International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Benchmarking Segmentation Models for Identifying Impacted Teeth in Panoramic Radiographs

G Harika[1], M.Bhagyalakshmi[2]

[1]PG Scholar, Department of Computer Science and Engineering, PVKK Institue of Technology Anantapuramu, India

[2]Assistant Professor, Department of Computer Science and Engineering, PVKK Institue of Technology Anantapuramu, India.

---***---

Abstract - The paper provides a new comparative performance of the pixel-based segmentation models to accurately identify the affected teeth in panoramic radiographical images. Using the analysis of the Dental Panoramic Radiographs dataset of Kaggle that provides the full set of images that will be specific to this task the work examines the work of five advanced segmentation algorithms: U-Net, DeepLabv3+, U-Net2, U-Net++, and SwinUNet. Intersection over Union (IoU) is the most important evaluation parameter and it is an effective measure of a segmentation accuracy. It is important to note that U-Net++ performs better than the rest with a score of 0.8838, and the U-Net2 is close behind with a score of 0.8734. On the contrary, such models as DeepLabv3+ and SwinUNet have relatively poor segmentation results. This research presents a very efficient method whereby the latest and advanced deep learning structures are used to recognize affected areas of the teeth using radiographs. The suggested methodology will provide a tremendous improvement in the precision and effectiveness of the dental practitioners in identifying the teeth affected and hence improve the automation of dental diagnosis. This project makes a significant contribution to the medical image field especially in the design of computerized systems to process panorama radiographies during dental practice.

***Key Words*: Affected teeth, Pixel-based segmentation, panorama radiograph, deep learning, U-Net2, U-Net++, IoU, medical imaging, dental diagnosis.**

1. INTRODUCTION

The correct identification of affected teeth in panoramic radiographs is a critical factor in the diagnosis of the dentist. The appearance of the affected teeth especially in children, may often lead to the possibility of early intervention to avoid the possible complications that may arise like misalignment, infection and retarded eruption of teeth. Such teeth can be manually identified in radiographs, which is subject to error and a lengthy process and that is why, automated means of segmentation are necessary. The field of medical image segmentation has been rapidly developed especially due to the emergence of deep learning methods which have enabled the process of detecting the anatomic structures with greater accuracy. Many architectures have been suggested in recent years when it comes to image segmentation tasks, and such models as U-Net, DeepLabv3+, and SwinUNet have become some of the most useful options.

U-Net architecture that was originally used to segment biomedical images has achieved a lot of popularity because of its ability to produce robust results and also because it is efficient in making pixel-wise predictions. Nevertheless, the later variants like U-Net++, U-Net2, and SwinUNet have made additional improvements by adding new changes to the segmentation accuracy, including deep supervision, skip pathways, transformer-based attention mechanisms, and so on. Such advancements are vital in solving the problems of the multifaceted structure and variability of the panorama radiographs which are usually filled with noise, artifacts and redundant structures.

This paper is a comparative analysis of five segmentation models U-Net, DeepLabv3+, U-Net2, U-Net++, and SwinUNet to understand their efficiency in identifying affected teeth. Evaluating the performance metrics like intersection over union (IoU) we are able to identify which models have the best performance in terms of dental image segmentation tasks to perform. U-Net++ and U-Net2 were the most popular on the tested models, and it is possible to conclude that these models are the most appropriate to use in automated dental radiograph analysis due to their architectural improvements. The results of this analysis are important in terms of the model choice to be used in automated dental diagnosis, which will open the way to more efficient and effective clinical processes.

The study is also a contribution to the increasing literature on medical image segmentation, in which deep learning has demonstrated a significant enhancement of the diagnostic potential in a range of healthcare areas.

This project is novel because it compares advanced segmentation structures with application to the detection of affected teeth in panoramic radiographs. In contrast to the earlier studies, this study comprehensively evaluates the performance of the U-Net++, U-Net2, DeepLabv3+, and SwinUNet on dental images and provides a comprehensive performance benchmark. The originality of the research approach involves taking advantage of the models of deep supervision and transformer-based improvements to the particulars of dental radiograph segmentation including noise, overlapping structures, and variability of tooth anatomy. This study does not only take a step forward in improving automated dental diagnostics, but it also helps in bettering AI-based clinical decision support.

