 DOI: 10.21608/adjalexu.2025.372617.1614

DEEP LEARNING-BASED ASSESSMENT OF JAW-BONE DENSITY FOR DENTAL IMPLANT PLANNING: A DIAGNOSTIC ACCURACY STUDY

Sara I. Madian[1*] MSc, Hassan Abouelkheir[2] PhD, Marwan Torki[3] PhD, Noha M. Elkersh[4] PhD.

ABSTRACT

INTRODUCTION: Dental implants have transformed restorative dentistry, offering a reliable solution for tooth replacement. Their success depends on primary implant stability, which is closely tied to bone density. Misch's classification system provides a precise method for assessing bone density. Cone Beam Computed Tomography (CBCT) has emerged as a preferred tool due to its lower radiation and cost-effectiveness. Recent advancements in deep learning, particularly Vision Transformers, show promise in analyzing CBCT images for bone density classification.

AIM: This research focused on designing and evaluating Vision Transformer (ViT) models to classify jawbone density from CBCT scans.

MATERIALS AND METHODS: A comprehensive dataset of 5,545 CBCT images, extracted from 500 scans, was organized into training, validation, and testing groups. Binary masks were utilized to isolate regions of interest, and the images were categorized into five density types following the Misch classification. Several ViT architectures were trained and assessed, with performance evaluated using key metrics, including accuracy, sensitivity, specificity, loss, and area under the curve (AUC).

RESULTS: The SwinV2 model delivered the best overall performance, achieving the highest accuracy (85.65%) and specificity (90.13%), along with a strong AUC (0.73) and the lowest loss (0.8905). The ViTamin model excelled in sensitivity, while the XciT model also performed well, showcasing its reliability. The integration of binary masks improves model outcomes, emphasizing their value in refining classification tasks.

CONCLUSIONS: The SwinV2 model proved to be the most effective for jawbone density classification. The use of binary masks significantly enhanced model accuracy.

KEYWORDS: Artificial Intelligence, Cone Beam Computed Tomography, Convolutional Neural Networks, Deep Learning, Density Classification.

RUNNING TITLE: Deep Learning for Jawbone Density Assessment.

1-Assistant Lecturer of Oral Medicine, Oral Periodontology, Oral Diagnosis, and Oral Radiology, Faculty of Dentistry, Alexandria University.
2-Professor of periodontology, oral medicine oral diagnosis, and oral radiology, Faculty of Dentistry, Alexandria University.
3-Professor of Computer and Systems Engineering, Faculty of Engineering, Alexandria University.
4-Lecturer of periodontology, oral medicine oral diagnosis, and oral radiology, Faculty of Dentistry, Alexandria University.

Corresponding author:
sara.mohamed.dent@alexu.edu.eg

INTRODUCTION

Dental implants have revolutionized restorative dentistry by providing a reliable solution for replacing missing teeth (1). A critical factor in their success is primary implant stability, which significantly influences osseointegration; the process of bone integration that ensures the implant's long-term stability and functionality (2).

Research has shown a strong link between primary implant stability and local bone density, with lower bone density often leading to higher failure rates, while higher density is associated with better outcomes (3). This relation allows surgeons to evaluate the potential success of the procedure before surgery and modify the treatment plan according to the patient's bone quality (4).

To better understand bone density, Misch introduced a classification system in 1988, dividing bone mineral density into four categories (D1-D4) based on the microscopic composition of compact and cancellous bone. The D1 type represents compact bone with minimal cancellous bone, while the D4 type consists of mostly cancellous bone with little to no cortical bone (5). Later, Misch expanded this classification by incorporating Computed Tomography (CT) scans and Hounsfield Units (HU) to provide a more precise assessment of bone density (6).

Computed tomography scans are frequently employed as a preoperative tool for evaluating bone quality and quantity before implant placement. Hounsfield Units (HU), derived from CT scans, measure bone density by calculating the de-

gree of X-ray attenuation per voxel, providing a detailed representation of the bone's density (7).
Based on these HU values, Misch categorized bone density into five groups: D1 bone, the densest, has values above 1250 HU; D2 ranges from 850 to 1250 HU; D3 falls between 350 and 850 HU; D4 spans 150 to 350 HU; and D5, the least dense, has values below 150 HU (6).

Cone Beam Computed Tomography (CBCT) has become the preferred imaging modality for dental applications because it offers lower radiation exposure and is more cost-effective than conventional CT scanners (8). In CBCT scans, the grayscale, or voxel value, reflects the level of X-ray attenuation. As a result, CBCT manufacturers often convert grayscale values into Hounsfield Units (HU) for standardization (8, 9).

