



# OPEN Automated classification of midpalatal suture maturation stages from CBCTs using an end-to-end deep learning framework

Omid Halimi Milani<sup>1</sup>, Lauren Mills<sup>2</sup>, Amanda Nikho<sup>2</sup>, Marouane Tliba<sup>1</sup>,  
Veerasathpurush Allareddy<sup>2</sup>, Rashid Ansari<sup>1</sup>, Ahmet Enis Cetin<sup>1</sup> & Mohammed H. Elnagar<sup>2,3</sup>✉

Accurate classification of midpalatal suture maturation stages is critical for orthodontic diagnosis, treatment planning, and the assessment of maxillary growth. Cone Beam Computed Tomography (CBCT) imaging offers detailed insights into this craniofacial structure but poses unique challenges for deep learning image recognition model design due to its high dimensionality, noise artifacts, and variability in image quality. To address these challenges, we propose a novel technique that highlights key image features through a simple filtering process to improve image clarity prior to analysis, thereby enhancing the learning process and better aligning with the distribution of the input data domain. Our preprocessing steps include region-of-interest extraction, followed by high-pass and Sobel filtering for emphasis of low-level features. The feature extraction integrates Convolutional Neural Networks (CNN) architectures, such as EfficientNet and ResNet18, alongside our novel Multi-Filter Convolutional Residual Attention Network (MFCRAN) enhanced with Discrete Cosine Transform (DCT) layers. Moreover, to better capture the inherent order within the data classes, we augment the supervised training process with a ranking loss by attending to the relationship within the label domain. Furthermore, to adhere to diagnostic constraints while training the model, we introduce a tailored data augmentation strategy to improve classification accuracy and robustness. In order to validate our method, we employed a k-fold cross-validation protocol on a private dataset comprising 618 CBCT images, annotated into five stages (A, B, C, D, and E) by expert evaluators. The experimental results demonstrate the effectiveness of our proposed approach, achieving the highest classification accuracy of 79.02%, significantly outperforming competing architectures, which achieved accuracies ranging from 71.87 to 78.05%. This work introduces a novel and fully automated framework for midpalatal suture maturation classification, marking a substantial advancement in orthodontic diagnostics and treatment planning.

The development of intelligent systems has revolutionized the potential for automating complex tasks in orthodontic diagnosis and treatment planning. These systems aim to streamline intricate workflows that require the integration of objective data and subjective clinical expertise. Orthodontic diagnostics typically involve gathering information from patient interviews, clinical examinations, and the analysis of various diagnostic records, such as dental casts, radiographs, and photographs. Among these, cone-beam computed tomography (CBCT) has emerged as an essential imaging modality<sup>1</sup>, providing high-resolution, three-dimensional (3D) reconstructions of craniofacial structures. CBCT has proven indispensable across a wide range of orthodontic applications, including 3D diagnosis, virtual treatment planning, localization of impacted and supernumerary teeth, bone thickness assessment, root resorption analysis, and temporomandibular joint evaluation<sup>2</sup>.

Recent advances in CBCT technology and image-enhancing software allow CBCT radiation dose to be reduced<sup>1</sup>. The American Academy of Oral and Maxillofacial Radiology panel concluded that the use of CBCT in orthodontic treatment should be justified on an individual basis based on clinical presentation. This statement provides general recommendations, specific use selection recommendations, optimization protocols, radiation-dose, and risk-assessment strategies for CBCT imaging in orthodontics (including diagnosis, treatment, and

<sup>1</sup>Department of Electrical and Computer Engineering, University of Illinois Chicago, Chicago, IL, USA. <sup>2</sup>Department of Orthodontics (M/C 841), College of Dentistry, University of Illinois Chicago, 801 S. Paulina Street, RM 131, Chicago, IL 60612-7211, USA. <sup>3</sup>Department of Orthodontics, Faculty of Dentistry, Tanta University, Tanta, Egypt. ✉email: melnagar@uic.edu

assessment of outcomes)<sup>1</sup>. Furthermore, with the transformation into fully digital workflows and introduction of CAD/CAM orthodontic appliances, the use of CBCT in orthodontic practice is expected to grow. Many appliance design software (e.g., Invisalign, Spark, Suresmaile, Brius, Dolphin 3D) now allow the integration of the CBCT with intraoral scans for appliance design and staging. When required, CBCT should be performed using the smallest possible field of view needed for the specific clinical scenario<sup>3</sup>. If the patient is exposed to CBCT radiation, the patient should not be exposed to routine 2D radiography.

The estimate of a patient's skeletal maturity is crucial in determining the optimal treatment timing<sup>4</sup> (e.g., for planning the appropriate time for use of rapid maxillary expansion, functional appliances, orthognathic surgery, and implants). Several methods have been used to determine skeletal age. Hand-wrist radiographs have been proposed to estimate a patient's bone maturation stage with relatively high accuracy. However, it has several drawbacks, including the additional amount of radiation required<sup>5</sup>. To address these limitations, alternate methods have been developed to assess the patient's skeletal maturation, including the Midpalatal Suture Maturation Stages (MPSMS). The midpalatal suture has been described as an end-to-end type of suture, consisting of the palatine bone posteriorly and the maxillary bone anteriorly. This suture has characteristic changes in its morphology during growth, becoming more interdigitated, then fusing from the posterior aspect to the anterior.

The anatomy and development of the midpalatal suture is an important consideration for orthodontists when treatment planning for maxillary transverse discrepancy. Maxillary transverse deficiency is a type of craniofacial malformation that can be caused by many factors, including mouth breathing, digit sucking or pacifier habits, or atypical oral motor patterns during speaking and swallowing<sup>6</sup>. Over 20% of the global population exhibits this malformation<sup>7</sup>. Different treatment options are available for the treatment of transverse deficiency and include use of tooth-borne Rapid Maxillary Expansion (RME), micro-implant-assisted rapid palatal expansion (MARPE), and surgically-assisted rapid palatal expansion (SARPE)<sup>8,9</sup>. All treatment modalities involve widening the maxilla by separating the midpalatal suture, and technique varies depending on appliance selection, activation method, and invasiveness<sup>6</sup>. It is well-documented in the literature that attempts to widen the maxilla at an inappropriate time in the patients growth can cause unnecessary trauma, as well as increased side effects such as external root resorption, dehiscence or fenestration of buccal cortical bone, dental tipping, gingival inflammation, gingival recession, relapse, and unsuccessful expansion<sup>10–12</sup>. To avoid adverse side effects, a practitioner must choose the appropriate treatment protocol based on the patency of the suture. RME is a simple procedure that is commonly rendered in growing children. When growth begins to cease, RME becomes less successful. Individuals past their pubertal growth spurt with unfused midpalatal sutures may exhibit successful expansion using MARPE, and SARPE is indicated when the patient has a suture that is entirely fused.

However, the start and the advancement of fusion of the midpalatal suture vary greatly with age and sex. Persson and Thilander observed fusion of the midpalatal suture in subjects ranging from 15 to 19 years old<sup>13</sup>. However, other histological and radiographic studies showed that patients at ages 27, 32, 54, and even 71 years have no signs of suture fusion. Such findings indicate that variability in the developmental stages of fusion of the midpalatal suture is not related directly to chronological age, particularly in young adults<sup>13–15</sup>. When treatment planning for transverse discrepancy, this variability can make it difficult to determine whether to utilize a simple approach, such as RME, or to use a more invasive approach which may require surgery.

Several methods have been proposed to assess midpalatal suture maturation and ossification status, including histological methods and imaging methods. The histological result is the golden standard but is not practical for in vivo routine clinical examination<sup>14</sup>. Occlusal radiography was used by Revelo and Fishman in some studies; however, occlusal radiographs are not reliable for analyzing midpalatal suture morphology because the vomer and structures of the external nose overlay the midpalatal area, potentially leading to false radiographic interpretations of suture fusion<sup>16</sup>. Because of this superimposition of adjacent structures, occlusal radiographs were eventually replaced by CT<sup>17</sup>, which provides a more accurate visualization of the midpalatal suture. In 2013, Angelieri et al. introduced a MPSMS classification method using CBCT. Based on morphological features of the midpalatal suture in one axial section, five maturation stages were classified: A, B, C, D, and E<sup>18</sup>.

The classification approach utilized by Angelieri et al. evaluates a single axial section, which is interpreted by a human practitioner. Classification of the developmental stage has low simplicity and is prone to subjectivity among evaluators<sup>19</sup>. Accurate appraisal of the development of the midpalatal suture is also affected by the experience of the examiner. Evaluators who have a high experience level (e.g., orthodontists) are more likely to classify the developmental stage correctly. However, even within groups of experienced evaluators, inter-rater reliability is weak and accurate evaluation cannot be relied upon<sup>20</sup>. Reliable midpalatal suture classification is important to ensure accurate diagnosis, low error, and consistent treatment outcomes. Given the high amount of variability between different evaluators, the current method is not qualified enough to be used routinely in clinical settings. For this reason, it is necessary to seek a methodological improvement that results in accurate and consistent outcomes<sup>19</sup>.

