Journal of Imaging Informatics in Medicine
https://doi.org/10.1007/s10278-026-02053-4

Transformer and Attention Enhanced Deep Learning Approach for CBCT-Based Mental Foramen Classification and Segmentation

Osman Güler[1] · **Mustafa Teke**[2] · **Zafer Civelek**[2]

Accepted: 4 June 2026
© The Author(s) 2026

Abstract
Determining the location and morphological characteristics of the mental foramen is of critical clinical importance in dental surgical procedures. In this study, the mental foramen was classified and segmented using deep learning methods on CBCT images. The study utilized a unique and manually annotated CBCT dataset. An ensemble deep learning model incorporating Transformer and Attention mechanisms was developed for classification. A UNet structure, also incorporating Transformer and Attention mechanisms, was proposed for segmentation. The models were trained and tested on images containing variations in the anatomical location, shape, and size of the mental foramen. The proposed DAFNet model achieved 95.20% validation accuracy and 91.25% test accuracy with an AUC of 0.9648, while the proposed TA-UNet segmentation model achieved a Dice score of 0.8804 and IoU of 0.8386, outperforming conventional deep learning architectures. The findings demonstrate that the mental foramen can be automatically detected with high accuracy in CBCT images and its segmentation can be reliably performed.

Keywords CBCT · Mental foramen · Deep learning · Artificial intelligence · Explainable artificial intelligence

Introduction

The mental foramen (MF) is an important anatomical feature of the mandible found on the buccal surface of the mandible, underneath the anterior molars [1]. It is an anatomical passage through which the mental nerve, a branch of the inferior alveolar nerve (a key branch of the mandibular nerve from the fifth cranial nerve), travels along with the mental artery and vein.

It innervates the soft tissues surrounding the lower lip, mandible, and lower anterior teeth [2]. Determining the location and number of foramina in the MF region is crucial before implant or other surgical procedures [3]. Figure 1 shows the mandibular foramen, mandibular canal, and the MF.

In recent years, artificial intelligence (AI) techniques such as computer vision, machine learning, deep learning (DL), transfer learning (TL), and ensemble learning have been widely used in medical image classification. These AI methods and techniques used in healthcare require large amounts of medical image data to provide accurate and reliable results [4]. Many studies have been conducted using DL architectures in healthcare for the diagnosis and early detection of diseases such as various types of kidney cancer [5], skin lesions [6], liver cirrhosis [7], hand tremors [8], gastrointestinal disease [9], and dental radiography [10].

In light of relevant studies, panoramic radiography and cone-beam computed tomography (CBCT) techniques appear to be the most commonly used imaging methods for detecting anomalies in teeth. Because panoramic radiography only shows the oral and dental structure in a single view, it is difficult to determine the true three dimensional (3D) image of the entire oral structure and the anomaly [11]. Therefore, additional examinations can be performed using CBCT to confirm the anomaly in a 3D view. However, in CBCT images, which are commonly preferred for mandibular canal and MF detection, the anomaly part occupies a very small portion of the CBCT image, making it difficult to distinguish [12].

✉ Mustafa Teke
mustafateke@karatekin.edu.tr

[1] Departmant of Electronic and Automation, Technical Sciences Vocational School, Gazi Universty, Ankara, Turkey

[2] Department of Electrical and Electronics Engineering, Faculty of Engineering, Çankırı Karatekin University, Çankırı, Turkey

Published online: 16 June 2026

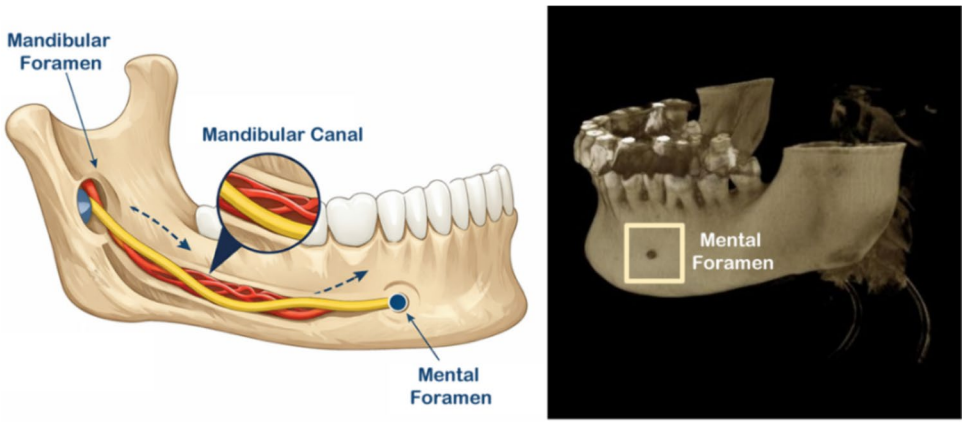


Fig. 1 Mandibular foramen, mandibular canal, and MF

Currently, AI has been applied extensively to the medical field for diagnosis and segmentation and classification tasks in images [13]. Specifically, there have been several breakthroughs through successful classification and segmentation of images by applying DL and convolutional neural networks (CNN) [14]. In dental health analysis, tooth and anatomical structure identification is a laborious process since this task is done by dentists manually. Automation of this task by developing AI systems capable of examining images to focus on anatomical structure rather than pathology can help dentists [15]. Several researchers have explored different AI methods for segmentation and anomaly identification in dental images.

Dasanayaka et al. [16] executed the segmentation task utilizing the UNet model with a custom dataset featuring 1000 images of dental panoramic tomography, each measuring 512×512 pixels. They divided the dataset evenly into 80% training, 10% validation, and 10% testing. In the concluding phase, the model achieved a Dice score of 0.991 during the validation stage. The Dice score averaged 0.987 during the testing phase.

Kats et al. [17] employed a single fully convolutional network based on UNet along with a cascaded architecture for detecting and segmenting MF on a custom dataset consisting of segmentation masks from 112 digital panoramic X-ray images. Seventy percent of the data was allocated for training, 15% for validation, and 15% for testing. Consequently, the highest results for Dice, sensitivity, recall, MF-based true positive rate (MFTPR), and MF-based false positive rate (MFFPR) were 49.51%, 71.13%, 68.24%, 87.81%, and 14.08%, respectively, in individual networks, whereas the cascaded network performance reached 31.77%, 23.92%, 88.83%, and 93.75%, respectively, for Dice, sensitivity, recall, and MFTPR.

Jaskari et al. [18] utilized CNN segmentation on a collection of 637 CBCT volumes where radiologists annotated the mandibular canals. The suggested model identified the mandibular canals of the voxel-level marked cluster with mean curve distance and mean symmetric surface distance measuring 0.56 mm and 0.45 mm, respectively.

Steybe et al. [19] introduced a multi-scale 3D CNN stack method utilizing a patchwork toolbox for the automated segmentation of head CT images. The segmentation outcome produced Dice values of 0.61 for the MF and 0.98 for the mandible.

Edvardsen et al. [20] examined the identification of the MF utilizing EfficientNet-B0 and RetinaNet approaches within a dataset of dental panoramic radiographs for automated evaluation of mandibular cortical width. A collection of 5197 dental panoramic radiographs was divided into 80% for training and 20% for testing. The EfficientNet-B0 model obtained an average accuracy of 0.30, whereas the RetinaNet model attained an average accuracy of 0.23. In the end, they showed that the EfficientNet-B0 model identified the MF with greater effectiveness.

Murugan et al. [21] focused on MF classification employing a CNN on a proprietary dataset composed of 468 digital panoramic images. The model demonstrated a performance of 0.991 on the validation set. The statistical findings derived from SPSS software indicated that type 4 MF appears more often along the long axis of the second premolar.

