Discover Applied Sciences

Brief Communication

Tooth segmentation in panoramic dental radiographs using deep convolution neural network -Insights from subjective analysis

Suvarna Bhat[1,2] · **Gajanan K. Birajdar**[1] · **Mukesh D. Patil**[1]

Received: 25 August 2024 / Accepted: 10 February 2025
Published online: 31 March 2025

Abstract

In the last few years, dentistry has witnessed a phenomenal advancement in artificial intelligence. The importance of teeth segmentation in dental radiographs has increased since it enables medical practitioners to conduct examinations more precisely and accurately in dentistry and helps them develop the most effective treatment strategy for their patients. In this research work, TUFT and UFBA dental data sets have been combined and used to train UNet, UNet ++ with ResNet 50 pre-trained model, UNet with mobileNet as encoder, and Deeplabv2 models for teeth segmentation. Evaluation of their performance in teeth segmentation using panoramic dental radiographs is carried out. Also, a subjective analysis of the model's predicted mask output from the practitioners is carried out. UNet++ with the combined data set and after fine-tunning hyperparameter gives the best results (IOU 0.8619 and Dice coefficient 0.9258). Also, it is observed that the use of the post-processing technique, 'Residual dense spatial-asymmetric attention' for deblurring the output images improved the result. According to the findings of the subjective study, the practitioner's satisfaction index is 4.2 on a scale of 5, which emphasizes the need for practitioners feedback in model building to ensure the clinical usability of the proposed system.

Article Highlights

- Our proposed work segments the teeth from the panoramic dental radiograph, which enables medical practitioners to do more exact and accurate dental examinations, as well as establish the most efficient treatment plan for their patients.
- The proposed methodology achieves 96.56 percent accuracy in segmenting teeth from panoramic dental radiographs using UNET++ model and post-processing technique.
- The subjective analysis highlights the importance of incorporating expert feedback into model development to ensure the clinical viability of AI systems in dentistry.

Keywords Medical image processing · Deep learning · Teeth segmentation

Gajanan K. Birajdar and Mukesh D. Patil authors have contributed equally to this work.

✉ Suvarna Bhat, suvarna.bhat@vit.edu.in; Gajanan K. Birajdar, gajanan.birajdar@rait.ac.in; Mukesh D. Patil, mukesh.patil@rait.ac.in | [1]Department of Electronics Engineering, Ramrao Adik Institute of Technology, D.Y. Patil deemed to be University, Nerul, Navi Mumbai, Maharashtra 400706, India. [2]Department of Computer Engineering, Vidyalankar Institute of Technology, Vidyalankar Marg, Mumbai, Maharashtra 400037, India.

1 Introduction

The research area of dentistry needs adaptive approaches and faster techniques of teeth segmentation, which is the primary stage in any diagnosis [1]. There are different types of radiographs, including occlusal, bitewing, periapical, and panoramic radiographs. Panoramic dental radiograph has gained popularity and importance as a diagnostic tool since its introduction. It is a specialized tomographic technique for providing a flat picture of the curved surface of the jaw. They are preliminary screening radiographs used to evaluate dentition and bone support, locate impacted teeth, and see where dental implants are placed [2]. They have advantages over other types of radiographs as they result in less radiation exposure, offer patients better comfort, and have a larger field of vision. For this research study panoramic dental radiograph dataset had been used.

The dentist can examine the entire tooth anatomy and develop a patient's treatment plan by examining panoramic radiographs. The evaluation is done empirically, meaning that the surgeon dentist directly affects their capacity to diagnose their patients through this exam, as there are no automated resources to assist the professionals in the study of the images. Tooth segmentation is essential for comprehending dental aesthetic patterns. Precise segmentation is prone to human mistakes since panoramic radiographs also show details originating from the bones of the facial and nasal areas [3]. Individual teeth in panoramic dental radiographs are difficult to segment due to overlapping teeth, variation in shape, size, and positions as well as noise and artifacts. The research has attempted to explore the possibility of designing practical algorithms that are general enough to be applicable in teeth segmentation [4].

The automation of teeth segmentation may be regarded as the first and most challenging step in the development of an Artificial Intelligence (AI) system capable of analyzing radiographs and distinguishing diseases from anatomical features. As a result, the initial step should be as precise as possible [5].

Deep learning is a novel approach for the segmentation of anatomical features and later in clinical decision-making in the case of teeth segmentation [6]. This advancement may improve virtual surgical planning, orthodontic planning, forensic identification [7], and the ability to distinguish normal anatomical structures from diseases. As a result, the combination of AI tools and expert analysis has the potential to revolutionize oral healthcare, allowing for truly precise and personalized dentistry. With computerized suggestions for complex situations, better treatment planning, and disease and outcome prediction, dentist's workflow will become more efficient [8].

In recent years, deep learning-based approaches have achieved amazing success in medical image segmentation, particularly dental image analysis. These approaches have the ability to automate the teeth segmentation task with high efficiency. However, considering the diversity and complexity of dental structures in panoramic radiographs, identifying a suitable deep-learning architecture for segmenting teeth remains a challenging task [9].

According to the literature study, deep learning models function better in healthy mouths with healthy teeth. Variations in patient-to-patient teeth, an artifact used for restoration and prostheses, poor image quality caused by some conditions such as noise, low contrast, homogeneity in a region that represents teeth and not teeth, space existing for missing teeth, and limitations of acquisition methods, which sometimes result in unsuccessful segmentation [10]. Because of factors such as the number of teeth per jaw, the proximity of neighboring tooth structures, the difference in density within a tooth, and tooth development, teeth segmentation is more difficult than bone structure segmentation [11]. As a result, the current study shows that there is still scope to identify an adequate approach for the segmentation of dental radiographs that may be utilized as a foundation for the construction of a specialized system to aid in dental diagnostics. This work aims to evaluate the segmentation of computational vision models and examine how well they segment the teeth on panoramic radiographs.

The paper is organized as follows Sect. 2 presents different methods used for teeth segmentation from the literature review. Section 3 explains the dataset and methodology used. Experimental analysis is mentioned in Sect. 4. Subjective analysis in shown in Sect. 5, and finally he result and discussion are covered in Sect. 6.

Main contribution of this paper:

- Utilization of the U-Net, UNet++, UNet with MobileNetV3 as the encoder, and DeepLabV3 combined with post-processing technique, to achieve enhanced accuracy in teeth segmentation.
- The proposed framework is systematically evaluated using different hyperparameters including varying batch sizes and different optimizers to optimize performance and ensure robust teeth segmentation results.

- The deep learning models were rigorously evaluated on TUFT, UFBA, and combined dental radiograph datasets, with performance metrics such as precision, recall, dice coefficient, and Intersection over union to assess segmentation accuracy.
- A subjective analysis of the predicted teeth-segmented mask was also carried out, incorporating evaluation and feedback from medical experts to validate the segmentation result.

2 Related work

The computerized analysis approach entails automatic teeth segmentation as an initial stage of processing [1]. Research demonstrates tooth segmentation has been used in a variety of situations, including significant dental treatments, varied brightness, and extreme teeth occlusion. Segmentation in dental imaging involves splitting and digitizing images into teeth pixels for easier processing and teeth detection [10]. The purpose of segmentation is to separate teeth from the background. The teeth segmentation task can be split between two studies: conventional methods that use existing knowledge and dental radiograph attributes and machine learning-based approaches that are data-driven.

Silva et al [10] employed a mask recurrent convolutional neural network (MASK R-CNN) to automate teeth extraction with a limited number of training dental radiographs. In [3], end-to-end neural network systems were proposed for tooth categorization, in which PANet architecture achieved the best result. A mask RCNN deep learning technique is used to recognize and localize the teeth on panoramic radiographs in the research study to automate tooth segmentation [12]. The teeth extraction method suggested in [1] achieved a good outcome by combining VGG-16 CNN and heuristic algorithm.

A two-stage attention segmentation network (TSASNET) on dental panoramic X-ray images to overcome the challenges caused by poor contrast and uneven intensity distribution in the tooth boundary and tooth root segmentation task is proposed in [13]. Authors of [14] used an innovative approach of automatic joint tooth segmentation using the pioneer UNet model, in this work authors proposed the use of a bounding box prior to the level of skip connection.

Authors of [15] introduced Swin-UNet, a transformer-based U-shaped encoder-decoder architecture with skip connections. They also introduced the PLAGH-BH dataset, which is used in the study to evaluate the teeth segmentation performance of Swin-UNet. Researchers are constantly looking for new approaches to solve challenges faced in teeth segmentation due to variations in patient-to-patient teeth, the presence of restoration, and poor image quality.

