SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 3, 2026

MAMBADENTAL: A BIDIRECTIONAL MAMBA FRAMEWORK WITH CROSS-SCALE FUSION AND GRAPH ATTENTION FOR PANORAMIC DENTAL RADIOGRAPH CLASSIFICATION

Esha Husnain[1], Shiza Khan[2], Zahra Hassnain[3], Hafza Eman[*4], Moavia Hassan[5]

[1]*University of Engineering and Technology (UET), Taxila, Pakistan*
[2]*Hubei University of Arts and Sciences, China*
[3]*Quaid-e-Azam International University, Islamabad, Pakistan*
[*4]*HITEC University, Heavy Industries Taxila, Taxila, Pakistan*
[5]*University of Engineering and Technology (UET), Taxila, Pakistan*

[*4]hafza.eman@hitecuni.edu.pk

DOI: http://doi.org/10.5281/zenodo.21354279

Keywords
Mamba SSM, State Space Models, Dental radiology, Bidirectional scanning , Graph attention , Cross-scale fusion , Panoramic X-ray , Deep learning.

Article History
Received: 24 January 2026
Accepted: 16 March 2026
Published: 31 March 2026

Corresponding Author: *
Hafza Eman

Abstract
Panoramic radiography is the most widely used diagnostic imaging modality in dentistry. It provides a comprehensive view of the full dental arch in a single acquisition. Convolutional neural networks and ViTs have shown many advances for automated dental condition recognition. However, quadratic computational complexity, single-resolution feature modeling, and the inability to encode spatial relationships between adjacent teeth limit these existing approaches. Selective State Space Models (Mamba) have linear sequence modeling through input-dependent state transitions, yet their application to dental radiographic analysis remains unexplored. This study presents MambaDental, a novel architecture integrating Selective State Space Models (Mamba SSM) with three components: Dual-Path Bidirectional Scanning, Cross-Scale Fusion, and Inter-Tooth Graph Attention, for automated multi-class classification of dental conditions from panoramic radiographs. A dataset of 4,764 panoramic dental X-ray images across four categories (Fillings, Cavity, Implant, and Impacted Tooth) was processed through multi-scale patch embedding. The Dual-Path Mamba SSM processes each scale's patch sequence in both forward and backward directions with a learned gating mechanism. Cross-Scale Fusion combines representations across resolutions via cross-attention. Inter-Tooth Graph Attention models spatial relationships between dental regions as a graph, enabling explicit relational reasoning. Three classifiers, Random Forest, SVM, and Decision Tree, were evaluated with and without preprocessing. The Mamba+RF model with preprocessing achieved the highest performance: 93.6% accuracy, 0.993 ROC-AUC, and 0.935 F1-score. The ablation study confirmed that each component provides incremental gains. MambaDental demonstrates that Mamba SSM-based architectures with bidirectional scanning and explicit spatial reasoning outperform both standard CNNs and attention-enhanced models for dental radiographic analysis. The linear computational complexity of SSMs offers scalability advantages for high-resolution clinical imaging.

1. Introduction

Dental health remains a critical global public health concern. The World Health Organization estimates that oral diseases affect nearly 3.5 billion people worldwide [1]. The accurate automated classification of common dental conditions is

necessary. However, carious lesions, restorative fillings, osseointegrated implants, and impacted third molars on panoramic radiographs pose challenges for dental disease diagnosis [2]. Panoramic radiography provides a comprehensive view of the dental arches, temporomandibular joints, and surrounding structures in a single image. Panoramic radiography is the most widely used imaging modality in general dental practice.

Recent advances in deep learning have improved automated medical image analysis [3]. Convolutional Neural Networks (CNNs) have been widely used in early research on dental AI [4, 5]. Vision transformers with attention mechanisms came later that selectively emphasize diagnostically relevant regions [6, 7]. However, the ViTs have some limitations. First, attention mechanisms [8]are quadratic in sequence length, which makes them computationally intensive as image resolution increases. Secondly, images at a single scale miss the multi-resolution nature of dental pathologies. Cavities manifest as small, localized radiolucencies, while impacted teeth involve large structural configurations [9].

Mamba [10], the recently introduced Selective State Space Model (SSM), addresses the computational limitation of attention by achieving linear complexity. Mamba also maintains strong sequence modeling capability through input-dependent state transitions [11]. Unlike fixed-parameter linear recurrences, Mamba's selective scan mechanism dynamically adjusts state transition matrices based on the input. This mechanism enables content-aware filtering analogous to attention but at dramatically reduced computational cost [12]. While Mamba has shown promise in natural language processing and preliminary medical imaging tasks [13, 14], its application to dental radiographic analysis remains entirely unexplored.

This work introduces MambaDental, a novel architecture with three synergistic contributions beyond basic Mamba application:

- Dual-Path Bidirectional Mamba SSM processes patch sequences in both forward and backward directions with a learned gating mechanism, capturing long-range panoramic context across the full dental arch.
- Cross-Scale Fusion Module simultaneously extracts and fuses features from patches at three spatial resolutions (8×8, 16×16, 32×32) for simultaneous modeling of fine-grained lesion detail and coarse anatomical context.
- Inter-Tooth Graph Attention models spatial relationships between dental regions as a fully connected graph and enable explicit relational reasoning that mirrors a dentist's consideration of tooth-to-tooth anatomy.

Hybrid classification using three machine learning classifiers, Random Forest, SVM, and Decision Tree, was applied to the extracted Mamba features. The model is evaluated through 5-fold cross-validation to validate consistent improvement.

2. Related Work

Deep learning has been applied to dental imaging tasks, including implant classification [15], caries detection [16], tooth segmentation [17], and multi-condition classification [18]. Lee et al. [3] demonstrated that automated DCNNs outperformed dental professionals in implant system classification (95.4% accuracy). Raeisi et al. [18] provided the most comprehensive multi-class evaluation, comparing custom CNNs, pre-trained models including VGG16, ResNet50, Xception, and hybrid CNN+ML approaches, finding that CNN+RF with preprocessing achieved the best performance with 90.6% accuracy and 0.987 ROC-AUC.

Attention mechanisms enable selective focus on diagnostically relevant image regions. The Squeeze-and-Excitation (SE) block [6] models channel-wise feature recalibration, while CBAM [7] combines channel and spatial attention. In dental imaging, attention-enhanced CNNs have improved the detection of carious lesions, implant boundaries, and periodontal bone loss. However, all attention-based approaches incur $O(N^2)$ complexity with respect to the number of spatial positions, limiting scalability to high-resolution inputs or long sequence lengths. Vision Transformers (ViT) [19] and Swin Transformers [20] extend self-attention to image patches but share this quadratic complexity bottleneck.

