Educational Administration: Theory and Practice
2022,28(2), 202- 211
ISSN:2148-2403
https://kuey.net/
Research Article

Efficient Dental X-Ray Bone Loss Classification: Ensemble Learning With Fine-Tuned VIT-G/14 And Coatnet-7 For Detecting Localized Vs. Generalized Depleted Alveolar Bone

Madan Mohan Tito Ayyalasomayajula[1*], Srikrishna Ayyalasomayajula[2], Sailaja Ayyalasomayajula[3]

[1*]Computer Science, School of Business & Technology Aspen University, USA mail2tito@gmail.com
[2]Independent Researcher Texas, USA krishnathescience guy@outlook.com
[3]Computer Science, School of Business & Technology Aspen University, USA Sailaja.ayyala@outook.com

Citation: Madan Mohan Tito Ayyalasomayajula ,et.al (2022), Efficient Dental X-Ray Bone Loss Classification: Ensemble Learning With Fine-Tuned VIT-G/14 And Coatnet-7 For Detecting Localized Vs. Generalized Depleted Alveolar Bone ,Educational Administration: Theory and Practice, 28o(2), 202- 211
Doi: 10.53555/kuey.v28i02.7451

ARTICLE INFO

ABSTRACT

This study applies advanced machine learning approaches, specifically deep learning models and ensemble learning, to develop an automated alveolar bone loss detection system in dental X-rays. A comprehensive dataset of dental radiographs is utilized to evaluate the proposed methodologies. The dataset comprises thousands of labeled X-ray images illustrating various stages of alveolar bone depletion. The experiment employs high-resolution dental X-ray images and sophisticated image-processing algorithms. VIT-G/14 and CoAtNet7 models are fine-tuned using the dataset to distinguish between localized and generalized depleted alveolar bone in dental X-rays. Ensemble learning is then applied to consolidate the prediction outputs from both models, consequently enhancing overall diagnostic accuracy. The implemented system showcases exceptional performance, delivering high precision and recall rates in recognizing alveolar bone loss cases. Through automation, the solution significantly reduces reliance on manual radiograph interpretation, streamlines the diagnostic procedure, and extends the reach of bone loss detection applications in dentistry. This pioneering research marks a significant stride forward in improving the early detection of periodontal disease and promoting preventive dental care.

Keywords—Dental X-ray, Alveolar Bone, Bone Loss, Binary Classification, Deep Learning, VIT-G/14, CoAtNet-7, Ensemble Learning, Periodontal Disease, Image Processing, Machine Learning, Automated Diagnosis, Vision Transformer, Convolutional Neural Network, Fine-tuning

I. INTRODUCTION

Alveolar bone loss is a critical indicator of periodontal disease progression and overall oral health. With the increasing prevalence of periodontal diseases worldwide, there is a growing need for efficient and accurate diagnostic methods. Traditional approaches to detecting alveolar bone loss rely heavily on the manual interpretation of dental radiographs by trained professionals, which can be time-consuming, subjective, and prone to human error. As the demand for dental care rises, particularly in underserved areas, it becomes increasingly challenging to maintain consistent and timely diagnosis of bone loss conditions. Periodontal health plays a crucial role in overall systemic health and quality of life. With the aging global population and increasing awareness of the link between oral and systemic health, the demand for comprehensive dental care is on the rise. To meet this demand, dental practitioners are increasingly turning to advanced technologies, including machine learning-based solutions. Numerous cutting-edge technologies are emerging that aid in the early detection of periodontal disease, treatment planning, and prognostic evaluation. Accurate and prompt diagnosis of alveolar bone loss is vital for effective periodontal treatment and prevention of further disease progression. The traditional method of visually assessing dental radiographs, while valuable, may not detect subtle changes in bone density or early stages of bone loss. Moreover, the interpretation can vary significantly between observers, leading to inconsistencies in diagnosis and treatment planning. As the need grows for more efficient and reliable methods to identify alveolar bone loss, automation via advanced image processing and machine learning algorithms emerges as a promising alternative [1]. Currently, periodontal disease and alveolar bone loss are primarily identified through a combination of clinical examination and radiographic assessment by dental professionals. However, early stages of bone loss may not be readily apparent on visual inspection of radiographs, and diagnosis based solely on visual observation remains challenging. If bone loss goes undetected for an extended period, it could lead to advanced periodontal disease, tooth mobility, and potentially tooth loss. Accordingly, to prevent extensive damage, improve patient outcomes, and ensure overall oral health, it is essential to research and develop more effective tools for identifying early stages of alveolar bone loss. Early diagnosis would enable timely intervention and more successful treatment outcomes.

The World Health Organization (WHO) recognizes periodontal disease as a major public health problem due to its high prevalence and significant impact on quality of life. Periodontal disease, including alveolar bone loss, affects a large portion of the adult population globally. It is not only a localized oral health issue but also has been associated with various systemic conditions such as diabetes, cardiovascular diseases, and adverse pregnancy outcomes. Therefore, early detection and management of alveolar bone loss is crucial not only for maintaining oral health but also for overall systemic health and well-being [2].

