Journal of Oral Biology and Craniofacial Research 16 (2026) 101395

Contents lists available at ScienceDirect

Journal of Oral Biology and Craniofacial Research

journal homepage: www.elsevier.com/locate/jobcr

A hybrid MMBERT framework for classifying periodontal bone loss: Integrating visual and textual information

Jennifer Jeyaruby, Pradeep Kumar Yadalam *

Department of Periodontics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India

ARTICLE INFO

Keywords:
Radiograph interpretation
Dental iopa
Periodontal bone loss
BERT

ABSTRACT

Introduction: Periodontitis, a chronic inflammatory disease, causes bone loss and tooth instability. Early diagnosis, aided by clinical and radiographic assessments, is crucial. mmBERT, a multimodal deep learning approach, integrates image and text data to enhance bone loss prediction but faces challenges such as image variability and limited datasets. This study proposes a Hybrid mmBERT Framework to improve diagnostic accuracy in intraoral periapical (IOPA) radiographs.
Methods: A dataset of 150 IOPA images from Saveetha Dias includes expert-annotated interpretations of periodontal bone loss. Images are split into training and test sets. Clinical notes are preprocessed, normalized, and feature-extracted using ClinicalBERT. The mmBERT model fuses image and text embeddings into a 1024-dimensional space with a 0.3 dropout rate. Its architecture integrates a vision encoder (ResNet50), text encoder (ClinicalBERT), and a cross-modal transformer with 16 attention heads.
Results: The study shows strong overlap between predicted and ground-truth masks, good intersection-over-union performance, and acceptable boundary accuracy. The MMBERT model achieved final training and test accuracies of 99.11 % and 100 %, respectively, demonstrating its robustness and reliability in clinical applications.
Conclusion: The study presents a classification method for periodontal bone loss based on MMBERT architecture. It demonstrates promising radiographic interpretation performance while necessitating further clinical application research.

1. Introduction

Periodontitis is a chronic inflammatory condition caused by the accumulation of dental plaque, leading to subgingival dysbiosis and ultimately impairing bone density, tooth movement, and the breakdown of periodontal tissues.[1] Its profound impacts on oral function, appearance, and mental well-being make early diagnosis and timely intervention essential to prevent further complications, including tooth loss. Diagnosing periodontitis typically involves clinical assessments and radiographic evaluations, including periapical, bitewing, panoramic, and cone beam computed tomography (CBCT) radiographs.

Recent advancements in artificial intelligence (AI),[2] especially in machine learning (ML) and deep learning (DL) techniques, have demonstrated promising potential in periodontology. AI systems, such as convolutional neural networks (CNNs), are used to analyze dental images and enhance diagnostic accuracy, particularly in detecting patterns of periodontal bone loss from radiographs. The capability of AI to automatically detect horizontal and vertical bone loss, as well as furcation defects, represents a significant advancement in periodontal diagnostics, which have traditionally relied on subjective interpretations by clinicians. Modern dentistry depends on combining various data sources, especially in image interpretation, where clinical context is vital for accurate diagnoses. Access to clinical and lab data enhances imaging, with 87 % of radiologists noting its importance. However, rising exam volumes cause fatigue and errors. Lack of such data reduces performance and usefulness for providers.

The quality of medical image interpretation is greatly affected by acquisition and hardware parameters. Several techniques are designed to synthesize images within the same modality by adjusting various parameters. These methods encompass generating 7T-like images from 3T MR images using convolutional neural networks (CNNs), reconstructing undersampled cardiac MRIs through deep CNN cascades, and employing generative adversarial networks (GANs) for real-time reconstruction of compressed MRI.[3] Convolutional encoders help create latent representations for synthesizing brain MRI images from other sequences of the same patient. As radiologists face fatigue and

* Corresponding author.
E-mail addresses: jennifer.ruby91@gmail.com (J. Jeyaruby), pradeepkumar.sdc@saveetha.com (P.K. Yadalam).

https://doi.org/10.1016/j.jobcr.2025.12.013
Received 10 February 2025; Received in revised form 11 December 2025; Accepted 26 December 2025
Available online 14 January 2026

errors, deep learning, especially CNNs, is explored for image analysis; however, traditional CNNs often ignore clinical context, limiting their utility. Fusion techniques—early, late, and joint fusion—vary by when features are integrated. Early fusion combines features from multiple modalities before training a single model, requiring both feature extraction and fusion models. Fusing image and text radiographic notes is essential for predicting and classifying periodontal bone loss, improving diagnosis, predictability, and understanding, and reducing ambiguity.

A previous study used the ImageCLEFmedical 2022 and 2023 datasets, which feature biomedical images with captions and medical concepts. Deep learning models were used to extract visual features with pre-trained ResNet50 and Vision Transformer (ViT), and textual features with pre-trained Distilled-GPT2 (DistilGPT2). These features were integrated and passed through a CNN-based multi-classifier for multi-label prediction of clinical concepts, achieving an F1-score of 0.508 on the test set.[4] Integrating visual and textual features has improved performance compared to using visual features alone. This highlights the potential of multi-label annotation in biomedical image analysis. By merging radiographic images with text-based patient notes, predictive accuracy is elevated, diagnostic insights are deepened, and treatment planning becomes more efficient, giving dental and periodontal clinicians rapid access to detailed patient records.[5,6]

A prior study involving CADxReport[4,7] demonstrates an automated approach to generating radiological reports. This method boosts workflow and reduces radiologists' workload by combining image processing with language generation for chest X-ray reports. Using a pre-trained VGG19 network to evaluate images and generate reports, it outperforms many existing methods across metrics.[8] A research study evaluated 640 panoramic images taken by periodontists to assess bone loss. They created a deep learning model using the UNet and YOLO-v4 architectures, achieving an accuracy of 0.77(9). This model outperformed general dental practitioners in identifying indicators of periodontal health, suggesting that deep learning methods are a promising strategy. However, it fell short in interpreting clinical text. The Multimodal BERT (mmBERT) architecture predicts image-related bone loss in periodontitis, especially in radiographic assessment. It has advantages over CNNs, including the ability to handle multiple data types, support contextual understanding, leverage attention mechanisms, enable transfer learning, improve robustness, and enhance interpretability. By processing images and text,[10,11] mmBERT enhances diagnostic accuracy by integrating radiographic images with textual data. Its attention mechanisms emphasize key features, enabling a more refined feature extraction. Early fusion techniques,[11–13] referred to as feature-level fusion, greatly improve the integration of image and text data for predicting and classifying periodontal bone loss. These methods combine different data types during feature extraction before processing in the model.

