Hierarchical Feature Learning for Accurate Pixel-Wise Segmentation of Cardiac Structures in CT Images: A Comprehensive Evaluation of U-Net Variants

Shafiya Qadeer Memon[1], Sania Bhatti[1], Muhammad Moazzam Jawaid[2], Mehran Memon[3], Gulzar Usman[4]
[1]Department of Software Engineering, Mehran University of Engineering & Technology Jamshoro, Pakistan.
[2]De Montfort University, Leicester United Kingdom
[3]Sindh Institute of Ophthalmology Visual Sciences (SIOVS) Hyderabad
[4]Liaquat University of Medical & Health Sciences
*__Correspondence__: shafiya.memon@faculty.muet.edu.pk

Citation | Memon. S. Q, Bhatti. S, Jawaid. M. M, Memon. M, Usman. G, Memon. S. Q, "Hierarchical Feature Learning for Accurate Pixel-Wise Segmentation of Cardiac Structures in CT Images: A Comprehensive Evaluation of U-Net Varients", IJIST, Vol. 8 Issue. 2 pp 728-742, April 2026
Received | March 17, 2026 **Revised |** April 19, 2026 **Accepted |** April 24, 2026 **Published |** April 29, 2026.

Precise segmentation of cardiac structures from computed tomography angiography (CTA) images is essential for the early analysis of cardiovascular conditions. This study presents a robust deep learning model for precise segmentation of cardiac structures from CTA images, with particular emphasis on pixel-level accuracy and hierarchical convolutional feature learning to delineate thin and low-contrast vessels. Four different variants of the U-Net architecture such as Vanilla U-Net, U-Net4, Attention U-Net, and U-Net++ were evaluated on a publicly available CTA image dataset. The Vanilla U-Net had the best overall accuracy of 99% and precision of 98%, indicating excellent feature learning with very few false positives. However, its recall of 91% was slightly lower, indicating some under-segmentation of the distal vessels. Attention U-Net and U-Net4 had the same accuracy of 98% and F1-score of 95%. Attention U-Net improved precision to 98% with attention-driven feature refinement, while U-Net4 improved recall to 93% with multi-scale feature aggregation. U-Net++ had a slightly lower accuracy of 97% but the best recall of 94%, indicating its dense skip connections that enable the refinement of fine details of the vessels. For all models, the Dice score varied from 92-94% and IoU from 88-89%, while AUC scores were all above 0.99, indicating excellent segmentation performance. Without performing any statistical test or confidence interval calculation, the results show a consistent pattern across all evaluation metrics considered in the study, which suggests that the comparisons between models are reliable. These results demonstrate that by leveraging deep learning-based feature learning accurate and continuous coronary artery segmentations can be obtained. The proposed method has promising potential in improving the non-invasive diagnosis of CAD and reducing the dependence on invasive procedures.

Keywords: Segmentation, Hierarchical Deep Features, Data-Efficient Learning, Structure-Preserving Segmentation, Medical Image Analysis, U-Net Variants

Introduction:

Cardiovascular diseases still rank among the top causes of death worldwide. This implies that detecting and accurately evaluating the heart at an early stage is of critical importance. The most commonly employed imaging technique today is CT scan technology, primarily because of the detailed and high-quality images it produces of the heart and the vessels. To make the best out of this information, it is critical to accurately detect and draw different parts of the heart, especially the chambers and vessels. This will help doctors accurately analyze cardiac structures for diagnosis and treatment planning and even diagnose and treat various conditions [1][2].

However, it is not easy to perform this segmentation of the heart manually. This process is time-consuming and requires specialized knowledge and expertise. Moreover, even with specialized knowledge and expertise, it is still possible to obtain varying results from different experts. The fact that this process is complicated by the complicated structure of the heart, the low contrast of various tissues, and the thin and curvy nature of some of the vessels makes it even more complicated. The traditional approaches, which depend on models or anatomical templates, are not always effective for each patient, as they fail to accommodate the anatomical differences between individuals [3][4][5][6].

Recently, artificial intelligence, particularly deep learning, has begun to transform the process of analyzing medical images. For instance, convolutional neural networks are proving to be extremely efficient tools. They can easily learn from images, making them an efficient solution. Architectures such as the encoder-decoder style, which is similar to the popular U-Net, have gained traction due to their efficiency in analyzing images. They can process images from start to finish, providing precise results. Even though this has been enhanced, there are a few issues. For example, small structures like blood vessels are difficult for the model to recognize, and there is a data imbalance since they only occupy a very small percentage of the image. In addition, the vessels are not very prominent, and the image quality varies. Also, there is a lack of labeled data for training. Even advanced models struggle to capture fine structural details and maintain vessel continuity, especially when the image quality is very poor [7][8][9][10].

Moreover, recent comparative studies suggest that there is no universal segmentation model that would produce outstanding results in all cases. The image quality, variation in anatomy, and imbalance of classes continue to play a role in the precision of segmentation at the pixel level. As a result, there is a growing need for optimized hybrid deep learning models that would incorporate all these factors into a single framework [11][12][13][14].

To address these difficulties, this research aims to develop a robust deep learning model capable of precisely segmenting cardiac structures from CT scans, paying particular attention to enhancing accuracy at the pixel level. This is achieved through the identification of smaller, less contrasted, and highly complex regions via the use of layer-by-layer feature extraction and highly optimized variants of the U-Net model.

Objectives:

The main objective of this study is to develop a robust deep learning-based framework for accurate cardiac CT image segmentation with improved pixel-level precision. The study also aims to enhance the segmentation of small and low-contrast cardiac structures through hierarchical feature extraction and optimized U-net architectures. Furthermore, different U-net based models are evaluated and compared to identify the most effective architecture for cardiac image segmentation.

