



IMPROVED PANORAMIC DENTAL X-RAY SEGMENTATION USING MSOA-SE-RES-DENSE U-NET++ ARCHITECTURE

Sankar Murugesan¹, Sivaprakash P², Vimala Ithayan J³, Dilip Charaan R M⁴, Vijayanandh T⁵,
Santhoshkumar S P⁶

^{1,2}Professor, School of Electrical and Communication, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.

^{3,4,5,6}Associate Professor, School of Computing, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.

¹<https://orcid.org/0000-0002-6256-4931>, ²<https://orcid.org/0000-0002-2777-8889>, ³<https://orcid.org/0000-0002-4647-0863>,

⁴<https://orcid.org/0009-0002-1332-4592>, ⁵<https://orcid.org/0000-0002-6307-4272>, ⁶<http://orcid.org/0000-0001-8531-759X>

Email: *drsankarm@veltech.edu.in, sivabtl85@gmail.com, vimala.ithayan@gmail.com, dilip_charaan@yahoo.co.in,
drvijayanandht@veltech.edu.in, spsanthoshkumar16@gmail.com

ARTICLE INFO

Article History

Received: January 19, 2026

Reviewed: July 22, 2026

Accepted: August 22, 2026

Published: September 30, 2026

Keywords:

Panoramic dental X-rays

Mantis Shrimp Optimization

Algorithm,

SE-Res-Dense U-Net++

Resize

CLACHE

Intensity Normalization

ABSTRACT

Panoramic dental X-rays are the most common imaging source for oral conditions; however, their reading is often delayed by artefacts, and poor contrast. A precise segmentation of dental structures represents the core for clinical diagnosis and treatment planning; yet, current solutions frequently fail to generalize and be robust. To address this issue, introduce a novel MSOA-SE-Res-Dense U-Net++ architecture integration the features of Squeeze-and-Excitation (SE) blocks, residual connections with U-Net ++, dense feature propagation, and nested skip pathways for enhancing the feature and gradient stability. The hyperparameters are tuned through the Mantis Shrimp Optimization Algorithm (MSOA) to faster convergence and better generalization. An exhaustive pre-processing process that comprises resizing, Contrast-Limited Adaptive Histogram Equalization (CLAHE), and intensity normalization is aimed at standardizing the input quality. From the Python software, experimental study using "Collection of Dental X-ray Images for Analysis" dataset demonstrates the effectiveness of approach in terms of accuracy (98.00%), precision (96.50%), recall (95.50%), and F1-score (96.50%), thereby outperforming the existing models. The accomplished results signify the promise of the presented framework in giving dependable clinical segmentation for dental diagnostics and treatment planning.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Medical image processing has been assessed and developed into a vital tool for researchers in clinical domains, including dentistry, because of the ongoing development of medical radiological imaging during the past few decades [1]. Orthopantomography (OPT), often known as panoramic radiography, is a standard imaging technology for dental X-rays [2]. Dentists use panoramic dental X-rays to visualize the patient's oral health and used to detect dental disorders such as caries, cavities, gum disease, restorations, implants, damaged teeth, and oral cancer [3], [4]. Without imaging X-rays, the dentist unable to discover the problem until the condition has progressed to a severe stage. OPT is a valuable imaging technology that offers doctors with diagnostic information, decision-making capabilities, and organizational insights; yet, it has significant limitations [5]. The lack of three-dimensionality, possible artefacts, overlapping structure, and geographical inhomogeneity are affecting doctors' ability to evaluate data accurately [6]. Furthermore, OPT images have been distorted and magnified, affecting measuring accuracy. The position of the patient's head, the dimension and shape of the jaws, and anatomical variances are lead to image distortion and change the true dimensions of the structures under examination [7], [8]. Medical professionals use dental X-rays to diagnose and classify tooth disorders, identify anatomical features, and segment tooth structures [9]. Tooth segmentation is employed in the dental profession for recognizing portions of the jaws or teeth, identifying dental tumours and infections in the jaw bone, however it faces many obstacles and challenges.

These issues exist in every health care sector, yet they are more prevalent in dentistry [10], [11]. These issues include noise, insufficient contrast, and irregular exposure. The segmentation process is challenging by additional factors including body tissue similarities, permanent artefacts like as tooth fillings and dental implants, and temporary artefacts including as orthodontic braces, damaged teeth, teeth crowding, space between teeth, and missing teeth [12], [13]. Due to variations in image quality and anatomical variations, a segmentation method which performs effectively in one dataset or imaging condition is fail to perform efficiently on another [14]. Manual segmentation is inefficient and needs extensive knowledge of dental anatomy and proficiency with specialized imaging techniques. It also has a high inter-operator variability, which reduces uniformity among practitioners [15]. In modern technologies, Deep Learning (DL) have exposed ability for improving the accurate prediction of dental image interpretation. These algorithms represent a substantial improvement in the detection and classification of tooth disease in panoramic dental X-ray images, as compared to former rule-based methods and clinician diagnoses [16], [17].

I.1 LITERATURE SURVEY

By [18] have created a Hybrid DL Model that utilized YOLOv8 for object detection, and the Segment Anything Model (SAM) for high-quality segmentation accuracy of restorative components in panoramic dental X-ray images. The Hybrid DL Model is effective at recognizing fillings, crowns, root canals, and implants, resulting in increased diagnostic accuracy, although its effectiveness is bounded by YOLO detection accuracy, and detection errors are propagating to SAM resulting in inaccurate and incomplete segmentation. In turn [19] have utilized a Mask Region-based Convolutional Neural Network (Mask-RCNN) Model for refined and reliable automatic teeth segmentation in panoramic dental X-ray images which successfully localized tooth boundaries. The method demonstrated sound performance in segmentation, as well as potential applicability in clinical practice including orthodontics and implantology. Some limitations of the study is a relatively small and homogenous dataset inclusive of some deciduous classes. According to [20] have proposed an Efficient Encoder-Decoder Network (EED-Net) for automating the maxillofacial segmentation of panoramic dental X-ray images to assist the diagnostic process. The authors incorporated the use of residual encoders, and U-Net to enhance accuracy. Nevertheless, the study is constrained by a result of limited image resolution because of memory constraints, and limited application for multi-class segmentation and extracting high-resolution ROIs. By [21] have implemented a DL based CariesNet to automatically segment tooth caries lesions from oral panoramic X-ray images. The framework is built on a U-shaped encoder-decoder framework that adds a partial encoder module to classify between shallow, moderate, and deep caries. However, it is limited in terms of generalization to new medical variants. According to [22] developed a multiclass teeth segmentation model comprised of an M-Net architecture, Swin Transformers (ST), and Teeth Attention Block (TAB), resulting in improved multi-class tooth segmentation accuracy on panoramic dental x-ray images. TAB improves attention on complex dental structures, while multiscale supervision and squared Dice loss approaches reduce class imbalance problem that result in increased performance. However, the study is limited due to few deceased samples, which indicates that more and differences samples are required to generalize model to tooth-related conditions. In turn [23] have introduced a Two-Stage Attention Segmentation Network (TSASNet) for fully automatic tooth segmentation of panoramic dental x-ray images. The architecture implements an attention module to segment initially and fine segmentation using a fully convolutional network in the second stage. Limitations include sensitivity to noise presented in images and reliant on panoramic images of optimal quality; therefore, robustness must be improved for images of different quality.

I.2 RESEARCH GAPS

Despite substantial advances in dental X-ray segmentation using DL, existing approaches continue to confront issues such as low generalisation across various datasets, sensitivity to noise and aberrations, and difficulty effectively delineating overlapping or irregular tooth features. Many past research used small or homogeneous datasets, which limited clinical usefulness, and optimization algorithms frequently converged prematurely, lowering robustness. These shortcomings emphasise the need for a more adaptable architecture and improved optimisation techniques to accomplish consistent, high-accuracy segmentation in real-world dental imaging situations.

I.3 CONTRIBUTION

- A systematic technique to pre-process images through resizing, CLAHE, and intensity normalization to better visibility of smaller dental features and ensures that the quality of inputs is maintained consistently across datasets.
- SE-Res-Dense U-Net++ segmentation is presented as the combination of SE modules, residual connections U-Net ++, dense feature block, to accomplish strong segmentation of Panoramic Dental X-rays.
- Hyperparameters are optimized with the use of MSOA, resulting in quick convergence and less of a tendency to over fit and an increase in the ability to generalize.

II. PROPOSED SYSTEM DESCRIPTION

Panoramic dental X-ray images are gathered from “Collection of Dental X-ray Images for Analysis” dataset and clinical sources for the purposes of this study. The dataset consists of high-resolution images containing a range of dental pathologies such as caries, cavities, restorations, missing teeth, and irregular tooth borders. The system begins with a pre-processing step, where panoramic dental X-ray images are resized to a uniform size for consistency across the dataset. After the images are resized, CLAHE is applied to enhance image contrast and visibility of subtle dental structures such as enamel, dentin, and tooth borders. Then, intensity normalization is performed to achieve a more constant distribution of pixel intensity across images by alleviating exposure variation and stabilizing the model.

