BULLETIN OF STOMATOLOGY AND MAXILLOFACIAL SURGERY
Volume 22, N 2

DOI:10.58240/1829006X-2026.22.2-18

ORIGINAL RESEARCH

EVALUATION OF AN ARTIFICIAL INTELLIGENCE MODEL FOR THE DETECTION OF DENTAL ABSCESS ON ORTHOPANTOMOGRAMS

Hebah Al badani[1], Sam Abd Alkareem Da'er[2], Hisham Haider[3], Mohanad Odhah[4], Haifaa Al hussini[5]

[1] Master's Candidate in Oral and Maxillofacial Surgery, Faculty of Dental College- Sana'a University, Sana'a, Yemen hebahalbadani1@gmail.com

[2] Associated professor of Oral and Maxillofacial Surgery Faculty of Dental College - Sana'a University, Sana'a, Yemen samdaer@yahoo.com

[3]Associated professor of Artificial Intelligence, Faculty of Computer Science College - Amran University and Al-Razi University, Sana'a, Yemen. hesham_haider@yahoo.com

[4] Engineer of Artificial Intelligence, Faculty of Computer Science College- Al-Razi University, Sana'a, Yemen awdamohanned@gmail.com

[5] Master's Candidate of Oral and Maxillofacial Surgery, Faculty of Dental College- Sana'a University, Sana'a, Yemen haifaamoh14@gmail.com

***Corresponding author:** Master's Candidate in Oral and Maxillofacial Surgery, Faculty of Dental College- Sana'a University, Sana'a, Yemen, hebahalbadani1@gmail.com

***Received:** Feb 3 2026; **Accepted:** Mar 3;2026; **Published:** Mar 5,2026*

Abstract

Background: Dental X-rays are fundamental for diagnosing intra-bony lesions, such as dental abscesses. Recently, deep learning techniques have shown promise in enhancing radiographic interpretation accuracy and efficiency.

Objective: This study aimed to develop and evaluate a deep learning model for the automated detection of dental abscesses on orthopantomograms.

Materials and Methods: A classification model based on the EfficientNetB3 architecture was developed. The model was trained and validated on a unique dataset of 238 confirmed abscess cases, which was expanded to 714 images using a multilevel data augmentation strategy to improve generalization. The performance of the model was then evaluated on a separate, independent test set of 40 cases.

Results: The EfficientNetB3 model achieved 96.40% validation accuracy on the augmented dataset. The model demonstrated 83.3% recall (95% CI: 68.6%–93.0%) and 53.0% precision (95% CI: 38.5%–67.1%) for abscess detection on the independent test set. This corresponds to a true positive rate of 87.5% and a false negative rate of 12.5%.

Conclusion: This study demonstrates that employing a specialized model for the classification task constitutes an effective methodological approach for clearly the Detection of abscesses. Nevertheless, the model's performance is substantially constrained by the size and class distribution of the training dataset, thereby reaffirming that access to large-scale, well-balanced data is fundamental for the development of robust and reliable clinical predictive models.

Keywords: Artificial Intelligence (AI), Deep Learning, EfficientNetB3, Dental X-ray, Periapical Abscess, Classification.

INTRODUCTION

An abscess is a collection of pus that forms within tissues, commonly resulting from a bacterial infection. Dental abscesses are a frequent and serious concern in dentistry, often arising from an infection that spreads from a tooth to the surrounding tissues [1].

It usually occurs as a result of dental cavities, trauma, extensive fillings, or unsuccessful root canal therapy. Upon breaching the intact pulp chamber, the root canals' colonization involves a varied array of bacterial agents. These microorganisms can form biofilms in root canals, thereby making the "biofilm concept" plausible in such infections[2].

Acute dental abscesses account for approximately 60% of all non-traumatic dental emergencies [3]. If left untreated, it can lead to significant complications and even mortality [4-6]. Dissemination of pathogens from endodontic abscesses to nearby tissues may result in fascial plane infections [7].

An accurate diagnosis is essential for optimal dental outcomes. This process relies on the professional experience of the clinician and diagnostic tools that provide critical information. Orthopantomograms

(OPGs) are widely accepted as a diagnostic tool due to their ability to capture all orofacial structures in a single image. This modality provides both convenience and reduced professional experience of the clinician. Identifying lesions from OPGs caused by dental diseases and systemic conditions helps doctors plan treatment and refer patients for specialist care [8].