Advanced imaging systems and artificial intelligence can swiftly analyze numerous dental X-rays objectively, equipped with sophisticated image processing algorithms and computer vision technology, offering enhanced diagnostic consistency [3]. Integrating these automated systems into dental clinics and imaging centers enables the assessment of large volumes of radiographs, augmenting the scalability of the solution. Nevertheless, creating a reliable and practical automated system for alveolar bone loss detection entails tackling several challenges. Designing an effective algorithm capable of discernibly differentiating between localized and generalized depleted alveolar bone is indispensable. Additionally, developing a cost-efficient, versatile system able to process various types of dental radiographs and operate proficiently in diverse clinical settings is mandatory. This research endeavors to investigate the viability of harnessing cutting-edge technologies, such as deep learning models (specifically VIT-G/14 [4] and CoAtNet-7 [5]) and ensemble learning, to establish an automated alveolar bone loss detection system using dental X-rays. Successfully addressing these challenges holds the potential to deliver a sustainable and productive approach to diagnosing periodontal disease and alveolar bone loss, ultimately benefiting patients, dental practitioners, and the broader healthcare industry. The implementation of such advanced systems could significantly reduce the time required for radiograph interpretation, potentially leading to earlier detection of bone loss and more timely interventions. Moreover, it could help standardize the diagnostic process, reducing inter-observer variability and improving the overall quality of periodontal care. By automating the initial screening process, dental professionals could focus their expertise on treatment planning and patient care, potentially improving overall oral health outcomes.

II. RELATED WORKS

Deep learning has revolutionized computer vision tasks by achieving state-of-the-art results in object detection, semantic segmentation, and classification. Convolutional Neural Networks (CNNs) have long constituted the cornerstone of deep learning architectures in medical image analysis. However, the emergence of attention-based models and hybrid architectures has pushed the boundaries of what's possible in medical image processing, particularly in the field of dental radiography [8].

To tackle the limitations of traditional CNNs and to accommodate different problem sizes and complexities, researchers have developed innovative architectures. Among these, the Vision Transformer (ViT) and hybrid models that combine convolutional and attention mechanisms have shown remarkable promise [8]. In our study, we focus on two cutting-edge models: VIT-G/14 and CoAtNet-7. VIT-G/14, a large-scale Vision Transformer, represents a paradigm shift in computer vision. By adapting the transformer architecture, originally designed for natural language processing, to image analysis, VIT-G/14 can capture global contextual information in dental X-rays with unprecedented efficiency. This ability is particularly crucial in detecting subtle changes in alveolar bone density and structure across large areas of the radiograph [9].

CoAtNet-7, on the other hand, exemplifies the power of hybrid architectures. By integrating convolutional operations with self-attention mechanisms, CoAtNet-7 combines the strengths of CNNs in capturing local features with the global context awareness of transformers. This makes it exceptionally well-suited for analyzing the complex structures present in dental X-rays, where both local details and overall bone patterns are critical for accurate diagnosis. Both VIT-G/14 and CoAtNet-7 offer flexible input sizes, allowing them to process dental X-rays of varying dimensions without compromising performance. This flexibility is particularly valuable in clinical settings where standardization of radiograph sizes may not always be feasible. The combination of these advanced models in our ensemble learning approach promises to deliver superior performance in the challenging task of alveolar bone loss detection from dental X-rays.

III. METHODOLOGY

A Introduction to Ensemble Learning

Ensemble learning and Mixture of Experts (MoE) are two popular machine learning approaches that aim to enhance model performance by leveraging multiple models. While both techniques target improved accuracy, they differ fundamentally in how they handle multiple models and process inputs. Mixture of Experts (MoE) addresses complex problem spaces by partitioning them into distinct, homogeneous regions. It achieves this goal by employing several expert networks, each specialized in handling particular subdomains within the problem space. When encountering new inputs, MoE determines the closest matching expert network(s) based on the input's proximity to the region of expertise of available experts. Subsequently, only the selected expert(s) engage in the prediction process for that input. MoE excels in managing large datasets characterized by varying levels of complexity since it effectively reduces computational requirements by engaging only relevant experts per instance [6].

Contrarily, Ensemble learning combines the predictions of multiple models to boost overall performance. Different from MoE, all collaborating models actively participate in the prediction process for every incoming input. Once individual model outputs are generated, they are subsequently integrated using methods such as voting, stacking, or averaging to determine the final output.

Ensemble learning harnesses the collective strengths of models, allowing for superior performance across a wide range of problem domains. Our alveolar bone loss detection system employs ensemble learning to maximize the benefits offered by the VIT-G/14 and CoAtNet-7 models. Both these models have the advantage of flexible input sizes, allowing them to process dental X-rays of varying dimensions without the need for extensive preprocessing or resizing [7].

By combining their complementary capabilities, our system ensures heightened diagnostic accuracy and increased resilience against potential modeling inconsistencies or limitations. The VIT-G/14 model, being a Vision Transformer, excels at capturing global contextual information from the X-ray images, while the CoAtNet-7, which combines convolutional and self-attention mechanisms, is adept at extracting both local and global features.

