Chapter 7

Attention-Based UNet Deep learning model for Plaque Segmentation in Carotid Ultrasound for Stroke Risk Stratification: An Artificial Intelligence Paradigm

Summary

Stroke and cardiovascular diseases (CVD) are affecting the world population predominantly. Early detection of such events may prevent the burden of death and costly surgery. Conventional methods are neither automated nor clinically accurate. Artificial Intelligence-based methods of automatically detecting and predicting the severity of CVD and stroke in their early stages are of prime importance. This study proposes an attention-channel-based UNet deep learning (DL) model that identifies the carotid plaques in the internal carotid artery (ICA) and common carotid artery (CCA) images. Our experiments consist of 970 ICA images from the UK, 379 CCA images from diabetic Japanese patients, and 300 CCA images from post-menopausal women from Hong Kong. We combined both CCA images to form an integrated database of 679 images. A rotation transform was applied to 679 CCA images doubling the database for the experiments. Cross-validation K5 (80% training: 20% testing) protocol was applied for accuracy determination. The results of the Attention-UNet model are benchmarked against UNet, UNet++, and UNet3P models. Visual plaque segmentation shows improvement in Attention-UNet results compared to the other three models. The correlation coefficient (CC) value for Attention-UNet is 0.96 compared to 0.93, 0.96, and 0.92 for UNet, UNet++ and UNet3P models. Similarly, the AUC value for Attention-UNet is 0.97 compared to 0.964, 0.966, and 0.965 for other models. Conclusively, the Attention-UNet model is beneficial in segmenting very bright and fuzzy plaque images that are hard to diagnose using other methods. Further, we present a multi-ethnic, multi-center, racial bias-free study of stroke risk assessment.

7.1 Introduction

In past decades stroke and other cardiovascular diseases (CVD) have emerged as fatal diseases across the globe. For developing countries like India, stroke has emerged as a pandemic [6, 7, 159, 244] in recent years with 105 to 152 cases per 0.1 million population [5]. In developed countries like the USA, 795,000 people experience a stroke, and approximately 240,000 get a transient ischemic stroke yearly [269]. Various cardiovascular risk factors such as diabetes [168, 270], smoking, hypertension, hypercholesterolemia, and irregular lifestyle have been discussed in many studies leading to higher stroke and CVD events [1, 2, 55, 269].

A major cause of stroke and CVD is atherosclerosis disease of the arteries [8], particularly in intimal and media walls . Different stages of plaque constitute soft plaque, such as lipid, macrophages,

and hard plaque such as calcium [10], so-called plaque morphology, which is helpful in the diagnosis of atherosclerosis disease severity [152, 245]. Three primary techniques available for carotid plaque identification are magnetic resonance imaging (MRI) [271], computed tomography (CT) [187], and ultrasound (US) imaging [216]. Based on cost-effectiveness, non-radiative, portability, and device handling properties, ultrasound is more prevalent in carotid artery imaging [216], . Different grayscale levels generated by ultrasound echoes segregate parts of the carotid artery such as the lumen area, lumen intima layer, media-adventitia layer, and plaque such as fibrin, fibrosis, hard and soft plaque [141]. This also helps the sonographer or radiologist mark the plaque area by delineating the lumen intima (LI) and media adventitia (MA) borders of the carotid arteries [272].

Delineating the carotid plaque using the manual technique is tedious and error-prone. Thus researchers have proposed various automated methods of delineating the carotid plaques [35]. The computerized method for carotid plaque and intima-media thickness (IMT) delineation involves edge-based [73, 76] methods including first order absolute moment (FOAM) [273] and local maxima [274], anisotropic Gaussian derivative filtering [41, 275] and non-maximum suppression [276], snake-based [79, 82, 277] or dual-snake-based [79], scale-space [38, 40] (or AtheroEdge™ 1.0, 2.0, AtheroPoint LLC, CA, USA) [41], shape-based approach [80, 81], and level-set [211, 247, 278]. A state-of-the-art automated segmentation system, AtheroEdge™ 2.0, was developed by AtheroPoint LLC, Roseville,, CA, USA for cIMT measurement at bulb edge points [145]. Several methods have been published that benchmarks computerized method against manual methods [75, 115] along with the variability studies [273].