Moreover, various models will be compared to determine the best model for cardiac segmentation. Cardiovascular diseases remain among the main causes of death on a global scale, which makes it necessary to pay special attention to their prompt detection and proper diagnosis. This research is crucial since it aims to deal with the challenges associated with

obtaining high pixel-level accuracy in cardiac structures segmentation. With the help of deep learning and hierarchical feature extraction, the algorithm is capable of delivering high-quality results and avoiding mistakes and inconsistencies that typically arise in manual segmentation. As a result, this leads to a clear boundary between the structures and their exact visualization. Such an approach may facilitate further diagnosis and even contribute to better treatment options. Moreover, comparing various U-Net-based algorithms will provide interesting insights for future studies. The key contributions are as follows:

The proposed model employs a deep learning framework that seeks to boost the segmentation of cardiac CT images through improved accuracy at the pixel level.

The design employs hierarchical feature extraction to enable the network to learn the general picture and detailed features of the image simultaneously.

Convolutional Neural Network (CNN) architecture is employed as the backbone of the model to ensure better feature representation and performance in segmentation.

The loss function is hybridized using two functions, Binary Cross-Entropy and Dice Loss, to ensure pixel accuracy and overlapping regions.

The architectures tested include vanilla U-Net, U-Net++, Attention U-Net, and other variants of U-Net.

The results indicate better accuracy at the pixel level, small organ and low contrast object segmentation, and accurate boundaries with structure consistency.

The rest of the paper is organized as follows: Section 2 describes the limitation of the current studies to identify the research gap. Section 3 presents the overall methodology steps were performed to conduct study. In Section 4 key findings were interpreted and discussed. Section 5 concluded the main findings and provides future directions.

Literature Review:

Pre-preservation of organs at risk (OARs) is a healthy tissue around cancer that is to be preserved during radiotherapy (RT). A deep learning-based automated heart segmentation was suggested in this present study [15]. Atlas-based segmentation and Concatenation Block-U-Net were used to segment the coarse segmentation and image processing to segment the fine segmentation. Public database of 36 cases of CT scans of patients to undergo radiotherapy. The model obtain 95.25% of the Dice similarity coefficient, 87.95% of the Jacquard (JAC) Index, 96.71% of sensitivity, and 99.39% of accuracy. The mean DSC of the entire heart is over 0.95. There was no remarkable change in analysis of variance among the four age groups; less than one year, 1y to 4y, 5y to 9y, 10y to 14y. Mean DSC of every chamber trained in a lump is 0.78-0.88. When these DSC are trained individually, they are 0.80-0.85 [16]. Similarly, a new preprocessing stage is suggested in this paper [14], which aids in the accurate isolating of the blood vessels. Another area that has to be tested is the influence of such preprocessing steps on the other algorithms. To start with, the preprocessing steps should be deployed and tested. The proposed steps then have to go through the preprocessing steps to remove the blood vessels in binarization. The background images in the retina are characterized by the identification and localization of the vessels in the retina. The most significant semantic segmentation model to Cardiac MRI is UNet with attention-based models. The suggested approach Detail Preserving Attention UNet (DPA-UNet) improves the extent of clinical diagnosis to find the necessary components of analysis accurately [17]. According to another study [18], a network structure of the encoder-decoder network was suggested, Adaptive Feature Medical Segmentation Network (AFMS-Net) with two variants of encoders, including Single Adaptive Encoder Block (SAEB) and Dual Adaptive Encoder Block (DAEB), was suggested. The proposed theory, AFMS-Net, delivers the state-of-the-art performance on a range of benchmarking datasets, such as BRATS 2021, ATLAS 2021, and ISLES 2022. Different architectures, i.e. U-net, Unet++, Attention U-net, and ResuNet++ were employed. The accuracy was increased by Unet, which was up to 0.988. More than 100

papers connected with cardiac image segmentation with the help of deep learning techniques are discussed, including the most common types of imaging, i.e., magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, and the most common structures of interest in the cardiovascular system, i.e. the ventricles, atria and vessels [19]. This systematic review was aimed at investigating the different techniques that can be applied in the automatic segmentation of the organs at risk in thoracic computer topographies and discussing the most appropriate technique which gives the highest accuracy in terms of segmentation among all the techniques that were proposed. The different methods, data sets, precision, and other difficulties were presented in the framework of this research field [20]. The study employed in this paper [21] relied on the 4D contrast-enhanced cardiac CT images of 1509 patients who were selected to undergo transcatheter aortic valve implantation with 21,605 3D images. The data were randomly split in the development set of N = 12 and the test set of N = 1497. The automatic segmentation showed a mean value of DSC = 0.89 ± 0.10 and ASSD = 1.43 ± 1.45 mm in 12 patients in 3D and DSC = 0.89 ± 0.08 and ASSD = 1.86 ± 1.20 mm in 81 patients in 2D.

Table 1. Limitation of Existing Studies

Study	Methods	Dataset	Key Findings	Limitations
[15]	Atlas-based + U-Net (Concatenation Block)	CT(36 Patients)	DSC =95.25% Accuracy =99.39%	Limited dataset, weak generalization
[16]	Deep CNN based Heart segmentation	CT	DSC = 0.78-0.88 (Chambers)	Inconsistent performance across structures
[14]	Preprocessing + vessel enhancement	Retinal / Vascular images	Improved vessel visibility	Not fully integrated with DL models
[17]	DPA-U-Net	MRI	Improved clinical feature extraction	High computational complexity
[18]	AFMS-Net (SAEB & DAEB)	BRATS, ATLAS, ISLES	State-of-the-art performance	Complex architecture, high resource demand
[19]	Systematic Review	More than 100 Papers related to cardiac	In depth review was explored	No practical analysis was performed
[20]	Systematic Review	Literature Review	Dataset, Models	No practical analysis was performed
[21]	3D CT segmentation model	4D CT	DSC = 0.89	Requires Large Annotated dataset
Proposed study	Vanilla U-Net, U-Net++, Attention U-Net	CT Binary Mask	Accuracy upto 99%	Limited Multi-Center Validation

Table 1 summarizes benchmark results (ImageCAS and ASOCA), which show that existing models demonstrate superior performance to existing models both in terms of Dice score improvements and boundary delineation. However, some significant challenges remain unsolved. For example, only a few automated segmentation algorithms for coronary arteries have been developed that can operate under fully supervised learning models with annotated 3D images. In fact, a majority of the existing solutions require the time-consuming, expensive, and clinically difficult manual image annotation process. This difficulty is multiplied due to the

thin, highly curved and closely spaced coronary arteries and low-contrast surrounding tissues of the surrounding tissue. Semi-supervised and positive-unlabeled learning approaches have been explored; however, their success remains limited and have shown little to no applicability when evaluating multiple centers' varying imaging protocols. Furthermore, there is limited research examining whether explicitly providing classical vessel enhancement responses as structured priors can improve deep network sensitivity to thin branches, enhance morphological consistency, and reduce fragmentation without requiring additional annotation effort. The potential of combining learned hierarchical features with analytically derived vascular cues in an end-to-end trainable framework therefore remains underexplored.

