

A Multimodal Detection for Dental Disease Classification

Mrs. J. Princess Bala¹, Ms. R. Jenifer²

¹Assistant Professor, Computer Science and Engineering, Jayaraj Annapackiam CSI College of Engineering,

²Research Scholar, Computer Science and Engineering, Jayaraj Annapackiam CSI College of Engineering.

Abstract

The project presents an AI-assisted dental disease classification system using the dental-disease-detection dataset, aimed at automating the identification of four dental conditions: Caries, Crown, Filling, and Periapical Lesion. It implements and evaluates three deep learning algorithms: ResNet50, MaxViT, and Swin V2 Tiny model, each with unique approaches to feature extraction and processing. ResNet50 utilizes deep convolutional layers, MaxViT integrates convolution with attention mechanisms, while Swin V2 Tiny employs a hierarchical Transformer architecture. The models are trained on preprocessed and augmented images, and their performance is measured through accuracy, precision, recall, F1-score, and confusion matrix metrics. The system offers computer-aided diagnostic support to enhance the efficiency of dental X-ray analysis and assist professionals in disease detection.

Keywords: dental x-ray images, deep learning, ResNet50, MaxViT, Swin V2 Tiny, image classification, transfer learning, computer-aided diagnosis, medical image analysis.

1. Introduction

Dental diseases are one of the most common problems relating to oral health and affect both the teeth and the surrounding tissues. It is important to identify these conditions early on in order to choose the right treatment and to avoid further complications. Dental X-rays, since they show the internal structure of the teeth, are commonly used by dental professionals for diagnosis. Yet manual examination of X-ray images is time-taking and prone to variation because of the quality of the image, the clinician's experience, and the fact that different dental conditions can look similar to each other.

The recent progress in Artificial Intelligence (AI) and Deep Learning (DL) has made it possible to carry out an automated analysis of medical images. Deep learning models are able to learn complex visual features directly from images and then classify them into separate disease categories. For this project, the dental-disease-detection dataset is used in order to create a multi-class dental disease classification system, the system concentrating on four types: Caries, Crown, Filling, and Periapical Lesion.

The three state-of-the-art deep learning architectures ResNet50, MaxViT, and Swin V2 Tiny are implemented and compared. ResNet50 uses deep convolutional layers to carry out feature extraction; MaxViT combines convolution and attention mechanisms in order to capture both local and global features; and Swin V2 Tiny makes use of a hierarchical Transformer architecture together with shifted-window attention in order to learn the contextual relationships in dental X-ray images.

The system involves image preprocessing, augmentation, model training, classification, and prediction. Performance of the model is assessed by means of accuracy, precision, recall, the F1-score, the confusion matrix, and the performance for each class. Moreover, the system determines the disease category and gives a confidence score for each input X-ray image. This method is intended to provide an efficient computer-aided diagnostic support system for the initial identification of dental disease.

2. Literature Review

Recent research has shown that deep learning methods are becoming more and more commonly used for automatically detecting dental diseases from X-ray images. Conventionally, dental diagnosis relies on a manual examination by dental professionals, a process that can take a lot of time and require considerable clinical expertise. Convolutional Neural Networks (CNNs) have been extensively studied for the analysis of dental images since they are able to automatically extract meaningful features from medical images. ResNet50 featured residual connections, which allow deeper networks to be trained, and has been successfully used in a variety of medical image classification tasks. More recently, attention-based architectures such as MaxViT have combined convolutional operations with attention mechanisms in order to capture both the local details and the global relationships in images. Transformer-based models have also become important in the area of medical image analysis, with Swin Transformer and Swin V2 Tiny enabling hierarchical feature learning via shifted-window attention. These architectures are capable of identifying complex spatial patterns that could be useful for telling similar dental conditions apart. There have been previous studies in which deep learning has been applied to the detection of conditions such as caries, restorations, crowns, and periapical lesions from dental radiographs. However, due to differences in the datasets, image quality, disease categories, and model architectures, direct comparison is difficult.