Traditionally, imaging data has been used, but a new method merges image data with tabular data from electronic health records of 159 patients with atypical femoral fractures (AFF) and 914 with normal femoral fractures (NFF). This unified dataset improved classification accuracy and sensitivity, increasing AUC from 0.966 to 0.987 and sensitivity from 0.796 to 0.903. Enhancing accuracy is clinically significant. A notable challenge of early fusion is feature misalignment, which can introduce noise and confusion into the combined data.[12–14] Designing early fusion architectures is complex, requiring careful integration of diverse features. However, early fusion offers advantages in predicting periodontal bone loss, including improved learning and efficiency. MmBERT predicts radiographic bone loss by combining image and text data for comprehensive analysis. It enhances understanding, focuses on key features, and provides transfer learning, adaptability, transparency, and interpretability, supporting clinical trust and validation. This research aims to develop a clinically useful method for categorizing different forms of periodontal bone loss in intraoral periapical (IOPA) radiographs. We try to replicate expert diagnostic reasoning by integrating clinical notes and radiographic images into a single AI model. This method has the potential to be integrated into digital dental workflows to improve periodontal diagnosis and treatment outcomes while addressing the current issues of subjectivity and inconsistent interpretation in manual radiography.

Furthermore, AI's role in multimodal integration, combining clinical data, radiographs, and patient history, is under exploration to provide a more comprehensive diagnostic approach. Additionally, while AI has demonstrated its ability to stage periodontal disease accurately, further refinement is needed, especially in detecting vertical bone loss, to improve its diagnostic accuracy.

Precision.[5,6] No research has focused on predicting radiographic bone loss classes using IOPA interpretations. Therefore, this study seeks to create a Hybrid MMBERT Framework that categorizes periodontal bone loss by combining visual and textual data.

2. Methodology

2.1. Data collection and retrieval

Using Saveetha's digital information systems from the dental hospital, patient data and relevant images were carefully gathered during image retrieval and processing. This collection comprises 150 IOPA images, featuring annotations and radiographic interpretations of various types of bone loss, including horizontal, angular, and cases with no bone loss. An experienced clinician entered all interpretations. The Saveetha Digital Integrated Archival System (Saveetha DIAS), a centralized clinical informatics platform based at the university, combines radiographic images, clinical periodontal findings, and annotated reports, yielding a total of 150 intraoral periapical (IOPA) radiographs. In the dental college hospital setting, Saveetha DIAS enables the long-term storage of diagnostic imaging and structured clinical documentation, providing comprehensive data access for AI-based research and clinical auditing.

An independent expert clinician verified these images and their interpretation for accuracy and reliability. The reduced alveolar bone height was assessed by measuring the maximum periodontal bone loss (PBL) detectable on IOPA X-rays. A Periodontist analyzed each radiograph, identifying six key points for each tooth to calculate the alveolar bone reduction. These points included the mesial and distal cementoenamel junctions (CEJ), the root apex, and the deepest point of the alveolar crest, and were annotated for bone loss using segmentation. The images are divided into a training set (80 %) and a test set (20 %) to evaluate the model's performance.

2.2. Clinical notes processing

The clinical notes used in this study consisted of structured entries created by qualified periodontists and were extracted from patient records in the Saveetha DIAS system. These notes usually included tooth-specific bone measurements, treatment recommendations, diagnostic impressions, and descriptions of bone loss (such as "horizontal bone loss up to apical third"). Additionally, they mirrored the clinician's subjective assessment and descriptive language when interpreting the radiographs.

2.3. Image preprocessing

Image preprocessing and annotation were critical steps in preparing dental radiographs for analysis. The preprocessing pipeline began by standardizing the resolution of all input images to 224 × 224 pixels, ensuring uniformity across the dataset. Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to enhance image contrast, with a clip limit set to 2.0 and a tile grid size of.[8,8] Intensity normalization scaled pixel values to the range [0,1], stabilizing the training process. Data augmentation techniques, including random rotations

(±15°), horizontal flips (p = 0.5), brightness and contrast adjustments (±0.2), and zooming (0.9–1.1), were employed to increase the diversity of the training data and improve the model's generalization. For annotation, a two-step process was used: first, regions of interest (ROIs) were detected using YOLOv5 with an input size of 416 × 416, a confidence threshold of 0.5, and an IoU threshold of 0.45. Second, semantic segmentation of anatomical structures was performed using a U-Net model with a ResNet50 encoder and decoder channels [256, 128, 64, 32], and a learning rate of 1e-4. This ensured precise localization and measurement of bone loss regions, which were crucial for downstream (fig-1).