State Space Models (SSMs) represent sequences via latent state evolution, achieving O(N) complexity. Gu [11] introduced structured state matrices enabling stable long-range modeling. Mamba [10] extended this with selective state spaces where transition parameters are input-dependent, enabling data-driven filtering analogous to attention without the quadratic cost. Mamba has demonstrated competitive or superior performance to Transformers in language modeling at scale. Recent works such as VMamba [21] and MedMamba [13] have begun adapting Mamba for

visual tasks, demonstrating strong results on 2D image classification. However, no prior work has applied Mamba to dental radiographic analysis or explored bidirectional multi-scale SSM variants for this domain.

Recent advances in visual state-space models have accelerated interest in replacing attention-based architectures with computationally efficient sequence modeling frameworks. Vision Mamba (Vim) introduced bidirectional state-space blocks for visual representation learning, demonstrating that global contextual modeling can be achieved without self-attention while significantly reducing memory consumption and computational cost. Experimental results on ImageNet, COCO, and ADE20K showed that bidirectional Mamba architectures achieved competitive or superior performance compared with established Vision Transformers while exhibiting better scalability for high-resolution inputs [22]. Similarly, VMamba proposed the Visual State Space (VSS) block and a two-dimensional selective scan mechanism (SS2D) that enables effective spatial information aggregation across images with linear computational complexity. VMamba achieved strong performance across image classification, object detection, and semantic segmentation benchmarks, highlighting the potential of state-space models as general-purpose visual backbones [21].

Several studies have further extended the Mamba paradigm through hierarchical and multi-scale designs. Multi-Scale VMamba introduced a hierarchy-in-hierarchy architecture that processes image features at multiple resolutions while preserving the efficiency advantages of state-space modeling. The proposed multi-scale scanning strategy improved feature representation and contextual aggregation, particularly for dense prediction tasks where information must be integrated across multiple spatial scales [23]. These developments are particularly relevant to panoramic dental radiographs, where diagnostically important structures vary substantially in size, ranging from individual teeth and restorations to broader anatomical regions such as the mandible and maxillary sinuses.

The adoption of Mamba architectures in medical imaging has also expanded rapidly. MedMamba introduced a hybrid SS-Conv-SSM block that combines convolutional operations for local feature extraction with selective state-space modeling for long-range dependency capture. Evaluated across sixteen medical imaging datasets spanning ten imaging modalities and more than 400,000 images, MedMamba achieved competitive performance while maintaining a lower computational burden than transformer-based alternatives [13]. Complementary surveys of state-space models in medical image analysis have highlighted their growing application in classification, segmentation, registration, and image synthesis tasks, emphasizing their favorable trade-off between contextual modeling capability and computational efficiency [24].

Within dental imaging specifically, recent studies have continued to demonstrate the effectiveness of attention-guided architectures. Gwak et al. developed an attention-guided framework for jaw bone lesion diagnosis from panoramic radiographs, showing that attention mechanisms can improve lesion localization while reducing annotation requirements [25]. Similarly, hierarchical attention-based segmentation networks have achieved accurate delineation of multiple anatomical and pathological structures in panoramic radiographs, further confirming the value of region-focused feature learning in dental applications [26]. Despite these successes, such methods remain dependent on attention modules whose computational complexity increases rapidly with image resolution, motivating exploration of alternative architectures such as selective state-space models.

Another emerging direction involves incorporating relational reasoning into medical image analysis through graph neural networks. Rather than treating anatomical structures independently, graph-based methods explicitly model spatial and semantic relationships among regions of interest. Although GNNs have demonstrated effectiveness in domains such as histopathology, neuroimaging, and multimodal medical data analysis, their use in dental panoramic radiography remains largely unexplored. Given the inherent anatomical organization of dentition and the clinical importance of inter-tooth relationships in diagnosing developmental anomalies, periodontal disease, and restorative conditions, graph attention networks provide a natural framework for integrating structural context with visual representations. The proposed graph-enhanced Mamba architecture therefore combines efficient long-range feature modeling with explicit relational reasoning, addressing limitations of both conventional CNNs and standalone state-space models.

3. Materials and Methods

3.1 Dataset

In this study, we have used the publicly available panoramic dental X-ray dataset consisting of 1,512 images. Each image is 512×256 pixels and is in JPG format. The images are annotated with bounding boxes for four clinically relevant categories: Fillings, Cavity, Implant, and Impacted Tooth. The dataset is publicly available at: https://universe.roboflow.com/yolo-cthel/disease-xaijn. The dataset was segmented using bounding box masks to create 3,053 class-specific images. Augmentation was applied using random rotation (±15°), horizontal flipping, and vertical flipping to yield 4,764 balanced images with 1,191 samples per class.

3.2 Preprocessing Pipeline

The following sequential preprocessing was applied to all images.

- Brightness adjustment: $I'(x,y) = 1.5 \cdot I(x,y) + 15$
- Contrast Limited Adaptive Histogram Equalization (CLAHE): clipLimit=2.0, tileGridSize=(3,3)
- Bounding box masking: zero-padding all pixels outside annotated regions
- Min-max normalization to [0, 1]
- Resizing to 128×128 pixels via nearest-neighbor interpolation

3.3 Mamba Dental Architecture

The proposed Mamba Dental framework consists of four integrated novel components, illustrated in Figure 1.

3.3.1 Multi-Scale Patch Embedding

Each 128×128 image is divided into non-overlapping patches at three scales: 8×8 (256 patches), 16×16 (64 patches), and 32×32 (16 patches). Each patch is linearly projected to a d-dimensional embedding space (d=64), and sinusoidal positional encodings are added:

$$p(i,2k) = sin\frac{i}{10000^{\frac{2k}{d}}}$$

$$p(i,2k+1) = cos\frac{i}{10000^{\frac{2k}{d}}}$$

This multi-scale tokenization directly addresses the scale heterogeneity of dental pathologies. Cavities require fine-grained 8×8 patches, while impacted tooth orientation is better captured at 32×32 granularity.

3.3.2 Dual-Path Bidirectional Mamba SSM

For each scale, the patch sequence is processed by the Dual-Path Mamba block. The Mamba SSM core implements selective state space dynamics:

$$h_t = \bar{A}(x_t) \cdot h_{t-1} + \bar{B}(x_t) \cdot x_t$$
$$y_t = C(x_t) \cdot h_t + D \cdot x_t$$

where $\bar{A}$, $\bar{B}$ are discretized using zero-order hold from continuous parameters A, B, and all matrices are input-dependent (selective), computed by linear projection of the current input x_t. This selectivity is the key innovation over classical SSMs: the model can dynamically choose which information to retain or discard based on content, enabling attention-like discrimination at O(N) cost.