However, many of today's segmentation algorithms focus primarily on optimizing voxel overlap-based metrics such as the Dice Similarity Coefficient, none of these prior models consider for the real-world clinical issues associated with ensuring clinical reliability, domain generalizability, and uncertainty estimation. Furthermore, most existing studies use standard metrics that do not take into account the effects of low-contrast imaging conditions.

Material and Methods:

This section describes a robust deep learning model for precise segmentation of cardiac structures from CT scans from CT scans, paying particular attention to enhancing accuracy at the pixel level framework in detail. The entire process begins with the collection of dataset, followed by preprocessing, intensity normalization, feature extraction, model training, and evaluation of multiple U-Net variants as shown conceptually in Figure 1.

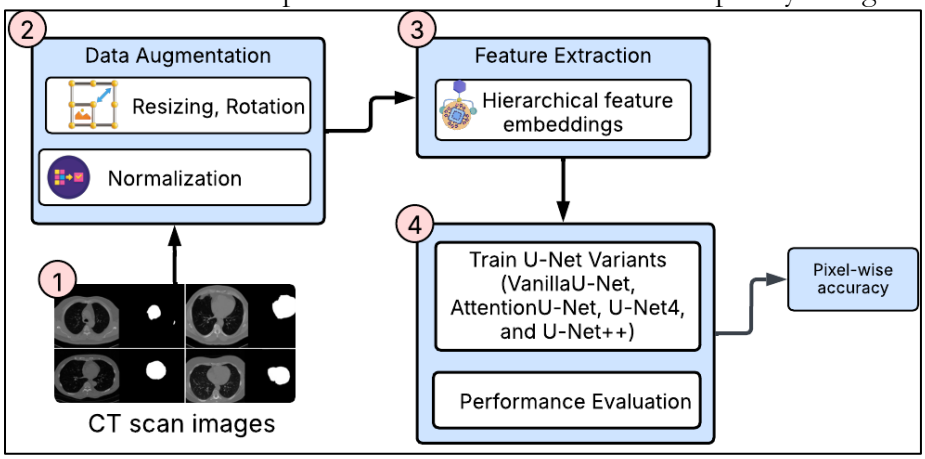

Figure 1. Conceptual overview of proposed model

Figure 1 illustrates the overall flow of the proposed framework for segmenting cardiac CT images. First, the dataset is collected by acquiring the CT images and their binary masks from a public Kaggle repository. After collecting the data, it undergoes preprocessing and augmentation where normalization is performed, and various random transformations are applied. Normalization and random transformations help in improving the generalization capacity of the model while avoiding overfitting. It should be noted that the spatial relationships between each image and its corresponding mask during this stage. Upon completing the preprocessing phase, the data is input into the feature extraction step. Hierarchical convolutional layers learn the necessary features at various depths, both at fine texture level and contextual level of the cardiac CT slice. Subsequently, the feature-extracted data is used for where various U-Net variants are trained various variants of the U-Net, including the Vanilla U-Net, U-Net++, Attention U-Net, and modified U-Net architecture, are trained via encoder-decoder mechanism. Upon completion of the training process, the performance of the trained models will be assessed through the experimentation and evaluation phase by measuring the difference between the predicted segmentation mask and the annotated one. These performance measures will be assessed quantitatively using metrics

like Dice Similarity Coefficient, Intersection over Union, Accuracy, and AUC. A comparison will then be made to select the best performing U-Net model for cardiac CT segmentation.

Dataset Collection: The dataset used in this study was obtained from publically available kaggle repository [22]. The dataset contains CT scan images and their respective binary masks. The dataset includes 19 different subjects where each subject contains images and their masks. Even though the initial dataset contained 19 subjects, the number of samples was reduced to 15 to train deep learning algorithms on images acquired using high resolution CT scans due to limited hardware capacity. This decision was made due to the large amount of slices per each subject and corresponding segmentation masks that increased the memory consumption and time spent on the training process. To ensure that experiments remained feasible yet had enough diversity in terms of the sample population, it was decided to use a random selection of 15 subjects from the original pool of 19 patients. Table 2 shows the image and corresponding mask samples per subjects.

Table 2. Per subject dataset samples

Subjects	Images Samples	Masks samples
100051	131	131
100053	112	112
100056	132	132
100058	148	148
100065	122	122
100067	137	137
100069	142	142
100072	63	63
100072	143	143
100075	152	152
100079	66	66
100080	142	142
100082	124	124
100089	174	174
100091	131	131

Preprocessing and Data Augmentation:

To boost the model's capacity to generalize and minimize overfitting, especially when dealing with the relatively small amount of annotated medical images, a stochastic data augmentation technique is used during training. This method can be considered as a way to augment the training using set of label-preserving transformations to the input images and their corresponding masks [23].

This ensures spatial alignment between the image and its corresponding mask. The medical imaging datasets are generally small and do not contain the entire range of anatomical variability that could be observed in real-world settings. This has the potential to overfit the training data by the deep convolutional networks, which learn the patterns that do not apply in new cases.