Consequent to recent advances in computing power and availability of big data, Artificial Intelligence (AI) has evolved rapidly to become part of daily life. Machine Learning (ML) is a subset of AI that refers to the computer's ability to learn without being explicitly programmed. Deep Learning (DL) is the most recent subset of Machine Learning, with networks capable of learning unsupervised from unstructured data. The development of automated image analysis has attracted significant interest due to its potential to reduce human errors and biases, as well as to save time and effort in diagnosis. The application of DL to study human growth and development from radiographs is a promising idea that has been applied to evaluation of CBCT images. Using AI to classify MPSMS could systematize the process, reduce error with clinicians, and enhance the treatment planning process. While Deep Learning (DL) has become the leading tool for image recognition and classification in computer vision applications<sup>21–23</sup>, applying these models to medical imaging data like CBCT scans presents unique challenges not encountered with natural scene images, for which these models were originally designed.

Medical images often exhibit high dimensionality, increased noise, and subtle contrasts. Unlike the high-level semantic features (objects, scenes) found in natural images, the diagnostically relevant features in medical images are typically low-level and semantically simple, such as slight variations in tissue density or small structural anomalies. Consequently, standard image recognition models may not effectively capture these critical features. This necessitates specialized model design adaptation and preprocessing techniques focused on enhancing low-level feature extraction, thereby aligning the learning process with the essential diagnostic information in medical imaging.

The key contributions of this paper are as follows:

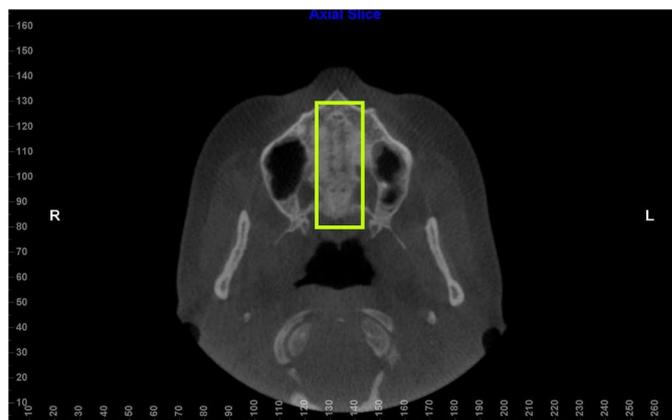
- **Development of a multi-stream deep-based architecture** tailored to classifying Midpalatal Suture Maturation Stages, with the aim of integrating an automated classification system that operates without manual intervention.
- **Proposal of a combined filtering approach** that enhances different parts of the input images based on expert feature judgment. This involves using a combination of filters to make diagnostically relevant features more prominent.
- **Implementation of noise removal on the latent representation** obtained from parallel streams with differently filtered inputs. This involves establishing a latent feature space in the Discrete Cosine Transform (DCT) domain.

## Materials and methods

### Preparing the data set

In this research, we conducted a retrospective study utilizing archived data. All data used in the study was deidentified to ensure confidentiality. The University of Illinois Chicago's Office for the Protection of Research Subjects (OPRS) Institutional Review Board (IRB) granted an exemption from the requirement for informed consent, providing IRB-exempt status to the study with the assigned Study ID: (STUDY2022-1048), all methods were performed by the relevant guidelines and regulations of the University of Illinois Chicago's OPRS and IRP. CBCT scans were collected according to the following inclusion criteria: patient assuming a natural head position with Frankfort horizontal plane horizontal to the ground, jaws immobilized using a chin holder, teeth occluded in intercuspal position with facial muscles relaxed, and utilization of extended field of view (FOV). The exclusion criteria included syndromes affecting the craniofacial and cervical column structures, history of trauma to the head and neck, and previous history of head and neck surgery.

On Dolphin Imaging Software (DolphinImaging Version 12, [www.dolphinimaging.com](http://www.dolphinimaging.com)), CBCT DICOM files were oriented to the following reference planes: sagittal-palatal plane (ANS-PNS), coronal-most inferior points of the nasal floor, and axial-mid-palatal suture (MPS). Our project's region of interest (RoI)—the MPS—can be best visualized from the axial plane, rendering the other two planes as nonessential for staging. Consequently, one axial slice at the palatal plane was extracted for each CBCT. The axial section slices were labeled and classified according to MPSMS using the method proposed by Angelieri et al. (Stages A, B, C, D, E). In Stage A, the midpalatal suture can be visualized as a straight, high-density line with little or no interdigitation. In Stage B, the suture becomes a scalloped high-density line. Here, there may be some areas with two scalloped, high-density lines running parallel to each other, separated by small low-density spaces. In Stage C, the midpalatal suture appears as two parallel, high-density lines, which run close to each other. These lines are present in both the maxillary and palatine bones. In Stage D, the midpalatal suture cannot be visualized in the palatine bone, as fusion has occurred in this area. In the maxillary bone, the suture can still be visualized as two parallel high-density lines. In Stage E, the midpalatal suture has fused in both the palatine and maxillary bones. The parasutural bone density is increased and the suture can no longer be visualized<sup>18</sup>. Figure 1 shows an axial slice image sample from our data set with MPS located in the center. Depicted in Fig. 1, the midpalatal suture is accurately cropped and resized to a standardized dimension of 210x70.



**Figure 1.** Axial CBCT slice showing the midpalatal suture (MPS) region, which has been cropped and standardized for analysis. This region serves as the input to our classification model.

Labeling and classification were performed by trained evaluators, including two orthodontists and an oral and maxillofacial radiologist, after reviewing the classification method<sup>18</sup>. To investigate inter-rater reliability, a crosstabulation was computed. The Kendall's tau-b statistic assesses the strength and direction of the association between ordinal ranked variables. It was between 0.877 and 0.986, indicating a strong significant positive association between the investigators, suggesting high inter-rater reliability. This is important for ensuring the validity and accuracy of the research findings.

The initial data set consisted of 1200 CBCT scans (592 female, 447 male, 161 unknown). Age distribution of the dataset was as follows: 221 children ages 6 to 10, 525 adolescents ages 11 to 16, 60 young adults ages 17 to 21, 230 adults ages 22 or older, and 164 scans with an unknown age demographic. MPSMS for each CBCT scan was classified, resulting in the following distribution: 12.4% stage A, 23.8% stage B, 26.3% stage C, 17.3% stage D, and 20.2% stage E. The 1200 CBCTs were then curated by the primary evaluators to exclude images of ambiguous diagnostic quality. The final data set comprises 618 carefully classified images, containing a MPSMS distribution of 13.9% stage A, 22.8% stage B, 25.7% stage C, 16.2% stage D, and 21.4% stage E.

One important consideration regarding the data set is the potential imbalance in sample distribution across different classes. To address this issue, effective strategies can be employed, including data augmentation. Data augmentation is a widely used technique that helps to overcome model overfitting by generating additional samples during each iteration of model training. In this study, we applied random affine transformations to the input images with the following parameters<sup>24</sup>. These techniques introduce subtle variations to the existing samples, effectively expanding the data set.

Through the use of data augmentation and a weighted data sampler, our model was better prepared to handle the distribution of data typical in medical imaging scenarios.

### Deep neural networks used in this study

Convolutional Neural Networks (CNNs) are a class of artificial intelligence models capable of learning and recognizing complex patterns in images. In this study, we leverage CNNs to automatically identify the stage of midpalatal suture maturation using selected slices from CBCT scans. It showcases proficiency in sorting images into categories, establishing itself as a cornerstone tool across different computer vision applications<sup>25–28</sup>. The fundamental concept behind CNNs is their capability to utilize two convolution operations, enabling them to uncover insights from images. These convolutional layers help the model identify important patterns in the images, such as the shape or density changes in the midpalatal suture.