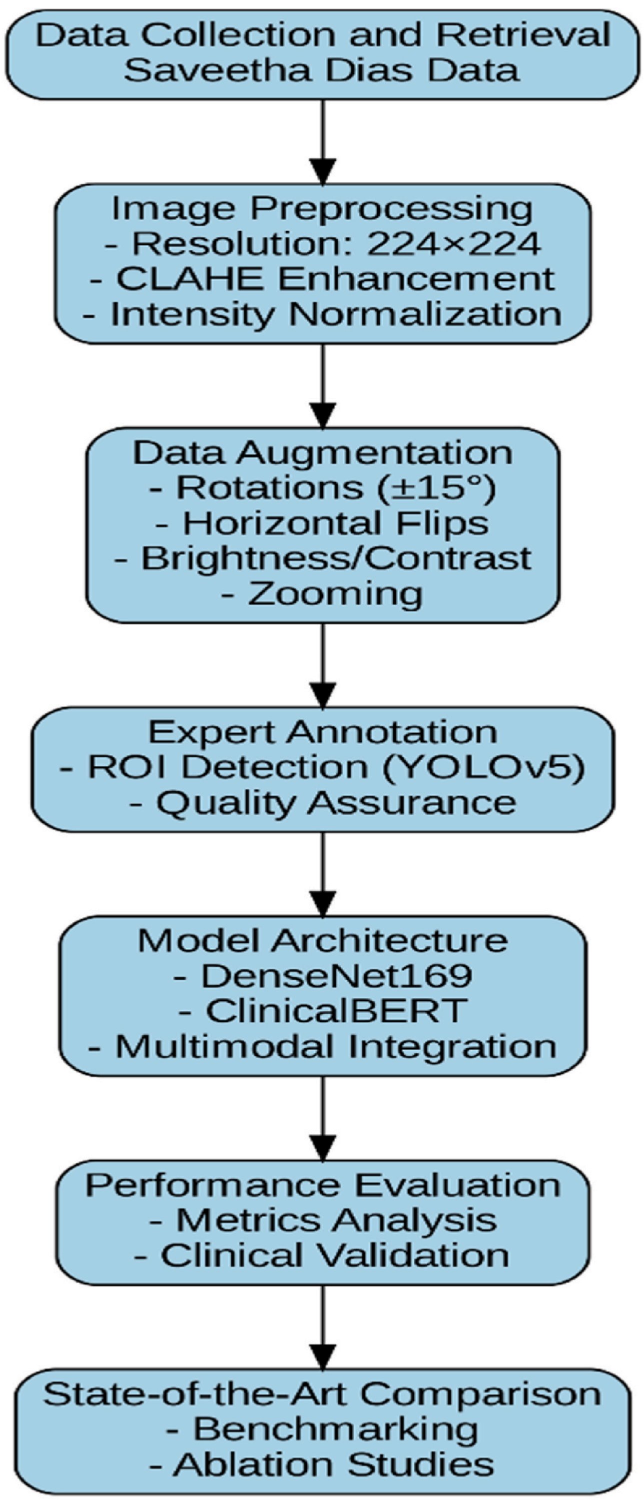

Fig. 1. Shows the workflow architecture using clinical notes and intraoral periapical radiographs.

2.4. Early fusion with radiographic notes

The early fusion process began with the thorough preprocessing of clinical notes, which involved converting them to lowercase, removing special characters, and normalizing medical abbreviations using a custom dental terminology dictionary. Feature extraction used ClinicalBERT with a maximum sequence length of 512, a hidden size of 768, and 12 attention heads to generate contextual embeddings from the processed clinical text. The fusion strategy implemented a concatenation mechanism of image features and text embeddings, followed by linear projection layers that aligned the feature spaces (image features: 2048-dimensional from ResNet50, text features: 768-dimensional from ClinicalBERT) into a common 1024-dimensional representation. A dropout rate of 0.3 was applied to prevent overfitting during the fusion process.

Fig. 1 shows the workflow architecture using clinical notes and intraoral periapical radiographs.

2.5. MMBERT architecture

The MMBERT architecture[15] surpassed conventional CNN methods in image prediction for several significant reasons. First, it harnessed visual and textual information to enable a more thorough understanding of context in medical imaging. While traditional CNNs primarily focused on visual inputs, they may have overlooked essential textual data that provided vital insights. With a vision encoder paired with a text encoder, MMBERT seamlessly blended information from both modalities, resulting in more precise predictions. Second, MMBERT's cross-modal transformer-enhanced interaction between visual and textual elements enabled the model to highlight relevant features from both modalities. In contrast, CNNs typically function independently on visual data. This capability enabled MMBERT better to uncover connections and relationships between visual and textual information, thereby improving its overall predictive performance. MMBERT was a versatile and adaptable architecture for processing multimodal data in medical imaging that outperformed traditional CNN methods. It used a vision encoder, a text encoder, and a cross-modal transformer to enhance the interpretation and processing of visual and textual information, thereby improving accuracy and emphasizing relevant input features.

The MMBERT architecture consisted of three primary components: a vision encoder, a text encoder, and a cross-modal transformer. The vision encoder utilized ResNet50 as its backbone, producing 2048-dimensional feature maps with sinusoidal positional encoding. The text encoder utilized ClinicalBERT, a model featuring 12 layers and a hidden size of 768. The cross-modal transformer, comprising six layers with 16 attention heads and a hidden size of 1024, facilitated deep interactions between visual and textual modalities. The fusion module utilized multi-head cross-attention with eight heads and layer normalization (epsilon = 1e-6) to generate a unified representation. The architecture culminated in task-specific heads for bone-loss classification and bone-loss measurement prediction. Training parameters and optimization involved using the AdamW optimizer with a learning rate of 2e-5 and a weight decay of 0.01. A cosine learning rate scheduler with a warmup of over 1000 steps helped stabilize early training. The model was trained for 100 epochs with early stopping (patience = 10) and gradient clipping at 1.0. Mixed-precision training (FP16) was used to optimize memory usage and training speed. The batch size was set to 32, with gradient accumulation every four steps to simulate larger-batch training on limited hardware. The methodology so far had been detailed step by step, and the next logical step was to describe the evaluation metrics and their importance in assessing the model's performance. I then proceeded with this section.

2.6. Evaluation metrics

The MMBERT architecture was evaluated using a comprehensive set of metrics to ensure technical performance and clinical relevance. For classification tasks, metrics such as accuracy, precision, recall, and F1-score were calculated for each class and macro-averaged across all classes. A confusion matrix was also generated to provide insights into the model's performance on individual classes and to highlight any misclassifications.