Feature Extraction:

The feature extraction module is intended to generate representations from cardiac CT slices that are highly discriminative and structurally rich [24]. Unlike conventional segmentation methods that rely solely on deep convolutional features, this approach leverages two distinct sources of information:

Hierarchical Convolutional Feature Embeddings:

The hierarchical convolutional feature extraction module is intended to learn multi-scale semantic representations from the input cardiac CT slices. These embeddings, in contrast

to purely handcrafted features, encode the local texture patterns as well as global contextual information, which is critical to the accurate segmentation of cardiac structures [25][26]. Given the dual-channel input $X \in \mathbb{R}^{2 \times H \times W}$ (normalized image + feature map) the encode compute as set of hierarchical features f correspond to multi-scale feature maps at successive encoder levels. The final feature set can be calculated using equation (8):

$$\Phi_{deep}(X) = \{f_1, f_2, \ldots\ldots f_n\} \quad (8)$$

Model Development:

Figure 2. U-Net Variants

Fig. 2 depicts the 4 variants of U-net applied in this paper. The three-level U-Net architecture is the same in all four models, and the input has two channels, namely the original CT scan and the corresponding vessel enhancement map. The models vary in the depth, attention mechanism, and skip connection to enhance performance of segmentation.

Experimental Setup:

The experiments were performed using Kaggle Notebook and well-known Python environment [27]. The Adam optimizer was used to train all the models and the learning rate was 0.001. The hybrid loss based on the Binary Cross-Entropy Loss and Dice Loss was used to mitigate the problem of class imbalance and come up with a more precise boundary. Pixel-level prediction accuracy is improved by the Binary Cross-Entropy Loss, whereas the Dice Loss encourages a better level of overlap between the predictions and the ground truth masks. Training was done in batch size of 8 and where possible, computation was done using a CUDA-compatible GPU. However, the U-Net networks used in this study had different levels of computational complexity and training characteristics owing to differences in the depth of the network, attention layers, and skip connections. Vanilla U-Net and similar basic architectures required low computational capabilities and usually took shorter times for training. Contrastingly, more complex models, such as Attention U-Net and U-Net++, consumed high computational resources because of their sophisticated processes of extracting features. Understanding these disparities gave an understanding of how to balance accuracy and computational resources in clinical settings.

Hybrid Loss Function:

This is a hybrid loss that is a combination of Binary Cross-Entropy (BCE) Loss function and Dice Loss function. The BCE Loss optimizes the cost of each pixel independently and in the process the distribution of the probability of every pixel in the mask is predicted correctly and the quality of the prediction of the pixels in the mask is maximized [28][29]. Conversely, the Dice Loss functional quantifies the extent of overlap between the predicted mask and the actual mask leading to maximization of overlap between the two masks regardless of the imbalance in classes. This loss function is well suited to segment thin-walled vessels and small cardiac structures. The hybrid loss defined as

$$\mathcal{L}_{hybrid} = \alpha BCE(y, y') + \beta DiceLoss(y, y') \quad (9)$$

Where, α and β are weighting coefficients that controls the contribution of BCE loss and Dice Loss respectively.

The BCE loss is calculated as:

$$BCE(y, y') = -\frac{1}{N}\sum[y \log(y') + (1-y) \log(1-y')] \quad (10)$$

The Dice Loss is defined as:

$$DiceLoss(y, y') = 1 - \frac{2\sum(y.y') + \epsilon}{\sum y + \sum y' + \epsilon} \quad (11)$$

Here;
y is ground-truth mask (0 or 1)
y' is pedicted probability (0-1)
α, β is weight factors
ϵ is small smooth term
N is total number of pixel

The BCE term plays an important role in achieving better results for pixel classification accuracy, while the Dice term seeks to maximize the overlapping between the predicted masks and the actual masks while maintaining their boundaries. With this combination, the proposed loss function is able to deliver robust and reliable results, especially in challenging conditions like low-contrast areas and thin blood vessels.

Table 3. Division of Samples

Training Samples	Validation Samples	Testing
1346	178	395

The division of the dataset is described in Table 3. The dataset was split into 70% for training, 20% for testing, and the remaining 10% for validation.

Results and Discussion:

This aims to propose vessel-enhancement model by utilizing U-net variants models. The models were train and tested over publically available CT scan images and binary mask dataset.

Table 4. Performance Evaluation of Models

Models	Accuracy	Dice	IoU
Vanilla	99%	93%	88%
Unet4	98%	94%	89%
AttentionUnet	98%	93%	89%
UnetPP	97%	92%	88%

Table 3 shows a comparison of the four U-Net-based architectures on standard segmentation evaluation metrics. The Vanilla U-Net had the highest accuracy of 99% which shows a very high level of agreement with the ground truth and very few false positives. Both U-Net4 and Attention U-Net had similar accuracy of 98%. The U-Net++ had slightly lower accuracy of 97% For all architectures, the Dice scores were 92-94% and IoU scores were 88-89%, which show a very high level of overlap with the ground truth segmentations.

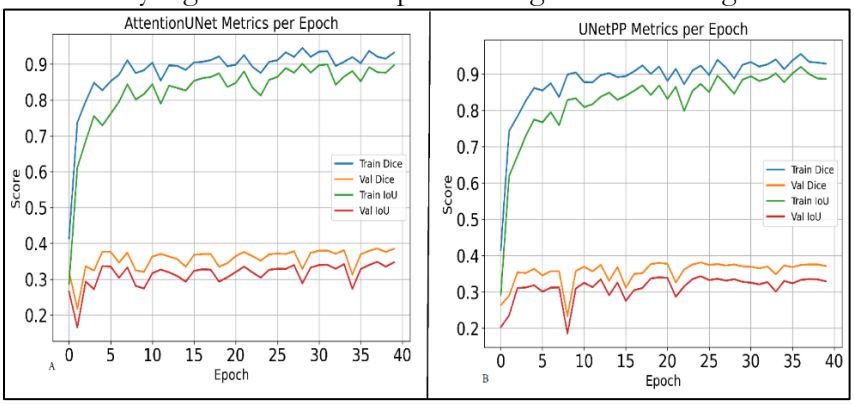

Figure 3. Loss curves for UNet models

Figure 3 illustrates the segmentation accuracy of Attention U-Net and UNet++ on dual-channel input data per epoch. Figure 3 (A) illustrates the train and validation loss for AttentionUNet model. The models train rapidly, with Dice and IoU scores above 0.9 in the initial 10 epochs, signifying their efficiency in capturing the minute details of the spatial information. This indicates that the model is learning useful features and that optimization is