Multiple studies have evaluated the precision of CBCT voxel values in measuring bone density (9-11). For example, a study by Parsa et al. (12) compared CBCT with multislice CT (MSCT) and micro-computed tomography (micro-CT), showing a strong agreement between CBCT and MSCT results. These findings indicate that CBCT is a reliable tool for evaluating bone density in potential implant areas, making it a practical alternative to more advanced imaging techniques.

Over the past few years, deep learning, a branch of artificial intelligence (AI), has become increasingly popular for analyzing radiographic images (13). Within the domain of deep learning, artificial neural networks (ANNs) have seen significant growth in use and recognition. ANNs consist of interconnected units known as neurons, which are structured into multiple layers. In medical and dental applications, convolutional neural networks (CNNs) and vision transformers (ViT), along with their variations, are among the most widely utilized types of ANNs (14).

Vision transformers represent a major advancement in the field of deep learning, demonstrating remarkable capabilities in processing both natural language and visual images. Although effective at transmitting and storing data, its most notable feature is its capacity to understand long-range relationships within information (15).

Vision transformers function by splitting an input image into smaller segments, which are treated as tokens, similar to how transformers process words in textual data. These image segments are converted into fixed-length vectors through linear embedding and paired with positional embeddings to maintain spatial details (16). This approach allows ViTs to process visual information with high precision and efficiency (17).

According to existing literature, no studies have explored the use of deep learning, particularly vision transformers, for classifying jawbone density using CBCT scans. To address this gap, the primary objective of this research was to design and implement a deep learning model trained on a comprehensive dataset of CBCT images. The model's accuracy was evaluated to confirm its reliability and effectiveness in classifying bone density, ensuring its potential for practical application in clinical settings.

The null hypothesis of this research was that there is no statistically significant difference between the developing vision transformer model and the manual method for bone density classification using CBCT images.

MATERIALS AND METHODS

Sample size estimation

Sample size was planned based on 95% confidence level to detect the accuracy of an artificial intelligence model in CBCT-based implant planning. Roongruangsilp and Khongkhunthian (18) reported that the panoramic accuracy of the original model used on 300 images is 60% (6/10) [95% confidence interval= 75.83, 84.80]. The required sample size was calculated to be 455 CBCT scans, increased to 500 to make-up for processing problems (19).

Software

MedCalc Statistical Software version 19.0.5 (MedCalc Software bvba, Ostend, Belgium; https://www.medcalc.org; 2019).

Dataset Collection

Following approval from the Research Ethics Committee of the Faculty of Dentistry at Alexandria University (IRB NO: 00010556-IORG0008839), a dataset comprising 500 CBCT scans was compiled. These scans were sourced from a private radiology center after ethical approval from the center's internal review board, ensuring compliance with data privacy and patient confidentiality. Given the retrospective and anonymized nature of the dataset, individual patient consent was waived in accordance with national ethical guidelines for non-interventional research. The scans were captured using the Green X CBCT machine (Green X Ct, Vatech, Hwaseong, Republic of Korea). The imaging process adhered to specific parameters: a voltage of 90kVp, current of 10mA, full 360° rotation, and an exposure time of 9 seconds. The field of view varied between 8*5 cm and 9*16 cm, with voxel sizes set at 120, 200, and 300 microns.

Each scan underwent a thorough evaluation to ensure compliance with the study's inclusion criteria. Eligible scans included patients aged 18 years or older, regardless of gender, who were free from systemic diseases and exhibited single or multiple edentulous spaces in either the maxilla or mandible, whether anterior or posterior. Scans were excluded if they displayed artifacts at the measurement sites or revealed pathological lesions.

Dataset Preparation

All CBCT images were initially saved in the Digital Imaging and Communications in Medicine (DI-

COM) format. These DICOM files were then imported into Blue Sky Plan® version 4.12.13 software (mdi Europa GmbH, Langenhagen, Germany) for analysis and to generate dental volume reconstruction (DVR). A panoramic view was created by drawing a panoramic curve in the axial view, starting from the right condyle, passing through the center of each tooth, and ending at the left condyle. Cross-sectional images were subsequently generated at the required positions.

In the reconstructed panoramic view, all edentulous areas were identified and marked. Bone density measurements were manually performed on the cross-sectional images using the software's measurement tools by two oral radiologists calibrated on the assessment method. To expand the dataset, a standardized measurement of six millimeters was taken for each edentulous space, with measurements conducted at the mid-center of the edentulous bone. The images were then exported in JPG format to ensure uniformity.

