IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 February 2024, accepted 15 March 2024, date of publication 25 March 2024, date of current version 29 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3380027

RESEARCH ARTICLE

Semantic Segmentation on Panoramic Dental X-Ray Images Using U-Net Architectures

RAFIATUL ZANNAH[1], MUBTASIM BASHAR[1], RAHIL BIN MUSHFIQ[1], AMITABHA CHAKRABARTY[1], SHAHRIAR HOSSAIN[2], AND YONG JU JUNG[3], (Member, IEEE)

[1]Department of Computer Science and Engineering, Brac University, Dhaka 1212, Bangladesh
[2]Department of Computer Science and Engineering, George Mason University, Fairfax, VA 22030, USA
[3]School of Computing, Gachon University, Seongnam 13120, South Korea

Corresponding author: Yong Ju Jung (yjung@gachon.ac.kr)

ABSTRACT The field of medical image analysis is in a constant state of evolution, particularly in the challenging tasks of segmenting organs, diseases, and abnormalities. Therefore, in the realm of dental disease diagnosis, image segmentation plays a crucial role in addressing the difficulties faced by dentists worldwide when diagnosing dental diseases with the naked eye. One prominent deep neural network architecture, known as U-Net, originally designed for biomedical image segmentation, has seen multiple variations and advancements aimed at improving its performance. However, the lack of comparative studies has made it challenging to assess the effectiveness of these U-Net variants in segmenting dental X-ray images. The primary objective of this research is to conduct a comprehensive performance comparison among various U-Net architectures for dental image segmentation. Specifically, we examine six U-Net architecture variants: Vanilla U-Net, Dense U-Net, Attention U-Net, SE U-Net, Residual U-Net, and R2 U-Net. These variants employ configurations with two and three convolutional layers in both the encoder and decoder blocks. Our evaluation metrics include Accuracy, Dice coefficient, F1 score, and IoU (Intersection over Union). Among these U-Net variants, Vanilla U-Net, with two convolutional layers, demonstrated the highest level of performance, achieving an Accuracy of 95.56% and an IoU score of 88% on the validation set. Notably, this model also exhibited a shorter processing time compared to the other architectures. Conversely, when employing three convolutional layers, the Dense U-Net variant emerged as the top performer, achieving an Accuracy of 95.94% and an IoU score of 89.07% on the validation set. Furthermore, the segmentation process successfully isolates the teeth from the surrounding structures, which holds promise for improving disease detection through the development of automated disease diagnosis models.

INDEX TERMS Dental, semantic segmentation, data annotation, OPG image, U-Net, U-Net variants, dice coefficient, IoU, architecture comparison.

I. INTRODUCTION

Dental diseases (i.e., bone loss, decay, gum disease, etc.) have become the most common phenomenon in the field of medical science and therefore are increasing highly in recent times. Dental health is just one aspect of oral health, including our body's overall health and well-being [1]. However, when we only talk about dental health, teeth are one of the body parts that are overused and have a high resistance to damage and long-lasting architectural durability. Still, it is vulnerable to various illnesses [2]. Dental diseases affect nearly 90% of individuals in the United States, as mentioned in a survey of the National Health and Nutrition Examination [3]. Furthermore, according to specific epidemiological statistics, dental disease is more prevalent in communities with poor socioeconomic positions [4]. Even then, Health and Retirement Studies show that even if an individual's wealth drops by 50% in the US, they pursue dental care for their well-being.

According to the World Health Organization, [5] Oral diseases, although largely preventable, impose significant health challenges globally, affecting individuals across their

The associate editor coordinating the review of this manuscript and approving it for publication was Tony Thomas.

lifespan and leading to pain, discomfort, disfigurement, and even mortality. The Global Burden of Disease 2019 estimates that there are 3.5 billion individuals affected by oral illnesses, with untreated dental caries in permanent teeth being the most common health issue. The cost of treating oral health conditions is high, often falling outside the scope of universal health coverage (UHC). In many low- and middle-income countries, inadequate services are available for the prevention and treatment of oral health issues. These diseases arise from modifiable risk factors common to various non-communicable diseases, such as sugar consumption, tobacco and alcohol use, poor hygiene, and the underlying social and economic determinants. Resolving these issues is essential to reducing the prevalence of oral illnesses worldwide.

A wide range of dental care is given through hands-on treatment or radiographic images. Since the invention of X-ray imaging, oral X-ray images having three types (Bitewing, Panoramic, and Periodical) have been widely employed [6]. These dental X-rays reveal the hidden inner part of teeth that a dentist's eye can hardly detect. Among the three types of X-ray radiography in dental anatomies, panoramic X-ray, known as the Orthopantomogram, is less time-consuming, has the lowest radiation value, and causes less patient discomfort. Therefore, using OPG images or panoramic X-rays can significantly detect dental diseases. Medical image segmentation is one of the essential steps in medical image analysis. The segmentation results' shape, size, and total area can provide crucial information for understanding early signs of potentially deadly diseases, as described in [7].

In the last few years, different medical fields such as dentistry [8], brain [9], liver [10], breast [11], lung [12] are explored through artificial intelligence. The Convolutional Neural Network architectures have shown excellent performance in detection and segmentation. Where [13] U-Net architecture is a popular model for medical image segmentation effective for both 2D and 3D images for its pixel-wise classification behavior. Furthermore, the variants under consideration include the normal U-Net model called the Vanilla U-Net [14], Dense U-Net [15] with dense layers interrogation, Attention U-Net [16] with attention mechanism, SE U-Net [17] with the squeeze and excitation (SE) block, Residual U-Net [18] with residual connections and R2 U-Net [19] with residual and recurrent connections. By investigating these variants, we aim to comprehensively analyze their respective strengths and weaknesses in the context of teeth segmentation. Analyzing this will contribute to our understanding of the suitability of different U-Net variants for this specific medical imaging task and guide the selection of the optimum model for accurate and efficient teeth segmentation in OPG X-ray images.

According to [20], In terms of the number of training samples needed, memory requirements, and computing time, U-Net has excelled in earlier research. For this reason, we selected U-Net, and its variants to perform semantic segmentation on our collected OPG images of teeth. This study determines how these models perform on our dataset. Contributions of this paper are as follows:

- A dataset of 389 dental panoramic x-ray images was made from scratch with a proper annotation that was used with proper pre-processing methods in this study.
- Six advanced U-Net models with layer variations were implemented with their respective parameters, and outputs were deeply analyzed.
- All six models are U-Net-based models where specific layer or block-based differences are made to have good accuracy and high performance.
- The effectiveness and performance of each of the six models were evaluated using four different metrics Accuracy metrics, F1-Score, Dice Coef, and IoU, with the plots against the mean epoch time(MET), including the confusion metrics table, training, and validation curves.
- We found minimal differences Based on the profoundly constructive and comprehensive comparison of the dataset's six U-Net-based models. The Dense U-Net (layer 3) model and the R2 U-Net(layer 3) model have the most optimum model in terms of accuracy and predictions which can be concluded for further studies.

After presenting the context, motivation, and objectives of our research, we will now delve into the 'Related Works' section where we have conducted an extensive review of the existing literature that informed our research. Following that, we will provide a detailed description of the dataset we utilized in the 'Data Set' section. Subsequently, we will elaborate on the intricacies of our methodology, including the various U-Net variants, in the 'Research Methodology' section. We will then discuss the evaluation metrics employed to assess our models' performance. After that, we will showcase the results of our experiments, followed by a comprehensive discussion and analysis of our findings. In the concluding sections of this paper, we will summarize our key conclusions and pinpoint potential directions for future research endeavors. To commence this journey, we will first elucidate the insights garnered from our review of related works in the following section.

II. RELATED WORKS

In this era of AI, Deep learning has become a promising aspect in the field of medical science due to its efficiency in results and ability to work with complex and large data in various imaging domains. One particular area of that is the deep learning techniques for semantic segmentation in the field of Dentistry. On the other hand, among the different medical imaging modalities, panoramic X-ray imaging has gained significant attention in terms of working with various deep-learning models. By analyzing existing studies deeply, this review provides valuable insights and identifies gaps for further research, contributing to the enhancement of dental imaging and analysis of various deep learning techniques.

As mentioned in [21], it resolves the challenge of initializing the tooth model by itself, and the findings

demonstrate that the tooth morphologies may be extremely closely matched. The teeth segmentation problem is solved in two phases using RFRV-CLMs. The first stage is estimating a few teeth and related mandibular areas that are being used to begin the search for individual teeth, and phase two involves searching each tooth individually. Again, the paper [22] focuses on CNN, which is based on the U-Net model and is used to create a model for teeth segmentation from panoramic images over a dataset of 1500 images. They made the following alterations to the U-Net architecture: they applied batch normalization before every max pooling, up-sampling, and concatenation layer instead of dropout during training which achieved a dice score of 0.936%.

Furthermore, in the paper [23], the authors utilized a dataset comprising 1500 panoramic x-ray images obtained from ivisionlab for their detection experiments. The primary aim of the study was to augment the DNS Panoramic Images dataset by identifying cavities in panoramic images and creating binary ground truth representations for evaluation purposes. The authors extended the DNS dataset by detecting cavities in the panoramic images and generating binary ground truth images. They employed three variations of the U-Net architecture, U-Net, U-Net++, and U-Net3+ – to improve the delineation of cavity boundaries. U-Net3+ exhibited exceptional performance, achieving a testing accuracy of 95%.