Bağ et al. [22] aimed to categorize maxillary and mandibular anatomical features seen in children's panoramic radiographs utilizing 2D CNN and YOLO-v5 models. In total, 981 pediatric patient images were annotated for 9 distinct pediatric anatomical reference points, which included the maxillary sinus, orbit, mandibular canal, MF, mandibular foramen, mandibular incisura, articular process, and condylar and coronoid processes, achieving the highest F1 Scores from orbit (1), mandibular incisura (0.99), maxillary sinus (0.98), and mandibular canal (0.97).

Raza et al. [23] employed two types of masks, circular and square, to enhance MF detection and segmentation from a proprietary dataset of 1000 panoramic radiographic images. Classical UNet, ResUNet++, UNet Attention, and LinkNet models were employed for the identification and

segmentation of MF. They claimed that the traditional UNet model outperformed other models, achieving a Dice coefficient of 0.79 and an intersection over union (IoU) of 0.67 on the test data.

Çelebi et al. [24] examined the efficacy of YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11, YOLO World, and RT-DETR models for MF identification and localization using a collection of 100 anonymized panoramic radiological images. Following the tests, the RT-DETR-x architecture attained the best performance, recording 89.0% accuracy, 85.0% sensitivity, and 85.6% mAP50.

Ünal and Namdar Pekiner [25] examined CBCT data from 300 patients utilizing DL techniques. This research assessed the relationship between MC-MM3 and automatically identified the impacted mandibular third molar, mandibular canal, MF, and mandibular foramen. The DL model utilized the nnU-NetV2 architecture. For MC assessment, the model reached an accuracy of 0.99, a sensitivity of 0.75, a precision of 0.78, a Dice score of 0.76, a Jaccard index of 0.62, and an AUC of 0.88. For MF evaluation, the associated values for accuracy, sensitivity, precision, Dice, Jaccard, and AUC were 0.99, 0.64, 0.66, 0.64, 0.57, and 0.82, respectively. For the evaluation of the mandibular foramen, the accuracy, sensitivity, precision, Dice, and AUC metrics were 0.99, 0.79, 0.68, 0.71, and 0.90, respectively.

Numerous research efforts have explored the automated identification and segmentation of the mental foramen employing artificial intelligence methods. Initial research focused mainly on CNNs used for panoramic radiographs. Recently, deep learning techniques have been utilized on CBCT data to examine anatomical features in the mandible. Additional research has investigated object detection and localization methods like YOLO-based frameworks and EfficientNet models to recognize anatomical landmarks, including the MF, in dental X-rays. Automated MF detection in the literature is typically viewed as an enhancement of mandibular canal segmentation from CBCT images [26–28]. In various research, panoramic radiographs have frequently been utilized for the detection of MF [29, 30]. There are few studies that specifically focus on MF segmentation. Most existing methods depend on conventional CNN architectures and fail to fully leverage recent innovations like transformer-based feature modeling and attention techniques. These constraints clearly emphasize the necessity for clinically relevant DL models developed with CBCT data focused on precise MF identification and segmentation in CBCT images. This research suggests a segmentation and classification method utilizing DL with a distinctive CBCT dataset.

In the proposed study, a unique and custom MF CBCT dataset was collected, and ground truth masks were created by expert dentists. Furthermore, while MF is generally considered part of mandibular canal segmentation in the literature, in this study, it was directly identified as the target structure, and both classification and segmentation studies were performed. In the study, the proposed DL model was tested on CBCT images containing different position, shape, and size variations of MF, thus evaluating the generalizability of the model to anatomical diversity. This study strengthens the bridge between dental radiology and artificial intelligence applications by revealing the potential use of the proposed DL-based approach for automatic MF detection in clinical processes.

Methodology

This paper designs a two-stage DL architecture to identify and classify the boundary intricacies of anomalies related to the MF in CBCT scans pertaining to dental and oral health. In the proposed methodology, the first step involves the classification of images in terms of the absence/presence of anomalies, and in the second step, the imaging predicted as containing anomalies, the segmentation of the anomaly is performed. For the classification task, the proposed model is termed DAFNet (Dense-Efficient Attention Fusion Network), which fuses two TL models as feature extractors and is further augmented with attention and transformer frameworks. For the segmentation task, a variant of the UNet architecture called TA-UNet is designed and augmented with transformer and attention mechanisms.

Dataset

In this study, MF detection was performed on a dataset consisting of a total of 1264 CBCT images, including 632 anomaly images and 632 normal images, created by the Çankırı Karatekin University Oral and Dental Health Application and Research Center [31]. A total of 3000 CBCT images were retrospectively examined from the institutional archive at the beginning of the investigation. To improve diagnostic precision, an endodontist and a dentomaxillofacial radiologist conducted the screening separately. A total of 632 CBCT images with an AMF that satisfied the following requirements—namely, adequate image quality and region of interest completeness—were selected for study. A total of 632 CBCT scans with normal MF anatomy that were matched for sex and age were chosen at random to serve as controls. Thus, 1264 CBCT scans made up the total sample size for statistical analysis. A maxillofacial radiologist and an endodontist separately examined a first subset of 20 CBCT scans to evaluate the uniformity of the annotation procedure.

The experts were unaware of each other's ratings and any underlying clinical information. Inter-observer agreement was assessed by calculating the kappa coefficient, which

showed high reliability between observers ($\kappa = 0.88$). Following this, the experts annotated the remaining dataset, and consensus labelling was reserved for instances of disagreement. The radiological features employed to differentiate between anomaly MFs and nutrient foramina included continuity with the mental canal, cortical borders, and visualization in all axial, coronal, and sagittal planes.

The CBCT scans included in this study had to fulfill the following criteria:

1) They had to cover the mandibular premolar region on both sides.
2) They had to have good image quality, allowing reliable visualization of the MF region.
3) They had to be from patients who had no history of mandibular surgery and pathology in the region of the MF area.

CBCT scans that satisfied any of the following requirements were not included:

(1) Low spatial resolution or strong motion artifacts
(2) Metallic artifacts that block the MF region
(3) Significant pathological lesions in the area of interest, prior mandibular trauma, surgical intervention, or (4) insufficient anatomical coverage of the region being assessed

The Castellini X-Rad1 US TR10 Plus device was used to obtain the CBCT scans with the following settings: tube voltage of 90 kVp, tube current of 5 mA, and exposure time of 7 pulses per second. The field of view (FOV) and voxel size were chosen at 10×8 mm and 0.3 mm, respectively. A Lenovo 15.0 510 80 SR monitor was used to assess the pictures in suitable lighting.

Since CBCT imaging produces volumetric data, the images used in this study were extracted from different slices and viewpoints of the 3D scans. Therefore, although multiple images may originate from the same patient, they correspond to different anatomical perspectives and distances within the mandibular region. The dataset was divided at the image level to preserve the diversity of training samples. The images were rescaled to 224×224 pixel dimensions for classification and normalized to the range [0, 1]. To enhance the generalization ability of the proposed model, we increased the number of images in the dataset. Data augmentation techniques including rotation, translation, and zooming were applied only to the training subset after the dataset splitting process in order to avoid information leakage into validation and test datasets. Sample images from the dataset are shown in Fig. 2.

In the generated MF dataset, 632 anomaly images were resized to 256×256 pixels for boundary detection and segmentation tasks. The bounding boxes were then used by expert dentists to generate mask images for training. The ground truth masks were manually annotated by a dentomaxillofacial radiologist and an endodontist, and disagreements were resolved through consensus. It is worth pointing out that Fig. 3 displays sample images, bounding boxes, and mask images.