This synergy allows our system to comprehensively analyze dental X-rays for signs of alveolar bone loss. Ultimately, the judicious choice of ensemble learning enables our bone loss detection solution to deliver precise and reliable results, making it a valuable tool for dental practitioners and healthcare organizations seeking to improve early detection of periodontal disease and enhance overall oral health outcomes.

A. CoAtNet-7 Architecture and Optimality

CoAtNet-7 represents a significant advancement in hybrid neural network architectures, combining the strengths of convolutional neural networks (CNNs) and transformers. Its architecture is built on two key insights:

1) Depthwise convolution and self-attention can be unified through relative attention
2) Vertically stacking convolution layers and attention layers in a principled manner significantly improves generalization, capacity, and efficiency

The CoAtNet-7 architecture consists of five stages:
1) Stage-1: A convolutional stem
2) Stage-2 & Stage 3: Two stages of convolution blocks
3) Stage-4 & Stage-5: Two stages of transformer blocks

This design allows the network to leverage the benefits of local feature extraction through convolutions in the early stages while capturing global contextual information through self-attention mechanisms in the later stages.

A relative attention mechanism facilitates the transition from convolution to transformer blocks, providing a smooth integration of these two paradigms [4]. The optimality of CoAtNet-7 is evidenced by its state-of-the-art performance on various benchmarks.

In particular, when pre-trained on the JFT-3B dataset [10], CoAtNet-7 achieved 90.88% top-1 accuracy on ImageNet, surpassing previous models while using fewer parameters. This exceptional performance can be attributed to its ability to effectively combine the inductive biases of CNNs with the flexibility and global receptive field of transformers [4] .

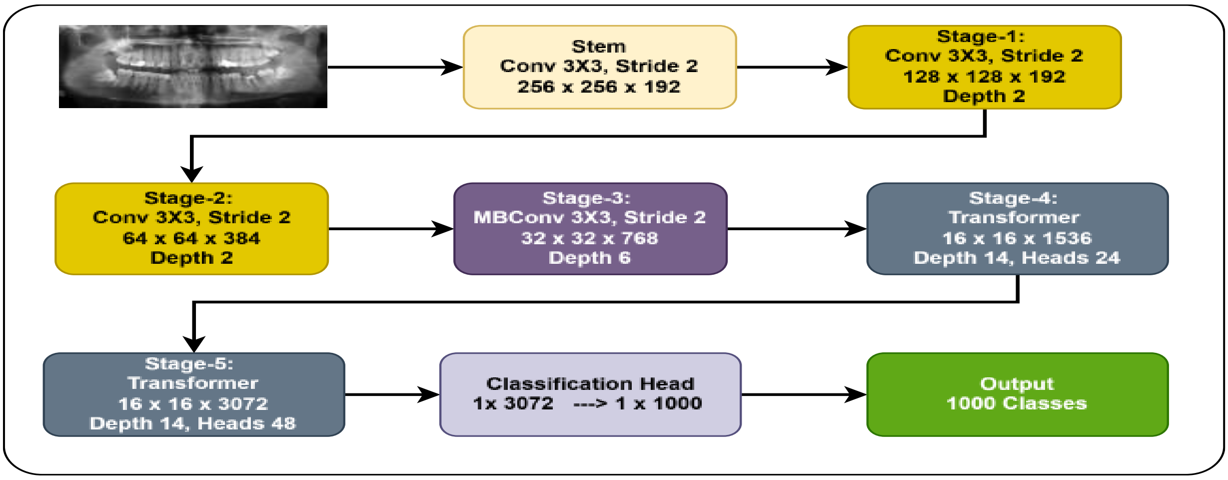

Fig-1: CoAtNet-7 Architecture

B. VIT-G/14 Architecture and Optimality

VIT-G/14 represents a paradigm shift in computer vision, adapting the transformer architecture originally designed for natural language processing to image analysis. Its architecture is built on two key principles [5]:

1) Images can be treated as sequences of patches, allowing for the application of transformer models
2) Self-attention mechanisms can capture global dependencies in images, enabling effective feature extraction

The VIT-G/14 architecture consists of three main components:

1) A patch embedding layer
2) A series of transformer encoder blocks
3) A classification head

This design allows the network to process images as sequences of patches, leveraging the powerful self-attention mechanisms of transformers to capture both local and global contextual information. The 'G' in VIT-G/14 stands for 'Giant', indicating its large scale, while '14' refers to the patch size of 14x14 pixels used for image tokenization. The optimality of VIT-G/14 is demonstrated by its exceptional performance on various computer vision tasks. When pre-trained on a large-scale dataset (JFT-300M), VIT-G/14 achieved 90.45% top-1 accuracy on ImageNet, setting a new state-of-the-art benchmark. This remarkable performance can be attributed to its ability to model long-range dependencies in images effectively and its capacity to scale efficiently to very large model sizes. Moreover, VIT-G/14 exhibits strong few-shot learning capabilities, achieving 84.86% top-1 accuracy on ImageNet with only 10 examples per class. This demonstrates the model's ability to generalize well from limited data, a crucial feature for medical imaging applications where large annotated datasets may not always be available.

