Development of an artificial intelligence model for standardized evaluation of apical periodontitis

Park, Yudam

Department of Dentistry

Graduate School

Yonsei University

Development of an artificial intelligence model for standardized evaluation of apical periodontitis

Advisor Kim, Sunil

A Master's Thesis

Submitted to the Department of Dentistry and the

Committee on Graduate School of Yonsei University

in partial fulfillment of the

Requirements for the degree of

Master of Dental Science

Park, Yudam

June 2025

Development of an artificial intelligence model for standardized evaluation of apical periodontitis

This certifies that the Master's Thesis of Park, Yudam is approved.

Committee Chair	Kim, Euiseong
Committee Member	Kim, Sunil
Committee Member	Kim, Ukseong

Department of Dentistry
The Graduate School
Yonsei University
June 2025

Table of Contents

List of Figures

List of Tables

Abstract

Development of an artificial intelligence model for standardized evaluation of apical periodontitis

Park, Yudam, D.D.S.

Department of Dentistry
The Graduate School, Yonsei University
(Directed by Professor Kim, Sunil, D.D.S., M.S.D., Ph.D.)

Apical periodontitis is a common inflammatory condition of endodontic origin that requires accurate radiographic evaluation for proper treatment planning. The Periapical Index (PAI) is a widely used scoring system for assessing lesion severity; however, it is subject to interobserver variability and diagnostic inconsistency. This study aimed to develop an artificial intelligence (AI) model specifically designed for automated PAI scoring, utilizing expert-annotated periapical radiographs to classify lesions according to the standardized PAI system.

A total of 8,506 annotated radiographs were used to train a ResNet50-based convolutional neural network (CNN). To improve performance and generalizability, the model incorporated contrast enhancement, dataset-specific normalization, and extensive data augmentation. A soft2-encoded cross-entropy loss function was used to account for the ordinal nature of PAI scores. Stratified five-fold cross-validation demonstrated strong agreement with expert scorings, achieving a quadratic weighted kappa (QWK) of 0.729 and consistent performance across all five PAI categories.

The proposed model provides standardized and objective lesion evaluation, which may reduce diagnostic variability and support general practitioners in making more informed clinical decisions. It also has potential applications in dental education by offering visual feedback and consistent references for student training. Future work should focus on multi-institutional data expansion, clinical validation, and integration with outcome-predictive AI models to establish a comprehensive decision-support system for endodontic treatment planning.

Keywords: Artificial intelligence; Periapical Index (PAI); Apical periodontitis; Convolutional neural network; Radiographic diagnosis

Development of an artificial intelligence model for standardized evaluation of apical periodontitis

Park, Yudam

Department of Dentistry
The Graduate School, Yonsei University
(Directed by Professor Kim, Sunil)

I. Introduction

Apical periodontitis is a prevalent dental condition characterized by inflammation of the periapical tissues, primarily resulting from bacterial infection of the dental pulp (Kakehashi et al., 1965). Without treatment, periapical inflammation may persist and contribute to gradual loss of periodontal support. Persistent or unresolved apical periodontitis remains a significant challenge in endodontics, often necessitating complex retreatment procedures or tooth extraction (Nair, 2006).

Endodontic treatments can effectively preserve natural teeth and have demonstrated favorable long-term survival rates comparable or superior to dental implants in many clinical scenarios (Setzer & Kim, 2014; Torabinejad & Goodacre, 2006). Nevertheless, many general dentists frequently opt for extraction followed by dental implant placement, particularly in complex or severe cases (Lee et al., 2020). This preference often arises from difficulties in accurately assessing the severity and prognosis of periapical lesions, which requires considerable clinical experience and expertise. Clinical decisions regarding tooth preservation versus extraction significantly depend on various factors, including tooth position, root canal filling status, lesion size, and characteristics (Lee et al., 2020).