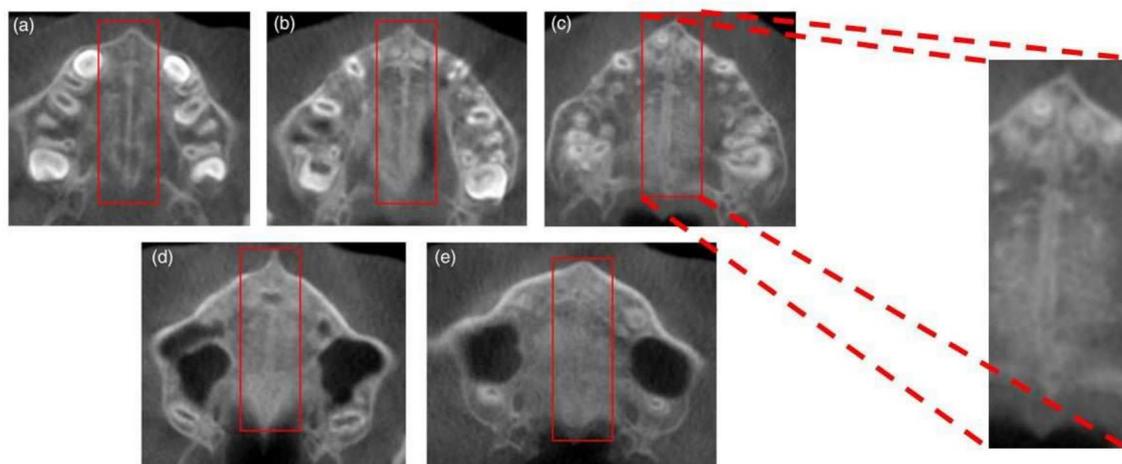
Recent improvements in image processing models—especially those used in medical imaging—have made it possible to more accurately interpret X-ray scans using AI-based methods like CNNs<sup>29</sup>. Out of these structures, ResNet is especially effective at tackling the issue of vanishing gradients in deep CNNs. With the addition of connections, data moves smoothly through the network, making ResNet capable of being more intricate and expressive, while achieving better results in image classification. Another noteworthy advancement in the field of CNN designs is MobileNet<sup>30</sup>. MobileNet models are designed to be light and efficient, making them a great choice for use on devices with limited processing capabilities, such as smartphones and embedded systems. They use depthwise convolutions to decrease the number of parameters and computations compared to traditional convolutions, leading to quicker inference times and optimal performance for real-time applications. Similarly, ShuffleNet is a network specifically created for mobile devices with restricted computing power. This design incorporates two operations: pointwise group convolution and channel shuffle, to lower computational expenses while maintaining accuracy.

EfficientNet is a CNN design that merges the attributes of ResNet and MobileNet while tackling their shortcomings<sup>31</sup>. A new method called compound scaling has been introduced, which optimizes the network's depth, width, and resolution while simultaneously using a scaling factor. This approach allows users to strike a balance between model size and accuracy, making EfficientNet a versatile and effective choice for image classification tasks. The network design combines depthwise convolutions from MobileNet and residual connections from ResNet, leading to top-notch performance on datasets with enhanced computational efficiency. In this study, we focused on using EfficientNet B0 as our base model while also comparing it with other commonly used CNN architectures.

The model pipeline employed in our study is illustrated in Fig. 2. As a first step, we extracted the RoI and removed any irrelevant parts from the CBCT images. By focusing only on the central region, we ensure that the model processes and classifies the most relevant information.

After extracting the RoI, the images underwent a process called data augmentation. This method involved creating samples by applying transformations such as rotation, auto contrast, and translation to enhance the diversity of the training data. Data augmentation is crucial for reducing overfitting and enhancing the model's capacity to generalize to new images. The modified images were then input into the model for classification. The model generated five results, indicating the probabilities of the input image belonging to each class. These probabilities reflected how confidently the model assigned each image to a category.

In the field of applications, tasks involving categorization often benefit from machine learning solutions that are driven by data. In these cases, datasets are carefully put together through the analysis of medical images. While simpler algorithms can sometimes produce acceptable outcomes, the superiority of models based on CNNs are clearly seen in terms of performance and their ability to decrease the need for manual intervention. These CNN models, which could automatically grasp features from the data, offer an alternative, especially in the medical sector where accuracy and efficiency are crucial. By utilizing CNNs, the precision and dependability of categorizations is improved while significantly reducing the work required. Cross entropy loss, which is commonly used for classification tasks, measures the performance of a classification model whose output is a probability value between 0 and 1. It is particularly useful in evaluating the accuracy of the model's predictions against the actual labels (targets). The cross entropy loss function effectively penalizes incorrect classifications,



**Figure 2.** Overview of the image preparation pipeline with stages A–E. The RoI is extracted from the axial CBCT slice.

Filter no.	Filter description	Filter function	Application
Filter 1	High-pass Filter (Base)	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	Enhances edges and fine details in the image.
Filter 3	High-pass Filter (Sensitive)	$\begin{bmatrix} -2 & -2 & -2 \\ -2 & 16 & -2 \\ -2 & -2 & -2 \end{bmatrix}$	Captures finer variations and subtle boundaries.
Filter 5	High-pass Filter (More Sensitive)	$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 24 & -3 \\ -3 & -3 & -3 \end{bmatrix}$	Detects very subtle changes in anatomical structures.
Filter 2	Sobel Filter (Base Sensitivity)	$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$	Highlights vertical edges and boundaries.
Filter 4	Sobel Filter (Sensitive)	$\begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix}$	Enhances detection of finer vertical edges.
Filter 6	Sobel Filter (More Sensitive)	$\begin{bmatrix} -5 & 0 & 5 \\ -20 & 0 & 20 \\ -5 & 0 & 5 \end{bmatrix}$	Captures very subtle vertical anatomical boundaries.

**Table 1.** Summary of image filters used for preprocessing to enhance the model's ability to detect subtle anatomical differences.

with higher penalties for predictions that are both incorrect and confident. This makes it a robust measure for training models in multi-stage classification problems.

Lastly, cross entropy loss has become a standard choice in various machine learning applications due to its effectiveness in differentiating between classes and its straightforward implementation. It is particularly valuable in deep learning, where it serves as a critical component in training neural networks for tasks such as image recognition, natural language processing, and more.

### Applied pre-filters

Before feeding the images into the AI model, we applied filters that highlight important visual features that human experts also rely on when reviewing images, such as boundary edges and structural changes in the surrounding bone. They are represented by specific 3x3 matrices and include high-pass filters and Sobel filters. These steps enhance the model's ability to detect subtle anatomical patterns that are critical for stage classification<sup>32</sup>.

By combining the outputs of these filtered images with the original 2D images, the input data for the network was enriched. This approach enabled the model to better recognize patterns and key details, improving its performance in tasks such as object recognition and classification. A total of six filters were used, including high-pass filters and Sobel filters, each capturing different aspects of the image's details (see Table 1). This variety was crucial, as it allowed the model to capture information across different orientations and frequency ranges. These filters mimic how a human examiner focuses on the edges or shape of the suture in an image and help the model focus on the most informative parts of the images.

The Residual Block is a fundamental building block of the network, composed of two convolutional layers with batch normalization and ReLU activation. The first convolutional layer in each residual block is designed to expand the number of output channels (64, 128, 256, and 512 channels in successive blocks), allowing the model to capture increasingly complex features. The second convolutional layer maintains the same number of input and output channels, ensuring the integrity of the learned features while applying further transformations. Additionally, a skip connection is included for residual learning, which helps improve gradient flow and convergence. This architecture not only enhances the model's ability to learn from deeper layers but also mitigates the vanishing gradient problem by providing an alternate path for the gradient to flow through.

### Pre-trained based models

In our research, we utilized several pre-trained models to evaluate their performance on the image classification task. Among these models were MobileNet, ResNet, and ShuffleNet. MobileNet, designed for efficiency, is particularly suitable for deployment on devices with limited processing power. It achieves this by employing depthwise separable convolutions, which significantly reduce the number of parameters and computational load compared to traditional convolutional layers. ShuffleNet, another model tailored for mobile devices, incorporates pointwise group convolution and channel shuffle operations. These techniques further minimize computational expenses while maintaining competitive accuracy. ResNet, on the other hand, addresses the challenge of vanishing gradients in deep networks through the use of residual connections, enabling the construction of very deep networks that can still be trained effectively.