With the emergence of machine learning (ML), researchers used a single-layer feed-forward network (SLFN) for IMT boundary estimation [170]. While it uses knowledge of the database, it is also not fully automated as it involves some human interventions, particularly for grayscale feature extraction. This has been noticed well during the image-based tissue characterization in many field of view [83, 121, 139, 279, 280]. Under the class of AI, more recently, the deep learning (DL)-based methods have emerged over the past decade [93, 149] offering advantages of automated feature extraction. The DL-based methods used cascaded and sometimes parallel convolutional layers for feature extraction, which can extract fine details of the objects, which is sometimes not possible through ML-based feature extraction [36, 37, 76, 281, 282].

Technically, DL-based methods fall into two categories: supervised and unsupervised. Supervised learning models learn from the labels (pixel or categorical labels or gold standard) provided along with the database, while unsupervised learning-based models don't use such labelled information. Cardiovascular imaging has widely used supervised learning-based methods. Cuadrado-Godia *et al.* [43] presented a two-stage model for cIMT and plaque measurement. Zhou *et al.* presented UNet and UNet++

models for plaque segmentation on multi-institutional databases [71, 98]. Meshram *et al.* [100] applied the dilation factor on convolutional layers of the UNet model to study the effect on plaque segmentation. Recently, few studies have emerged hybrid deep learning (HDL) models [97, 118, 250, 283, 284]. There is no specific definition or criteria for a hybrid deep learning model. Still, any modification in the basic encoder-decoder arms or skip connections of the fundamental UNet architecture may fall under the HDL category. Keeping the spirit of HDL, Jain *et al.* were the first to propose few hybrid deep learning (HDL) models for ICA plaque segmentation [96]. The authors showed HDL models are a fusion of two solo deep learning models with the feature extraction capability of two SDL models. Few more faster, small, and low parameter HDL models were proposed by them recently [99]. Another study by Jain *et al.* used the same HDL model on CCA plaque segmentation [96]. Thus, HDL models are gaining more importance due to their better feature extraction capability.

Attention channel maps have been recently introduced in UNet-based deep learning [285, 286], however, they have not been tried in the carotid ultrasound framework. Further, no study has benchmarked UNet against other models for carotid ultrasound framework. Therefore, we hypothesize that attention channel maps, when added to the skip connection of UNet architectures, will improve the performance visually and quantitatively. Figure 7.1 shows the global system diagram of plaque segmentation using the attention-based UNet-based DL paradigm (named as AtheroEdge™ 3.0, AtheroPoint LLP, Roseville, CA, USA).

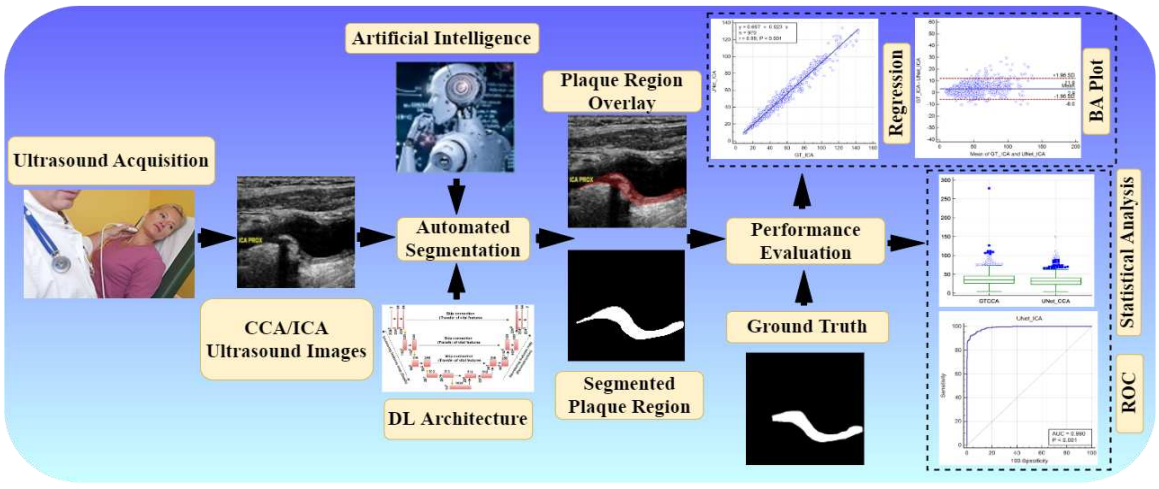

Figure 7.1 Global system (under the class of AtheroEdge™ 3.0) of carotid plaque segmentation.