The Periapical Index (PAI) scoring system provides a standardized method for evaluating lesion severity on a scale ranging from 1 (healthy) to 5 (severe periodontitis) (Ørstavik et al., 1986). However, manual PAI scoring is inherently subjective, relying heavily on the clinician's experience and judgment. Studies have reported moderate intraobserver agreement ($\kappa = 0.48$) and fair interobserver agreement ($\kappa = 0.39$), highlighting that even experienced clinicians may face challenges maintaining consistency (Tarcin et al., 2015). Additionally, the accuracy of manual PAI assessments can occasionally be compromised by limitations inherent to periapical radiographs, such as two-dimensional representation and overlapping anatomical structures, potentially complicating lesion interpretation (Patel et al., 2015). Such variability and limitations can result in inconsistent clinical decisions, including unnecessary tooth extractions. Therefore, the involvement or guidance of experienced endodontists or oral radiologists is frequently recommended to enhance diagnostic accuracy and consistency in clinical practice and research settings.

In recent years, artificial intelligence (AI) has emerged as a powerful tool in medical diagnostics, offering significant potential for automating complex diagnostic tasks, enhancing accuracy, and reducing human error (Schwendicke et al., 2020). In dentistry, convolutional neural networks (CNNs), a subset of deep learning, have been successfully applied to various diagnostic fields such as detecting dental caries, periodontal disease, and root fractures, demonstrating promising diagnostic performance comparable to experienced clinicians (Hung et al., 2020). Furthermore, AI-driven models have expanded into the field of endodontics, where accurate prediction of treatment outcomes and prognosis can significantly impact clinical decision-making. Recently, Lee et al. (2023) proposed PRESSAN-17, a deep convolutional neural network (DCNN) designed to predict endodontic outcomes using preoperative periapical radiographs. Building upon this, Hwang (2024) extended the model to incorporate a significantly larger dataset and clinical features across various tooth types. The modified model improved predictive accuracy and sensitivity, highlighting the scalability and adaptability of DCNNs in endodontic prognosis prediction. These advancements illustrate the expanding role of AI in clinical endodontics and the need for further integration with structured diagnostic tools such as PAI for comprehensive diagnostic support. A recent review by Fontenele and Jacobs (2025) also emphasized the potential of AI in image-based diagnostics and treatment planning in endodontics, while highlighting the current lack of models tailored for standardized lesion classification.

However, the application of AI in endodontics, particularly in the evaluation of periapical lesions, remains underexplored. Existing AI models in endodontics have primarily focused on predicting treatment outcomes based on preoperative radiographs, often overlooking the critical intermediate step of lesion classification through PAI scoring. Although Lee et al.'s study marked a significant step forward by demonstrating that AI can learn relevant features for prognosis prediction, the study implicitly relied on accurate lesion assessment as part of the dataset construction. This highlights a key limitation: before attempting to predict prognosis, the development of a reliable AI model capable of accurately evaluating periapical lesions via PAI scoring must precede.

This gap is significant, as accurate PAI scoring is essential for effective treatment planning and prognosis assessment (Kirkevang et al., 2017). Without reliable and standardized lesion evaluation, the accuracy of subsequent predictive models is compromised. By automating PAI scoring, AI has the potential to eliminate subjectivity, ensure consistency, and provide clinicians with a reliable tool for endodontic decision-making. The present study addresses this gap by developing an AI model specifically designed for automated PAI scoring, utilizing expert-annotated periapical radiographs to classify lesions according to the standardized PAI system.

2. Materials and Methods

2.1. Ethical considerations

This study was approved by the Institutional Review Board (IRB) of the Yonsei University Dental Hospital (2-2024-0064). The IRB waived the need for individual informed consent, as this study featured a non-interventional retrospective design, and all the data were analyzed anonymously. The study was presented in accordance with the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) (Mongan et al., 2020).

2.2. Data Collection

Periapical images were collected from the database of Yonsei University Dental Hospital. Patients who received nonsurgical root canal treatment or retreatment between 2008 and 2015 were included in the database. For each case, preoperative radiographs obtained prior to treatment, as well as postoperative radiographs obtained after the procedures, were separately collected. The inclusion criteria were as follows :

1) Presence of a preoperative periapical radiograph obtained within 3 months before treatment initiation.