Only a few studies in the open literature have attempted to address the issue of deblurring medical images in recent years. The majority of these solutions use traditional image processing techniques to eliminate blurring by employing sharpening filters [16]. Image deblurring algorithms based on deep learning have improved significantly during the last decade. These methods have shown considerable improvement in the removal of non-medical images. Notably, deep deblurring algorithms beat traditional counterparts across a wide range of data sets[17]. To recover essential details while learning medical image deblurring, the authors [16] proposed a network, that includes a unique residual dense block with spatial-asymmetric attention.

From the literature study teeth segmentation in dental radiographs describes a variety of methodologies from conventional image processing techniques to advanced deep learning models, aimed at improving accuracy in complex dental radiographs. Deep learning frameworks such as MASK R-CNN, PANet, and UNet-based models (particularly Swin-UNet and TSASNET) have been successfully implemented, resulting in significant improvements in teeth localization, segmentation, and categorization. These models frequently use specialized architectures and novel strategies such as attention mechanisms and bounding box priors to improve performance on panoramic X-rays and other radiographs. Furthermore, deep learning-based deblurring algorithms have outperformed typical sharpening filters, notably in medical photos, by recovering finer details using techniques like residual dense blocks and asymmetric attention. In general, researchers are continuing to investigate and create robust segmentation and deblurring methods that are customized to the specific constraints of dental radiographs, with the goal of enhancing diagnostic accuracy and efficiency.

3 Material and methodology

3.1 Framework

In this research work, we aim to evaluate and contrast the effectiveness of four advanced deep learning architectures for teeth segmentation on panoramic dental radiographs: UNet, UNet++, UNet with mobileNet V3 encoder, and Deeplab.

These architectures were chosen based on their abilities to handle various aspects of segmentation task, such as fine boundary localization, and multi-scale contextual information, and computing efficiency. Employing these models research study determines the most effective model for teeth segmentation in challenging panoramic radiographs. Figure 1 shows the proposed system.

Experiment with two different datasets separately as well as when combined as shown in Fig. 1. In addition, as illustrated in Fig. 1, our suggested model was tested for several hyperparameter tuning, including different batch sizes, and optimizers. Further post-processing techniques were applied to enhance the result. Subjective analysis was employed to calculate the practitioner's satisfaction index.

3.2 Dataset

In this research work, we have used two datasets namely: the Tufts University Faculty of Technology School Dental Radiograph Dataset (TUFT) dental dataset [18] and Universidade Federal da Bahia-Universidade Estadual de Santa Cruz (UFBA-UESC) dataset [10]. Tuft's dental data set was published in december 2021 and is available on request. Radiographs were acquired at the Tuft University research center. The data set contains a total of 1000 panoramic dental radiographs along with labeled teeth masks. It is observed that the data set has various categories of radiographs such as radiographs with all 32 teeth, and radiographs with dental treatments like filling, implants, and root canals. There are some radiographs with no teeth, are also included in the dataset. The inclusion criteria for the dataset was the optimum diagnostic quality of the radiographs with no error or minimal error [18]. Another UFBA dental dataset is publicly available and it has 1500 panoramic radiographs of 10 different categories. These radiographs were acquired at the Universidade Estadual Sudoeste Bahia (UESB) [10]. Images are classified in terms of the structure of the teeth. Before passing data to models, all the radiographs were resized to 512 px height by 512 px width and normalized.

3.3 Different deep learning models for teeth segmentation

3.3.1 UNet

UNet is a widely recognized model for medical image segmentation that has shown effectiveness in a variety of medical imaging modalities. Its symmetric encoder-decoder architecture refer, Fig. 2 which includes skip connections, enables it to capture tiny details in radiographs and segment complicated structures such as teeth. Also, its potential to preserve spatial information through skip connections makes it ideal for segmenting complex dental structures in panoramic dental radiographs.

The UNet network can be divided into, a contracting path and an expansive path. As mentioned, the contracting path, which is the first section, employs the standard CNN design. A ReLU activation unit, a max pooling layer, and two consecutive 3x3 convolutions comprise each block of the contracting path. There are various iterations of this arrangement.

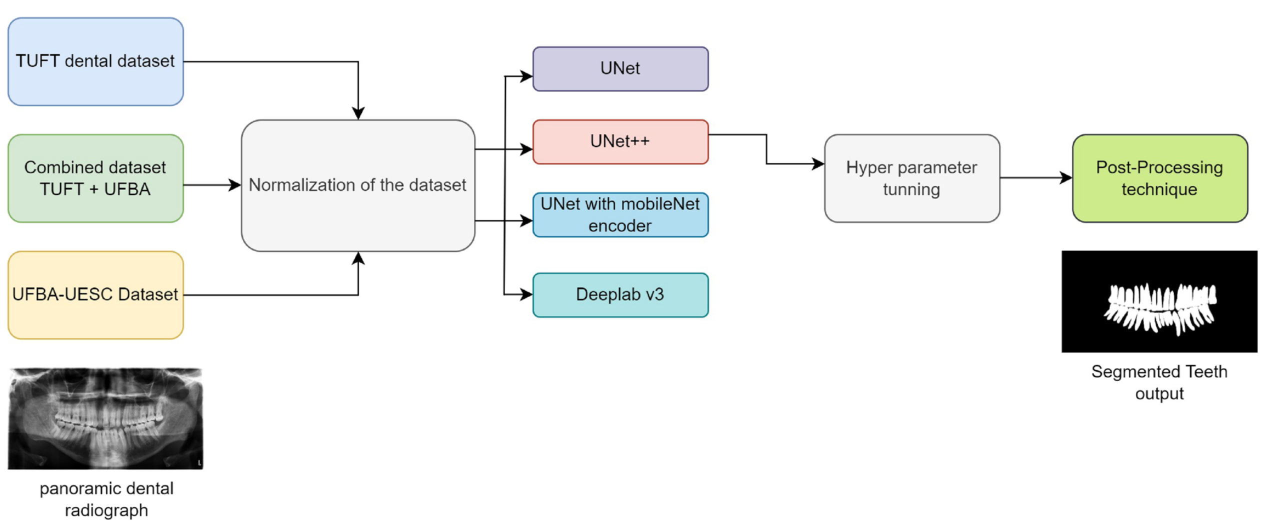


Fig. 1 Proposed framework for teeth segmentation

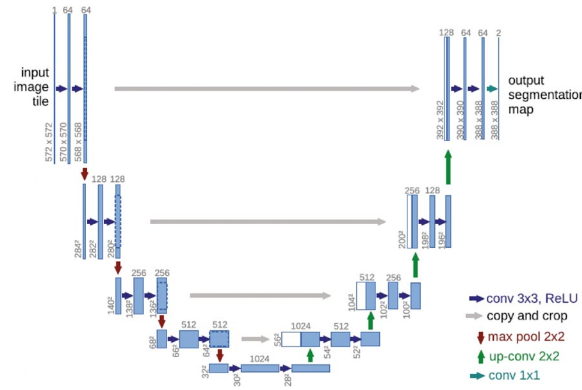


Fig. 2 U-Net architecture [19]

The unique aspect of U-Net is its expansive path, where the feature map is upsampled using a 2x2 up-conversion at each level. After that, the upsampled feature map is cropped and concatenated with the feature map from the matching layer in the contracting path. Two subsequent 3x3 convolutions and ReLu activation follow next [19].

The following equation provides the network's energy function:

$$E = \sum_{x \varepsilon \Omega} w(x) log(p_{l(x)}(X))[19]. \tag{1}$$

where Pk is the final feature map after the pixel wise SoftMax function has been applied [19].

$$Pk(x) = \frac{exp(ak(X))}{\sum_{k'=1}^{K} exp(a_{k'}(X))}[19]. \tag{2}$$

And ak (X) stands for channel k activation [19].

3.3.2 UNet++

Author's of [20] proposed a new modified UNet architecture known as UNet++ for better medical image segmentation as shown in Fig. 3. Its design consists of a deeply-supervised encoder-decoder network with nested, dense skip paths connecting the encoder and decoder sub-networks. The semantic gap between the feature maps of the encoder and decoder subnetworks is intended to be reduced by the newly developed skip paths. The main aim behind the UNet++ architecture is to bridge the semantic gap between the feature map of the encoder and decoder before the fusion [21]. It enables finer segmentation, especially in radiographs with complicated object boundaries. Panoramic dental radiographs frequently show densely packed teeth with modest borders. UNet++'s dense connection aids in better capturing these boundaries, resulting in more exact segmentation, which is required for correct dental analysis.