By considering the benefits of both residual networks (ResNet) and DenseNet, they suggest an effective network architecture in this study [20]. While using much fewer model parameters than DenseNet, this approach adds more skip connections than ResNet.For the ISIC 2018 dataset and the brain MRI dataset, they gained a mean dice coefficient of 0.861 and 0.8643, respectively. Again, in [24], they provide a model for segmentation called the MFFRU-Net. They create an easy-to-use multi-scale feedback mechanism. They used a public image dataset to assess their proposed MFFRU-Net which got a 96.78% accuracy rate and a 98.56% AUC, respectively.

In the paper [25], the author used various U-Net architectures such as Dense U-Net, Attention U-Net, 3D U-Net, etc on different types of medical images like MRI, Microscopy, Dermoscopy, etc, and specifies U-Net as a context-based learning model and emphasizes its suitability for medical images.

Research done in [26] is centered around the segmentation of the 3D image. For end-to-end learning of tooth instance segmentation in 3D point of iOS cloud data, a model named Mask-MCNet is introduced in this research. The suggested model separates the points that are relevant to each distinct tooth instance while also predicting each tooth's 3D bounding box to localize each tooth instance. This property results in highly precise segmentation that is necessary for clinical practice by preserving the intricate context of data. Among the two datasets, the first dataset includes 120 optical images of odontiasis from 60 adult patients, including lower and upper jaw images. The second dataset consists of 48 optical images of 24 adult people and is exclusively used to assess the robustness of MCNet's to various scanner types. The outcomes demonstrate that the Mask-MCNet beats modern models by reaching a tooth instance segmentation score of 98% IoU. Similarly, the paper [27] proposes a hierarchical multi-step model that automatically identifies and segments 3D individual teeth from dental CBCT images. To get over the computational difficulty posed by high dimensional data, it generates panoramic photos of the upper and lower jaw images. Following that, 3D individual teeth lose- and tight-ROIs are captured from the acquired 2D images. They got a 93.35% F1 score and a 94.79% Dice coefficient percentage for the study.

The study in [28] looks towards lightweight deep learning techniques for segmenting dental X-ray images. This research proposes a novel lightweight knowledge distillation neural network technique. They propose an attempt to retrieve reliable data from a teacher network using a knowledge network. They referred to it as a knowledge consistency neural network for simplicity (KCNet). In total, 1321 dental panoramic images were employed in this research. As their student and teacher networks, respectively, they selected U-Net and ESPNet-v2. They got the IoU score of 80.4% and 7 the Dice coefficient of 89%. Similarly, the study done in [29] assesses the precision and effectiveness of deep learning-based automatic teeth segmentation using a DGCNN-based algorithm. Three different methods were used to compare electronic dental models. Five hundred sixteen dental models were used to train a deep learning system to segment teeth, and 30 dental domains were used to evaluate the precision and efficiency of the segmentation. The accuracy of tooth segmentation was 97.26%, 97.14%, and 87.86% for the AS, LS and DS, respectively

Furthermore, the study [30] shows the viability of the SWin-U-Net CNN model for segmenting teeth on panoramic X-rays. SWin-U-Net is an encoder-decoder system that uses transformers and is shaped like a U with skip connections. In SWin-U-Net, a symmetric encoder-decoder structure is built using jump connections. It uses a local to a global strategy for self-attention. Moreover, it builds a patch-expanding layer to increase sampling and feature dimension without using convolution or interpolation techniques. For research purposes, 100 panoramic radiographs of adult patients were randomly chosen. They achieved an accuracy of 88.52% using SWin-U-Net.

Moreover, according to [31], the CNN-Transformer Architecture UNet network is proposed as a proficient and successful approach for segmenting dental cone-beam computed tomography (CBCT) images. The study showcases the model's robust performance and adaptability to external datasets, achieved through innovative architectural design and strategic fine-tuning. The researchers gathered a dataset comprising 200 CBCT scans, with annotations provided for 45 of them to facilitate network training. Following the training phase, the model demonstrated outstanding performance, as evidenced by notable metrics, including a Dice

Similarity Coefficient (DSC) of 87.12%, Intersection over Union (IoU) of 78.90%, Hausdorff Distance 95 (HD95) of 0.525 mm, and Average Symmetric Surface Distance (ASSD) of 0.199 mm. With a similar dataset, in paper [32], their investigation provides a thorough evaluation of the combined segmentation outcomes generated by three convolutional neural network (CNN) models in the construction of a maxillary virtual patient (MVP) from cone-beam computed tomography (CBCT) images. The dataset consisted of 40 CBCT scans with varied scanning parameters. By integrating three independently validated CNN models, the study successfully achieved comprehensive segmentation encompassing the maxillary complex, maxillary sinuses, and upper dentition. Expert qualitative assessments yielded high scores, with 85% of segmentations rated within the range of 7 to 10, while the remaining 15% fell between 3 and 6. The automated segmentation process demonstrated efficiency, with an average processing time of 1.7 minutes. Key quantitative metrics, such as a Dice Similarity Coefficient (DSC) of 99.3%, indicated outstanding alignment between automated and refined segmentation. Furthermore, the consistency in refinements among observers showcased a 95% Hausdorff distance of 0.045 mm. This research highlights the promising potential of integrated CNN models in the precise and efficient creation of maxillary virtual patients from CBCT scans, substantiated by robust qualitative and quantitative assessments.

In the paper [33], the Teeth U-Net model is introduced to tackle challenges associated with dental panoramic X-ray image segmentation. The key contributions involve integrating a Squeeze-Excitation Module within both the encoder and decoder, coupled with incorporating a dense skip connection to narrow the semantic gap between them. To address issues related to irregular tooth shapes and low image contrast, a multi-scale aggregation attention block (MAB) is applied in the bottleneck layer, effectively extracting teeth shape features and integrating multi-scale features adaptively. A Dilated Hybrid self-Attentive Block (DHAB) is also formulated to capture dental feature information across a broader perceptual field. The proposed model exhibits noteworthy outcomes in three comparative experiments, showcasing elevated Accuracy, Precision, Recall, Dice, Volumetric Overlap Error, and Relative Volume Difference metrics for dental panoramic X-ray teeth segmentation 98.53%, 95.62%, 94.51%, 94.28%, 88.92%, and 95.97%, respectively. The study confirms the algorithm's effectiveness through its application to clinical dental panoramic X-ray image datasets, highlighting its potential for precise and robust teeth segmentation.

According to [7], a faster version of R-CNN conducts instance segmentation of teeth. First, features from ResNet101 are extracted, and these features are combined to form an FPN that defines anchors and extracts ROIs. The segmentation requires quick weight adjustment with the values of $103, 1$ as $0.9, 2$ as 0.999, and 108, which the Adam optimizer can provide. The SGD is used to fine-tune the weights without any momentum, with 106 as the learning rate. The MSCOCO dataset is utilized, which contains 193 buccal panoramic x-ray images divided into 10 categories. After training with these images, the Mask RCNN achieved 98% accuracy and a 0.88 f1-score. Similar utilization of this algorithm has been mentioned in the image segmentation phase in [34], where they used it from the sample library. The Al model successfully reached 90% of diagnosis accuracy. The paper demonstrates making up an intelligent dental Health-IoT system that is organized and has 3 layers of services. After making the training data set using the semi-automatic labeling method, the clinical images were labeled by the detector, classifying them into 7 types of dental diseases. The detector's function includes visual enhancement, coarse localization, and classification.

According to [35], the author presents the Adaptive Feature Fusion UNet (AFF-UNet), designed to enhance semantic segmentation in remote sensing imagery (RSI). AFF-UNet integrates dense skip connections, an adaptive feature fusion module, a channel attention convolution block, and a spatial attention module. Assessment of public RSI datasets, especially the Potsdam dataset, revealed AFF-UNet's outperformance compared to DeepLabv3+, achieving a 1.09% increase in average F1 score and a 0.99% improvement in overall accuracy. Visual outcomes demonstrated reduced confusion among classes, improved segmentation of diverse object sizes, and enhanced object integrity. AFF-UNet effectively tackles challenges in RSI, providing optimized accuracy for semantic segmentation.

As per [36], CariesNet is a U-shape network with an extra full-scale axial attention mechanism. It is used for segmenting caries types from dental Radiographics. From 1159 x-rays, three types of labeling are applied to 3217 caries locations. The feature extraction process from multi-level CNN is combined with the U-shaped framework. Experiments reveal that their technique can segment three degrees of caries with a mean Dice coefficient of 93.64% and a 93.61% accuracy.

In the paper [37], inspired by U-Net architectures, the author introduces the Neural Architecture Search (NAS) system, aligning it with U-Net models. The application involves Down Sampling and Up-Sampling on medical images using a U-like backbone, referred to as NAS-U-Net.

According to the paper [38], the author introduces the MH U-Net, a U-Net architecture with extraction and aggregation of multi-scale features that is very efficient in medical image segmentation. To have efficient gradient flow and fewer training parameters the Densely connected blocks are used like Dense U-Net architecture.

As mentioned in paper [39], the study presents a cloud-based Convolutional Neural Network (CNN) model for automated segmentation of dental implants and prosthetic crowns in Cone-Beam Computed Tomography (CBCT) images. The dataset consists of 280 maxillomandibular jawbone CBCT scans. The CNN model, trained on expert-based semi-automated segmentation, demonstrated

high efficiency, requiring 60 times less time than the semi-automated approach. Results indicated strong segmentation performance with high dice coefficient similarity scores (0.92 for implants, 0.91 for implants with restoration) and low root mean square deviation values (0.080.09 mm for implants). Finally, the study shows the model's accuracy and clinical significance in dental imaging.

In the paper [40], the author introduced ELU-Net, a lightweight U-Net model with deep skip connections at a large scale, extending from the encoder to the fully extracted features of the decoder, similar to U-Net++. Furthermore, the paper emphasizes the use of a distinct loss function to enhance the efficiency of brain tumor detection.

