Automated classification of oral diseases on dental radiographs using pretrained Convolutional Neural Network models

Raúl Humberto Moguel Dominguez [1] and Ismael Eliezer Pérez Ruiz [2]

[1,2] Universidad Modelo, Yucatán 97305, MEX

15222609@modelo.edu.mx

Abstract. Dental diseases such as caries, bone loss, and periapical lesions are among the most common oral health issues, and their early detection is essential to avoid long-term complications. Due to the visual similarity of these conditions in dental radiographs, accurate diagnosis can be challenging, even for experienced professionals. This study explores the application of convolutional neural networks for the automated classification of dental radiographs into four categories: healthy, caries, bone loss, and periapical lesion. Transfer learning was employed using four state-of-the-art pretrained convolutional neural networks architectures: InceptionV3, ResNet50, EfficientNetB0, and MobileNetV2. A custom dataset was compiled and preprocessed, although limited image availability posed a major constraint during training. The models were adapted by incorporating task-specific classification layers aimed at identifying dental conditions. To ensure a fair comparison, all models were trained using a standardized pipeline, with identical preprocessing procedures, input resolution, batch size, and training settings. Experimental results revealed that InceptionV3 achieved the highest validation accuracy and consistency, outperforming the other models, which showed greater sensitivity to data scarcity. Performance was further evaluated using accuracy and loss along with confusion matrices, confirming InceptionV3's superior class-wise performance and more stable training behavior. Despite the limited dataset, the findings highlight the feasibility of integrating convolutional neural networks-based diagnostic tools in dental screening workflows, while emphasizing the need for larger, high-quality radiographic datasets for robust performance.

Keywords: Convolutional Neural Network, Oral diseases, Automated classification

1 Introduction

Oral diseases remain a major public health issue worldwide, affecting people across all ages. Among the most prevalent conditions are dental caries, bone loss, and periapical lesions, each of which can cause significant pain, tooth loss, and reduced quality of life if not detected and treated early [1], [2]. Dental caries result from the demineralization of enamel due to bacterial acid production, while bone loss commonly occurs as a consequence of periodontal disease, leading to the destruction of alveolar bone [3]. Periapical lesions develop as inflammatory responses to infections in the dental pulp, frequently necessitating endodontic intervention [4]. Although these diseases differ in etiology, they often exhibit highly similar visual features in dental radiographs, making it difficult to distinguish among them, even for experienced clinicians [5]. Radiographic imaging remains the primary diagnostic tool, yet overlapping appearances and subtle morphological differences pose a diagnostic challenge.

Convolutional neural networks (CNNs) have recently gained attention in medical image analysis due to their ability to autonomously learn complex spatial hierarchies and patterns from raw images, improving diagnostic accuracy [6]. Pretrained CNN architectures like InceptionV3, ResNet50, EfficientNetB0, and MobileNetV2 have been successfully applied in various domains through transfer learning, which helps to mitigate data scarcity by leveraging knowledge from large datasets [7].

Nevertheless, despite the growing use of CNNs in medical imaging, their application to dental radiographic diagnosis is still limited, with few studies conducting comprehensive comparisons of multiple CNN models for the classification of common dental diseases [8]. This study aims to address this gap by evaluating the performance of four leading CNN architectures on a custom dental radiograph dataset.

2 Methodology

2.1 Dataset

Two publicly available datasets from Kaggle were used for this study. The first, the X-Ray Panoramic Dataset, contains panoramic dental radiographs labeled with over 30 different conditions. From this dataset, only images corresponding to the three target pathologies dental caries, bone loss, and periapical lesions were extracted. The second dataset, the Dental OPG X-ray Dataset, was used to obtain images of healthy teeth.

Due to data imbalance and availability issues, the dataset had to be constrained. Although over 1000 images were available for the pathological classes, only 223 images of healthy teeth were found. To maintain balance, a subset of the pathological images was randomly selected to match the number of healthy samples, resulting in a total dataset of 892 images (223 for each class).