2) Periapical radiographs clearly demonstrating the apical region without overlapping adjacent

anatomical structures, enabling accurate PAI scoring.

Only permanent teeth were included in this study. Cases involving retained primary teeth, or unclear root apex visibility due to overlapping anatomical structures were excluded.

The total number of data points in the input dataset was 8506. All patient data were anonymized, and each exported radiographic image was labeled with a randomly generated serial number.

After data collection, PAI assessment prior to AI model training was independently performed by two experienced endodontists. In cases of disagreement, the final PAI score was determined through consensus discussion. The inter-observer reliability, measured by Cohen's kappa coefficient, was 0.445. These finalized consensus scores served as the reference standard for training and evaluating the AI model, and AI-generated scores were subsequently compared against these human expert evaluations. The dataset included diverse cases from mild to severe apical periodontitis, ensuring the AI model’s broad applicability.

2.3. Image Annotation and Preprocessing

All annotations were performed by a single experienced endodontist in a blinded manner with respect to treatment outcomes. Specialized software (Image J, National Institutes of Health, Bethesda, Maryland, USA) was used to ensure consistency in labeling. The evaluator followed a predefined annotation protocol that specified guidelines for selecting the region of interest (ROI) and identifying apical radiolucency. Ambiguous cases were flagged during annotation and re-evaluated after consensus discussion with other endodontist. The preprocessing steps included:

- **Cropping:** Images were cropped to focus on the target tooth, following specific guidelines to ensure that the region ROI was accurately captured. The cropping criteria for maxillary teeth were as follows:

  - **Inferior Boundary:** The lower boundary was set to include the entire crown of the target tooth.
  - **Lateral Boundaries:** The left and right boundaries were set to include the pulpal horn of the adjacent teeth. If the target tooth was not fully included within this range, the boundaries were extended until the target tooth was fully captured. In cases where an adjacent tooth was missing, the boundary was set symmetrically based on the available adjacent tooth.
  - **Superior Boundary:** The upper boundary was set to include the lesion margin plus an additional 2 mm superiorly. If the lesion extended beyond the radiographic field, the boundary was extended to the uppermost edge of the radiograph.

  For mandibular teeth, the same criteria were applied in reverse, with the superior and inferior boundaries inverted.

- **Contrast Normalization:** Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to enhance image contrast, making subtle lesions more visible.
- **Resizing:** Images were cropped to focus on the target tooth and resized to a standard dimension of 224 x 224 pixels.
- **Data Augmentation:** Data augmentation techniques were applied using the Albumentations library, including geometric transformations, intensity adjustments, Gaussian Blur, Elastic Transform, Grid Distortion, Optical Distortion, and Random Gamma adjustments to prevent model overfitting and ensure robustness. These augmentations were performed on-the-fly during training to prevent overfitting and improve the model's ability to generalize to new, unseen data.

2.4. AI Model Development

A CNN based on the ResNet50 architecture was utilized, employing:

- **Cross-Validation:** Stratified cross-validation was performed to maintain consistent class distribution, ensuring robust model generalization. Multiple dataset partitions were averaged to minimize data distribution bias and effectively handle class imbalance.

- **Transfer Learning:** Due to the limited size of the dataset for fully training a large-scale deep learning model, transfer learning was employed. The model was initialized with parameters pretrained on a large-scale dataset (ImageNet), followed by fine-tuning to the periapical radiograph dataset. Since periapical radiographs differ visually from ImageNet images, feature reuse alone was insufficient. Therefore, initial layers were kept fixed, while intermediate and upper layers were fine-tuned using a layer-wise fine-tuning approach to balance generalizability of pretrained features and adaptation to the target domain.

- **Loss Function:** Soft-encoded cross-entropy (SCE) with additional regularization terms was used to reflect the ordinal nature of PAI scoring and enforce unimodal output distribution.