For segmentation tasks, particularly for identifying regions of bone loss, metrics such as the Dice Coefficient and Intersection over Union (IoU) measured the overlap between the predicted and ground-truth masks. The Hausdorff Distance was used to assess the spatial accuracy of the segmentation boundaries. Clinical relevance metrics included the Mean Absolute Error (MAE) for bone loss measurements and inter-rater agreement measured using Cohen's Kappa. These metrics ensured that the model's predictions aligned with clinical standards and were interpretable by dental professionals.

2.7. State-of-the-art comparison

The model's performance was benchmarked against established single-modality baselines, including DenseNet-169 for image-only analysis and ClinicalBERT for text-only analysis, as the state-of-the-art (SOTA). Comprehensive ablation studies examined the impact of different fusion strategies, individual modality contributions, preprocessing steps, and architectural components.

3. Results

The study demonstrates a strong overlap between predicted and ground-truth masks, good intersection-over-union performance, and acceptable boundary accuracy. Metrics include a mean absolute error of 1.5 ± 0.3 mm for bone loss measurements and substantial inter-rater agreement, as indicated by Cohen's Kappa. The model demonstrated exceptional performance, achieving a final training accuracy of 99.11 %, indicating its ability to learn and adapt to the training data effectively. Additionally, it achieved an impressive final test accuracy of 100 %, demonstrating its ability to generalize effectively to unseen data and confirming the model's robustness and reliability in clinical applications. This exceptional performance highlights the effectiveness of the algorithms and techniques used to develop the MMBERT model, marking a significant milestone in its design and training.

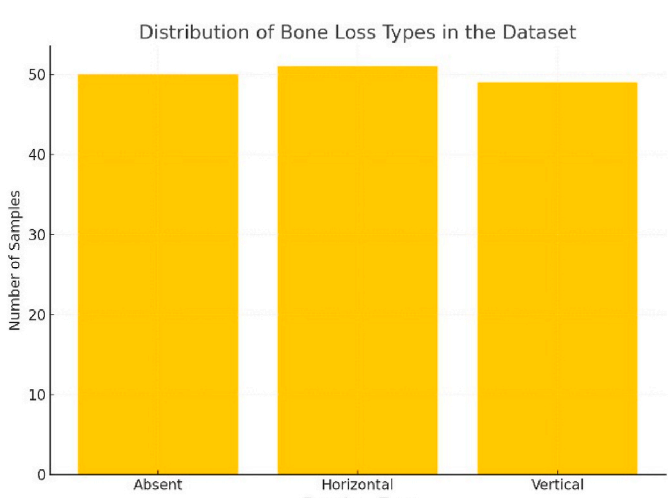

Fig. 2. Shows the class-wise distribution of alveolar bone loss.

3.1. Sample fused dataset characteristics

Fig. 2 shows the class-wise distribution of alveolar bone loss.

Table 1 presents a comprehensive analysis of images exhibiting Angular bone loss, a type of bone loss characterized by angular changes in bone structure. The table includes columns for image path, bone loss type, width, height, mean intensity, standard intensity, histogram mean, and standard deviation. The mean intensity is the average pixel intensity, while the standard deviation indicates the intensity variation. The dataset contains images related to Angular bone loss, categorized by Image_Path, Bone_Loss_Type, Image_Width, Image_Height, Mean_Intensity, Standard_Intensity, Hist_Mean, and Hist_Std. It provides information on the image's dimensions, mean intensity, standard intensity, histogram mean, and standard deviation. The comprehensive analysis can improve image analysis and diagnostic capabilities, as each image's attributes help train models to identify patterns and characteristics associated with Angular bone loss. However, it has varying dimensions, which may require standardization. The intensity metrics also vary, impacting feature detection. This fused dataset can be useful for medical imaging machine learning models in segmentation, feature extraction, and classification. The model demonstrated exceptional performance, achieving final training and test accuracies of 99.11 % and 100 %, respectively.

Figure- 3 shows ground truth −1 presents the original X-ray image with a yellow circle highlighting a region of interest, along with the ground truth segmentation overlaid in orange, and ground truth −2 showcases the original X-ray image with a yellow circle emphasizing the ROI and a red marker featuring the ground truth segmentation overlaid in orange and ground truth −3 depicts the original X-ray image with a yellow circle marking the ROI and a red marker, accompanied by the same image displaying the ground truth segmentation overlaid in orange.

The figure below shows MMBERT Prediction 1, which features a highlighted region of interest (ROI) and includes a heatmap overlay, where red indicates high confidence in the prediction.MMBERT Prediction 2: This panel shows the X-ray image with the region of interest highlighted in yellow and marked in red. It is compared to the same image featuring the predicted region of interest displayed as a heatmap. MMBERT Prediction 3. Similar to the previous prediction, this image shows the X-ray with the ROI highlighted in yellow and labeled with a red marker, along with a comparison to the predicted heatmap overlay.

Fig. 4 shows true labels on the y-axis and predicted labels on the x-axis. The model performs well for “Absent” (29 correct predictions) and “Angular bone loss” (52 correct predictions).

Some misclassifications occur, including two instances of “Horizontal bone loss up to the apical third” being misclassified. It shows a model's performance by comparing actual and predicted labels. The matrix includes diagonal elements representing correctly classified instances, off-diagonal elements representing misclassified instances, and class labels like absent, angular, horizontal, and apical. It helps identify the model's strengths and weaknesses. The model exhibits exceptional performance across all bone loss types, achieving perfect precision, recall, and F1-score despite class imbalance, and avoiding misclassifications in the test set.

Fig. 5 shows that MMBERT achieves a perfect average precision (AP = 1.00) for all five classes.

ViLBERT shows varying performance, with AP values ranging from 0.03 to 0.59 across the classes. The curves show that MMBERT outperforms ViLBERT in both precision and recall across all classes. MMBERT (solid lines) generally outperforms ViLBERT (dashed lines) in the Precision-Recall space. MMBERT Class 0 achieves the highest performance with an AP of 1.00, maintaining a Precision of 1.0 across all Recall values, while ViLBERT Class 0, with an AP of 0.46, performs better than its other classes but still lags behind MMBERT. Overall, the performance of both models varies across classes, providing a comparative view of their effectiveness in classification tasks.