The dataset used in this study was retrospectively collected from the Çankırı Karatekin University Oral and Dental Health Application and Research Center. All CBCT scans were anonymized before analysis. The dataset is not publicly available but can be shared upon reasonable request.

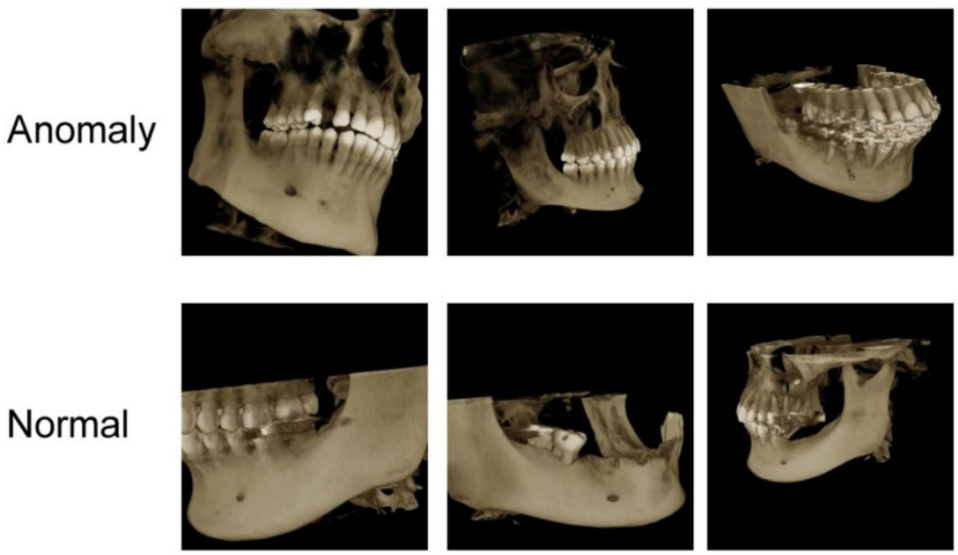


Fig. 2 Sample images of dataset

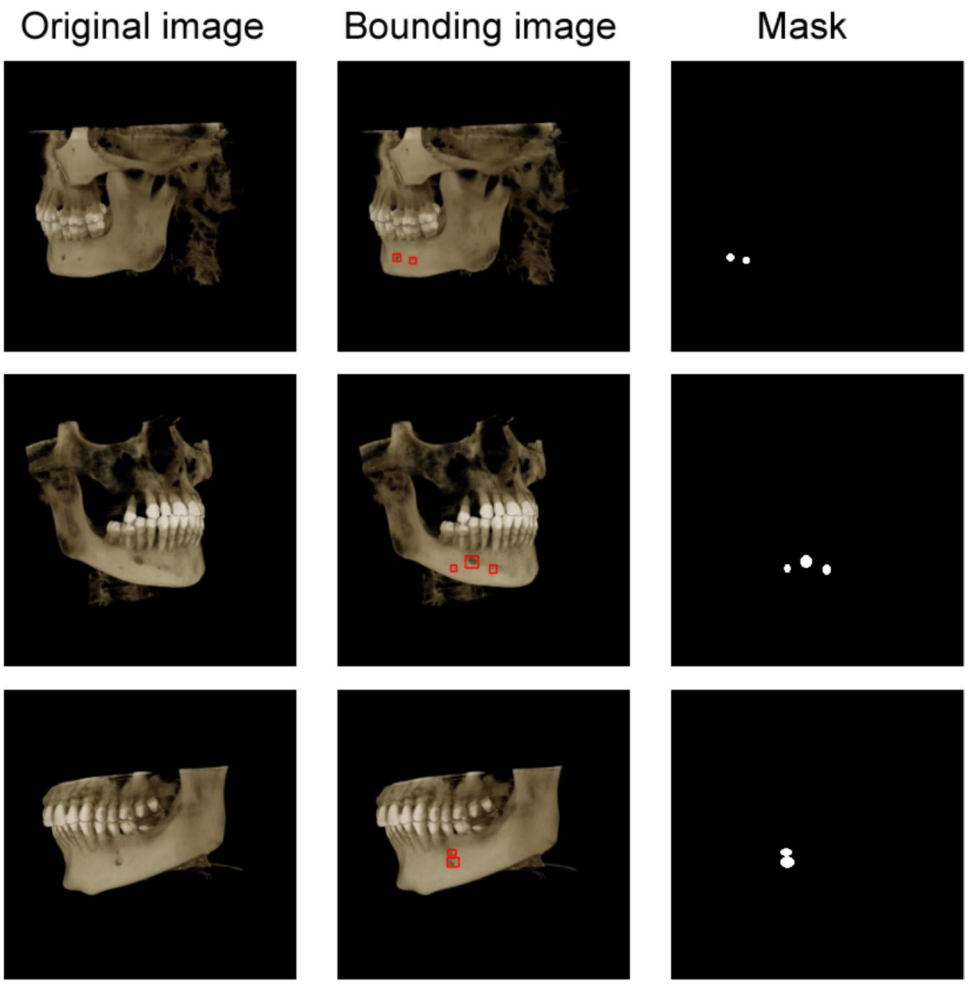


Fig. 3 Anomaly images, bounding boxes and masks

Proposed Model

In this research, there was two-phased workflow: MF classification and image segmentation from CBCT images. A hybrid architecture based on TL models, transformer model, and attention mechanism was proposed for MF classification. There were two different TL models used for this proposed architecture. They include DenseNet121 [32], EfficientNetB0 [33], InceptionV3 [34], MobileNetV2 [35], ResNet101 [36], ViTB16 [37], and Xception [38] from Keras libraries. The two most appropriate models for this proposed architecture were DenseNet121 and EfficientNetB0. In DenseNet121 and EfficientNetB0, parallel feature extractors are used at the input layer, followed by global average pooling layers for acquiring feature vectors.

The combined feature vectors are fed into a Transformer Encoder Block that has a multi-head attention component to extract the overall information. This is represented by the following code block. Thereafter, a Channel Attention Block is used to enhance the discriminative ability of the features. The Channel Attention Block is responsible for identifying the relationships between the features. The unnecessary information in the features is eliminated by the Channel Attention Block. The dense layer and Sigmoid activation function were employed as classifiers.

The Grad-CAM (Gradient-weighted Class Activation Mapping) technique was incorporated to improve clinical dependability and guarantee the transparency of the proposed DAFNet model's decision-making process. Grad-CAM is unique in that it can interpret CNN pictures by emphasizing areas of the image that are important for making decisions [39].The regions the model concentrates on when generating anomalous judgments were displayed by merging the activation maps from the DenseNet and EfficientNet branches. Figure 4 shows the architecture of the suggested DAFNet classification model.

