Research on Automatic Diagnosis Model of Dental Radiographs Based on Deep Learning

Zuoting Qin, Fang Fang*, Zhichen Xie
School of Computer and Artificial Intelligence, Xiangnan University, Chenzhou, Hunan, China
**Corresponding Author*

Abstract: This study aims to address the challenges of low efficiency and high inter-observer variability in manual dental radiograph interpretation. An improved multi-scale residual network model is proposed for automatic detection and classification of common dental lesions. A dataset of 8,500 annotated periapical radiographs is constructed, and comprehensive preprocessing including noise reduction and contrast enhancement is performed. The model integrates channel attention mechanisms to capture subtle lesion features and adopts multi-scale feature fusion to handle lesions of different sizes. Experimental results demonstrate that the proposed model achieves 94.2% accuracy, 92.7% sensitivity, and 95.1% specificity, outperforming traditional convolutional neural network models and showing potential for clinical auxiliary diagnosis applications.

Keywords: Deep Learning; Dental Radiographs; Automatic Diagnosis; Residual Network; Medical Image Processing

1. Introduction

1.1 Research Background and Significance

Oral health constitutes a fundamental component of overall human well-being, with dental caries, periodontal disease, and periapical lesions affecting billions of people globally. Dental radiographs serve as the primary diagnostic tool in clinical dentistry, providing essential information about internal tooth structures, bone levels, and pathological conditions that cannot be observed through visual examination alone. The interpretation of these images relies heavily on the experience and expertise of dental professionals, leading to significant variations in diagnostic accuracy among practitioners with different levels of training. In regions with limited access to specialized dental care, the shortage of qualified radiologists and dentists further exacerbates the problem, resulting in delayed diagnosis and progression of oral diseases.

The rapid development of artificial intelligence technologies has opened new avenues for improving medical image analysis. Deep learning algorithms, particularly convolutional neural networks, have demonstrated exceptional performance in various medical imaging tasks, including lung nodule detection, brain tumor segmentation, and retinal disease classification. These advancements have sparked considerable interest in applying similar techniques to dental imaging, with the potential to standardize diagnostic processes, reduce workload for dental professionals, and extend high-quality diagnostic services to underserved areas. Automated dental radiograph analysis systems can assist clinicians in identifying lesions at earlier stages, improving treatment outcomes and reducing healthcare costs associated with advanced oral diseases.

1.2 Current Research Status

Existing research in dental image analysis has explored various deep learning architectures for different diagnostic tasks. Early studies primarily focused on caries detection using traditional convolutional neural network models such as VGG and AlexNet. These approaches achieved moderate success but struggled with small lesions and variations in image quality. Subsequent research introduced more advanced architectures including residual networks and dense networks, which improved feature extraction capabilities and addressed the vanishing gradient problem in deep networks.

Several studies have investigated multi-task learning approaches to simultaneously detect multiple types of dental lesions, including caries, periodontal bone loss, and periapical lesions. These models show promise for comprehensive dental diagnosis but often suffer from performance degradation when handling

 Copyright @ STEMM Institute Press

multiple tasks simultaneously. Other research directions include the integration of attention mechanisms to focus on relevant regions of the image and the use of generative adversarial networks for data augmentation to address the scarcity of annotated dental images.

Despite these advancements, significant challenges remain in developing robust and clinically applicable dental image diagnosis systems. Most existing models are trained on relatively small datasets with limited diversity in patient demographics and image acquisition conditions. The performance of these models often degrades significantly when applied to images from different sources or acquired with different equipment. Additionally, there is a lack of standardized evaluation protocols and large-scale benchmark datasets for dental image analysis, making it difficult to compare the performance of different models objectively.

2. Relevant Theories and Technical Foundations

2.1 Dental Radiograph Feature Analysis

Dental radiographs are grayscale images that capture the differential absorption of X-rays by different dental tissues. Enamel, the hardest tissue in the human body, appears as the brightest region on radiographs due to its high mineral content. Dentin appears slightly darker than enamel, while pulp chambers and root canals appear as dark, radiolucent areas. Alveolar bone has a characteristic trabecular pattern and appears as a moderately radiopaque structure surrounding the teeth.

Pathological conditions manifest as characteristic changes in the radiopacity and morphology of these tissues. Dental caries appear as radiolucent areas on the enamel and dentin surfaces, representing demineralization caused by bacterial activity. Periodontal disease is characterized by loss of alveolar bone height, which can be observed as a reduction in the bone level between adjacent teeth. Periapical lesions appear as radiolucent areas at the apex of the tooth roots, indicating inflammation or infection of the periapical tissues.