performed well. On the other hand, the slight fluctuations in the validation loss are indicative of certain sensitivity to data fluctuations, which has been commonly observed in attention-based models applied to medical image segmentation in previous research works. Conversely, in Figure 3 (B), the losses curves for UNetPP model are represented. In the initial epochs, the IoU and Dice scores plummeted and continued to fluctuate with the increasing number of epochs while the train and validation loss remained below a value of 0.3. Despite this the model maintains relatively low loss values (<0.3), indicating the multi-scale feature aggregation is effective but requires more stable optimization. Both Attention U-Net and UNet++ make use of their ability to aggregate features at various scales and the attention mechanism for context refinement, which is quite advantageous for the models to capture information at different scales.

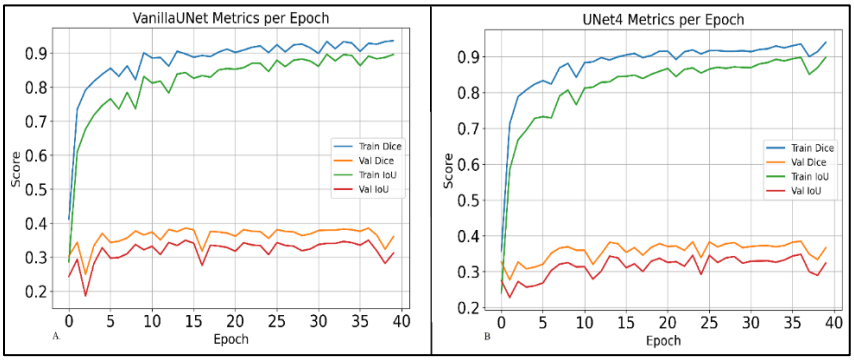

Figure 4. Loss graph

In Figure 4, the performance of both Vanilla U-Net and UNet4 was assessed based on training and validation results and calculated using both Dice and IoU scores over 40 epochs (duration of training). In the loss functions, as shown in Figure 4 (A), both the Dice and IoU loss functions for the Vanilla U-Net were close to the target (0.9), however; these scores fluctuated. In terms of fluctuation of the training and validation losses, these losses were lower than that of Dice or IoU around 0.3 (fluctuates). The Dice scores and IoU scores reached above the target levels for the training and testing of the same two models and were approximately greater than or equal to 0.9 and 0.85 respectively, indicating efficiency of both models at hierarchical and contextual feature extraction. The validation Dice and IoU scores for both models reached 0.35 and 0.32 respectively. Both models are based on encoders and decoders; therefore, UNet4 has more skip connections in the architecture, resulting in improved representation of spatial features. The UNet4 exhibits slightly better convergence properties than the Vanilla U-Net, primarily due to its deeper architecture and enhanced skip connections that contribute to preserving features on various scales. The findings also demonstrate that increasing the depth of the U-Net does not necessarily guarantee stable validation performance when compared to previous iterations of the architecture in literature.

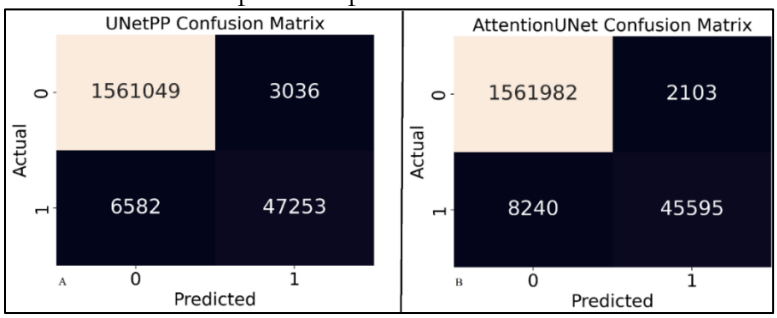

Figure 5. Confusion Matrix for UNetPP and AttentionUNet Models

Fig 5 shows the segmentation performance of models. In Fig 5 (A) the analysis of UNetPP model is shown where the model correctly segmented 1,608,302 out of 1,617,920

samples, with only 9,618 incorrect classifications, which is an accuracy of about 99.41%. This architectural feature is beneficial for improved boundary detection and multi-scale feature fusion, leading to the model's superior performance. Moving to Fig 5 (B) the confusion matrix for AttentionUNet model is presented where the model rightly predicted 1607577 samples with only 10343 misclassifications, which is an accuracy of about 99.36%. In comparison to U-Net performance from previous research, the slight gain in performance emphasizes the importance of incorporating the multi-scale feature fusion approach along with the use of attention, as in U-Net++. However, the difference between the two models is marginal, meaning that at this point any future improvements would be determined more by data quality than any modifications made to the model structure.