To ensure reproducibility, the main architectural and training hyperparameters of DAFNet are explicitly reported. The model combines DenseNet169 and EfficientNetB0 backbones, whose pooled features are projected into a 1024-dimensional embedding space and used as tokens for a transformer encoder with 1 layer and 4 attention heads (key dimension = 64, feed-forward dimension = 1024). A channel attention block (reduction ratio = 8) is applied after

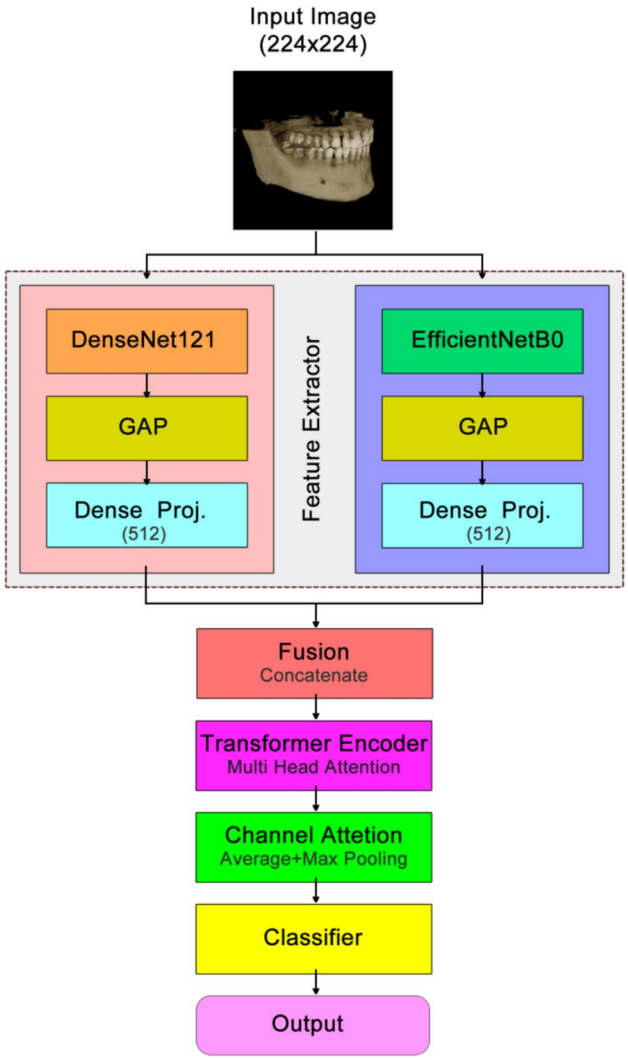


Fig. 4 Diagram of proposed DAFNet classification approach

the transformer output, followed by fully connected layers with 256 and 128 neurons. The model is trained using the Adam optimizer with a learning rate of 1×10^{-4}, batch size of 10, and 30 training epochs. Binary cross-entropy is used as the loss function, and input images are resized to 224×224 pixels.

In the segmentation process, the second stage of the proposed structure, a UNet based TA-UNet architecture with customized encoder and skip connections was developed to precisely determine the boundaries of the MF in images detected as anomalies. The proposed model uses the MobileNetV2 backbone as the encoder, which has high computational efficiency. This allows for faster convergence and high performance with fewer parameters. Vision Transformer blocks were used to model global inter-pixel dependencies at the bottleneck point between the encoder and decoder. By applying a multi-headed attention mechanism to the image patches, the local receptive field limitation of convolutional layers is overcome, allowing for a better understanding of the holistic context of the anatomical structure.

In this study, skip connections, a key element in the success of UNet architectures, were strengthened with CBAM (Convolutional Block Attention Module). CBAM reduces noise in the information transmitted to the decoder by filtering the low-level feature maps from the encoder along both the channel and spatial axes, ensuring that only relevant structural details such as edges and texture are transmitted. In the decoder stage, a segmentation mask is generated using stepwise upsampling and convolution blocks. The architecture of the proposed TA-UNet segmentation model is given in Fig. 5.

Experimental Results

In this study, TensorFlow and Keras libraries were used for the detection of MF from CBCT images. Training was performed using the Adam optimization algorithm; "Binary Crossentropy" was preferred for classification, and a loss function that also tracks Dice and IoU coefficients was preferred for segmentation. To prevent overfitting, "Early Stopping" and "ReduceLROnPlateau" callback mechanisms, which dynamically reduce the learning rate, were used. In performance evaluation, accuracy, precision, recall, F1-Score, and ROC-AUC metrics were calculated for classification, and Dice similarity coefficient (DSC), Jaccard Index (IoU), and Boundary F1 Score metrics were calculated for segmentation. The dataset for classification was divided into three parts: training (764), validation (400), and test (100). After data splitting, data augmentation was applied to the training data. Therefore, 4000 images were used for training, 400 for validation, and 100 for testing.

To enable generalization and fair comparison on the dataset, training was first conducted using commonly used transfer learning models from the Keras library in the TensorFlow environment. In addition, five-fold cross-validation was applied to evaluate the robustness and stability of the models. The experimental results of the transfer learning models and the proposed DAFNet model for MF classification, including both five-fold cross-validation results and final test performance, are presented in Table 1.

From a careful observation of the values in Table 1, it is clear that the DenseNet121 model shows relatively balanced and consistent performance in terms of validation and test accuracy rates, indicating a reliable predictive performance among the transfer learning models. Though the InceptionV3 model shows one of the highest values in terms of validation accuracy, a decrease in test accuracy is also observed, indicating a relatively limited generalization performance of this model. In contrast, it is clear from the above

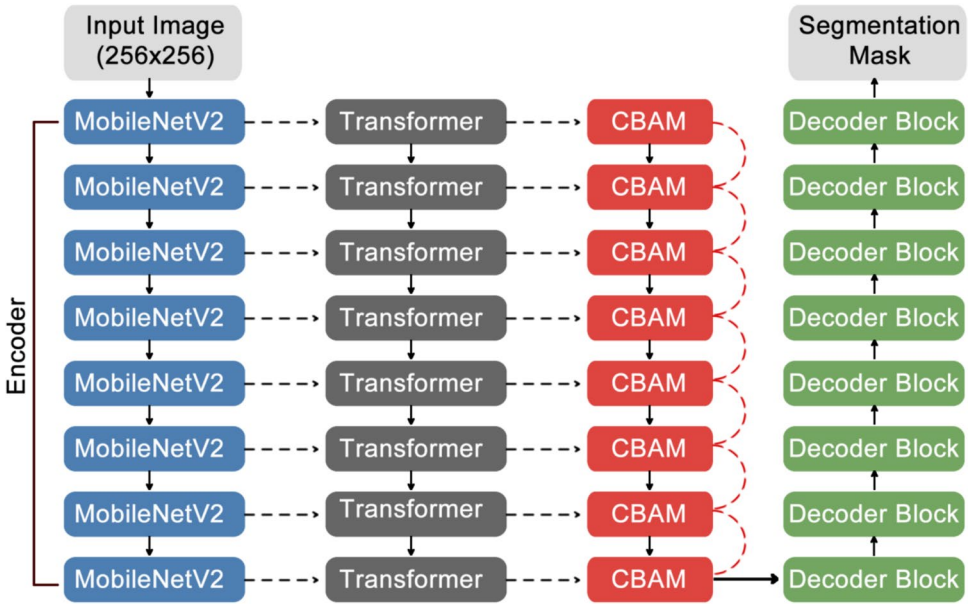


Fig. 5 Diagram of proposed TA-UNet segmentation approach

Table 1 Results of TL models and the proposed DAFNet model

Model	Validation accuracy (mean ± SD)	Test accuracy	Precision	Recall	F1 Score	AUC	MCC	BAS	*p*-value (DAFNet, t-test)
DenseNet121	87.80 ± 3.52	83.00	0.85	0.83	0.83	0.9088	0.6835	0.83	0.011
EfficientNetB0	79.48 ± 6.46	71.00	0.74	0.71	0.71	0.71	0.69	0.71	0.005
InceptionV3	91.22 ± 0.90	86.00	0.87	0.86	0.86	0.9248	0.7252	0.86	0.00095
MobileNetV2	84.62 ± 2.62	79.00	0.79	0.79	0.79	0.8904	0.580	0.79	0.0015
ResNet101	86.82 ± 3.10	85.00	0.85	0.85	0.85	0.9244	0.7012	0.85	0.005
ViT-B16	69.20 ± 2.99	68.00	0.60	0.68	0.68	0.608	0.745	0.68	< 0.001
Xception	83.40 ± 0.73	83.00	0.84	0.83	0.83	0.923	0.6665	0.83	< 0.001
DAFNet	**95.20 ± 0.87**	**91.25**	**0.94**	**0.88**	**0.90**	**0.9648**	**0.9012**	**0.91**	—

table that the proposed DAFNet model shows relatively better performance in comparison to the other architectures in terms of both validation and test data sets. For example, it shows the highest value in terms of 95.20% validation accuracy and 91.25% test accuracy when compared to the performance of the other architectures. Furthermore, it is also clear from the above table that relatively 0.94 precision and 0.88 recall values are observed, indicating balanced performance in terms of prediction by reducing false positive rates while keeping high true positive rates, and a relatively 0.90 F1-score is also observed. The 0.9648 AUC value indicates the high capacity of the model to distinguish between the positive and negative classes, whereas the 0.9012 MCC and 0.91 BAS values indicate the reliable predictions of the model, even for the potentially imbalanced classes.