The layout of this study is as follows: Section 7.2 presents database selection, preparation, and baseline characteristics. Section 7.3 presents the architecture of All UNet models used in this study. Methodology and experiments are presented in section 7.4. The results are presented in section 7.5. Section 7.6 shows the performance evaluation, Section 7.7 presents the discussions, and the paper concludes in section 7.8.

7.2 Database Selection, Preparation, and Baseline Characteristics

We considered multi-institution, multi-ethnic databases for our study. Hence, our experiments are free from data selection bias. We considered three databases in this research work, DB1: ICA database from the United Kingdom; DB2: CCA database from Japan; DB3: CCA database from Hong Kong. Detailed descriptions of databases are given here.

7.2.1 DB1: UK ICA Database

For UK ICA database description refer to the section **3.2.1.** The image size is changed to 224x224.

7.2.2 DB2: Japanese Diabetic CCA database

For Japanese CCA database description refer to the section **4.2.1.** For baseline characteristics refer to Table 4.1.

7.2.3 DB3: Hong Kong Post-menopausal Women CCA Database

For HongKong CCA database description refer to the section **5.2.1.2.**

7.2.4 Data Preparation and Augmentation Technique

Data preparation and augmentation play a vital role in DL-based systems. First, we removed the non-relevant information in ultrasound scans, such as patient ID (name), age, date, and model of the machine etc., by cropping the grayscale area only. This is very popular in image processing and is standardized [144, 211]. Secondly, we converted all images of DB1, DB2, and DB3 into equal-sized images of size 224×224, compatible with the first layer of the DL models. DB1 and DB2 are from different institutions and countries. Thus, we merged both databases to provide our experiments a multi-ethnic, multi-institutional database. Therefore, we combined the CCA images of DB2 and DB3 into a single folder, DB23 and achieved a new DB23 totaling 379+300 = 679 images. We also used a data augmentation technique on our combined CCA DB23 to enhance the number of images in the experiments. For this purpose, we used a rotation transform [-15^0 to $+15^0$] on all images of combined CCA DB23. Finally, we accomplished a database DB2A, where A stands for augmentation of 679×2 = 1358 images.

7.2.5 Binary Mask Preparation for Supervised Learning

This work falls under supervised learning-based semantic segmentation of the atherosclerotic plaque from carotid ultrasound images. Thus, the models require pixel-label information for the training phase. We have a team of experienced sonographers and cardiologists who successfully identified the LI and MA layers of the carotid artery and delineated the plaque. Further, another sonographer verified the

same for any error. The delineated LI and MA closed borders (contours) in grayscale images are converted into a binary mask using a MATLAB® based program. Further, these binary masks are also converted into equal-sized images of 224x224, just like the greyscale image.

Inter- and intra-observer analysis always seeks time of the radiologist and is sometimes becomes expensive. Further, while inter- and intra-observer studies were not an integral part of this pilot design, but our observations has proven that such analysis leads to have variations between 1% to 5% variations [31, 110, 161, 218, 251, 287-292] and such ranges are very normal and qualify FDA 510 (K) regulations. We intend to add this as part of future studies

7.3 UNet Architectures

7.3.1 Basic UNet Model

Basic Unet architecture and description please refer to the section 6.3.1 and Figure 6.3 from previous chapter.

7.3.2 UNet++ Architecture

For UNet++ architecture and description please refer to the section 6.3.2 and Figure 6.5 from previous chapter.

7.3.3 UNet3P: Multiscale architecture

For UNet3P architecture and description please refer to the section 6.3.3 and Figure 6.6 from previous chapter.