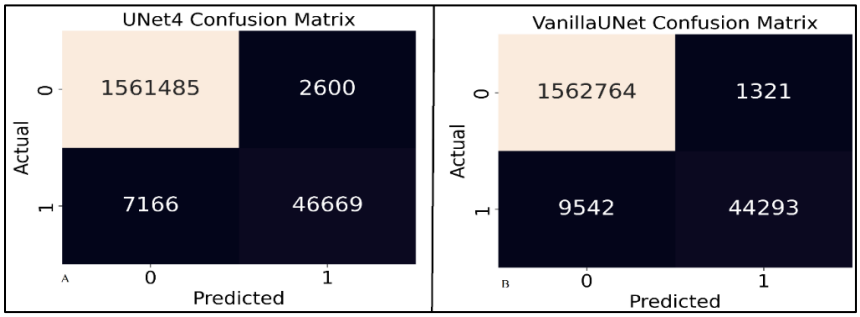

Figure 6. Confusion Matrix for U-Net Variants

Fig 6 illustrates the segmentation results of the various U-Net models based on confusion matrix analysis. Fig 6(A), UNet4 made 1,608,154 correct predictions out of 1,617,920 samples with 9,766 errors, resulting in an accuracy of about 99.40%. This is likely due to its deeper network architecture and superior feature representation capability. The increased depth enables the network to better grasp contextual information at various scales and to better define the boundaries of objects. In Fig 6(B), the Vanilla U-Net model made 1,607,057 correct predictions with 10,863 errors, yielding an accuracy of about 99.33%. The small improvement seen in UNet4 is likely due to its ability to extract more detailed features and better capture context at different scales. At the same time, the difference between the models is quite small, which indicates that all the U-Net variants are performing fairly consistently on this dataset. This also matches what has been reported in recent medical image segmentation studies, where adding more complexity to the model does not always lead to noticeable gains, especially when the dataset is already well prepared and preprocessed.

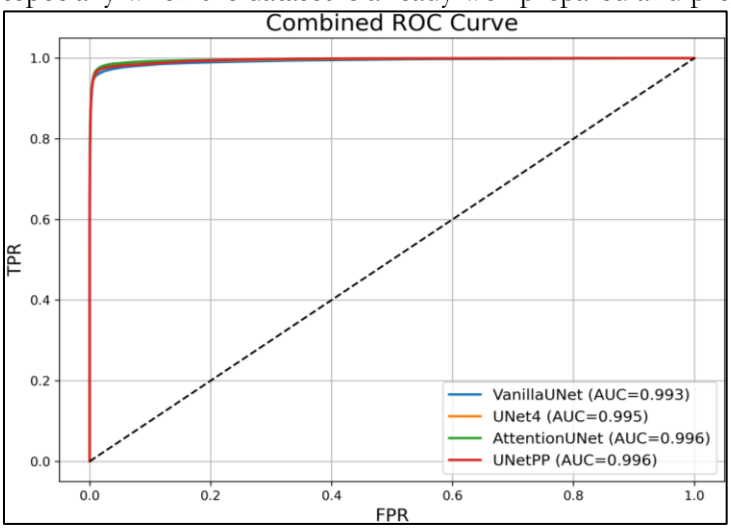

Figure 7. ROC analysis

The analysis of the ROC curve is depicted in fig 7 of the four U-Net-based models. Among them, U-Net++ and Attention U-Net presented the most successful outcome in the

AUC with the value of 0.996, which highlights the impressive ability of both to isolate foreground and background pixels. This level of AUC implies that the sensitivity and specificity of the model to various levels of decisions are almost optimal, and hence their very reliable classification performance. UNet4 was followed by Vanilla U-Net (0.993) and U-Net4 (0.995). Though these are a little lower, nevertheless, they are reflective of their outstanding segmentation capability. The minor variations in the AUC could be explained by the architectural disparity of the models. Considering the small variance between the AUCs, all four algorithms can be considered highly reliable when it comes to segmenting images. However, it seems that incorporating an attention mechanism and using multi-scale features does indeed provide some improvements in terms of classification stability. This finding is consistent with the current literature on the subject, which states that attention-based algorithms in the field of medical imaging provide slight improvements consistently.

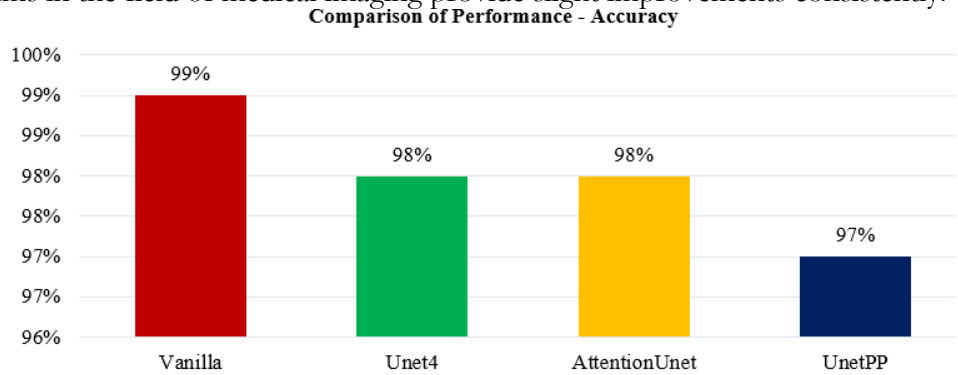

Figure 8. Comparison of Performance of Model accuracy wise.

Fig 8 represents the accuracy comparison of the four variants of the U-Net. Vanilla U-Net was the most accurate with 99 percent implying that it learnt features well and that it has an acceptable trade-off between FP and FN. U-Net4 and Attention U-Net were slightly less accurate with similar values of 98, which means that more detailed multi-scale pathways or attention modules enhance the learning of features without a significant effect on the agreement at a voxel level. The U-Net++ was slightly less accurate with 97% and this is probably because of its more complex architecture. Even though its thick skip connections are advantageous in terms of fine vessel details recording, they can also lead to a minor rise in FP, thus the minimal loss in accuracy.