UNet++ consists of an encoder and decoder which are connected through a series of nested dense convolutional blocks. The dense convolutional blocks bring the semantic level of the encoder feature maps closer to that of the feature maps in the decoder. The deep supervision model operates in the two modes of operation i.e. accurate mode and fast mode. Inaccurate mode the average output is taken from all segmentation branches. In the fast mode, the final segmentation map is selected from only one segmentation branch, which then determines the model pruning extent and speed gain. To summarize, UNet++ differs from the original UNet in three ways: UNet++ has convolution layers on skip pathways, dense skip connections, and deep supervision. When the feature map is represented by x, the skip connection unit operates as follows, with i and j standing in for the index down the contracting path and across the skip connections, respectively.

$$x^{i,j} = \begin{cases} \mathcal{H}\left(x^{i-1,j}\right), & j = 0 \\ \mathcal{H}\left(\left[\left[x^{i,k}\right]_{k=0}^{j-1}, \mathcal{U}(x^{i+1,j-1})\right]\right), & j > 0 \end{cases}$$

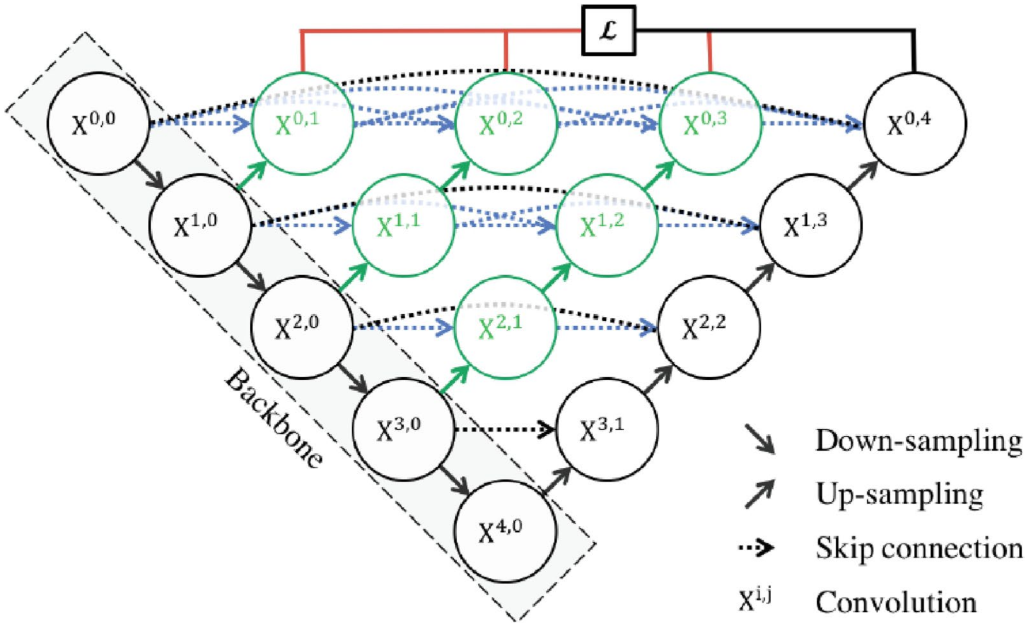


Fig. 3 UNet++ architecture [20]

3.3.3 U-Net-MobileNet V2

According to the original UNet architecture, the encoder extracts characteristics from the input images, and these resources are concatenated with the decoder so that the network performs the segmentation task. In these instances, using pre-trained networks in huge databases greatly aids segmentation because weights have already been trained with millions of learned resources and will not start from the beginning. Using mobileNet V3 as the encoder in the UNet architecture results in a lightweight and efficient model with high segmentation performance and lower computational costs.

Fig. 4 depicts the proposed UNet with MobileNet V2 model where pre-trained MobileNet V2 is applied as the encoder from our previous work. In this approach, the encoder will filter and learn the properties of radiographs that feed the network using compact depth convolutions. The inclusion of these inverted residual blocks reduces the number of

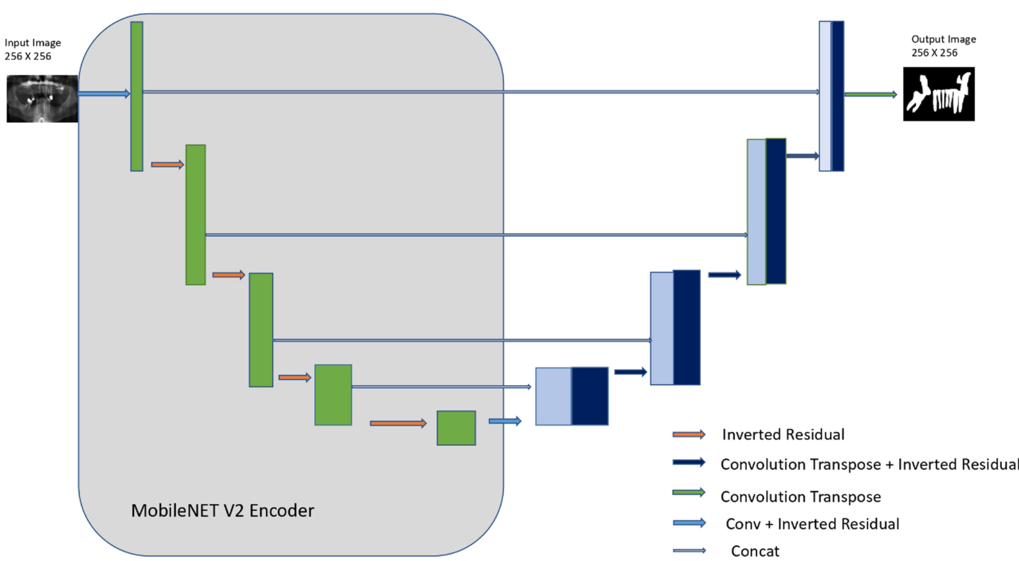


Fig. 4 UNet with MobileNet V2 as encoder

parameters and makes training the model easier and faster [22]. Up-sampling methods are employed in the decoding route to return the feature map to its original size.

UNet with MobileNet V3 balances accuracy and computational efficiency, making it suitable for deployment in resource-constrained situations such as clinics with limited computer power and high-resolution panoramic dental radiographs. This makes it suitable for practical and real-time applications in dental practices. Another advantage of the model is that the model will perform better and converge faster than it would without the usage of a pre-trained network.

3.3.4 DeeplabV3

DeeplabV3 excels in capturing contextual information using dilated convolutions and Atrous Spatial Pyramid Pooling (ASPP). This enables the model to properly handle varied item scales and intricate spatial found in panoramic radiography. Teeth on panoramic radiography exhibit great variation in size and shape. DeepLab V3's capacity to gather information across several scales improves its ability to reliably segment teeth, even when the structures are complicated or obscured by other anatomical features. The architecture refer Fig. 5 of DeepLabv3 consists of a backbone network for feature extraction (VGG, DenseNet, ResNet), an atrous convolution in the last few backbone blocks, an astrous spatial Pyramid Pooling network, to categorize each pixel with their correlating pixel-level classification, and a 1x1 convolution generates an output mask with the image's real size. This architecture significantly enhances performance on segmentation tasks by extracting dense feature maps for long-range contexts and capturing objects at various scales by exponentially expanding the receptive field without losing the spatial dimention[23].

3.4 Model training setup

Further, as per Fig. 1, all the deep learning models were trained on the training dataset, where we divided all datasets into 80:20 ratio for training and testing respectively. Also resized all images with size 512 px height by 512 px width before applying as input to all 4 deep learning models. The model training using a python programming setup is carried out on Google's cloud-based platform colab. The whole code is executed on a Windows 11 personal computer with an Intel Core i7 CPU and 16 GB of RAM. Google Colab is used to train, test, and validate models and features. The experiment analysis is carried out three times on the three distinct datasets indicated earlier. For post-processing our experimental details are as follows: According to authors of [16], there is no medical imaging data available with the pair of blurr and deblur output. Hence authors created a synthesized blurr dataset. We prepared the data set with a pair of ground truth and output image generated by the model.

3.5 Application of post-processing technique

Sharp edges are expected along the teeth's border and in the background area. To address this problem, it is crucial to investigate and improve the underlying causes of blur in the output images. One of the possible solutions to this problem is the optimization of the model, we have optimized the U-Net++ model and its parameters to ensure it captures the fine details and edges of the teeth accurately. Also, we have used a combination of two different source datasets to achieve diversity helped in enhancing the segmentation performance and reducing blur. Further, we planned to use the post-processing technique to mitigate the blur in segmented images.