Although some of them focus on other types of medical images [20], [24] and some provide both teeth segmentation as well as disease classification [7], [34]. The studies done in [27] and [29] are based on 3D image segmentation, which is not our area of concern at this moment as we are focusing on 2D panoramic x-rays. Some of the research introduced new novel models for segmentation [22], [27], [29] by combining features from U-Net variants or by introducing new features.

This study will compare six U-Net variations on dental panoramic X-ray images to evaluate their segmentation performance based on the architecture complexity and total model training time. Though some studies above used modified U-Net structure and a combination of U-Net variants for segmentation, our study aims to compare some U-Net architecture variants to determine which network model is best for our domain for segmentation operation. The effect of the number of convolutional layers will also be examined by comparing two and three layers per block as a simple complementary experiment alongside the more complex architectural changes. While a two-layer architecture may provide a balance between simplicity and effectiveness, a three-layer architecture has a higher level of model complexity, which may have the potential to capture more intricate patterns and details of panoramic X-ray images. While deeper architecture may benefit our research, it often comes with increased computational demands. With this in mind, we want to see if deeper networks with three convolutional layers make a bigger impact in capturing intricate patterns and give better accuracy in segmentation. Thus, We want to evaluate the impact of the number of convolutional layers by analyzing two and three levels per block. This study aims to develop observations about semantic dental panoramic image segmentation via U-Net architectures to be used on any dental dataset and save much time for later research.

III. METHODS

Here we discuss the U-Net architecture variants we utilized, and the testing, validation, and evaluation techniques employed in this section's study on panoramic X-ray images in detail.

A. PROCEDURE

As shown in figure 1, our research starts with collecting data from a local dental clinic. To remove unwanted backgrounds, we cleaned 389 images and resized them to 1024 × 1024 pixels. We later patched each picture to 256 × 256 pixels with a non-overlapping approach to avoid losing any pixels throughout the model training. Furthermore, we used this patched data and split them into train test and validation sets with an 8 : 1 : 1 ratio.

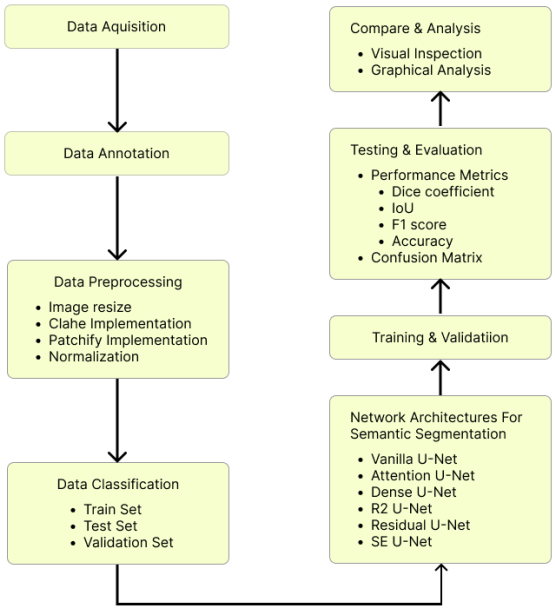

FIGURE 1. Work plan.

We used six different U-Net variants for training. We used four accuracy matrices (dice coefficient, IoU, f1 score, and accuracy) on the validation set. Then we generated confusion matrix tables on a test set to evaluate the model's performance on unseen data. Moreover, we compared each variant's accuracy over time to find an optimal solution for our domain. Furthermore, we compared the results of all models by visual inspection and graphical analysis.

B. DATASET

1) DATA COLLECTION

Initially, we wanted to use raw resources to proceed with our idea, and we started fieldwork. After being consulted in many dental clinics, we convinced one of the renowned dentists of Bogura, Bangladesh, Dr. Ashique Mahmud Igbal (BDS, Dhaka Dental College), of IQBAL'S Dental Clinic. He allowed us to gather the OPG images of the patients, with the condition of not sharing the personal information of his patients, as shown in Figure 2. Finally, we gathered 389 OPG images of X-rays with the best quality. In the first stage, we captured 200 images of different patients. In the last 2nd stage, we captured 189 images. The device used is a Xiaomi Redmi Note 9 Pro with a 64 MP camera for capturing the raw images of the patients. In the image collection, we found the versatility of patients in terms of age

and gender. Nevertheless, we mostly got images of middle-aged patients.

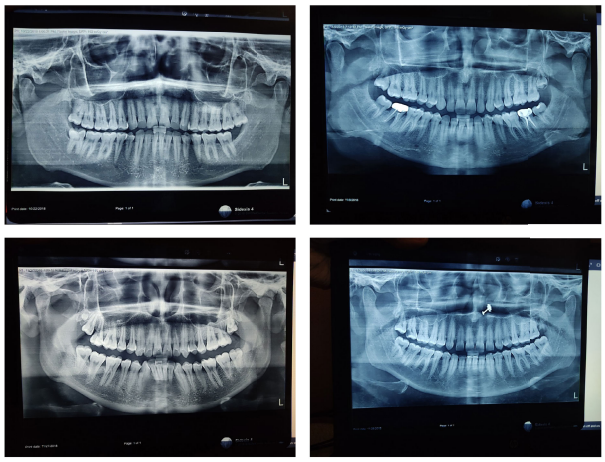

FIGURE 2. Data sample.

2) DATA DESCRIPTION

In the two stages of data collection, we get a total of 389 images in our dataset from different individuals. From visual inspection, we can see that some images have "blue" tints, and some have "grey" tints. These images are of various aged people, including children, men, and women. Our dataset shows that some X-rays contain 32 teeth; some have less than that, which differs from age to age, man to woman, or even child. Children aged 6 to 10 tend to have different patterned and missing teeth. Overall, our dataset contains a good variety of data.

3) DATA ANNOTATION

Label Studio is used to label the 389 panoramic x-ray images using the "Semantic Segmentation Using Mask" module for the data labeling of the OPG images. At the very beginning, after selecting our color format (R= 255 G= 76 B= 66) for our labeling, we set up some minor changes in the Labeling Interface. After the data was imported into the label studio, we set up the region named 'Tooth'. Then, for the labeling, we eventually masked up the regions or areas that are precisely the teeth areas of a particular X-ray image. We masked the images one by one and generated each image's 'Ground Truth' or mask. Finally, we got precisely 389 masks for each X-ray image after annotation.

4) DATA PRE-PROCESSING

a: IMAGE RESIZE

After creating a dataset from scratch, some steps are implemented before extracting teeth as the primary region within images as shown in figure 3. The collected raw X-ray images have dimensions of 4000 × 3000 pixels. As we are using different variants of the U-Net network model for the Semantic segmentation process, the raw images of 4000 × 3000 pixels dimension are unsuitable to take as input for the architectures. We used manual cropping to make squared-sized shapes, eliminating irrelevant backgrounds and resizing the images into 1024 × 1024 pixel dimensions for the convenience of U-Net models.

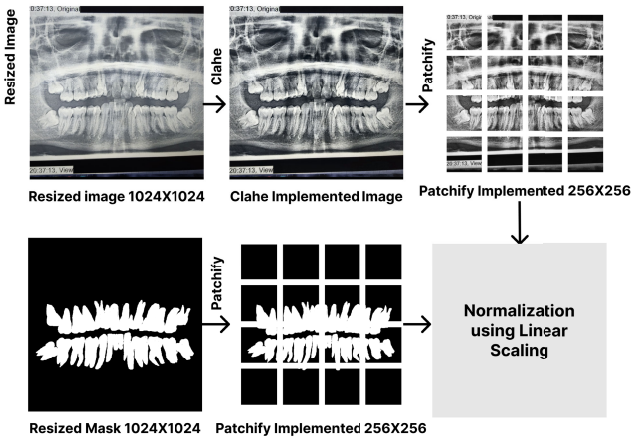

FIGURE 3. Pre-process steps.

b: CLAHE IMPLEMENTATION

In the following process, after getting the 1024 × 1024 sized images, we used the CLAHE [41] (Contrast Limited Adaptive Histogram Equalization) a modern technique that improves the visibility of any sort of digital image with lower contrast and low lighting to ensure the clarity and quality of the X-ray images. Therefore, we can avoid the over-amplification of noise in each image beyond a certain threshold enhancing the local contrast of each digital image. We used the 'clip-limit' of 2.0, which refers to the threshold and Grid size of (8,8) for our implementation purpose. We randomly took a picture to visualize the luminosity as shown in figure 4. After applying CLAHE we can see that the histogram is equalized evenly in figure 5.