Table 4. Comparison of proposed models with existing models

Work	Models	Dataset	Accuracy
[30]	Vanilla U-net, U-net32, U-net16	DCA1 and ICA	Vanilla U-net = 98.4%, U-net32 =98.5% U-net16 = 98.5%
[31]	CAS-Unet	Private dataset (150 images obtained from 50 patients)	96.91%
Proposed work	VanillaU-Net, AttentionU-net. U-netPP, U_net4	CT scan images binary masks	VanillaU-net = 99% U-net4 = 98% AttentionU-net = 98% U-netPP = 97%

Table 4 presents a comparison of the proposed models with existing models in terms of accuracy. Previous studies have already proven that models based on U-Net are very effective in the segmentation of coronary vessels. Based on publicly available datasets such as DCA1 and invasive coronary angiography images, the accuracy rate of the U-Net and its smaller variants U-Net32 and U-Net16 was above 98.4% [22]. These findings prove the effectiveness and reliability of encoder-decoder architecture, particularly in the segmentation of 2D vessels. Moreover, CAS-Unet has also proven to be very effective. When tested on a private dataset consisting of 150 images from 50 patients, it attained an accuracy rate of 96.91%

[23], proving that this architecture can be very effective in real-world clinical angiography images. In this research, different variants of U-Net were tested using images from CT scans and their corresponding binary masks. The original U-Net model attained the highest accuracy of 99%. U-Net4 and Attention U-Net attained an accuracy of 98%, while U-Net++ attained 97%. These findings prove that convolutional encoder-decoder architecture is still very effective in vascular segmentation tasks. However, they also suggest that more complex architectures may not necessarily provide a significant boost in performance if the training data is properly prepared and consistently annotated.

Discussion:

Experimental results shown in Table 3 and Figures 3-8, all variants of the U-Net architecture perform at a very high level. The Dice index values vary between 92% and 94%, IoU between 88% and 89%, and AUC values are all above 0.99. The largest differences between models are not in their ability to segment vessels from the background, but in their trade-off between precision and recall. The original U-Net had the highest overall accuracy of 99%, which is a strong indication of its agreement with the ground truth at the voxel level and its ability to suppress false positives. Attention U-Net had the best AUC value of 0.996, which indicates better foreground-background separation. However, the difference between the training and validation curves in Figs 3 and 4 indicates some degree of overfitting. This implies that the performance is affected not only by the architecture but also by other aspects, for example the quality of the data.

The performances of the proposed models when compared with state-of-the-art models given in Table 4 can be observed to be either equal or better compared to previous methods. As discussed in [8] and [9] the models achieved accuracies ranging from 96.91% to 98.5%. On the other hand, the proposed model has obtained an accuracy of 99% when applied on U-Net based binary mask segmentation of CT scans, without compromising its stability among various U-Nets. From the above comparisons, it can be infer that while architecture modifications may improve the performance of the model, the preparation of the dataset may equally play a key role.

The results indicate that the optimized baseline models are still very competitive. The more complex architectural changes are only beneficial for the detection of small vessels.

The results obtained in this study are aligned with the research objectives outlined at the start. The objective of enhancing pixel-level segmentation is justified by the high performance of the models developed in this study, which gave Dice scores of 92-94%, IoU of 88-89% and achieved accuracies of up to 99%. The other objective of enhancing segmentation in small vessels and low contrast images has been realized, especially with the models based on Attention U-Net and the hybrid loss technique, where boundaries were preserved in a higher capacity. Furthermore, comparing different models such as Vanilla U-Net, U-Net++, Attention U-Net and the modified U-Net was an opportunity to show the compromise between complexity and effectiveness of segmentation.

Practical Implications: These findings have several implications. Firstly, the high Dice scores (92-94%) and high AUC values above 0.99 indicate that U-Net-based models are capable of segmenting the coronary vessels reliably. Such performance makes them suitable for use in clinical settings for vascular analysis on CT scans.

Limitations: However, despite the success in obtaining excellent numerical values, two important drawbacks still exist. First, the difference between the training and validation curves suggests the potential for overfitting, which means that these models are not very good at generalizing to new data or multi-center scans performed with different imaging protocols.

Another limitation of this study is that no formal statistics, like significance test, or the calculation of p-values, was done to confirm any differences in the performances of the different models tested. The aim of this study was to conduct an analysis based on the widely

used measures of segmentation accuracy under the same experimental conditions. Future studies would include the use of statistical tests for the evaluation of the differences in larger samples and from multiple centers.

Conclusion:

This research aims to develop a robust deep learning model capable of precisely segmenting cardiac structures from CT scans, paying particular attention to enhancing accuracy at the pixel level. The proposed framework integrates hierarchical convolutional feature learning to preserve connectivity, improve boundary delineation, and enhance the detection of small and low-contrast vessels. The Vanilla U-Net had the highest overall accuracy of 99%, indicating strong generalization capability for feature extraction. The Dice and IoU metrics were in the range of 92-94% and 88-89%, respectively, with AUC > 0.99, indicating the robustness and reliability of the approach. Future studies may include several directions to further advance the Vessel-Enhanced Multi-Variant U-Net model. Future studies would integrate the topology-aware loss functions or graph constraints to further improve the continuity of small distal branches and stenotic areas, which currently suffer from under-segmentation problems in the standard U-Net. Moreover, the study would extend the model to semi-supervised or self-supervised learning paradigms, which would alleviate the need for dense human annotation and improve the generalizability of the model to multi-center datasets with diverse imaging protocols.

CRediT Authorship Contribution Statement:

Shafiya Qadeer Memon: Software, Formal analysis & Conceptualization.
Dr. Sania Bhatti: Review-Edit Draft, Supervision.
Dr. Muhammad Moazzam Jawaid: Supervision, Visualization & Conceptualization.
Mehran Memon: Data curation & Visualization.
Dr. Gulzar Usman: Supervision, Writing, review & Editing.

Declaration of Competing Interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment:

This work was carried out as part of a PhD study and supported by a Pakistan Science Foundation-funded research project (PSF/CRP/S-MUET/Cons-172).

References:

[1] "Cardiovascular diseases (CVDs)." Accessed: May 09, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] "High Blood Pressure and Cardiovascular Disease - PubMed." Accessed: May 09, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31865786/

[3] "Deep learning-based automatic segmentation of images in cardiac radiography: A promising challenge - PubMed." Accessed: May 09, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/35487181/

[4] R. Birjais, "Challenges and Future Directions for Segmentation of Medical Images Using Deep Learning Models," *Deep Learn. Appl. Med. Image Segmentation Overview, Approaches, Challenges*, pp. 243–264, Jan. 2024, doi: 10.1002/9781394245369.CH10.