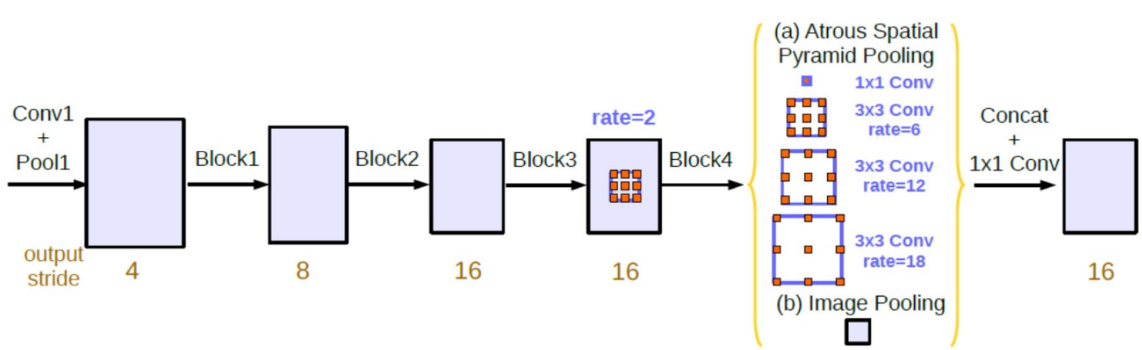


Fig. 5 Architecture of Deeplabv3 [23]

We conducted a literature review on different post-processing techniques to overcome blur issues. In the literature, it is observed that there are several techniques available to deblur the images. These techniques are mainly categorized under traditional methods and learning-based techniques. To restore the sharpness and clarity of the segmented teeth, image deblurring algorithm such as deconvolution and non-blind deblurring algorithms approaches can be used. These procedures can help to improve visual quality, refine tooth borders, and deliver more precise segmentation results.

We applied residual dense spatial-asymmetric attention proposed in [16] to deblur the output images obtained from the UNet++ model. The proposed approach attempts to convert blurry medical images (IB) into N: $I_B \rightarrow I_D$. Here the mapping function (N) learns to deblur medical image (I_B) as I_D. The network is divided into three sub-networks for learning medical image deblurring at various image resolutions. Individual branch output is upscaled and included in the subsequent sub-network as follows:

$$I^i h^i = N(I_B^i), (I_D^{i+1\uparrow}, h^{i+1\uparrow}; \theta)[16]. \tag{3}$$

I_B^i and I_D^i represents the blurry input and deblurred output of the i^{th} scale, respectively. θ represents the proposed network N's training parameters, while h represents the hidden states acquired. Authors of [16] proposes RD-SAM in each sub-network's encoder and decoder for precise feature learning. To accelerate residual features authors, connected encoder and decoder with a skip connection.

The suggested RD-SAM consists of a residual dense block followed by spatial-asymmetric in the following order:

$$X'_M = R_m(D_m(X))[16]. \tag{4}$$

In the equation 4, R(.) and D(.) residual dense block and spatial-asymetric block pursue local–global attentaion.

Objective function

The network N which is parameterized with weights W, tries to minimize training loss by appropriating the given P pairs of training images $\left\{ I_B^t, I_G^t \right\}_{t=1}^{p}$ as follows:

$$W *= argmin \frac{1}{p} \sum_{p}^{t=1} \mathcal{L}(G(I_M^t), I_G^t)\ [16]. \tag{5}$$

$\mathcal{L}$ signifies the perceived objective function, which is

$$\mathcal{L} = ||I_G - I_D||_1^2\ [16]. \tag{6}$$

Here I_G and I_D provide the original image as well as upgraded segmented radiograph.

3.6 Evaluation parameters

In this work, the deep learning model's performance for teeth segmentation in panoramic dental radiographs is assessed using both objective and subjective metrics. These metrics examine the segmentation result's accuracy, efficiency, and usability. The evaluation metrics used in this study are Pixel Accuracy (PA), Intersection over Union (IoU), and Dice Coefficient (Dice) [24]. While pixel accuracy is not a realistic criterion for assessing the true effectiveness of a segmentation model, it is nonetheless used as a general indicator. The pixel accuracy is computed by dividing the number of correctly identified pixels by the total number of pixels, as indicated in the formula 7. Dice coefficient (dice score) and IoU are typical measures for evaluating segmentation model performance. On the other hand, dice coefficient (also known as dice score) and IoU are two of the most often used measures for assessing segmentation model performance. The dice coefficient, as stated in formula 8, assesses the similarity between predicted and ground truth segmentation masks and is calculated twice the intersection of predicted and ground truth masks divided by the total of their areas. In formula A is the set of pixels in predicted segmentation and B is the set of pixels in the ground truth. Similarly, the intersection over union (IoU) in formula 9 evaluates how well the segmentation matches the ground truth. The value is determined by dividing the expected and ground truth masks by their union.

$$Accuracy = \frac{No\ of\ correct\ pixels}{Total\ number\ of\ pixels}[24]. \tag{7}$$

Table 1 Segmentation result on the Tuft dental panoramic dataset

Model	IOU	Disc coeficient	Accuracy
U-Net	0.8300	0.9121	0.9423
U-Net++	0.8242	0.9014	0.9323
U-Net-MobileNet V2	0.8134	0.8761	0.9265
DeeplabV3	0.5582	0.6412	0.6634

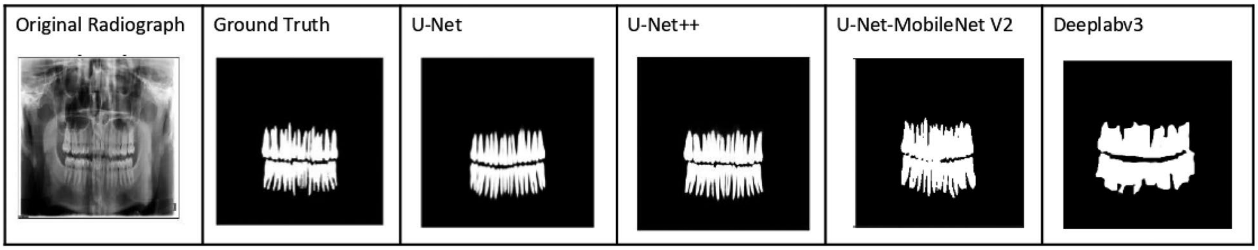


Fig. 6 Qualitative comparision between U-Net, U-Net++, U-Net-MobileNet V2 and Deeplabv3 showing teeth segmentation for TUFT panoramic radiograph dataset

$$Dice\ Coefficient = \frac{2|A \cap B|}{|A| + |B|}[24]. \tag{8}$$

$$IOU = \frac{|A \cap B|}{|A| \cup |B|}[24]. \tag{9}$$

4 Experimental results

4.1 Analysis using deep learning models on TUFT dental dataset

As shown in Fig. 1 all of the above-mentioned deep learning models are evaluated on the TUFT data set first. The data set consisted of 1000 dental panoramic radiographs and ground truth masks, which were split into training (80 percent) and validation (20 percent) sets. The models were trained on Googlecolab GPUs. For training, use of SGD optimizer with an initial learning rate of $1X10^{-3}$. UNet and UNet++ architectures with the ResNet50 as a backbone network modified UNet with mobileNet V2 as an encoder, and the Deeplabv3 model have been used. All the models are trained for 50 epochs with a batch size of 8. The most significant metrics for analyzing the accuracy of the segmentation algorithms, to evaluate the performance of each model, namely Intersection Over Union (IOU), Disc Coefficient, and Accuracy are used. Table 1 shows the results obtained after experimentation.

Based on the result mentioned in Table 1 and visual inspection of the output mask of the teeth segmentation by deep learning models shown in Fig. 6, for the Tuft dental dataset, UNet and UNet++ perform marginally better. We examined the UFBA dental dataset for all of the deep learning models stated above; in this case, UNet and UNet++ architectures outperformed. As a result, we concluded that conduction of further experimental analysis using solely UNet and UNet++.