The number of trainable parameters (P) after each stage is calculated using the following formula $P = [(m \times n \times d) \times b \times k]$, where m is the height of the convolutional filter, n is the width of the convolutional filter, d is the number of filters in the previous layer, b is the bias, and k is the number of filters in the current layer. Table 2 shows the number of training parameters in different parts of the UNet, UNet++, UNet3P architectures.

7.3.4 Attention-based UNet model

The concept of attention mechanism was earlier proposed by Bahdanau *et al.* [293] and Luong *et al.* [294]. Further, the same concept was integrated with the UNet by Oktay *et al.* [285] for segmentation of the Pancreas. We have utilized the attention-UNet for plaque segmentation as the atherosclerotic plaque has a very fuzzy nature, which is challenging to segment using other UNet models in many cases. Figure 7.2 below shows an attention block used in the place of skip connection of the UNet model.

The attention gate has two inputs and one output. One of the inputs is “input features” ‘x’ from the same encoder level, and the second input is gating features ‘g’ from the lower decoder level. Both inputs

have their inherent property to carry the features. The input feature ‘x’ from the same encoder level comes from a shallow network level; therefore, it contains spatial feature information. The gating signal comes from the decoder level, which is at a deeper level compared to the input signal ‘x’. Thus gating signal provides better feature representation. Further, the input signal is downsampled by a stride of ‘2’ to make it compatible with the gating signal. The downsampled signal ‘x’ is then added with the gating signal ‘g’ and passed through the rectified linear unit (ReLU) activation function. The ReLU is a non-linear activation function which removes the negative values from the input, i.e. it gives the output if it has positive values.

Further, both ‘x’ and ‘g’ signals are passed through a 1 x 1 convolutional operation to acquire the weights from the combined weight signal. The combined weight signal passes through a Sigmoid activation function. Sigmoid is a ‘S-shaped’ non-linear curve defined by Sig(x) = 1/(1+exp(-x)). Due to its large slope value, it is able to transform the falling input values to between ‘0’ and ‘1’. Finally, the combined weight signal is upsampled to the same scale as the input signal ‘x’ and multiplied by it elementwise.

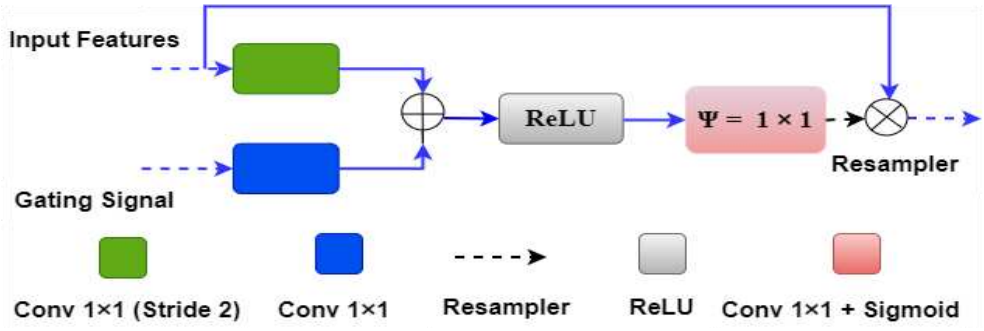

Figure 7.2 Attention block with input features, gating signals and weighted output.

We can understand the above attention mechanism with an example in Figure 7.3, which shows the first attention gate at the skip connection between the first encoder and decoder stages. The input features ‘x’ of size 224x224x64 come from the first encoder stage, and the gating signal ‘g’ of size 112x112x128 comes from one lower decoder level. Input signal ‘x’ is downsampled by green block (convolutional 1x1 with #filter = gating signal filter i.e. 128) to the size 112x112x128. Also, the gating signal passes through the blue box (#filters =128) with 1x1 convolution, so it acquires the same shape as ‘x’, i.e. 112x112x128. Both signals are added and applied to the ReLU activation function, where any nonlinearity is removed from the combined signal.