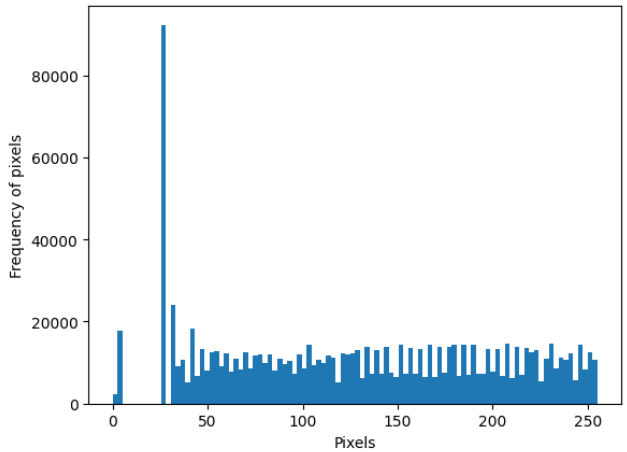

FIGURE 4. Before implementing CLAHE.

c: PATCHED IMAGES

U-Net architecture with patch-based data is more convenient in the case of images with large dimensions, giving better accuracy. We need to convert the 1024 × 1024 sized images

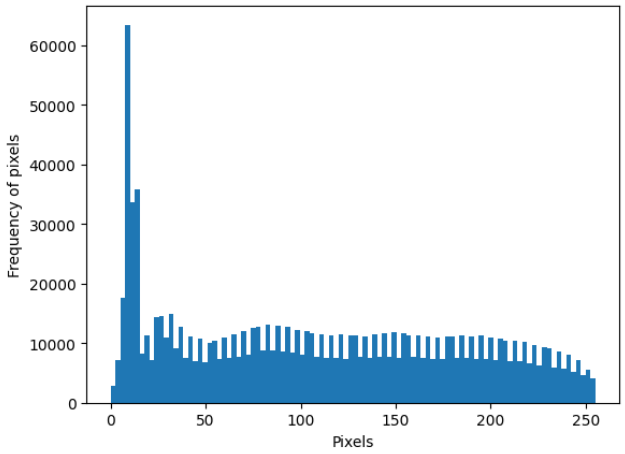

FIGURE 5. After Implementing CLAHE.

into patched images with 256 × 256 pixels for the proposed models to work using patchify, dividing each image into 4 × 4 sub-parts or patches. Finally, after the patchify, we get 389 × 16 = 6224 patched images and masks with 256 × 256 pixels.

d: NORMALIZATION

Our proposed method of normalization of the data is Linear Scaling. As the images are converted into an array, their values are 0 − 255 for each pixel. We divide the data by 255 to normalize our data, considering 0 as the minimum value, and 255 as the maximum value.

$$\text{Scaled Pixel Value} = \frac{\text{Pixel Value} - \text{Min Value}}{\text{Max Value} - \text{Min Value}} \quad (1)$$

5) DATA CLASSIFICATION

With a random state of 42, we divided the dataset into an 8 : 1 : 1 ratio for the train, test, and validation set. We get 6224 patches on both image and mask data from the pre-processing step. After splitting the data, we get 4978 patches on the train set and 623 patches on both the validation and test set.

C. NETWORK ARCHITECTURES FOR SEMANTIC SEGMENTATION

U-Net was introduced in 2015 by Ronneberger et al. [14] to perform segmentation on biomedical images. According to [25], U-Net is primarily used in different domains of biomedical images, such as CT scans, MRIs, microscopy, and X-rays. This architecture is mainly used for segmentation purposes. It computes the separation of borders in morphological order as mentioned in [14]. The weight map is then computed with the following equation.

$$w(x) = w_c(x) + w_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right) \quad (2)$$

The weight map, $w_c : \Omega \to \mathbb{R}$, is employed to balance the frequencies of distinct classes in our problem. The

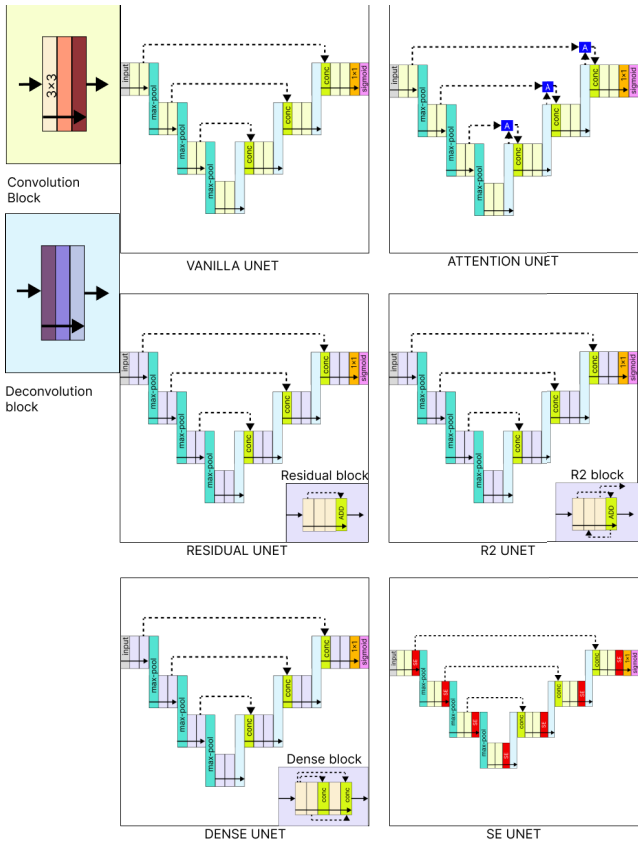

FIGURE 6. U-Net architecture variations.

functions $d_1 : \Omega \to \mathbb{R}$ and $d_2 : \Omega \to \mathbb{R}$ measure the distances from a particular point x to the closest and second closest cell boundaries, respectively. In our experimental implementation, we set w_0 to 10 and σ to approximately 5 pixels.

However, many more applications have been seen. So, the potential of this architecture is increasing. Different variants of U-Net have been introduced to the world since the first introduction of U-Net. We have used six different architecture variants (figure 6) of this model to compare model performance for our domain. Furthermore, the key features of the mentioned architectures are given in table 1.

D. TRAINING AND ARCHITECTURE PARAMETERS

To train our architectures and evaluate the performance we used a computer with the given configurations:

- CPU: AMD Ryzen 9 5950X 16-Core Processor
- RAM: 64 GB
- Storage: TEAM TM8FP6256G
- GPU: NVIDIA GeForce RTX 3080 Ti

The principal deep learning framework employed for model development is TensorFlow, version 2.10.1, ensuring compatibility with the NVIDIA GeForce RTX 3080 Ti GPU by utilizing the ''tensorflow-gpu'' package. The segmentation algorithms extensively rely on well-established Python libraries for various purposes, including numerical

TABLE 1. Summary of U-Net architecture variations.

Variant	Ref.	Key Features
Vanilla U-Net	[14]	Symmetric structure, skip connections, upsampling, and a smaller number of parameters.
Attention U-Net	[16]	Attention mechanisms (self, channel, spatial attention), adaptive feature aggregation.
Dense U-Net	[15]	Dense connections, improved feature reuse, and high-resolution feature maps.
R2 U-Net	[19]	Residual and recurrent connections, increased feature diversity.
Residual U-Net	[42]	Residual connections, improved gradient flow.
SE U-Net	[43]	Squeeze and Excitation blocks, recalibrate the channel-wise feature response.

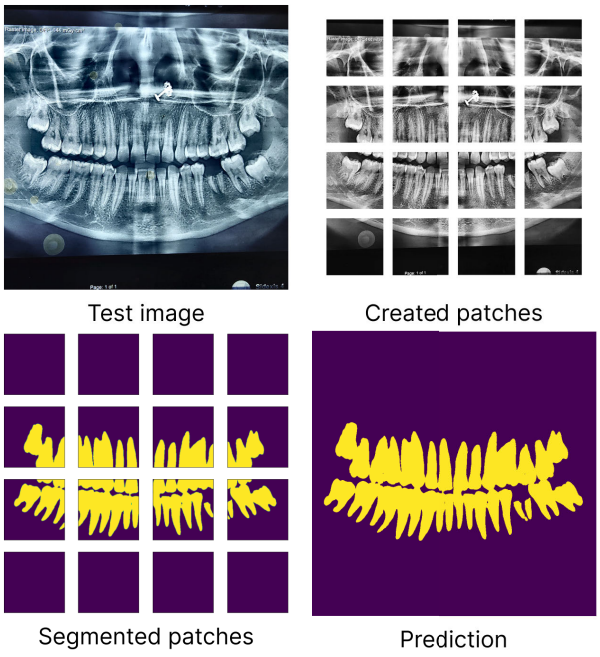

FIGURE 7. Prediction steps.

computing (NumPy), data manipulation (Pandas), image processing (Pillow), and neural network construction (Keras). In addition, other auxiliary libraries were incorporated to enhance the coding experience. The code was executed on a machine with the specified library versions and dependencies to facilitate reproducibility and transparency in the research process.

Our network architectures use a 1×1 convolution layer with stride one followed by a sigmoid activation. The output of these networks gives us binary classification probabilities corresponding to each pixel in the original input teeth X-ray images. All the networks use 2×2 max-pooling and transposed convolution on the downsampling and upsampling, respectively. Furthermore, Network architecture consists of four downsampling and four upsampling layers. The first downsampling layer starts with 16 filters, and as the network goes deep, the filter size increases twice the previous amount. However, the filter size decreases by half the prior amount for the upsampling layers. We use two and three convolution operations per block for performance comparison for each of these encoder and decoder blocks. These convolution operations are followed by batch normalization and ReLU activation functions. For training

TABLE 2. Confusion matrix results.