Subsequently, bone density values were categorized into five classes based on the classification system proposed by Misch (6) (Figure 1). The bone density class and its corresponding JPG image were recorded and prepared for use in training the model (Figure 2).

Model Training

The dataset consisted of 5,545 JPG images derived from CBCT scans, which were split into three subsets: 4,645 (from 350 scans) images for training, 300 (from 50 scans) for validation, and 600 (from 100 scans) for testing.

The training procedure involved several key steps. First, a binary mask was created for each input image to identify the specific region where bone density classification was required. Since bone density could vary within the same image, the mask served as a guide, directing the model to focus on the relevant area. This mask was integrated as a fourth channel alongside the original RGB image. To maintain uniformity, all images were resized to 336 × 336 pixels before processing.

After preparing the training data, multiple pre-trained models were utilized to predict bone density classifications within the masked regions. Each model was trained independently on a P100 GPU, using categorical cross-entropy loss as the objective function. The Adam optimizer was employed with a learning rate of 0.00003, and a batch size of 1 was used for training. The number of epochs ranged between 25 and 50, depending on the experiment. A variety of Vision Transformers models were applied to accomplish this task as follows:

*CaiT (Class-Attention in Image Transformers) (**20**):*

The CaiT model introduces significant advancements to transformer-based architectures, making them more efficient for image classification tasks. By incorporating two key innovations, LayerScale and class-attention layers, it reduces both the number of parameters and computational complexity. LayerScale enhances the training of deeper models by integrating a learnable diagonal matrix into residual blocks, which stabilizes optimization and improves performance as the model's depth increases. Meanwhile, the class-attention layers separate the attention mechanisms for image patches and class embeddings, enabling more precise and effective class representation. These features collectively enhance the model's efficiency and accuracy in handling image classification.

*XciT (Cross-covariance Image Transformer) (**21**):*

This model incorporates cross-covariance attention, a mechanism designed to enhance efficiency by focusing on feature channels instead of tokens. This approach significantly reduces computational complexity, enabling the processing of high-resolution medical images without compromising scalability. By shifting attention to channels rather than tokens, the model achieves greater efficiency while maintaining its ability to handle detailed and large-scale medical imaging data.

*Swin V2 (**22**):*

Swin Transformer V2 addresses limitations in the original Swin Transformer, particularly in terms of training stability and scalability. Key enhancements include the implementation of residual post-normalization, which stabilizes training by regulating activation levels throughout the network layers. Furthermore, the model incorporates a log-spaced continuous position bias, enabling greater adaptability to varying window sizes and improving its versatility for diverse tasks. The architectural design of the Swin V2 model is illustrated in (Figure 3).

*CrossViT (Cross-Attention Vision Transformer) (**23**):*

The architecture of this model features a dual-branch design, allowing it to analyze image patches at varying scales. This approach enables the simultaneous extraction of detailed local features and broader contextual information. By leveraging cross-attention between patches of different sizes, the model seamlessly merges local and global insights, enhancing its ability to classify complex patterns. Furthermore, the integration of linear token fusion optimizes computational efficiency, making it an ideal choice for medical imaging applications where resource conservation is critical.

*ViTamin (Vision Transformer for Vision-Language tasks) (**24**):*

ViTamin represents a cutting-edge hybrid model specifically engineered to handle large-scale image-text datasets with high efficiency. By integrating Mobile Convolution Blocks (MBConv) and Transformer Blocks, it effectively extracts high-resolution features for image classification tasks. The model demonstrates exceptional generalization

capabilities, even with limited data, making it particularly well-suited for clinical applications where annotated datasets are often scarce. Additionally, the incorporation of GeGLU (Gated Linear Units) enhances parameter efficiency, delivering a balance of accuracy and computational performance that is ideal for analyzing intricate dental imaging patterns.

Model Testing

Throughout model training, no instability or divergence was observed. Training and validation losses decreased smoothly and consistently, and early stopping was employed where appropriate. This ensured that the training process was robust and free from overfitting.

The models' ability to predict bone classes was evaluated by comparing their predictions on the test dataset against the actual class labels. Performance metrics, including sensitivity, specificity, accuracy, loss, and area under the curve (AUC), were calculated to assess their effectiveness. These metrics were computed for each individual class and then averaged across all classes to determine the overall performance of the models.