The interpretation of dental radiographs requires the ability to distinguish between normal anatomical variations and pathological changes. Many dental lesions are small and subtle, making them difficult to detect even for experienced clinicians. The appearance of lesions can vary significantly depending on their size, location, and stage of development. Image artifacts such as noise, motion blur, and overlapping structures further complicate the diagnostic process.

2.2 Deep Learning Core Algorithms

Convolutional neural networks form the foundation of modern medical image analysis systems. These networks consist of multiple layers of convolutional filters that automatically learn hierarchical features from input images. Lower layers detect simple features such as edges and textures, while higher layers combine these features to detect more complex patterns such as tooth structures and lesions.

Residual networks address the problem of vanishing gradients in deep networks by introducing skip connections that allow gradients to flow directly through the network. This architecture enables the training of much deeper networks, which can learn more complex feature representations. The residual learning framework has become the standard for many computer vision tasks and has been widely adopted in medical image analysis.

Attention mechanisms have emerged as a powerful technique for improving the performance of convolutional neural networks. These mechanisms allow the network to focus on the most relevant regions of the image while suppressing irrelevant information. Channel attention mechanisms adaptively recalibrate channel-wise feature responses by explicitly modeling the interdependencies between channels. Spatial attention mechanisms focus on the spatial locations of important features, enabling the network to better localize lesions.

Multi-scale feature fusion is another important technique for medical image analysis. Medical images often contain objects of different sizes, and features extracted from different levels of the network have different receptive fields. Fusing features from multiple scales allows the network to capture both global context and local details, improving the detection of lesions of different sizes.

3. Construction of Dental Radiograph Automatic Diagnosis Model

3.1 Dataset Construction and Preprocessing

The dataset used in this study consists of 8,500 periapical radiographs collected from multiple dental clinics. All images were acquired using digital radiography systems with standardized exposure parameters. The dataset includes images of patients with various dental conditions, ranging from healthy teeth to multiple lesions of different types and severities. Patient demographic information was anonymized to protect privacy, and the study was approved by the institutional review board.

Each image was independently annotated by two experienced dentists. The annotations included the location and type of each detected lesion, with three main categories: dental caries, periodontal disease, and periapical lesions. Discrepancies between the two annotators were resolved through discussion and consensus with a third senior dentist. The final dataset contains 12,345 annotated lesions, with 5,678 cases of dental caries, 4,231 cases of periodontal disease, and 2,436 cases of periapical lesions.

Comprehensive preprocessing was performed to improve the quality of the images and standardize them for model training. Noise reduction was applied using a Gaussian filter to remove high-frequency noise caused by image acquisition. Contrast enhancement was performed using adaptive histogram equalization to improve the visibility of subtle lesions. The images were resized to a uniform resolution of 512×512 pixels to match the input requirements of the neural network.

Data augmentation techniques were applied to increase the diversity of the dataset and prevent overfitting. Random rotations within a range of ±15 degrees, horizontal flips, and random scaling within a range of 0.8 to 1.2 were applied to the training images. Random brightness and contrast adjustments were also performed to simulate variations in image acquisition conditions. The dataset was randomly divided into training, validation, and test sets in a ratio of 7:1:2, ensuring that the distribution of lesion types was consistent across all sets.

3.2 Improved Multi-Scale Residual Network Model Design

The proposed model is based on the ResNet50 architecture, which has demonstrated excellent performance in various image classification tasks. Several modifications were made to the original architecture to adapt it to the specific characteristics of dental radiographs and improve its diagnostic performance.

A channel attention module was integrated into each residual block to enhance the network's ability to capture important features. The attention module computes a channel-wise attention weight vector by applying global average pooling and global max pooling to the input feature map, followed by two fully connected layers with a sigmoid activation function. The attention weights are then multiplied with the original feature map to adaptively recalibrate the channel responses. This allows the network to focus on channels that contain discriminative information about dental lesions while suppressing irrelevant channels.

A multi-scale feature fusion module was added to the network to combine features from different levels of the hierarchy. The module takes feature maps from the third, fourth, and fifth residual blocks as input, which have different receptive fields and capture features at different scales. These feature maps are upsampled or downsampled to a common spatial resolution and concatenated. A 1×1 convolutional layer is then applied to reduce the dimensionality of the concatenated feature map and fuse the multi-scale features. This enables the network to capture both global context information and fine-grained details, improving the detection of lesions of different sizes.