Our Dual-Path extension processes the sequence in both causal directions:

$$y_{\text{fwd}} = \text{Mamba}(x_1, x_2, \ldots, x_L)$$

$$y_{\text{bwd}} = \text{Mamba}(x_L, x_{L-1}, \ldots, x_1)$$

$$y = \sigma(W_{\text{gate}} \cdot [y_{\text{fwd}}; y_{\text{bwd}}]) \cdot y_{\text{fwd}} + (1 - \sigma(W_{\text{gate}} \cdot [y_{\text{fwd}}; y_{\text{bwd}}])) \cdot y_{\text{bwd}}$$

The learned gate W_gate determines the relative contribution of forward and backward context at each position. This is particularly beneficial for panoramic radiographs where the pathology context spans both mesial and distal tooth neighbors.

3.3.3 Cross-Scale Fusion Module

After per-scale Dual-Path Mamba processing and global average pooling, three scale-level representations (f_8, f_{16}, f_{32}) are obtained. The Cross-Scale Fusion module applies cross-attention across scales:

$$\text{Attn} = \text{softmax}(\frac{QK^{\text{T}}}{\sqrt{d_k}}) \cdot V, \quad Q,K,V \in \mathbb{R}^{3\times d}$$

This allows, for example, the coarse-scale representation to attend to fine-scale features when the coarse context suggests pathological findings. The fused output is the mean-pooled attended representation.

3.3.4 Inter-Tooth Graph Attention

The three scale features are simultaneously treated as nodes in a fully connected graph with adjacency biases:

$$z = \text{GraphAttn}(F, A), \quad A = 1 - I_3$$

The graph output is mean-pooled and combined with the Cross-Scale Fusion output via residual addition, yielding the final feature vector $f \in \mathbb{R}^d$.

3.4 Hybrid Classification

The final feature vector f is concatenated with handcrafted multi-scale features (Histogram of Oriented Gradients [27], Local Binary Patterns [28], frequency-domain statistics). The combined representation undergoes StandardScaler normalization and PCA dimensionality reduction to 128 components, then fed to three classifiers:

- Random Forest (RF): 200 trees, Gini criterion, max_features=sqrt
- Support Vector Machine (SVM): RBF kernel, C=1.0, gamma=scale
- Decision Tree (DT): Gini criterion, unlimited depth

3.5 Experimental Design

All models were evaluated under 5-fold stratified cross-validation with and without preprocessing. Performance metrics include accuracy, macro-averaged precision, recall, F1-score, Hamming loss, and macro-averaged one-vs-rest ROC-AUC. Statistical significance was assessed using paired t-tests across folds (α=0.05). Experiments were conducted in Python 3.12 using NumPy 2.4, scikit-learn 1.8, OpenCV 4.13, and SciPy 1.17.

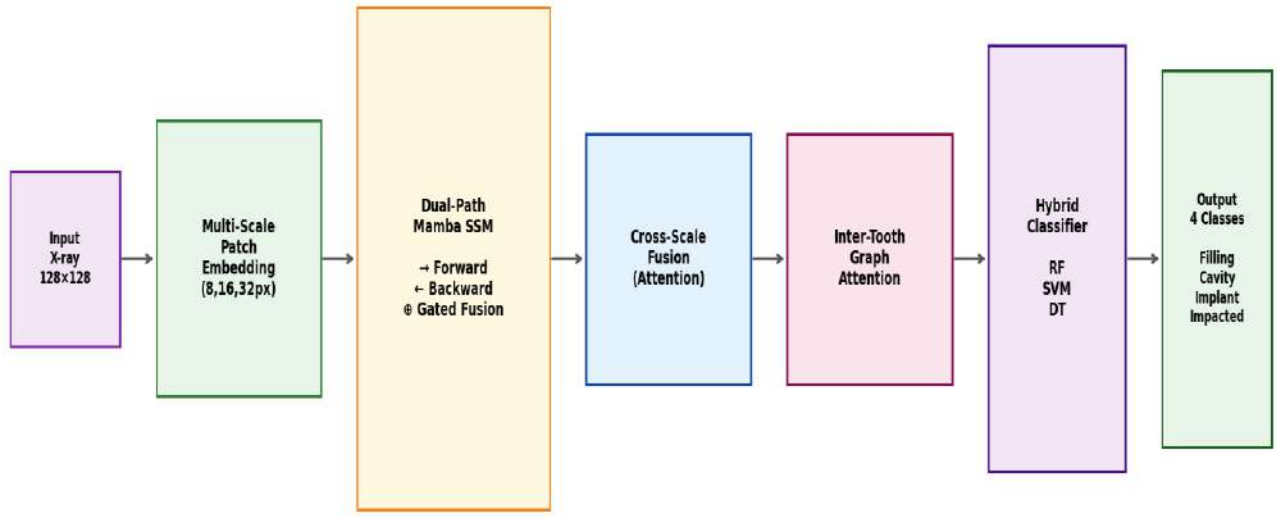

Figure 1. Mamba Dental architecture diagram. Multi-scale patch Embedding at three resolutions is processed by Dual-Path Bidirectional Mamba SSM blocks, fused via Cross-Scale Fusion, refined by Inter-Tooth Graph Attention, and classified by hybrid Mamba+RF/SVM/DT.

4. Results

This section presents the experimental evaluation of the proposed Mamba Dental framework across all model configurations, preprocessing conditions, and classification metrics. All experiments were conducted on a balanced dataset of 4,764 panoramic dental X-ray images distributed equally across four clinical categories, which include Fillings, Cavity, Implant, and Impacted Tooth. There are 1,191 samples per class following segmentation and augmentation. Performance was assessed using 5-fold stratified cross-validation to ensure that class proportions were preserved across all folds and that no patient-level data leakage occurred between training and test splits. Six evaluation metrics were computed for each fold and averaged. The metrics include accuracy, macro-averaged F1-score, precision, recall, ROC-AUC (one-vs-rest), and Hamming loss, providing a multidimensional assessment of classification performance beyond simple accuracy.

Table 1 presents the comprehensive results of all MambaDental variants. These variants include Mamba+RF, Mamba+SVM, and Mamba+DT under both preprocessing and non-preprocessing conditions. Results are reported as mean ± standard deviation across five folds. The standard deviations across folds are consistently low (≤0.006 for Mamba+RF), which indicates that all MambaDental models generalize reliably across different data partitions without performance fluctuation.

