Imaging Science in Dentistry 2024; 54: 257-63
https://doi.org/10.5624/isd.20240020

Classification of mandibular molar furcation involvement in periapical radiographs by deep learning

Katerina Vilkomir[1], Cody Phen[2], Fiondra Baldwin[2], Jared Cole[2], Nic Herndon[1], Wenjian Zhang[2,*]

[1]*Department of Computer Science, East Carolina University, Greenville, NC, USA*
[2]*School of Dental Medicine, East Carolina University, Greenville, NC, USA*

ABSTRACT

Purpose: The purpose of this study was to classify mandibular molar furcation involvement (FI) in periapical radiographs using a deep learning algorithm.

Materials and Methods: Full mouth series taken at East Carolina University School of Dental Medicine from 2011-2023 were screened. Diagnostic-quality mandibular premolar and molar periapical radiographs with healthy or FI mandibular molars were included. The radiographs were cropped into individual molar images, annotated as "healthy" or "FI," and divided into training, validation, and testing datasets. The images were preprocessed by PyTorch transformations. ResNet-18, a convolutional neural network model, was refined using the PyTorch deep learning framework for the specific imaging classification task. CrossEntropyLoss and the AdamW optimizer were employed for loss function training and optimizing the learning rate, respectively. The images were loaded by PyTorch DataLoader for efficiency. The performance of ResNet-18 algorithm was evaluated with multiple metrics, including training and validation losses, confusion matrix, accuracy, sensitivity, specificity, the receiver operating characteristic (ROC) curve, and the area under the ROC curve.

Results: After adequate training, ResNet-18 classified healthy vs. FI molars in the testing set with an accuracy of 96.47%, indicating its suitability for image classification.

Conclusion: The deep learning algorithm developed in this study was shown to be promising for classifying mandibular molar FI. It could serve as a valuable supplemental tool for detecting and managing periodontal diseases.
(Imaging Sci Dent 2024; 54: 257-63)

KEY WORDS: Mandible; Molar; Periodontitis; Radiography; Deep Learning

Introduction

Artificial intelligence (AI) is defined as the ability of a machine to imitate human intelligence to perform sophisticated tasks, such as object recognition, problem-solving, and decision-making.[1,2] Convolutional neural networks (CNNs), the latest and core model of artificial neural networks and deep learning, have recently demonstrated excellent performance in computer vision, including mapping, localization, tracking, and facial recognition.[3] In the medical field, radiology is considered one of the forefront specialties where AI is readily applicable, since digitally coded images can be easily translated into computer language.[4]

Numerous AI diagnostic models have been developed to detect diseases, including pulmonary nodules,[5] coronary artery calcification,[6] cerebral aneurysms,[7] and colon polyps.[8] In the dental field, studies have investigated the accuracy and efficacy of AI in diagnosing caries,[9] alveolar bone loss,[10,11] periapical lesions,[12] and maxillofacial cysts and/or tumors,[13,14] with promising results.

Furcation involvement (FI) refers to the loss of alveolar bone between the roots of multirooted teeth.[15] It is a common finding in periodontitis cases and a predictor of poorer

Received February 13, 2024; Revised May 13, 2024; Accepted June 26, 2024
Published online August 12, 2024
*Correspondence to : Prof. Wenjian Zhang
East Carolina University School of Dental Medicine, Mail Stop 701, 1851 MacGregor Downs Rd, Greenville, NC 27834 USA
Tel) 1-252-737-7888, E-mail) zhangw20@ecu.edu

Copyright ⓒ 2024 by Korean Academy of Oral and Maxillofacial Radiology
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Imaging Science in Dentistry · pISSN 2233-7822 eISSN 2233-7830

long-term prognosis compared with teeth without FI.[16-18] Traditionally, dentists rely on periodontal probing, mobility tests, and X-ray examinations to detect the presence of FI.[19,20] However, limited accessibility, complex anatomy, variable force and angle in periodontal probing, and inconsistent experiences and training of the operators negatively impact the accuracy and reproducibility of FI diagnosis.[21-25]

To date, limited studies have attempted to detect molar FI by deep learning.[26,27] With the rapid development of this technology, it is imperative to test different deep learning mechanisms to achieve the highest diagnostic precision for FI.