S.N O	TITLE	AUTHOR	YEAR & JOURN AL	METHOD	REMARKS	DISEASE ACCURAC Y
1.	Multimodal Swin Transformer With Attention-Based Fusion of Panoramic Bone Loss Features for Age Estimation	Shaohua Tang, Seung-Min Kim, Ji-Hyeong Ko, Hyeong-Jae Kim, Kyu-Hwan Lee, Dong-Hun Han, Young-Jin Jung	2026 IEEE	Dual-Channel Swin Transformer + FiLM Gated Fusion + CBAM Attention	The study demonstrates the effectiveness of Swin Transformer and attention mechanisms for dental image analysis.	It considers age estimation

2.	Deep Learning Applications for Dental-Disease Classification Using Intraoral Photographic Images: Current Status and Future Perspectives	A. M. Mutawa, Yacoub Yousef Altarakemah, and Karthiga Thirupathy	2026, AI (MDPI), Volume 7, Article 85	Using deep learning models such as CNN, ResNet, U-Net, Vision Transformer, GAN, and Federated Learning.	DL shows promising results for dental disease detection, but challenges include limited data, class imbalance, image variability, and poor external validation.	Less than 85%
3.	A Feature Selection and Enhanced Reuse Framework for Detection and Classification of Dental Diseases in Panoramic Dental X-Ray Images	Zhiyuan Wu, Huangjian, and Zhengsong	2025, IEEE Access	CNN using selective receptive field fusion, multi-scale feature extraction, and feature reuse with skip fusion.	Achieved improved dental disease detection performance, increasing mAP@50 from 0.410 to 0.631. The proposed DENIMPACT dataset also supports impacted tooth detection.	Accuracy = 63.1%
4.	Deep Learning Applications for Dental-Disease Classification Using Intraoral Photographic Images:	A. M. Mutawa, Y. Y. Altarakemah, and K. Thirupathy	2026, AI (MDPI)	Vision Transformer and Transfer Learning	Deep learning provides promising results, but challenges such as limited datasets, class imbalance, and image-	Accuracy = 79%

	Current Status and Future Perspectives				quality variations remain.	
--	--	--	--	--	----------------------------	--

3. Hardware and Software Requirements

3.1 Hardware

- Processor: Intel Core i5/i7
- RAM: Minimum 8 GB (16 GB recommended)
- Storage: Minimum 20 GB of free disk space
- GPU: NVIDIA GPU with CUDA support (recommended for deep learning training)
- GPU Memory: Minimum 8 GB VRAM

3.2 Software

- Operating System: Windows, Linux
- Programming Language: Python 3.x
- Development Platform: Google Colab / Jupyter Notebook
- Deep Learning Framework: PyTorch
- Libraries: NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, OpenCV
- GPU Support: CUDA
- Dataset Source: Kaggle Dental Disease Detection Dataset