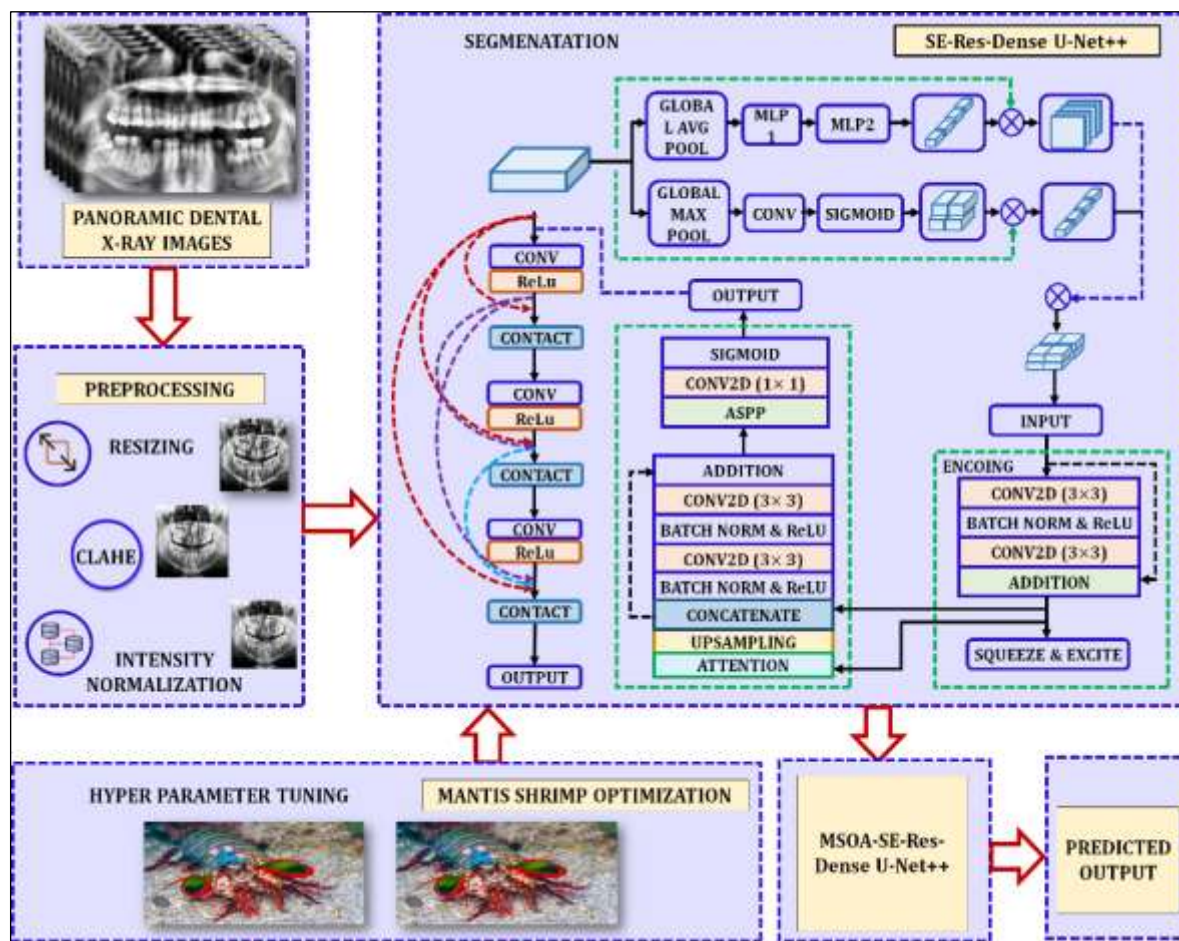


Figure 1: Block of panoramic dental x-ray segmentation using MSOA-SE-Res-Dense U-Net++.
Source: Authors, (2026).

The images are then processed through the SE-Res-Dense U-Net++ segmentation network which comprises SE blocks to recalibrate channel-wise feature responses, residual connections with U-Net ++ for gradient flow, and dense feature propagation to learn both local and global contexts. Furthermore, hyperparameter tuning is also performed using the MSOA to make convergence faster, reduce overfitting, and improve generalization. The final architecture, known as MSOA-SE-Res-Dense U-Net++, results in very accurate segmentation to allow for precise delineation of dental structures and irregular tooth borders in Figure 1.

III. PROPOSED SYSTEM METHODOLOGY

III.1 DATA PREPROCESSING

To improve and maintain the quality and consistency of panoramic dental X-rays, data pre-processing is used. The data pre-processing for panoramic dental X-rays consists of the following steps in Figure 2: resizing, CLAHE, and intensity normalisation for the improved visibility of subtle dental structures.

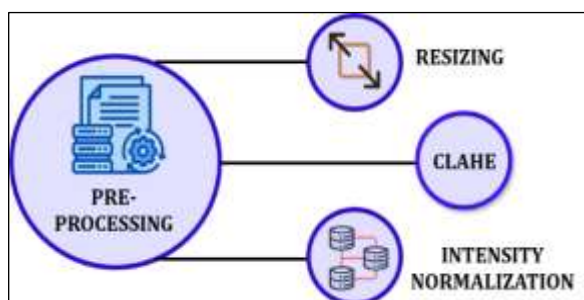


Figure 2: Pre-processing stages.
Source: Authors, (2026).

III.1.1 Resize

In the first step of the segmentation structure, the images are resized to be uniform across the public and clinical datasets. The original images collected are of high resolution (4000×3000 pixels) and contain very large non-informative background that is not useful for the purposes of dental anatomical studies. In order to focus the learning process on clinically relevant structures, an informed cropping operation to collect the dental arch, leaving out redundant background is done.

After cropping, each of the images are resized to 1024×1024 pixels. The size is chosen because it supported preserving fine anatomical details such as enamel boundaries and tooth contours, but also optimized for computational efficiency. The standardized dimension of an image creates uniformity in terms of feature extraction and assurances stable learning behavior throughout the dataset which helps improve the overall robustness of the proposed segmentation architecture.

III.1.2 CLAHE

After the images have been reduced in size, CLAHE is utilized to help describe the tiny dental structures required by the proposed model. The dental radiographs experience low-contrast due to exposure related variability, patient position, and anatomical density. CLAHE is again applied with a clip limit of 2.0, along with a grid size of 8×8 , in order to increase local contrast without increasing noise. This emphasizes additional subtle regions related to segmentation, such as enamel boundaries, dentin interfaces, and irregular edges of the tooth, all of which are useful features for the model to segment accurately.

III.1.3 Intensity Normalization

The following processing step is intensity normalization, which aims to standardize the grayscale intensity distributions of all images. There is a lot of variability in brightness and contrast of panoramic dental X-rays, all of which influences by acquisition parameters. Therefore, this step standardizes the input to the segmentation network. Min-Max normalization is used in this study to scale the pixel intensities to the range of 0-1 with the following transformation:

$$I_{norm}(x) = \frac{I(x) - I_{min}}{I_{max} - I_{min}} \quad (1)$$

This normalization ensures that the learning process is stabilized, provides better information flow, and increases the generalizability of the model. With the assignment of all images into a consistent intensity scale, the network is better able to learn structural patterns and accurately identify and detect irregular tooth borders.

III.2 SE-RES-DENSE U-NET++

SE-Res-Dense U-Net++ is an improved structure to help segment panoramic dental X-Rays. It uses SE blocks, residual links, and use of dense features within the U-Net++ structure to enhance both network's capability to extract features from an image and to give better stability when calculating gradients. This type of architecture allows network to capture both very fine local details and also to combine this information with larger areas of context that are important for correctly delineating a tooth or bone structure.

III.2.1 Squeeze and Excitation Residual Network (SER-Net)

The SER-Net is made up of stacked SE-Residual modules that complete the feature extraction operation and a decoder that ultimately reconstructs the refined features as illustrates in Figure 3. Each SE-Residual module consists of a SE block on Figure 4, and a residual structure that minimizes feature over-fitting and enhances the informative feature channels. The input images are then processed through multiple iterations of SE-Residual modules that initially extract the features, the refined features then being further processed in the decoder that aims to enhance the discriminative representation, and the output segmentation generated in two separate stages of transposed convolution followed by the Softmax operation.

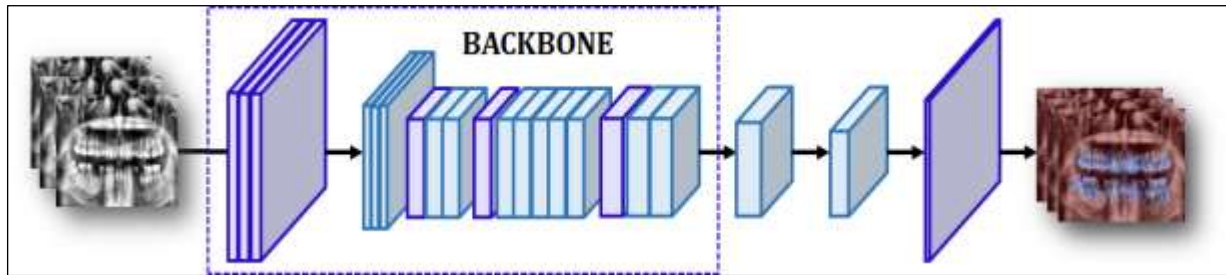


Figure 3: SER-Net Architecture.
Source: Authors, (2026).

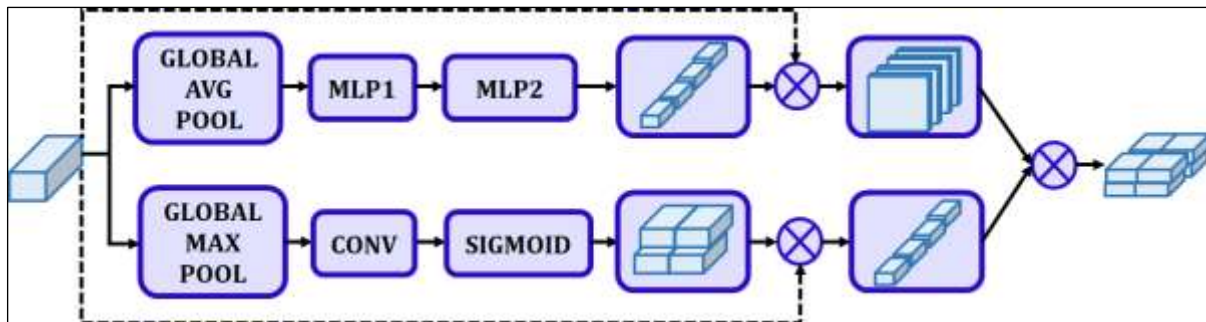


Figure 4: Squeeze and Excitation Block.
Source: Authors, (2026).

III.2.2 SE-Residual Module

SE Block, along with two variations of SE-Residual modules in Figure 5, are tailored explicitly to the panoramic dental x-ray image modality. The two versions include SE-Residual-A architecture and the SE-Residual-B architecture; in both configurations, the SE Block is integrated into two distinct residual structures. The SE Block is composed of a Channel Squeeze-and-Excitation (CSE) block and Spatial Squeeze-and-Excitation (SSE) block, which enables feature representation at the interdependencies of the characteristics of channel and spatial separation.