Table 2 displays the results obtained from the ANOVA test, which is performed to statistically verify the significance of the performance variation among the compared models. As different deep learning models were tested using the accuracy results obtained from the five-fold cross-validation method, the one-way ANOVA test is performed to verify the variation of accuracy among the different models.

Table 2 ANOVA test results

Test	*F*-value	*p*-value
One-way ANOVA	30.57	1.93×10^{-12}

To statistically evaluate the performance differences between the models, a one-way ANOVA test was conducted using the five-fold cross-validation results. The analysis revealed a statistically significant difference between the models with an F value of 30.57 and a significance level of p less than 0.001. Furthermore, paired t-tests comparing the proposed DAFNet with the baseline architectures confirmed that the proposed model significantly outperformed the other deep learning models with p values lower than 0.05. These findings demonstrate that

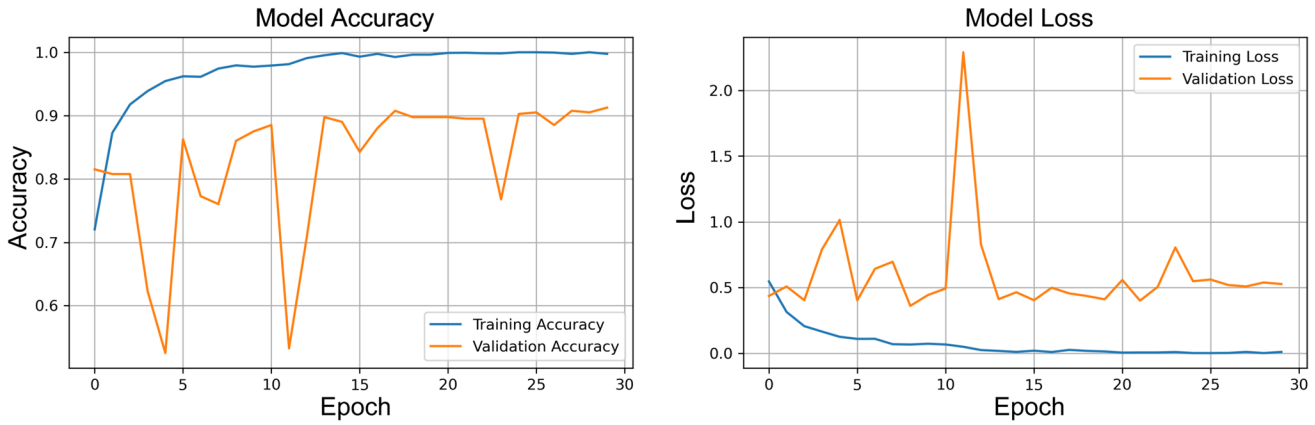


Fig. 6 Accuracy and Loss graphs of proposed DAFNet model

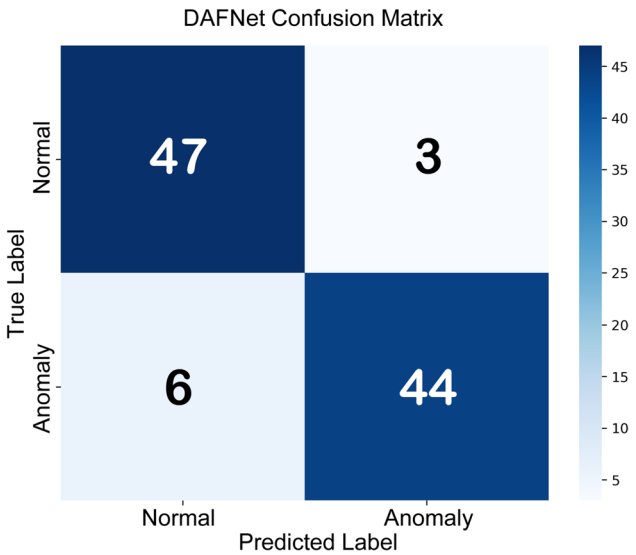


Fig. 7 Confusion matrix of proposed DAFNet model

the proposed DAFNet architecture provides statistically superior performance and stronger generalization capability compared with widely used transfer learning models.

The training and validation accuracy and loss graphs of the proposed DAFNet model for anomaly detection on the MF dataset are given in Fig. 6 and the confusion matrix in Fig. 7.

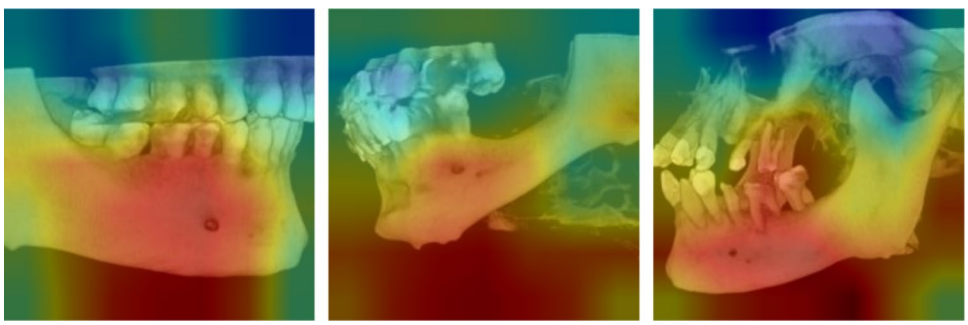
Figure 6 shows that when the accuracy and loss values of the model are examined, it is seen that the model has a high learning capacity. Although the validation rate fluctuates, its overall upward trend and exceeding 90% indicate that the model has a strong generalization ability. When the loss curves are examined, it is seen that the training loss decreases steadily and approaches zero, while the validation loss shows sudden increases from time to time but remains within a stable range. This indicates that the model follows a balanced learning process without overlearning.

Figure 7 shows that the confusion matrix of the DAFNet model correctly classified 47 normal instances and 44 anomaly instances, while only 3 normal instances were incorrectly classified as anomalies and 6 anomaly instances were incorrectly classified as normal. This indicates that the model exhibits balanced performance for both classifications.

Figure 8 shows the Grad-CAM outputs of the DAFNet model. In the color maps, red and yellow tones represent the regions that contribute most to the model's classification decision; blue and green tones represent relatively less influential areas.

When Fig. 8 images are examined, it is seen that the proposed DAFNet model focuses particularly on the mandibular bone structure, suspected lesion areas, and

Fig. 8 Grad-CAM images of the proposed DAFNet model



surrounding anatomical areas. The absence of significant activation in the background regions indicates that DAFNet makes its classification decisions based on clinically significant anatomical structures rather than random pixels.

In conclusion, the proposed DAFNet model offers high accuracy, stable, balanced, and reliable performance. These results strongly support the idea that the proposed model is suitable for clinical applications.

In the second phase of the study, the average results of the five-fold cross-validation of the segmentation process performed on anomaly images with the UNet model and the proposed TA-UNet model are shown in Table 3 and statistical test results are given in Table 4.