2.2 Pre-processing

To ensure compatibility with the selected convolutional neural network (CNN) architectures, all radiographic images were first resized to 224×224 pixels. This resolution is the standard input size required by most pretrained. Given that the radiographs in the datasets were grayscale, but the pretrained models expect three-channel RGB images, all images were converted to RGB format. Additionally, pixel values were normalized by scaling the intensity values to the range [0,1], which helps improve convergence speed and model stability during training.

To address the class imbalance where the healthy class was significantly underrepresented data augmentation was applied on EfficientNetB0 and MobileNetV2 during training. The augmentation included random rotations, horizontal flips, zooming, and slight translations. The final dataset was split into training (75%), validation (10%), and testing (15%) sets using a stratified split to preserve class distribution across all subsets.

2.3 CNN Architectures

The choice of convolutional neural network (CNN) architectures in this study was guided by their proven performance in image classification tasks. Four architectures were evaluated: InceptionV3, ResNet50, EfficientNetB0, and MobileNetV2. InceptionV3 employs parallel convolutional operations to extract multi-scale features efficiently, making it particularly suitable for detecting subtle patterns in medical images. ResNet50 introduces residual connections that ease the training of deeper networks by mitigating vanishing gradients. EfficientNetB0 balances network depth, width, and resolution using compound scaling, optimizing performance with fewer parameters. MobileNetV2 is designed for lightweight applications, using depth wise separable convolutions to reduce complexity.

2.4 Training

To enable the use of pretrained convolutional neural networks (CNNs) while adapting them to the classification of dental radiographs, a transfer learning approach was employed. For each the convolutional base was balanced with the same weights, and a number of layers were frozen to preserve previously learned low-level features. Specifically, 146 layers were frozen in ResNet50, 145 in MobileNetV2, 200 in EfficientNetB0, and 300 in InceptionV3.

On top of each frozen base, a custom classification head was added, consisting of a Global Average Pooling layer to reduce dimensionality, followed by one or more Dense layers and a final output layer with a softmax activation for multi-class prediction across the four target classes: healthy, caries, bone loss, and periapical lesion.

The dataset was divided into 14 batches with a batch size of 64, resulting in 11 batches used for training, 1 for validation, and 2 for testing. The models were trained using the Adam optimizer and categorical cross-entropy loss, with early stopping based on validation accuracy to prevent overfitting.

3 Results

The goal is to compare the performance of the four architectures on the small dataset of dental radiographies using fine-tuning and to assess their generalization capacity based on standard evaluation metrics: accuracy, precision, recall, F1-score, and confusion matrix.

During training, ResNet50 showed progressive improvement on the training set, reaching a maximum accuracy of 69.4% by epoch 24. However, the validation loss increased significantly during the final epochs, while validation accuracy fluctuated peaking at just 60.9% indicating clear overfitting. The classification report revealed a weighted F1-score of only 0.12, with all predictions concentrated in the “Healthy” class, leaving the other three categories unrecognized.

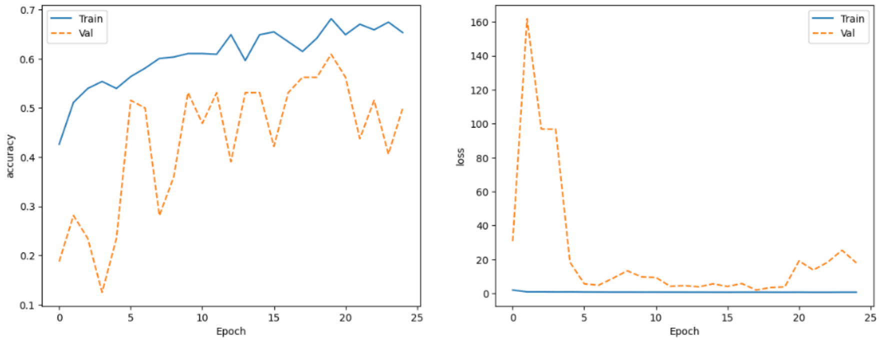

Figure 1. Accuracy and loss graph of the train performance for ResNet50. Source: Own elaboration

EfficientNetB0, although faster to train, demonstrated similar results. Training accuracy peaked at 42.9%, and validation loss rose steadily to 13.3. Validation accuracy remained low (up to 39) the model predicted only one class during testing. The weighted F1-score was again 0.12.