The objective of this study was to identify and train a deep learning algorithm to achieve adequate accuracy in the identification of mandibular molar FI. The exploration will have benefits for the clinical decision-making process and ultimately improve therapeutic outcomes for patients with periodontal disease.

Materials and Methods

Subjects

Patients who visited the East Carolina University School of Dental Medicine between July 2011 and October 2023 were screened. Those with diagnostic-quality full mouth series (FMXs), both with and without mandibular molar FI, were included in the study. Institutional review board (IRB) approval was obtained prior to the commencement of the study.

Imaging acquisition

The FMXs were captured using Instrumentarium Focus wall-mounted intraoral X-ray units (Instrumentarium Dental Inc., Charlotte, NC, USA) equipped with an XCP receptor-holding device (Dentsply Rinn, Elgin, IL, USA) and rectangular collimation. The exposure settings were 70 kVp and 7 mA, with the exposure time adjusted according to the anatomical location.

Imaging preparation

Mandibular premolar and molar periapical radiographs were cropped into individual tooth images, which were then manually annotated as "healthy" or "FI." The original dimensions of the periapical radiographs were 1644 × 1930 pixels, while the cropped images measured 1644 × 643 pixels. The pixel size of the radiographs was 21 × 21 μm, with a bit depth of 16. The annotation process was carried out by CP, FB, and JC, and subsequently verified by WZ based on the treatment notes in the electronic patient record system, Axium (Henry Schein, Melville, NY, USA). The data were segmented into 3 sets: training, validation, and testing, using a random partitioning method. The training dataset included 402 healthy teeth and 509 FI teeth. The validation dataset comprised 41 healthy teeth and 41 FI teeth, while the testing dataset contained 45 healthy teeth and 40 FI teeth. The distribution of the dataset was designed to present a comprehensive challenge to the model, incorporating 25 complicated cases across the training, validation, and testing sets. This strategy enabled the model to learn from a variety of examples and assess its robustness in handling challenging cases.

Image preprocessing and augmentation

The preprocessing started by resizing the images to 224 × 224 pixels, establishing a uniform input size for training, validation, and testing phases. To maintain consistency, normalization was applied using a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225]. This step ensured that the pixel values were scaled uniformly across all datasets.

Various PyTorch transformations were used to augment the training dataset, thereby increasing the sample size and reducing the risk of overfitting. Once the images were converted into tensors—the required format for PyTorch models—techniques such as random horizontal flipping, translations, scaling, and shearing were applied during the training phase. These methods enhanced data diversity and simulated minor positional changes. Additionally, adjustments to brightness, contrast, saturation, and hue were made randomly to bolster the model's capability to manage color variations.

The enhancements collectively provided a robust training experience, enhancing sampling and data diversity.[28] Consistent resizing and normalization ensured uniform and reliable pixel scale values across all datasets.

Model Architecture

ResNet-18, a CNN model, was chosen and adapted for the imaging analysis because of its proven effectiveness in image classification tasks. This model had been pre-trained on the ImageNet dataset. The PyTorch deep learning framework was used to further refine the model. The last fully connected layer of the pre-trained ResNet-18 model was substituted with a new linear layer that has 2 output features, "healthy" and "FI," which represent the 2 classes of teeth examined in the study.

Performance evaluation for classification

The training and evaluation process involved a combination of data loading, optimization, early stopping, and com-

prehensive evaluation metrics.

To efficiently load images for training and testing, the PyTorch DataLoader was utilized. It created batches of 8 samples, resulting in 114 batches for training and 11 batches for both the validation and testing datasets. The CrossEntropy Loss was chosen to monitor the loss function, while the AdamW optimizer was employed to enhance the model's performance, using a learning rate of 1e-5 and a weight decay of 1e-3.

An early stopping technique was implemented as a precaution against overfitting. The model's performance was assessed for validation loss at each epoch. Early stopping would be triggered if there was no improvement in loss for 3 consecutive epochs, or if the improvement was less than a minimum delta of 0.01.