The segmentation performances of the models were evaluated using Dice and IoU metrics obtained from five-fold cross-validation, together with additional evaluation metrics including precision, recall, F1 score, boundary F1 score, and mean average precision. The results show that the TA-UNet model achieved significantly higher segmentation performance compared with the conventional UNet architecture. The mean Dice score of TA-UNet was 0.8804 ± 0.038, whereas UNet achieved a mean Dice score of 0.6763 ± 0.037. Similarly, the average IoU value obtained by TA-UNet was substantially higher than that of UNet. The proposed model also produced higher precision, recall, and F1 score values, indicating more accurate detection of MF regions and better balance between false positive and false negative predictions. Furthermore, the TA-UNet architecture achieved higher BF1 and mAP values, demonstrating improved boundary localization and overall segmentation reliability. Statistical analysis confirmed the superiority of TA-UNet, where paired *t*-test and Wilcoxon signed-rank test results indicated statistically significant differences between the models. These findings demonstrate that the attention-based TA-UNet architecture provides more accurate and robust segmentation performance compared with the baseline UNet model.

Figure 9 shows the segmentation results of the UNet and proposed TA-UNet models on a sample image.

When Fig. 9 is examined, it can be seen that the UNet model can generally detect anomaly regions, but the predicted region does not completely match the ground truth. In contrast, the segmentation results of the TA-UNet model show a high agreement with the ground truth in terms of both location and shape. It is clearly seen that the predicted regions cover the target area more accurately, and the boundaries are clearer and more consistent.

Table 4 Statistical test results

Test	*p*-value	Result
Paired t-test (Dice)	0.0009	Significant difference
Wilcoxon test	0.031	Significant difference

In the experimental studies conducted, graphical analyses based on various metrics were performed to evaluate the segmentation performance of the proposed TA-UNet model. Performance curves of the TA-UNet model are presented in Fig. 10.

When Fig. 10 is examined, a stable increase in accuracy and Dice coefficient is observed in the training process of the TA-UNet model. From the graphs, the accuracy, Dice, and IoU increases shown in the training period represent a successful validation process by the model.

It is observed that the TA-UNet model is more accurate in comparison to the UNet model. This reveals the effectiveness of the new model compared to the previous one.

Discussion

In this work, classification and segmentation models have been developed regarding the detection of MF in the images obtained from the CBCT modality. The proposed DAFNet model was successful in the classification task of MF with the highest accuracy in classification with 95.20% validation accuracy, 91.25% test accuracy, and AUC 0.9648. Additionally, when other evaluation metrics were taken into account, it could be argued that the method was successful in preventing the increase of false positive cases and false negative cases. In the segmentation task, the proposed TA-UNet model was successful in achieving the highest Dice measure of 0.8804 and IoU measure of 0.8386. If the measures obtained in the work are taken into account and when considered relative to the studies in the literature regarding the detection of MF, the level of success obtained in the proposed classification and segmentation models is considerable.

It was shown that DAFNet and TA-UNet performed very well on MF detection and segmentation. CBCT imaging was also a better option than previously published models using

Table 3 Results of segmentation performance

Model	Dice (Val) Mean ± SD	IoU (Val) Mean ± SD	Precision	Recall	F1 Score	IoU (Test)	BF1	mAP
UNet	0.6763 ± 0.037	0.5144 ± 0.041	0.8186	0.6441	0.7001	0.5640	0.9469	0.5232
TA-UNet	0.8804 ± 0.038	0.8386 ± 0.016	0.9005	0.9375	0.9078	0.8464	0.9962	0.8409