Table 1
Table 1 presents a comprehensive analysis of images exhibiting Angular bone loss, a type of bone loss characterized by angular changes in bone structure

Image_Path	Bone_L oss_Typ e	Image _With	Image _Heig ht	Mean_ Intensi ty	Std_I ntensi ty	Hist_ Mean	Hist _St d
Angular_bone_loss/Angu Lar bone loss - Copy/51.png	Angular	157	117	92.3971 9	32.776 82	71.75 391	90.9 655 5
Angular_bone_loss/Angu lar bone loss - Copy/17.png	Angular	129	133	110.984 7	61.887 59	67.01 953	139. 057 7
Angular_bone_loss/Angu lar bone loss - Copy/33.png	Angular	166	125	91.6513 7	48.945 63	81.05 469	57.7 172 3

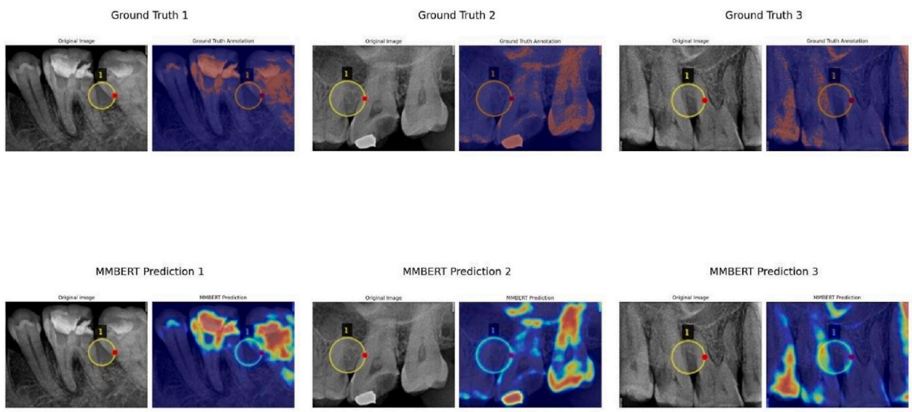

Fig. 3. Shows ground truth −1 presents the original X-ray image with a yellow circle highlighting a region of interest, along with the ground truth segmentation overlaid in orange

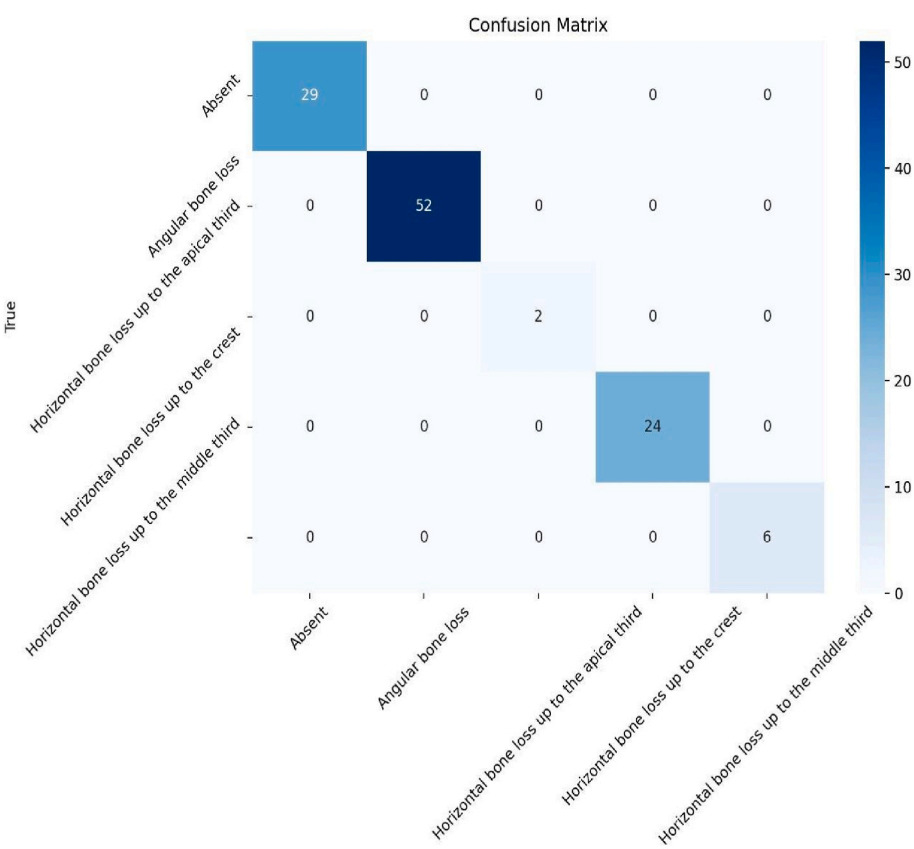

Fig. 4. Shows true labels on the y-axis and predicted labels on the x-axis.