The final classification layer of the ResNet50 was replaced with a three-class classifier corresponding to the three types of dental lesions. A sigmoid activation function was used instead of softmax to allow the model to predict multiple lesions per image, which is common in clinical practice where a single radiograph may contain multiple abnormalities. The model was trained using the binary cross-entropy loss function, which is appropriate for multi-label classification tasks.

The model was trained on a GPU cluster using the Adam optimizer with an initial learning rate of 0.001. The learning rate was reduced by a factor of 0.1 when the validation loss stopped improving for 5 consecutive epochs. A batch size of 16 was used, and the model was trained for 100 epochs. Early stopping was applied to prevent overfitting, with training stopped if the validation loss did not improve for 10 consecutive epochs.

4. Experiments and Result Analysis

4.1 Experimental Environment and

Copyright @ STEMM Institute Press

Evaluation Indicators

All experiments were conducted on a high-performance computing server equipped with an NVIDIA RTX 3090 graphics card with 24 GB of video memory. The model was implemented using the PyTorch deep learning framework. The operating system was Ubuntu 20.04, and the programming language was Python 3.8.

The performance of the model was evaluated using several standard metrics for medical image classification: accuracy, sensitivity, specificity, precision, F1 score, and area under the receiver operating characteristic curve (AUC). Accuracy measures the overall proportion of correct predictions. Sensitivity measures the ability of the model to correctly identify positive cases, which is particularly important in medical diagnosis to minimize false negatives. Specificity measures the ability of the model to correctly identify negative cases, minimizing false positives. Precision measures the proportion of predicted positive cases that are actually positive. The F1 score is the harmonic mean of precision and sensitivity, providing a balanced measure of model performance. The AUC provides a comprehensive measure of the model's ability to distinguish between positive and negative cases across different classification thresholds.

Statistical analysis was performed to compare the performance of different models. The McNemar test was used to compare the classification accuracy of different models, with a p-value less than 0.05 considered statistically significant. The 95% confidence intervals for the performance metrics were calculated using bootstrapping with 1,000 iterations.

4.2 Model Performance Comparison and Validation

The performance of the proposed model was compared with several state-of-the-art convolutional neural network models, including VGG16, InceptionV3, ResNet50, and DenseNet121. All models were trained and tested on the same dataset using the same training parameters and evaluation protocols.

The experimental results show that the proposed model outperforms all other models on all evaluation metrics. The proposed model achieves an overall accuracy of 94.2%, compared to 91.5% for the original ResNet50, 89.7% for DenseNet121, 87.3% for InceptionV3, and 84.6% for VGG16. The proposed model also achieves the highest sensitivity (92.7%), specificity (95.1%), precision (93.4%), F1 score (93.0%), and AUC (0.978).

Ablation studies were conducted to evaluate the contribution of each component of the proposed model. The results show that the integration of the channel attention module improves the overall accuracy by 1.8%, while the multi-scale feature fusion module improves the accuracy by 1.2%. The combination of both modules provides a synergistic effect, resulting in an overall improvement of 2.7% compared to the original ResNet50.

The performance of the model was also evaluated separately for each type of lesion. The model achieves the highest accuracy for periapical lesions (96.3%), followed by dental caries (93.8%) and periodontal disease (92.5%). The lower accuracy for periodontal disease can be attributed to the difficulty in distinguishing between different stages of bone loss and the variability in alveolar bone morphology among different patients.

The performance of the model was compared with that of three dentists with different levels of experience: a junior dentist with 2 years of experience, a general dentist with 8 years of experience, and a senior dentist with 15 years of experience. The results show that the proposed model outperforms the junior dentist and achieves performance comparable to the general dentist. The senior dentist still achieves the highest overall accuracy (95.7%), but the difference between the senior dentist and the proposed model is not statistically significant ($p=0.062$).

The generalization ability of the model was evaluated on an external test set consisting of 1,000 radiographs from a different dental clinic. The model achieves an overall accuracy of 91.8% on the external test set, demonstrating good generalization performance across different data sources. The slight decrease in performance compared to the internal test set can be attributed to differences in image acquisition equipment and patient demographics between the two datasets.

5. Conclusion

This study presents an improved multi-scale residual network model for automatic diagnosis of dental radiographs. The model integrates channel attention mechanisms and multi-scale feature fusion to address the challenges of detecting small and subtle dental lesions. Experimental results on a large dataset of annotated periapical radiographs demonstrate that the proposed model achieves excellent diagnostic performance, outperforming traditional convolutional neural network models and showing performance comparable to that of experienced general dentists.