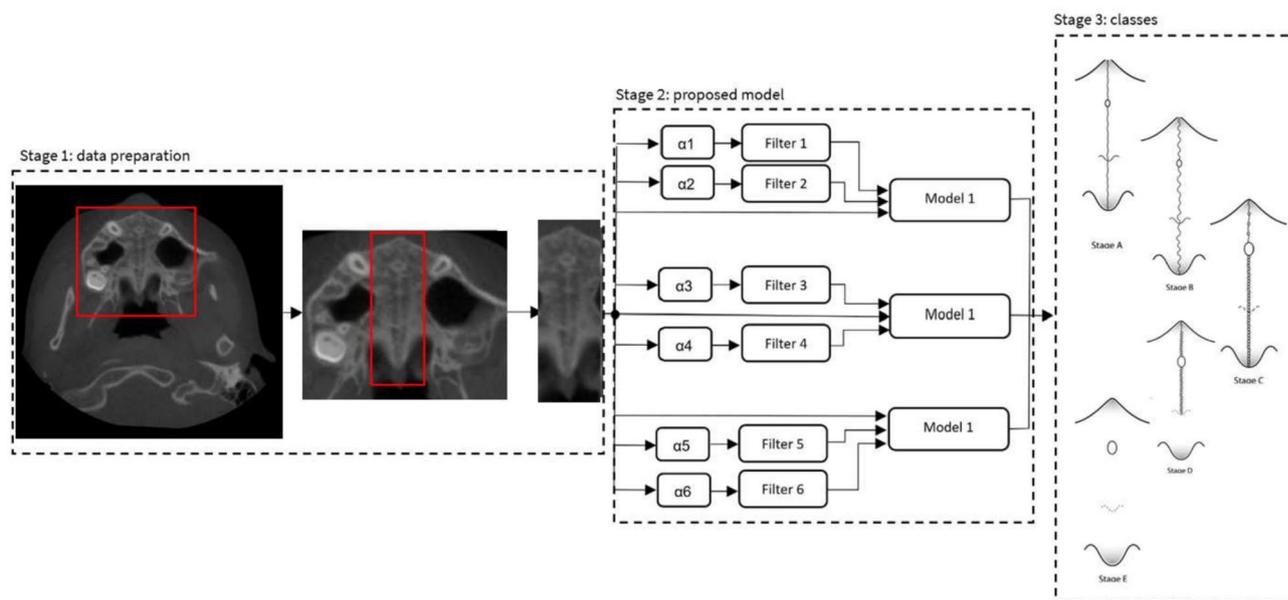
Among the pre-trained models evaluated, ResNet and EfficientNet demonstrated superior performance. ShuffleNet and MobileNet achieved lower accuracy rates of 56.96% and 58.22%, respectively. Given these results, we decided to further explore models based on EfficientNet and ResNet18, focusing on improving the architecture to enhance performance. Our proposed architecture integrates parallel structures, as illustrated in the accompanying figure, to leverage the strengths of both ResNet and EfficientNet. This hybrid approach aims to combine the depth and residual capabilities of ResNet with the efficient scaling properties of EfficientNet, potentially leading to even higher accuracy and robustness in image classification tasks.

The CBCT image processing pipeline consists of three stages (see Fig. 3). First, a relevant slice of the CBCT image is selected and cropped to focus on the desired RoI. Next, a pretrained model (referred to as “Model 1”) classifies the cropped RoI. Six variables ( $\alpha_1$  to  $\alpha_6$ ) are used to dynamically update the filters, enhancing feature extraction and allowing the model to adapt to the specific characteristics of the input image. Finally, the model predicts probabilities for predefined classes, and the class with the highest probability is selected as the target class.

### Introduction to DCT layer

In this section, we delve into the design and implementation of the Discrete Cosine Transform (DCT) layer within our neural network model. This layer is pivotal in enhancing the model's ability to accurately classify the MPS class by leveraging frequency-specific analysis and filtering.

The DCT layer is integral to our neural network's architecture, serving as a powerful tool for transforming input data into the frequency domain. This transformation facilitates the identification and extraction of significant



**Figure 3.** End-to-end pipeline for processing CBCT images. The workflow includes extracting the midpalatal suture region, applying pretrained models with adaptive filters, and classifying the image into one of the maturation stages (A–E).

frequency components, which are crucial for analyzing and classifying MPS. The DCT layer transforms the image into a format that makes it easier for the model to detect subtle changes in texture and bone fusion.

For a given input time sequence  $x = [x_0, x_1, \dots, x_N]$ , the DCT representation  $X = [X_0, X_1, \dots, X_N]$  is defined as follows:

$$X[0] = \sqrt{\frac{1}{N}} \sum_{n=0}^{N-1} x[n] \quad (1)$$

$$X[k] = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x[n] \cos\left(\frac{(2n+1)k\pi}{2N}\right), \quad \text{for } k = 1, 2, \dots, N-1 \quad (2)$$

### Soft-thresholding nonlinearity

To further refine the representation and suppress noise, we incorporated a trainable soft-thresholding function within the DCT layer. Unlike hard thresholding, which sets values below a certain threshold to zero, soft thresholding smoothly reduces the magnitude of values around the threshold. This approach maintains more information and provides a smoother transition, potentially leading to better generalization. The soft-thresholding function  $F_T(x)$  is defined as:

$$F_T(x) = \text{sign}(x) \cdot \text{ReLU}(|x| - T) \quad (3)$$

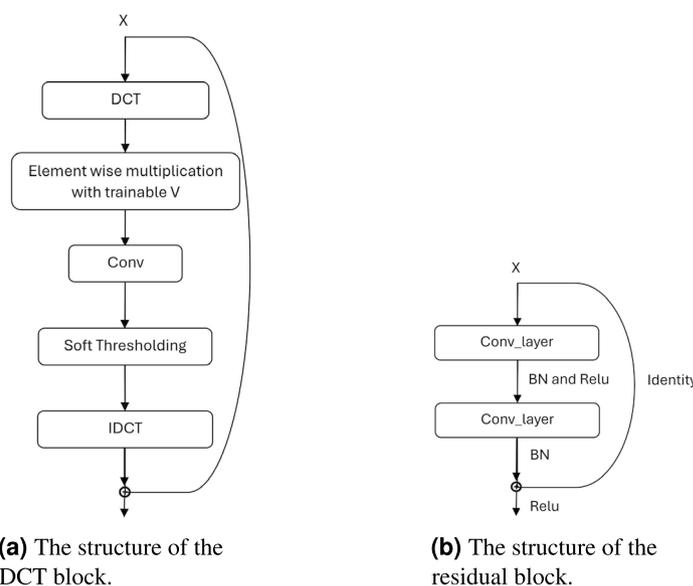
Here,  $T$  is a trainable threshold parameter optimized during training via backpropagation. After applying the soft-thresholding function, the filtered frequency components  $\tilde{X}$  are transformed back into the time domain using the inverse DCT:

$$\tilde{x}_n = \sqrt{\frac{1}{N}} \tilde{X}[0] + \sqrt{\frac{2}{N}} \sum_{k=1}^{N-1} \tilde{X}[k] \cos\left(\frac{(2n+1)k\pi}{2N}\right), \quad \text{for } n = 0, 1, \dots, N-1 \quad (4)$$

Figure 4 illustrates how the DCT layer is seamlessly integrated into the neural network alongside the residual block structure. By strategically placing the DCT layer, the network gains a powerful ability to analyze frequency-specific features, leading to more precise and reliable predictions. This enhancement is particularly valuable for MPS classification, where capturing subtle patterns makes a significant difference. Experimental results clearly show that models incorporating the DCT layer outperform traditional approaches, demonstrating its effectiveness in improving accuracy and robustness.

### Multi-filter convolutional residual attention network (MFCRAN)

This paper presents a novel neural network architecture combining convolutional, attention, and DCT layers for image classification tasks. The proposed model leverages the strengths of various layers and mechanisms to



**Figure 4.** (a) The Discrete Cosine Transform (DCT) block enhances the model's ability to identify frequency-based patterns. (b) The residual block helps retain important features while allowing deeper learning by enabling shortcut connections.

enhance feature extraction, denoising, and classification accuracy. Experimental results demonstrate the efficacy of the architecture.

As shown in Fig. 5, the proposed model incorporates additional architectural enhancements for improved classification. The model uses six variables ( $\alpha_1$  to  $\alpha_6$ ) to dynamically update filters for feature extraction. These filtered features are then processed through a convolutional layer, followed by multiple residual blocks that preserve key features while reducing redundant information.

A multi-head attention layer is added to capture long-range dependencies, allowing the model to focus on critical regions within the input. Finally, the features are passed through a DCT and a fully connected layer to generate the probabilities for the predefined classes. This updated design improves the robustness and accuracy of the classification process.

In our study, we delved deeper into exploring the utilization of DCT layers, demonstrating enhancements in model performance. We designed our experiments to maintain a consistent weight count across different setups. The combination of DCT layers consistently outperformed the traditional Multilayer Perceptron (MLP) arrangement across datasets and tasks. Enhancements were evident in metrics like accuracy and confusion matrix outcomes, indicating that our approach excels in capturing data patterns and interactions. The fusion of DCT layers enhances feature extraction capabilities significantly.

### Modified margin-based negative logarithmic loss (MNRL)

To improve accuracy, we introduced **Modified Margin-based Negative Logarithmic Loss (MNRL)**. MNRL is a loss function that penalizes incorrect predictions based on the discrepancy between the model's predicted classification and the true classification. Determination of this discrepancy is achieved by incorporating a distance matrix between the two classifications.

The model produces a set of logits, representing the raw outputs across five classes before activation. For any given data point  $i$ , let the logits vector  $z_i = [z_{i1}, z_{i2}, \dots, z_{i5}]$  denote these scores for each class. The correct label for each data point  $i$  is given by  $y_i \in \{1, 2, \dots, 5\}$ .