2. Literature Review

The semantic segmentation of the dental panoramic radiograph has become the focal point of automated diagnosis in dentistry since it enables the specification of individual teeth and pathological images on a pixel-by-pixel basis. Initial research has demonstrated that convolutional neural networks (CNNs) are very superior to regular handcrafted filters as they use hierarchical features without necessarily using handcrafted filters [1]. The majority of dental applications make use of the U Net architecture, a type of architecture that was designed to deal with biomedical image segmentation because of its symmetric encoder-decoder design with skip connections that have the ability to retain spatial information during up -sampling and down -sampling [2].

Successful applications have been done in the dental field using U Net variants to automatically extract teeth in multi-faceted dental panoramic radiographs. it has been demonstrated to succeed in binary segmentation tasks with mean Intersection over Union (IoU) and Dice scores high and better than the traditional ones [3], [4]. An example is U Net based model that has good segmentation and parallel detection of dental agenesis in paediatric orthopantomograms with Dice coefficient of 0.8773 and high validation accuracy [3]. Similarly, the variants of the attention gate U Net with a joint loss have also improved the degree of IoU and pixel accuracy on annotated dental X ray datasets [4].

Scholars have generalized U Net to learn contextual semantics and low contrast boundaries. Teeth U Net also combines squeeze and excitation, and multi scale attention to enhance representation of features and subdivision of complex teeth structures more closely, and results in a high precision and overlap with baseline models [5]. Other studies by U Net have been conducted on panoramic X ray images and they make architectural adjustments in dealing with jagged tooth structures and irregular backgrounds [6]. Multiple U Net encoder backbones benchmarks also underscore the fact that marginal gains are obtained in increasing model complexity, and architectural efficiency is required [7].

In addition to U Net, DeepLabv3+ has been tested in the context of lesion detecting and lesion segmentation. Despite the fact that this model demonstrates better contextual capture through atrous spatial pyramid pooling, in comparative studies, DeepLabv3+ and U Net tend to be equally good in terms of IoU measure, and DeepLabv3+ is better concerning recall but worse concerning precision with dental lesion segmentation [8]. DeepLab variants will also be relevant in the future in the task of capturing multi scale features in semantic segmentation inherent to non-medical imaging, as has been demonstrated in the wider biomedical task [9].

Other advancements to the research of segmentation models include transformer-based models like Swin Unet which combines the self-attention mechanism to represent long range dependencies which CNNs cannot capture. Their capability to balance between the global context and local features has made Swin Unet architectures attain state of the art results in multi organ medical segmentation tasks [10]. The classification Dual Swin Transformer U Net (hybrid transformer U Net) and other variations also use long range feature fusion to push segmentation accuracy in a range of medical imaging applications [11].

Surveys of deep learning segmentation algorithms classify U Net, its extension UNet++, and transformer-based models as baseline applications in recent research, as they are applied across a variety of biomedical tasks because they have strong performance measures such as iou and dice coefficient [12], [13]. Further studies specifically investigate Swin Transformer adoption with dental panorama information using both the self-supervised learning and masked image modeling methods to enhance instance segmentation tasks [14].

Taken together, these publications have solidified that deep learning models can greatly improve pixel level segmentation of panoramic radiographs, but still cannot rival the models such as DeepLabv3+ and transformer models to the specific task of dental work, particularly in the fine structure localization and the treatment of occlusions [8], [10], [14].

Despite the U Net and its variations ruling dental segmentation studies and providing high levels of IoU, few studies have read into the transformer-based segmentation (e.g., Swin Unet) specifically applied to dental panoramic imagery and little has been done in comparison between the long-range attention and the pixel level based CNNs. Moreover, existing literature does not contain thorough research on the issue of occlusion management and lightweight optimization of real time dental diagnosis.

3. Dataset

Adult Tooth Segmentation Dataset is a set of dental radiographic images that were created with the purpose of adult tooth segmentation and analysis. This data consists of high-resolution panorama X-ray images, which image the entire set of adult teeth and annotations are provided that permit proper segmentation. The main objective of this dataset is to support the creation of machine learning algorithms to be able to automatically detect and extract single teeth out of dental radiographs, which is essential to numerous dental diagnostic applications. Proper tooth segmentation is necessary in the detection of conditions such as tooth decay, misalignment, and other health conditions of the mouth. The dataset could be an important source of information to researchers and developers of computer vision algorithms in medical imaging, especially in dental health. With such a dataset, scientists are able to train models that can help dental practitioners enhance the accuracy of diagnoses, treatment plans, and care of patients. The quality of images and the presence of annotations on the dataset render it a crucial resource to develop the sphere of dental image analysis and contribute to the development of the automated system, which can facilitate the accurate recognition and assessment of dental diseases in adults.