Sensitivity, calculated as TP/(TP+FN) ×100, measured the models' capability to correctly identify each bone class. Specificity, calculated as TN/(TN+FP) ×100, evaluated their ability to avoid incorrect classifications. The AUC, derived using a one-vs-rest multiclass classification approach, summarized the balance between sensitivity and specificity across different classification thresholds. Accuracy, calculated as (TP+TN)/(TP+TN+FP+FN) ×100, represented the overall proportion of correct predictions made by the models.

The evaluation process utilized true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) obtained from the confusion matrix. Data analysis was conducted using MedCalc Statistical Software version 19.0.5 (MedCalc Software bvba, Ostend, Belgium; https://www.medcalc.org; 2019).

Reliability Assessment

To ensure consistency in density classification, calibration was conducted for two examiners. Both inter-examiner and intra-examiner reliability were assessed, with the intraclass correlation coefficient (ICC) ranging from 0.833 to 0.998. This range indicates a high level of reliability, demonstrating excellent agreement between the examiners as well as consistent performance over time.

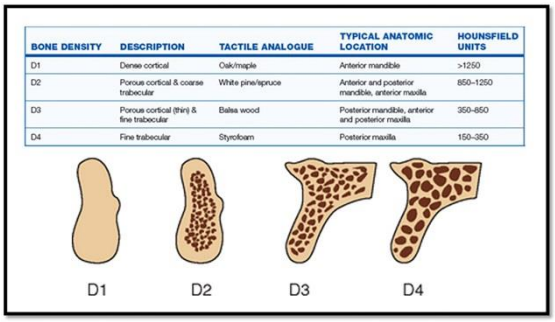

Figure 1: Misch Bone Density Classification.

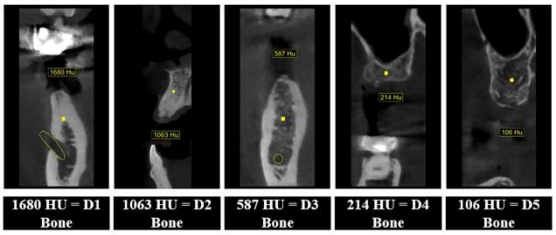

Figure 2: Composite Figure of Misch Density Classes on Cross-sectional CBCT Images.

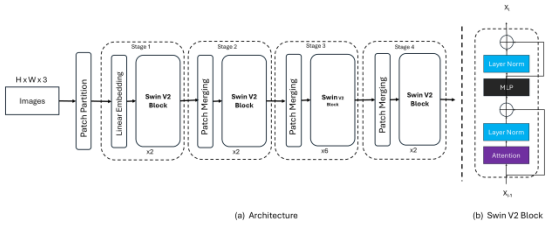

Figure 3: The Architectural Layers of the Swin V2 Model.

RESULTS

Table 1 provides a detailed comparison of the performance of various ViT models in classifying jawbone density using masked CBCT images. Each model was evaluated based on accuracy, sensitivity, specificity, AUC, and loss metrics to ensure a thorough assessment. Among the models, SwinV2 emerged as the top performer, achieving the highest accuracy and specificity, the second-best AUC, and the lowest loss, establishing it as the most effective for this task. The ViTamin model recorded the highest sensitivity and AUC, while the XciT model demonstrated strong performance with the second-best accuracy, specificity, and loss values.

In contrast, Table 2 presents the results when the binary mask-indicating the region for dental bone density calculation was excluded. The removal of the mask led to a decline in all evaluation metrics, highlighting its significant impact on model performance. Notably, the CrossViT model exhibited the best performance under these conditions, though its results were still lower compared to those achieved with the mask. These findings underscore the critical role of the binary mask in enhancing the models' effectiveness.

Table 3 shows a classification report and confusion matrix for the best-performing model (SwinV2 with the density mask input). Most misclassifications occur between adjacent classes (e.g., D2 misclassified as D3, D4 as D3 or D5), which is clinically understandable due to the gradual nature of bone density transitions. D1 bone, while having perfect precision, suffers from low recall, indicating that although the few predictions made for D1 are correct, the model is hesitant to classify samples as D1. D5 bone shows both high precision and recall, confirming it as the most confidently identified class.

The average inference time for each architecture was recorded using a standardized setup and averaged over multiple forward passes on the test dataset. The results are summarized as: CaiT model: 0.0100 Sec, XciT model: 0.0260 Sec, SwinV2 model: 0.0256 Sec, CrossViT model: 0.0125 Sec, and ViTamin model: 0.0086 Sec. These results demonstrate that ViTamin and CrossViT models are notably faster in terms of inference speed, making them suitable for real-time or resource-constrained clinical applications. However, SwinV2, despite slightly longer inference times, offers superior performance in terms of diagnostic accuracy and robustness.