4.2 Analysis using deep learning models on combined dental dataset

Further, experimental analysis with a combined dataset as mentioned in figure1 is carried out. In combined Tuft and UFBA dental datasets, a total of 2,500 radiographs are there. As both radiographs are from different sources, pre-processing step like resizing the images to 512X512 is conducted. In this step, the experiment is continued with UNet and UNet++ architectures. Both models use ResNet50 as the backbone network with weights learned from pre-training on Imagenet. This experiment continued with an SGD optimizer, 50 epochs, and a batch size of 4.

Table 2 Segmentation result on the combined dataset

Model	IOU	Disc coeficient	Accuracy	Precision	Recall
UNet	0.8613	0.9253	0.9653	0.9103	0.9414
UNet++	0.8619	0.9258	0.9656	0.9188	0.9337

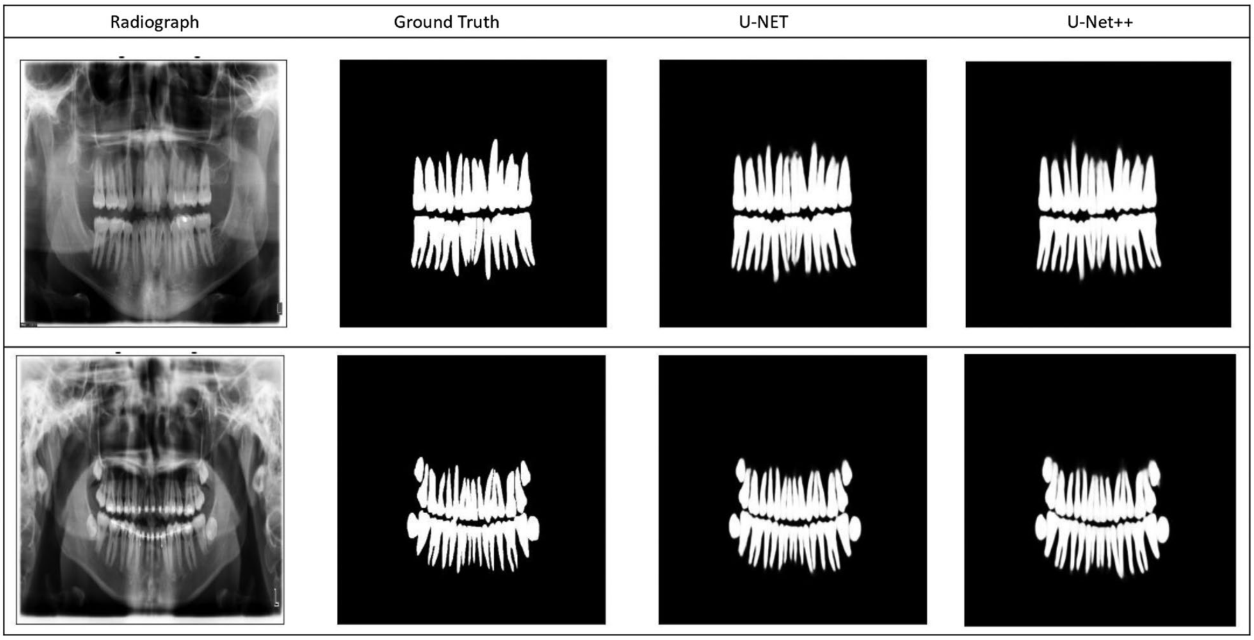


Fig. 7 Qualitative comparision between UNet, UNet++ showing teeth segmentation for combined panoramic radiograph dataset

To evaluate the performance of both the models and to analyze the accuracy of the segmentation, the most significant matrics with parameters namely Intersection Over Union, Disc coefficient, Accuracy, Precision, and Recall are calculated. The result of the experiment is shown in Table 2. Based on evaluation matrix parameters and visual inspection shown in Fig. 7 of the output mask produced by both modules UNet++ outperforms.

4.3 Evaluation of optimized UNet++ on combined dataset

We further tested the UNet++ model for several hyperparameter tunning instances to fine tune it.

4.3.1 Analysis of result based on diffrent batch sizes

Experiment with different batch sizes carried out. From the preceding phase, it was determined that the UNet++ model produces better results based on evaluation parameters and visual examination. However, during this study, we employed a common batch size of 8. To improve the result, now we will experiment with batch sizes 4, 8, and 16 for the UNet++ model and try to discover a suitable batch size based on the result. Figure 8 shows an analysis of training loss and accuracy for batch sizes 4, 8, and 16. Figure 8a shows training loss with batch size 4, the maximum loss value is 0.084, and over the number of epochs, it decreases. Figure 8b shows a training loss above 0.083 with batch size 8 and Fig. 8e shows a training loss close to 0.114 with batch size 16. Figure 8b,d, and f concludes that all the batch sizes have training accuracy above 0.97.

For different batch sizes i.e. 4,8 and 16 the evaluation parameters were examined for the validation dataset as illustrated in Fig. 9. This graph illustrates that batch size 8 performs slightly higher on almost all parameters including Dics coefficient, Precision, Recall, IOU, and Accuracy.

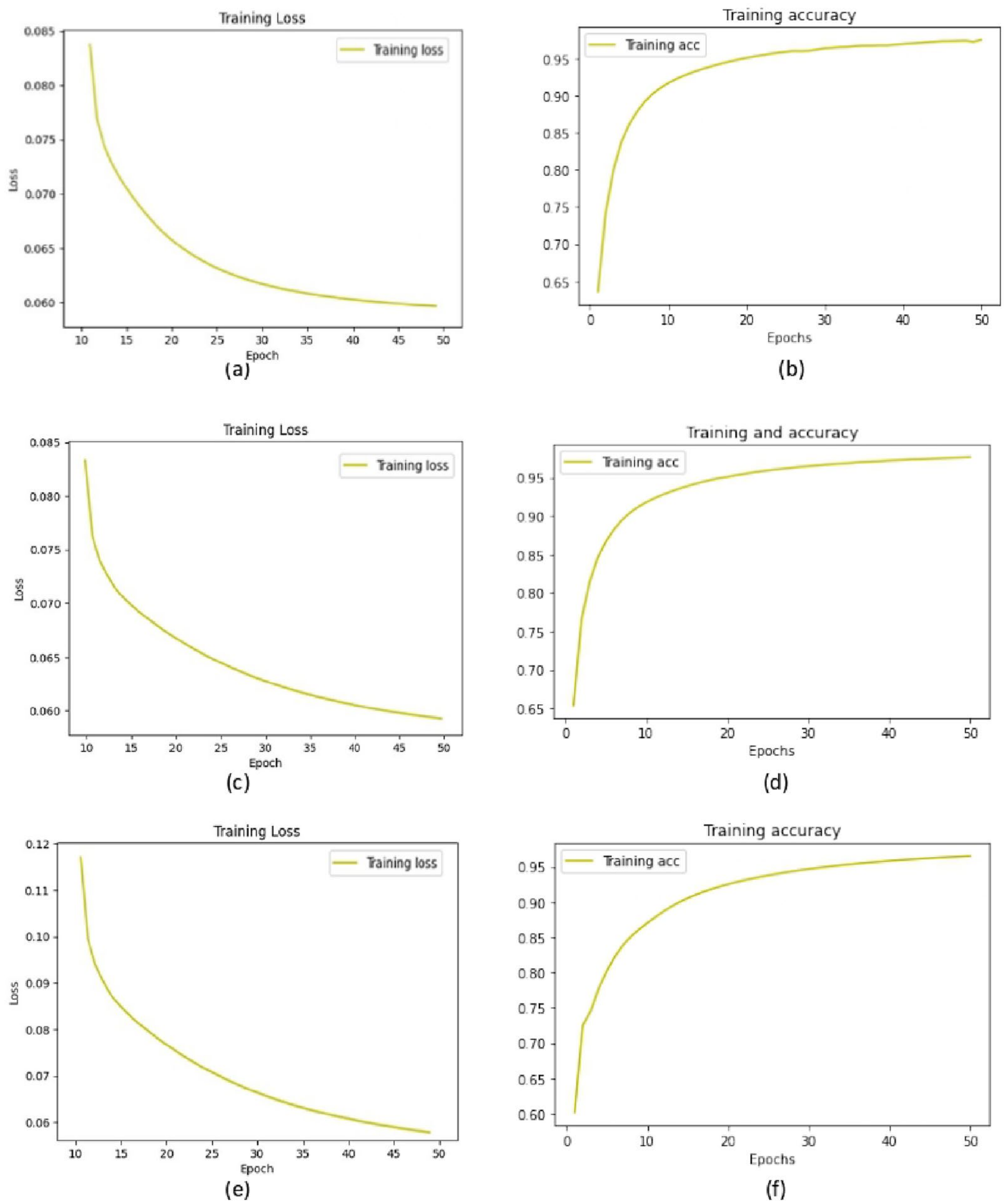
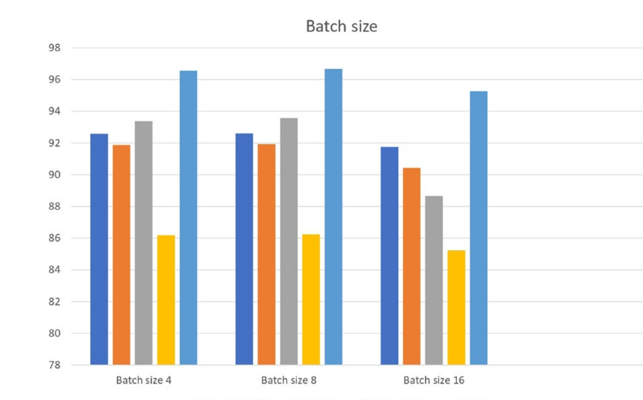


Fig. 8 Training loss and accuracy

Fig. 9 Evaluation parameters



4.3.2 Analysis of result based on different optimizers

This section contains the outcomes of SGD, Adam, and Nadam optimizers with epoch 50 and batch size of 8.