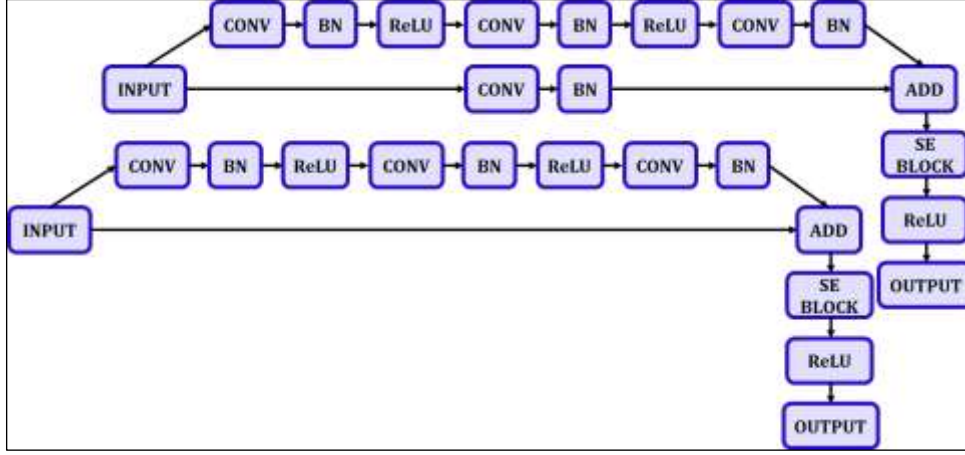


Figure 5: SE-Residual Modules.
Source: Authors, (2026).

Within the framework of segmentation of panoramic x-ray image candidate sites, these SE Blocks act as channel and spatial adaptations to aid feature representation or recalibration. The adaptive routing of these mechanisms promotes further propagation of clinically relevant features within panoramic dental x-ray images namely, tooth contours, enamel-dentin boundaries, cavities, restorations, and jawbone morphology. Furthermore, these features modify the design to alleviate network degradation and vanishing gradients when completing deep feature extraction. The input feature maps are expressed as,

$$X = [x_1, x_2, \dots, x_c, \dots, x_C] \quad (2)$$

Where, $x_c \in \mathbb{R}^{H \times W}$. H, W and C represents height, width, and number of channels in the feature maps and c indicates c^{th} channel.

III.2.3 Channel Squeeze and Excitation Block

The input feature maps X are initially passed via CSE Block, which effectively models the interdependencies across channels. The CSE Block encodes the information along the spatial dimensions to create a multi-channel descriptor. This means the global contextual cues on tooth structure, jawbone density or variations in radiographic intensity to be collective and transported into deeper layers of network. This is accomplished by applying the global average pooling operation, which computes pooled channel-wise descriptor.

$$Z = [z_1, z_2, \dots, z_c, \dots, z_C] \quad (3)$$

Obtained by spatially shrinking the feature maps,

$$X = [x_1, x_2, \dots, x_c, \dots, x_C], x_c \in \mathbb{R}^{H \times W} \quad (4)$$

The CSE Block's output is generated via channel-wise multiplication of the feature mappings (X) and modulation vector (V):

$$U_{CSE} = V \odot X = MLP_2(MLP_1(Z)) \odot X = MLP_2\left(MLP_1\left(\frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j)\right)\right) \odot X \quad (5)$$

Here, $\odot$ denotes channel wise multiplications and MLP_1 and MLP_2 represents shared perceptron layer.

III.2.4 Resnet Connection in U-Net ++

Residual UNet++ in Figure 6, is an extension of the UNet++ segmentation framework that incorporates residual units into its attribute framework while retaining the premise of a unique dense skip-pathway. The changes still retain the attributes from UNet++ that includes a dense arrangement of self-nested feature connections to link both the encoding and the decoding nodes across a wide range of semantic paths. The UNet++ decoder receives multi-level feature maps from the ResNet encoder through nested feature connections, which facilitates the preservation of fine dental structures include tooth boundaries, enamel edges, and jaw contours.

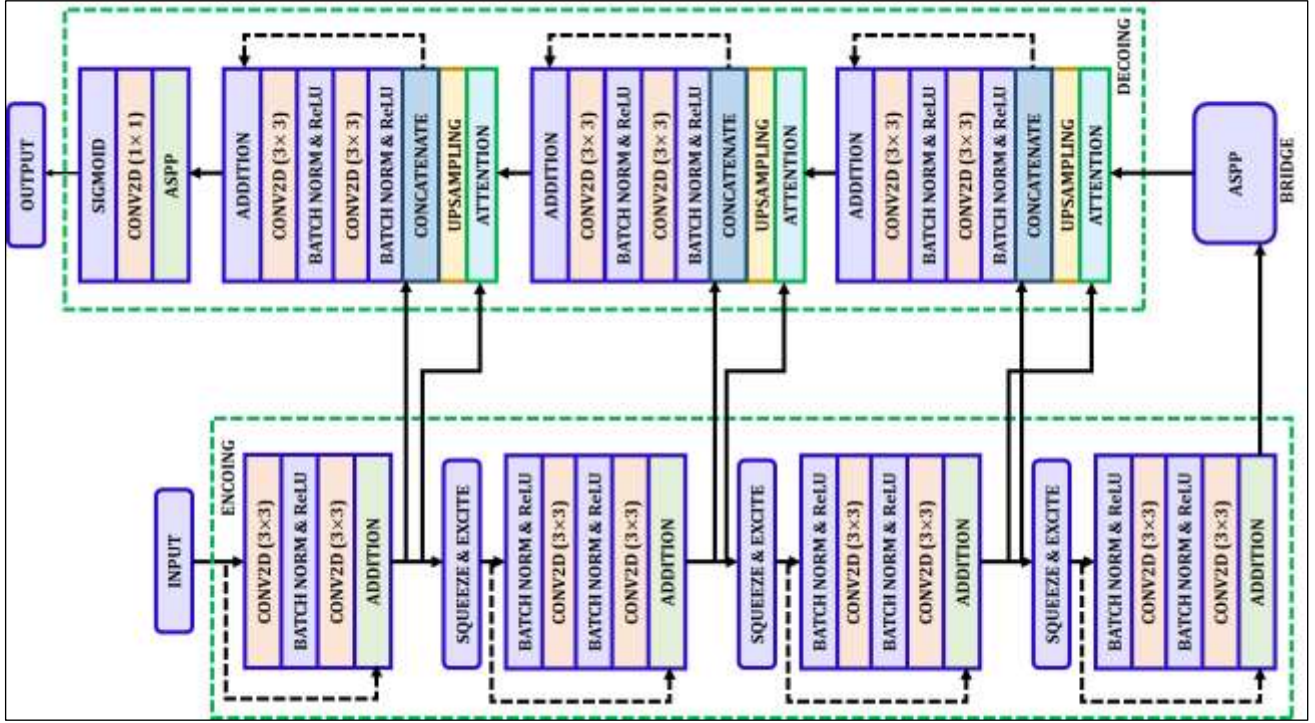


Figure 6: Architecture of ResUNet++.
Source: Authors, (2026).

Let $x_{i,j}$ denote the output at encoder–decoder depth i and dense-block level j . The feature transmission is defined as:

$$x_{i,j} = \begin{cases} H(x_{i-1,j}), & j = 0 \\ H([x_{i,0}, x_{i,1}, \dots, x_{i,j-1}, T(x_{i+1,j-1})]), & j > 0 \end{cases} \quad (6)$$

Here, $H(\cdot)$ is a convolution followed by ReLU activation, and $T(\cdot)$ represents the upsampling of deeper feature maps. The concatenation operator $[\cdot]$ refers to the union of multi-resolution information. The simultaneous spatial representation of encoder features and decoder features is semantically aligned, which is paramount to accurately segment overlapping dental structures in panoramic radiographs. The integration of ResNet residual blocks with UNet++ dense connections improve gradient flow and stability during training, as well as the model's proficiency in detecting subtle radiographic reveal variations, including caries, cavities, and irregular tooth borders.

III.2.5 Dense Blocks (DBs)

To facilitate feature reuse and reduce the chance of vanishing gradients in the proposed panoramic dental X-ray segmentation network, DBs in Figure 7, are used in the architecture. Each DB connects each layer that follows to the feature maps of all preceding layers to allow for the most efficient flow of information through the network, which also helps the network build more robust representations of fine-grained dental structures.

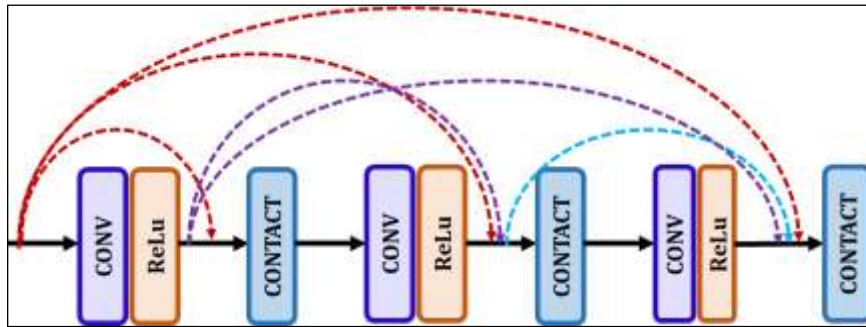


Figure 7: Structure of Dense Block.
Source: Authors, (2026).

Let x_l be the initial input to a DB. Each block consists of L convolutional layers, where the l^{th} layer applies a nonlinear transformation denoted by $F_l(\cdot)$:

$$y_l = F_l(x_l) \quad (7)$$

For $l \geq 2$, the input x_l is gathered via concatenating the outputs of all previous layers:

$$x_l = [y_{l-1}, \dots, y_l, x_l] \quad (8)$$

This dense structure of connectivity guarantees that low-level spatial boundary information and high-level anatomical features are exchanged or propagated through the network together. By connecting the two DBs, the segmentation framework achieves improvements in gradient stability, better contextual representations, and enhanced accuracy to segment teeth, jawbone, and surrounding anatomical tissues.

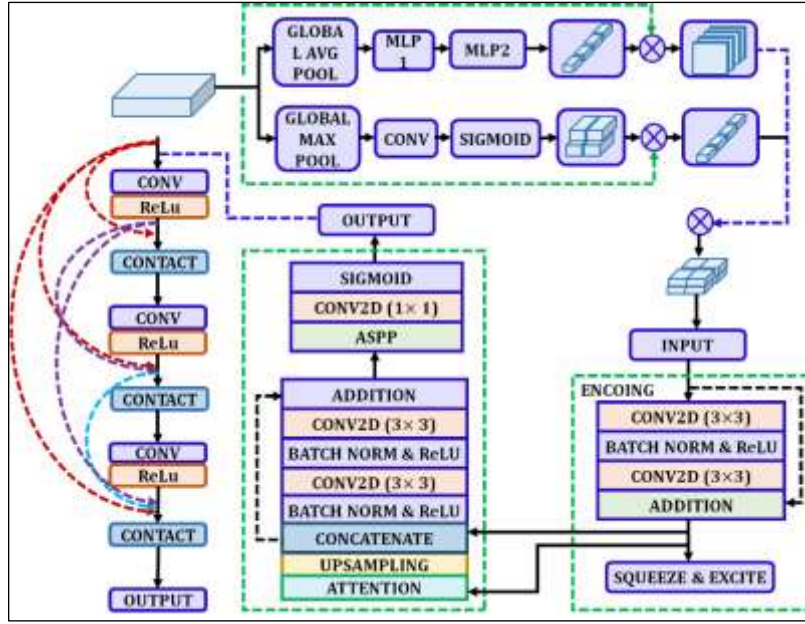


Figure 8: Architecture diagram of B. SE-Res-Dense U-Net++.