Artificial Intelligence (AI) technology has its origins in the 1900s, with the term being introduced by McCarthy et al. in 1950. AI is short for "intelligent machines that can think and act like people.[9,10]. AI learns from what people know and performs tasks that normally require human intelligence [11].

Artificial intelligence (AI) has been integrated into the healthcare field in recent years, changing how patients are diagnosed and treated. AI shows promise in helping with morphological diagnostics, which could improve the quality of diagnostic processes [12]. AI uses datasets, mathematical algorithms, and statistical models. In the field of dental pathologies, these datasets can contain visual elements comprising radiographs, text, and audio data [13].

The EfficientNetB3 architecture, introduced by Tan and Le in 2019, addresses a key challenge in deep learning: scaling CNNs for better accuracy and efficiency. Instead of arbitrarily increasing the network depth, width, or resolution, EfficientNet proposes a principled compound scaling strategy. It uses a neural architecture search (NAS) to find a baseline network and then uniformly scales it, achieving an excellent balance between accuracy, computational cost, and inference speed, making it highly suitable for practical applications such as medical image classification [14].

Kumar et al. recently examined 172 OPGs using a support vector machine (SVM) classifier that relied on handcrafted features. Although this approach showed some success, it was limited by the modest dataset size and its reliance on manual feature engineering [15]. Esmaeilyfard et al. (2025) employed a CNN to analyze 150 CBCT images. Although CBCT offers higher three-dimensional resolution, the relatively small sample size limited the generalizability of its results [16]. These studies highlight a persistent challenge in the field: the trade-off between dataset size, image quality, and methodological sophistication. Data scarcity has constrained many studies, which can compromise the robustness and clinical applicability of the resulting models.

This study aimed to improve the diagnostic accuracy of dental abscess lesions by processing panoramic images. This study evaluates the performance of an EfficientNetB3 model, enhanced by a multilevel data augmentation strategy, on a unique dataset of orthopantomograms (OPGs). We hypothesize that our model can achieve improved diagnostic performance while addressing some of the limitations inherent in manual, feature-based methods by using a fully automated deep learning approach on a larger dataset than many prior studies.

2. MATERIALS AND METHODS

2.1. Study Design and Ethical Considerations

This study employed a retrospective, multicenter, diagnostic accuracy design. All cases meeting the criteria based on panoramic radiographs were included in the study and were archived in the radiographic centers in Sanaa, Yemen. The study was conducted in accordance with the Declaration of Helsinki and received ethical approval from the Medical Ethical Committee of the Faculty of Dentistry, Sana'a University (OMFS:13/06/2024). The dataset comprised approximately 503 radiographic images presenting osseous lesions, which were filtered to isolate abscess lesions, resulting in 278 images. These images were randomly divided into 238 cases for model training and 40 images for model testing.

2.2. Selection of OPG images and data collection

The inclusion criteria for this study were the presence of a clearly depicted panoramic image and a complete patient file with basic demographic information. The exclusion criteria included radiographs of poor diagnostic quality, such as those with significantly pronounced overlapping teeth, diffuse or distorted images, or significant patient motion artifacts.

From the initial cohort, 503 radiographs were identified to contain osseous lesions. These findings were then reviewed by a board-certified oral and maxillofacial surgeon with over 10 years of experience to establish the reference standard diagnosis. This process resulted in the selection of 278 cases confirmed as dental abscesses. The remaining 225 cases consisted of non-abscess lesions. The initial dataset of 238 cases was partitioned into a training set (80%) and a validation set (20%). This 80/20 split was strategically chosen to provide a sufficiently representative sample for achieving stable performance metrics during the EfficientNetB3 architecture fine-tuning. An independent test set of 40 cases was subsequently utilized to rigorously evaluate the model's generalizability. These cases were entirely held out during the training and validation phases; the model had no prior exposure to these images. Furthermore, these 40 cases remained in their raw format and did not undergo any data augmentation to ensure the integrity of the testing process and reflect real-world clinical performance.

2.3. Methods of Analysis:

Methodological Framework:

A framework based on the creation of distinct and specialized models, each functioning in parallel, was chosen to handle the diagnostic problem's multifaceted nature. The most suitable architecture can be selected for each challenge.