Figure 10a shows the training loss using the SGD optimizer, the maximum value is 0.0855, and over the number of epochs it decreases. Figure 10c shows the training loss using Adam optimizer, the maximum value is 0.084, and over the number of epochs it decreases and Fig. 10e shows training loss using Nadam optimizer, the maximum value is 0.087 it also decreases over the epochs. Figure 10b, d, and f all three optimizers have an accuracy of more than 0.95 but Adam optimizer has slightly higher accuracy i.e. 0.97. Hence in terms of training loss and accuracy, Adam optimizer outperforms.

Graph of Fig. 11 indicates that Adam optimizer outperforms all other parameters, including Disc coefficient, Precision, Recall, and Accuracy.

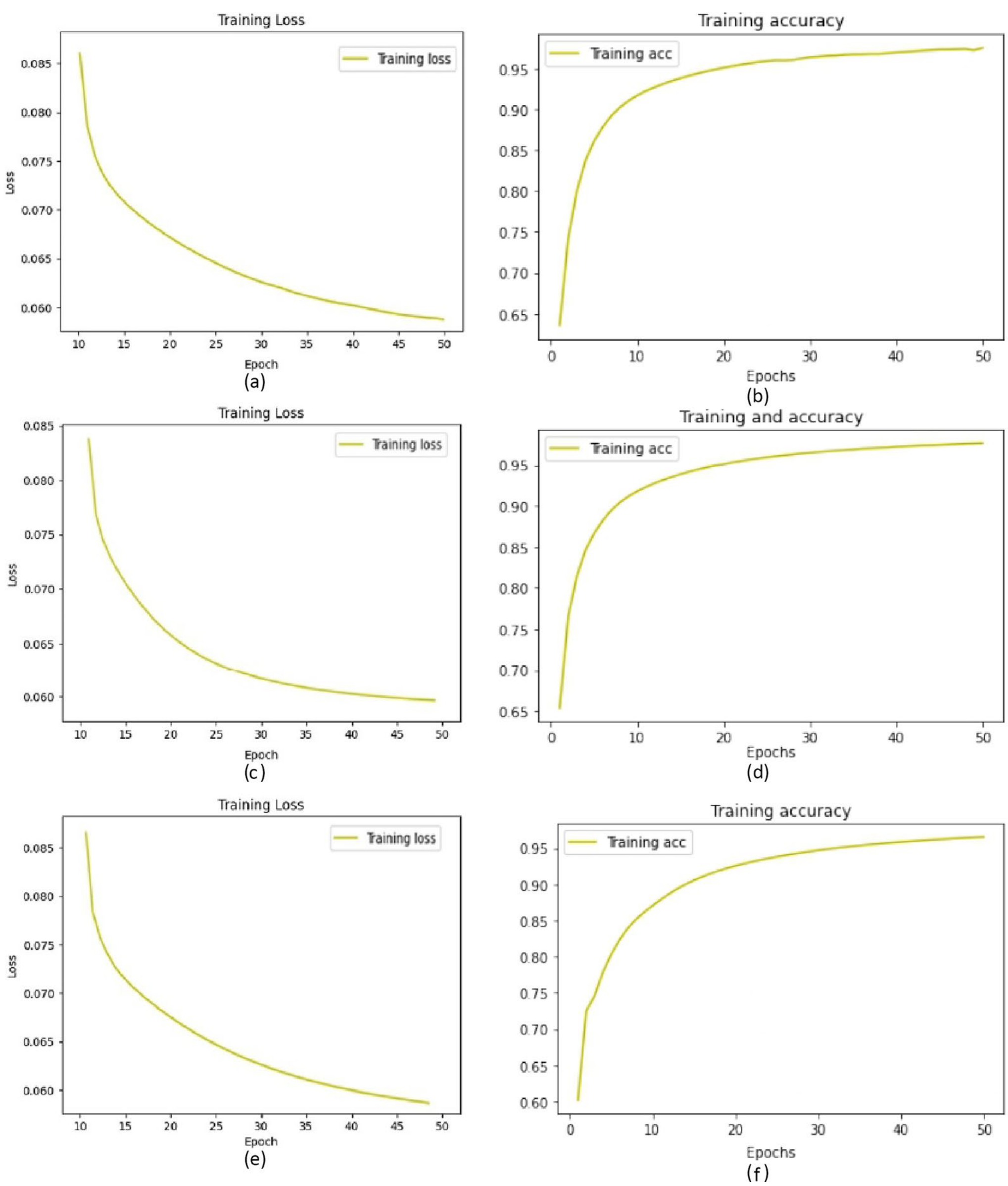
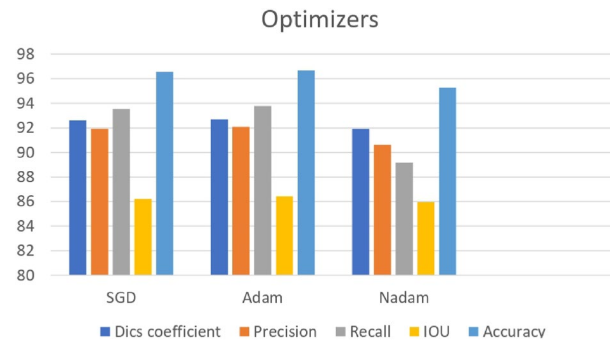


Fig. 10 Training loss and accuracy

Fig. 11 Evaluation parameters



4.4 Analysis of result after post processing technique

According to the preceding section's qualitative and visual inspection of the U-Net ++ model's output, the contour of the teeth area on the dental radiograph is somewhat blurry. We applied the residual dense spatial-asymmetric attention method to deblur the output. To measure the quantitative performance of the model, we evaluate the model's output with evaluation matrics PSNR and SSIM. PSNR, for example, assesses the signal-to-noise ratio, whereas SSIM evaluates structural information. We compare the result with an original output of the model and deblur output based on PSNR and SSIM evaluation metrics as shown in Table 3. Figure 12 shows a qualitative comparison of model output and deblur output, it is clear that deblur output exceeds model output in recovering salient information. In addition to recovery details, the deblur process can remove maximum blur while displaying no visually distracting artifacts.