Source: Authors, (2026).

SE-Res-Dense U-Net++ in Figure 8 improves the robustness of the segmentation process to artefacts, overlapping structures, and variability between humans and their teeth. It improves the visibility of small features within the X-ray image of a tooth such as enamel lines, contours, etc. which ultimately enhances the ability for a practitioner to use it clinically. Further, hyperparameters are tuned using the MSOA which accelerated the rate of convergence, reduced the probability of overfitting, and improved the generalization of this network.

III.3 MANTIS SHRIMP OPTIMIZATION ALGORITHM (MSOA)

MSOA is used to optimize hyperparameters of SE-Res-Dense U-Net++ segmentation model in Figure 9. MSOA exploits polarized-light visual processing characteristics from the shrimp and its adaptive foraging, attacking, and sheltering behaviours to explore and exploit hyperparameter sets through polarization-based decision logic. The MSOA acts to quickly search the parameter space to inform systematic improvements in feature recalibration, dense connectivity, and nested skip-path optimized segmentation.

III.3.1 Initialization

In the initial stage of the optimization procedure, MSOA produces collection of multidimensional candidate solutions, where individual candidate solution is a hyperparameter set that employed in the SE-Res-Dense U-Net++ segmentation model.

$$X_{i,j} = lb_{1,j} + rand_{i,j} \cdot (ub_{1,j} - lb_{1,j}) \quad (9)$$

The initial candidate population $X_{i,j}$ is obtained by uniformly sampling a hyperparameter within a predefined set of limits defined as lower and upper bounds $lb_{1,j}$ and $ub_{1,j}$. The hyperparameter space is initialized to ensure sufficient diversity in the search for effective optimization of learning rate of the segmenting properties of the SE-Res-Dense U-Net++.

III.3.2 Foraging Strategies

During the exploration phase, MSOA executes a Brownian-motion-based search to enhance the positions of each candidate hyperparameter set of the SE-Res-Dense U-Net++ model in accordance with the Langevin-inspired position rule.

$$\mu \ddot{q} = -\frac{dw}{dq} - \int_0^t q'(\tau) \zeta(t-\tau) d + R(t) \quad (10)$$

$$\mu \ddot{q} = -\zeta o \dot{q} + R(t) \quad (11)$$

The equation of motion only involves acceleration $\ddot{q}$ and velocity, it is rewritten using $v = \dot{q}$, where the dynamic friction kernel is represented with $\zeta(t)$ and random force is designated as $R(t)$; a diffusion constant D is also introduced for scaling the stochastic exploration component while the hyperparameters are being tuned.

$$\mu\dot{v} = -\zeta ov + R(t) \quad (12)$$

$$\mu\dot{v} = -\zeta ov + D \cdot R(t) \quad (13)$$

Subsequently, update the current position by exchanging random movement defined by the Langevin equation. For the purposes of this work, set $\zeta(t)$ to equal 1, thus equation is simply be:

$$x_i(t+1) = x_{best} - v + D \cdot R(t) \quad (14)$$

$$v = x_t(t) - x_{best} \quad (15)$$

$$R(t) = x_r(t) - x_i(t) \quad (16)$$

Here $x_i(t+1)$ stands updated position of mantis shrimp in $(t+1)^{th}$ iteration; x_{best} stands best position; v is defined as difference between position $x_i(t)$ in the t iteration and x_{best} ; and $R(t)$ is the position difference between $x_r(t)$ and current vector $x_i(t)$.

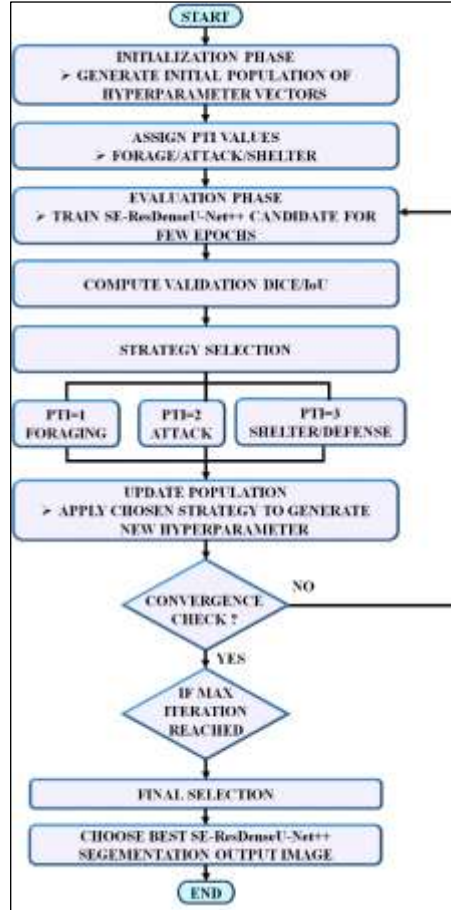


Figure 9: Flowchart for MSOA for tuning SE-Res-Dense U-Net++.

Source: Authors, (2026).

III.3.3 Attack Strategies

The mantis shrimp's strike is a very fast-moving strike with a high impact that break very hard shells. The mantis shrimp strike similarly with a circular parametric equation to inform the dynamic search behavior of the adaptive component when tuning the SE-ResDense U-Net++ model.

$$x_t = r * \cos\theta \quad (17)$$

Where the vector r is the exploratory radius in the search space, analogous to the mantis shrimp's front appendages, and θ is the angular motion of the strike. Finally, the mantis-inspired update rule for the tuning of the SE-ResDense U-Net++ model modelled as follows:

$$x_i(t+1) = x_{best} \cos\theta \quad (18)$$

Where $x_i(t+1)$ represents the updated candidate parameter set for SE-ResDense U-Net++ in the $(t+1)^{th}$ iteration, x_{best} is the best parameter set, and θ is an arbitrary angular factor in the range $0 \in [\pi, 2\pi]$.

III.3.4 Defense or Shelter Strategies

The defense–shelter behavior models an adaptive adjustment step involving avoidance of poor-performing solutions and an approach of good-performing solutions to stabilize and further tune SE-ResDense U-Net++.

$$\begin{cases} \text{Defense: } x_i(t+1) = x_{best} + k * (x_{best}) \\ \text{Shelter: } x_i(t+1) = x_{best} - k * (x_{best}) \end{cases} \quad (19)$$

Here, $x_i(t+1)$ represents the new solution for iteration $(t+1)$, x_{best} is the best solution, and k is a random scaling factor $k \in [0,0.3]$. The method allows for movement inward or outward from x_{best} in a controlled manner to support tuning while simulating defensive (convergent) or sheltering (divergent) behaviours the parameter. The proposed framework enhance the quality of current dental segmentation methods and produce a basis for incorporation into clinical dental diagnostic practices.

IV. RESULT AND DISCUSSION

The public dataset named “A Collection of Dental X-ray Images for Analysis” is used to assess the suggested MSOA-SE-ResDense U-Net++ model. This dataset is sourced from Kaggle and contains of 64 panoramic dental X-ray images in high resolution and in addition to that 64 segmentation masks representing ground-truth. These images are provided in JPEG/PNG formats and show the panorama of teeth, jawbone, and adjacent structure in detail. Accompanying each X-ray image is a binary mask that indicates the relevant teeth areas, making the dataset appropriate for supervised semantic segmentation. Each image reliability offers consistent quality across samples, high clarity of proteins anatomy, and yield adequate training/validation data for optimizing segmentation models specifically for dental images. This study provides a quantity of images and a number of segmentation masks to establish a benchmark for reliably objective and informative dataset models, validating whether segmentation output is sufficiently capturing the clinical anatomy.

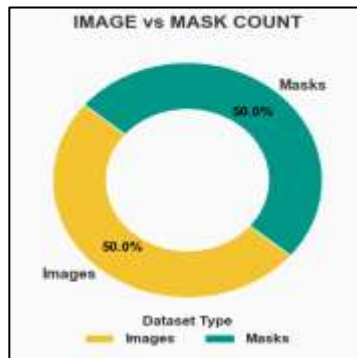


Figure 10: Distribution of Image Counts and Mask Counts.
Source: Authors, (2026).

Figure 10 depicts visualizes the distribution of dataset samples. The data is shown to consist of an equal number of 64 images and 64 segmentation masks, both accounting for 50% of the data. The yellow indicates the input X-ray images, and green indicates the ground-truth masks. This balanced distribution provides reassurance that every image in the dataset has a corresponding accurate annotated mask, establishing consistency and reliability for the supervised segmentation model training and evaluation.

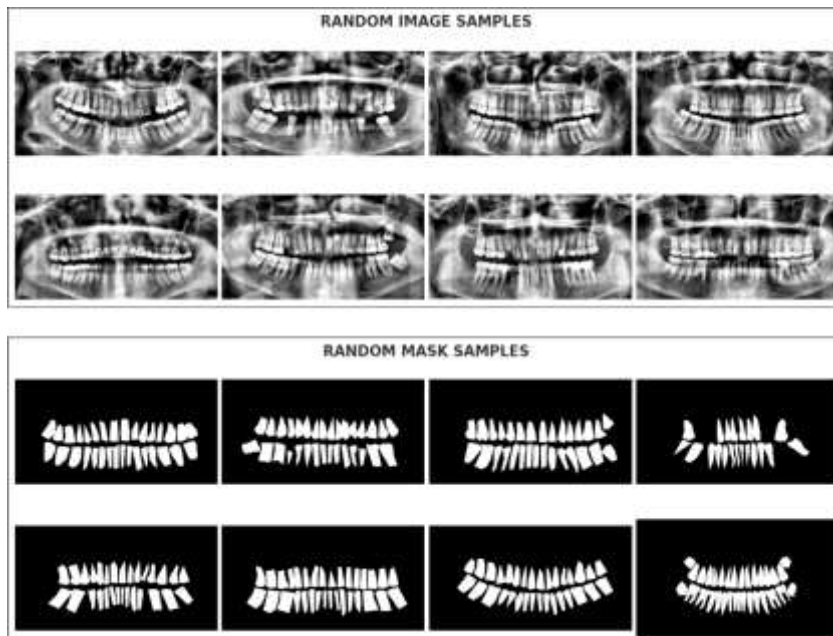


Figure 11: Panoramic X-ray Images and Radom Masks samples.
Source: Authors, (2026).