For testing purposes, a variety of images with different levels of complexity were input into the modified ResNet-18 algorithm. The performance of this algorithm was evaluated using several parameters, including training and validation losses, confusion matrix, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, receiver operating characteristic (ROC) curve, and area under the ROC curve.

Results

Throughout the training process, the model's performance was closely monitored. The training loss showed a steady decrease, indicating effective learning. It started at

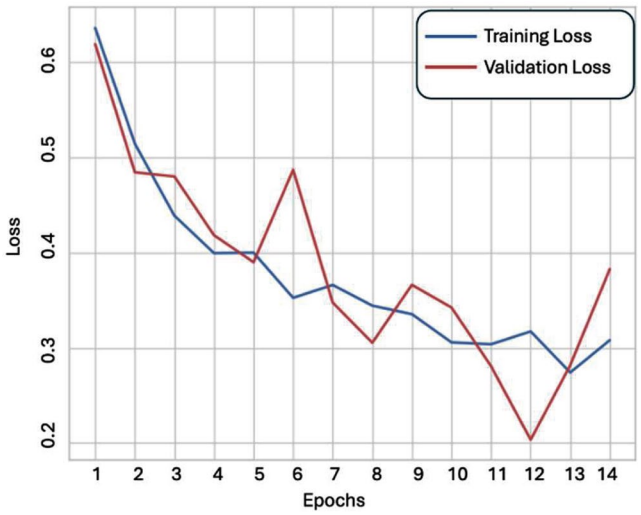

Fig. 1. Training and validation losses, indicating improvement in the ResNet-18 model performance during the training phase.