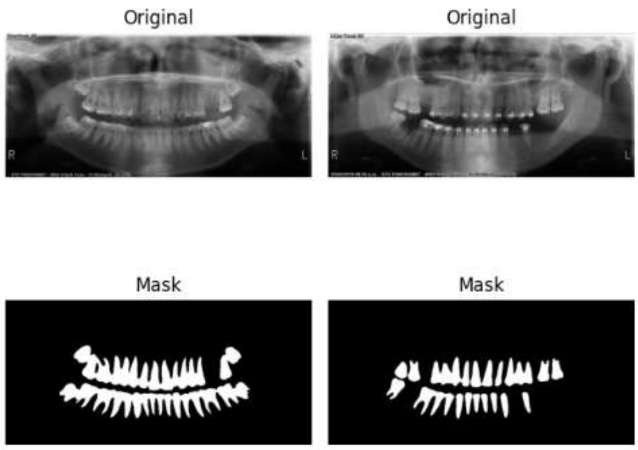

Fig -1 : Sample Dataset

4. Proposed Methodology

4.1 Data Preprocessing

The importance of effective data preprocessing is more valid when using them to solve more complex tasks in medical image processing such as panoramic radiographs. In the research present at hand, preprocessing is a basic process that the algorithms that are to be tested include U-Net, U-Net++, U-Net2, DeepLabV3+ and SwinUNet to allow the input images to be formatted and normalized to be fed into the models.

The initial preprocessing process lies in image resizing that normalizes the input size of all the models. This is done to all images to get them to 256x512 pixels to have uniformity and consistency, particularly in models where the dimensions of the input have to be constant. After the resizing process, the images are transformed into the tensors by way of To Tensor function which allows them to be compatible with the neural networks to process them further.

The DeepLabV3+ model includes a special preprocessing step, i.e., the image normalization. The pixel values are made to have a certain mean of [0.485, 0.456, 0.406] and the standard deviation of [0.229, 0.224, 0.225]. The pretrained version of DeepLabV3+ relies on these values to align itself with the pretraining condition to minimise internal covariate shift, speed up model convergence in the training step, and, in the end, perform better.

Meanwhile, other models do not use normalization. These models aim at normalizing the size of the images and transforming them into tensors, where the value of the pixels is normalized to lie in the range 0 to 1 after the resizing process. This simple preprocessing guarantees the quality of the inputs to all the models regarding training and segmentation activities.

4.2 Working Process

U-Net, U-Net++, U-Net2, DeepLabV3+ and SwinUNet were chosen due to their success in image segmentation tasks particularly in the medical imaging sector in this project. These models have achieved improved outcomes in the segmentation of complicated structures of any medical radiograph. The U-Net and its variants (U-Net+, U-Net2) are also of the encoder-decoder design that is efficient in precise segmentation. DeepLabV3 + has atrous convolution and SwinUNet has transformer-based attention, which improve multi scale context and accuracy of segmentation, respectively. The advantage of each of the models is that they are applicable to address the diverse image size, noise and boundary complexities of panorama radiographs which contributes to their high ability to detect the affected teeth.

U-Net is realized on the principle of encoder-decoder scheme, where the encoder successively sketches the objects of the input image with the assistance of a bank of convolutional layers. The output segmentation map is then up sampled by the decoder. It is a special architecture which is customized to extract detail at the fine-grain in segmentation tasks. The network objective of the training is to minimize the pixel-wise loss by means of a loss measure like the Dice coefficient, or binary cross-entropy. The U-Net usually uses gradient-based optimization algorithms like adam in the instance of the optimization process. The objective is to minimize:

$$L_{\text{U-Net}} = -\sum_i y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)$$

where y_i is the ground truth, and $\hat{y}_i$ is the predicted output.

U-Net++ expands on this by employing thick skip roads between the encoder and decoder to enable the model learn to better capture correlation of features, and to learn better. The skip connections, concatenated, facilitate the network to learn more rich feature representations, which enhances better segmentation. In the process of training, the model adjusts its learning rate to emphasize regions that are hard to segment to optimize with time using the same loss functions as U-Net but with the added advantage of these extra connections.