2.5. Model Evaluation

The model's performance was assessed using:

- **QWK Scores:** Quadratic weighted kappa (QWK) was the primary metric, complemented by accuracy measurements. QWK is particularly suited for ordinal classification tasks as it incorporates the distance between predicted and actual classes into the evaluation, penalizing larger discrepancies more heavily. This allows for a nuanced assessment of model performance in cases where class distinctions carry meaningful ordinal relationships, making QWK ideal for evaluating the reliability of PAI scoring models.

- **Confusion matrix:** To evaluate classification performance in detail, a confusion matrix was generated using the predictions of the trained model on the test dataset. The confusion matrix visualizes the agreement between predicted PAI scores and the expert-annotated ground truth labels, allowing for a granular assessment of classification accuracy across ordinal categories. Given the ordinal nature of the PAI scoring system, the matrix was carefully analyzed to assess whether the model tended to misclassify into adjacent scores rather than making large errors across non-neighboring classes.

2.6. Visualization of Model Predictions

Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to visually interpret the trained model's predictions. This method aggregates gradient-weighted feature maps from convolutional layers to highlight regions within the input image that were most influential in the model's decision-making process.

Representative cases were selected based on the model’s prediction confidence and the level of agreement with expert PAI scores. Both high-confidence correct predictions and borderline misclassifications were included to provide a comprehensive overview of the model's behavior across a range of scenarios.

The resulting Grad-CAM heatmaps were superimposed onto the corresponding preprocessed periapical radiographs to visualize the regions of diagnostic focus. Highlighted areas were considered clinically valid if they overlapped with the apical radiolucency regions identified by human experts.

3. Results

The distribution of PAI scores within the dataset is presented in Table 1. Among the total cases, the most frequent PAI score was 3, accounting for 2,504 cases (29.44%), followed by scores of 2 (2,155 cases; 25.34%) and 4 (1,776 cases; 20.88%). Cases with scores of 5 and 1 were less prevalent, comprising 1,161 (13.65%) and 910 (10.70%) cases, respectively.

Table 1 The Distribution of PAI Scores

PAI score	1	2	3	4	5
Number of cases(%)	910(10.70%)	2155(25.34%)	2504(29.44%)	1776(20.88%)	1161(13.65%)

Table 2 presents the classification performance metrics of the proposed model across five cross-validation folds. The AI model achieved a QWK score of 0.7297 ± 0.0100 and an accuracy of 0.5441 ± 0.0230 across five-fold stratified cross-validation, indicating stable and reliable performance. Individual fold QWK values ranged from 0.7174 to 0.7445, while accuracy ranged from 0.5132 to 0.5724.

Table 2 Performance of the AI model across 5-fold cross-validation.

	Accuracy	**QWK**
Fold 1	0.5296	0.7174
Fold 2	0.5691	0.7381
Fold 3	0.5132	0.7253
Fold 4	0.5724	0.7445
Fold 5	0.5362	0.7233
Avg(std)	0.5441(0.0230)	0.7297(0.0100)

As shown in the confusion matrix (Figure 1), the model achieved the highest classification accuracy for PAI score 3, with 1,173 instances correctly predicted. Score 5 also showed favorable performance with 688 correctly classified cases. Most misclassifications occurred between neighboring classes, especially between scores 2 and 3, and scores 4 and 5, consistent with the ordinal structure of the PAI scale. For example, score 2 was frequently misclassified as 3 (492 cases), and score 4 was sometimes predicted as 5 (213 cases), while score 3 also received misclassifications from both directions (601 predicted as 3 from score 2; 313 from score 4).