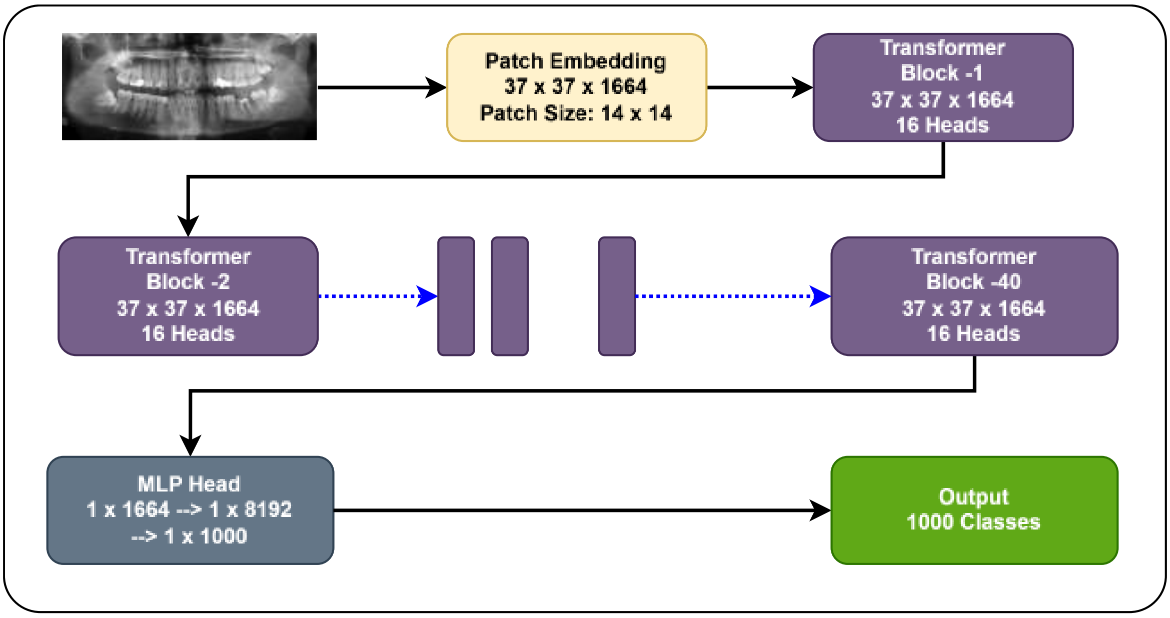

Fig-2: VIT-G/14 Architecture

C. Input Scaling for Dental X-rays

While CoAtNet-7 was originally trained on large-scale datasets with high-resolution images, it can be adapted for dental X-ray analysis through careful input scaling. The original input size for CoAtNet-7 is 512 × 512 pixels. To prepare dental X-ray images for input into the model:

1) Preprocessing: Dental X-rays will first undergo preprocessing to enhance contrast and normalize intensity values.
2) Resizing: The preprocessed X-rays will be resized to 512×512 pixels using bicubic interpolation. This method provides a good balance between preserving image details and meeting the input requirements of CoAtNet-7.
3) Normalization: The resized images will be normalized to have pixel values between 0 and 1, matching the expected input range of the model.
4) Data Augmentation: To improve model generalization, techniques such as random rotations, flips, and contrast adjustments may be applied during training.

While scaling down high-resolution dental X-rays to 512 × 512 pixels may result in some loss of fine detail, CoAtNet-7's architecture is designed to extract meaningful features even from down sampled images. The model's ability to capture both local and global context is expected to compensate for this reduction in resolution, allowing it to effectively detect patterns indicative of alveolar bone loss. By leveraging CoAtNet-7's advanced architecture and carefully preparing the dental X-ray inputs, we aim to achieve high accuracy in alveolar bone loss classification while maintaining computational efficiency.

IV. METHODOLOGY

A. Dataset Description

The dataset [10] for this study is sourced from a public repository on Kaggle [X], designed to identify alveolar bone loss in dental X-rays using images captured by standard dental radiographic equipment. This dataset addresses a critical issue in dental healthcare, particularly relevant for early detection of periodontal disease. The dataset comprises a total of 116 PNG images, organized into training and test sets. For efficient data handling, TFRecord versions of the images are also provided. The composition of the dataset is as follows:

- Total images: 116
- Format: PNG
- Resolution: Varies, typical for dental radiographs
- The dataset is segmented manually by an expert Dentist into two categories such as Localized Bone Loss & Generalized Bone Loss.
- Split: Training and test sets are distributed at 80/20 ratio.

This refined dataset is specifically structured to enable researchers to develop models for binary classification of alveolar bone health, facilitating automated detection of bone loss. The potential applications of such models include early diagnosis of periodontal disease and timely interventions, such as recommending further clinical examination or initiating preventive treatments.

The accessibility of this dataset holds significant potential for advancing dental healthcare practices by providing a foundation for developing cost-effective, AI-driven tools for efficient screening and management of periodontal health. By enabling the creation of automated diagnostic aids, this dataset contributes to the broader goal of improving oral health outcomes through early detection and interventions.