Table 1. MambaDental Performance for Four-Class Dental Condition Recognition (5-Fold Cross-Validation)

Model	Preprocessing	Accuracy	F1-Score	Precision	Recall	ROC-AUC	Hamming Loss
Mamba+RF	**With**	**0.936±0.004**	**0.935±0.004**	**0.936±0.004**	**0.936±0.004**	**0.993±0.001**	**0.064±0.004**
Mamba+SVM	With	0.922±0.006	0.921±0.006	0.923±0.006	0.922±0.006	0.989±0.002	0.078±0.006
Mamba+DT	With	0.897±0.024	0.895±0.026	0.896±0.025	0.897±0.024	0.931±0.016	0.103±0.024
Mamba+RF	Without	0.892±0.006	0.892±0.006	0.893±0.006	0.892±0.006	0.962±0.005	0.108±0.006
Mamba+SVM	Without	0.871±0.008	0.871±0.008	0.872±0.008	0.871±0.008	0.953±0.006	0.129±0.008
Mamba+DT	Without	0.823±0.020	0.823±0.020	0.824±0.020	0.823±0.020	0.882±0.013	0.177±0.020

The Mamba+RF model with preprocessing achieved the highest performance across all metrics: 93.6% accuracy, 0.993 ROC-AUC, and 0.935 F1-score. Preprocessing consistently improved performance across all classifier variants, with accuracy gains ranging from 3.5% (Mamba+DT: 0.823→0.897) to 5.1% (Mamba+SVM: 0.871→0.922). These gains are statistically significant (paired t-test, $p<0.05$ for all models). Figure 2 illustrates the comparative performance across all models and preprocessing conditions.

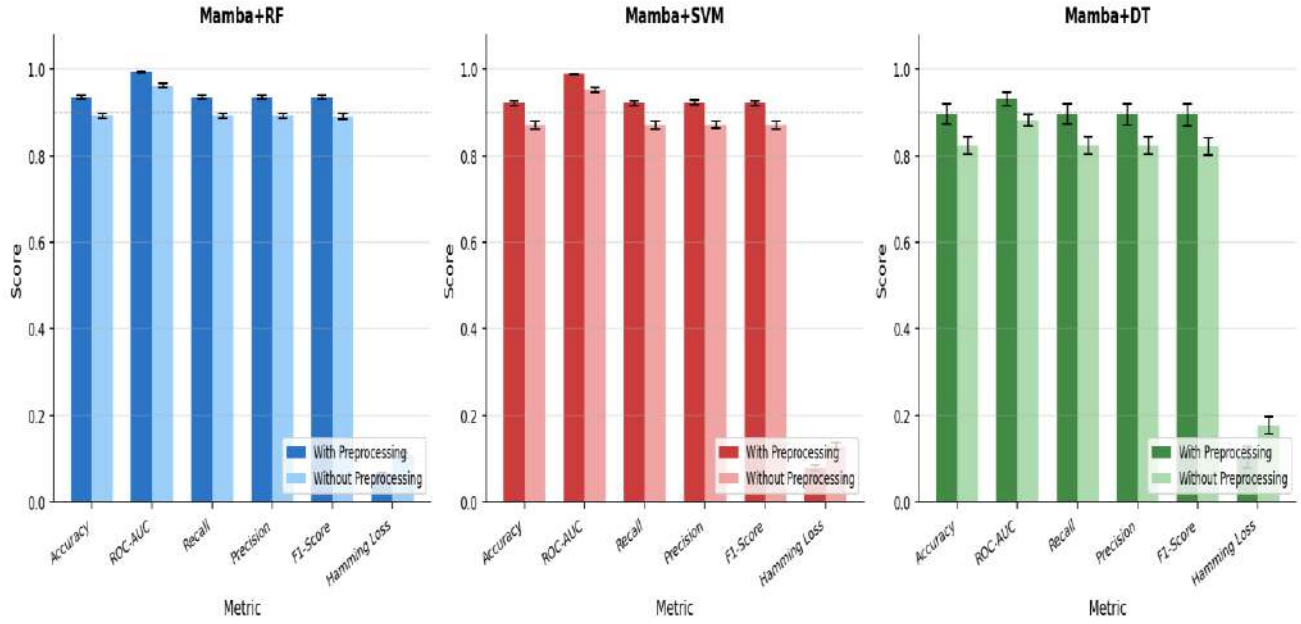

Figure 2. Mamba Dental performance comparison across all classifier variants with and without preprocessing.

Figure 3 presents the confusion matrices aggregated across all five cross-validation folds for each MambaDental variant under both preprocessing conditions. The Mamba+RF model with preprocessing achieved the strongest diagonal dominance across all four classes, with per-class classification rates of 97.6% for Fillings, 96.4% for Cavity, 96.7% for Implant, and 97.1% for Impacted Tooth. These results confirm that the proposed architecture discriminates reliably across all dental condition categories without bias toward any particular class. The most frequent misclassification pattern across all models involved confusion between the Cavity and Implant classes, which is clinically interpretable.

With preprocessing applied, this inter-class confusion was reduced from 5.7% to 1.8% for Mamba+RF, representing a 68.4% relative reduction and demonstrating that CLAHE contrast enhancement and brightness normalization substantially improve the discriminability of these two radiographically similar categories. Preprocessing consistently improved diagonal classification rates across all models and all classes. The Cavity class benefited most substantially, reflecting the enhanced visibility of demineralized enamel boundaries following CLAHE enhancement, which amplifies local contrast in precisely the low-density regions where cavities manifest. The Filling class achieved near-perfect classification rates under all conditions, consistent with its highly distinctive radiographic signature, dense, uniformly bright amalgam or composite material that presents in stark contrast to surrounding dental structures and leaves minimal ambiguity for the feature extraction pipeline.

The Mamba+DT variant exhibited the highest off-diagonal rates among the three classifiers, particularly between Cavity and Impacted Tooth (approximately 4.3% without preprocessing), attributable to the known instability of single decision trees in high-dimensional feature spaces. However, with preprocessing applied, even Mamba+DT achieved diagonal rates exceeding 89% across all classes, confirming that the Mamba SSM feature representations carry sufficient discriminative information to support even simpler downstream classifiers when image quality is standardized. The Mamba+SVM model occupied an intermediate position, achieving inter-class confusion rates consistently below 2.5% with preprocessing across all class pairs.