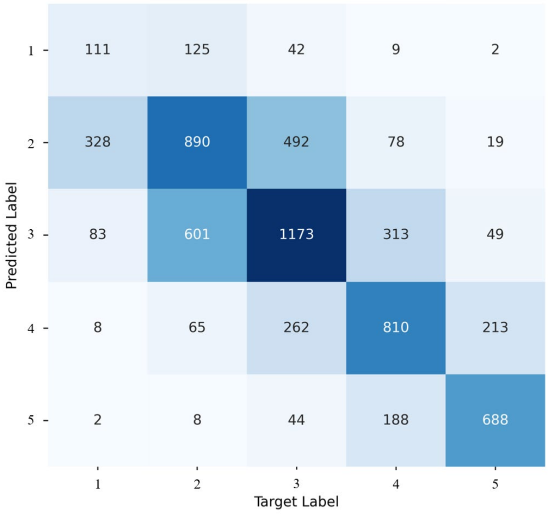

Figure 1 Confusion matrix of the AI model's predictions versus expert-assigned PAI scores.

Visualization of Model Predictions

To provide visual interpretability of the model's decision-making process, Grad-CAM was utilized. Representative examples are shown in Figure 2. Across multiple test samples, Grad-CAM consistently emphasized the periapical region, particularly around the apex of the root, where radiolucent signs of apical periodontitis typically appear. For cases predicted correctly with high confidence (e.g., PAI scores 3 and 4), the highlighted areas matched well with the lesion sites annotated by experienced endodontists.

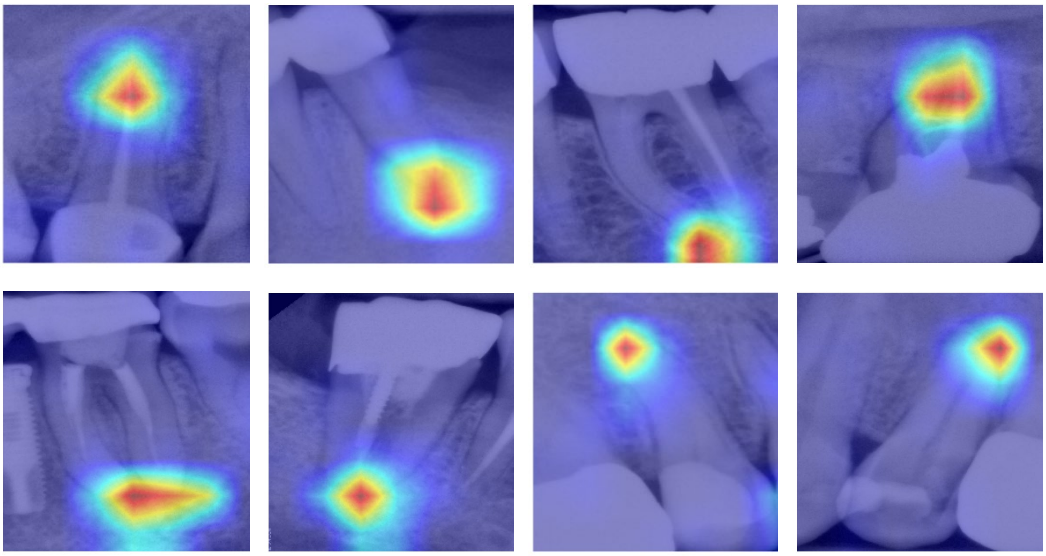

Figure 2 Representative Grad-CAM visualizations of AI predictions for PAI scoring.

4. Discussion

Recent advances in AI have enabled deep DCNNs to extract clinically relevant features from periapical radiographs and support decision-making in endodontics (Schwendicke et al., 2020). A representative study by Lee et al. (2023) introduced PRESSAN-17, a self-attention–enhanced DCNN model that accurately detected key clinical features and predicted three-year endodontic outcomes from preoperative radiographs. Similarly, Hwang (2024) developed a DCNN-based model to forecast endodontic prognosis from intraoral radiographs, further demonstrating the potential of AI in prognosis-driven decision support. Their results demonstrated the potential of AI models to assist clinicians in formulating prognosis and treatment decisions based on radiographic information.

However, while these models focused primarily on outcome prediction, they often required manual cropping of the region of interest and lacked interpretability in intermediate diagnostic steps.