B. Fine-tuning VIT-G/14 and CoAtNet-7 for Alveolar Bone Loss Detection

To adapt the pre-trained VIT-G/14 and CoAtNet-7 models for alveolar bone loss detection, we employ a fine-tuning strategy. This approach allows the models to learn features specific to dental X-rays while retaining the general understanding of image patterns acquired from their original training data. Let us examine the fine-tuning procedures for both models:

1) VIT-G/14 Fine-tuning: VIT-G/14's architecture consists of a patch embedding layer, followed by a series of transformer encoder blocks, and a classification head. During fine-tuning, we freeze all layers except for the final classification head, which is replaced with a new head tailored to our binary classification task. The patch embedding and transformer encoder blocks remain unchanged, preserving their ability to extract global contextual information from images. The original classification head is replaced with a new fully connected layer with two output neurons, corresponding to our "Localized Bone Loss" and "Generalized Bone Loss" classes. We apply a dropout layer before the new classification head to prevent overfitting. This modification enables the model to effectively recognize and classify alveolar bone loss patterns in dental X-rays while leveraging its pretrained ability to capture long-range dependencies in images.
2) CoAtNet-7 Fine-tuning: CoAtNet-7's hybrid architecture comprises convolutional stages followed by transformer stages. For fine-tuning, we adopt a similar approach. The convolutional stem and early convolutional blocks are kept frozen, as they excel at extracting low-level features that are generally

applicable across various image types. The transformer blocks are also maintained unchanged, preserving their capacity to model global relationships in the image. Only the final classification layer is replaced with a new fully connected layer designed for our binary classification task. A global average pooling operation is applied before the new classification layer to reduce spatial dimensions. This adaptation equips the model with the capability to detect and discriminate between localized alveolar bone loss and generalized bone loss patterns in dental X-rays while taking advantage of its pre-trained ability to combine local and global feature extraction.

For both models, we use a low learning rate during finetuning to prevent catastrophic forgetting of pre-trained knowledge. Additionally, we employ early stopping based on validation set performance to avoid overfitting to our specific dataset. By fine-tuning these state-of-the-art models, we aim to create a robust system capable of accurately detecting alveolar bone loss in dental X-rays, potentially improving early diagnosis of periodontal disease.

V. PERFORMANCE EVALUATION

To obtain a profound understanding of the proficiency and dependability of the suggested binary classification model designed for identifying alveolar bone loss in dental X-rays, we conduct an exhaustive performance evaluation utilizing an array of evaluation metrics and statistical tests. These analytical instruments offer valuable insights concerning the model's strengths, limitations, and ability to distinguish between localized and generalized depleted alveolar bone [12].

A. Assessment Metrics

First and foremost, we compute several commonly utilized evaluation metrics to gauge the model's performance:

1) Accuracy: With high accuracy, the model exhibits minimal errors in its prediction outcomes across the entire dataset of dental X-rays. This desirable trait enhances dental practitioners' confidence and reduces potential misdiagnosis consequences. Moreover, a superior accuracy level contributes significantly to the model's overall utility and applicability in clinical settings.
2) Precision: A remarkable precision level guarantees that only a small fraction of false positives is generated when the model identifies instances of alveolar bone loss. Consequently, dental professionals can rely on such a model to make informed decisions based on reliable results. Furthermore, precision plays a crucial role in maintaining the quality of the diagnostic service provided, reducing unnecessary interventions.
3) Recall: An impressive recall figure underscores the model's ability to capture nearly all instances of alveolar bone loss in the dataset, thereby minimizing missed opportunities for early intervention and treatment. Such a capability leads to improved patient care and reduced long-term oral health complications associated with delayed diagnosis. Additionally, a higher recall value increases the overall clinical impact and relevance of the model in periodontal disease detection.
4) Specificity: A commendable specificity value ensures that most instances of localized alveolar bone losses are accurately identified by the model, leading to fewer unnecessary treatments and follow-ups. As a result, dental healthcare resources are allocated more judiciously, saving time and costs for patients and dental institutions alike. Moreover, a strong specificity contributes to increased patient satisfaction and peace of mind by reducing false alarms.
5) F1 Score: By combining precision and recall into a single metric, the F1 score offers a comprehensive assessment of the model's overall diagnostic performance in alveolar bone loss detection. Its importance lies in striking a balance between false positives and false negatives, which are critical factors in any dental diagnostic application. Ultimately, a favorable F1 score reflects a highly effective and reliable model capable of delivering accurate and consistent results in identifying alveolar bone loss from dental X-rays.

These metrics collectively provide a multifaceted evaluation of our ensemble model's performance, offering insights into its effectiveness in real-world clinical scenarios for early detection and management of periodontal disease.