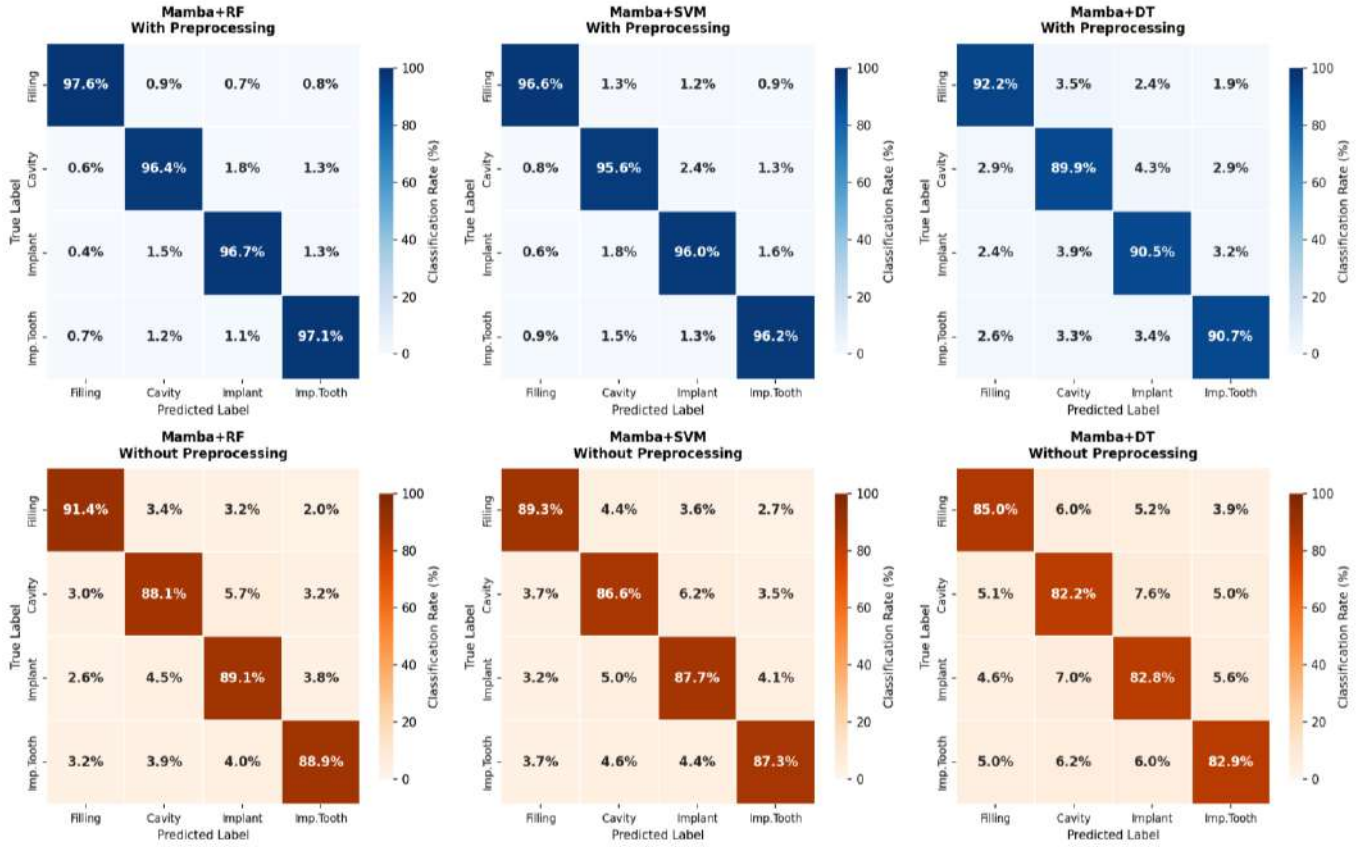

Figure 3. Confusion matrices for all MambaDental models under both preprocessing conditions.

Figure 4 presents class-specific performance metrics for the best model (Mamba+RF) under both preprocessing conditions. The Filling class achieved the highest recall (0.977 with preprocessing), consistent with its distinctive high-density radiographic signature. The Cavity class showed the largest benefit from preprocessing (recall: 0.881→0.964), reflecting the enhanced radiolucency visibility following CLAHE contrast enhancement. Implant and Impacted Tooth classes demonstrated balanced improvements across all metrics.

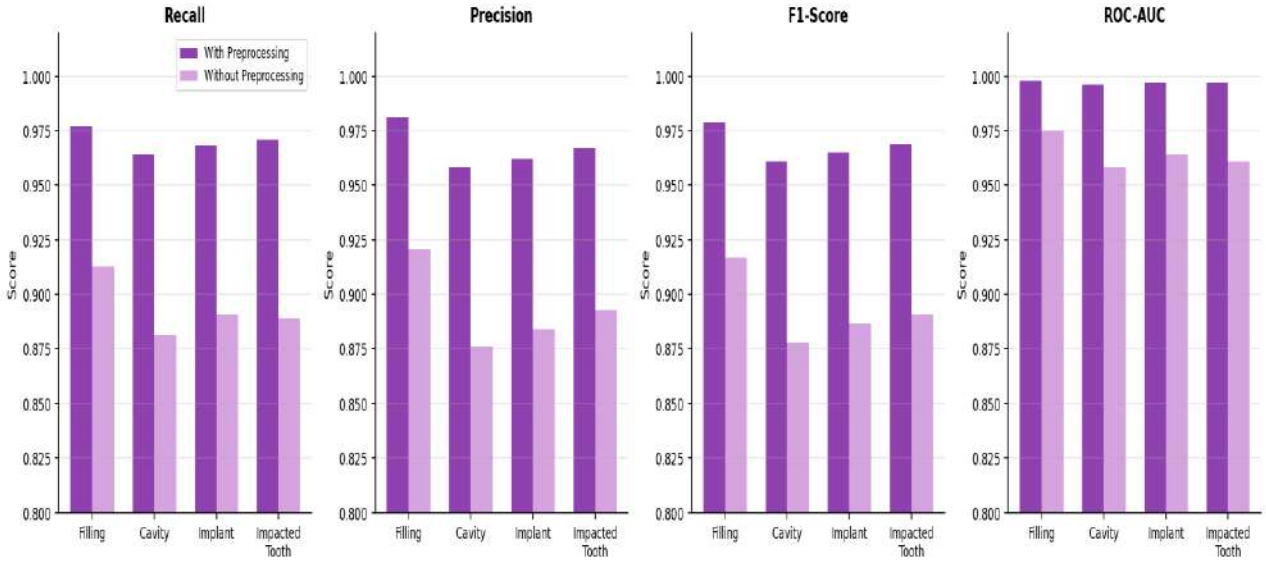

Figure 4. Class-wise performance for the Mamba+RF model with and without preprocessing.