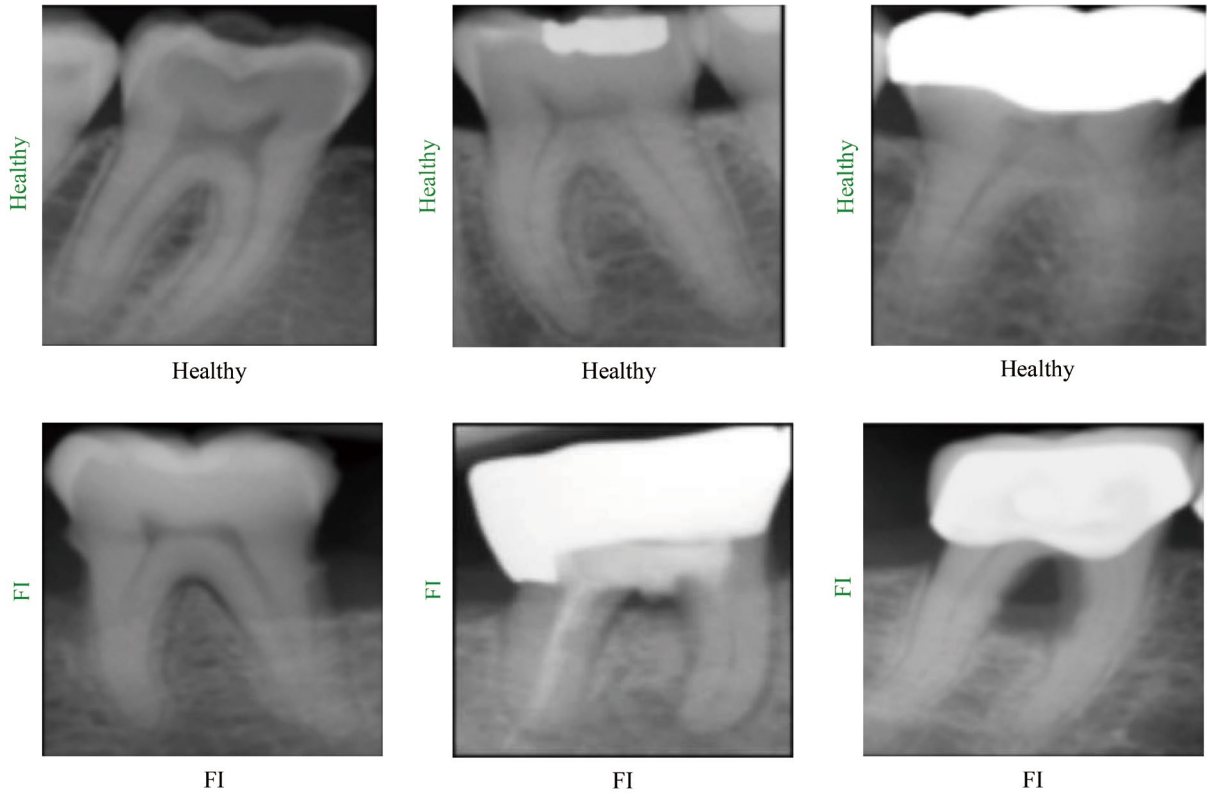

Fig. 2. Demonstration of correctly classified healthy versus furcation involvement (FI) molars by ResNet-18 model in the training dataset. Predicted values are displayed on the vertical plane, and the actual ground truth is shown on the horizontal plane for comparison. A correctly classified value is labeled in green.

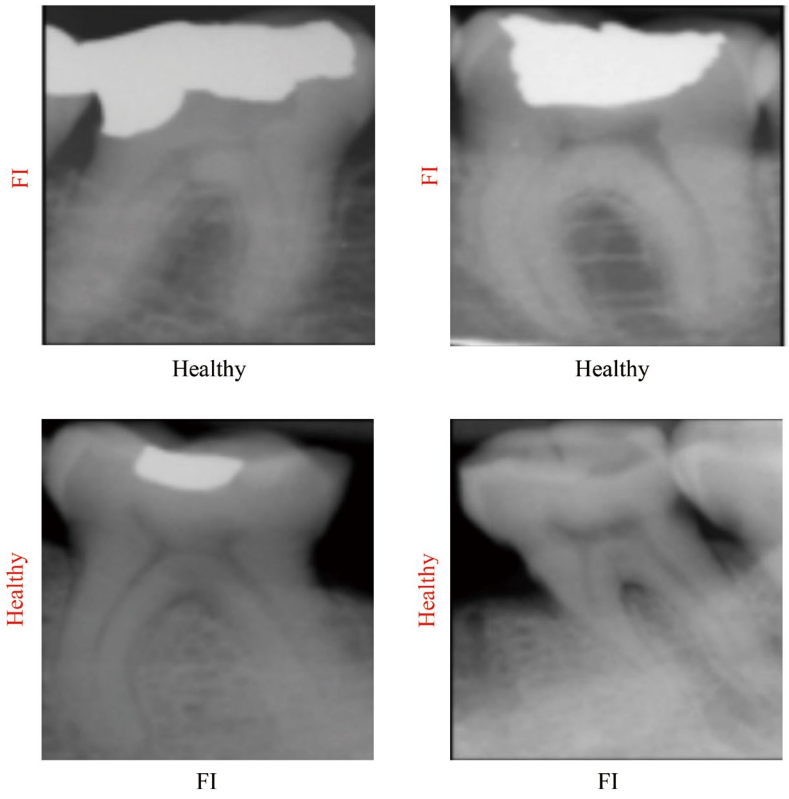

Fig. 3. Demonstration of incorrectly classified healthy versus furcation involvement (FI) molars by the ResNet-18 model in the training dataset. Predicted values are displayed on the vertical plane, and the actual ground truth is shown on the horizontal plane for comparison. An incorrectly classified value is labeled in red.

0.6356 in the first epoch and steadily decreased to 0.2740 by the 13th epoch. Similarly, the validation loss decreased, reflecting a consistent improvement in model performance. Although it rose to 0.4868 during the sixth epoch, it decreased to 0.2035 by the 12th epoch (Fig. 1). The training process was terminated by the early stopping mechanism at the 15th epoch to avoid overfitting. At this point, the training and validation accuracy had increased by 25.36% and 28.05%, respectively.