### Distance matrix

The MNRL loss leverages a predefined distance matrix  $D$  of dimension  $5 \times 5$ , where  $D_{ab}$  represents the distance between class  $a$  and class  $b$ . This matrix encodes the relationships between classes, with smaller distances for closely related classes and larger distances for more distinct classes. The specific structures and applications of the distance matrices used are detailed in Table 2.

### Softmax activation

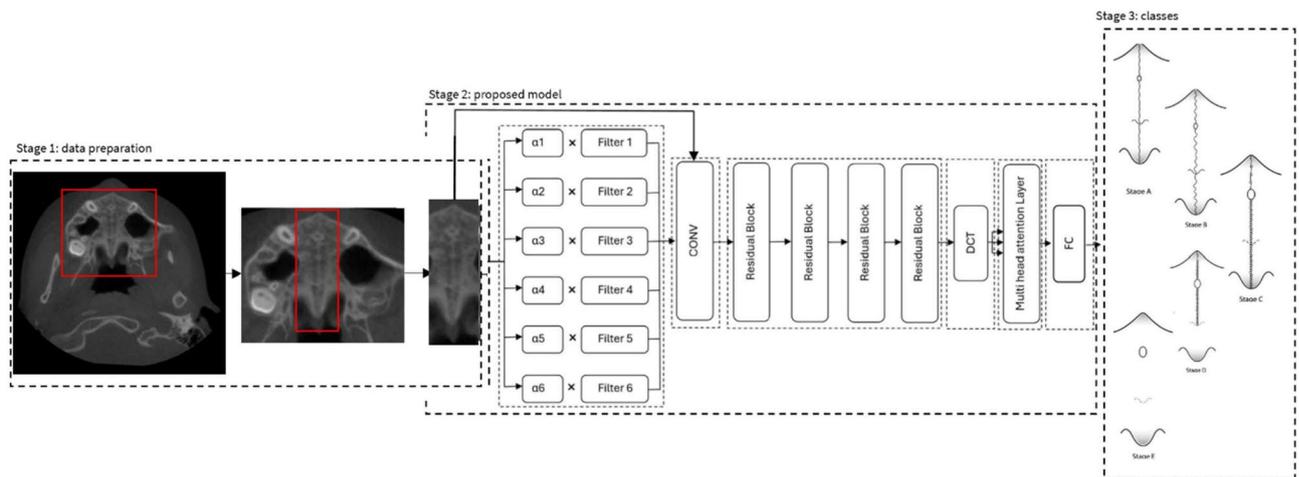
To obtain probabilities  $p_i$  for each class prediction, the logits are converted using the softmax function:

$$p_{ij} = \frac{e^{z_{ij}}}{\sum_{k=1}^5 e^{z_{ik}}} \quad (5)$$

where  $p_{ij}$  represents the probability assigned to class  $j$  for data point  $i$ .

### Positive loss component

The positive loss component  $\mathcal{L}_{\text{pos}}$  focuses on the probability assigned to the correct class  $y_i$ . It is defined as:



**Figure 5.** Architecture of the proposed model with dynamic filters, residual blocks, multi-head attention, and DCT for enhanced feature extraction and classification. The model learns to associate image patterns with maturity stages.

Matrix	Distance matrix	Application
Linear Distance Matrix	$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{bmatrix}$	Enforces a linear penalty structure, assigning greater penalties as class distances increase, encouraging the model to avoid distant misclassifications.
Refined Proximity Penalty Matrix	$\begin{bmatrix} 0 & 2 & 1 & 0.5 & 0.2 \\ 2 & 0 & 2 & 1 & 0.5 \\ 1 & 2 & 0 & 2 & 1 \\ 0.5 & 1 & 2 & 0 & 2 \\ 0.2 & 0.5 & 1 & 2 & 0 \end{bmatrix}$	Applies higher penalties for closer class misclassifications and lower penalties for more distant ones, refining the model's ability to distinguish closely related classes.

**Table 2.** Different distance matrices used in the customized loss function. These matrices assign penalties for misclassifications based on the discrepancy between the predicted stage and the correct one, improving model learning.

$$\mathcal{L}_{\text{pos}} = - \sum_i \log p_{iy_i} \quad (6)$$

This term encourages the model to assign high probabilities to the correct class for each data point.

### Negative loss component

The negative loss component  $\mathcal{L}_{\text{neg}}$  penalizes the model for assigning high probabilities to incorrect classes, especially those that are distant from the correct class. For each incorrect class  $j \neq y_i$ , the negative probability is weighted by the distance between the predicted class and the true class:

$$\mathcal{L}_{\text{neg}} = \sum_i \sum_{j \neq y_i} D_{y_i j} \cdot p_{ij} \quad (7)$$

where  $D_{y_i j}$  is the distance between the true class  $y_i$  and the incorrect class  $j$ .

### Overall loss calculation

The total loss  $\mathcal{L}_{\text{MNRL}}$  for a batch of data points is the sum of the positive and negative loss components:

$$\mathcal{L}_{\text{MNRL}} = \frac{1}{N} \sum_{i=1}^N \left( -\log p_{iy_i} + \sum_{j \neq y_i} D_{y_i j} \cdot p_{ij} \right) \quad (8)$$

where  $N$  is the number of data points in the batch. This loss function balances the emphasis on correct classification confidence with penalties for incorrect predictions based on their distance from the true class.

### Evaluation of different distance matrices

Various distance matrices were tested to assess the effect of class relationships on model performance. The Linear Distance Matrix introduces a straightforward penalty structure based on increasing distance, while the Refined Proximity Penalty Matrix was applied to the fully trained model to further enhance results by imposing more significant penalties on closer classes while decreasing penalties for distant classes. This fine-grained penalty adjustment helps the model differentiate better between closely related classes.

### Results

We implemented a k-fold cross-validation technique to validate the effectiveness of our model. This method involves dividing the dataset into  $k$  subsets (folds), with each fold serving as both a training and testing set in rotation. By repeating this process  $k$  times, we obtain a comprehensive evaluation of our proposed methodology on various data subsets. K-fold cross-validation ensures the model's generalizability to diverse datasets and assesses whether the model learns contextual information or merely memorizes specific examples. Unlike traditional training and testing data separation, k-fold cross-validation dynamically determines training and testing sets for each fold, enhancing the robustness of the evaluation. We adopted  $k = 5$  for our experiments.

In the 5-fold cross-validation, we utilized a dataset comprising 618 high-quality images. The split ratio of 4:1 ensures non-overlap between the training and validation folds. Images are divided into the respective folds before augmentation, which is then applied independently. This approach guarantees that each training and validation fold remains distinct throughout the training process. The iterative selection of model components shapes the final structure of our model, ensuring its capability to accurately classify the midpalatal suture across diverse datasets.

This method helps determine if the model comprehended the image context or merely memorized instances. The augmentation process was then applied to the training set to increase the diversity of the input data and improve the robustness of the model. This method ensures that the training and validation sets remained distinct throughout the training phase.

We computed the accuracy metric and summarized the outcomes in a table to analyze Convolutional Neural Network (CNN) models. To optimize model performance, we systematically explored a range of hyperparameters, including learning rates [0.005, 0.001, 0.0005], loss functions, and optimizers (Adam and AdamW). The best performance was obtained using a learning rate of 0.0005 with the AdamW optimizer ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 1e-08$ , weight decay = 0.01). A `MultiStepLR` scheduler was applied with milestones at epochs 50, 70, and 90 and decay factor of  $\gamma = 0.1$ , enabling dynamic learning rate adjustment. Hyperparameter selection was guided by validation accuracy across folds. All reported results are based on this configuration.

Table 3 compares the performance of various machine learning models under specific conditions employing the Focal loss function<sup>33</sup>. Each row represents a different model: `Parallel_EfficientNet_b0`, `Parallel_ResNet18`, `MFCRAN_FC`, `MFCRAN_DCT_FC`, and `MFCRAN_DCT`. The table presents the performance of each model with pre-filtering. Additionally, it includes baseline models such as `ResNet18`, `ResNet34`, `ResNet50`, `EfficientNet_b0`, and `MobileNetV3` to provide a broader perspective on performance.

`MFCRAN_DCT_FC_MNRL` achieved the highest classification performance with an accuracy of 79.02%, outperforming all baselines and variations. Among the baselines, `ResNet18` and `ResNet34` showed identical accuracy (73.33%); however, `ResNet34` requires nearly double the number of parameters. In contrast, deeper `ResNet50` and lightweight `MobileNetV3` underperformed, likely due to overfitting and limited feature capacity, respectively. The inclusion of parallel filtering improved the performance of `ResNet18` and `EfficientNet_b0`, while `MFCRAN` models further benefited from architectural enhancements, including DCT and the modified MNRL loss. These enhancements consistently improved classification metrics across accuracy, precision, recall, and F1-score, underlining the impact of each component. Overall, the results support the effectiveness of our modular design in improving classification performance with only a modest increase in complexity.