Architecture	Conv. Layer	TN [%]	FP [%]	FN [%]	TP [%]
Vanilla U-Net	2	0.98	0.02	0.13	0.87
	3	0.98	0.02	0.12	0.88
Attention U-Net	2	0.97	0.028	0.1	0.9
	3	0.98	0.024	0.1	0.9
Dense U-Net	2	0.98	0.018	0.13	0.87
	3	0.98	0.019	0.11	0.89
R2 U-Net	2	0.98	0.025	0.11	0.89
	3	0.98	0.024	0.087	0.91
Residual U-Net	2	0.98	0.019	0.14	0.86
	3	0.98	0.018	0.12	0.88
SE U-Net	2	0.98	0.016	0.14	0.86
	3	0.98	0.019	0.12	0.88

purposes, the models are initially compiled using the Adam optimizer with a learning rate set to 0.0001. A batch size of 16 is employed, and all models undergo training for a total of 100 epochs. Subsequently, the models are fitted using 4,978 patched image and mask pairs for training and validated against 623 patched image and mask pairs. To evaluate the models' performance on the validation set, we utilize metrics such as "accuracy," "dice coefficient," "f1 score," and "IoU" (Intersection over Union). Finally, the dice_coef_loss function is used to calculate the loss.

$$A = \sum_{j=1}^{k} \frac{m_j n_j + \delta}{\sum_{j=1}^{k} (m_j + n_j + \delta)} \tag{3}$$

$$B = \sum_{j=1}^{k} \frac{(1 - m_j)(1 - n_j) + \delta}{\sum_{j=1}^{k} (2 - m_j - n_j + \delta)} \tag{4}$$

$$L = 1 - (A - B) \tag{5}$$

where, the predicted value is m_j, and the corresponding ground-truth value is n_j is the corresponding ground-truth value. To compare the networks, we consider these metrics with the best epoch.

E. PREDICTION ON TEST IMAGES AFTER TRAINING

After training all of our architectures, we predicted some test images using all of the models and each of their variants. The steps are similar for all architectures.

As shown in figure (7), first, we need to create patches out of the image that we want to predict. To do so, we will take a dental x-ray image as our input that we would like to segment i.e., predict using our model. Next, we divide our image into $4 \times 4 = 16$ patches, each with a 256×256 pixel size. Also, we need to normalize all the values before feeding the patches for prediction. As we

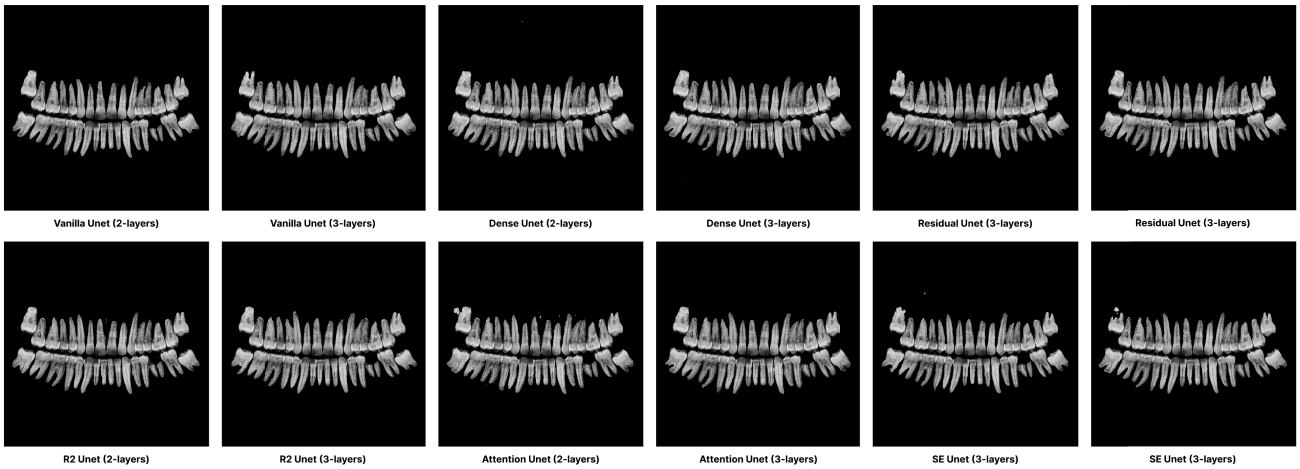

FIGURE 8. Visual inspection on all models.

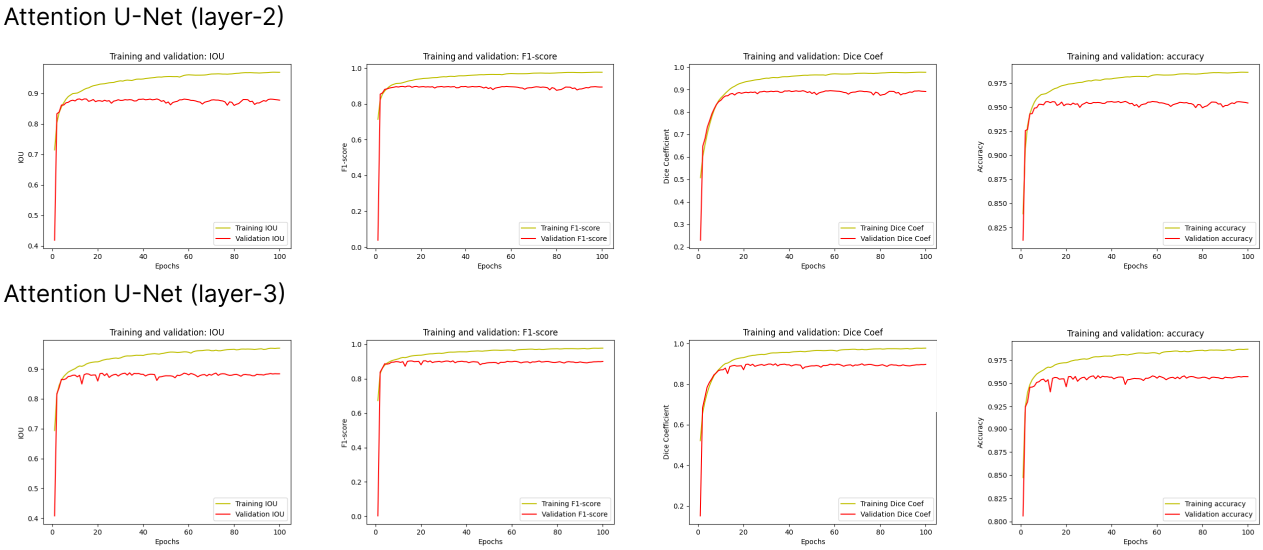

FIGURE 9. Training vs accuracy curves of attention U-Net 2 and 3 layer.

have already trained our model, we can use it to predict and segment an image. So, now we will use our model to predict each of the patches. After predicting all the patches, we will be left with 16 segmented patches for an image. As illustrated in figure (7), after we get the 16 segmented patches, we will reconstruct the entire image. In simpler words, we will put together all 16 segmented patches and form the entire image with a final size of 1024 × 1024 pixels.

IV. EXPERIMENTS AND RESULTS

For segmentation purposes, we used six variations of U-Net Architecture. For each variant, we used 2 and 3-layered architecture. Furthermore, we used tabular, visual, and graphical analysis to evaluate the results of multiple U-Net architectures.

A. EVALUATION METHODS

As we aim to train segmentation models, more than common metrics such as accuracy or f1 score are needed to evaluate our models properly. Alongside these metrics, we have considered our two-layer and three-layer variants on dice coefficient and IoU for further clarification by comparing the training and validation curves. Furthermore, we used a confusion matrix as a tabular analysis to evaluate the performance of our U-Net variants.

1) PERFORMACE METRICS

We have used accuracy, Dice coefficient, F1 score, and Intersection over Union (IoU) to evaluate the performance of our binary image segmentation models. By using these four metrics to evaluate our models' performance, we can better understand how well the models perform on the image

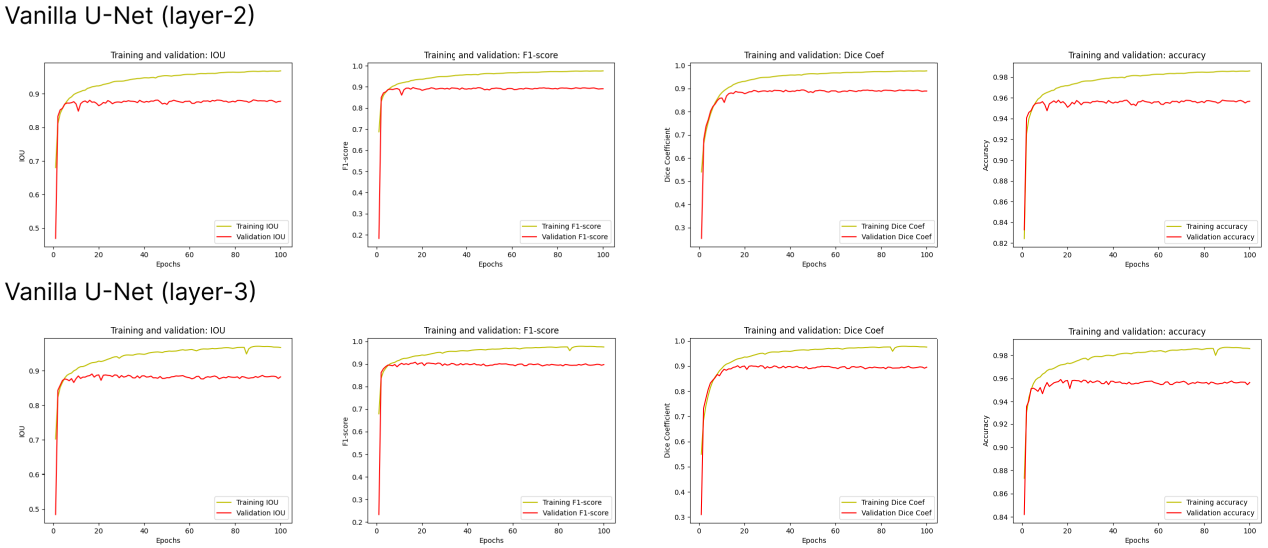

FIGURE 10. Training vs accuracy curves of vanilla U-Net 2 and 3 layers.