The evaluation process involved inputting batches of images into the ResNet-18 model for classification. The predicted values were displayed on the left side of each image, while the actual ground truths were labeled below the images for comparison. Correctly classified values were marked in green (Fig. 2), and incorrectly classified values were marked in red (Fig. 3).

For the testing dataset, all evaluation metrics exceeded 95% (Table 1). The model's capacity to distinguish between healthy and FI classes was assessed using the ROC curve, illustrating the trade-off between sensitivity and specificity. The AUROC was determined to be 0.99 (Fig. 4), demonstrating the model's excellent discriminatory ability.

Table 1. The comprehensive evaluation metrics for the ResNet-18 model on the test dataset

Metric	Value
Accuracy	0.96
Sensitivity	0.95
Specificity	0.98
Positive predictive value	0.97
Negative predictive value	0.96
Area under ROC curve	0.99

ROC: receiver operating curve

To better understand the model's performance and identify potential areas for improvement, a detailed analysis was conducted on the misclassified images from the testing set. The confusion matrix provided an in-depth look at the model's accuracy, indicating that the majority of images were correctly classified (as shown on the main diagonal). There were only 3 misclassification errors: 1 healthy image was incorrectly labeled as FI, and 2 FI images were mistakenly classified as healthy (Fig. 5).

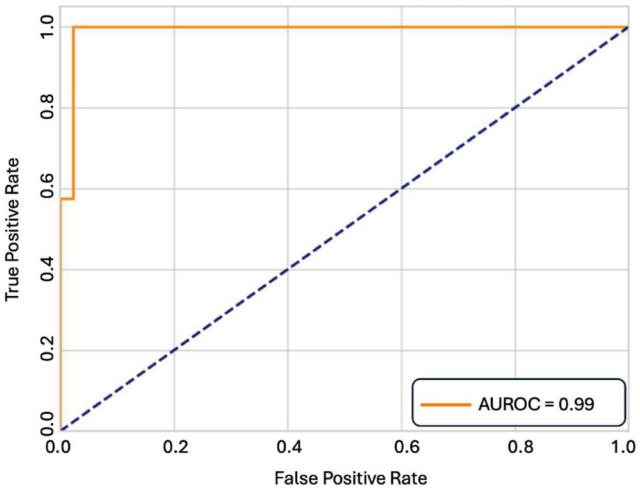

Fig. 4. Receiver operating characteristic curve demonstrating the model's balance between sensitivity and specificity, with an area under the receiver operating characteristic curve (AUROC) value of 0.99, indicating its ability to distinguish between healthy and furcation involvement (FI) cases.

Discussion

Molars with FI have a higher risk for tooth loss during supportive periodontal treatment.[15,29] The early diagnosis of FI is beneficial to achieve long-term survival of periodontally compromised teeth.[30] Clinically and radiographically, identifying FI at its initial stage presents significant challenges.[25] This study identified an optimized deep learning algorithm, ResNet-18, which demonstrated satisfactory performance in the identification of mandibular molar FI. This finding supports the beneficial role of AI in disease detection within a dental context.

Few studies have successfully identified furcation involvement with the assistance of AI. Mao et al.[26] utilized gray-level adjustment, Gaussian high-pass filtering, and adaptive thresholding to segment periapical radiographs. This approach enabled them to isolate each tooth in the radiographs and achieve an accuracy of 94.97% in detecting furcation involvement using the CNN learning model GoogLeNet.