Abscess Classification using EfficientNetB3: EfficientNet, introduced by Tan and Le in 2019, addresses a key challenge in deep learning by scaling Convolutional Neural Networks (CNNs) for superior accuracy and efficiency. Instead of heuristically increasing the network depth layers, channels, or input resolution, EfficientNet proposes a principled compound scaling method. It is a neural architecture search (NAS) to find an optimal baseline network (EfficientNetB0) and then uniformly scales depth, width, and resolution using a fixed set of scaling coefficients determined by a grid search. This coordinated scaling facilitates parameter efficiency and computational performance compared to scaling these variables independently. The EfficientNet series (B0-B7) offers a range of model parameter scales and classification accuracy. EfficientNetB3 strikes an excellent balance between accuracy, computational cost, and inference speed, making it highly suitable for practical applications such as medical image classification.

Classification:

The EfficientNetB3 architecture was chosen to address the clinical question of whether an abscess is present or not because of its proven ability to extract fine details from medical images. For this binary task, we utilized transfer learning to leverage pre-trained knowledge, modifying the final layers of the model to focus on classifying images as either 'Abscess' or 'Non-abscess, as shown in **Figure**

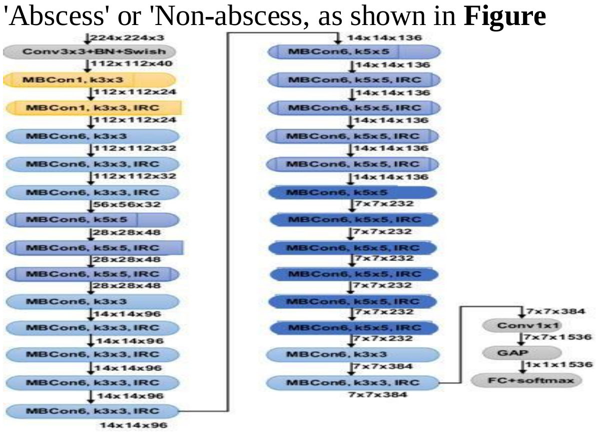

Figure 1. Diagram illustrating the EfficientNetB classification model architecture, showing the frozen and trainable parts and the added layers.

EfficientNetB3 excels at distinguishing between different types of maxillofacial lesions based on their radiographic appearance, as shown in **Figure 2.**

Figure 2. User interface of the AI-based "X-ray Analyzer" system demonstrating binary classification. The image shows a panoramic radiograph (OPG) being analyzed, with the final diagnostic output displayed as "Result: abscess." This illustrates the model's application in real-time clinical decision support for detecting periapical lesions.

Innovative Data Augmentation Strategy:

A targeted and multi-level data augmentation strategy was designed to combat the overfitting problem, which was evident in preliminary experiments. This sequential process was engineered to simulate different generations of realistic imaging conditions, which was critical in improving the generalizability of the model. The dataset was expanded from 238 to 714 through data augmentation. This augmentation preserved the original class balance while substantially increasing the dataset size, thereby enhancing the training process's robustness.

3. RESULTS

The experimental findings are based on the evaluation of the proposed deep learning models. The results are structured around a principal task, which is the classification of 'Abscess' or 'Non-abscess' using EfficientNetB3. By presenting the task in detail. The use of spreadsheet software (e.g., Microsoft Excel) has enhanced data processing and organization efficiency. Subsequently to a careful screening process, the final study cohort consisted of 278 patients with confirmed dental abscesses; 161 (57.9%) were male and 117 (42.1%). The training set included 238 patients, and the independent test set comprised 40 patients.

The distribution according to the OPG is shown in **Table 1**. Due to data augmentation, the dataset grew to 714 cases.

This addition maintained the original class balance while significantly increasing the dataset size, enhancing the training process's robustness.

Table 1. Distribution of OPG study sample (n=278)

Variables	Train	Test	Total	%
Abscess	238	40	278	100
Male	138	23	161	57.9
Female	100	17	117	42.1

3.1Model Performance

During the training phase, the accuracy of the model on the augmented validation set increased from an initial 65.48% to a final stable accuracy of 96.40%, as shown in Figures 3; A, B. The training accuracy reached 99.8%, while the validation accuracy plateaued, indicating that the model was well-fitted to the training data without significant overfitting.