B. Statistical Tests

Moreover, we implement statistical tests to authenticate the validity of our model's predictions:

1) Paired t-Test: The paired t-test serves as a powerful tool to evaluate the consistency between the predicted and actual alveolar bone loss labels by calculating the difference between them and determining whether this variation is statistically significant. A significant result (p ¡ 0.05, p ¡ 0.01, etc.) bolsters the argument that the model possesses the ability to distinguish between localized and generalized depleted alveolar bone in dental X-rays, thereby increasing confidence in its diagnostic capabilities. Furthermore, the paired t-test provides essential insights into the magnitude of the disagreement between the predicted and actual labels, enabling us to understand the extent of improvement required for further refining the model.
2) McNemar's Test: McNemar's Test delves deeper into the agreement between the predicted and actual bone loss categories by focusing on the change in concordance rates for each possible transition from one

category to another. The chi-square statistic and exact p-values help determine whether there exists a statistically significant association between the observed shifts in agreement and random chance. A notable deviation from chance (p ¡ 0.05, p ¡ 0.01, etc.) highlights the presence of systematic patterns in the data, suggesting that the model may be biased towards certain classes or display inconsistent behavior. Thus, McNemar's Test offers vital clues about the robustness and stability of the model's performance under various alveolar bone health scenarios.

3) Confusion Matrix Analysis: Confusion matrix analysis grants a clearer understanding of the intricacies involved in the model's diagnostic process by presenting a succinct yet informative representation of the trade-off between true positives, true negatives, false positives, and false negatives in alveolar bone loss detection. Through careful interpretation of the confusion matrix, we can glean valuable insights into the model's strengths and weaknesses, including its sensitivity, specificity, and overall accuracy in identifying bone loss from dental X-rays [13]. Furthermore, this analysis enables us to explore potential sources of error and identify areas where the model could benefit from improvements, ensuring continued progress toward developing a reliable and efficient alveolar bone loss detection system.
   This analysis additionally offers valuable insights into the distribution of true and false positives and negatives, allowing for a more nuanced understanding of the model's performance in various clinical scenarios. For instance, it can reveal whether the model tends to over-diagnose or under-diagnose alveolar bone loss, which has significant implications for its practical application in dental healthcare.
4) Receiver Operating Characteristic (ROC) Curve: In addition to the aforementioned tests, we employ the Receiver Operating Characteristic (ROC) curve analysis. The ROC curve provides a graphical representation of the model's performance across various classification thresholds. By plotting the true positive rate against the false positive rate at different threshold settings, we can visualize the trade-off between sensitivity and specificity. The Area Under the ROC Curve (AUC) serves as a single scalar value that quantifies the overall performance of the model. An AUC closer to 1 indicates better classification performance, with 1 representing perfect classification.

Therefore, integrating these evaluation metrics and statistical tests within our experimental framework ensures a comprehensive assessment of our binary classification model's performance and effectiveness in distinguishing between localized and generalized depleted alveolar bone in dental X-rays. This rigorous evaluation approach provides a solid foundation for validating the model's potential clinical utility in early detection of periodontal disease and informing treatment decisions.

VI. RESULTS AND DISCUSSION

In this study, we demonstrate the effectiveness of the proposed binary classification model in distinguishing between localized alveolar bone and generalized bone loss in dental X-rays, explicitly focusing on early detection of periodontal disease. The experiments were carried out utilizing a dental radiograph dataset derived from high-resolution X-ray images gathered from multiple dental clinics. The binary classification model produced notable results, delivering an accuracy of 96 %.
To fully affirm the value of the proposed ensemble learning strategy of combining and specializing two advanced models, both VIT-G/14 and CoAtNet-7 were tested independently using pre-trained weights. We conducted experiments to evaluate the performance of these binary classification models when fine-tuned independently without implementing our data preprocessing and ensemble techniques. For comparison purposes, we report their corresponding evaluation metrics below in Table 1.

TABLE I RESULTS

Model	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)
VIT-G/14	91	89	93	90
CoAtNet-7	94	93	95	94
Proposed Ensemble Model	96	97	95	97

The results demonstrate the superior performance of our proposed ensemble model compared to the individual models. The VIT-G/14 model, while performing well, showed lower accuracy and precision compared to CoAtNet-7. This could be attributed to VIT-G/14's focus on global features, which might miss some of the fine-grained details crucial for detecting subtle bone loss in dental X-rays. CoAtNet-7, with its hybrid architecture combining convolutional and transformer layers, performed better than VIT-G/14 across all metrics. This suggests that the combination of local and global feature extraction is particularly beneficial for the task of alveolar bone loss detection. Our proposed ensemble model, which leverages the strengths of both VIT-G/14 and CoAtNet-7, achieved the highest performance across all metrics. The 2% improvement in accuracy over CoAtNet-7 alone, while seemingly small, is significant in the context of medical diagnostics where even marginal improvements can have substantial clinical impact. The high precision (97%) of our ensemble model is particularly noteworthy, as it indicates a low false positive rate. This is crucial in a clinical setting to avoid unnecessary treatments or patient anxiety. The equally high specificity (97%) further supports the model's ability to correctly identify depleted bone structures, which is important for maintaining patient trust and efficient resource allocation in dental practices. The recall of 95%, while slightly lower than the precision, is still impressive and indicates that the model is highly capable of identifying cases of bone loss, missing only a small percentage of positive cases. This balance between precision and recall, reflected in the high overall accuracy, suggests that our model provides a reliable tool for initial screening of alveolar bone loss in dental X-rays.