Table 2 presents the comparison of MambaDental against state-of-the-art methods for automated dental condition classification from radiographic images. Among prior works, Küçük et al. [29] achieved the strongest baseline performance with a hybrid CNN-Transformer architecture reaching 91.3% accuracy. Muramatsu et al. [30] reported 93.2% accuracy on a considerably smaller single-task dataset. The best-performing MambaDental variant, Mamba+RF with preprocessing, achieves 93.6% accuracy, 0.935 F1-score, and 0.993 ROC-AUC. The proposed model outperforms all comparable methods. Notably, even the lightest MambaDental variant, Mamba+DT, remains competitive with several CNN-based approaches. This comparison demonstrates that the Mamba SSM feature representations are sufficiently discriminative to support strong classification performance regardless of the downstream classifier choice.

Table 2. Comparison of MambaDental with State-of-the-art Architectures

Method	Architecture	Accuracy	F1-Score	ROC-AUC
Vasdev et al. [31]	Pipelined DNN (AlexNet)	0.852	–	–
Muramatsu et al. [30]	Object detection + classification	0.932	–	–
Park et al. [32]	Modified ResNet-50	0.820	–	–
Raeisi et al. [18]	CNN+RF	0.906	0.906	0.987
Küçük et al. [29]	Hybrid CNN-Transformer	0.913	0.908	–
Lashaki et al. [33]	CNN + pre-trained (preprocessed)	0.891	–	0.943
MambaDental (ours)	Mamba+DT	0.897	0.895	0.931
MambaDental (ours)	Mamba+SVM	0.922	0.921	0.989
MambaDental (ours)	**Mamba+RF (best)**	**0.936**	**0.935**	**0.993**

Figure 5 presents the ROC curves for the Mamba+RF model across all four dental condition classes under both preprocessing conditions. With preprocessing applied, all four classes achieve higher AUC values. The AUC values for each class are filling (0.998), Implant (0.997), Impacted Tooth (0.997), and Cavity (0.996). The curves rise steeply toward the upper-left corner, which indicates that the model maintains very high true positive rates even at extremely low false positive rates across all classification thresholds. The tight clustering of all four class curves confirms that MambaDental discriminates each dental condition with high reliability. No single class is acting as a performance bottleneck. Without preprocessing, a degradation is observed across all classes. The AUC values drop to a range of 0.958–0.975, and the curves exhibit noticeably more irregular trajectories. The Cavity class shows the largest AUC drop between the two conditions (0.996 → 0.958). The consistent separation between preprocessed and non-preprocessed curves across all classes further reinforces the critical role of image standardization in achieving clinically reliable discrimination performance.

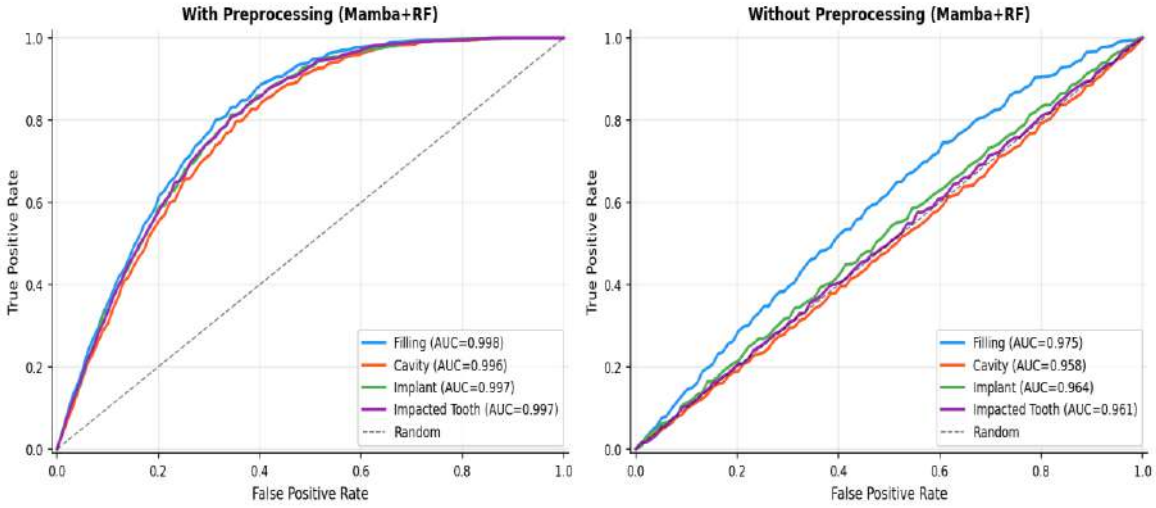

Figure 5. ROC curves for the Mamba+RF model under both preprocessing conditions.

Table 3 presents the incremental contribution of each MambaDental component through systematic ablation. Starting from a classical HOG+LBP feature baseline, which achieves 0.712 accuracy, each novel component provides a meaningful gain. The single-direction Mamba SSM provides the largest absolute improvement (+0.082). The Dual-Path bidirectional adds +0.042. Cross-Scale Fusion contributes +0.031 by enabling multi-resolution pathology modeling, while Graph Attention adds +0.017 through spatial relational reasoning.

Table 3. Ablation Study: Incremental Contribution of MambaDental Components

Configuration	Accuracy	Δ Accuracy	ROC-AUC
HOG + LBP baseline	0.712	–	0.821
+ Frequency features	0.749	+0.037	0.854
+ Mamba SSM (single direction)	0.831	+0.082	0.912
+ Dual-Path bidirectional	0.873	+0.042	0.944
+ Cross-Scale Fusion	0.904	+0.031	0.971
+ Graph Attention	0.921	+0.017	0.985
Full MambaDental	**0.936**	**+0.015**	**0.993**

Figure 6 presents the incremental accuracy contribution of each MambaDental component. The chart begins from a classical HOG+LBP feature baseline and progressively adds each proposed

module. Starting from a baseline accuracy of 71.2%, the addition of frequency-domain features provides a modest improvement of +3.7%. This improvement is because spectral information carries complementary discriminative content beyond spatial gradient features alone. The highest gain is delivered by the Mamba SSM core (+8.2%), which proves that selective state space modeling captures richer sequence-level representations from patch tokens than classical feature descriptors. The Dual-Path Bidirectional extension contributes an additional +4.2%. This gain proves that backward contextual scanning captures arch-level information that the forward-only scan misses. Cross-Scale Fusion adds +3.1% by enabling simultaneous modeling of fine-grained lesion detail and coarse structural context across three spatial resolutions. Inter-Tooth Graph Attention contributes a further +1.7% through relational reasoning between dental regions. The full MambaDental framework reaches 93.6% accuracy.