[5] "Frontiers | Advancements in cardiac structures segmentation: a comprehensive systematic review of deep learning in CT imaging." Accessed: May 09, 2026. [Online]. Available: https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2024.1323461/full

[6] "Deep learning-based automatic segmentation of cardiac substructures for lung cancers - PubMed." Accessed: May 09, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38122850/

[7] "Advances in Medical Image Segmentation: A Comprehensive Review of Traditional,

Deep Learning and Hybrid Approaches." Accessed: May 09, 2026. [Online]. Available: https://www.mdpi.com/2306-5354/11/10/1034

[8] "Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images - ScienceDirect." Accessed: May 09, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260721002170

[9] "Coronary artery segmentation under class imbalance using a U-Net based architecture on computed tomography angiography images | Scientific Reports." Accessed: May 09, 2026. [Online]. Available: https://www.nature.com/articles/s41598-021-93889-z

[10] "Coronary Artery Segmentation Based on Transfer Learning and UNet Architecture on Computed Tomography Coronary Angiography Images | IEEE Journals & Magazine | IEEE Xplore." Accessed: May 09, 2026. [Online]. Available: https://ieeexplore.ieee.org/document/10175517

[11] "An analytics-driven review of U-Net for medical image segmentation | Request PDF." Accessed: May 09, 2026. [Online]. Available: https://www.researchgate.net/publication/395766139_An_analytics-driven_review_of_U-Net_for_medical_image_segmentation

[12] R. Azad *et al.*, "Medical Image Segmentation Review: The Success of U-Net," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 10076–10095, 2024, doi: 10.1109/TPAMI.2024.3435571.

[13] "A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation - Jiangtao - 2025 - IET Image Processing - Wiley Online Library." Accessed: May 09, 2026. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70019

[14] "U-Net-Based Medical Image Segmentation - PMC." Accessed: May 09, 2026. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9033381/

[15] "An automatic approach for heart segmentation in CT scans through image processing techniques and Concat-U-Net | Request PDF." Accessed: May 09, 2026. [Online]. Available: https://www.researchgate.net/publication/358567283_An_automatic_approach_for_heart_segmentation_in_CT_scans_through_image_processing_techniques_and_Concat-U-Net

[16] "Automated heart segmentation using U-Net in pediatric cardiac CT | Request PDF." Accessed: May 09, 2026. [Online]. Available: https://www.researchgate.net/publication/355081653_Automated_heart_segmentation_using_U-Net_in_pediatric_cardiac_CT

[17] V. Subha, G. Gomathi, and A. Manivanna Boopathi, "An exploration of ventricle regions segmentation and multiclass disease detection using cardiac MRI," *Int. J. Imaging Syst. Technol.*, vol. 34, no. 1, Jan. 2024, doi: 10.1002/IMA.22938.

[18] "Frontiers | Adaptive Feature Medical Segmentation Network: an adaptable deep learning paradigm for high-performance 3D brain lesion segmentation in medical imaging." Accessed: May 09, 2026. [Online]. Available: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2024.1363930/full

[19] "Frontiers | Deep Learning for Cardiac Image Segmentation: A Review." Accessed: May 09, 2026. [Online]. Available: https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2020.00025/full

[20] M. Ashok and A. Gupta, "A Systematic Review of the Techniques for the Automatic Segmentation of Organs-at-Risk in Thoracic Computed Tomography Images," *Arch.*

Comput. Methods Eng. 2020 284, vol. 28, no. 4, pp. 3245–3267, Sep. 2020, doi: 10.1007/S11831-020-09497-Z.

[21] "Deep learning-based whole-heart segmentation in 4D contrast-enhanced cardiac CT - PubMed." Accessed: May 09, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/35026571/

[22] "Heart CT & MRI Dataset." Accessed: May 09, 2026. [Online]. Available: https://www.kaggle.com/datasets/ziya07/heart-ct-and-mri-dataset

[23] "A Comprehensive Survey of Deep Learning Approaches in Image Processing." Accessed: May 09, 2026. [Online]. Available: https://www.mdpi.com/1424-8220/25/2/531

[24] "Feature extraction and feature selection in medical images - ScienceDirect." Accessed: May 09, 2026. [Online]. Available: https://www.sciencedirect.com/science/chapter/edited-volume/abs/pii/B9780443159992000086

[25] X. Guo, J. Hu, T. Lu, G. Li, and R. Xiao, "A novel vessel enhancement method based on Hessian matrix eigenvalues using multilayer perceptron," *Biomed. Mater. Eng.*, vol. 36, no. 2, pp. 83–97, Mar. 2025, doi: 10.1177/09592989241296431.

[26] O. O. Sule, S. Viriri, and A. Abayomi, "Effects of image enhancement techniques on cNns based algorithms for segmentation of blood vessels: A review," *2020 Int. Conf. Artif. Intell. Big Data, Comput. Data Commun. Syst. icABCD 2020 - Proc.*, Aug. 2020, doi: 10.1109/ICABCD49160.2020.9183896.

[27] "Lemon-Flavored Gummy Candies: Sourness, Flavor and Overall Acceptance Optimization Using Lattice-Simplex Mixture Design Implemented with Python Programming Language." Accessed: May 09, 2026. [Online]. Available: https://www.mdpi.com/2673-9976/40/1/41

[28] "A novel sub-differentiable hausdorff loss combined with BCE for MRI brain tumor segmentation using UNet variants | Scientific Reports." Accessed: May 09, 2026. [Online]. Available: https://www.nature.com/articles/s41598-025-33136-x

[29] "A lightweight segmentation model based on dilated multi-scale residual attention U-Net for brain tumor segmentation - ScienceDirect." Accessed: May 09, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095219762501927X

[30] "Lightweight U-Net for Blood Vessels Segmentation in X-Ray Coronary Angiography - PubMed." Accessed: May 09, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/40278022/

[31] "(PDF) CAS-Net: A Novel Coronary Artery Segmentation Neural Network." Accessed: May 09, 2026. [Online]. Available: https://www.researchgate.net/publication/355252457_CAS-Net_A_Novel_Coronary_Artery_Segmentation_Neural_Network