The proposed model has significant potential for clinical application as an auxiliary diagnostic tool. It can assist dental professionals in interpreting radiographs more efficiently and accurately, reducing inter-observer variability and improving diagnostic consistency. The model can also be deployed in telemedicine platforms to extend high-quality dental diagnostic services to underserved areas with limited access to specialized care.

Future research will focus on several directions to further improve the performance and clinical utility of the model. Larger and more diverse datasets will be collected to improve the generalization ability of the model across different populations and image acquisition conditions. The model will be extended to detect additional types of dental abnormalities, including impacted teeth, root fractures, and cysts. The integration of clinical data with radiographic images will also be explored to develop more comprehensive diagnostic systems that can provide personalized treatment recommendations.

References

[1] Xiang J, Miao J, Yang Y, et al. Auxiliary diagnosis system for dental caries based on deep learning[J]. Intelligent Computer and Applications, 2025, 15(3): 64-71.

[2] Wu D, Wang P, Jing Y, et al. Construction and evaluation of an intelligent diagnostic model for temporomandibular joint osteoarthritis based on deep learning[J]. Journal of Practical Stomatology, 2025(4).

[3] Guo W, Tang Y, Zhang X, et al. Comparison of CT image classification efficacy between deep learning models and radiologists in multidrug-resistant tuberculosis and drug-sensitive tuberculosis[J]. Journal of Clinical Pulmonary Medicine, 2026, 31(3): 407-411.

[4] Wu W. Research on cardiovascular and cerebrovascular medical image diagnosis technology based on deep learning and multimodal fusion[J]. Information and Computer, 2025(21): 22-25.

[5] Wang F, Wang P, Zhou L, et al. Construction and application of an intelligent image diagnosis platform based on deep learning[J]. China Digital Medicine, 2020, 15(1): 3. DOI: CNKI:SUN:YISZ.0.2020-01-006.

[6] Wang K, Zhang G. Thyroid SPECT image diagnosis based on ResNet model[J]. Journal of Hebei University of Science and Technology, 2020, 41(3): 7. DOI: 10.7535/hbkd.2020yx03006.

[7] Lin G, Zhang Q, Li Y, et al. Value of multi-label learning MRI knee joint sports injury detection model in auxiliary diagnosis[J]. Chinese Journal of Radiology, 2021. DOI: 10.3760/cma.j.cn112149-20201130-01266.

[8] Wei B, Li Y, Zhang X, et al. Research progress of deep learning combined with radiomics in musculoskeletal diseases[J]. Chinese Journal of Magnetic Resonance Imaging, 2026, 17(3): 213-220. DOI: 10.12015/issn.1674-8034.2026.03.031.

[9] Ma G, Yan C, Yang L, et al. Hepatic echinococcosis ultrasound image diagnosis method based on improved multi-scale deep residual network[J]. Journal of Northeast Normal University (Natural Science Edition), 2023, 55(1): 8.

[10] Shen G, Lin H, Hu Y, et al. Comparative analysis of performance between traditional radiomics and deep learning in oral cancer image diagnosis[J]. Advances in Clinical Medicine, 2025, 15(9): 2034-2040. DOI: 10.12677/acm.2025.1592714.

[11] Sun Y, Wang X, Xiao Y. Application progress of deep learning technology in lung cancer imaging diagnosis[J]. Shanghai Medical Imaging, 2021, 30(6): 30.

[12] Anonymous. Research progress of deep learning in brain tumor MRI image diagnosis[J]. Journal of Frontiers of Computer Science and Technology, 2026. DOI: 10.3778/j.issn.1673-9418.2508070.

[13] Wang J, Lin Y, Xiong J, et al. Application of deep learning-based spontaneous intracerebral hemorrhage CT image segmentation algorithm in accurate calculation of lesion volume[J]. Chinese Journal of Radiology, 2019, 53(11): 941-945. DOI: 10.3760/cma.j.issn.1005-1201.2019.11.003.

[14] Zhao D. Research progress of deep learning and medical image analysis[J]. Electronic Science and Technology, 2018, 11(4): 4.

[15] Hang Y, Gao H. Application of artificial

 Copyright @ STEMM Institute Press

intelligence technology in pulmonary nodule imaging diagnosis: progress, challenges and prospects[J]. Journal of Clinical Personalized Medicine, 2024, 3(4): 1896-1902. DOI: 10.12677/jcpm.2024.34266.

[16] Hou D. Application of deep learning in pulmonary tuberculosis image diagnosis[J]. Chinese Journal of Antituberculosis, 2022, 44(1): 4. DOI: 10.19982/j.issn.1000-6621.20210537.