Table 1: Performance Comparison of Different Vision Transformer Models for Jawbone Density Classification on Test Dataset.

Model	Accuracy	Sensitivity	Specificity	AUC	Loss
CaiT	83.04%	55.78%	88.39%	0.72	0.9459
XciT	85.02%	49.86%	89.39%	0.7	0.9223
Swin V2	**85.65%**	55.48%	**90.13%**	0.73	**0.8905**
CrossViT	84.83%	54.71%	89.26%	0.72	0.9285
ViTamin	84.25%	**58.26%**	89.22%	**0.74**	0.9299

Bold values indicate the best score, while underlined values indicate the second-best score in each column.

Table 2: Performance of different models where the input does not include the fourth channel representing the binary mask (the region where density is calculated).

Model	Accuracy	Sensitivity	Specificity	AUC
CaiT	75.94%	39.84%	84.96%	0.53
XciT	74.67%	36.67%	84.17%	0.51
Swin V2	74.35%	41.90%	83.97%	**0.55**
CrossViT	**77.27%**	**43.17%**	**85.79%**	0.52
ViTamin	74.67%	36.67%	84.17%	0.5

Bold values indicate the best score.

Table 3: Classification report and confusion matrix for the best-performing model (SwinV2 with the density mask input).

	Precision	Recall	F1-score	support
D1	1	0.166667	0.285714	30
D2	0.355932	0.456522	0.4	46
D3	0.651822	0.712389	0.680761	226
D4	0.457627	0.432	0.444444	125
D5	0.810945	0.802956	0.806931	203
Accuracy	0.64127	0.64127	0.64127	0.64127
Macro avg	0.655265	0.514107	0.52357	630
Weighted avg	0.65954	0.64127	0.635216	630

DISCUSSION

The integration of deep learning models and CBCT imaging for classifying jawbone density marks a major leap forward in dental diagnostics. This research investigated the effectiveness of various vision transformer architectures in performing this classification task. Among the models tested, SwinV2 stood out with superior performance metrics, including the highest accuracy and specificity, along with the lowest loss.

Jawbone quality plays a critical role in the success of dental implant treatments. Research has identified instances of cluster failures, which may be associated with inferior bone quality (25-27). The biomechanical condition of the alveolar bone influences key factors such as implant placement, number, abutment selection, and prosthesis design (28). As a result, adopting a quantitative scale, rather than relying on absolute values, could provide clinicians with a more efficient method for categorizing bone quality.

In this study, bone density quantification was based on the Misch classification system (6), which is currently the most widely adopted approach. Adhering to the ALADA (As Low As Diagnostically Accepted) principle (29) CBCT scans were utilized for bone density assessment instead of traditional CT scans. The effectiveness of CBCT for this purpose has been validated in numerous studies (8, 10, 12).

In the field of dental implantology, CBCT scans have been widely studied for their role in evaluating alveolar bone density. Liu et al. (10) demonstrated that CBCT images offer reliable data on bone density, making them a valuable tool for preoperative assessments in implant procedures. While CBCT grayscale values (GVs) differ from traditional Hounsfield Units (HUs), they operate on a similar principle, where radiation attenuation correlates with tissue density. This allows bone density to be effectively represented using CBCT GVs (10).

Moreover, Ahmed et al. (30) explored the use of CBCT for measuring alveolar bone density in HUs, confirming its effectiveness as a preoperative imaging tool. Their findings revealed significant regional variations in bone density, consistent with the Misch classification system. These insights assist clinicians in selecting appropriate implant types, surgical techniques, loading protocols, and success rates. Additionally, Nomura et al. (31) further noted that although CBCT GVs are generally higher than CT HUs, both metrics maintain a positive correlation with bone density, reinforcing their utility in clinical evaluations.

Recent innovations in bone density assessment include the work of Kwon et al. (2015) (32), who introduced a texture mapping technique using a graph-cut algorithm to visualize alveolar bone density distribution in CBCT images. This method segments bone regions based on predefined grayscale thresholds and applies texture patterns to these segments. However, its accuracy depends on precise threshold adjustments, which can be challenging due to patient-specific variations in bone density. Additionally, while effective for smaller datasets, the method's processing time increases with more segmentation levels, limiting its scalability for larger datasets.