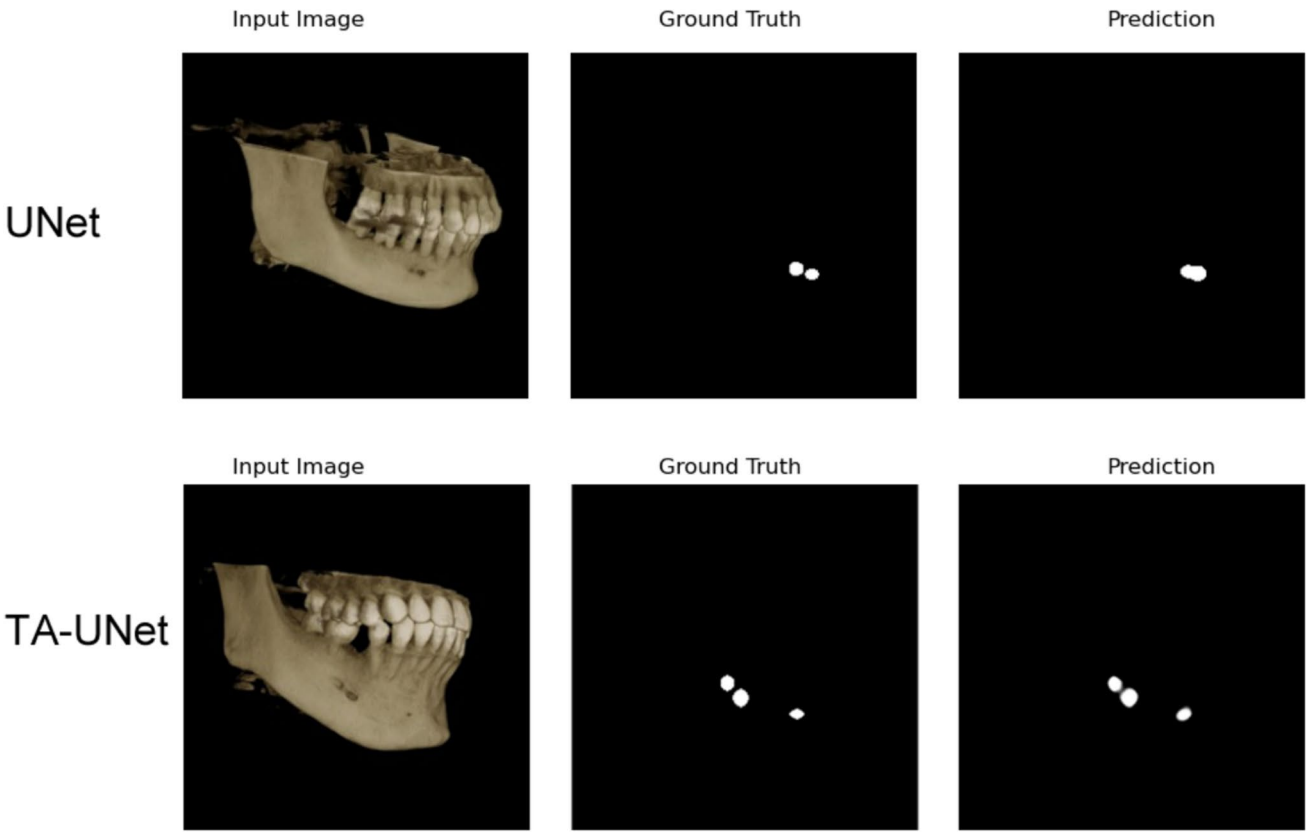


Fig. 9 Segmentation images of UNet and proposed TA-UNet models

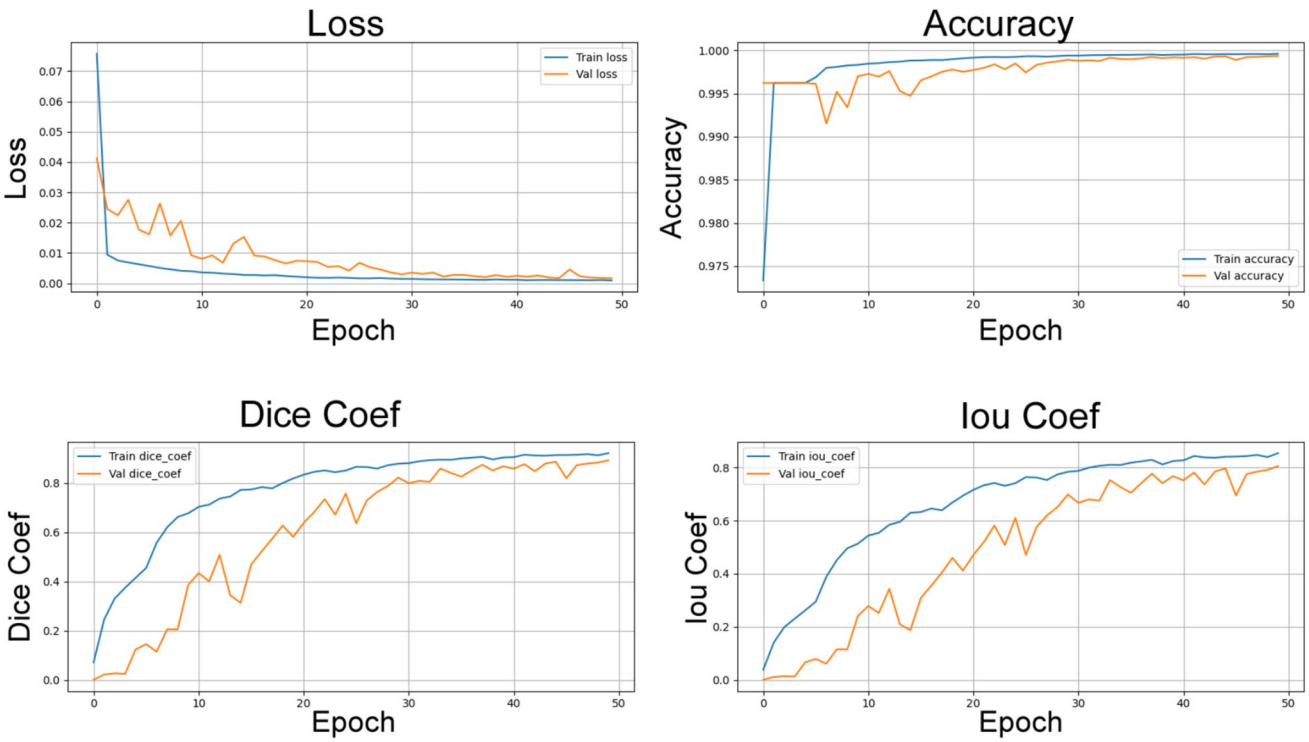


Fig. 10 TA-UNet performance metrics

panoramic scanners for accurate visualization of structures. The combination of transformers and attentional network architectures enabled the extraction of both local and global features of structures within an image. These different types of features likely led to more accurate classification and segmentation results achieved from our models. Of those evaluated, DAFNet performed the best when compared to baseline models, receiving the highest score in accuracy, AUC, and MCC. This is thought to have occurred largely due to the ability of DAFNet to capture many different size features within the CBCT images and the complexity of how they are organized. CNN-based frameworks including DenseNet121 and ResNet101 obtained competitive results as they still have good feature extraction capabilities. However, transformer frameworks like ViT-B16 produced inferior results compared to other networks likely since they struggle when provided insufficient data or are affected by small datasets.

Grad-CAM graphical representations show researchers the different areas of the model's focus in predicting outcomes. In most cases, the focal point of the predicted outcome is on the anatomical region of the mental foramen and clinically interpretable from the models. In rare situations, the model's focus and attention may extend to areas, resulting in misclassification or lower accuracy of localization. Identifying the patterns of attention will also enable the researcher to better identify limits of the model and provides more transparent information for the possible application of AI-assisted systems in a clinical environment.

From a clinical viewpoint, AI-assisted systems have the potential to aid dentists and radiologists in the analysis of CBCT images, especially when it is important to accurately locate the mental foramen. Proper identification and localization of this anatomical structure is critical to the performance of many dental procedures, including implant planning, surgical intervention, and the administration of local anesthesia. Consequently, automated detection systems could reduce the workload of the clinician and provide better clinical decision making.

There are a number of limitations associated with the proposed models that must be considered despite their promising results. The first limitation is that the dataset is relatively small compared to many large-scale medical imaging datasets. Data augmentation techniques were used in this study to increase variability and improve robustness, but these techniques cannot fully replicate the statistical diversity of larger datasets.

The second limitation is that although the images were produced using the 3D CBCT volumes, and those images were analyzed using 2D slices, the data collected from the two-dimensional slices is only able to capture contextually three-dimensional information to a limited extent.

In addition, the slices representing multiple anatomical levels may have originated from the same patient, leading to potential correlation within the entire dataset. It will be important for future studies to incorporate patient-level data splitting into their analyses to help eliminate this potential source of bias. Furthermore, given that the dataset was derived from a single facility, it may not be applicable to other populations, imaging devices, or clinical settings. Because of this, an external validation study using data sets from multiple organizations would provide greater evidence of the validity of the proposed methodology.

Finally, while there were good results for both classification and segmentation, the proposed system should be evaluated on larger and more independent datasets, and tested in clinical studies before it can be added to everyday use in a clinical setting.

Conclusion

This research study examines MF using deep learning classification and segmentation applied using transformer and attention models on CBCT image data. The efficacy of each model was confirmed by their ability to accurately distinguish MF from non-MF as well as their ability to accurately define MF boundaries at each individual pixel level.

The model results from this work demonstrate that they can be effectively utilized by clinicians to assist them in the interpretation of CBCT images, thus supporting their decision making when it comes to planning for dental implant procedures, assessing surgical risks, and preventing nerve injury. In addition, methodological success within this study provides the basis for a clinical application of these methodologies.

Future research for this project will utilize multi-center datasets with data collected from different CBCT machines to help develop generalizable models, as well as develop user-friendly software applications that will make these models more user-friendly for practical applications. Continued work along these lines should help to provide enhanced clinical utility for MF detection using AI methods.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data Availability The dataset will be shared upon request.

Declarations

Ethics Approval Not applicable. This study used publicly available datasets and did not involve any experiments with human participants or animals requiring ethical approval.

Consent to Participate Not applicable. The study was based entirely on secondary data obtained from publicly accessible sources.