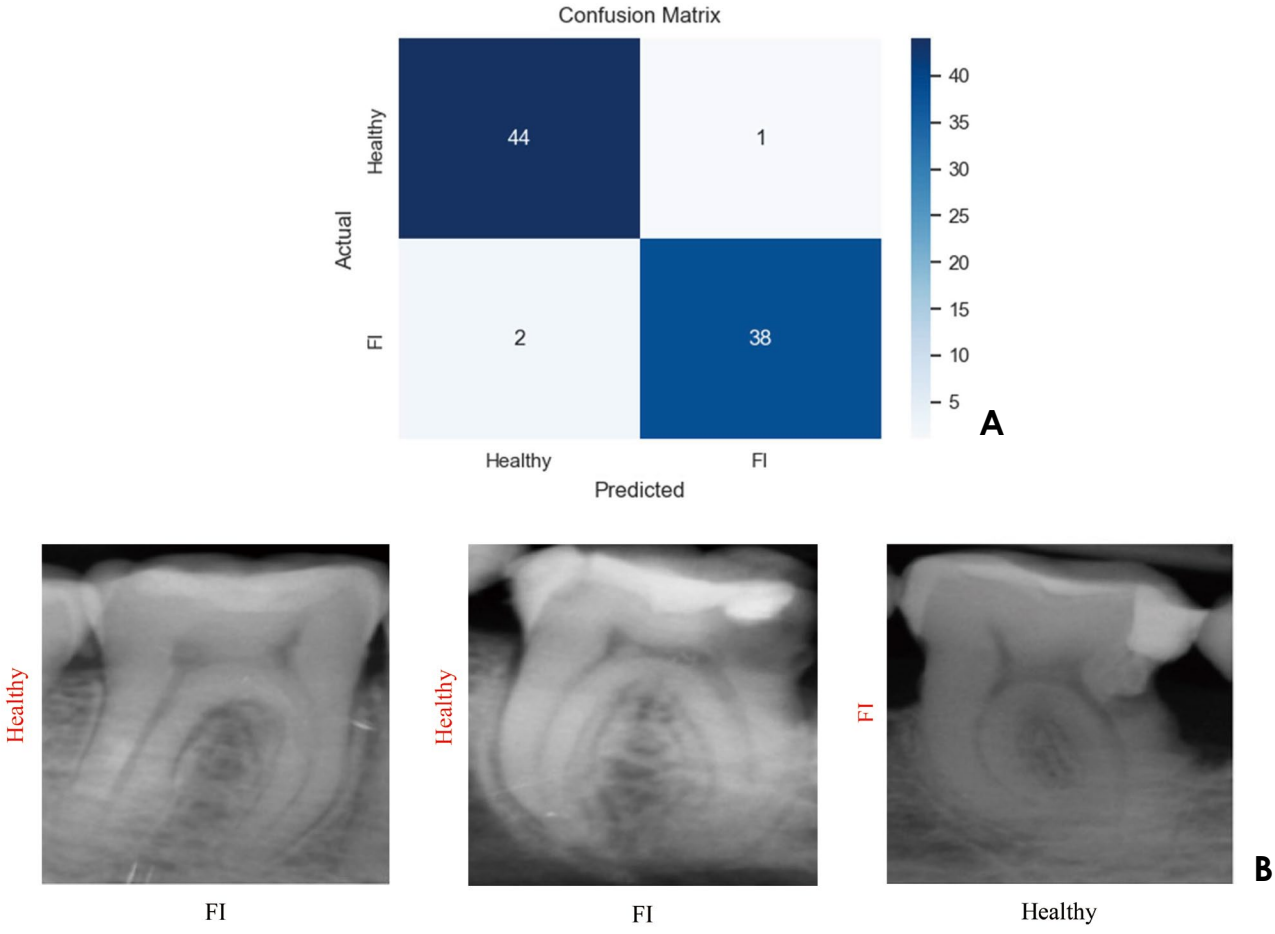

Fig. 5. A. A confusion matrix of the testing data shows that the majority of images are correctly classified, with only 1 healthy tooth misclassified as furcation involvement (FI) and 2 FI teeth misclassified as healthy. B. The 3 misclassified images. Predicted values are displayed on the vertical plane, and the actual ground truth is shown on the horizontal plane for comparison.

The present study attained an accuracy of 96.47% in identifying molar furcation involvement using ResNet-18 in the testing dataset. Collectively, these studies underscore the potential and practicality of employing AI to aid in periodontal diagnosis.