A study by Sorkhabi et al. (33) explored CNN-based methods for alveolar bone density classification. Their dataset consisted of 207 target areas extracted from CBCT scans of 83 patients, divided into training (110), validation (54), and testing (43) subsets. Although the results were promising, the study acknowledged limitations, including the small dataset size and dependence on subjective clinical annotations.

Moreover, Xiao et al. (34) developed an AI model to classify jawbone density at implant sites using CBCT images. Their dataset included 605 PNG images derived from DICOM files of 70 patients. The model, built on the Nested-UNet architecture, categorized bone density into five types based on Hounsfield Unit (HU) ranges: Type 1 (1000–2000), Type 2 (700–1000), Type 3 (400–700), Type 4 (100–400), and Type 5 (−200–100). While the model achieved high accuracy, closely matching expert classifications, the study was limited by its small sample size of only 605 images from 70 CBCT scans. A larger dataset would be necessary to evaluate the system's reliability across diverse clinical scenarios.

Another study by Luo et al. (35) investigated the relationship between dataset size and the performance of deep learning models in classification tasks. Their findings emphasized that larger datasets generally enhance classification accuracy. Aligning with this insight, the current study utilized a significantly larger dataset of 5,545 images from 610 CBCT DICOM files. This dataset was divided into 4,645 training images, 300 validation images, and 600 testing images, ensuring a robust evaluation of the model's performance.

During the training phase, several pre-trained deep learning models were employed, with the SwinV2 model standing out due to its exceptional performance. It achieved the highest accuracy (85.65%), a strong AUC (0.73), and the lowest loss (0.8905), making it the most effective model in this study. These results are consistent with earlier research by Shamshad et al. (36) , which highlighted the advantages of transformer-based architectures in medical imaging. Similarly, a systematic review by Takahashi et al. (2024) (37) found that Vision Transformers (ViTs) outperform traditional convolutional neural networks (CNNs) in medical image classification tasks, further supporting the superior performance of the SwinV2 model observed in this study.

Notably, the ViTamin model demonstrated high sensitivity, suggesting its potential for identifying low-density areas, which is critical for early detection of conditions like osteopenia and osteoporosis. Meanwhile, the XciT model also performed well, achieving the second-best accuracy, specificity, and loss values, indicating its reliability for similar tasks.

Notably, a key finding was the significant impact of the binary mask in directing the model's attention to relevant jawbone regions. This aligns with the work of Fu et al. (38), who demonstrated the importance of incorporating depth information as an additional mask in segmentation tasks. The noticeable decline in accuracy when the mask was excluded highlights the necessity of including region-specific information to achieve precise classifications in medical imaging applications.

Error analysis for different models revealed that misclassifications predominantly occurred between adjacent bone density categories, especially in borderline cases. These errors were often associated with subtle radiographic features,

image quality issues, or model attention to non-diagnostic regions. This highlights the need for further refinement, including improved preprocessing.

Although the SwinV2 model achieved the highest overall accuracy and strong performance in our study, there remains substantial room for improvement in its ability to consistently differentiate between all five bone density classes, especially the less common ones.

One major challenge is the imbalance in the dataset, where certain bone density classes (such as D4 and D5) are underrepresented. This likely led to lower classification accuracy for those categories. To address this issue, more targeted data augmentation methods could be applied to boost the model's ability to learn from limited examples and generalize better.

Another area for improvement involves enhancing feature extraction. More advanced techniques that can capture fine-grained differences in bone texture and structure could improve classification. For example, using a loss function that guides attention mechanisms toward clinically important regions of the jaw may help the model focus on relevant features and reduce errors caused by irrelevant image areas.

Additionally, using ensemble methods, by combining outputs from different model versions or architectures, could help lower bias and variance in the predictions. However, computational constraints during this study limited the use of larger models or multiple ensembles.

Moreover, increasing the size and diversity of the dataset, particularly by including more examples of the minority classes and patients from varied backgrounds, would greatly enhance model robustness. A more balanced and comprehensive dataset would allow the model to learn more accurate representations for each bone density level.

While the SwinV2 model demonstrated high classification performance, it is important to note that Vision Transformers are often regarded as "black box" models due to their complex internal mechanisms, such as self-attention and the use of query-key-value operations for capturing long-range dependencies across input image regions, making their decision-making processes difficult to interpret. This can pose challenges in clinical adoption, where model transparency is essential.