MobileNetV2, however, showed the best relative performance. Despite not having the highest training accuracy (which peaked at 56.9%), it exhibited greater stability during validation, achieving a maximum validation accuracy of 59.4% with a more controlled loss curve. On the test set, MobileNet reached an accuracy of 50.8%. The weighted F1-score was 0.36, significantly better than the others.

In contrast, InceptionV3 achieved the best overall performance among the evaluated architectures. During training, it reached a maximum accuracy of 80.4%, with a validation accuracy of 85.9% and a validation loss of 0.3795—indicative of strong convergence and generalization. On the test set, it achieved an outstanding accuracy of 93%, with a weighted F1-score of 0.93.

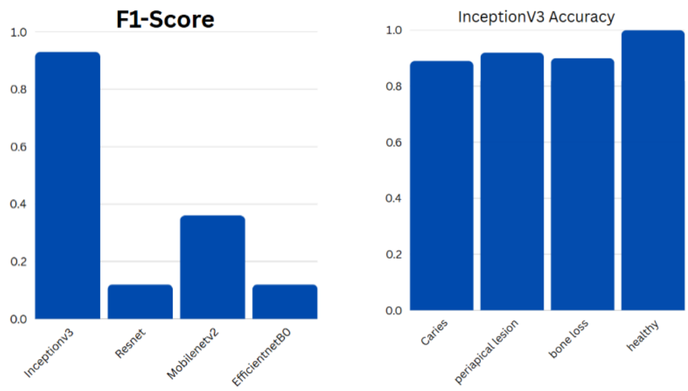

Figure 2, 3. Global F1-Score for InceptionV3, ResNet50, EfficientNetB0, and MobileNetV2. InceptionV3 individual class accuracy. Source: Own elaboration

4 Discussion

In this study, InceptionV3 achieved superior performance in classifying dental radiographs, with an accuracy of 93% and a weighted F1-score of 0.93. This aligns with findings who reported high accuracy using CNN-based models for classifying deciduous and permanent teeth from panoramic images [9].

MobileNetV2, while more lightweight, attained a test accuracy of 50.8% and a weighted F1-score of 0.36. Despite its lower performance compared to InceptionV3, its efficiency makes it suitable for applications with limited computational resources [10]. ResNet50 and EfficientNetB0 exhibited lower accuracies of 27% and 39%, respectively. These results may be attributed to overfitting and sensitivity to data imbalance [11]. Data augmentation was essential to address class imbalance, particularly given the underrepresentation of healthy samples, and it follows established practices to enhance model generalization in medical imaging [12] [13].

These findings highlight the relevance of model selection based on dataset size, balance, and diagnostic complexity. InceptionV3 emerges as a promising architecture for automatic dental diagnosis from radiographs, particularly in identifying early-stage pathologies such as caries or periapical lesions [14] [15].

5 Conclusion

This study demonstrates the effectiveness of deep learning models, particularly convolutional neural networks, in the automatic classification of dental radiographs for the detection of pathologies such as caries and periapical lesions. By leveraging pretrained architectures like EfficientNetB0 and MobileNetV2, combined with appropriate preprocessing steps such as image resizing, RGB conversion, and normalization we ensured model compatibility and enhanced training performance

The results highlight the importance of selecting the appropriate CNN architecture, as different models vary in complexity, accuracy, and computational cost. EfficientNetB0, for instance, provided a balance between accuracy and efficiency.

The study also highlighted the importance of adequate preprocessing steps and data augmentation to address class imbalance and improve model generalization. These combined strategies ensured that the models, especially InceptionV3, could effectively learn from the limited and heterogeneous dataset.