Radiographically, FI is diagnosed by the presence of a triangular radiolucency at the furca area and/or when the bone level falls below the furcation.[31] This model produced 3 misclassifications in the testing dataset, which presented diagnostic challenges even for experienced clinicians. Two cases of FI were incorrectly identified as healthy; both exhibited radiographic features very similar to those of healthy teeth, including a slightly widened periodontal ligament space at the furca and a small radiolucent bone marrow space nearby. The key distinguishing feature in these cases was the marginal alveolar bone level, which was slightly higher than the furcation, leading to their classification as healthy. In the third case, a healthy tooth was misclassified as FI. Here, the furca was filled with bone, but the bone level, particularly on the distal surface, was significantly lower than the furcation level. The radiographic appearance was so ambiguous that consultation with Axium notes was necessary to achieve definitive classifications. This underscores the complexity of disease diagnosis, a multifactorial process that involves not only imaging, which may show subtle changes, but also clinical examination, lab results, and other factors. To improve the model's performance, incorporating more challenging images into the system would be beneficial to improve its ability to differentiate between conditions.

The present study focused on the AI classification of mandibular molar FI. Detecting FI in maxillary molars using periapical radiographs presents a greater challenge, primarily because the palatal root is superimposed over the furca area, obscuring its visibility.[32] Including maxillary molar FI detection in future AI studies would provide a more comprehensive understanding of AI's benefits in diagnosing FI in both maxillary and mandibular molars.

Despite increasing interest in the application of AI in dental care, caution must be taken to avoid over-reliance on AI-generated information. Glick et al.[27] investigated the performance of dental students in identifying molar FI with or without AI assistance. The results indicated that third-year dental students were more prone to diagnostic errors due to their overreliance on AI-generated labels, compared to fourth-year dental students who had more advanced training. This finding highlights the psychological impact of the AI-clinician relationship in the diagnostic process, particularly for novice clinicians.

In conclusion, this study successfully identified and trained a deep learning model, ResNet-18, which demonstrated the capability to classify mandibular molar FI with satisfactory accuracy. The model's performance could be further enhanced by training it with additional data on incipient FI and maxillary molar FI, as well as incorporating an auto-segmentation function. While the implementation of AI in dental care holds significant potential, it should be regarded as a valuable supplementary tool rather than a replacement for clinicians' critical reasoning and decision-making.

Conflicts of Interest: None

References

1. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Ann Surg 2018; 268: 70-6.
2. Wong SH, Al-Hasani H, Alam Z, Alam A. Artificial intelligence in radiology: how will we be affected? Eur Radiol 2019; 29: 141-3.
3. Sklan JE, Plassard AJ, Fabbri D, Landman BA. Toward content based image retrieval with deep convolutional neural networks. Proc SPIE Int Soc Opt Eng 2015; 9417: 94172C.
4. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol 2018; 15: 504-8.
5. Liu Y, Balagurunathan Y, Atwater T, Antic S, Li Q, Walker RC, et al. Radiological image traits predictive of cancer status in pulmonary nodules. Clin Cancer Res 2017; 23: 1442-9.
6. Schuhbaeck A, Otaki Y, Achenbach S, Schneider C, Slomka P, Berman DS, et al. Coronary calcium scoring from contrast coronary CT angiography using a semiautomated standardized method. J Cardiovasc Comput Tomogr 2015; 9: 446-53.
7. Arimura H, Li Q, Korogi Y, Hirai T, Katsuragawa S, Yamashita Y, et al. Computerized detection of intracranial aneurysms for three-dimensional MR angiography: feature extraction of small protrusions based on a shape-based difference image technique. Med Phys 2006; 33: 394-401.
8. Wang S, Yao J, Summers RM. Improved classifier for computer-aided polyp detection in CT colonography by nonlinear dimensionality reduction. Med Phys 2008; 35: 1377-86.
9. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent 2018; 77: 106-11.
10. Lin PL, Huang PW, Huang PY, Hsu HC. Alveolar bone-loss area localization in periodontitis radiographs based on threshold segmentation with a hybrid feature fused of intensity and the H-value of fractional Brownian motion model. Comput Methods Programs Biomed 2015; 121: 117-26.
11. Lin PL, Huang PY, Huang PW. Automatic methods for alveolar bone loss degree measurement in periodontitis periapical radiographs. Comput Methods Programs Biomed 2017; 148: 1-11.
12. Okada K, Rysavy S, Flores A, Linguraru MG. Noninvasive