4. Methodology

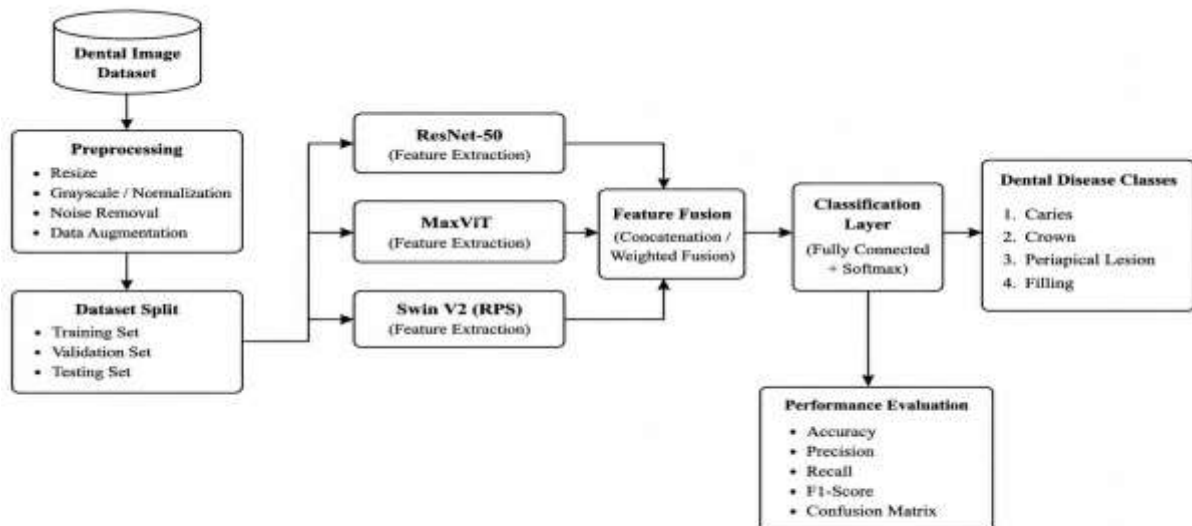


Figure 1: Architecture Diagram for Proposed Model

5. Proposed Model

Swin Transformer (Shifted Window Transformer) is a deep learning architecture designed for computer vision tasks. Unlike traditional Convolutional Neural Networks (CNNs), which mainly use convolution

operations to extract image features, Swin Transformer uses self-attention mechanisms to understand relationships between different regions of an image. It divides an input image into small patches and processes them using window-based multi-head self-attention (W-MSA). In the next layer, the windows are shifted, creating Shifted Window Multi-Head Self-Attention (SW-MSA). This allows information to flow between neighboring windows while reducing the computational cost compared with global self-attention. Swin Transformer uses a hierarchical architecture, where feature representations are gradually merged as the network becomes deeper. This enables the model to learn both fine-grained local features and higher-level global features. Because of these properties, Swin Transformer has been applied to medical image analysis and classification.