It is important to note that, beyond diagnostic performance in dental implant planning, clinical integration is a key consideration for real-world utility. The SwinV2 model could be embedded within dental centers' systems to automatically assess bone density from radiographs as part of routine workflows. Such integration could facilitate early detection of osteopenia or osteoporosis, enable case prioritization, and function as a decision-support tool to augment radiologist performance.

Although the findings of this study are encouraging, they are constrained by the limited diversity of the dataset, which may affect the generalizability of the results. To address this, future research should focus on incorporating more varied datasets, encompassing different demographic groups and CBCT imaging devices, to ensure the models' applicability across a wider range of cases. Furthermore, expanding the scope to include other imaging modalities, such as computed tomography (CT) or magnetic resonance imaging (MRI), could provide a more comprehensive evaluation of the models' capabilities.

Another promising direction for future work is the development of hybrid models that leverage the strengths of vision transformers and convolutional neural networks. By combining the global context capture of ViTs with the local feature extraction capabilities of CNNs, such hybrid approaches could potentially achieve even greater performance improvements in medical image analysis tasks. These advancements would not only enhance the robustness of the models but also broaden their applicability in clinical settings.

CONCLUSIONS

The findings of this study reveal that advanced Vision Transformer models are highly capable of classifying jawbone density when applied to CBCT images. A notable improvement in performance was observed with the inclusion of a binary mask, which helps the models concentrate on specific regions of interest.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

FUNDING

No dedicated funding was obtained for this work.

REFERENCES

1. Al-Omiri M, Hantash RA, Al-Wahadni A. Satisfaction with dental implants: a literature review. Implant Dent. 2005;14:399-408.
2. Alfaraj TA, Al-Madani S, Alqahtani NS, Almohammadi AA, Alqahtani AM, AlQabbani HS, et al. Optimizing Osseointegration in Dental Implantology: A Cross-Disciplinary Review of Current and Emerging Strategies. Cureus. 2023;15:e47943.
3. Cassetta M, Stefanelli L, Di Carlo S, Pompa G, Barbato E. The accuracy of CBCT in measuring jaws bone density. Eur Rev Med Pharmacol Sci. 2012;16:1425-9.
4. Rios HF, Borgnakke WS, Benavides E. The use of cone-beam computed tomography in management of patients requiring dental implants: an American Academy of Periodontology best evidence review. J Periodontol. 2017;88:946-59.