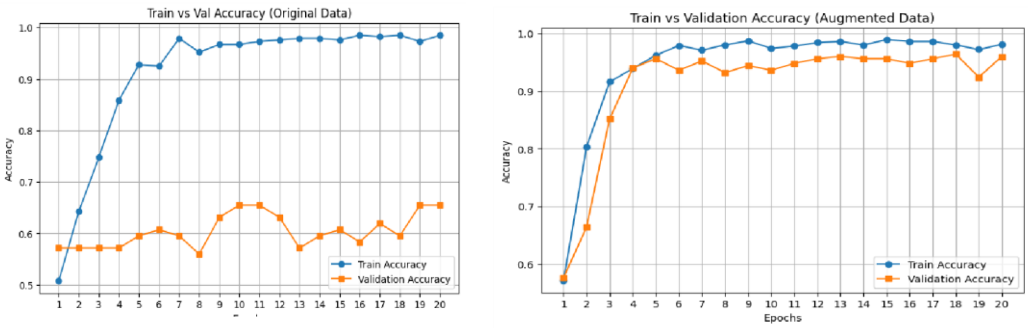

Figure 3. Training and validation accuracy curves. (A) Before data augmentation, showing unstable convergence due to limited data. (B) After data augmentation, showing a stable increase in validation accuracy to 96.40% and training accuracy to 99.8%, indicating a well-fitted model.

3.2 Test of the Classification model using EfficientNetB3:

We evaluated the model's performance on the independent test set of 40 cases. The comprehensive performance metrics for the test set are detailed in Table 2. The model achieved sensitivity (recall) of 87.5% and specificity of 61.3%. The precision (positive predictive value) was 53.0%. The greatest success in abscess classification was probably brought about by the abundance of abscess samples in the training dataset, which enhances model learning. **Table 2** presents the confusion matrix for the test set.

Table 2. Confusion Matrix for the Test Set (n=40)

	Predicted: Abscess		Non-Abscess		Actual Total		Recall
	N	%	N	%	N	%	
Actual: Abscess	35(TP)	87.5	5(FN)	12.5	40	44.4	83,3%
Precision	53%			75%	-		-

Note: The "(TP)" IS = true positives
"(FN)" IS = false negatives

Hebah Al badani, Sam Abd Alkareem Da'er, Hisham Haider, Mohanad Odhah, Haifaa Al hussini. Evaluation of an Artificial Intelligence Model for the Detection of Dental Abscess on Orthopantomograms.Bulletin of Stomatology and Maxillofacial Surgery.2026;22(2)18-24 doi:10.58240/1829006X-2026.22.2-18

4. DISCUSSION

The diagnosis of osseous lesions, such as abscesses, involving the maxilla and mandible, was the focus of this study. Before performing oral biopsies, the diagnostic procedure was performed using invasive methods. To facilitate and expedite clinical decision-making, the goal was to provide an approximate probability percentage that indicated the likelihood that a lesion belonged to a particular pathological class.

An AI model was trained to identify abscess-type lesions using OPG images that two specialists confirmed. This limitation aligns with the findings of Kown et al. 2020), who reported that cysts and tumors were ultimately classified based on histopathological diagnoses from biopsies, which also required confirmation by specialists [17].

Our study builds upon previous work by employing a fully automated deep learning pipeline on a larger initial dataset than several prior studies [15,16]. By using the EfficientNetB3 architecture with transfer learning, we leveraged a powerful, pre-trained model, which is a significant advantage over methods relying on manual, handcrafted features[15]. Furthermore, the use of a multi-level data augmentation strategy facilitated the expansion of our training set, a common and necessary technique to improve model generalization when working with limited medical imaging data. The clear separation of an independent test set, while small, is a methodological strength that provides a more estimative assessment of performance than relying on validation data alone.

In contrast to previous research, the current study's implementation utilizes a larger dataset and a more advanced architecture, which are notable improvements. For instance, Kumar et al. (2023) used a smaller dataset and a traditional SVM approach, which is less powerful than modern deep learning models [15]. While Esmaeilyfard et al (2025) used a CNN, their sample size was also smaller [16]. Although our findings align with the broader literature in demonstrating the potential of AI in dental radiology, they also serve as a crucial reminder that high accuracy on a validation set does not always translate to clinically acceptable performance on a real-world test set.

The present study has similarities with other retrospective studies by Kumar in 2023 [15] Regarding the type of radiograph and the primary objective of using AI, both studies employ OPG and classify odontogenic lesions, including dental cysts, abscesses, and tumors. Kumar [15] analyzed 172 OPG images through a support vector machine (SVM) classifier, accomplishing suitable sensitivity and specificity using handcrafted feature extraction techniques such as GLCM, GLRLM, and wavelet analysis. However, this approach was constrained by the modest dataset and by its reliance on manual features, which confines the ability to capture more complex radiographic patterns. In contrast, the present study utilized a substantially larger dataset of 503 OPG osseous lesion images, including 238 abscesses (57.6%) before augmentation. This broader and more illustrative distribution reflects real-world clinical practice, thereby enhancing AI-founded radiographic lesion classification. This study evaluated the classification model's effectiveness in evaluating odontogenic lesions, particularly abscesses. The model exhibited an 87.5% true positive rate for detecting abscesses, resulting in a 12.5% false negative rate. These results demonstrate the efficacy of the model in identifying the types of lesions present.