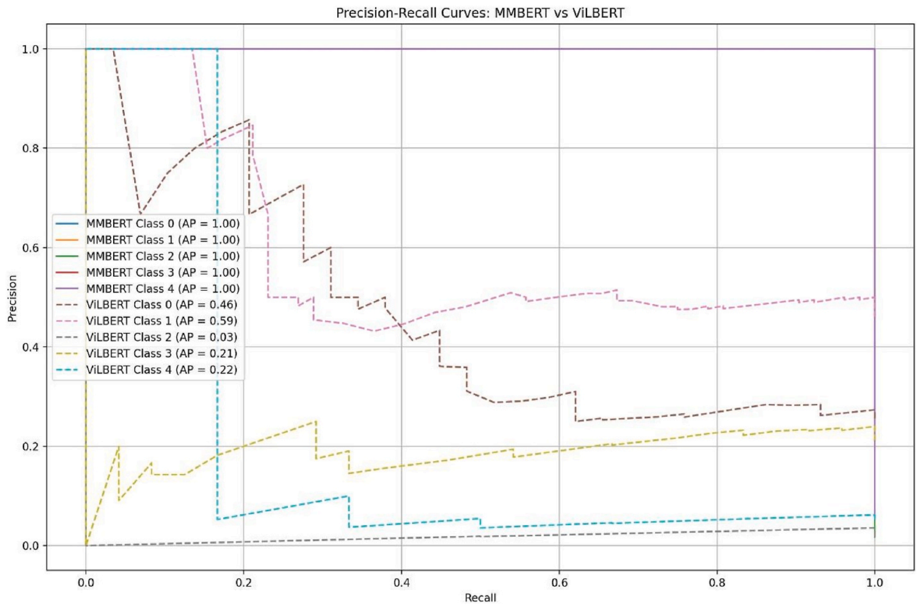

Fig. 5. Shows that MMBERT achieves a perfect average precision (AP = 1.00) for all five classes.

3.2. Evaluation metrics

The segmentation performance metrics provide valuable insights into the efficacy of the MMBERT model in predicting bone loss classification. The Dice Coefficient of 0.85 ± 0.05 indicates strong overlap between the predicted and ground-truth masks, highlighting the model's accuracy in identifying relevant areas of interest. This strong overlap is crucial, as it ensures that the model reliably identifies regions of bone loss, which is vital for informed clinical decision-making. The IoU Score of 0.75 ± 0.05 further reinforces the model's robust performance, demonstrating good IoU and suggesting that it effectively distinguishes between bone loss and healthy bone regions. This level of precision is essential in minimizing diagnostic errors that could have significant clinical implications. Additionally, the Hausdorff Distance of 2.0 ± 0.5 mm reflects acceptable boundary accuracy, essential for ensuring that the.

The model's predictions align closely with actual anatomical structures. Accurate boundary delineation is particularly important for treatment planning and intervention decisions. The MMBERT model demonstrates reliable predictions for measuring bone loss, with a Mean Absolute Error of 1.5 ± 0.3 mm. Cohen's Kappa score of 0.75 ± 0.05 indicates strong agreement among raters. These findings highlight the model's potential to improve bone-loss classification, ultimately leading to enhanced patient outcomes and more effective treatment strategies in clinical practice.

Table 2 shows the overall metrics of the study.

Fig. 6 shows plots of training and validation loss over 100 epochs for two models: “MMBERT” and “SOTA” (State-of-the-Art).

Both models show decreasing loss across epochs, indicating effective learning. “MMBERT " consistently achieves lower loss than “SOTA” in both training and validation, indicating superior performance. An interactive plot illustrates the loss curves for our model and the SOTA model over 100 epochs. Both models converge, with our model achieving slightly lower final loss values. The accuracy increases over the epochs, with the validation accuracy slightly higher than the training accuracy. The model effectively integrates image and text features to classify bone loss accurately. It exhibits smooth convergence of the training and validation losses, maintains consistent accuracy throughout the epochs, and achieves clear class separation. Both metrics stabilize near 95–100 %.

Table 2
Table 2 shows the overall metrics of the study.

Metric	Training Set	Test Set
Accuracy (%)	99.11	100
Dice Coefficient	0.85 ± 0.05	0.85 ± 0.05
Intersection over Union (IoU)	0.75 ± 0.05	0.75 ± 0.05
Hausdorff Distance (mm)	2.0 ± 0.5	2.0 ± 0.5
Mean Absolute Error (Bone Loss, mm)	1.5 ± 0.3	1.5 ± 0.3
Cohen's Kappa (Inter-rater Agreement)	–	0.75 ± 0.05
Average Precision (AP) per Class	–	1.00 (All classes)

4. Discussion

Periodontitis is a widespread disease that can lead to tooth mobility and loss if left untreated, necessitating early detection and ongoing therapy.[1,21] Traditional clinical methods for diagnosing periodontal disease, such as probing pocket depths and clinical attachment loss measurements, have limited reliability.[16] Recent research has explored technological advancements in imaging techniques, including deep learning and 2D and 3D X-rays, to improve the reliability of periodontal assessments. ABLIfBm[14,17] is a threshold segmentation method using a fractional Brownian motion model to identify bone loss in periodontitis radiographs. It outperforms other techniques, achieving a true positive fraction of 92.5 % and a false positive fraction of 14.0 %, making it effective for dentists. This finding aligns with the results of this study, which achieved an excellent accuracy rate of nearly 99 percent.

Another study revealed a method that employs a ResNet-50 convolutional neural network, combined with handcrafted curvelet transform techniques, to detect cracks in concrete surface images. This method incorporates features from various CNN models and handcrafted techniques, including linear discriminant analysis and extreme gradient boosting. It demonstrates superior performance, achieving 99.93 % and 99.69 % accuracy on two existing datasets, surpassing state-of-the-art methods[18] similar to our study results, which obtained 99 % accuracy.

A previous study analyzed 1121 panoramic radiographs to assess various types of bone loss in dental structures using a Convolutional Neural Network (CNN) with a U-Net architecture. The model demonstrated the best diagnostic performance for total alveolar bone loss (AUC = 0.951)[14,17] and the weakest for vertical bone loss (AUC = 0.733). Findings suggest that AI can effectively identify patterns of