differential diagnosis of dental periapical lesions in cone-beam CT scans. Med Phys 2015; 42: 1653-65.
13. Abdolali F, Zoroofi RA, Otake Y, Sato Y. Automated classification of maxillofacial cysts in cone beam CT images using contourlet transformation and spherical harmonics. Comput Methods Programs Biomed 2017; 139: 197-207.
14. Yilmaz E, Kayikcioglu T, Kayipmaz S. Computer-aided diagnosis of periapical cyst and keratocystic odontogenic tumor on cone beam computed tomography. Comput Methods Programs Biomed 2017; 146: 91-100.
15. Nibali L, Zavattini A, Nagata K, Di Iorio A, Lin GH, Needleman I, et al. Tooth loss in molars with and without furcation involvement - a systematic review and meta-analysis. J Clin Periodontol 2016, 43: 156-66.
16. Goodson JM. Antimicrobial strategies for treatment of periodontal diseases. Periodontol 2000 1994; 5: 142-68.
17. McFall WT Jr. Tooth loss in 100 treated patients with periodontal disease. A long-term study. J Periodontol 1982; 53: 539-49.
18. Svärdström G, Wennström JL. Prevalence of furcation involvements in patients referred for periodontal treatment. J Clin Periodontol 1996; 23: 1093-9.
19. Graetz C, Plaumann A, Wiebe JF, Springer C, Sälzer S, Dörfer CE. Periodontal probing versus radiographs for the diagnosis of furcation involvement. J Periodontol 2014; 85: 1371-9.
20. Shaker ZMH, Parsa A, Moharamzadeh K. Development of a radiographic index for periodontitis. Dent J (Basel) 2021; 9: 19.
21. Abbas F, Hart AA, Oosting J, van der Velden U. Effect of training and probing force on the reproducibility of pocket depth measurements. J Periodontal Res 1982; 17: 226-34.
22. Brägger U. Radiographic parameters: biological significance and clinical use. Periodontol 2000 2005; 39: 73-90.
23. Eickholz P. Reproducibility and validity of furcation measurements as related to class of furcation invasion. J Periodontol 1995; 66: 984-9.
24. Eickholz P, Hausmann E. Accuracy of radiographic assessment of interproximal bone loss in intrabony defects using linear measurements. Eur J Oral Sci 2000; 108: 70-3.
25. Müller HP, Eger T. Furcation diagnosis. J Clin Periodontol 1999; 26: 485-98.
26. Mao YC, Huang YC, Chen TY, Li KC, Lin YJ, Liu YL, et al. Deep learning for dental diagnosis: a novel approach to furcation involvement detection on periapical radiographs. Bioengineering (Basel) 2023; 10: 802.
27. Glick A, Clayton M, Angelov N, Chang J. Impact of explainable artificial intelligence assistance on clinical decision-making of novice dental clinicians. JAMIA Open 2022; 5: ooac031.
28. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. J Med Imaging Radiat Oncol 2021; 65: 545-63.
29. Graetz C, Schützhold S, Plaumann A, Kahl M, Springer C, Sälzer S, et al. Prognostic factors for the loss of molars - an 18-years retrospective cohort study. J Clin Periodontol 2015; 42: 943-50.
30. Reddy MS, Aichelmann-Reidy ME, Avila-Ortiz G, Klokkevold PR, Murphy KG, Rosen PS, et al. Periodontal regeneration - furcation defects: a consensus report from the AAP Regeneration Workshop. J Periodontol 2015; 86(2 Suppl): S131-3.
31. Jolivet G, Huck O, Petit C. Evaluation of furcation involvement with diagnostic imaging methods: a systematic review. Dentomaxillofac Radiol 2022; 51: 20210529.
32. Alasqah M, Alotaibi FD, Gufran K. The radiographic assessment of furcation area in maxillary and mandibular first molars while considering the new classification of periodontal disease. Healthcare (Basel) 2022; 10: 1464.