The success of InceptionV3 can be attributed to its unique architectural design, which balances depth and computational efficiency through its inception modules, allowing for multi-scale feature extraction. This advantage was clearly reflected in the model's superior validation accuracy (85.94%) and strong classification metrics across all classes, including caries, periapical lesions, and healthy cases.

Overall, the results confirm that carefully selecting the right CNN architecture, like InceptionV3, is critical for achieving high diagnostic accuracy in medical image analysis tasks. Future work should focus on expanding datasets for a better performance.

References

1. J. R. Petersen, “Global burden of oral diseases: emerging concepts, management and integration with general health,” International Dental Journal, vol. 68, no. 2, pp. 67–75, Apr. 2018 https://onlinelibrary.wiley.com/doi/10.1111/idj.12322
2. A. Tonetti et al., “Periapical lesions: epidemiology, diagnosis and treatment,” Journal of Endodontics, vol. 44, no. 5, pp. 559–566, May 2018 https://www.sciencedirect.com/science/article/pii/S0099239918300143
3. P. K. Dash, R. Mishra, and A. Jena, “A comprehensive review on dental caries and its management,” Journal of Conservative Dentistry, vol. 22, no. 4, pp. 345–353, Jul. 2019. https://www.jcd.org.in/article.asp?issn=0972-0707;year=2019;volume=22;issue=4;spage=345;epage=353;aulast=Dash
4. L. Litjens et al., “A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017. https://www.sciencedirect.com/science/article/pii/S1361841517301135
5. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
6. C. Szegedy et al., “Rethinking the inception architecture for computer vision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 2818–2826. https://openaccess.thecvf.com/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html
7. M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *arXiv preprint arXiv:1905.11946*, 2019. https://arxiv.org/abs/1905.11946
8. F. Azam, F. Majeed, and A. Nisar, “A systematic review on the application of deep learning techniques in dental radiographs,” *Computers in Biology and Medicine*, vol. 143, p. 105247, 2022. https://www.sciencedirect.com/science/article/pii/S0010482522002992
9. Soni, S., et al., "Deep Learning Models for Classification of Deciduous and Permanent Teeth From Digital Panoramic Images," *Cureus*, vol. 15, no. 1, 2023. https://www.cureus.com/articles/198261-deep-learning-models-for-classification-of-deciduous-and-permanent-teeth-from-digital-panoramic-images
10. I. D. S. Chen et al., "Deep Learning-Based Recognition of Periodontitis and Dental Caries in Dental X-ray Images," *Bioengineering*, vol. 10, no. 8, p. 911, 2023. https://www.mdpi.com/2306-5354/10/8/911
11. F. Azam, F. Majeed, and A. Nisar, "A systematic review on the application of deep learning techniques in dental radiographs," *Computers in Biology and Medicine*, 2022.
12. Papasratorn, D., et al. "Investigation of the Best Effective Fold of Data Augmentation for Training Deep Learning Models for Recognition of Contiguity Between Mandibular Third Molar and Inferior Alveolar Canal on Panoramic Radiographs," Clinical Oral Investigations, vol. 27, no. 7, pp. 3759–3769, 2023. https://pubmed.ncbi.nlm.nih.gov/37043029/
13. Chen, X., et al. "Recent Advances and Clinical Applications of Deep Learning in Medical Image Analysis," arXiv preprint arXiv:2105.13381, 2021. https://arxiv.org/abs/2105.13381
14. Yi, X., Walia, E., & Babyn, P. "Generative Adversarial Network in Medical Imaging: A Review," arXiv preprint arXiv:1809.07294, 2018. https://arxiv.org/abs/1809.07294
15. IMAIOS. "Convolutional Neural Networks Architectures for Classification in Medical Imaging," IMAIOS Blog, 2021. https://www.imaios.com/en/resources/blog/classification-of-medical-images-the-most-efficient-cnn-architectures