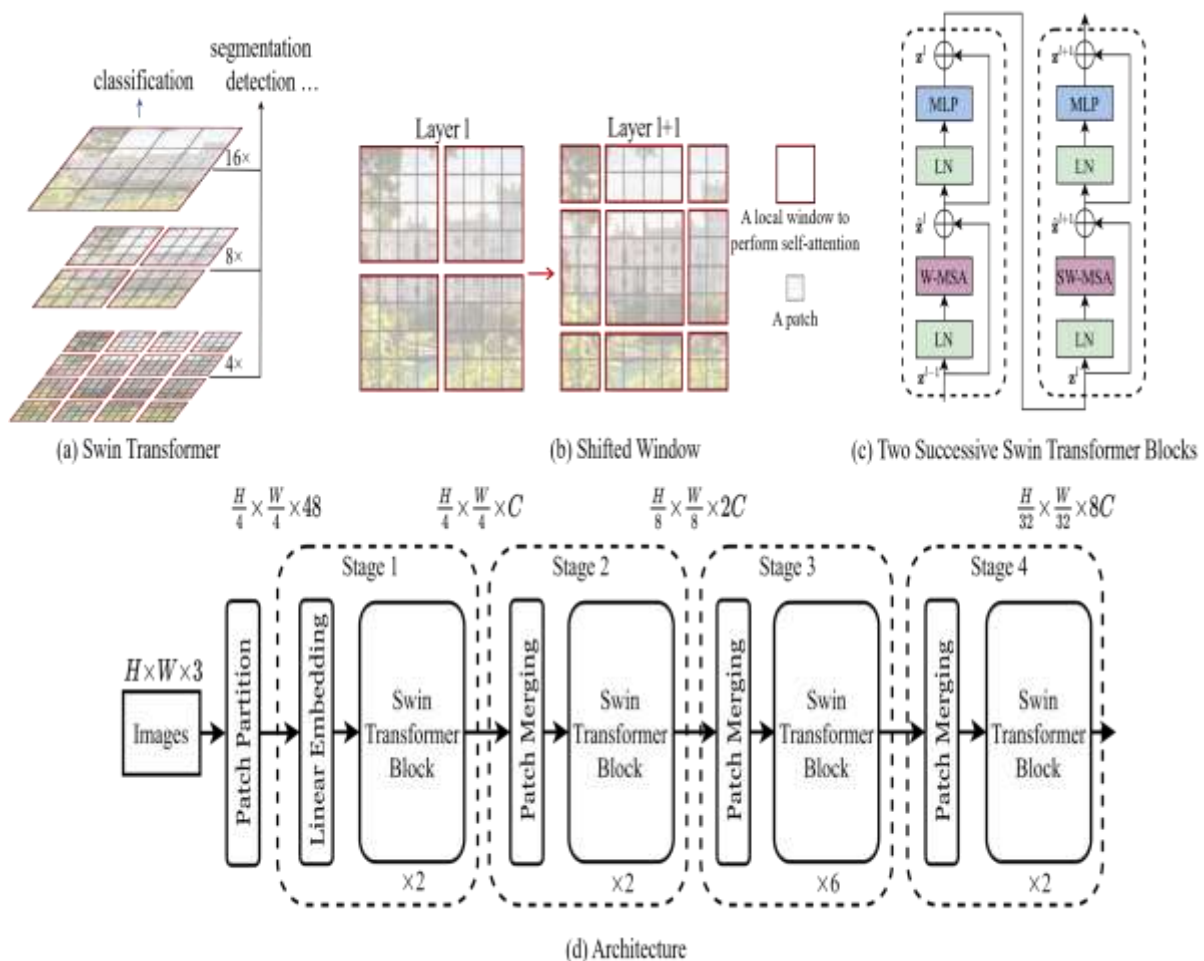


Figure 2: Architecture Diagram for Swin Transformer

Swin v2 is a model with 3 billion parameters that is concerned with methods for scaling a vision model to billions of parameters; it includes techniques such as combining residual-post-norm with cosine attention to improve training stability, and the use of log-spaced continuous position bias to better deal with the different image resolutions encountered between pre-training and fine-tuning. These enhancements allow very large models (up to 3 billion parameters) to be trained efficiently and enable them to process high-resolution images.

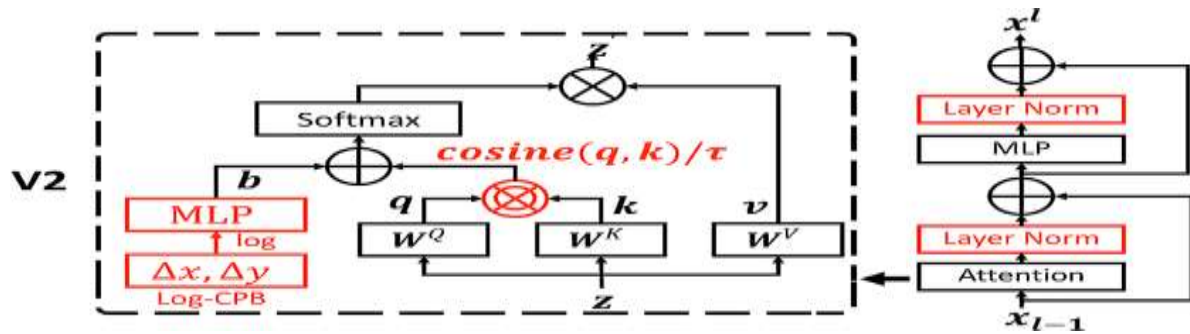


Figure 3: Architecture Diagram Extended from Swin Transformer from Swin V2 Tiny Model