U-Net2 also builds on U-Net++ with a deeper hierarchical residual learning that also enables the model to learn more detailed patterns. It makes the circulation of gradient easier in training hence, effective learning. Optimization The process resembles U-Net++ except that it has improved gradient flow because of the residual blocks.

DeepLabV3+ makes use of atrous convolution to increase receptive field without loss of spatial resolution. This will enable the model to retain features of a multi-scale at varying resolutions and this is especially critical in the detection of small, complex features such as impacted teeth. The loss function used in the model during training is analogous to that of U-Net, except that, in this case, the rate at which the convolution is dilated is as follows:

$$y = \sum_i w_i \cdot x_{i+d}$$

where d is the dilation factor which enables the model to increase its field of view so that it can view the context better.

SwinUNet uses transformer-based attention mechanism, which is effective to capture the long-range dependencies in the data. Transformers in contrast to traditional convolutional operations have self-attention mechanisms to learn contextual relationships throughout an image. The method will allow SwinUNet to divide complex structures in radiographs in a better way. The model attached during training is the development of the relationship between remote parts of the image which enhances its capacity to cope with varying image resolutions and hazy backgrounds.

$$Q, K, V = \text{Attention}(X) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Q, K, V denotes the query, key and value matrices, respectively, and d k is the size of the key. The attention mechanism is trained that learns the relevant parts of the image to be used in the process of segmentation and weighted these parts more during training. This enhances the quality of the model in highlighting the finer details in difficult images such as panoramic radiographs.

Training iterations are performed on all these models with backpropagation employed in changing the weights of the network and optimizing the parameters of the model to the greatest possible segmentation accuracy. The training process of each model is designed to fit the medical imaging-specific challenges, namely, the complexity of the segmentation that must be performed on the panoramic radiograph images to identify the affected teeth.

4.3 Model Evolution Metrics

1) Intersection over Union (IoU)

The overlap of the predicted segmentation mask and the ground truth mask is measured using a value. It was calculated by determining the areas of intersection (overlap) and areas of union of the two sets. The IoU is in the range of 0 to 1 where a higher value will illustrate an optimal performance. The IoU is the most useful in the situation of evaluating the segmentation model as it takes into account the false positives and false negative.

Formula for IoU:

$$\text{IoU} = \frac{\text{Intersection}}{\text{Union}} = \frac{|\boldsymbol{A} \cap \boldsymbol{B}|}{|\boldsymbol{A} \cup \boldsymbol{B}|}$$

Where:

- $|\boldsymbol{A} \cap \boldsymbol{B}|$ is the area of overlap between the predicted and ground truth masks.
- $|\boldsymbol{A} \cup \boldsymbol{B}|$ is the area covered by either the predicted mask or the ground truth mask.

2) Dice Similarity Coefficient (DICE)

It is a measure of how two sets are similar that is It is an indicator of the similarity between two sets which is normally used in image segmentation tasks. It is calculated as: twice the area under the masks of the predicted and ground truth/ total of the area of the two mask images. DICE ranges between 0 (no overlap) to 1 (perfect overlap), and the larger the value, the higher performance is obtained by the segmentation. It is fully applicable particularly when the models are evaluated on unequal data.

Formula for DICE:

$$\text{DICE} = \frac{\mathbf{2} \times \text{Intersection}}{\text{Size of A} + \text{Size of B}} = \frac{\mathbf{2}|\boldsymbol{A} \cap \boldsymbol{B}|}{|\boldsymbol{A}| + |\boldsymbol{B}|}$$

Where:

- $|\boldsymbol{A}|$ is the number of pixels in the predicted mask.
- $|\boldsymbol{B}|$ is the number of pixels in the ground truth mask.
- $|\boldsymbol{A} \cap \boldsymbol{B}|$ is the number of pixels in the intersection of both masks.