Consent for Publication All authors have reviewed and approved the final version of the manuscript and consent to its publication.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Beshtawi, K. R., & Qirresh, E. (2025). A rare case of an anatomical variant of nonexistent mental foramen. *Radiology Case Reports*, *20*(2), 1082-1086.
2. Alrahabi, M., & Zafar, M. (2018). Anatomical Variations of Mental Foramen: A Retrospective Cross-Sectional Study. *International Journal of Morphology*, *36*(3).
3. Wei, X., Gu, P., Hao, Y., & Wang, J. (2020). Detection and characterization of anterior loop, accessory mental foramen, and lateral lingual foramen by using cone beam computed tomography. *The Journal of Prosthetic Dentistry*, *124*(3), 365-371.
4. Liu, L., Zhang, X., & Xu, Z. (2024). An adaptive weight search method based on the Grey wolf optimizer algorithm for skin lesion ensemble classification. *International Journal of Imaging Systems and Technology*, 34(2), e23049.
5. Teke, M., Etem, T. Cascading GLCM and T-SNE for detecting tumor on kidney CT images with lightweight machine learning design. *Eur. Phys. J. Spec. Top*. 234, 4619–4634 (2025). https://doi.org/10.1140/epjs/s11734-025-01842-2
6. Reis, H. C., & Turk, V. (2025). DSCIMABNet: A novel multi-head attention depthwise separable CNN model for skin cancer detection. *Pattern Recognition*, 159, 111182.
7. Savaş, S. (2025). Explainable Artificial Intelligence for Diagnosis and Staging of Liver Cirrhosis Using Stacked Ensemble and Multi-Task Learning. *Diagnostics*, 15(9), 1177.
8. Önder, M., Daldal, N., Polat, K., & Dogan, M. U. (2025). Internet of thing-based hand tremor monitoring system and automated detection of hand tremor frequency. *Computers and Electrical Engineering*, 122, 110002.
9. Sivari, E., Bostancı, G. E., Güzel, M. S., Açıcı, K., Aşuroğlu, T., & Erçelebi Ayyıldız, T. (2023). A New Approach for Gastrointestinal Tract Findings Detection and Classification Deep Learning-Based Hybrid Stacking Ensemble Models. *Diagnostics*, 13(4), 720–0.
10. Sivari Resul, E., Senirkentli, G. B., Bostanci, E., & Oduncuoglu, B. F. (2025). APD-FFNet: a novel explainable deep feature fusion network for automated periodontitis diagnosis on dental panoramic radiography. *Dentomaxillofacial Radiology*, twaf034.
11. Jeoun, B. S., Yang, S., Lee, S. J., Kim, T. I., Kim, J. M., Kim, J. E., ... & Yi, W. J. (2022). Canal-Net for automatic and robust 3D segmentation of mandibular canals in CBCT images using a continuity-aware contextual network. *Scientific reports*, *12*(1), 13460.
12. Lv, J., Zhang, L., Xu, J., Li, W., Li, G., & Zhou, H. (2023). Automatic segmentation of mandibular canal using transformer based neural networks. *Frontiers in Bioengineering and Biotechnology*, *11*, 1302524.
13. Yeshua, T., Ladyzhensky, S., Abu-Nasser, A., Abdalla-Aslan, R., Boharon, T., Itzhak-Pur, A., ... & Nadler, C. (2023). Deep learning for detection and 3D segmentation of maxillofacial bone lesions in cone beam CT. *European Radiology*, *33*(11), 7507–7518.
14. Lo Giudice, A., Ronsivalle, V., Spampinato, C., & Leonardi, R. (2021). Fully automatic segmentation of the mandible based on convolutional neural networks (CNNs). *Orthodontics & Craniofacial Research*, *24*, 100-107.
15. Beser, B., Reis, T., Berber, M. N., Topaloglu, E., Gungor, E., Kılıc, M. C., ... & Bayrakdar, I. S. (2024). YOLO-V5 based deep learning approach for tooth detection and segmentation on pediatric panoramic radiographs in mixed dentition. *BMC medical imaging*, *24*(1), 172.
16. Dasanayaka, C., Dharmasena, B., Bandara, W. R., Dissanayake, M. B., & Jayasinghe, R. (2019, December). Segmentation of mental foramen in dental panoramic tomography using deep learning. In *2019 14th conference on industrial and information systems (ICIIS)* (pp. 81–84). IEEE.
17. Kats, L., Vered, M., Blumer, S., & Kats, E. (2020). Neural network detection and segmentation of mental foramen in panoramic imaging. *Journal of Clinical Pediatric Dentistry*, *44*(3), 168-173.
18. Jaskari, J., Sahlsten, J., Järnstedt, J., Mehtonen, H., Karhu, K., Sundqvist, O., ... & Kaski, K. (2020). Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes. *Scientific reports*, *10*(1), 5842.
19. Steybe, D., Poxleitner, P., Metzger, M. C., Brandenburg, L. S., Schmelzeisen, R., Bamberg, F., ... & Russe, M. F. (2022). Automated segmentation of head CT scans for computer-assisted craniomaxillofacial surgery applying a hierarchical patch-based stack of convolutional neural networks. *International journal of computer assisted radiology and surgery*, *17*(11), 2093–2101.
20. Edvardsen, I. P., Teterina, A., Johansen, T., Myhre, J. N., Godtliebsen, F., & Bolstad, N. L. (2022). Automatic detection of the mental foramen for estimating mandibular cortical width in dental panoramic radiographs: the seventh survey of the Tromsø Study (Tromsø7) in 2015–2016. *Journal of International Medical Research*, *50*(11), 03000605221135147.
21. Murugan, A. J., Anuradha, G., Lakshmi, K. C., & Swathi, K. V. (2023). Deep learning-based categorization of mental foramen location using digital panoramic imaging. *Journal of Indian Academy of Oral Medicine and Radiology*, *35*(4), 567-571.
22. Bağ, İ., Bilgir, E., Bayrakdar, İ. Ş., Baydar, O., Atak, F. M., Çelik, Ö., & Orhan, K. (2023). An artificial intelligence study: automatic description of anatomic landmarks on panoramic radiographs in the pediatric population. *BMC Oral Health*, *23*(1), 764.
23. Raza, H., Ali, M., Singh, V. K., Wahjuningrum, A., Sarig, R., & Chaurasia, A. (2024). Segmentation of mental foramen in orthopantomographs: a deep learning approach. *arXiv preprint arXiv:2408.04763*.
24. Çelebi, A., Küçük, D. B., Adaş, G. N., Küçük, A. Ö., Türkoğlu, M., & Atıl, F. (2025). Detection of Mental Foramen in Panoramic Radiographs with Artificial Intelligence-based Systems. *Journal of Health Institutes of Türkiye*.
25. Ünal, S. Y., & Namdar Pekiner, F. (2025). Evaluation of the mandibular canal and the third mandibular molar relationship by CBCT with a deep learning approach. *Oral Radiology*, *41*(2), 222-230.

26. Cipriano, M., Allegretti, S., Bolelli, F., Di Bartolomeo, M., Pollastri, F., Pellacani, A., ... & Grana, C. (2022). Deep segmentation of the mandibular canal: a new 3D annotated dataset of CBCT volumes. *Ieee Access*, 10, 11500–11510.
27. Yuce, F., Buyuk, C., Bilgir, E., Çelik, Ö., & Bayrakdar, İ. Ş. (2025). Deploying a novel deep learning framework for segmentation of specific anatomical structures on cone-beam CT. *Oral Radiology*, 1–9.
28. Lahoud, P., Diels, S., Niclaes, L., Van Aelst, S., Willems, H., Van Gerven, A., ... & Jacobs, R. (2022). Development and validation of a novel artificial intelligence driven tool for accurate mandibular canal segmentation on CBCT. *Journal of dentistry*, 116, 103891.
29. Gaudin, R., Otto, W., Ghanad, I., Kewenig, S., Rendenbach, C., Alevizakos, V., ... & von See, C. (2024). Enhanced osteoporosis detection using artificial intelligence: A deep learning approach to panoramic radiographs with an emphasis on the mental foramen. *Medical Sciences*, 12(3), 49.
30. Nakamoto, T., Taguchi, A., & Kakimoto, N. (2022). Osteoporosis screening support system from panoramic radiographs using deep learning by convolutional neural network. *Dentomaxillofacial Radiology*, 51(6), 20220135.
31. Ovuz, Z., Gur, E., & Etem, T. (2026). Automated identification of accessory mental foramen using cbct and convolutional neural networks. *International Dental Journal*.
32. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).
33. Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. *In International conference on machine learning* (pp. 6105–6114). PMLR.
34. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826).
35. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint* arXiv:1704.04861
36. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
37. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929.
38. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251–1258).
39. Fayyaz, A. M., Abdulkadir, S. J., Talpur, N., Al-Selwi, S. M., Hassan, S. U., & Sumiea, E. H. (2025). Grad-CAM (Gradient-weighted Class Activation Mapping): A systematic literature review. *Computers in Biology and Medicine*, 198, 111200.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.