The present study's model enables objective, consistent assessment of lesion severity without requiring manual preprocessing. By targeting the PAI score directly, the model offers interpretable outputs aligned with clinical evaluation practices. This not only increases its potential for seamless clinical integration but also provides a foundation for future extensions into outcome prediction models. Moreover, the inclusion of Grad-CAM visualizations enhances trust and interpretability, allowing clinicians to understand the basis of AI-generated decisions. Ultimately, this study's model bridges a practical and diagnostic gap, complementing existing approaches while improving real-world applicability.

The QWK is a robust metric for evaluating agreement in ordinal classification tasks, as it considers both the order of categories and the degree of disagreement between predicted and actual labels. Unlike simple accuracy, which treats all misclassifications equally, QWK penalizes predictions more severely when they deviate further from the true class, making it particularly suitable for tasks such as PAI scoring, where classes are ordered and clinically graded.

In this study, the proposed AI model achieved a QWK score of 0.729, representing substantial agreement with expert-assigned PAI scores. This result is notable given the intrinsic difficulty of PAI scoring, which requires the evaluator to distinguish between subtle differences in lesion severity across five ordinal categories. In previous studies outside of dentistry, QWK has been widely used to assess AI performance in similarly complex classification tasks. For instance, Araújo et al. (2020) applied QWK to evaluate their DR|GRADUATE system for diabetic retinopathy grading, reporting values ranging from 0.71 to 0.84 depending on the dataset, despite variability in image quality and inter-rater uncertainty. Additionally, Swiecicki et al. (2021) developed a deep learning-based algorithm for assessing knee osteoarthritis severity in radiographs using the Kellgren-Lawrence grading system. Their model achieved a QWK of 0.9066 when compared to expert annotations, demonstrating the effectiveness of QWK in evaluating AI performance in medical imaging tasks with ordinal scales. Although these models achieved comparable or slightly higher QWK scores, it is important to consider the unique challenges of PAI classification, including the radiographic complexity of periapical lesions. Therefore, the QWK of 0.729 attained in this study can be considered a clinically meaningful result that supports the model's potential as a decision-support tool in endodontics.

To improve the model's QWK performance, several steps were taken throughout this study. First, data distribution was balanced by curating a dataset with relatively uniform representation across all five PAI scores. Additionally, the number of radiographs was significantly increased compared to previous studies, enhancing model generalizability. Preprocessing techniques such as contrast enhancement using CLAHE and dataset-specific intensity normalization further improved image quality for learning.

From a modeling perspective, key architectural enhancements were introduced. A SCE loss function with ordinal regularization was employed to better reflect the structure of the PAI scale. A layer-wise fine-tuning approach allowed the model to effectively leverage pretrained ImageNet weights while adapting to the nuances of grayscale dental radiographs. Furthermore, advanced data augmentation techniques, including distortion-based transformations, were added to improve robustness and reduce overfitting. The final model thus incorporates both data-level and model-level optimizations aimed at maximizing ordinal classification performance in the context of PAI scoring. These design choices collectively contributed to the achieved QWK score and improved the clinical utility of the system.

Compared to the study by Moidu et al. (2022), which developed a YOLOv3-based CNN for automated PAI classification, our study offers several key improvements. Their model showed limited accuracy for intermediate and severe lesions, whereas the model developed in this study demonstrated more balanced performance across all five PAI categories, particularly for scores 3, 4, and 5. This was enabled by a larger, more evenly distributed dataset and advanced preprocessing, including CLAHE, dataset-specific normalization, and distortion-based augmentation. In addition, our model utilized a soft-encoded loss function tailored for ordinal classification, allowing for more appropriate handling of the graded nature of PAI scores. The integration of Grad-CAM also enhanced model transparency, making it more suitable for clinical decision support and educational use.

The proposed AI model holds significant potential for clinical implementation. By integrating automated PAI scoring into dental radiographic workflows, clinicians can obtain consistent, objective assessments of lesion severity, reducing diagnostic variability across practitioners. This capability is particularly valuable for general dentists who may not have specialized endodontic training, promoting more standardized and evidence-based treatment decisions. Such decision-support tools have been shown to enhance diagnostic efficiency and confidence without undermining the clinician's role in final judgment. Beyond clinical application, this study's model shows potential as an educational tool for training purposes. Integrating AI-assisted PAI scoring into dental training programs can help students learn to assess periapical lesions with greater consistency. By comparing their judgments with model predictions and Grad-CAM visualizations, students can deepen their understanding of radiographic interpretation and improve their diagnostic accuracy in endodontics.