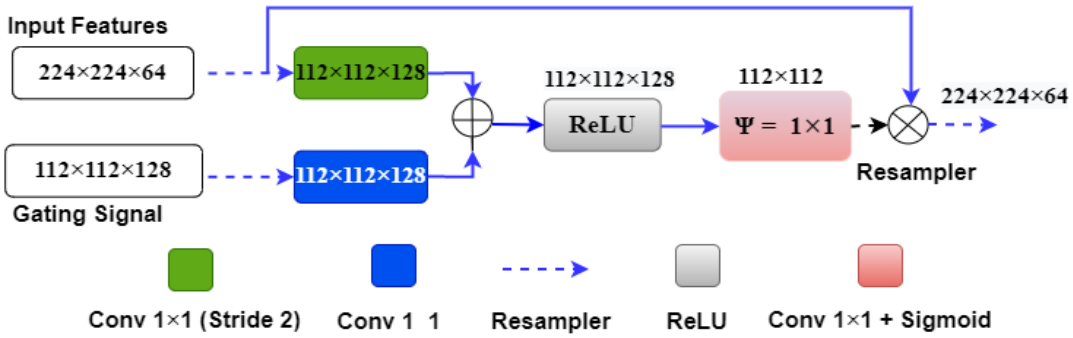

Figure 7.3 The attention gate mechanism between first encoder and decoder layer.

Further, the combined signal is passed through the psi (Ψ) block, which is a 1x1 convolutional block to acquire the weights from the combined signal. These weights are the attention gate weights generated by resampling the shallow and deep features from the encoder and decoder stages. These weights are upsampled to size 224x224 by upsampler and multiplied to the input signal 'x' of the same size. Figure 7.4 shows the complete attention-UNet model with four encoder and decoder stages. In this architecture, four attention blocks have been used in place of the skip connection between the encoder and decoder stages.

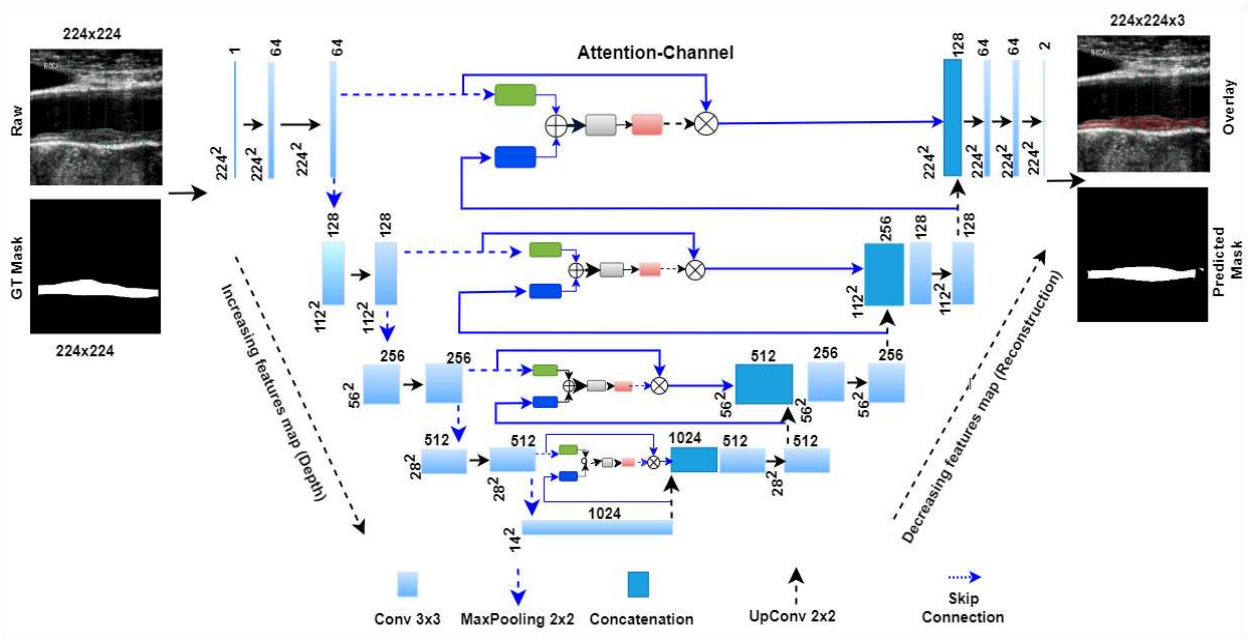

Figure 7.4 Four-stage Attention-UNet model.