Figure 11 shows randomly selected examples from the dental X-ray dataset, which includes both the original panoramic radiographs and respective ground-truth segmentation masks. The first two rows show X-ray images with high resolutions that depict various dental structures and differences and variations in tooth alignment with restorations and missing teeth. The last two rows show the associated binary masks representing the annotated individual teeth and their boundaries. These highlight the variability of anatomical patterns in the dataset and provide segmentation challenges for developing a robust and generalizable DL models.

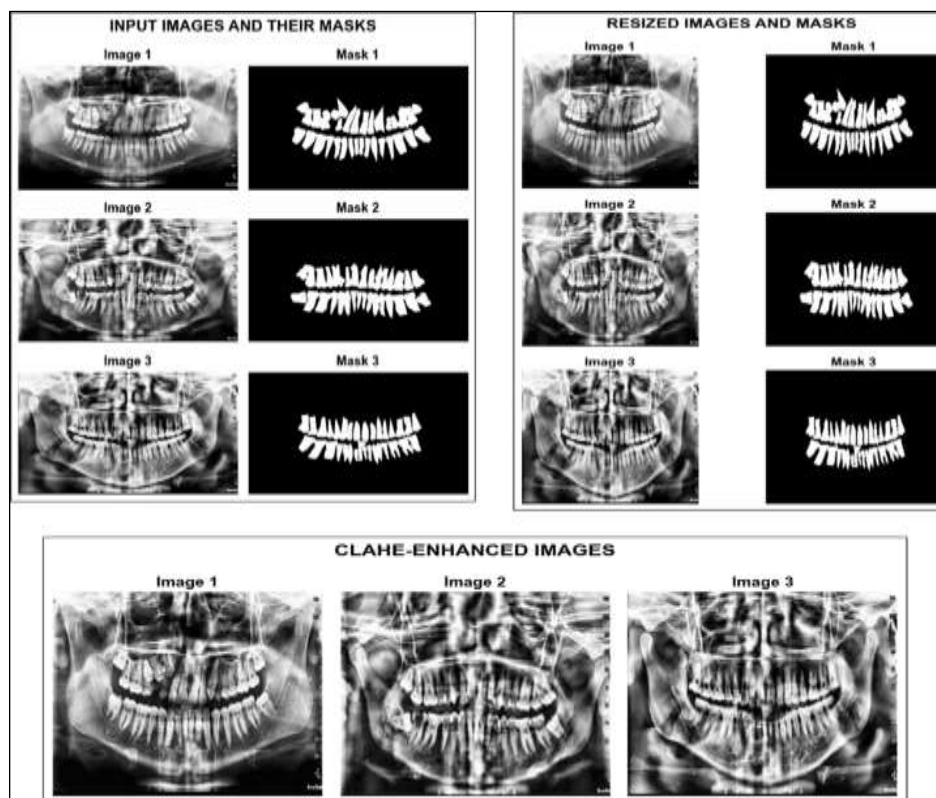


Figure 12: Pre-processing structure: Original, Resized, and CLAHE-Enhanced Dental X-ray Images.
Source: Authors, (2026).

Pre-processing applied to panoramic dental X-ray images before segmentation shows in Figure 12. The first block shows input images and associated ground-truth masks that represent original quality of the radiograph and annotated tooth structures. The second block provides the resized images and masks displayed in which the samples are normalized to have the same resolution to introduce uniformity to the model. The last block presents the images that are enhanced using CLAHE. The CLAHE enhancement helps improve contrast, introduces lighter features such as enamel edge, dentin edge, and tooth margins. This pre-processing ensures better visibility and prepares the data for the proposed MSOA-SE-Res-Dense U-Net++ segmentation model.

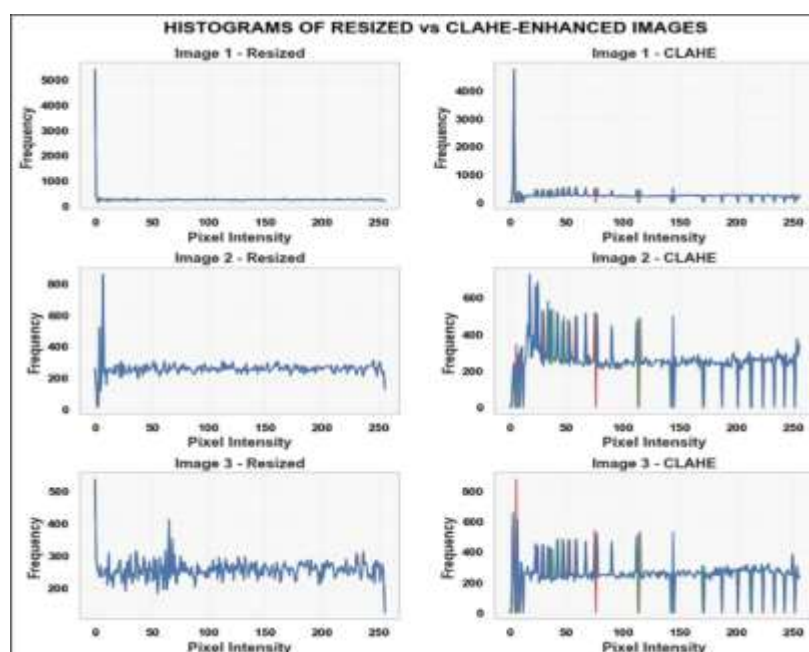


Figure 13: Histogram Comparison of Resized Images and CLAHE Enhanced Images.
Source: Authors, (2026).

Figure 13 represents the pixel intensity distribution of X-ray images with and without CLAHE enhancement. Histograms for the resized images seen in the left row. The pixel intensities for resized images are concentrated mostly in the lower pixel range with limited contrast and little structural visibility. The right histogram displays the histograms to CLAHE enhanced images with a more uniform and maximally expanded redistribution of pixel intensities throughout the dynamic range of pixel intensities. The redistribution of pixel values in the pixel intensity histogram demonstrates CLAHE's effectiveness at enhancing local contrast, increasing representation of small anatomical details of anatomical features, and increasing the visibility of tooth boundaries and more evident internal structures. The histograms clearly demonstrate the action CLAHE contributes toward improved feature representation prior to segmentation.

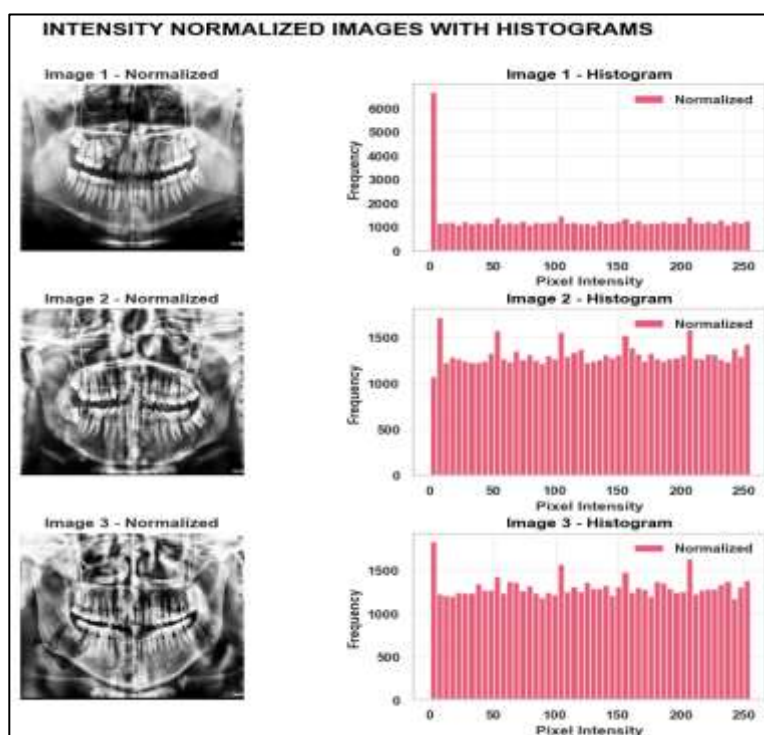


Figure 14: Intensity Normalized Images with Corresponding Histograms.
Source: Authors, (2026).

Figure 14 illustrates three example panoramic dental X-ray images after they have undergone the intensity normalization step in the pre-processing pipeline. The left images displays the normalized images, while right images displays consistent pixel intensity histograms for the images. The intensity normalization step redistributes the pixel values within a more uniform range, reducing the variability in exposure, and enhances the visibility of dental structures. The histograms demonstrate a more uniform and extended intensity distribution, representing improved stability and contrast throughout the dataset. The normalization ensures that the subsequent segmentations, where the MSOA-SE-Res-Dense U-Net++ architecture is used for consistent lighting condition.

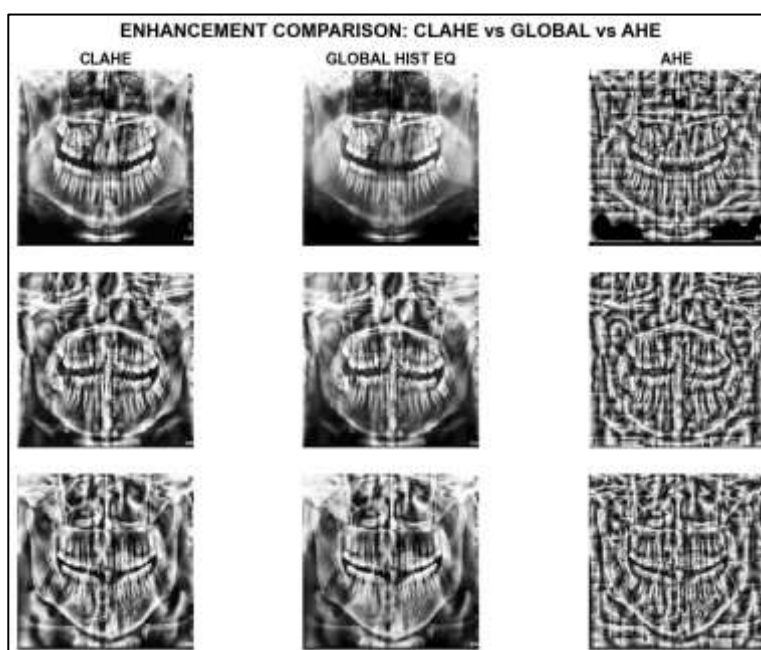


Figure 15: Comparative Image Enhancement Methods: CLAHE, Global Histogram Equalization, and AHE.
Source: Authors, (2026).