While our AI model demonstrates strong performance, some limitations should be considered. First, the training data was derived from a single institution, which may limit generalizability. To address this, future studies should conduct external validation using independent datasets to evaluate the model's performance across diverse clinical settings and patient populations. Second, while our model achieved a high QWK score, further optimization through deep learning architectures, such as attention mechanisms, may enhance performance. Additionally, real-time AI-assisted diagnostic tools should be developed and tested in clinical settings to evaluate their practical impact on treatment planning. Finally, further research should focus on refining the AI model to reduce potential biases, improve interpretability, and ensure seamless integration into existing dental imaging systems.

The integration of the proposed model with predictive models for long-term treatment outcomes could provide a comprehensive decision-support system for endodontic care. Another promising avenue involves developing image-generation models capable of reconstructing anticipated post-treatment radiographic outcomes based on preoperative imaging data. This innovation would provide clinicians with valuable visual representations to facilitate patient communication and informed decision-making, potentially transforming endodontic diagnostics by coupling predictive analytics with intuitive visual interpretation.

5. Conclusion

This study developed and validated an AI model for automated PAI scoring that achieved expert-level accuracy. By providing consistent and objective assessments of periapical lesions, the model has potential to support standardized diagnostic workflows in clinical endodontics. These findings highlight the role of AI in enhancing diagnostic reliability and promoting precision in endodontic treatment planning.

6. References

Araújo, T., Aresta, G., Mendonça, L., Penas, S., Maia, C., Carneiro, Â., Mendonça, A. M., & Campilho, A. (2020). DR| GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images. *Medical Image Analysis*, *63*, 101715.

Fontenele, R. C., & Jacobs, R. (2025). Unveiling the power of artificial intelligence for image-based diagnosis and treatment in endodontics: An ally or adversary? *International endodontic journal*, *58*(2), 155–170.

Hung, K., Yeung, A. W. K., Tanaka, R., & Bornstein, M. M. (2020). Current applications, opportunities, and limitations of AI for 3D imaging in dental research and practice. *International Journal of Environmental Research and Public Health*, *17*(12), 4424.

Hwang, J. (2024). Forecasting endodontic outcomes with deep convolution neural network by analyzing intraoral radiographs. Master's thesis, Graduate School, Yonsei University, Seoul.

Kakehashi, S., Stanley, H., & Fitzgerald, R. (1965). The effects of surgical exposures of dental pulps in germ-free and conventional laboratory rats. *Oral surgery, oral medicine, oral pathology*, *20*(3), 340–349.

Kirkevang, L. L., Ørstavik, D., Bahrami, G., Wenzel, A., & Væth, M. (2017). Prediction of periapical status and tooth extraction. *International endodontic journal*, *50*(1), 5–14.

Lee, J., Kang, S., Jung, H.-I., Kim, S., Karabucak, B., & Kim, E. (2020). Dentists' clinical decision-making about teeth with apical periodontitis using a variable-controlled survey model in South Korea. *BMC Oral Health*, *20*, 1–10.

Lee, J., Seo, H., Choi, Y. J., Lee, C., Kim, S., Lee, Y. S., Lee, S., & Kim, E. (2023). An endodontic forecasting model based on the analysis of preoperative dental radiographs: a pilot study on an endodontic predictive deep neural network. *Journal of Endodontics*, *49*(6), 710–719.

Moidu, N. P., Sharma, S., Chawla, A., Kumar, V., & Logani, A. (2022). Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. *Clinical oral investigations*, *26*(1), 651–658.

Mongan, J., Moy, L., & Kahn Jr, C. E. (2020). Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. In (Vol. 2, pp. e200029): Radiological Society of North America.