Throughout the experiments conducted in this study, `EfficientNet`—with pre-filtering and an average accuracy—of 71.87% demonstrates reliable and robust performance on the test data. `ResNet18` consistently achieved an average of around 73.33% accuracy, indicating that the task is well-suited for conventional convolutional neural networks (CNN).

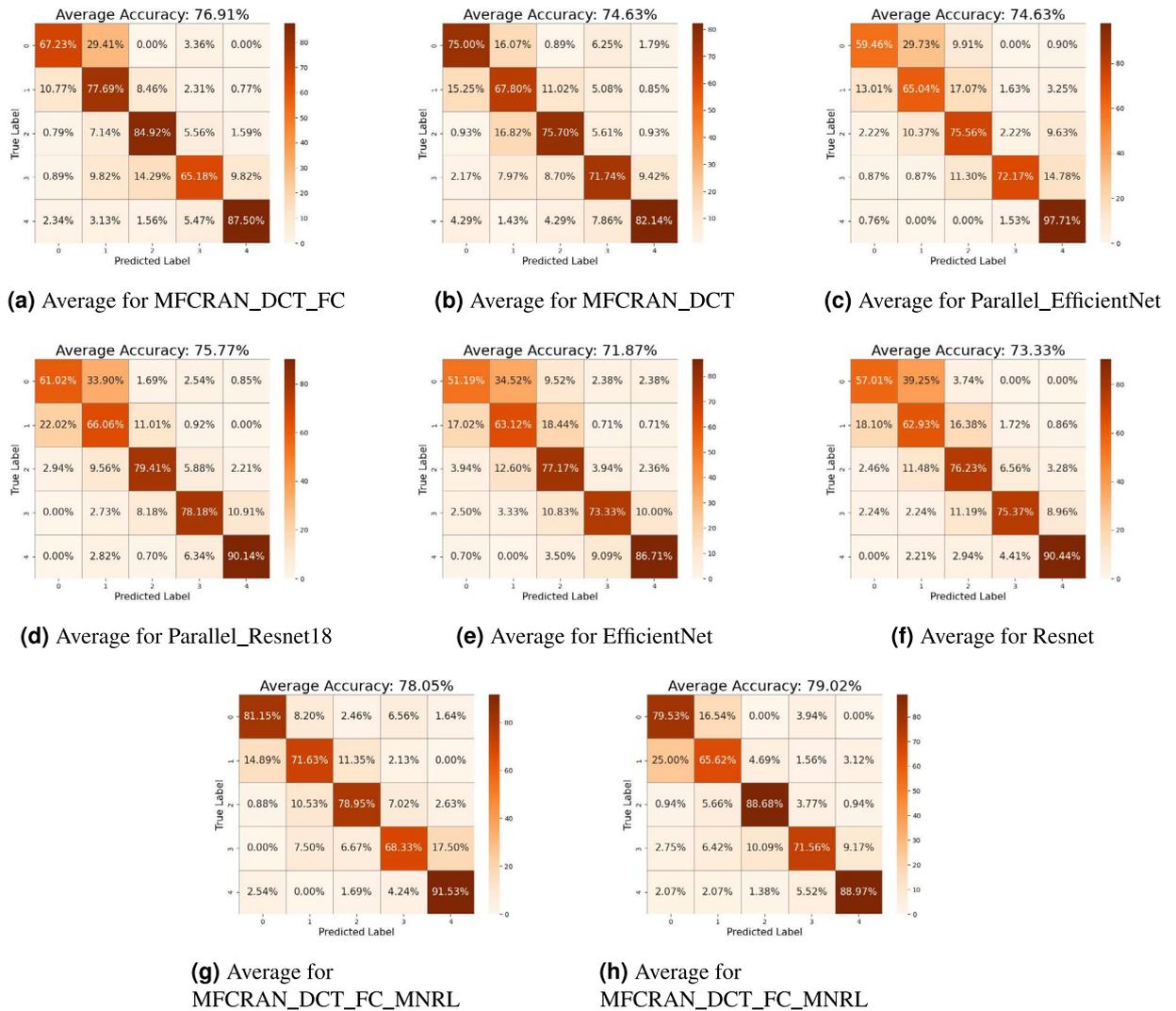
The experiments also reinforced the significance of incorporating preprocessing filters. The pinnacle performance across all experiments was achieved by `EfficientNet` with filters and a mean accuracy level of 74.63%. High pass filters, which enhance high-frequency components by emphasizing edges and fine details, and Sobel filters in the x-direction, which detect horizontal edges with different sensitivities, were applied and fed into this model. We also show the confusion matrices of all the models in Fig. 6. Given our small dataset, we observed that smaller models like `EfficientNet` achieved comparable results to larger ones like `ResNet18`. Furthermore, performance comparisons of models are provided in Table 3. Additionally, Fig. 6 demonstrates that classification errors are predominantly between adjacent stages, particularly in mid-maturation phases and at transitions such as A-B and D-E. This aligns with clinical observations, where differentiation is more challenging due to subtle morphological differences.

The `MFCRAN` architecture integrates pre-filters, attention modules, and a Discrete Cosine Transform (DCT) layer while maintaining computational efficiency suitable for clinical use. For instance, although `ResNet18` has approximately 11.2 million parameters, it achieves a lower accuracy of 73.33%. In contrast, our `MFCRAN_DCT_FC_MNRL` model reaches 79.02% accuracy with only 12.6 million parameters. This demonstrates that `MFCRAN` provides superior performance with a modest increase in model size, while remaining considerably lighter than deeper architectures like `ResNet34` (21.3M), `ResNet50` (23.5M), and `ConvNeXt` (27.8M), which yield lower accuracies and higher computational demands. These results indicate that `MFCRAN` is both effective and computationally efficient, making it a strong candidate for deployment in real-world clinical settings.

A critical component contributing to this improvement is the Discrete Cosine Transform (DCT) layer, positioned after the final convolutional block and before the fully connected layer. The DCT converts spatial features into frequency space, enabling the model to focus on informative frequency components while suppressing irrelevant noise. This transformation enhances the signal-to-noise ratio, enables more effective feature selection, and disentangles redundant patterns, allowing the model to concentrate on clinically

Model	Accuracy	Precision	Recall	F1 Score
ResNet18	73.33%	73.29	72.40	72.54
ResNet34	73.33%	74.83	73.34	73.58
ResNet50	69.59%	72.05	69.84	69.91
MobileNetV3	68.62%	69.02	67.91	68.25
EfficientNet_b0	71.87%	70.89	70.30	70.44
Parallel_ResNet18	75.77%	75.53	74.96	75.04
Parallel_EfficientNet_b0	74.63%	74.51	74.48	74.43
MFCRAN_FC	74.63%	75.70	73.99	74.16
MFCRAN_DCT_FC	76.91%	77.83	76.50	76.76
MFCRAN_FC_MNRL	78.05%	77.99	78.32	78.00
MFCRAN_DCT_FC_MNRL	79.02%	78.95	78.87	78.79

**Table 3.** Performance metrics (accuracy, precision, recall, and F1 score) for all models used to classify midpalatal suture stages.



**Figure 6.** Average confusion matrices for different models. Each matrix shows how often the predicted classes matched the actual maturation stages (A–E), helping evaluate classification performance.

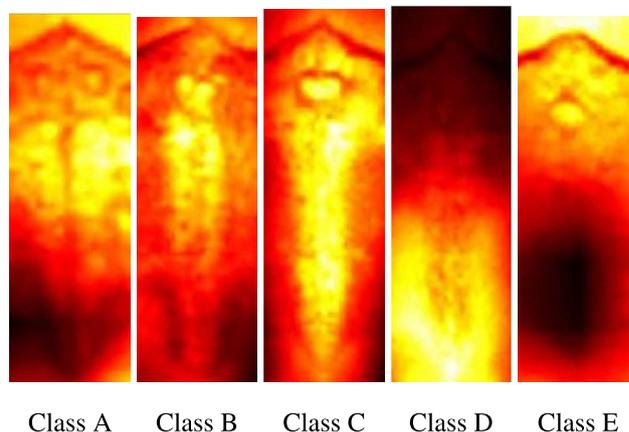
meaningful structures. Our results in Table 3 confirm that models incorporating the DCT layer consistently outperform their counterparts, offering both theoretical rationale and empirical support for its inclusion.