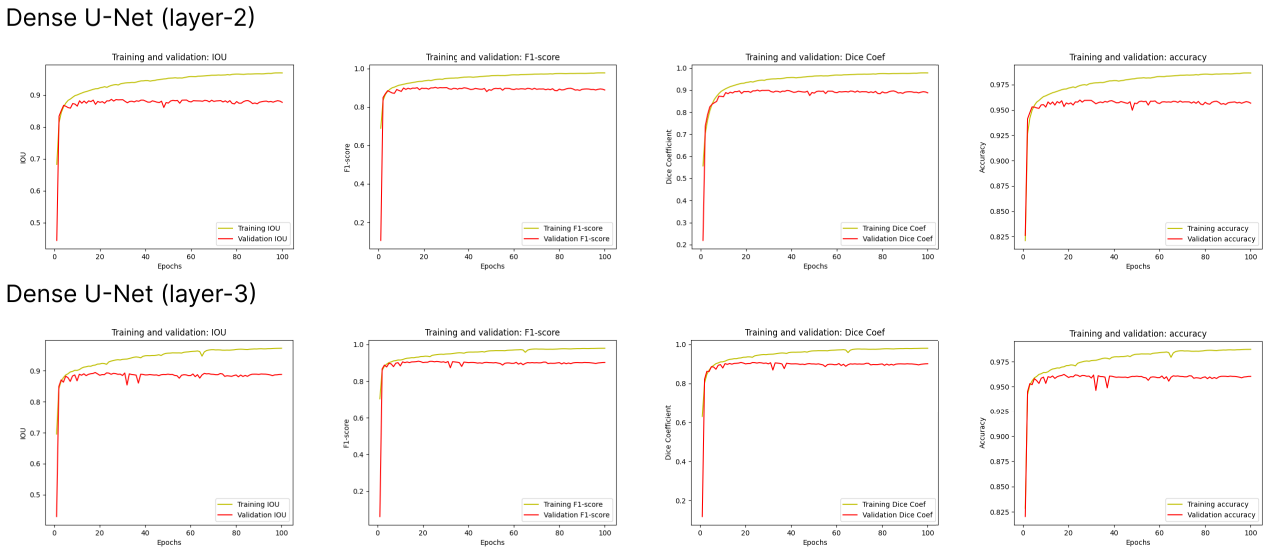

FIGURE 11. Training vs accuracy curves of Dense U-Net 2 and 3 layers.

segmentation task. From figure 9, 10, 11, 12, 13, 14, these curves represent the training vs. validation curve on the validation set. Now, for all four metrics, these validation curves are smooth for all U-Net variants for layers 2 and 3. Our model does not overfit, and all of these curves go up from 87% to 91%.

2) CONFUSION MATRIX

Overall, as shown in table 2 all models seem to be performing well in terms of accurately identifying teeth and background. This combination of high true positive and true negative rates is a positive indication that the models are performing well overall. Both two-layer and three-layer variants have similar performance, but the latter might be slightly more accurate as per their confusion matrices on the test sets.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2 \times Precision * Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{8}$$

In the above set of equations, and in the table

- TP = True Positive
- FP = False Positive

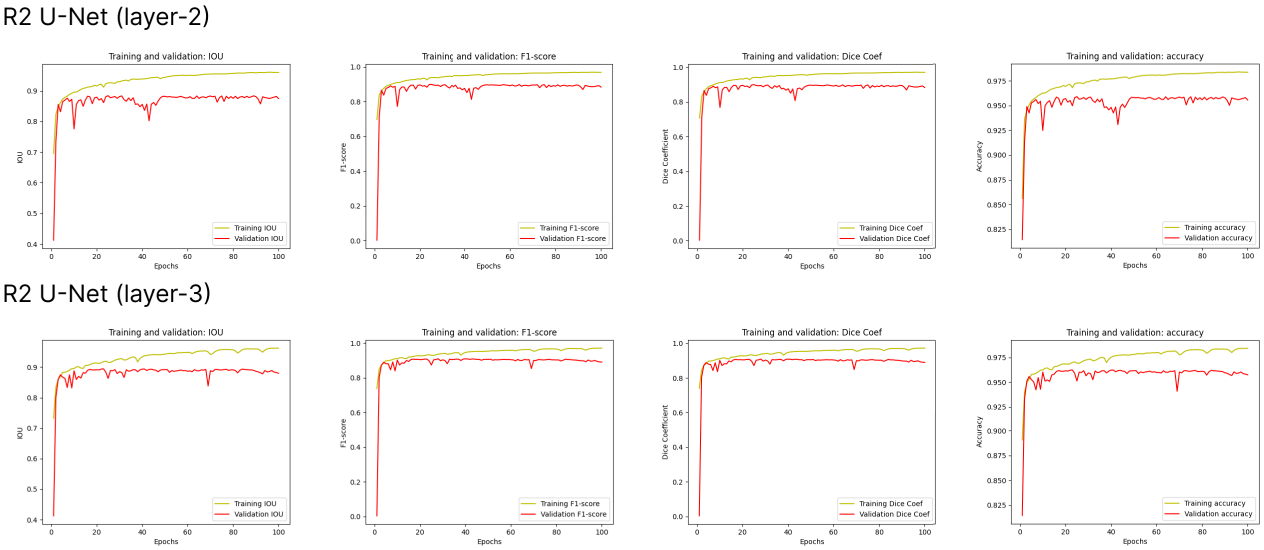

FIGURE 12. Training vs accuracy curves of R2 U-Net 2 and 3 layers.

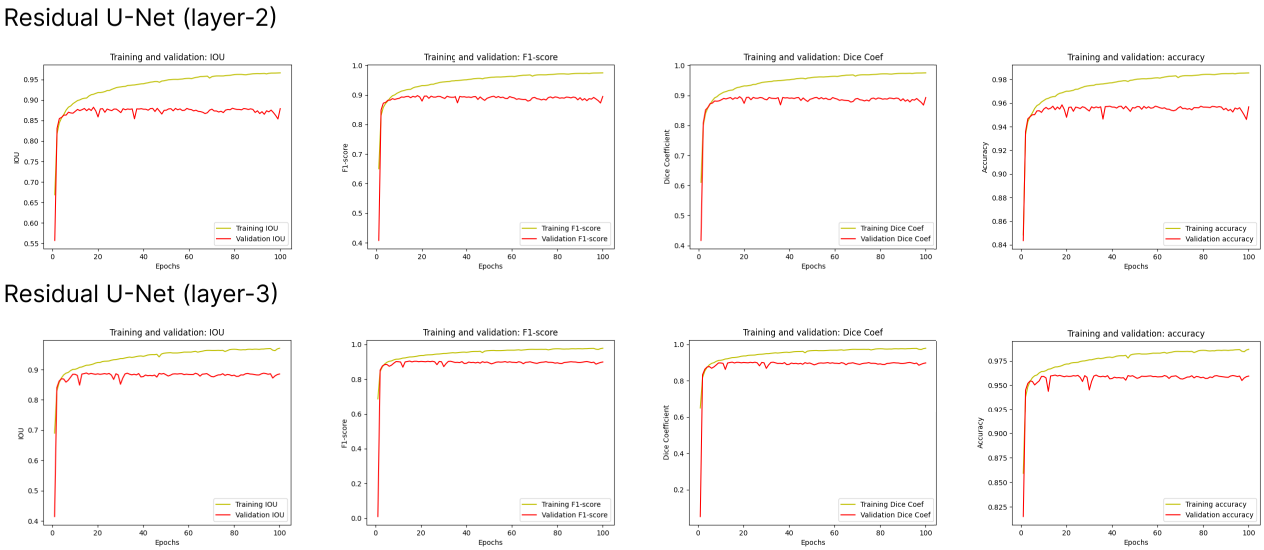

FIGURE 13. Training vs accuracy curves of residual U-Net 2 and 3 layers.

- FN = False Negative
- TN = True Negative

B. RESULTS

The segmentation performance of the OPG teeth image dataset was examined using six distinct U-Net variations in this study. Using two and three convolutional blocks per layer, comparisons of each design were also made.

1) VISUAL INSPECTION

Upon reviewing Figure 8, we conducted a visual assessment of the models' performance. The outcomes indicated that all models exhibited proficiency in accurately recognizing teeth within the test image. As our investigation progressed, we employed a technique known as "Bitwise-and" to isolate the regions where teeth had been segmented from the original test image. On closer scrutiny, we did observe isolated instances where tiny portions of gum were erroneously classified as teeth. It is worth emphasizing that these occurrences were infrequent and had negligible impact. Despite these minor challenges, the models showcased commendable overall performance, particularly in adeptly delineating the principal contours of teeth.