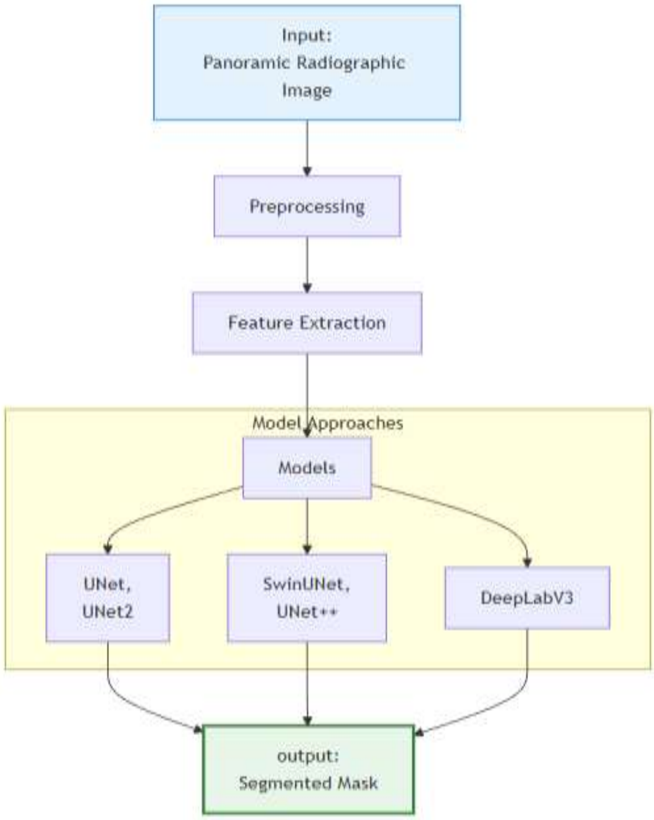

Figure 2 Proposed Methodology Work Flow

5. Results and Discussions

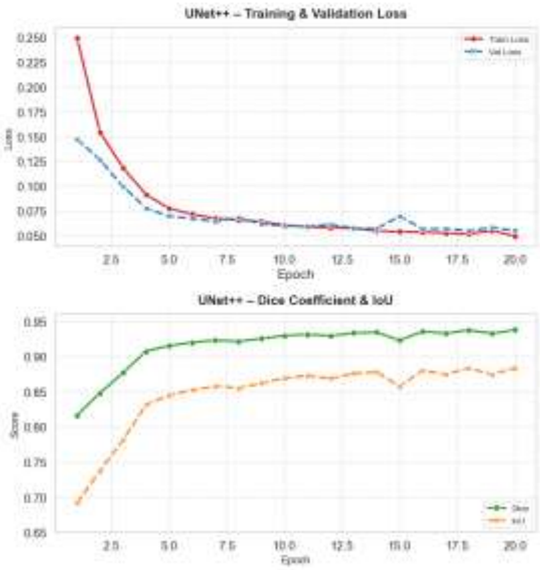

Figure 3 Unet++ Model Performance

Considering Figure 3, we are able to observe the trends of the performance of the U-Net++ model relative to 20 epochs. There is a progressive decrease in the loss of the train and this indicates that the model is learning and error minimizing with time. Similarly, the loss of validation is also on the same trend, which means that the model learns simultaneously well on unknown data. As far as the performance of segmentation is concerned, both Dice coefficient and the IoU are constantly increasing and followed by small fluctuations. It shows that the model is already optimizing its segmentation capabilities and it can reach a high performance with Dice and IoU almost equal to 0.94 at the 20 th epoch. These measures remain stable and it is through this that the strength of the model and its ability to identify the affected teeth in panoramic radiographs as the right segmentation of the teeth comes into focus.

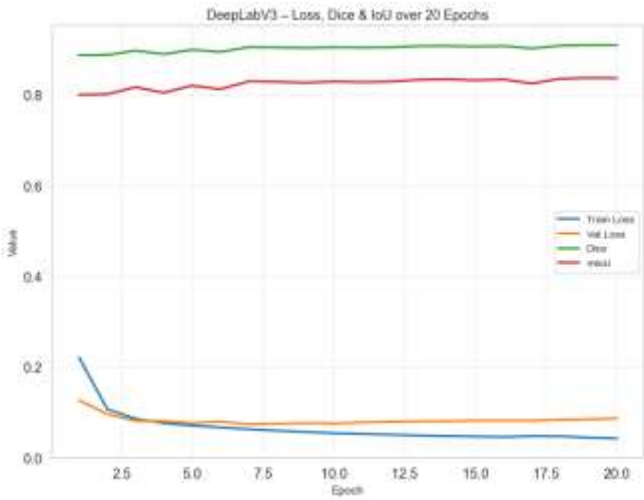

Figure 4 Deeplabv3 Model Performance

By referring to Figure 4, one can observe the performance indicators of DeepLabV3 model in 20 epochs. The decrease in the loss decreases significantly, the decrease in validation also decreases, and one can realize that the model learns and adapts to the data, and it generalizes successfully to unknown data. The Dice coefficient and an IoU slowly increase during training and the ultimate results of training attained values, which are closer to 0.90. These measures show that the model is performing better segmentation that has a higher overlap of the predicted and ground truth masks thus, providing a robust model to segmentation tasks.