5. Misch C. Bone classification, training keys to implant success. Dent Today. 1989;8:39-44.
6. Misch CE. Contemporary Implant Dentistry-E-Book: Contemporary Implant Dentistry-E-Book. 3rd ed. Canada: Mosby Elsevier; 2007.
7. Morar L, Băciuț G, Băciuț M, Bran S, Colosi H, Manea A, et al. Analysis of CBCT bone density using the Hounsfield scale. Prosthesis. 2022;4:414-23.
8. Hao Y, Zhao W, Wang Y, Yu J, Zou D. Assessments of jaw bone density at implant sites using 3D cone-beam computed tomography. Eur Rev Med Pharmacol Sci. 2014;18:1398-403.
9. Felicori SM, Gama RdSd, Queiroz CS, Salgado DMRdA, Zambrana JRM, Giovani ÉM, et al. Assessment of maxillary bone density by the tomodensitometric scale in Cone-Beam Computed Tomography (CBCT). J Health Sci Inst. 2015:319-22.
10. Liu J, Chen H-Y, DoDo H, Yousef H, Firestone AR, Chaudhry J, et al. Efficacy of cone-beam computed tomography in evaluating bone quality for optimum implant treatment planning. Implant Dent. 2017;26:405-11.
11. Radi IA-W, Ibrahim W, Iskandar SM, AbdelNabi N. Prognosis of dental implants in patients with low bone density: A systematic review and meta-analysis. J Prosthet Dent. 2018;120:668-77.
12. Parsa A, Ibrahim N, Hassan B, Motroni A, Van der Stelt P, Wismeijer D. Influence of cone beam CT scanning parameters on grey value measurements at an implant site. Dentomaxillofac Radiol. 2013;42:79884780.
13. Peng T, Zeng X, Li Y, Li M, Pu B, Zhi B, et al. A study on whether deep learning models based on CT images for bone density classification and prediction can be used for opportunistic osteoporosis screening. Osteoporos Int. 2024;35:117-28.
14. 1Nguyen TT, Larrivée N, Lee A, Bilaniuk O, Durand R. Use of artificial intelligence in dentistry: current clinical trends and research advances. J Can Dent Assoc. 2021;87:1488-2159.
15. Sarmadi A, Razavi ZS, van Wijnen AJ, Soltani M. Comparative analysis of vision transformers and convolutional neural networks in osteoporosis detection from X-ray images. Sci Rep. 2024;14:18007.
16. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, et al. A survey on vision transformer. IEEE Trans Pattern Anal Mach Intell. 2022;45:87-110.
17. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020.
18. Roongruangsilp P, Khongkhunthian P. The learning curve of artificial intelligence for dental implant treatment planning: a descriptive study. Appl Sci. 2021;11:10159.
19. Petrie A, Sabin C. Medical Statistics at a Glance. 3rd ed. Chichester, UK: Wiley-Blackwell; 2009.
20. Touvron H, Cord M, Sablayrolles A, Synnaeve G, Jégou H. Going deeper with image transformers. Proceedings of the IEEE/CVF international conference on computer vision. 2021:32-42.
21. Ali A, Touvron H, Caron M, Bojanowski P, Douze M, Joulin A, et al. Xcit: Cross-covariance image transformers. Advances in neural information processing systems. 2021;34:20014-27.
22. Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, et al. Swin transformer v2: Scaling up capacity and resolution. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022:12009-19.
23. Chen C-FR, Fan Q, Panda R. Crossvit: Cross-attention multi-scale vision transformer for image classification. Proceedings of the IEEE/CVF international conference on computer vision. 2021:357-66.
24. Chen J, Yu Q, Shen X, Yuille A, Chen L-C. ViTamin: Designing Scalable Vision Models in the Vision-Language Era. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024:12954-66.
25. Schwartz-Arad D, Laviv A, Levin L. Failure causes, timing, and cluster behavior: an 8-year study of dental implants. Implant Dent. 2008;17:200-7.
26. Jemt T, Book K, Lindén B, Urde G. Failures and complications in 92 consecutively inserted overdentures supported by Brånemark implants in severely resorbed edentulous maxillae: a study from prosthetic treatment to first annual check-up. Implant Dent. 1993;2:53.
27. Van Steenberghe D, Jacobs R, Desnyder M, Maffei G, Quirynen M. The relative impact of local and endogenous patient-related factors on implant failure up to the abutment stage. Clin Oral Implants Res. 2002;13:617-22.
28. Lee H, Jo M, Sailer I, Noh G. Effects of implant diameter, implant-abutment connection type, and bone density on the biomechanical stability of implant components and bone: A finite element analysis study. J Prosthet Dent. 2022;128:716-28.
29. Jaju PP, Jaju SP. Cone-beam computed tomography: time to move from ALARA to ALADA. Imaging Sci Dent. 2015;45:263-5.
30. Ahmed M, Ikram Y, Qureshi F, Sharjeel M, Khan ZA, Ataullah K. Assessment of jaw bone density in terms of Hounsfield units using cone beam computed tomography for dental implant

treatment planning. Pak Armed Forces Med J. 2021;71:221-27.
31. Nomura Y, Watanabe H, Honda E, Kurabayashi T. Reliability of voxel values from cone-beam computed tomography for dental use in evaluating bone mineral density. Clin Oral Implants Res. 2010;21:558-62.
32. Kwon K, Kang D-S, Shin B-S. Multiple texture mapping of alveolar bone area for implant treatment in prosthetic dentistry. Comput Biol Med. 2015;56:89-96.
33. Sorkhabi MM, Khajeh MS. Classification of alveolar bone density using 3-D deep convolutional neural network in the cone-beam CT images: A 6-month clinical study. Measurement. 2019;148:106945.
34. Xiao Y, Liang Q, Zhou L, He X, Lv L, Chen J, et al. Construction of a new automatic grading system for jaw bone mineral density level based on deep learning using cone beam computed tomography. Sci Rep. 2022;12:12841.
35. Luo C, Li X, Wang L, He J, Li D, Zhou J. How does the data set affect cnn-based image classification performance? 2018 5th international conference on systems and informatics (ICSAI). 2018:361-6.
36. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: A survey. Med Image Anal. 2023;88:102802.
37. Takahashi S, Sakaguchi Y, Kouno N, Takasawa K, Ishizu K, Akagi Y, et al. Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. J Med Syst. 2024;48:1-22.
38. Fu Y, Fan J, Xing S, Wang Z, Jing F, Tan M. Image segmentation of cabin assembly scene based on improved RGB-D mask R-CNN. IEEE Transactions on Instrumentation and Measurement. 2022;71:1-12.