Nair, P. (2006). On the causes of persistent apical periodontitis: a review. *International endodontic journal*, *39*(4), 249–281.

Ørstavik, D., Kerekes, K., & Eriksen, H. M. (1986). The periapical index: a scoring system for radiographic assessment of apical periodontitis. *Dental Traumatology*, *2*(1), 20–34.

Patel, S., Durack, C., Abella, F., Shemesh, H., Roig, M., & Lemberg, K. (2015). Cone beam computed tomography in e ndodontics–a review. *International endodontic journal*, *48*(1), 3–15.

Schwendicke, F. a., Samek, W., & Krois, J. (2020). Artificial intelligence in dentistry: chances and challenges. *Journal of dental research*, *99*(7), 769–774.

Setzer, F., & Kim, S. (2014). Comparison of long-term survival of implants and endodontically treated teeth. *Journal of dental research*, *93*(1), 19–26.

Swiecicki, A., Li, N., O'Donnell, J., Said, N., Yang, J., Mather, R. C., Jiranek, W. A., & Mazurowski, M. A. (2021). Deep learning-based algorithm for assessment of knee osteoarthritis severity in radiographs matches performance of radiologists. *Computers in biology and medicine*, *133*, 104334.

Tarcin, B., Gumru, B., Iriboz, E., Turkaydin, D. E., & Ovecoglu, H. S. (2015). Radiologic assessment of periapical health: comparison of 3 different index systems. *Journal of Endodontics*, *41*(11), 1834–1838.

Torabinejad, M., & Goodacre, C. J. (2006). Endodontic or dental implant therapy: the factors affecting treatment planning. *The journal of the american dental association*, *137*(7), 973–977.

Abstract (In Korean)

치근단 치주염의 표준화된 평가를 위한

인공지능 모델 개발

박 유 담

연세대학교 대학원

치의학과

(지도교수 김 선 일)

치수 감염에 기인한 대표적인 염증성 질환인 치근단 치주염은 적절한 치료 계획 수립을 위해 정확한 방사선학적 평가가 필수적이다. Periapical Index (PAI)는 병소의 중증도를 평가하는 데 널리 사용되는 지표이나, 평가자 간 변동성과 진단의 일관성 부족이라는 한계를 지닌다. 본 연구는 치근단 방사선 영상을 기반으로 병소를 표준화된 PAI 기준에 따라 자동 분류할 수 있는 인공지능(AI) 모델을 개발하고자 하였다.

총 8,506 장의 라벨링된 방사선 영상을 활용하여 ResNet50 기반의 합성곱 신경망(CNN)을 학습시켰으며, 성능 향상 및 일반화 능력 확보를 위해 CLAHE 대비 향상, 데이터셋 특이적 정규화, 다양한 형태의 데이터 증강을 적용하였다. 또한, PAI 점수의 서열적 특성을 반영하기 위해 소프트 인코딩된 교차 엔트로피 손실

함수(SCE)를 도입하였다. 교차 검증 결과, 본 모델은 전문가의 평가와 높은 일치도를 보였으며, quadratic weighted kappa (QWK) 계수는 0.729 로 나타나 전 PAI 등급에서 고른 성능을 확인하였다.

본 모델은 표준화되고 객관적인 병소 평가를 제공함으로써 진단의 일관성을 높이고, 일반 치과의사의 임상적 의사결정을 지원할 수 있다. 더불어, 시각적 피드백 및 일관된 기준 제공을 통해 치과 교육 분야에서도 효과적으로 활용될 수 있는 가능성을 지닌다. 향후 연구는 외부 데이터셋을 활용한 검증, 실제 임상 환경에서의 타당성 평가, 예후 예측 AI 모델과의 통합을 통해 근관치료에 활용 가능한 포괄적인 임상 의사결정 지원 시스템으로 확장되어야 할 것이다.

핵심 되는 말 : 인공지능, Periapical Index scoring, 치근단 병소, 합성곱 신경망, 딥러닝, 방사선 진단