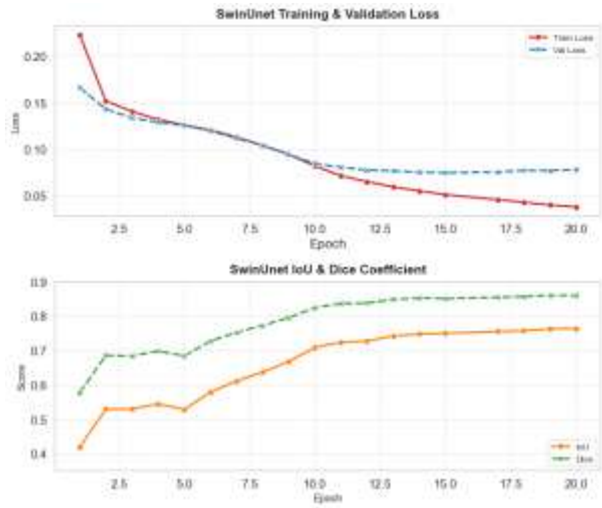

Figure 5 Swin Unet Model Performance

In accordance with Figure 5, we may observe the SwinUNet model performance during the training period, in 20 epochs. The loss of the train involves a slow decrease of the first epoch, which indicates the tendency to the fact that the model is learning more and making fewer mistakes. Likewise, the validation loss also decreases and it signifies that the model is generalizing correctly on unexplored data and that there is no overfitting in the model. The gap between train and validation loss reduces to quite small values towards the end of the epochs, again signifying the possibility of the model to generalize. Both Dice coefficient and IoU demonstrate an enormous enhancement in the results of segmentation across the epochs. In the first epoch, both dice and IoU begin with 0.5773 and 0.4188 respectively but in the final epoch, Dice and IoU have reached 0.8598 and 0.7636 respectively. This consistent upward trend in both these measures indicates that the model is well training its segmentation abilities, and has a high overlap with the ground truth and provides consistent guarded results to the segmentation problem.

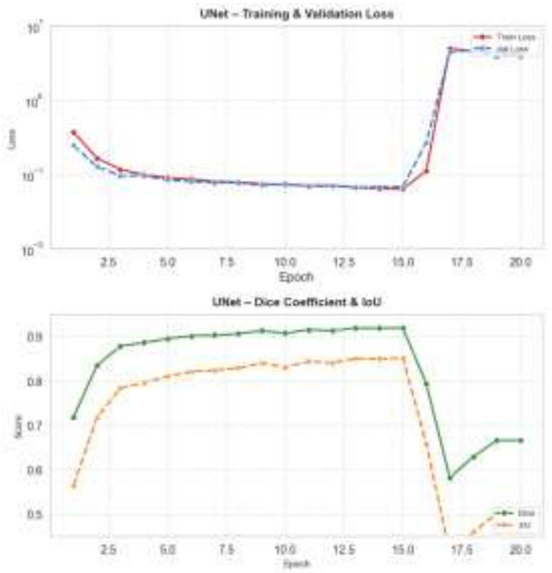

Figure 6 Unet Model Performance

In Figure 6 we can take the performance of U-Net model in 20 epochs into consideration. The tendencies of the train loss are clearly becoming negative, and it means that the model is learning effectively and reduces its mistakes with the growth of the number of epochs. However, the loss of validation increases strongly at epoch 10, potentially pointing to some overfitting or instability of the model at this point. However, the validation loss lacks any additional extreme variation and this shows improved generalization.

Both measurements of Dice coefficient and the IoU have been improved in the trend, which is the indication that the model is effective in the way it has optimized the results of segmentation. The two metrics are progressively increasing at start, Dice is heading towards 0.90, and at epoch 20, iou is heading towards greater than 0.80. The fact that the model can be improved with time demonstrates that it is an effective method of segmentation as it acquires to become more accurate in the realization in the division of areas of interest.