6. Data Flow Process

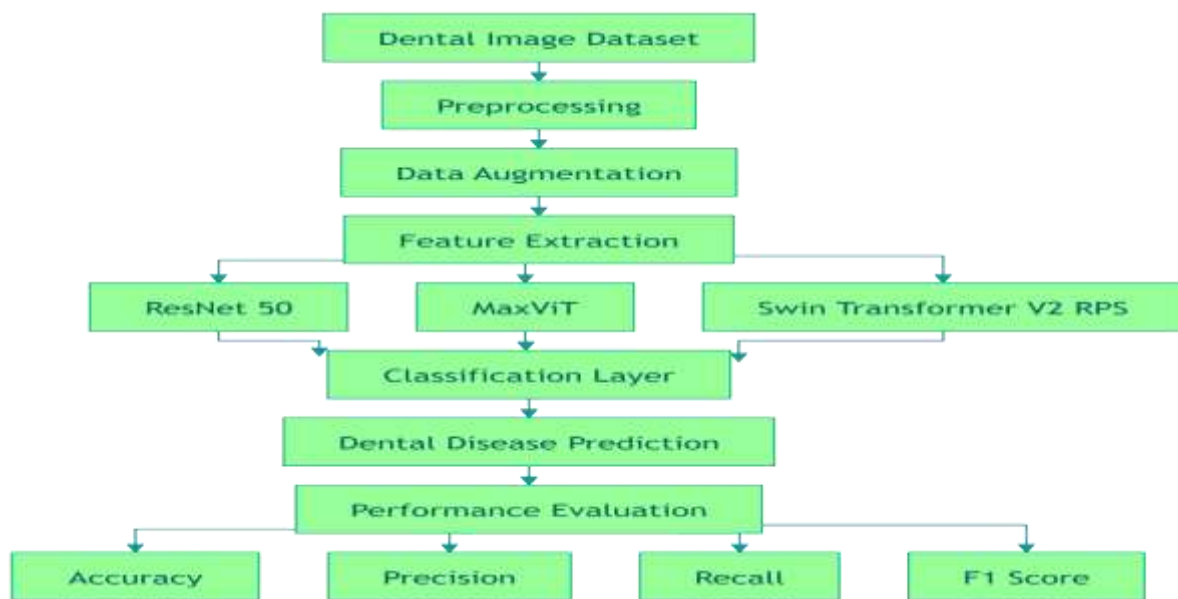
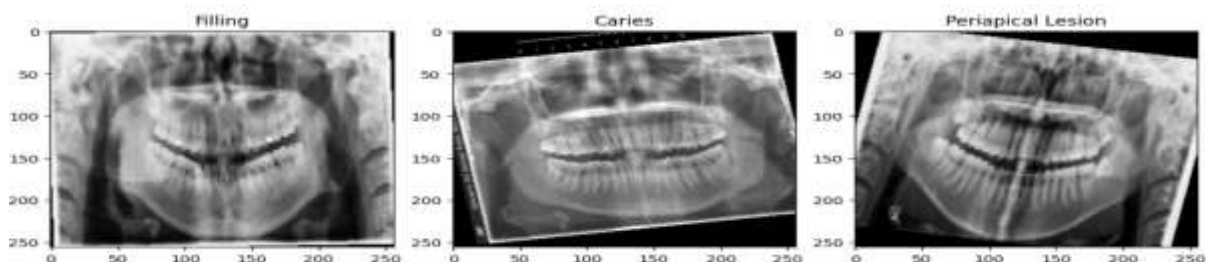


Figure 4: Data Flow Diagram

7. Implementation Process

7.1 Image Dataset Collection

The dataset has four classes, representing different health conditions and diseases: Caries, Crown, Filling, Periapical Lesion. The dataset consider dental X-ray images. In my dataset the total image instances with target classes have 17,711 images were collected from kaggle website and dataset name is “dental-disease-detection”.



Disease	Images
Caries	3065
Crown	4087
Filling	9947
Periapical Lesion	1612
Total	17,711

Table 1: Dataset Images for each Disease

7.2 Data Preprocessing

The image must be processed before sending to the algorithm. Preprocessing techniques are Resize images, normalize pixel values, remove noise, large image cropping, and vertical flipping and apply data augmentation. Input image as a 1615*840 and 640*640 pixels to Resizing on image is 256*256pixel.