Kumar, Pradeep R Kumar, et al. (2023) [15] utilized GLCM, GLRLM, and the wavelet model to assess their effectiveness in differentiating between lesion and non-lesion pictures by confusion matrix analysis. The GLCM model analyzed 80 OPG images, yielding 39 true negatives, 1 false negative, 0 false positives, and 40 true positives. Among the photos, 40 had lesions whereas 40 were deemed normal. The GLRLM model identified 40 true negatives, 2 false negatives, 2 false positives, and 36 true positives using the same number of images. The wavelet model recorded 30 true negatives, 4 false negatives, 3 false positives, and 43 true positives. These results were used to calculate accuracy, specificity, sensitivity, Matthews Correlation Coefficient (MCC), and precision rate for all three techniques, demonstrating the comparative performance of each method in lesion detection.

A Study by Kown et al. [17]A dataset of 1282 panoramic patient radiographs was randomly split into a training dataset (80%) and a test dataset (20%) before data augmentation. The training dataset included 280, 242, 240, and 184 panoramic images for dentigerous cysts, periapical cysts, odontogenic keratocysts, and Ameloblastomas, respectively. The test dataset contained 70, 60, 60, and 46 images for each lesion type. The training dataset was used to train the convolutional neural network (CNN), while the test dataset was employed to assess the performance of the final trained model. In this study, 503 cases were collected, with 238 images also used for both training and testing to evaluate the model's performance. Furthermore, this study surpasses the previous one by incorporating an additional phase, utilizing 40 cases in a separate test phase specifically focused on diagnosing the lesions. The summary is shown in **Table 3.**

Table 3. Summary Comparison Table

Aspect	Prior Studies	Our Current Research
Lesion Localization	Manual, impractical, and subjective.	Fully automated, mimics clinical reality, and is objective
Data Augmentation	Quantitative and random to increase the data size.	Methodical and qualitative (generational) to simulate real-world scenarios
Model Evaluation	The test set was derived from the original dataset.	Completely new and independent test set for generalization assessment

5. RECOMMENDATIONS, LIMITATIONS, and CONCLUSIONS:

5.1Recommendations

According to this study, outcomes and evaluation of model for finding and identifying dental lesions (abscesses), a few recommendations are proposed:

Medical institutions adopt systematic approaches to archiving clinical data of patients, thereby confirming a unified source that can enhance future research and provide accurate, precise scientific results.

Incorporate Explainable AI (XAI) Tools: Integrating Explainable AI tools (XAI), such as Grad-CAM or SHAP, can foster clinical trust, facilitate the visualization of the model's processes, and provide insights into its decision-making, which is particularly vital in situations where a misclassification could have serious consequences.

Integrate Clinical Metadata with Imaging Input: Providing the model with patient-relevant metadata —such as age, symptoms, and the lesion's location— may enhance classification accuracy.

5.2 Limitations of the Study

Expanding the sample size was pivotal for improving the robustness of the trained AI model. However, the number of available samples was excessively small compared to the necessary data. The augmentation technique was used to double the sample size to achieve a high degree of diagnostic accuracy. Not only that, but also the required clinical radiographic examination executed by an oral and maxillofacial surgery specialist to ensure obtaining relatively accurate results for the thesis.

Data Imbalance and Recall Variance. As noted in the conclusion, the model's performance was influenced by the balance of data. The disparity between the overall accuracy (96.40%) and the recall for abscesses (83.3%) suggests that the model may still face challenges in identifying positive cases within unbalanced clinical environments.

Regarding radiology centers, most of them do not retain or archive previous radiographs, especially panoramic X-rays, under the assumption that their clinical utility for the physician diminishes after six months to a year from the date of examination. This indicates a lack of awareness of the importance of preserving such data to support scientific research.

5.3Conclusion:

The EfficientNetB3 model achieved a high level of validation accuracy and clearly identified the locations of abscesses based on the exploratory findings. These results indicate that the modelling approach demonstrates high generalizability of periapical abscesses.