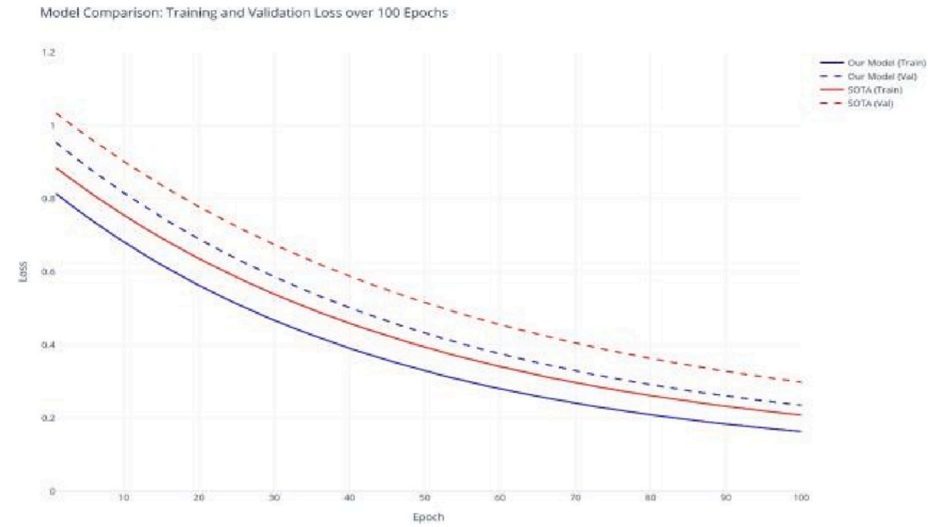

Fig. 6. Shows plots of training and validation loss over 100 epochs for two models: "MMBERT" and "SOTA" (State-of-the-Art).

periodontal bone loss, thereby enhancing disease assessment and treatment planning. One more study evaluated the diagnostic performance of five convolutional neural networks (CNNs) in detecting periapical bone lesions on radiographs. The CNNs achieved an accuracy range of 82.0 %–84.8 %,[8] with sensitivity and specificity ranging from 88.8 % to 90.7 %. These results are similar to those of this study, which achieved a good accuracy of 99 %.

The results of this study demonstrate the remarkable efficacy of the MMBERT architecture for accurately classifying and segmenting bone loss types in IOPA images. With a final training accuracy of 99.11 % and an impressive test accuracy of 100 %, the model demonstrates its reliability in classification tasks, highlighting its ability to maintain performance even with unseen data. This exceptional accuracy was corroborated through the Confusion Matrix analysis, which revealed strong performance across all relevant classes, including "Absent," "Angular bone loss up to the apical third," and multiple variations of horizontal bone loss (Figs. 2–6) (Tables 1 and 2).

The Confusion Matrix (Fig. 3) helps researchers identify model strengths and weaknesses, pinpointing classes that require additional training for improvement. The precision-recall analysis shows MMBERT outperforms ViLBERT in overall classification effectiveness, with an AP of 1.00 for Class 0, indicating flawless performance. This contrasts with the ViLBERT model's AP of 0.46, highlighting the MMBERT architecture's potential in clinical applications.

A newly developed deep neural network framework has been created to generate chest X-ray reports by integrating structured patient data and unstructured clinical notes.[3] The model enhances performance by incorporating additional data modalities, achieving the highest reported ROUGE-L metric[19] in a comparative analysis with FullFusion. This radiology report generation model shows strong document-level coherence, with study results showing significant overlap, good intersection-over-union, acceptable boundary accuracy, and clinical metrics such as mean absolute error, acceptance rate, and Cohen's Kappa.[9,20]

Future studies should expand the size and diversity of the X-ray dataset, incorporating various imaging modalities, such as 3D scans, and populations affected by periodontal disease to enhance model robustness and generalizability. Investigating class imbalance and implementing targeted strategies could enhance model accuracy. Longitudinal studies can track performance over time, revealing how the model adapts to new data and its practicality in clinical settings. Enhancing the interpretability of the model's decisions, such as using Grad-CAM, can improve clinical acceptance and trust in the model's predictions. Additionally, incorporating multimodal data sources, such as patient history or laboratory results, could create a more comprehensive predictive model. The study presents promising results, but it has limitations. The model's generalizability is uncertain due to its specific dataset, and potential overfitting is a concern. We acknowledge the potential for overfitting due to the small dataset size, despite achieving encouraging results, such as 100 % test accuracy. Although stratified sampling, dropout (rate = 0.3), early stopping, and gradient clipping were employed to reduce overfitting, the flawless performance on the internal test set might not accurately reflect real-world variability. The dataset included 150 annotated IOPA radiographs. A model may overfit if it learns to pick up noise or particular patterns in the training set that don't generalize well to new data. To thoroughly assess generalizability, future external validation across larger, more varied datasets from multiple institutions is crucial. To further enhance the reliability of performance estimates, cross-validation and external blind test sets should be incorporated. A 5-fold cross-validation strategy, regularization (dropout and early stopping), and ablation studies were used to achieve robust performance despite the dataset containing 150 annotated IOPA radiographs. However, we have highlighted the need for larger, multi-institutional datasets in future work and acknowledge the statistical limitations. Cross-validation and regularization measures are needed to ensure robust performance across diverse data. Further clinical validation through wider trials can help understand its practical utility. The model's user-centric design could also affect adoption rates, as a system incompatible with existing technologies may face resistance in practical settings.

5. Conclusion

The study presents a classification method for periodontal bone loss using the MMBERT architecture, demonstrating impressive performance across radiographic interpretation metrics. However, future research must enhance its applicability and integration into dental clinical practices. Addressing limitations and pursuing future directions will help realize AI's full potential in dentistry.

Author contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Conflicts of interest

In compliance with the ICMJE uniform disclosure form, all authors declare the following:

Payment and services information: All authors declare that they received no financial support from any organization for the submitted work.

Financial relationships: All authors declare that they have no financial relationships with any organizations that might be interested in the submitted work, either currently or within the previous three years.

Other relationships: All authors declare that they have no other relationships or activities that could be perceived as influencing the submitted work.

Ethical clearance statement

This study was designed as a retrospective, de-identified, radiograph-based computational analysis using data retrieved from the Saveetha Dental Information Archiving Software (Saveetha DIAS). The study did not involve direct human participation, clinical intervention, collection of biological samples, or access to personally identifiable patient information.