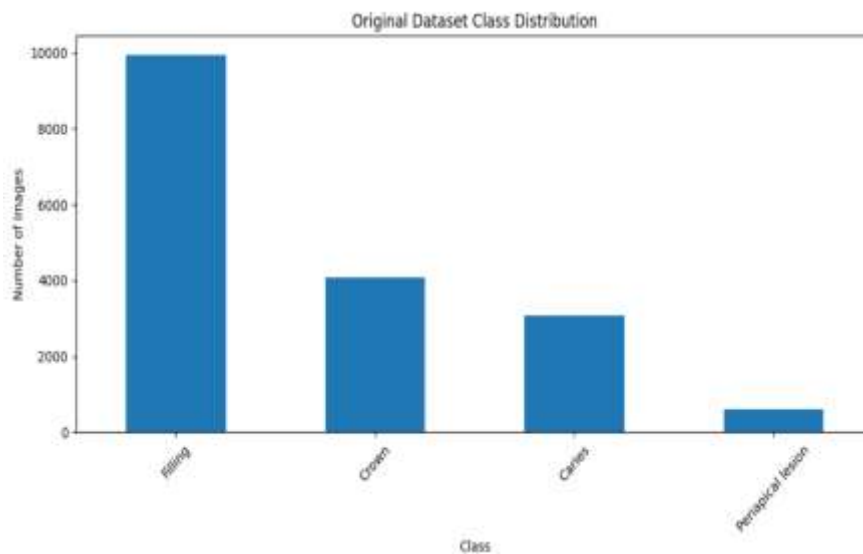


Figure 5: Dataset Class Distribution Bar Chart

7.3 Data Augmentation

Data augmentation is used to artificially increase the diversity of training images and reduce over fitting. For dental disease classification, appropriate augmentation techniques can improve the performance of both ResNet-50 and Max ViT, Swin v2 Tiny.

7.4 Feature Extraction

Preprocessed dental images are passed through ResNet-50, MaxViT, and Swin V2 Tiny to extract complementary spatial, texture, and attention-based features. The extracted features are then fused and forwarded to the classification layer.

- **ResNet-50 Feature Extraction:** Feed the dental images into ResNet-50 and extract deep CNN features.
- **MaxViT Feature Extraction:** Process the same images using MaxViT to obtain attention-based feature representations.

- **Swin v2 Tiny:** Swin V2 feature extraction means extracting the important patterns from a dental X-ray, such as tooth shape, edges, lesions, and abnormal regions, so that the model can identify the dental disease.

7.5 Dataset Splitting

DATASET	TRAINING	VALIDATION	TESTING
Caries	2145	460	460
Crown	2861	613	613
Filling	6963	1492	1492
Periapical lesion	428	92	92

Table 2: Training, Validating & Testing

7.5 Model Training

Train the complete network using a suitable loss function and optimizer, while monitoring training and validation performance. The fused features are used to train the classification model using the training dataset. The model learns to classify dental images into the four disease classes by minimizing the loss function and updating its weights over multiple epochs.