Table 3 Quantitative compariosn between deblurr output and model output

Evaluation matrics	Model output	Deblur model output
PSNR	32.42	34.39
SSIM	0.61	0.64

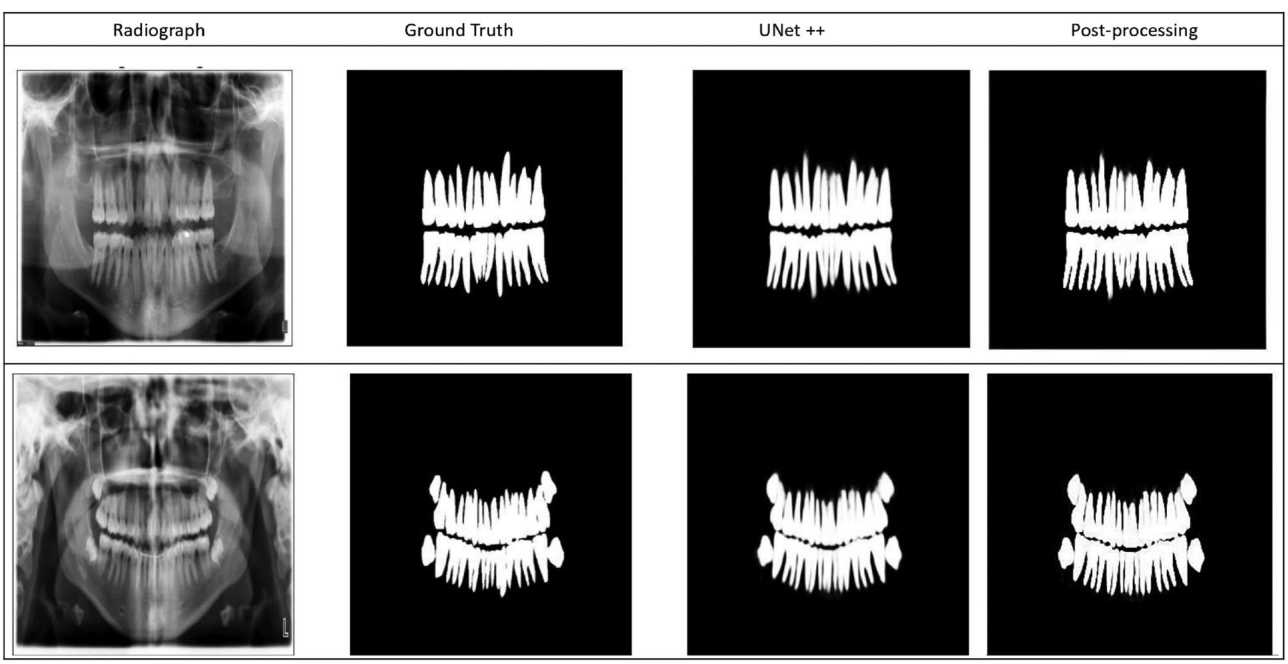


Fig. 12 Post-Processing output

Regular inspection and validation of the teeth segmentation deep learning model's performance are required to identify and address any potential sources of a blur. Collaboration between dental professionals and experts can contribute to constructing robust models that reduce blur and increase the overall accuracy and reliability of teeth segmentation in dental radiograph applications.

4.5 Comparison with state of art methods

After fine-tuning the UNet++ model comparison of the IOU, Disc coefficient, and accuracy with other state-of-art methods is carried out. The result analysis shown in Table 4 UNet++ model with a combined data set and fine-tunning of the model achieves improved overall accuracy compared to state-of-art approaches.

5 Subjective analysis

Subjective analysis for the output of the teeth segmentation deep learning model often entails visual inspection and evaluation of the segmentation findings. The result of the subjective analysis can be used to provide input for deep learning model improvement [26]. If there are any recurring issues or limitations identified, this insight can be used to refine the model architecture, training data, or other relevant parameters. Here the general approach, used to conduct the subjective analysis is mentioned. The earlier Sect. 4 established a parameter to evaluate the performance of the segmentation models. Also, Evaluation parameters for the overall segmentation performance are calculated. This step helps to provide objective measurements of the model's accuracy. While the quantitative measures provide valuable insight, subjective analysis is based on human judgment and professional knowledge. This approach, examines the segmentation result visually, taking into account aspects such as the clinical significance of segmentation accuracy, level of detail recorded, and overall segmentation quality.

The Satisfaction Index is intended to represent the subjective judgment of teeth segmentation performance. It assists in determining how effectively the automated segmentation satisfies the expectations and practical needs of experts. Because objective measurements (e.g., accuracy or dice coefficient) do not always fully convey dental segmentation accuracy, subjective feedback might provide valuable insights into the model's usability in real-world clinical contexts. A subjective analysis at the local dental clinic was conducted. First, we asked two dentists from the clinic to grade the output predicted mask of ten samples generated from the model on a scale of 0 to 5. The number of dentists included in conducting this subjective analysis was one senior and one junior dentist. Senior dentist with 20 years of experience and junior with 5 years of experience. The average score of satisfaction index for the test data received is 4.2. We also sought experts to provide feedback on expected output to determine the scope of improvement. We were told to try to distinguish between different grades of radiolucency and radiopacity.

Furthermore, we created a Google form with the explicit objective of the subjective analysis along with the models's 10 predicted sample output images and surveyed to solicit feedback from other dentists. The form responses provide the dentist's details, their year of experience, and their satisfaction index on the model's predicted output on a scale of 1 to 5. After floating the forms we received the following response.

As per the subjective analysis survey conducted, the mean years of dental experience of experts is 15.15 years and standard deviation is 8.12, which supports a rich, multi-dimensional approach to subjective analysis, leveraging diverse

Table 4 Comparison with state of art methods

Method	Dataset	IOU	Disc coeficient	Accuracy
Mask RCNN [10]	UFBA dataset	–	–	0.9208
UNet[25]	personal dataset	0.809	0.894	–
PANet [3]	UFBA-UESC	–	–	0.967
U-Net [18]	TUFT dataset	0.8662	0.9249	0.9519
SWin-UNet [15]	PLAGH-BH dataset	0.6956	–	0.8704
U-NET++	**TUFT and UFBA combined**	**0.8619**	**0.9258**	**0.9656**

UNet++ model gives improved results on the TUFT and UFBA combined dataset ascompared to others indicated in bold perspectives and balanced decision-making. The graph from Fig. 13 depicts the average of their ratings for each output image, the average satisfaction index is 4.087.

To compare the results of the UNET++ model with the expert's satisfaction index (subjective analysis) statistically, as we have the model's performance (e.g., Dice coefficient, IoU, accuracy) and the Satisfaction Index from human experts, we conducted paired comparisons to check if there's a significant difference between the U-Net++ segmentation performance and the human experts' subjective evaluations. 1st we conducted the normality test. As the data was normally distributed, we conducted the paired t-test, to compare model's performance. Obtained p-value for segmented output is greater than 0.05 in a statistical test, this means, the difference between the two groups being compared (model performance vs. human expert ratings) is not statistically significant.

6 Result and discussion

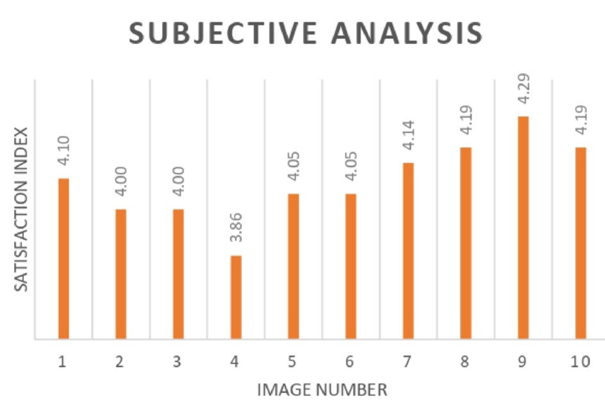
The dentist can analyze the overall dental structure and create the patient's treatment plan based on the images received with an X-ray. However, due to the lack of suitable automated tools to aid in the interpretation of dental X-rays, the evaluation is done empirically, which is based on the dentist's expertise. The complexity of analyzing dental X-rays remains high when dealing with extra-oral radiographs because these radiographs are not limited to a single section of teeth, as is the case of intra-oral radiographs. Extraoral radiographs show the temporomandibular regions and details formed by bones of the nose and face area, in addition to teeth. This process is significantly more challenging on panoramic radiographs because of its inherent limitation [5]. In our work, by combining objective and subjective analysis, we obtained a comprehensive evaluation of the teeth segmentation deep learning model and iteratively refined it for better results. Our methodology outperformed a newly established system in terms of teeth segmentation.

We have considerably enhanced convergence time and final accuracy for the segmentation challenge by combining two different data sets and fine-tuning the model with a post-processing application.

In our combined dataset, the UFBA data set of 1500 radiographs published by authors of [10] marked substantial differences when compared to other previously used datasets. It contains a wide range of radiographs with various attributes that are organized into ten categories based on the following general characteristics: variation in tooth structure, number of teeth, presence of restorations, presence of dental implants, presence of dental appliance, and presence of radiographs with more than 32 teeth. [10]. Secondly in the Tuft dental data set clinicians prescribed all of the digital panoramic radiographs used based on the patient's diagnostic and clinical needs [18]. This dataset includes the different categories of abnormality which include inflammation, dysplasia, developmental, benign tumor or cyst, etc [18]. Above mentioned two different types of dataset combinations, make it a diverse and large dataset which is very important to training the deep learning model. A broad data set ensures that the deep learning model experiences a diverse variety of instances, capturing distinct imaging modalities. This allows the model to learn to generalize more effectively and generate more accurate predictions on previously unseen data. Also deep learning model becomes more adaptive and less prone to overfitting when trained with big and diverse datasets. By introducing the model to a variety of situations, it learns to extract key properties that are constant across multiple instances of the same conditions, resulting in improved generalization.