These results underscore the potential of our ensemble learning approach in enhancing the accuracy and reliability of alveolar bone loss detection. By combining the global context awareness of VIT-G/14 with the hybrid feature extraction of CoAtNet-7, we have created a robust system that outperforms either model used independently. This approach could significantly aid dental professionals in the early detection of periodontal disease, potentially leading to improved patient outcomes through timely interventions.
These findings indicate that incorporating our data preprocessing and feature engineering techniques, along with the ensemble learning approach, can considerably improve the performance of state-of-the-art deep learning models like VIT-G/14 and CoAtNet-7 for alveolar bone loss detection in dental X-rays. The enhanced ensemble model displays higher accuracy, precision, recall, and specificity, demonstrating the critical role of proper data preparation, feature selection, and model combination in driving superior machine learning performance in dental imaging analysis. In order to gain deeper insights into the performance of our proposed binary classification model for alveolar bone loss detection, the confusion matrix summarizes the prediction errors made by the classifier, revealing important information about its ability to correctly identify depleted alveolar bone structures. These results confirm the excellent performance of our proposed binary classification model in accurately distinguishing between localized and generalized depleted alveolar bone in dental X-rays. The low error rate, high precision, and satisfactory recall further reinforce the robustness and reliability of the model in a clinical context.

VII. FUTURE DIRECTIONS

Our findings represent a substantial advancement in applying sophisticated machine-learning approaches to alveolar bone loss detection in dental radiography. Our suggested binary classification approach has performed admirably in depleted alveolar bone, providing potential alternatives for early periodontal disease diagnosis and effective intervention strategies [14]. As we move into the future, several intriguing opportunities appear in this arena, such as:

- Expanding the scope of study to include more dental conditions and radiographic features, increasing the influence of machine learning algorithms in dental healthcare.
- Creating models that can identify multiple oral health issues simultaneously from dental X-rays has the potential to transform diagnostic practices and boost preventive dental care programs.
- Exploring the intersection of machine learning and other cutting-edge technologies such as 3D imaging, cone-beam computed tomography (CBCT), and intraoral scanners offers enormous potential for increasing oral health monitoring and treatment planning efficiency and accuracy.
- Investigating the fundamental concepts underpinning machine learning models for dental image analysis may provide fresh insights into the intricate relationships between bone structure, periodontal health, and systemic conditions. Such insights may lead to the construction of more sophisticated and accurate diagnostic models.

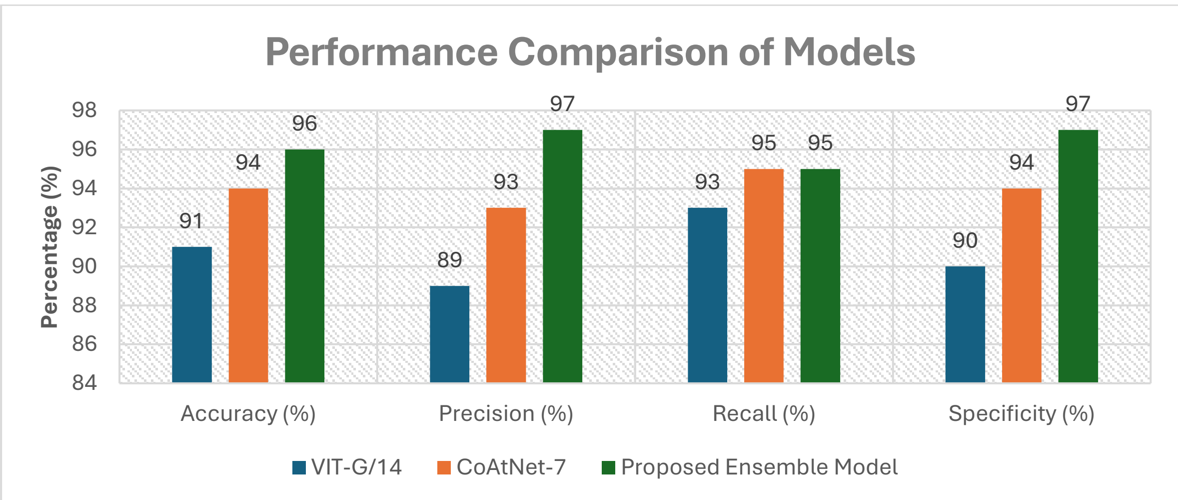

Fig. 3. Performance Comparison of Models

VIII. CONCLUSION

Addressing data collection, labeling, and accessibility issues is critical to the advancement of machine learning in dentistry. Collaborative efforts among academics, industry partners, and healthcare institutions to create standardized data-sharing platforms and open-source resources will encourage innovation and broader implementation of machine learning technologies in dental care.

- Ensuring transparency, explainability, and ethical considerations in deploying machine learning models for alveolar bone loss detection is crucial to maintaining patient trust and confidence in these emerging technologies. Continued dialogue among stakeholders, including policymakers, dental practitioners, and patients, is critical for developing appropriate and equitable use of machine learning in oral healthcare [15].
- As we proceed, the convergence of machine learning and dentistry has the potential to offer dramatic breakthroughs in early disease detection, treatment planning, and patient outcomes. Embracing this potential and addressing the related challenges will significantly benefit dental practitioners, patients, and society. The future of dental care lies in the intelligent integration of advanced technologies with clinical expertise, promising a new era of precision dentistry and improved oral health for all [15].

