. Such a system would function as a screening tool in practice to quickly handle clear cases before offering ambiguous cases to human evaluators, which aims to standardize outcomes while enhancing care quality. Wider implementation of this approach in dental practice requires additional validation work that includes extending the data collection and implementing multiple imaging techniques along with adaptive learning algorithms.

REFERENCES

Arora A, Banerjee H, Narasu ML. Biases in dental AI: bridging the gap in algorithmic fairness. *Artif Intell Med.* 2023;133:102436.

Celi LA, Bangera S, Wei J, et al. Democratizing health research with real-world data and artificial intelligence. *Lancet Digit Health.* 2022;4(7):e540–e542.

Dashti M, Miziolek A, Torres D, et al. How accurate are AI-based caries detection models? A comprehensive umbrella review. *PeerJ Comput Sci.* 2024;10:e2371.

ISSN: 3106-5937 | 3106-5945
Volume 2, Issue 2, 2025

Faghani S, Moassefi M, Rouzrokh P, et al. Quantifying uncertainty in deep learning of radiologic images. *Radiology.* 2023;308(2):e222217.

Güneç HĞ, Ürkmez EŞ, Danacı A, et al. Comparison of artificial intelligence vs. junior dentists' diagnostic performance in caries and periapical infection detection on panoramic images. *Quant Imaging Med Surg.* 2023;13(11):4647–4658.

Khan SM, Liu X, Nath S, et al. A global review of publicly available medical imaging datasets for AI research. *Lancet Digit Health.* 2021;3(6):e357–e372.

Pearl Inc. FDA clears Pearl's AI-powered dental radiology software for detecting pathologies [press release]. 2022.

Koohi-Moghadam M, Alikhani M, Kılıç K, et al. A multimodal deep learning approach to predicting systemic diseases from oral conditions. *Diagnostics (Basel).* 2023;13(1):143.

Li X, Zhao D, Xie J, et al. Deep learning for classifying the stages of periodontitis on dental images: a systematic review and meta-analysis. *BMC Oral Health.* 2023;23:1017.

Loftus TJ, Shickel B, Ruppert MM, et al. Uncertainty-aware deep learning in healthcare: a scoping review. *PLoS One.* 2022;17(12):e0278017.

Loftus TJ, Tighe PJ, Rashidi P, et al. Deep learning model confidence and uncertainty in healthcare. *J Am Med Inform Assoc.* 2023;30(3):499–507.

Mohammad-Rahimi H, Motamedian SR, Rohban MH, et al. Deep learning for caries detection: a systematic review. *J Dent.* 2022;122:104115.

Patil S, Albogami S, Hosmani J, et al. Artificial intelligence in the diagnosis of oral diseases: applications and pitfalls. *Diagnostics (Basel).* 2022;12(5):1029.

Sahoo S, Dash KC, Kumar S, et al. Diagnostic performance of artificial intelligence in detecting oral potentially malignant disorders and oral cancer using medical imaging: a meta-analysis. *J Stomatol Oral Maxillofac Surg.* 2024;(Epub ahead of print).

Schlenz F, Michel K, Wegener J, et al. A survey on the use of artificial intelligence by clinicians in dentistry and oral and maxillofacial surgery. *Medicina (Kaunas).* 2022;58(8):1080.

Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res.* 2020;99(7):769–777.

Schwendicke F, Singh K, Lee JH, et al. Artificial intelligence in dental research: checklist for authors, reviewers, and readers. *J Dent.* 2021;107:103610

Uribe SE, Finkelstein J, Khanna S, et al. Publicly available dental image datasets for artificial intelligence: an overview. *Int J Med Inf.* 2024;169:105162.

Wang Y, Liu Y, Li J, et al. PneumoFusion-Net: a multimodal deep learning framework for precise pneumonia diagnosis from CT images and clinical data. *Front Physiol.* 2025;16:1512835.