Figure 15 displays a qualitative comparison of three contrast enhancement methods including CLAHE, Global Histogram Equalization (Global HIST EQ), and Adaptive Histogram Equalization (AHE) applied to X-ray images. Each row represents a unique example of a dental x-ray, with the columns comparing the three enhancement techniques. CLAHE improved structural visibility and moderated contrast enhancement particularly around the enamel edge and tooth margins. Global HIST EQ improved global contrast but over brightness on the dense regions of the dental X-ray images. AHE increased local noise and grids, making the output undesirable for clinical-quality inputs. Overall, CLAHE provided the most consistent and diagnostically valuable enhancement of the three methods.

Table 1: Evaluation of Enhancement Filters for Pre-processing Panoramic Dental X-ray Images.

Filter	Entropy	CII	AMBE	NRM	RMS Contrast
CLACHE	8.41353	1.05709	3.25368	0.96264	79.70256
Global HIST EQ	7.60779	0.98367	4.71805	0.61124	74.19816
AHE	7.77999	0.99276	3.65274	0.89421	74.88383

Source: Authors, (2026).

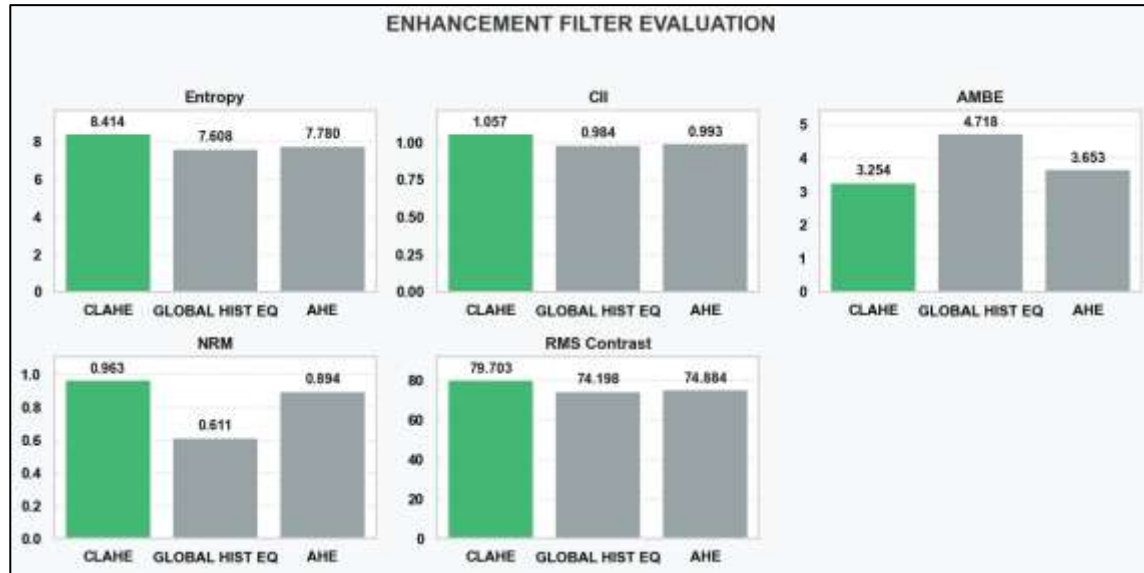


Figure 16: Quantitative evaluation of enhancement filters based on Entropy, CII, AMBE, NRM, and RMS Contrast measure.

Source: Authors, (2026).

Table 1 shows a comparative quantitative analysis of three image enhancement techniques, which are CLAHE, Global Histogram Equalization, and AHE using five commonly used performance metrics which are Entropy, Contrast Improvement Index (CII), Absolute Mean Brightness Error (AMBE), Normalized Root Mean (NRM), and RMS Contrast. CLAHE significantly provided better enhancement quality as demonstrated by higher entropy (8.414), higher CII (1.057), and best RMS contrast (79.703), which proves it provided better visibility of details and a greater distribution of contrast. To improve, CLAHE provided the lowest AMBE (3.254) confirming it distorted original brightness the least. In comparison, global histogram equalization demonstrated weaker performance due to lower NRM (0.611) and higher AMBE (4.718) values. AHE demonstrated excessive local amplification and moderated scores in all metric measures. The quantitative values confirm CLAHE pre-processing enhancement is the most successful enhancement technique for processing a panoramic dental x-ray. CLAHE produced a superior level of image quality for preprocessing compared to Global Hist EQ and AHE as depicted in Figure 16. Thus, CLAHE is selected as the enhancement technique used in the proposed MSOA-SE-Res-Dense U-Net++ segmentation framework.

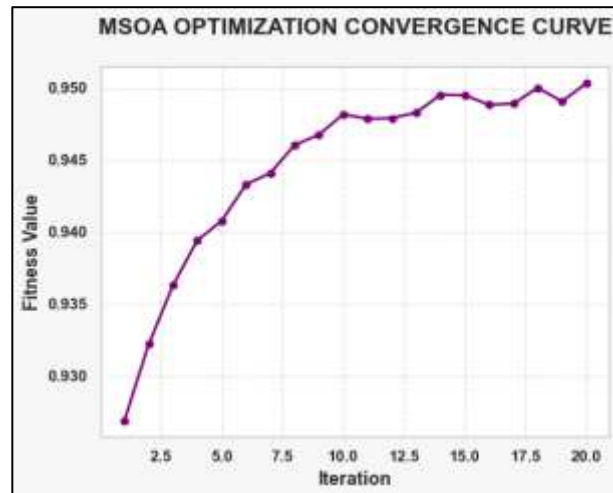


Figure 17: Convergence Curve of the MSOA.

Source: Authors, (2026).

Figure 17 shows the convergence behavior of the MSOA during the hyperparameter tuning process of the SE-ResDense U-Net++ segmentation model. The fitness value progressively improves through iterations, suggesting that the search space is explored and exploited successfully. There is a notable change in the fitness value during the initial iterations, followed by a more gradual approach to convergence in the fitness curve. The fitness values then approached their optimal fitness value (0.981), and the smooth and monotonic upward trend, suggesting that the algorithm converged in a reliable manner, while maintaining the capability to avoid converging prematurely, thus facilitating improved accuracy and generalization from segmentation.

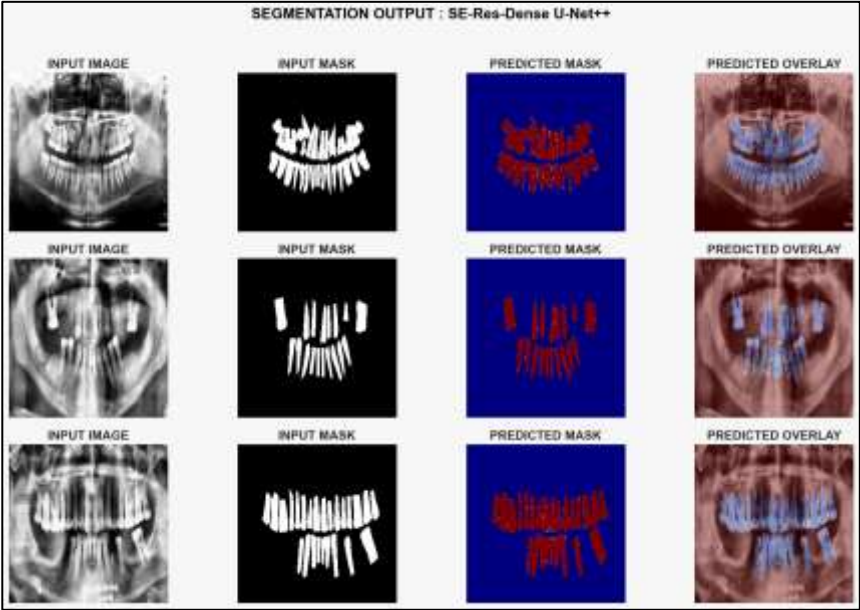


Figure 18: SE-Res-Dense U-Net++ segmentation output.
Source: Authors, (2026).

Figure 18 illustrates segmentation performance of SE-Res-Dense U-Net++ model on multiple samples. Each row corresponds to one test image, illustrating the entire segmentation pipeline from an input image to the model's prediction. The model demonstrated the ability to separate tooth contours, roots, and dental structure with high spatial fidelity, with good agreement with the ground-truth marks. The overlay results indicate that the model is capture fine-grain dental boundaries, even in regions containing background noise, overlapping anatomical structures, or low contrast.

Table 2: SE-Res-Dense U-Net++ Model Segmentation Performance Metrics.

SE-Res-Dense U-Net++	
Pixel Accuracy	0.980
Precision	0.965
Recall	0.955
F1-Score	0.965
Specificity	0.977
Dice	0.965
IoU	0.934

Source: Authors, (2026).

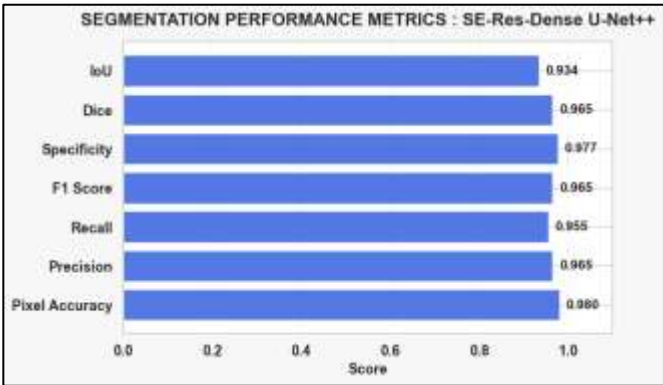


Figure 19: SE-Res-Dense U-Net++ Model Segmentation Performance Metrics.
Source: Authors, (2026).