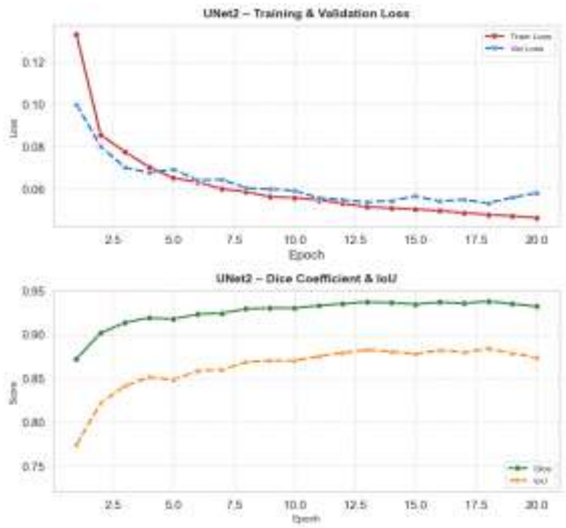

Figure 7 Unet2 Model Performance

Figure 7 indicates that the U-Net2 model is observed to experience a significant reduction in the loss and segmentation measure of the 20 epochs. The loss of the train commences with 0.1329 and decreases gradually implying the effective learning and minimization of error. Similarly, the loss in validation begins at 0.0999 and plateau at the termination of the training suggesting that it can extrapolate effectively with unobserved data and it does not overfit.

The metrics of Dice coefficient and IoU are continuously increasing and are equal to 0.9320 and 0.8734 respectively in the final epoch. These increasing values are a sign of the ability of the model to segment areas more precisely as the two measures are a measure of the mismatch between the predicted and ground truth segmentation. Despite the slight deviation in both measurements in the middle epochs, the overall tendency of the increasing number of the metrics demonstrates that the model is becoming more competent in its segmentation activities, therefore, becoming a helpful tool of determining the affected teeth in the medical images as precisely as possible.

Table 1 Model Comparison Table

Model	Best Epoch	Train Loss	Val Loss	Dice	IoU	Accuracy
UNet	15	0.063	0.067	0.919	0.851	0.919
UNet2	18	0.047	0.053	0.937	0.883	0.937

UNet++	20	0.049	0.055	0.937	0.883	0.937
DeepLabV3	19	0.044	0.084	0.910	0.837	0.910
Swin-UNet	19	0.040	0.077	0.860	0.763	0.860

Table 1 shows the results of different segmentation models measured by several indicators and accuracy. UNet2 and UNet++ have the highest overall performance with high Dice (0.937), IoU (0.883) and accuracy (93.7%). These models have low training and validation errors (0.047 and 0.053 on UNet2, 0.049 and 0.055 on UNet++. The models have low training and validation error (0.047 and 0.053 on UNet2 and 0.049 and 0.055 on UNet++) suggesting that they converged well during the training and generalization on unseen data. DeepLabV3, however, has a lower training loss (0.044) but a higher validation loss (0.084) which could indicate that DeepLabV3 was fitting the training data. Its Dice (0.910), IoU (0.837) and accuracy (91.0) are less than UNet2 and UNet++, which means that there is still an opportunity to improve its performance.

The best train loss (0.040) is obtained with Swin-UNet but with the lowest Dice (0.860) and IoU (0.763) which results in the low overall accuracy (86.0%). Such a discrepancy can indicate that this model is effective with the training data but has difficulties with extrapolating the data to previously unknown data, a fact that raises the question of a possible problem with the complexity of the model or the architecture used to handle segmentation tasks. These results highlight the usefulness of UNet2 and UNet++ as effective high-quality segmentation models.

6. CONCLUSIONS

Lastly, the results of the analysis of many deep learning frameworks, including UNet, UNet2, UNet++, DeepLabV3, and Swin-UNet, confirm the discrepancies in the training and validation performance of individual models, and a clear tendency towards the superior segmentation metrics. UNet2 has been best than the others as it has the low validation loss, highest Dice coefficient, and greatest IoU that implies that it can readily generalize to unknown data and yield very great accuracy in segmentation. The same was also done by Unet++ with regular results and good segmentation abilities but with slight increases in the validation loss. The DeepLabV3 that had the lowest training loss experienced greater challenges in the increased validation loss and this negatively affected its segmentation accuracy. Though the Swin-UNet demonstrated a low validation loss, it had poorer performance in terms of Dice and IoU scores which depicts a poor performance of the model in segmentation compared to others. Overall, the findings indicate the benefits of UNet2 and UNet++ in being used as medical image segmentation, but UNet2 was more effective in terms of accuracy and generalization.