DECLARATIONS

Funding

This research did not receive any specific grant or financial support from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors have no competing interests to declare.

Ethical Approval

The study was approved by the appropriate ethics committee and conducted according to relevant guidelines and regulations.

REFERENCES

1.Siqueira JF, Rôças IN. Microbiology and Treatment of Acute Apical Abscesses. Clin Microbiol Rev. 2013 Apr;26(2):255–73. doi:10.1128/CMR.00082-12
2.Shu M, Wong L, Miller JH, Sissons CH. Development of multi-species consortia biofilms of oral bacteria as an enamel and root caries model

system. Arch Oral Biol. 2000 Jan;45(1):27–40. doi:10.1016/S0003-9969(99)00111-9
3.Quiñonez C, Gibson D, Jokovic A, Locker D. Emergency department visits for dental care of nontraumatic origin. Community Dent Oral Epidemiol. 2009 Aug 7;37(4):366–71. 16. doi:10.1111/j.1600-0528.2009.00476.x
4.Green AW, Flower EA, New NE. Mortality associated with odontogenic infection! Br Dent J. 2001 26;190(10):529–30. doi:10.1038/sj.bdj.4801024
5.Robertson D, Smith AJ. The microbiology of the 17. acute dental abscess. J Med Microbiol. 2009 Feb 1;58(2):155–62. doi:10.1099/jmm.0.003517-0
6.Flynn TR. THE SWOLLEN FACE. Emerg Med Clin North Am. 2000 Aug;18(3):481–519. doi:10.1016/S0733-8627(05)70140-1
7.Gendron R, Grenier D, Maheu-Robert LF. The oral cavity as a reservoir of bacterial pathogens for focal infections. Microbes Infect. 2000 Jul;2(8):897–906. doi:10.1016/S1286-4579(00)00391-9
8.Bonfanti-Gris M, Garcia-Cañas A, Alonso-Calvo R, Salido Rodriguez-Manzaneque MP, Pradies Ramiro G. Evaluation of an Artificial Intelligence web-based software to detect and classify dental structures and treatments in panoramic radiographs. J Dent. 2022 Nov;126:104301. doi:10.1016/j.jdent.2022.104301
9.Dhinesh Dr V, Balasankari DrK, Ilayanila DrC, Sandhiya DrM, K DrV., Kumar DrMS. Artificial Intelligence in Oral Pathology-The New Era -A Review [Internet]. Zenodo; 2023. Available from: https://doi.org/10.5281/zenodo.8308098 doi:10.5281/zenodo.8308098
10.Ahmed AA, Abouzid M, Kaczmarek E. Deep Learning Approaches in Histopathology. Cancers (Basel). 2022 Oct 26;14(21):5264. doi:10.3390/cancers14215264
11.Ding H, Wu J, Zhao W, Matinlinna JP, Burrow MF, Tsoi JKH. Artificial intelligence in dentistry—A review. Frontiers in Dental Medicine. 2023 Feb 20;4. doi:10.3389/fdmed.2023.1085251
12.Naseri S, Shukla S, Hiwale K, Jagtap MM, Gadkari P, Gupta K, et al. From Pixels to Prognosis: A Narrative Review on Artificial Intelligence's Pioneering Role in Colorectal Carcinoma Histopathology. Cureus. 2024 Apr 27. doi:10.7759/cureus.59171
13.Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. Journal of Oral Pathology & Medicine. 2020 Oct 15;49(9):849–56. doi:10.1111/jop.13042
14.Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:190511946 [Internet]. 2019. Available from: https://arxiv.org/abs/1905.11946
15.Kumar VS, Kumar PR, Yadalam PK, Anegundi RV, Shrivastava D, Alfurhud AA, et al. Machine learning in the detection of dental cyst, tumor, and abscess lesions. BMC Oral Health. 2023 Nov 6;23(1):833. doi:10.1186/s12903-023-03571-1
16.Esmaeilyfard R, Esmaeeli N, Paknahad M. An artificial intelligence mechanism for detecting cystic lesions on CBCT images using deep learning. J Stomatol Oral Maxillofac Surg. 2025 Dec;126(6):102152. doi:10.1016/j.jormas.2024.102152
17.Kwon O, Yong TH, Kang SR, Kim JE, Huh KH, Heo MS, et al. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. Dentomaxillofacial Radiology.2020;1;49(8):20200185. doi:10.1259/dmfr.20200185