Table 2 displays the quantitative segmentation performance metrics of the SE-Res-Dense U-Net++ model for segmentation. The model achieves robust results across evaluation metrics; IoU of 0.934, Dice score of 0.965, Specificity of 0.977, F1-score of 0.965, Recall of 0.955, Precision of 0.965, and Pixel Accuracy of 0.980. These values are substantially high and showcase the model's skill at accurately delineating tooth boundaries, limiting false positive and false negative classifications, and obtaining robust segmentation performance across different dental structures as illustrates in Figure 19.

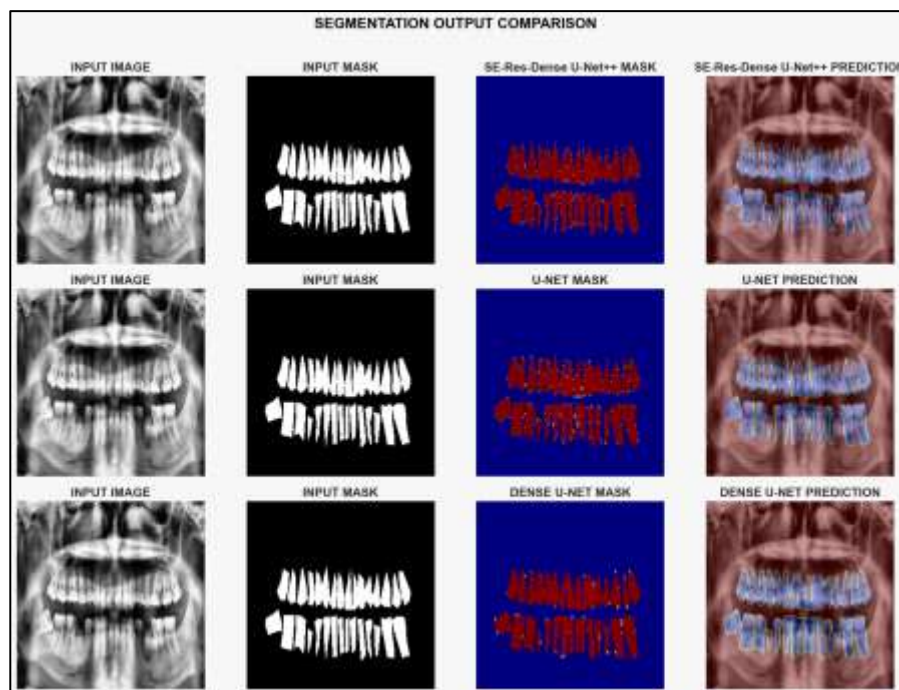


Figure 20: Segmentation Results across Various DL Models.
Source: Authors, (2026).

Figure 20 provides a visual comparison of segmentation results for three types of architectures: SE-Res-Dense U-Net++, U-Net, and Dense U-Net. Each input image displays the ground truth mask and output segmented image for each model. Proposed approach shows better precision for boundary segmentation along the tooth sites with clear separation between teeth compared to U-Net and Dense U-Net, which show softer edges and minor leakage across the boundaries. This part of the comparison illustrates the structure and robustness of proposed architecture.

Table 3: Segmentation Performance Metrics of U-Net Model.

U-Net	
Pixel Accuracy	0.960
F1-Score	0.961
Specificity	0.972
Dice	0.962
IoU	0.931
Precision	0.957
Recall	0.964

Source: Authors, (2026).

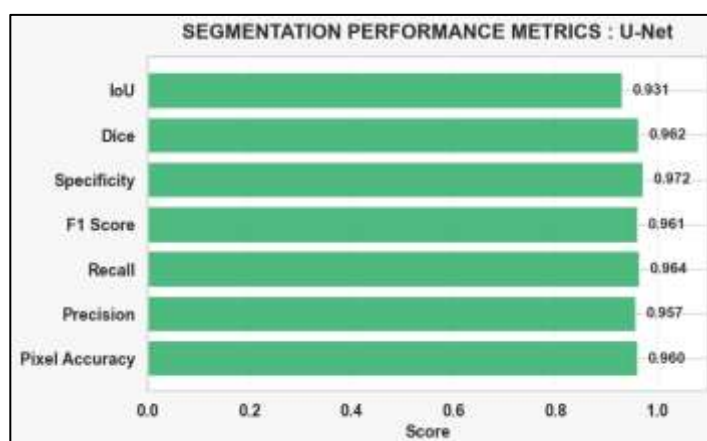


Figure 21: Segmentation Performance Metrics of the U-Net Model.
Source: Authors, (2026).

Summarizes segmentation performance of U-Net in Table 3 using evaluation metrics. The model achieves a Pixel Accuracy of 0.960, Precision of 0.957, Recall of 0.964, F1-score of 0.961, Specificity of 0.972, Dice coefficient of 0.962, and an IoU of 0.931. The results show in Figure 21 that U-Net has stable and reasonably accurate segmentation of tooth structures; however, it has lower performance compared to the proposed SE-Res-Dense U-Net++, which has better accuracy and boundary precision across all metrics.

Table 4: Dense U-Net Model Performance Metrics.

Dense U-Net	
Pixel Accuracy	0.940
Precision	0.936
Recall	0.945
F1-Score	0.940
Specificity	0.953
Dice	0.942
IoU	0.902

Source: Authors, (2026).



Figure 22: Segmentation Performance Metrics of the Dense U-Net Model.

Source: Authors, (2026).

Table 4 presents quantitative segmentation performance of Dense U-Net architecture produces an IoU of 0.902, Dice score of 0.942, Specificity of 0.953, F1 score of 0.940, Recall of 0.945, Precision of 0.936, and Pixel Accuracy of 0.940. These results shows in Figure 22, that Dense U-Net perform reliably in recognizing tooth structures, but when compared to the proposed SE-Res-Dense U-Net++, the model works lower across all metrics. This comparison validates that the proposed architecture creates improved segmentation accuracy and boundary preserving segmentation to more complicated dental structures.

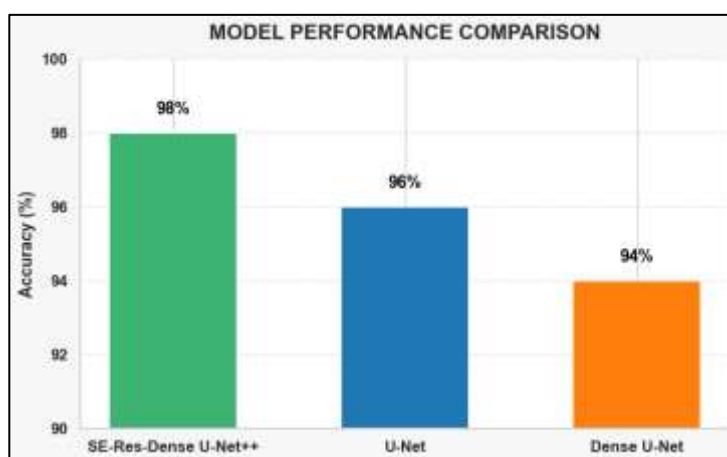


Figure 23: Evaluating Model Performance Using Segmentation Accuracy.

Source: Authors, (2026).

Figure 23 illustrates the segmentation accuracy for three DL models SE-Res-Dense U-Net++, U-Net, and Dense U-Net with the panoramic dental X-ray dataset. The SE-Res-Dense U-Net++ proposed achieves the highest accuracy of 98% and outperforms U-Net (96%) and Dense U-Net (94%). The outcomes are sufficient to show that using the combination of squeeze-and-excitation blocks, residual connections, dense feature propagation, and tuning of the hyperparameters using MSOA were beneficial to improved segmentation accuracy.

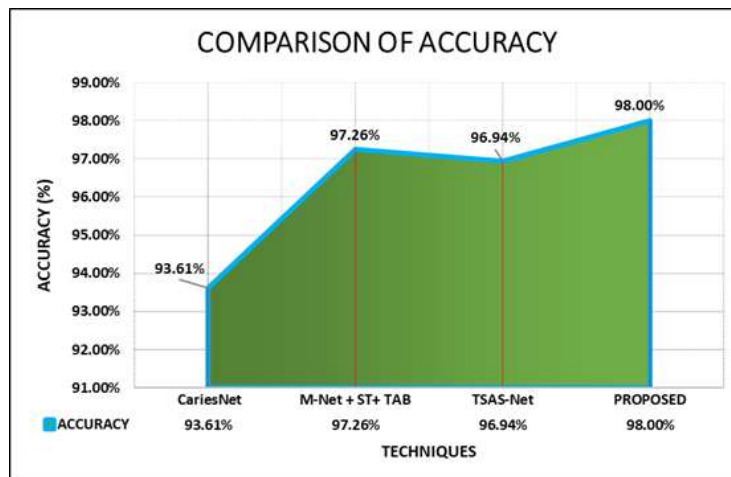


Figure 24: Accuracy Comparison of Different Classification Models.
Source: Authors, (2026).

In terms of accuracy, Figure 24 compared different classification models such as CariesNet, M-Net + ST + TAB, TSAS-Net and proposed model. The Proposed method is attain 98.00%, which shows that this method provides better performance than all other models listed above (M-Net + ST + TAB = 97.26%, TSAS-Net = 96.94%, CariesNet = 93.61%). The proposed method surpasses others because of its ability to maintain robustness over different image qualities and anatomical variances, enabling consistent segmentation accuracy even in difficult circumstances such as overlapping structures, restorations, and missing teeth.

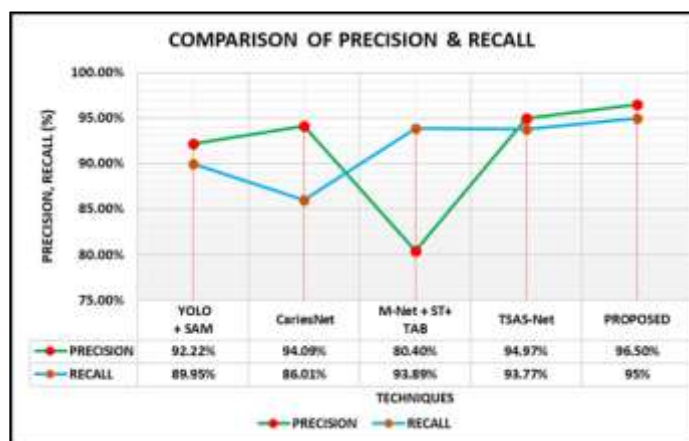


Figure 25: Visualizing Precision and Recall of Techniques.
Source: Authors, (2026).