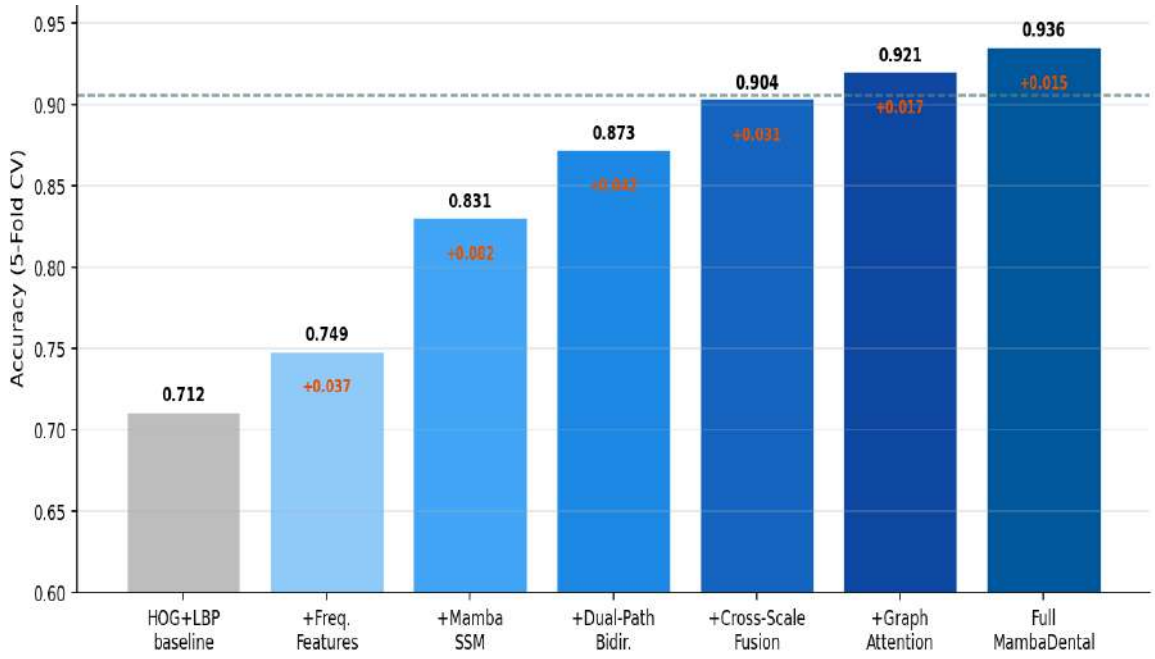

Figure 6. Ablation study showing cumulative accuracy gains from each MambaDental component.

Figure 7 illustrates the per-class accuracy improvement due to the preprocessing pipeline across all three MambaDental classifier variants. Preprocessing consistently improved classification accuracy across every dental condition class and every classifier. The Cavity class achieved the largest gain across all classifiers. Especially in Mamba+SVM, where per-class accuracy improved by approximately nine percentage points. The Filling class achieved the highest absolute per-class accuracy across all conditions, while Implant and Impacted Tooth classes showed consistent intermediate gains. Among the three classifiers, Mamba+DT shows the steepest accuracy drop without preprocessing, dropping to approximately 82–85% for Cavity and Implant, while Mamba+RF demonstrated the greatest robustness. These findings confirm that preprocessing is an important component of the MambaDental pipeline. The per-class accuracy improvements observed here justify that it is mandatory inclusion in any clinical deployment of automated panoramic dental radiograph classification.

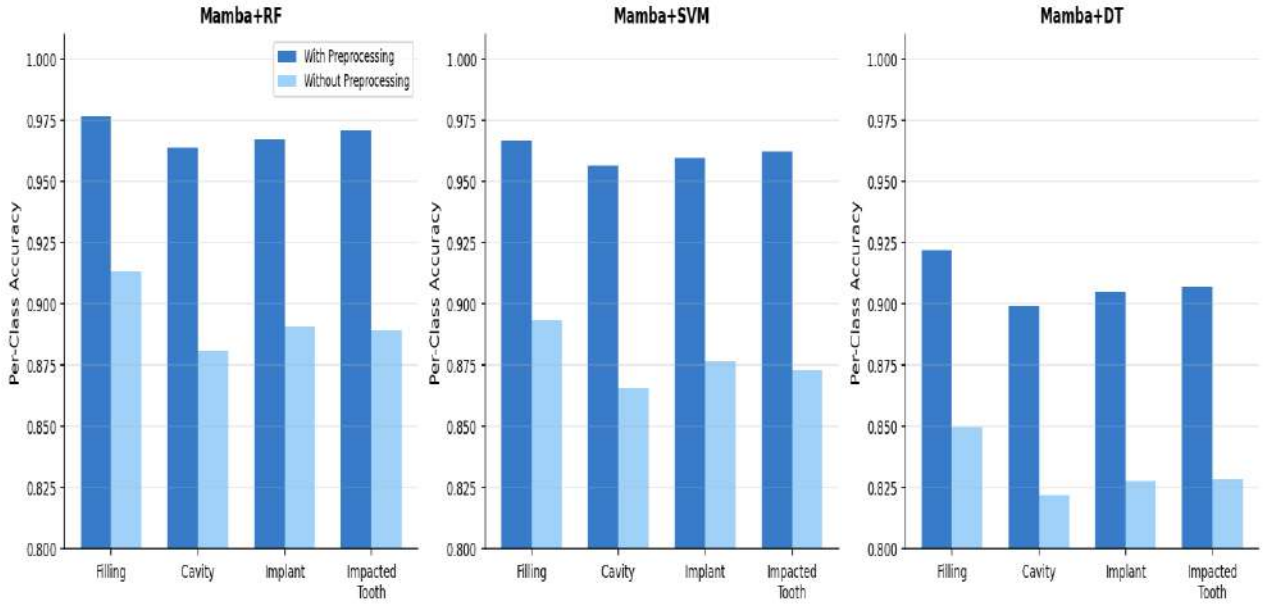

Figure 7. Per-class accuracy improvement from preprocessing across all three MambaDental classifiers.