Figure 6: Swin V2 Tiny Model Training and Validation

7.6 Model Evaluation

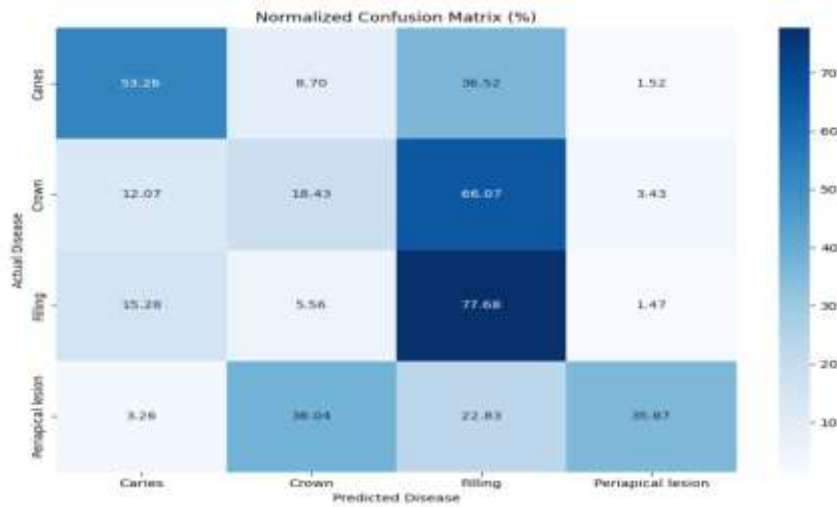


Figure 7: Confusion Matrix for Swin V2 Tiny Model

7.7 Classification Report Summary

METRIC	PRECISION	RECALL	F1 SCORE
Caries	0.8627	0.9412	0.9003
Crown	0.8214	0.7931	0.8070
Filling	0.9459	0.9138	0.9296
Periapical lesion	0.8233	0.8459	0.8345
Macro Avg	0.8634	0.8735	0.8678

Table 3: Swin V2 Tiny Model Report

7.8 Performance Evaluation Overall Accuracy

EPOCHS	RESNET 50	MAXVIT	Swin v2 Tiny
5	58.22%	75.94%	82.14%
10	60.18%	91.17%	95.29%
Final Accuracy	59.20%	85.26%	89.88%

Table 4: Accuracy

8. Conclusion

The comparative implementation of ResNet-50 and MaxViT demonstrates the potential of deep learning for improving the accuracy and efficiency of automated dental disease diagnosis. By applying appropriate image preprocessing, data augmentation, transfer learning, and model optimization techniques, the proposed system can provide reliable classification of caries, crown, fillings, and periapical lesions.

Overall, the developed system can serve as a computer-aided diagnostic tool to assist dentists in the early identification and classification of dental conditions. In the future, the system can be further improved by using larger and more diverse datasets, addressing class imbalance, incorporating explainable AI techniques, and developing a hybrid model that combines the strengths of ResNet-50 and MaxViT and Swin v2 Tiny for enhanced diagnostic performance.

References

1. Oprea, S.; Marinescu, C.; Lita, I.; Jurianu, M.; Visan, D.A.; Cioc, I.B. Image processing techniques used for dental X-ray image analysis. In Proceedings of the 2008 31st International Spring Seminar on Electronics Technology, Budapest, Hungary, 7–11 May 2008; pp. 125–129.
2. Ossowska, A.; Kusiak, A.; Świetlik, D. Artificial Intelligence in Dentistry—Narrative Review. *Int. J. Environ. Res. Public Health* 2022, 19, 3449.
3. Yu, Y.J. Machine learning for dental image analysis. arXiv 2016, arXiv:1611.09958. Tuzoff, D.V.; Tuzova, L.N.; Bornstein, M.M.; Krasnov, A.S.; Kharchenko, M.A.; Nikolenko, S.I.; Sveshnikov, M.M.; Bednenko, G.B. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofac. Radiol.* 2019, 48, 20180051.
4. Imangaliyev, S.; Veen, M.H.; Volgenant, C.; Keijser, B.J.; Crielaard, W.; Levin, E. Deep learning for classification of dental plaque images. In International Workshop on Machine Learning, Optimization, and Big Data; Springer: Cham, Switzerland, 2016; pp. 407–410.
5. Liu, L.; Xu, J.; Huan, Y.; Zou, Z.; Yeh, S.C.; Zheng, L.R. A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal. *IEEE J. Biomed. Health Inform.* 2019, 24, 898–906.
6. Khan, H.A.; Haider, M.A.; Ansari, H.A.; Ishaq, H.; Kiyani, A.; Sohail, K.; Muhammad, M.; Khurram, S.A. Automated feature detection in dental periapical radiographs by using deep learning. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 2021, 131, 711–720.
7. Tian, S.; Dai, N.; Zhang, B.; Yuan, F.; Yu, Q.; Cheng, X. Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. *IEEE Access* 2019, 7, 84817–84828.
8. Casalegno, F.; Newton, T.; Daher, R.; Abdelaziz, M.; Lodi-Rizzini, A.; Schürmann, F.; Krejci, I.; Markram, H. Caries detection with near-infrared transillumination using deep learning. *J. Dent. Res.* 2019, 98, 1227–1233.
9. Prajapati, S.A.; Nagaraj, R.; Mitra, S. Classification of dental diseases using CNN and transfer learning. In Proceedings of the 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), Dubai, United Arab Emirates, 11–14 August 2017; pp. 70–74.
10. Chen, H.; Zhang, K.; Lyu, P.; Li, H.; Zhang, L.; Wu, J.; Lee, C.H. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci. Rep.* 2019, 9, 3840.
11. Silva, G.; Oliveira, L.; Pithon, M. Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Syst. Appl.* 2018, 107, 15–31.
12. Muresan, M.P.; Barbura, A.R.; Nedevschi, S. Teeth Detection and Dental Problem Classification in Panoramic X-ray Images using Deep Learning and Image Processing Techniques. In Proceedings of the 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 3–5 September 2020; pp. 457–463.