The study protocol was reviewed by the Institutional Human Ethics Committee (IHEC), Saveetha Institute of Medical and Technical Sciences, Chennai, India. In accordance with institutional and national guidelines governing retrospective, anonymized data-based research, the committee determined that formal ethical approval was not required and granted an ethical exemption.

The scientific review board clearance number for this study is SRB/SDC/FACULTY/25/PERIO/312, dated November 28, 2025.

Sources of funding

Payment/services info: All authors have declared that no financial support was received from any organization for the submitted work.

Financial relationships: All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Other relationships: All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Ramamurthy J. Evaluation of antimicrobial activity of nanoformulated grape seed oil against oral microbes: an in vitro study. *World J Dent*. 2024;15:44–47.
2. Jacob D, Tk L. Digital decision making in dentistry: analysis and prediction of periodontitis using a machine learning approach. *Int J Next-Gener Comput*. 2022 Oct 31;13(3).
3. Huang SC, Pareek A, Zamanian R, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health records: a case study in pulmonary embolism detection. *Sci Rep*. 2020;10(1), 22147. https://doi.org/10.1038/s41598-020-78888-w.
4. Da-Ano R, Andrade-Miranda G, Tankyevych O, Visvikis D, Conze PH, Rest CC Le. Automated PD-L1 status prediction in lung cancer with multi-modal PET/CT fusion. *Sci Rep*. 2024 Jul;14(1), 16720.
5. Yuan H, Wu Y, Dai M. Multi-modal feature fusion-based multi-branch classification network for pulmonary nodule malignancy suspicion diagnosis. *J Digit Imag*. 2023 Apr;36(2):617–626.
6. Zhang Y, He X, Chan YH, Teng Q, Rajapakse JC. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. *Comput Biol Med*. 2023 Sep;164, 107328.
7. Gong B, Yan C, Bai J, Zou C, Gao Y. Hamming embedding sensitivity guided fusion network for 3D shape representation. *IEEE Trans Image Process*. 2020 Aug.
8. Ryu J, Lee DM, Jung YH, et al. Automated detection of periodontal bone loss using deep learning and panoramic radiographs: a convolutional neural network approach. *Appl Sci*. 2023;13(9). Available from: https://www.mdpi.com/2076-3417/13/9/5261.
9. Sabharwal A, Kavthekar N, Miecznikowski J, Glogauer M, Maddi A, Sarder P. Integrating image analysis and dental radiography for periodontal and peri-implant diagnosis. *Front Dental Med*. 2022;3. Available from: https://www.frontiersin.org/journals/dental-medicine/articles/10.3389/fdmed.2022.840963.
10. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med*. 2020;3(1):136. https://doi.org/10.1038/s41746-020-00341-z.
11. Qiu S, Chang GH, Panagia M, Gopal DM, Au R, Kolachalama VB. The fusion of deep learning models from MRI scans, the mini–mental state examination, and the logical memory test enhances the diagnosis of mild cognitive impairment. *Alzheimer's Dementia: Diagn Assess Dis Monitor*. 2018;10:737–749. Available from: https://www.sciencedirect.com/science/article/pii/S2352872918300654.
12. Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Sci Rep*. 2022 Oct;12(1), 17981.
13. Kumar A, Bhadauria HS, Singh A. Descriptive analysis of dental X-ray images using various practical methods: a review. *PeerJ Comput Sci*. 2021 Sep;7, e620. https://doi.org/10.7717/peerj-cs.620.
14. Jiang L, Chen D, Cao Z, Wu F, Zhu H, Zhu F. A two-stage deep learning architecture for radiographic staging of periodontal bone loss. *BMC Oral Health*. 2022 Apr;22(1):106.
15. Khare Y, Bagal V, Mathew M, Devi A, Priyakumar UD, Jawahar CV. MMBERT: multimodal BERT pretraining for improved medical VQA. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) [Internet]*; 2021:1033–1036. Available from: https://api.semanticscholar.org/CorpusID:233024902.
16. Anand D, Ramamurthy J, Kannan B, Priyadharsini JV, Arumugam P. N6-methyladenosine-mediated overexpression of TREM-1 is associated with periodontal disease. *Odontology*. 2024.
17. Kurt-Bayrakdar S, Bayrakdar İŞ, Yavuz MB, et al. Detection of periodontal bone loss patterns and furcation defects from panoramic radiographs using a deep learning algorithm: a retrospective study. *BMC Oral Health*. 2024 Jan;24(1):155.
18. Swarna RA, Hossain MM, Khatun MR, Rahman MM, Munir A. Concrete crack detection and segregation: a feature fusion, crack isolation, and explainable AI-Based approach. *J Imaging*. 2024 Aug;10(9).
19. Zakariah M, Al-Razgan M, Alfakih T. Dual vision Transformer-DSUNET with feature fusion for brain tumor segmentation. *Heliyon*. 2024 Sep;10(18), e37804.
20. Hasan M, Jani M, Rahman M. *Image and Text Feature-based Multimodal Learning for Multi- Label Classification of Radiology Images in Biomedical Literature*. 2024:679–686.
21. Swarna Meenakshi P, Sankari M. Formulation and evaluation of a novel herbal trio gel containing flaxseed extract, carbopol, and carboxymethyl cellulose. *Bioinformation*. 2023;19(5):540–545.

List of abbreviations

Term Definition

IOPA: Intraoral Periapical Radiograph
MMBERT: Multimodal Bidirectional Encoder Representations from Transformers
BERT: Bidirectional Encoder Representations from Transformers
IoU: Intersection over Union (segmentation accuracy metric)
Dice Coefficient: Statistical measure of similarity between predicted and actual segmentation
MAE: Mean Absolute Error
CLAHE: Contrast Limited Adaptive Histogram Equalization (used to enhance image contrast)
AP: Average Precision
FP16: 16-bit Floating Point Precision (used to speed up model training)
ResNet50: A deep convolutional neural network with 50 layers
UNet: A convolutional neural network architecture is commonly used for image segmentation