As mentioned in Sect. 3 we selected 4 different deep learning models for teeth segmentation. Further analyzing the result of the combined data set we concluded that UNet ++ gives better results. Next, we fine-tune the model

Fig. 13 Dentist's satisfaction index



by setting the hyperparameters. Due to the combined two datasets model was trained on diverse datasets. The data sets included a different variety of images. In our observation, there were panoramic radiographs with low, medium, and high contrast. In visual inspection, medium contrast gives the best results. This large data set also contained radiographs with only foreign parts present in the mouth like only implants are present, and radiographs without teeth. However model has shown satisfactory output for these cases. Due to a lack of clarity in the boundaries, two central incisors are frequently missed. This lack of clarity, which is induced by the imaging process, is frequent for central incisors in panoramic images [9]. Furthermore, tooth filling in both data sets influences tooth segmentation accuracy, especially when the fillings do not match the real profile of the teeth. But in both cases model has shown satisfactory results.

The blurred output radiographs of the U-Net++ model may make precisely detecting and defining the dental borders challenging. This blurriness can be attributable to a variety of variables, and the limitation of the deep learning model may be one of the reasons. The blurring effect may result in incorrect tooth segmentation and jeopardize the dependability of subsequent dental examinations and treatment. We aimed to utilize a post-processing approach to reduce blur in segmented radiographs. To identify and address any potential sources of blur, the performance of the teeth segmentation deep learning model must be evaluated and validated on a regular basis.

Collaboration among dental professionals and specialists can help to build robust models that can reduce blur and improve the overall accuracy and the reliability of teeth segmentation in dental radiographs. Therefore, subjective analysis of the deep learning model's segmented output in dental radiographs is essential for addressing their clinical viability and potential improvements. The satisfaction index of subjective analysis indicates the model correctly identifies and segments the teeth from the background. Dentists were pleased to see model maintains accuracy for diverse and complex instances, which include challenging anatomical variations. Also, it was advised to conduct a clinical validation study where the model's segmented output is used in real clinical settings. From the discussion with experts, it was concluded that deep learning models may require continuous monitoring and retraining to adapt to changing clinical scenarios.

Statistical analysis of human expert and AI model, implies that the model performs on par with human experts, which is a positive result in cases where we want the AI system to match expert-level performance. Both the models are independently performing equally well.

7 Conclusion and future scope

On panoramic dental radiographs, we successfully demonstrated the feasibility of the UNet++ deep learning model for teeth segmentation along with hyperparameter tuning and the use of the post-processing technique. UNet ++ also outperforms state-of-art methods listed in Table 4. The subjective analysis emphasises the significance of using expert feedback in model building to ensure the professional usability of AI systems in dentistry. The current development will serve as the foundation for future advances in AI-driven technologies that will accurately and automatically identify numerous dental problems, saving time and reducing human errors. The use of more resilient and understandable technology is a critical step towards a more precise and personalized dental practice.

Strategies should be developed to build AI datasets that are gathered, curated, and made available to researchers worldwide in a systematic, secure, and ethical manner. The data set annotation process must be governed through the development of standardized software and tools so that dataset properties are uniform and can be used by researchers worldwide.

Author Contributions S. Bhat conceptualized the research idea and designed the overall study, with input from G.K Birajdar and M.D. Patil. S. Bhat and G. K. Birajdar developed the methodology, including the adaptation of the model for the experiments and data collection. S. Bhat and M.D. Patil conducted the data analysis and interpretation of the results. S. Bhat drafted the initial manuscript, and G. K. Birajdar and M.D. Patil contributed to reviewing and editing the final draft. All authors reviewed and approved the final manuscript for submission.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. Authors declare they have no financial interests.

Data Availability The data presented in this study are available on request from the corresponding author.

Code Availability The code can be made available based on request.

Declarations

Competing interests The authors declare that they have no competing interests.

References

1. Tuzoff DV, Tuzova LN, Bornstein MM, Krasnov AS, Kharchenko MA, Nikolenko SI, Sveshnikov MM, Bednenko GB. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofacial Radiol. 2019;48(4):20180051.
2. Shah N, Bansal N, Logani A. Recent advances in imaging technologies in dentistry. World J Radiol. 2014;6(10):794.
3. Silva, B., Pinheiro, L., Oliveira, L., Pithon, M.: A study on tooth segmentation and numbering using end-to-end deep neural networks. In: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2020:164–171
4. Schwendicke Fa, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2020;99(7):769–74.
5. Leite AF, Gerven AV, Willems H, Beznik T, Lahoud P, Gaêta-Araujo H, Vranckx M, Jacobs R. Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs. Clin Oral Invest. 2021;25:2257–67.
6. Vinayahalingam S, Xi T, Bergé S, Maal T, De Jong G. Automated detection of third molars and mandibular nerve by deep learning. Sci Rep. 2019;9(1):9007.
7. Vijayakumari B, Vidhya S, Saranya J. Deep learning-based gender classification with dental x-ray images. Int J Biomed Eng Technol. 2023;42(1):109–21.
8. Lee J-H, Kim D-H, Jeong S-N, Choi S-H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 2018;77:106–11.
9. Haghanifar A, Majdabadi MM, Ko S-B. Automated teeth extraction from dental panoramic x-ray images using genetic algorithm. In: 2020 IEEE International Symposium on Circuits and Systems (ISCAS), 2020;1–5
10. Silva G, Oliveira L, Pithon M. Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives. Expert Syst Appl. 2018;107:15–31.
11. Shaheen E, Khalil W, Ezeldeen M, Casteele E, Sun Y, Politis C, Jacobs R. Accuracy of segmentation of tooth structures using 3 different cbct machines. Oral Surg Oral Med Oral Pathol Oral Radiol. 2017;123(1):123–8.
12. Lee J-H, Han S-S, Kim YH, Lee C, Kim I. Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. Oral Surg Oral Med Oral Pathol Oral Radiol. 2020;129(6):635–42.
13. Zhao Y, Li P, Gao C, Liu Y, Chen Q, Yang F, Meng D. Tsasnet: Tooth segmentation on dental panoramic x-ray images by two-stage attention segmentation network. Knowl-Based Syst. 2020;206: 106338.
14. Nader R, Smorodin A, De La Fourniere N, Amouriq Y, Autrusseau F. Automatic teeth segmentation on panoramic x-rays using deep neural networks. In: 2022 26th International Conference on Pattern Recognition (ICPR), 2022;4299–4305
15. Sheng C, Wang L, Huang Z, Wang T, Guo Y, Hou W, Xu L, Wang J, Yan X. Transformer-based deep learning network for tooth segmentation on panoramic radiographs. J Syst Sci Complexity. 2023;36(1):257–72.
16. Sharif S, Naqvi RA, Mehmood Z, Hussain J, Ali A, Lee S-W. Meddeblur: medical image deblurring with residual dense spatial-asymmetric attention. Mathematics. 2022;11(1):115.
17. Zhang K, Ren W, Luo W, Lai W-S, Stenger B, Yang M-H, Li H. Deep image deblurring: a survey. Int J Comput Vision. 2022;130(9):2103–30.
18. Panetta K, Rajendran R, Ramesh A, Rao SP, Agaian S. Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems. IEEE J Biomed Health Inform. 2021;26(4):1650–9.
19. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 2015;234–241
20. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain 2018
21. Zhou J, Abdel-Mottaleb M. A content-based system for human identification based on bitewing dental x-ray images. Pattern Recogn. 2005;38(11):2132–42.
22. 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, 2018;4510–4520.
23. Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 2017.
24. Mahran AMH, Hussein W, Saber SEDM. Automatic teeth segmentation using attention u-net 2023.
25. Nishitani Y, Nakayama R, Hayashi D, Hizukuri A, Murata K. Segmentation of teeth in panoramic dental x-ray images using u-net with a loss function weighted on the tooth edge. Radiol Phys Technol. 2021;14:64–9.

26. Christiansen F, Epstein E, Smedberg E, Åkerlund M, Smith K, Epstein E. Ultrasound image analysis using deep neural networks for discriminating between benign and malignant ovarian tumors: comparison with expert subjective assessment. Ultrasound Obstetrics Gynecol. 2021;57(1):155–63.