To enhance the model’s robustness and improve classification accuracy, we implemented a modified loss function. This new loss function was designed to prioritize the learning of harder-to-classify samples by adjusting gradient magnitude based on classification difficulty, which effectively prevents overfitting to easily recognized classes. By applying this loss function, the model undergoes a more balanced learning process, as it assigns a higher penalty to errors associated with complex samples and a lower penalty to simpler cases. This dynamic adjustment encourages the model to distribute attention more evenly across classes and minimizes the likelihood of model bias toward dominant classes.

To enhance the interpretability of our model and assess its clinical relevance, we employed Gradient-weighted Class Activation Mapping (Grad-CAM)-a widely used technique for visualizing the regions in an image that most influence a CNN’s prediction. Grad-CAM computes a localization heatmap by backpropagating the gradients of a target class into the final convolutional layers, thereby highlighting the spatial regions contributing most strongly to the classification outcome.

In our study, Grad-CAM was applied to visualize the model’s attention across different midpalatal suture maturation stages. The resulting attention maps consistently focused on clinically relevant anatomical structures, particularly the midpalatal suture region and adjacent maxillary bone, which are the key regions examined by clinicians during manual assessment. This alignment with expert reasoning enhances the trustworthiness and transparency of our model’s predictions.

Importantly, these attention maps confirm that the model is not relying on irrelevant or spurious features, but rather attends to areas that have direct morphological relevance to the classification task. This serves as an important step toward clinical adoption, where explainability is essential for integration into diagnostic workflows.



**Figure 7.** Example of Grad-CAM applied to our classification model for different classes (A–E). The highlighted regions indicate the areas of the image that the model considered important for its decision.

Thus, Grad-CAM not only helps in validating the internal logic of the model but also reinforces the clinical credibility of the proposed framework.

Incorporating Grad-CAM into our research workflow not only improved our understanding of the model's behavior but also provided a tool for diagnosing potential failure modes and biases (see Fig. 7).

## Discussion

Determination of midpalatal suture maturation through radiographic evaluation using CBCT is of utmost value in the dental field. This is especially useful in the treatment planning process for patients with maxillary transverse deficiency. The accurate and efficient appraisal of midpalatal suture maturation and ossification status is the basis for diagnosing the maxillary transverse developmental status<sup>34</sup>, thereby influencing success of treatment.

Tooth-borne RME is the most common and least invasive method to treat maxillary transverse deficiency. Classically, age has been used to determine success of expansion. Early histological studies have proposed that the midpalatal suture fuses in early childhood<sup>35</sup>. Bjork's growth studies have suggested that growth at the midpalatal suture ceases near the pubertal growth spurt<sup>10</sup>, with optimal timing of maxillary expansion occurring before 13 to 15 years old<sup>6,10</sup>. These findings indicate that RME may not be completed successfully in older ages due to increased interdigitation or fusion of the midpalatal suture. More recent histological and imaging studies, though, have found unfused midpalatal sutures in adult patients and a large amount of individual variation in midpalatal suture development when considering chronological age alone<sup>6</sup>.

The introduction of MARPE has increased the threshold at which clinicians can successfully complete palatal expansion, even in adult patients<sup>8</sup>. Jia et al. evaluated expansion using MARPE and RME in patients past their pubertal growth spurt. The researchers reported midpalatal suture separation in 100% of MARPE patients, compared to 86.7% in patients treated with a Hyrax RPE. MARPE was not only more successful in treating the transverse discrepancy, but also resulted in fewer adverse side effects to the dentition<sup>36</sup>.

It is difficult to reliably predict the success of RME and MARPE, and also unclear when to shift to a surgical approach when treating transverse discrepancy. There have been many attempts to determine when SARPE should be considered instead of RME or MARPE to achieve palatal expansion, with studies recommending a surgical approach starting anywhere from ages 14 to 25. There are currently no standardized clinical guidelines regarding timing of this treatment<sup>37</sup>. It is also known that age is not always directly related to suture fusion and cannot be used reliably as a sole determinant of midpalatal suture fusion<sup>19</sup>. For this reason, other predictors have also been considered, including cervical vertebral maturation, midpalatal suture density (MPSD), or appraising CBCT images directly and classifying according to MPSMS. Although the success of RME is multifactorial, many studies concluded that MPSMS is the most predictable parameter, and RME is successful at stages A, B, and C<sup>38–40</sup>. Shin et al. investigated the predictors of midpalatal suture expansion using MARPE in adults, concluding that age, palate length, and MPSM should be considered simultaneously as predictors of the success or failure of MARPE in young adults (Fig. 4)<sup>41</sup>. Other studies concluded that there is no correlation between successful MARPE expansion and MPSMS<sup>41</sup>.

Classification of the midpalatal suture using CBCT also presents a challenge in that accurate and consistent classification depends on the experience of the examiner<sup>20</sup>. Traditional methods rely heavily on these assessments, which can result in variations in diagnosis and differences in treatment outcomes. Developing AI for this task proposes an improvement in the precision of classification among different practitioners and standardization of treatment protocol. Gao et al. developed a CNN model analyzing the correlation between characteristics of midpalatal suture CBCT images and chronological age. Prediction efficiency was high for ages 4 to 10 (91%) and 17 to 23 (78.9%), but was much lower for the age ranges in between, especially in the 13 to 14 range (65.8%). This result attends to the variability in MPS maturation and ossification between individuals when evaluating based on age<sup>19</sup>. If chronological age cannot be used as an indicator for maturation, success of palatal expansion should instead be predicted using MPSMS. This study introduces an approach that utilizes learning techniques

to automate the classification process using MPSMS, aiming to reduce human error, enhance consistency, and optimize treatment for patients.

Our proposed deep neural network model incorporates layers, attention mechanisms<sup>42,43</sup>, and Discrete Cosine Transform (DCT) layers, demonstrating advancements in classification accuracy. The integration of DCT layers enables the model to effectively capture frequency-domain features, which are critical for identifying subtle patterns in CBCT images. Attention mechanisms further enhance the model by focusing on the most relevant regions of the input data, improving its interpretability and performance. The MFCRAN\_DCT\_FC model achieved an accuracy of 76.91%, showcasing its usefulness in clinical applications. The MFCRAN\_DCT\_FC\_MNRL model achieved an accuracy of 79.02%, demonstrating the additional value of the Modified Margin-based Negative Logarithmic Loss (MNRL) function in refining classification boundaries and addressing class imbalances. The integration of High Pass and Sobel filters was pivotal in enhancing the model's ability to identify and emphasize critical features in CBCT images. The High Pass filter sharpens the input data by highlighting high-frequency components, such as edges and fine details, while the Sobel filter further accentuates gradient information, aiding in the detection of structural patterns<sup>44,45</sup>. Together, these preprocessing steps improve the quality of input features, ensuring that the model can better distinguish between subtle variations within the images.

Moreover, employing data augmentation techniques has helped addressing class imbalances and mitigating overfitting issues, bolstering the model's resilience. These results emphasize the significance of integrating advanced image preprocessing and augmentation strategies when developing AI models for imaging purposes.

Our study is the first to utilize a deep learning algorithm to automatically classify midpalatal suture maturation stages. The dataset of 618 images, which were extracted from archived data, included a wide variety of ages; however, the race of the individuals was not specified. Future studies should concentrate on expanding the dataset to encompass a wider variety of CBCT images, including scans from a large collection of demographic backgrounds. The model's performance could be enhanced by exploring alternative deep learning architectures, investigating the utilization of imaging modalities, and incorporating modal data. Upon diversification of data and further improving the model, future directions could also include minimizing the role of human evaluators by automating the selection and retrieval of CBCT slices.

## Conclusion

This research introduces a deep learning-based method to automatically categorize the maturation stages of the midpalatal suture using CBCT images, a novel feat. This innovative model, which combines layers, attention mechanisms, and Discrete Cosine Transform (DCT) layers, outperformed other approaches. The MFCRAN\_DCT\_FC\_MNRL model achieved a classification accuracy of 79.02%, showcasing its usefulness in orthodontic diagnosis and treatment planning.

Integrating High pass and Sobel filters as preprocessing steps improved the model's ability to identify features in CBCT images, leading to enhanced classification accuracy. Further research will focus on enhancing the model and confirming its effectiveness with varied clinical data sets.

Leveraging learning techniques such as convolutional layers, attention mechanisms, and DCT layers provides a valuable asset for orthodontists. This method holds promise in enhancing the precision and effectiveness of diagnosing and planning treatments for transverse deficiencies.

## Data availability

All data generated during this study are included in this published article and its supplementary information files. The data analyzed in this study are available from The University of Illinois Chicago. Still, restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. However, they are available from the corresponding author on reasonable request.