2) GRAPHICAL ANALYSIS

Figure 9, 10, 11, 12, 13, 14 provides a visual summary of segmentation accuracy vs mean epoch train time vs the network parameters for all four metrics. The circle shapes

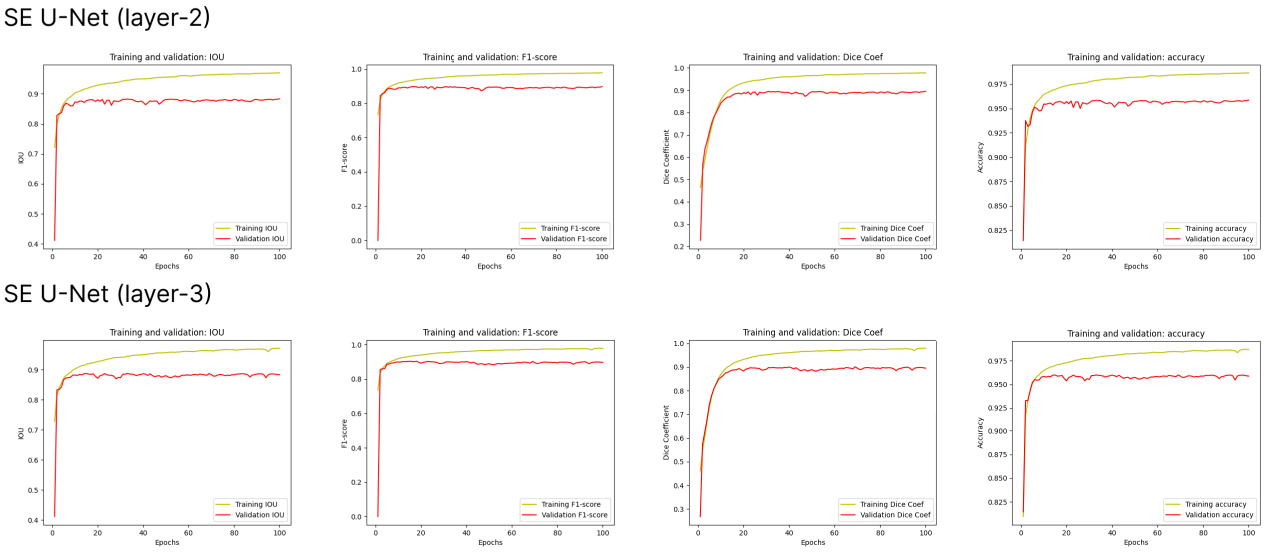

FIGURE 14. Training vs accuracy curves of SE U-Net 2 and 3 layers.

TABLE 3. Comparison of all architectures.

U-Net Architecture	Conv. Layer	Acc. [%]	F1 Score [%]	Dice Coef. [%]	IoU [%]	Mean Epoch Train Time [s]	Par-ams [10^6]
Vanilla U-Net	2	95.56	89.34	89.11	88.00	30	1.9
	3	95.65	89.69	89.53	88.22	40	2.9
Attention U-Net	2	95.44	89.37	89.12	87.76	37	2
	3	95.72	90.02	89.78	88.39	47	3
Dense U-Net	2	95.56	89.24	89.14	87.93	37	2.7
	3	95.94	90.45	90.33	89.07	62	5.4
R2 U-Net	2	95.51	89.41	89.22	87.78	48	2
	3	95.88	90.52	90.35	88.87	82	3
Residual U-Net	2	95.42	88.86	88.73	87.50	31	1.9
	3	95.75	89.82	89.73	88.50	42	2.9
SE U-Net	2	95.51	88.97	88.87	87.84	37	2.1
	3	95.89	90.28	90.15	88.96	48	3

represent the two-layer variants, and the square shapes represent the three-layer variants. Also, the sizes of the two shapes vary based on their network complexities.

Among the different architectures we studied, the three-layer versions of Dense U-Net and R2 U-Net performed the best. However, it's worth noting that Dense U-Net had the most complex structure, while R2 U-Net was the slowest to train and evaluate. One standout model among the three-layer versions is the SE U-Net. It is impressive because its training time is quite similar to most of the two-layer models, yet it achieves one of the highest segmentation accuracies. If we focus on finding a model that trains quickly and has fewer complicated parts, the two-layer Vanilla U-Net is a strong contender. It not only has the simplest architecture among the models but also the shortest training and evaluation times. While Attention U-Net and Residual U-Net might not be the best in terms of segmentation accuracy, their two-layer versions are some of the fastest to train.

To sum it up, the three-layer Dense U-Net and R2 U-Net are top performers, but they come with complexity and longer training times. The SE U-Net is notable for balancing good accuracy and reasonable training time. If we want a simpler and faster model without compromising much on accuracy, the two-layer Vanilla U-Net is a good choice. Attention U-Net and Residual U-Net might not have good accuracy, but their two-layer versions train faster.

Here in table 3 is a more detailed summary of the results obtained from our dataset. We can observe that the accuracy of segmenting teeth using the six U-Net architectures was quite similar. However, we noticed that the performance of each architecture showed a slight improvement when we added an extra layer for processing in each block. This improvement, though, came at the cost of making the models more complex, almost 1.5 times more complex for all the architectures. Additionally, we saw that both training and evaluation times for all the models increased significantly when we added this extra layer. This suggests that when we add this layer, we face a trade-off between factors like how fast the model learns, how complicated the model becomes, and how accurate the segmentation method is. Interestingly, despite the increase in complexity and longer training times, the improvement in segmentation accuracy from the other models was not very noticeable when compared to the standard Vanilla U-Net. This makes us think that opting for a simpler and faster architecture like the Vanilla U-Net might be the best choice for segmenting teeth in panoramic X-ray images.

In simpler terms, this study shows that all the U-Net models performed almost similarly on the segmentation operation. Adding an extra processing layer to each model made them more accurate, but also a lot more complex and slower to train. This made us realize that stability is required between the accuracy of the segmentation model, the efficiency of the model, and its complexity. In contrast to other models, the basic Vanilla U-Net ensures satisfactory accuracy, which is simpler and faster as well. Therefore, suggesting faster, simplified architecture, such as Vanilla U-Net, seems to be the optimal solution for OPG teeth segmentation.

V. CONCLUSION

In this study, we conducted a comprehensive and unbiased analysis, comparing six U-Net architectures, encompassing both two and three-layer variants, for the segmentation of teeth in panoramic x-ray radiographs using our dataset. We aim to contribute insights that can facilitate the development of successful segmentation models. All U-Net models employed in our study demonstrated commendable performance in teeth segmentation from dental panoramic x-rays in our dataset. Upon meticulous examination of the results, particularly using the Dice Coefficient as the primary accuracy metric, we observed minimal differences in performance among the U-Net models. However, recognizing the diverse conditions of clinical applications and time and complexity constraints, our study highlights certain adjustments. Our focus is on providing an in-depth and impartial analysis of U-Net models, offering valuable insights for segmentation approaches that can significantly impact the ongoing evolution of U-Net models, particularly in developing disease diagnosis models utilizing optimal segmentation methods. Notably, our findings indicate that the 3-layer variants of R2 U-Net (Recurrent Residual U-Net) and Dense U-Net exhibit superior performance, achieving Dice Coefficient percentages of 90.35% and 90.33%, respectively. While these models excel in performance and segmentation accuracy, their increased complexity and time requirements should be considered. Considering these factors, the 2-layer variant of the Vanilla U-Net model emerges as an optimal choice for clinical applications, offering a Dice Coefficient percentage of 88.00 with fewer layers and a more time-efficient approach. In conclusion, our study, featuring an impartial analysis and guidance on optimal model selection for teeth segmentation, holds significant potential for future research endeavors, saving valuable time in pursuing the most effective segmentation model. Despite the impressive performance of certain U-Net variants, the fundamental Vanilla U-Net remains a pragmatic choice for practical teeth segmentation in real-world applications due to its balanced performance, lower complexity, and faster processing speed.

CODE AVAILABILITY

The algorithms and code utilized in this study are openly accessible and can be found at https://github.com/rafiatulzannah/Semantic-Segmentation-Using-Panoramic-X-ray-Images

REFERENCES

[1] L. Fiorillo, "Oral health: The first step to well-being," *Medicina*, vol. 55, no. 10, p. 676, Oct. 2019.

[2] P. Amrollahi, B. Shah, A. Seifi, and L. Tayebi, "Recent advancements in regenerative dentistry: A review," *Mater. Sci. Eng., C*, vol. 69, pp. 1383–1390, Dec. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0928493116308827

[3] E. D. Beltrán-Aguilar, L. K. Barker, and M. T. Canto, "Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis–United States, 1988–1994 and 1999–2002," *MMWR Surveill. Summ.*, vol. 54, no. 3, pp. 1–43, 1988.

[4] S. Kiatpongsan, R. S. Huckman, and M. D. Hornstein, "The great recession, insurance mandates, and the use of in vitro fertilization services in the United States," *Fertility Sterility*, vol. 103, no. 2, pp. 448–454, Feb. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0015028214023024

[5] WH Organization. (2023). *Oral Health*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/oral-health

[6] C. W. Douglass, R. W. Valachovic, A. Wijesinha, H. H. Chauncey, K. K. Kapur, and B. J. McNeil, "Clinical efficacy of dental radiography in the detection of dental caries and periodontal diseases," *Oral Surgery, Oral Med., Oral Pathol.*, vol. 62, no. 3, pp. 330–339, Sep. 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0030422086900174

[7] X. Qin, M. Xu, C. Zheng, C. He, and X. Zhang, "Multi-scale feedback feature refinement U-net for medical image segmentation," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2021, pp. 1–6.

[8] Y. W. Chen, K. Stanley, and W. Att, "Artificial intelligence in dentistry: Current applications and future perspectives," *Quintessence Int.*, vol. 51, no. 3, pp. 248–257, 2020.

[9] T. O. Frizzell, M. Glashutter, C. C. Liu, A. Zeng, D. Pan, S. G. Hajra, R. C. N. D'Arcy, and X. Song, "Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: A systematic review," *Ageing Res. Rev.*, vol. 77, May 2022, Art. no. 101614.

[10] A. Mansur, A. Vrionis, J. P. Charles, K. Hancel, J. C. Panagides, F. Moloudi, S. Iqbal, and D. Daye, "The role of artificial intelligence in the detection and implementation of biomarkers for hepatocellular carcinoma: Outlook and opportunities," *Cancers*, vol. 15, no. 11, p. 2928, May 2023.

[11] J. S. Ahn, S. Shin, S.-A. Yang, E. K. Park, K. H. Kim, S. I. Cho, C.-Y. Ock, and S. Kim, "Artificial intelligence in breast cancer diagnosis and personalized medicine," *J. Breast Cancer*, vol. 26, no. 5, p. 405, 2023.