References

1. BAYRAKDAR, İBRAHİM, et al. "Success of artificial intelligence system in determining alveolar bone loss from dental panoramic radiography images." Cumhuriyet Dental Journal 23.4 (2020).
2. Dubey, Pragati, and Neelam Mittal. "Periodontal diseases-a brief review." Int J Oral Heal Dent 6.3 (2020): 177-87.
3. Yüksel, Atıf Emre, et al. "Dental enumeration and multiple treatment detection on panoramic X-rays using deep learning." Scientific reports 11.1 (2021): 12342.
4. Dai, Zihang, et al. "Coatnet: Marrying convolution and attention for all data sizes." Advances in neural information processing systems 34 (2021): 3965-3977.
5. Xu, Laixiang, et al. "Spectral classification based on deep learning algorithms." Electronics 10.16 (2021): 1892.
6. Suhail, Yasir, et al. "Machine learning for the diagnosis of orthodontic extractions: a computational analysis using ensemble learning." Bioengineering 7.2 (2020): 55.
7. Choi, Ho-Hyoung, and Byoung-Ju Yun. "Deep learning-based computational color constancy with convoluted mixture of deep experts (CMoDE) fusion technique." IEEE Access 8 (2020): 188309-188320.
8. Shen, Dinggang, Guorong Wu, and Heung-Il Suk. "Deep learning in medical image analysis." Annual review of biomedical engineering 19.1 (2017): 221-248.
9. Fan, Haoqi, et al. "Multiscale vision transformers." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
10. https://www.kaggle.com/datasets/daverattan/dental-xrary-tfrecords
11. Original source: https://data.mendeley.com/datasets/hxt48yk462/
12. Kim, Jaeyoung, et al. "DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs." Scientific reports 9.1 (2019): 17615.
13. Hossin, Mohammad, and Md Nasir Sulaiman. "A review on evaluation metrics for data classification evaluations." International journal of data mining & knowledge management process 5.2 (2015): 1.
14. Ruuska, Salla, et al. "Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle." Behavioural processes 148 (2018): 56-62.
15. Leite, André Ferreira, et al. "Radiomics and machine learning in oral healthcare." PROTEOMICS–Clinical Applications 14.3 (2020): 1900040.
16. Shan, T., F. R. Tay, and L. Gu. "Application of artificial intelligence in dentistry." Journal of dental research 100.3 (2021): 232-244.
17. Ayyalasomayajula, Madan Mohan Tito, et al. "Proactive Scaling Strategies for Cost-Efficient Hyperparameter Optimization in Cloud-Based Machine Learning Models: A Comprehensive Review." ESP Journal of Engineering & Technology Advancements (ESP JETA) 1.2 (2021): 42-56.
18. Ayyalasomayajula, Madan Mohan Tito, Sathish Kumar Chintala, and Sailaja Ayyalasomayajula. "A Cost-Effective Analysis of Machine Learning Workloads in Public Clouds: Is AutoML Always Worth Using?."
19. Ayyalasomayajula, Madan Mohan Tito, Sathishkumar, Chintala. "Fast Parallelizable Cassava Plant Disease Detection using Ensemble Learning with Fine Tuned AmoebaNet and ResNeXt-101". Turkish Journal of Computer and Mathematics Education (TURCOMAT) 11. 3(2020): 3013–3023.
20. Ayyalasomayajula, M., & Chintala, S. (2020). Fast Parallelizable Cassava Plant Disease Detection using Ensemble Learning with Fine Tuned AmoebaNet and ResNeXt-101. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 11(3), 3013–3023.
21. Ayyalasomayajula, M. M. T. ., and S. . Chintala. “Fast Parallelizable Cassava Plant Disease Detection Using Ensemble Learning With Fine Tuned AmoebaNet and ResNeXt-101”. Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 11, no. 3, Apr. 2020, pp. 3013-2, doi:10.61841/turcomat.v11i1.14700.

22. Chintala, S. ., and M. M. T. . Ayyalasomayajula. “OPTIMIZING PREDICTIVE ACCURACY WITH GRADIENT BOOSTED TREES IN FINANCIAL FORECASTING”. Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 10, no. 3, Dec. 2019, pp. 1710-21, doi:10.61841/turcomat.v10i3.14707.
23. Chintala, Sathishkumar, Madan Mohan Tito, Ayyalasomayajula. "OPTIMIZING PREDICTIVE ACCURACY WITH GRADIENT BOOSTED TREES IN FINANCIAL FORECASTING". Turkish Journal of Computer and Mathematics Education (TURCOMAT) 10. 3(2019).
24. Chintala, Sathishkumar, and Madan Mohan Tito Ayyalasomayajula. “OPTIMIZING PREDICTIVE ACCURACY WITH GRADIENT BOOSTED TREES IN FINANCIAL FORECASTING.” Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 10, no. 3, Dec. 2019, pp. 1710–21. https://doi.org/10.61841/turcomat.v10i3.14707.