The improvement of Swin-UNet model could be made through better performance on segmentation in the future, which may be through either changing the architecture or fine-tuning of hyperparameters. Moreover, further enhancement of DeepLabV3 and UNet++ generalization and segmentation might be achieved by using more sophisticated methods like attention mechanisms or hybrid models. Moreover, training on larger and more varied datasets may also be useful to enhance model resilience and segmentation accuracy between medical imaging of various type, and increase their clinical utility.

7. REFERENCES

1. D. Luo, W. Zeng, J. Chen, and W. Tang, "Deep learning for automatic image segmentation in stomatology and its clinical application," *Front. Med. Technol.*, vol. 3, p. 767836, 2021, doi: 10.3389/fmedt.2021.767836.
2. D. Koishiyeva, D. Mukhammejanova, J. W. Kang, and A. Mukasheva, "A review of deep learning approaches based on Segment Anything Model for medical image segmentation," *Bioengineering*, vol. 12, no. 12, p. 1312, Dec. 2025, doi: 10.3390/bioengineering12121312.
3. H. Tunç, N. Akkaya, B. Aykanat, and G. Ünsal, "U-Net-Based deep learning for simultaneous segmentation and agenesis detection of primary and permanent teeth in panoramic radiographs," *Diagnostics*, vol. 15, no. 20, Oct. 2025, doi: 10.3390/diagnostics15202577.
4. M. A. Makrygiannakis, A. Ntokos, and E. G. Kaklamanos, "Development and evaluation of an attention-gated U-net model for binary segmentation of teeth versus background in panoramic radiographs for orthodontic applications," *Eur. J.*

Orthod., vol. 48, no. 1, Dec. 2025, doi: 10.1093/ejo/cjaf114.

5. S. Hou, T. Zhou, Y. Liu, P. Dang, H. Lu, and H. Shi, "Teeth U-Net: A segmentation model of dental panoramic X-ray images for context semantics and contrast enhancement," *Comput. Biol. Med.*, vol. 152, p. 106296, Jan. 2023, doi: 10.1016/j.compbiomed.2022.106296.
6. J. H. Lee, S. S. Han, Y. H. Kim, C. Lee, and I. Kim, "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.*, vol. 129, no. 6, pp. 635–642, Jun. 2020, doi: 10.1016/j.oooo.2019.11.007.
7. A. Yalım, E. Aytugar, F. Kalabalık, and İ. Akdağ, "Comparative analysis of deep learning architectures for automatic tooth segmentation in panoramic dental radiographs: Balancing accuracy and computational efficiency," *Diagnostics*, vol. 16, no. 2, p. 336, Jan. 2026, doi: 10.3390/diagnostics16020336.
8. F. K. Çay et al., "DeepLabv3+ method for detecting and segmenting apical lesions on panoramic radiography," *Clin. Oral Investig.*, vol. 29, p. 101, 2025, doi: 10.1007/s00784-025-06156-0.
9. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with atrous separable convolution for semantic image segmentation," Accessed: Feb. 28, 2026.
10. H. Cao et al., "Swin-Unet: Unet-like pure transformer for medical image segmentation," Accessed: Feb. 28, 2026.
11. A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "DS-TransUNet: Dual Swin transformer U-Net for medical image segmentation," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, doi: 10.1109/TIM.2022.3178991.
12. W. Yao, J. Bai, W. Liao, Y. Chen, M. Liu, and Y. Xie, "From CNN to transformer: A review of medical image segmentation models," *J. Imaging Informatics Med.*, vol. 37, pp. 1529–1547, 123AD, doi: 10.1007/s10278-024-00981-7.
13. Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical image segmentation: A comprehensive review of deep learning-based methods," *Tomography*, vol. 11, no. 5, Apr. 2025, doi: 10.3390/tomography11050052.
14. A. A. Longin and J. Latecki, "Self-supervised learning with masked image modeling for teeth numbering, detection of dental restorations, and instance segmentation in dental panoramic radiographs," Accessed: Feb. 28, 2026.