13. Lee, J.-H.; Han, S.-S.; Kim, Y.H.; Lee, C.; Kim, I. Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 2019, 129, 635–642.
14. Jader, G.; Fontineli, J.; Ruiz, M.; Abdalla, K.; Pithon, M.; Oliveira, L. Deep instance segmentation of teeth in panoramic X-ray images. In *Proceedings of the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Parana, Brazil, 29 October–1 November 2018; pp. 400–407.
15. Kim, J.; Lee, H.S.; Song, I.S.; Jung, K.H. DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. *Sci. Rep.* 2019, 9, 17615.
16. Jain, A.K.; Chen, H. Matching of dental X-ray images for human identification. *Pattern Recognit.* 2004, 37, 1519–1532.
17. Liu, X.; Faes, L.; Kale, A.U.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdas, M.; Kern, C.; et al. A comparison of deep learning performance against healthcare professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *Lancet Digit. Health* 2019, 1, e271–e297.
18. Zhang, K.; Wu, J.; Chen, H.; Lyu, P. An effective teeth recognition method using label tree with cascade network structure. *Comput. Med. Imaging Graph.* 2018, 68, 61–70.
19. Schwendicke, F.; Golla, T.; Dreher, M.; Krois, J. Convolutional neural networks for dental image diagnostics: A scoping review. *J. Dent.* 2019, 91, 103226.
20. Johari, M.; Esmaili, F.; Andalib, A.; Garjani, S.; Saberhari, H. Detection of vertical root fractures in intact and endodontically treated premolar teeth by designing a probabilistic neural network: An ex vivo study. *Dentomaxillofac. Radiol.* 2017, 46, 20160107.
21. Fukuda, M.; Inamoto, K.; Shibata, N.; Aiji, Y.; Yanashita, Y.; Kutsuna, S.; Nakata, K.; Katsumata, A.; Fujita, H.; Aiji, E. Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. *Oral Radiol.* 2020, 36, 337–343.
22. Lee, C.T.; Kabir, T.; Nelson, J.; Sheng, S.; Meng, H.W.; Van Dyke, T.E.; Walji, M.F.; Jiang, X.; Shams, S. Use of the deep learning approach to measure alveolar bone level. *J. Clin. Periodontol.* 2022, 49, 260–269.
23. Pakbaznejad Esmaili, E.; Pakkala, T.; Haukka, J.; Siukosaari, P. Low reproducibility between oral radiologists and general dentists with regards to radiographic diagnosis of caries. *Acta Odontol. Scand.* 2018, 76, 346–350.
24. Abdalla-Aslan, R.; Yeshua, T.; Kabla, D.; Leichter, I.; Nadler, C. An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 2020, 130, 593–602.