Received: 13 January 2025; Accepted: 22 May 2025

Published online: 29 May 2025

## References

1. Olch, A. J. & Alaei, P. How low can you go? A CBCT dose reduction study. *J. Appl. Clin. Med. Phys.* **22**, 85–89 (2021).
2. Scarfe, W. C., Azevedo, B., Toghiani, S. & Farman, A. G. Cone beam computed tomographic imaging in orthodontics. *Aust. Dent. J.* **62**, 33–50 (2017).
3. Kapila, S. D. & Nervina, J. M. CBCT in orthodontics: Assessment of treatment outcomes and indications for its use. *Dentomaxillofac. Radiol.* <https://doi.org/10.1259/dmfr.20140282> (2015).
4. Hägg, U. & Taranger, J. Maturation indicators and the pubertal growth spurt. *Am. J. Orthod.* **82**, 299–309 (1982).
5. Fishman, L. S. Radiographic evaluation of skeletal maturation: A clinically oriented method based on hand-wrist films. *Angle Orthod.* **52**, 88–112 (1982).
6. Shayani, C. G., Vidal, M. Midpalatal suture maturation method for the assessment of maturation before maxillary expansion: A systematic review. *Diagnostics* **12** (2022).
7. Proffit, W. R., Fields, H., Larson, B. & Sarver, D. M. *Contemporary Orthodontics-E-Book: Contemporary Orthodontics-E-Book* (Elsevier Health Sciences, 2018).
8. Jones, J. P., Elnagar, M. H. & Perez, D. E. Temporary skeletal anchorage techniques. *Oral Maxillofac. Surg. Clin.* **32**, 27–37 (2020).
9. Allareddy, V., Maclaine, J. & Elnagar, M. H. Principles of orthodontic management of treacher collins syndrome. *Craniofacial Microsomia and Treacher Collins Syndrome: Comprehensive Treatment of Associated Facial Deformities*, 429–446 (2022).
10. Bishara, S. E. & Staley, R. N. Maxillary expansion: Clinical implications. *Am. J. Orthod. Dentofac. Orthop.* **91**, 3–14 (1987).
11. Garib, D. G., Henriques, J., Janson, G., de Freitas, M. R. & Fernandes, A. Y. Periodontal effects of rapid maxillary expansion with tooth-tissue-borne and tooth-borne expanders: A computed tomography evaluation. *Am. J. Orthod. Dentofac. Orthop.* **129**, 749–758 (2006).

12. Baysal, A. et al. Evaluation of root resorption following rapid maxillary expansion using cone-beam computed tomography. *Angle Orthod.* **82**, 488–494 (2012).
13. Persson, M. & Thilander, B. Palatal suture closure in man from 15 to 35 years of age. *Am. J. Orthod.* **72**, 42–52 (1977).
14. Melsen, B. Palatal growth studied on human autopsy material: A histologic microradiographic study. *Am. J. Orthod.* **68**, 42–54 (1975).
15. Korbmacher, H., Schilling, A., Püschel, K., Amling, M. & Kahl-Nieke, B. Age-dependent three-dimensional microcomputed tomography analysis of the human midpalatal suture. *J. Orofacial Orthop.* **68**, 364–376 (2007).
16. Wehrbein, H. & Yildizhan, F. The mid-palatal suture in young adults. A radiological–histological investigation. *Eur. J. Orthodont.* **23**, 105–114 (2001).
17. Revelo, B. & Fishman, L. S. Maturational evaluation of ossification of the midpalatal suture. *Am. J. Orthod. Dentofac. Orthop.* **105**, 288–292 (1994).
18. Angelieri, F., Franchi, L., Cevidan, L. H., Bueno-Silva, B. & McNamara, J. A. Jr. Prediction of rapid maxillary expansion by assessing the maturation of the midpalatal suture on cone beam ct. *Dent. Press J. Orthodont.* **21**, 115–125 (2016).
19. Lu Gao, L. Z. Z. S. Q. W., Zhiyu Chen & Yu, G. Midpalatal suture cbct image quantitative characteristics analysis based on machine learning algorithm construction and optimization. *Bioengineering* **9** (2022).
20. Sachin Chhatwani<sup>1</sup>, S. C. M. B. L. J. J., Annahita Arman & Danesh<sup>1</sup>, G. Performance of dental students, orthodontic residents, and orthodontists for classification of midpalatal suture maturation stages on cone-beam computed tomography scans—A preliminary study. *BMC Oral Health* **24** (2024).
21. Lee, J.-G. et al. Deep learning in medical imaging: General overview. *Korean J. Radiol.* **18**, 570–584 (2017).
22. Zhang, J., Li, C., Song, Q., Gao, L. & Lai, Y.-K. Automatic 3d tooth segmentation using convolutional neural networks in harmonic parameter space. *Graph. Models* **109**, 101071 (2020).
23. Thurzo, A. et al. Use of advanced artificial intelligence in forensic medicine, forensic anthropology and clinical anatomy. *Healthcare* **9**, 1545 (2021).
24. Pytorch transform. <https://pytorch.org/vision/stable/transforms.html> Accessed: 2023-07-13.
25. LeCun, Y. et al. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Netw.* **3361**, 1995 (1995).
26. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.* **25**, 1097–1105 (2012).
27. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
28. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3431–3440 (2015).
29. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
30. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4510–4520 (2018).
31. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* 6105–6114 (PMLR, 2019).
32. Vetterli, M. & Herley, C. Wavelets and filter banks: Theory and design. *IEEE Trans. Signal Process.* **40**, 2207–2232 (1992).
33. Milani, O. H. et al. A fully automated classification of third molar development stages using deep learning. *Sci. Rep.* **14**, 13082 (2024).
34. Priyadarshini, J. et al. Stress and displacement patterns in the craniofacial skeleton with rapid maxillary expansion—A finite element method study. *Prog. Orthod.* **18**, 1–8 (2017).
35. RA, L. The development, structure and growth pattern of the human mid-palatal suture. *J. Anat.* **108**, 31–41 (1971).
36. Jia H, Z. N. B. Y. L. S., Zhuang L. Comparison of skeletal maxillary transverse deficiency treated by microimpant-assisted rapid palatal expansion and tooth-borne expansion during the post-pubertal growth spurt stage: A prospective cone beam computed tomography study. *Angle Orthod.* **91**, 36–45 (2021).
37. Jang HI, C. J. K. K. C. J. C. N. L. K. C. J., Kim SC. Relationship between maturation indices and morphology of the midpalatal suture obtained using cone-beam computed tomography images. *Korean J. Orthod.* **46** (2016).
38. Tonello, D. L. et al. Midpalatal suture maturation in 11- to 15-year-olds: A cone-beam computed tomographic study. *Am. J. Orthod. Dentofac. Orthop.* **152**, 42–48 (2017).
39. Acar, Y. B. & Abuhan, E. Evaluation of the relationship between midpalatal suture maturation indicators as predictors for surgically-assisted and conventional rapid maxillary expansion. *J. Craniofac. Surg.* **32**, 678–681 (2021).
40. Grünheid, T., Larson, C. E. & Larson, B. E. Midpalatal suture density ratio: A novel predictor of skeletal response to rapid maxillary expansion. *Am. J. Orthod. Dentofac. Orthop.* **151**, 267–276 (2017).
41. Shin, H. et al. Predictors of midpalatal suture expansion by miniscrew-assisted rapid palatal expansion in young adults: a preliminary study. *Kor. J. Orthodont.* **49**, 360–371 (2019).
42. Xie, Y. et al. Attention mechanisms in medical image segmentation: A survey. *arXiv preprint arXiv:2305.17937* (2023).
43. Karacan, M. H. & Yücebas, S. C. A deep learning model with attention mechanism for dental image segmentation. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–4 (IEEE, 2022).
44. Agrawal, A. & Bhogal, R. K. Edge detection techniques in dental radiographs (sobel, t1fls & it2fls). In *Communication, Networks and Computing: First International Conference, CNC 2018, Gwalior, India, March 22–24, 2018, Revised Selected Papers 1*, 411–421 (Springer, 2019).
45. Dogra, A. & Bhalla, P. Image sharpening by gaussian and Butterworth high pass filter. *Biomed. Pharmacol. J.* **7**, 707–713 (2014).

## Acknowledgements

American Association of Orthodontists Foundation.

## Author contributions

O.M, M. T. Conducted the experiment and Wrote the manuscript. R.A. and A.C. contributed to the Study design and editing of the manuscript. V.A. Edited manuscript. L.M, A.N labeled and prepared data and Wrote the manuscript. M.H.E. Contributed to conception, funding acquisition, study design, data preparation, project director, and manuscript writing.

## Funding

M.H.E. American Association of Orthodontists Foundation. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Declarations

### Competing interests

All the authors have nothing to disclose.

### Consent for publication

All the authors gave full consent for publication.

### Additional information

**Correspondence** and requests for materials should be addressed to M.H.E.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025