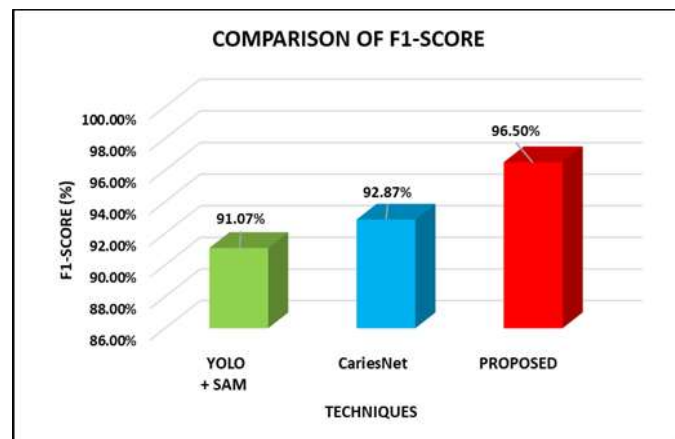


Figure 26: Visualizing F1-score of Techniques.
Source: Authors, (2026).

Figure 25 compares precision and recall of five different techniques: YOLO + SAM, CariesNet, M-Net + ST + TAB, TSAS-Net and proposed method. Proposed method exhibited the greatest outcomes compared to other models, with highest precision rates (96.50%) and recall rates (95.00%). Figure 26 presents a comparison of the F1-scores for 3 different detection techniques: YOLO+SAM, CariesNet and the Proposed Method. The proposed method attains 96.50%, the CariesNet and YOLO+SAM have 92.87% and 91.07%. It clearly shows proposed method is capable to identify more accurately and with a much more balanced detection ability than previously established methodologies.

V. CONCLUSION

This research presented a MSOA-SE-Res-Dense U-Net++ framework to produce efficient segmentation solution for panoramic dental X-Ray images. With the aid of MSOA-SE-Res-Dense U-Net++ segmentation, address several critical issues such as low contrast between tooth and background, overlapping anatomical structures, and variability of the anatomy within the jaw. Furthermore, it effectively addresses the problem of limited visibility of subtle dental structures. Pre-processing methods effectively maximizes the visibility of these subtle features and provides a platform for stable learning from a wide variety of datasets. The benchmarks used to demonstrate the effectiveness of proposed method yielded results of 98% accuracy, 96.5% precise, 95.5% recall, and 96.5% F1-score. This method received consistently better results than the current leading methods, confirming that proposed method achieves a superior trade-off between sensitivity and specificity and accurately delineate tooth contours, restorations, and jawbone structures with an excellent level of spatial fidelity. The MSOA-SE-Res-Dense U-Net++ model improves segmentation performance while creating a clear path towards incorporating technology into the real-world workflow of dental diagnosis, as well as enhanced clinical decision-making, treatment planning, and overall quality of care.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Sankar Murugesan, Sivaprakash P, Vimala Ithayan J, Dilip Charaan R M, Vijayanandh T, Santhoshkumar S P.
Methodology: Sivaprakash P, Vimala Ithayan J, Dilip Charaan R M, Vijayanandh T, Santhoshkumar S P.
Investigation: Sankar Murugesan, Sivaprakash P, Vimala Ithayan J.
Discussion of results: Sankar Murugesan, Sivaprakash P, Vimala Ithayan J, Dilip Charaan R M, Vijayanandh T, Santhoshkumar S P.
Writing – Original Draft: Sankar Murugesan.
Writing – Review and Editing: Sivaprakash P, Vimala Ithayan J, Dilip Charaan R M, Vijayanandh T, Santhoshkumar S P.
Resources: Sivaprakash P, Vimala Ithayan J, Dilip Charaan R M.
Supervision: Sivaprakash P, Vimala Ithayan J, Dilip Charaan R M, Vijayanandh T, Santhoshkumar S P.
Approval of the final text: Sankar Murugesan.

VII. REFERENCES

- [1] R. Zannah, M. Bashar, R. B. Mushfiq, A. Chakrabarty, S. Hossain, and Y. J. Jung, "Semantic segmentation on panoramic dental x-ray images using u-net architectures", IEEE Access, vol. 12, pp. 44598-44612, 2024.
- [2] S. Lin, X. Hao, Y. Liu, D. Yan, J. Liu, and M. Zhong, "Lightweight deep learning methods for panoramic dental X-ray image segmentation", Neural Computing and Applications, vol. 35, no. 11, pp. 8295-8306, 2023.
- [3] A. Haghanifar, M. M. Majdabadi, S. Haghanifar, Y. Choi, and S. B. Ko, "PaXNet: Tooth segmentation and dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier", Multimedia Tools and Applications, vol. 82, no. 18, pp. 27659-27679, 2023.
- [4] P. M. Jesi, and V. A. A. Daniel, "Differential CNN and KELM integration for accurate liver cancer detection", Biomedical Signal Processing and Control, vol. 95, pp. 106419, 2024.
- [5] S. R. Gadu, and C. S. Potala, "Vertebra Segmentation Based Vertebral Compression Fracture Determination from Reconstructed Spine X-Ray Images", International Journal of Electrical and Electronics Research, vol. 11, no. 4, pp. 1225-1239, 2023.
- [6] T. Ma, J. Li, Z. Dang, Y. Li, and Y. Li, "A Dual-Stream Dental Panoramic X-Ray Image Segmentation Method Based on Transformer Heterogeneous Feature Complementation", Technologies, vol. 13, no. 7, pp. 293, 2025.
- [7] B. Jaishankar, A. M. Ashwini, D. Vidyabharathi, and L. Raja, "A novel epilepsy seizure prediction model using deep learning and classification", Healthcare analytics, vol. 4, pp. 100222, 2023.
- [8] A. Bindhu, A. Ahilan, S. Vallisree, P. Maria Jesi, B. Muthu Kumar, N. K. Marriwala, and A. Q. M. Sabr, "Skin Cancer Diagnosis Using High-Performance Deep Learning Architectures", In International Conference on Emergent Converging Technologies and Biomedical Systems, pp. 693-703. Singapore: Springer Nature Singapore, 2023.
- [9] Y. S. Lim, D. Chun, J. Kim, J. Y. Lee, M. J. Ju, J. H. Park, and H. J. Jeon, "Dual convolutional neural network framework for segmenting dental caries in panoramic radiographs", The Journal of Prosthetic Dentistry, 2025.
- [10] A. Asaithambi, and V. Thamilarasi. "Classification of lung chest X-ray images using deep learning with efficient optimizers", In 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0465-0469. IEEE, 2023.
- [11] F. Oztekin, O. Katar, F. Sadak, M. Yildirim, H. Cakar, M. Aydogan, Z. Ozpolat, T. Talo Yildirim, O. Yildirim, O. Faust, and U. R. Acharya, "An explainable deep learning model to prediction dental caries using panoramic radiograph images", Diagnostics, vol. 13, no. 2, pp. 226, 2023.
- [12] D. Karthikeyan, R. Nivasini, K. T. Raja, D. Madhivadhani, B. Sivakumar, and A. K. Priya, "Hybrid Deep Learning Algorithm For MRI Brain Alzheimer's Disease Prediction", 2023 International Conference on Energy, Materials and Communication Engineering (ICEMCE), Madurai, India, pp. 1-6, 2023.
- [13] S. R. Gadu, "A Hybrid Generative Adversarial network with Quantum U-NET for 3D spine X-ray image registration", Healthcare Analytics, vol. 4, pp. 100251, 2023.
- [14] V. Thamilarasi, A. Asaithambi, and R. Roselin, "ENHANCED ENSEMBLE SEGMENTATION OF LUNG CHEST X-RAY IMAGES BY DENOISING AUTOENCODER AND CLAHE", ICTACT Journal on Image & Video Processing, vol. 15, no. 3, 2025.
- [15] S. Gadu, G. Himaja, Y. L. Vinanya, K. D. Sravya, K. S. Durga Aryan, Tasleem and C. A. Sai, "Detection of early-stage of lung cancer using machine learning classifiers", In AIP Conference Proceedings, vol. 3306, no. 1, p. 040001. AIP Publishing LLC, 2025.
- [16] A. P. Sunilkumar, B. K. Parida, and W. You, "Recent advances in dental panoramic x-ray synthesis and its clinical applications", IEEE Access, vol. 12, pp. 141032-141051, 2024.
- [17] R. Zannah, M. Bashar, R. B. Mushfiq, A. Chakrabarty, S. Hossain, and Y. J. Jung, "Semantic segmentation on panoramic dental x-ray images using u-net architectures", IEEE Access, vol. 12, pp. 44598-44612, 2024.
- [18] M. M. Khubrani, F. Jeribi, A. Tahir, and A. A. Metwally, "Panoramic Dental X-Ray Restorative Elements Segmentation using Hybrid Deep Learning", WSEAS Transactions on Computers, vol. 23, pp. 328-335, 2024.
- [19] G. Rubiu, M. Bologna, M. Cellina, M. Cè, D. Sala, R. Pagani, E. Mattavelli, D. Fazzini, S. Ibba, S. Papa, and M. Ali, "Teeth segmentation in panoramic dental X-ray using mask regional convolutional neural network", Applied Sciences, vol. 13, no. 13, pp. 7947, 2023.
- [20] Z. Kong, F. Xiong, C. Zhang, Z. Fu, M. Zhang, J. Weng, and M. Fan, "Automated maxillofacial segmentation in panoramic dental x-ray images using an efficient encoder-decoder network", Ieee Access, vol. 8, pp. 207822-207833, 2020.
- [21] H. Zhu, Z. Cao, L. Lian, G. Ye, H. Gao, and J. Wu, "CariesNet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image", Neural Computing and Applications, vol. 35, no. 22, pp. 16051-16059, 2023.
- [22] A. Ghafoor, S. Y. Moon, and B. Lee, "Multiclass Segmentation Using Teeth Attention Modules for Dental X-Ray Images", IEEE Access, vol. 11, pp. 123891-123903, 2023.
- [23] Y. Zhao, P. Li, C. Gao, Y. Liu, Q. Chen, F. Yang, and D. Meng, "TSASNet: Tooth segmentation on dental panoramic X-ray images by Two-Stage Attention Segmentation Network", Knowledge-Based Systems, vol. 206, pp. 106338, 2020.