[12] E. Dack, A. Christe, M. Fontanellaz, L. Brigato, J. T. Heverhagen, A. A. Peters, A. T. Huber, H. Hoppe, S. Mougiakakou, and L. Ebner, "Artificial intelligence and interstitial lung disease: Diagnosis and prognosis," *Invest. Radiol.*, vol. 58, no. 8, pp. 602–609, Aug. 2023.

[13] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of U-Net," 2022, *arXiv:2211.14830*.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assisted. Intervent.* Berlin, Germany: Springer, 2015, pp. 234–241.

[15] S. Cai, Y. Tian, H. Lui, H. Zeng, Y. Wu, and G. Chen, "Dense-UNet: A novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network," *Quant. Imag. Med. Surgery*, vol. 10, no. 6, pp. 1275–1285, Jun. 2020. [Online]. Available: https://qims.amegroups.com/article/view/43519

[16] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[17] L. Rundo, C. Han, Y. Nagano, J. Zhang, R. Hataya, C. Militello, A. Tangherloni, M. S. Nobile, C. Ferretti, D. Besozzi, M. C. Gilardi, S. Vitabile, G. Mauri, H. Nakayama, and P. Cazzaniga, "USE-net: Incorporating squeeze-and-excitation blocks into U-net for prostate zonal segmentation of multi-institutional MRI datasets," *Neurocomputing*, vol. 365, pp. 31–43, Nov. 2019.

[18] D. Li, D. A. Dharmawan, B. P. Ng, and S. Rahardja, "Residual U-net for retinal vessel segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1425–1429.

[19] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.

[20] M. Jafari, D. Auer, S. Francis, J. Garibaldi, and X. Chen, "DRU-net: An efficient deep convolutional neural network for medical image segmentation," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1144–1148. [Online]. Available: https://api.semanticscholar.org/CorpusID:216562414

[21] N. Vila Blanco, I. Tomás Carmona, and M. Carreira, "Fully automatic teeth segmentation in adult OPG images," *Proceedings*, vol. 2, p. 1199, Sep. 2018.

[22] T. L. Koch, M. Perslev, C. Igel, and S. S. Brandt, "Accurate segmentation of dental panoramic radiographs with U-NETS," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 15–19.

[23] S. S. Alharbi, A. A. AlRugaibah, H. F. Alhasson, and R. U. Khan, "Detection of cavities from dental panoramic X-ray images using nested U-Net models," *Appl. Sci.*, vol. 13, no. 23, p. 12771, Nov. 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/23/12771

[24] X. Qin, M. Xu, C. Zheng, C. He, and X. Zhang, "Multi-scale feedback feature refinement U-Net for medical image segmentation," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2021, pp. 1–6.

[25] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

[26] F. G. Zanjani, A. Pourtaherian, S. Zinger, D. A. Moin, F. Claessen, T. Cherici, S. Parinussa, and P. H. N. de With, "Mask-MCNet: Tooth instance segmentation in 3D point clouds of intra-oral scans," *Neurocomputing*, vol. 453, pp. 286–298, Sep. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231221001041

[27] T. J. Jang, K. C. Kim, H. C. Cho, and J. K. Seo, "A fully automated method for 3D individual tooth identification and segmentation in dental CBCT," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6562–6568, Oct. 2022, doi: 10.1109/TPAMI.2021.3086072.

[28] 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 Comput. Appl.*, vol. 35, no. 11, pp. 8295–8306, Apr. 2023.

[29] J. Im, J.-Y. Kim, H.-S. Yu, K.-J. Lee, S.-H. Choi, J.-H. Kim, H.-K. Ahn, and J.-Y. Cha, "Accuracy and efficiency of automatic tooth segmentation in digital dental models using deep learning," *Sci. Rep.*, vol. 12, no. 1, p. 9429, Jun. 2022.

[30] C. Sheng, L. Wang, Z. Huang, T. Wang, Y. Guo, W. Hou, L. Xu, J. Wang, and X. Yan, "Transformer-based deep learning network for tooth segmentation on panoramic radiographs," *J. Syst. Sci. Complex.*, vol. 36, no. 1, pp. 257–272, Oct. 2022. [Online]. Available: https://europepmc.org/articles/PMC9976655

[31] Z. Chen, S. Chen, and F. Hu, "CTA-UNet: CNN-transformer architecture UNet for dental CBCT images segmentation," *Phys. Med. Biol.*, vol. 68, no. 17, Aug. 2023, Art. no. 175042.

[32] F. Nogueira-Reis, N. Morgan, S. Nomidis, A. Van Gerven, N. Oliveira-Santos, R. Jacobs, and C. P. M. Tabchoury, "Three-dimensional maxillary virtual patient creation by convolutional neural network-based segmentation on cone-beam computed tomography images," *Clin. Oral Investigations*, vol. 27, no. 3, pp. 1133–1141, Sep. 2022.

[33] 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, Jan. 2023, Art. no. 106296.

[34] L. Liu, J. Xu, Y. Huan, Z. Zou, S.-C. Yeh, and L.-R. Zheng, "A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 898–906, Mar. 2020.

[35] X. Wang, Z. Hu, S. Shi, M. Hou, L. Xu, and X. Zhang, "A deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved UNet," *Sci. Rep.*, vol. 13, no. 1, pp. 1–13, May 2023.

[36] 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 Comput. Appl.*, vol. 35, no. 22, pp. 16051–16059, Aug. 2023.

[37] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44247–44257, 2019.

[38] P. Ahmad, H. Jin, R. Alroobaea, S. Qamar, R. Zheng, F. Alnajjar, and F. Aboudi, "MH UNet: A multi-scale hierarchical based architecture for medical image segmentation," *IEEE Access*, vol. 9, pp. 148384–148408, 2021.

[39] B. M. Elgarba, S. Van Aelst, A. Swaity, N. Morgan, S. Shujaat, and R. Jacobs, "Deep learning-based segmentation of dental implants on cone-beam computed tomography images: A validation study," *J. Dentistry*, vol. 137, Oct. 2023, Art. no. 104639.

[40] Y. Deng, Y. Hou, J. Yan, and D. Zeng, "ELU-net: An efficient and lightweight U-net for medical image segmentation," *IEEE Access*, vol. 10, pp. 35932–35941, 2022.

[41] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2014, pp. 2392–2397.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[43] A. Guha Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel squeeze & excitation in fully convolutional networks," 2018, *arXiv:1803.02579*.

RAFIATUL ZANNAH received the bachelor's degree in computer science and engineering from Brac University. With a passion for cutting-edge technologies and a strong foundation in computer science, Zannah has actively engaged in research and academic pursuits. Driven by a curiosity to explore and contribute to the evolving landscape of computer science, she has embarked on a journey of research and innovation. They have a keen interest in computer vision, deep learning, and machine learning, and their work reflects a commitment to advancing the field.

MUBTASIM BASHAR received the bachelor's degree in computer science and engineering from Brac University. He is a passionate Data Science enthusiast. His academic journey is flavored with hands-on experience in deep learning, machine learning, and data science. With a keen interest in pushing the boundaries of these domains, he focuses on the intersection of deep learning, image processing, and computer vision. His commitment to innovative research is evident in his academic achievements and professional pursuits, showcasing a genuine enthusiasm for advancing practical applications in these fields.

RAHIL BIN MUSHFIQ received the bachelor's degree in computer science and engineering from Brac University. He is a tech enthusiast with a keen interest in problem-solving, he is poised for a dynamic career. During University, he honed his skills in software development and emerged with a solid understanding of computer science principles. His passion lies in leveraging technology to create impactful solutions. With a foundation built on innovation and a commitment to excellence, he looks forward to contributing significantly to the tech industry.

AMITABHA CHAKRABARTY received the M.Sc. degree from the Department of Computer Science and Engineering, University of Rajshahi, in 2004, the M.Sc. degree in telecommunication engineering from Independent University, Bangladesh, and the Ph.D. degree from the Faculty of Engineering and Computing, Dublin City University, Dublin, Ireland, in 2012. He is currently a Professor with the Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh. He has published research papers in various national and international conferences and journals and book chapters. He is involved in active research having a number of graduate and undergraduate research groups in different research projects. His research interests include the Internet of Things (IoT), machine learning, deep learning, embedded systems, and switching theory. He is serving as a TCP member for various international journals and conferences. He is also serving as a senior judge for various national IT competitions.

SHAHRIAR HOSSAIN received the B.Sc. degree in computer science and engineering from Brac University, in 2021. He is currently pursuing the Ph.D. degree with the Department of Computer Science, George Mason University. He has attended several national and international competitions in the field of robotics. His current research interests include machine learning, deep learning, vision transformers, and robotics.

YONG JU JUNG (Member, IEEE) received the Ph.D. degree from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2005. From 2005 to 2010, he was a Principal Research Scientist with the Samsung Advanced Institute of Technology, contributing to 3D display processing for 3D TV. From 2010 to 2014, he was an Associate Research Professor with the Image and Video Systems Laboratory, Department of Electrical Engineering, KAIST. From 2014 to 2016, he was a Principal Engineer with the System LSI Division, Samsung Electronics, contributing to the development of image sensors and multi-camera solutions. Since 2016, he has been an Associate Professor with the School of Computing and the Director of the Computer Vision and Image Processing Laboratory (CVIP Laboratory), Gachon University. His current research interests include image processing, computer vision, and deep learning. He was a recipient of the Samsung Muhan Research Award in 2010 and the Best Innovation Award from Samsung Electronics in 2015. He co-organized special sessions on human 3D perception and 3D video assessments in DSP2011. He is an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.