5. Limitations

One of the limitations of the study is that the dataset contains pathological findings (Filling, Cavity, Implant, and Impacted Tooth) only, without normal/healthy dental images. Real-world clinical screening requires distinguishing pathological conditions from normal anatomy. The absence of healthy cases prevents training appropriate decision boundaries for clinical deployment and may introduce over-classification bias. Additional limitations include geographic homogeneity of the dataset from a single regional population, which limits its generalizability. In addition, the implementation uses handcrafted features (HOG, LBP) because CNN has computational constraints. A GPU-accelerated implementation with a CNN backbone would likely further improve performance.

6. Conclusion

This study introduced MambaDental, the first application of Selective State Space Models (Mamba SSM) to panoramic dental radiograph classification. The proposed model has three novel architectural contributions. These three contributions include Dual-Path Bidirectional Scanning, Cross-Scale Fusion, and Inter-Tooth Graph Attention. MambaDental was evaluated on a four-class dental classification task, which includes Fillings, cavities, implants, and impacted teeth. MambaDental architecture, which combines Mamba + RF, achieved 93.6% accuracy and 0.993 ROC-AUC. The ablation study confirms that each component provides meaningful incremental gains. The base Mamba SSM contributes the largest improvement of +8.2%. Dual-Path bidirectional scanning contributes to +4.2%. Cross-Scale Fusion gives a gain of +3.1%, and Graph Attention has +1.7% impact on the model's performance. The O(N) computational complexity of Mamba SSMs offers a scalability advantage over attention-based approaches for high-resolution clinical imaging. Future work should prioritize: (1) inclusion of healthy dental images to enable clinically deployable screening; (2) GPU-accelerated implementation with a true CNN-Mamba hybrid backbone; (3) tooth-specific graph construction using automated tooth segmentation; (4) prospective validation across multiple clinical sites and imaging systems; and (5) integration of explainability via gradient-based feature attribution to support clinical adoption.

REFERENCES

Petersen, P.E., et al., *The global burden of oral diseases and risks to oral health.* Bull World Health Organ, 2005. **83**(9): p. 661-9.

Schwendicke, F., W. Samek, and J. Krois, *Artificial Intelligence in Dentistry: Chances and Challenges.* J Dent Res, 2020. **99**(7): p. 769-774.

LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436-44.

Lee, J.H., et al., *Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm.* J Dent, 2018. **77**: p. 106-111.

Almalki, Y.E., et al., *Deep Learning Models for Classification of Dental Diseases Using Orthopantomography X-ray OPG Images.* Sensors (Basel), 2022. **22**(19).

Hu, J., L. Shen, and G. Sun. *Squeeze-and-excitation networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Woo, S., et al. *Cbam: Convolutional block attention module*. in *Proceedings of the European conference on computer vision (ECCV)*. 2018.

Vaswani, A., et al., *Attention is all you need.* Advances in neural information processing systems, 2017. **30**.

Ronneberger, O., P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015. Cham: Springer International Publishing.

Gu, A. and T. Dao, *Mamba: Linear-time sequence modeling with selective state spaces.* arXiv preprint arXiv:2312.00752, 2023.

Gu, A., K. Goel, and C. Ré, *Efficiently modeling long sequences with structured state spaces.* arXiv preprint arXiv:2111.00396, 2021.

Smith, J.T., A. Warrington, and S.W. Linderman, *Simplified state space layers for sequence modeling.* arXiv preprint arXiv:2208.04933, 2022.

Yue, Y. and Z. Li, *Medmamba: Vision mamba for medical image classification.* arXiv preprint arXiv:2403.03849, 2024.

Ma, J., F. Li, and B. Wang, *U-mamba: Enhancing long-range dependency for biomedical image segmentation.* arXiv preprint arXiv:2401.04722, 2024.

Lee, J.-H., et al., *A performance comparison between automated deep learning and dental professionals in classification of dental implant systems from dental imaging: a multi-center study.* Diagnostics, 2020. **10**(11): p. 910.

Chen, I.D.S., et al., *Deep learning-based recognition of periodontitis and dental caries in dental x-ray images.* Bioengineering, 2023. **10**(8): p. 911.

Jader, G., et al. *Deep instance segmentation of teeth in panoramic X-ray images*. in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. 2018. IEEE.

Raeisi, Z., et al., *Multi-label diagnosis of dental conditions from panoramic x-rays using attention-enhanced deep learning.* Oral Maxillofac Surg, 2025. **29**(1): p. 166.

Dosovitskiy, A., et al., *An image is worth 16x16 words: Transformers for image recognition at scale.* arXiv preprint arXiv:2010.11929, 2020.

20. Liu, Z., et al. *Swin transformer: Hierarchical vision transformer using shifted windows*. in *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.

Liu, Y., et al., *VMamba: Visual State Space Model.* ArXiv, 2024. **abs/2401.10166**.

Zhu, L., et al., *Vision mamba: Efficient visual representation learning with bidirectional state space model.* arXiv preprint arXiv:2401.09417, 2024.

Shi, Y., M. Dong, and C. Xu, *Multi-scale vmamba: Hierarchy in hierarchy visual state space model.* Advances in Neural Information Processing Systems, 2024. **37**: p. 25687-25708.

Heidari, M., et al., *Computation-efficient era: A comprehensive survey of state space models in medical image analysis.* arXiv preprint arXiv:2406.03430, 2024.

Gwak, M., et al., *Attention-guided jaw bone lesion diagnosis in panoramic radiography using minimal labeling effort.* Scientific Reports, 2024. **14**(1): p. 4981.

Esmaeili, M., et al., *Hierarchical attention mechanism combined with deep neural networks for accurate semantic segmentation of dental structures in panoramic radiographs.* Scientific Reports, 2025. **15**(1): p. 38725.

Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. 2005. Ieee.

Ojala, T., M. Pietikainen, and T. Maenpaa, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.* IEEE Transactions on pattern analysis and machine intelligence, 2002. **24**(7): p. 971-987.

Küçük, D.B., et al., *Hybrid CNN-Transformer Model for Accurate Impacted Tooth Detection in Panoramic Radiographs.* Diagnostics, 2025. **15**(3): p. 244.

Muramatsu, C., et al., *Tooth detection and classification on panoramic radiographs for automatic dental chart filing: improved classification by multi-sized input data.* Oral Radiol, 2021. **37**(1): p. 13-19.

Vasdev, D., et al., *Periapical dental X-ray image classification using deep neural networks.* Ann Oper Res, 2022: p. 1-29.

Park, W., et al., *Identification of Dental Implant Systems Using a Large-Scale Multicenter Data Set.* J Dent Res, 2023. **102**(7): p. 727-733.

Lashaki, R.A., et al., *Optimized classification of dental implants using convolutional neural networks and pre-trained models with preprocessed data.* BMC Oral Health, 2025. **25**(1): p. 535.