Table 7.1 Number of trainable parameters in different parts of the UNet architectures.

Model	Encoder Arm	Bottleneck Layer	Decoder Arm	Intermediate Stages	FC	Classification Layer	Total Parameters
UNet	4685376	14157824	12188480	None	1154	6	31032840
UNet++	4685376	14157824	12593984	1275008	None	130	32712322
UNet3P	4685376	14157824	1196800	1447040	1154	6	21488200
Attention-UNet	4685376	14157824	12535680	523684	None	130	31902694

FC = fully connected layer.

Table 7.1 comprises training parameters of different parts of all UNet architectures, such as encoder arm, decoder arm, bottleneck layer, intermediate stages, fully connected layer (FC) and classification layers. We can also compare the total number of parameters of individual models.

7.4 Methodology and Experiments

All six models are trained using the raw images and binary masks of the ICA DB1 and CCA DB2A mentioned in database section 2. Various experimental setup steps and key points are mentioned below:

7.4.1 Hyperparameter selection and optimization

The choice of hyperparameters is very standardized and well-established. The main parameters were (1) # of layers, (2) CV protocol, (3) # of epochs, (4) Learning Rate, (5) Batch Size, (6) Filter size. The number of layers is shown in block diagram of all DL models. Also, Table 7.1 shows the total number of parameters in major parts of the DL architecture. The hyperparameters of the experiments include 100 epochs, which we optimize after many experiments with the database. At this number of epochs, no significant change in loss function is observed, and the loss value converges. We selected batch sizes of 8 for UNet, UNet3P, Squeeze-UNet, and attention-UNet, and 4 for UNet++ and Fractal-UNet models. Further, we used a standard learning rate of 10^{-4} , convolutional filter size of 3x3, and bias value of 1 with the same padding.

7.4.2 Sparse categorical Cross-Entropy loss function

We used a sparse-categorical cross-entropy loss function to minimize the training loss. The loss function is defined as the following Eq. (7.1):

$$L(w) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \quad (7.1)$$

Where, w refers to the model parameters, e.g. weights of the neural network; N is the total number of pixels in an image, y_i is the true (actual) label; $\hat{y}_i$ is the predicted label.

7.4.3 K5 Cross-validation

We used a commonly used K5 cross-validation [224] method for training and testing. The complete database is divided into 80% training and 20% test datasets in this cross-validation system. In the first step, the training set is used to train the system and offline weight generation, and the test set is used to validate the system and segmentation parameters generation. Again, we switched the 20% test set back into the main dataset and fetched another 20% images for testing and 80% for training. We repeated the previous step. We repeated this training and testing experiments five times, thus, five offline models are

generated corresponding to each test data set. Now each test dataset is used with corresponding offline weight to generate test results. Thus, all images are tested at least once.

Since the data sets were relatively smaller and our past experiences in AI protocols showed strong validation results, this pilot study did not conduct the validation component. This method can be applied potentially in subsequent experiments.

Experiment 1 is conducted with 970 ICA images. Total images are divided into 776 training and 194 test images. The test set of 194 images is replaced with a unique set of 194 images from training set. Finally, we tested our system with five unique sets of the 194 test images. Similarly, experiment 2 is performed with 2 × 679 = 1358 images (DB2A) of Japanese and Hong Kong patients. The complete dataset is divided into five unique test datasets of 271,271,272, 272, and 272 images and their corresponding training sets. Each test set is used with the offline model generated after training on corresponding dataset. Once the test result of each image is generated, we carried out the arithmetic mean of all images by the following Eq (7.2).

$$X = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (7.2)$$

Where xi is extracted feature of image 'i', and 'X' is arithmetic mean of 'N' images.

7.5 Results

Table 7.2 Segmentation performance (in %) of all UNet models using ICA database DB1.

Model	Acc	Sens	Spec	Prec	MCC	DSC	JI
UNet	98.58±0.61	87.43±5.45	99.48±0.40	93.11±4.61	89.39±3.63	90.02±3.53	82.03±5.66
UNet++	98.51±0.65	85.72±6.56	99.53±0.38	93.70±4.55	88.74±4.14	89.32±4.16	80.94±6.45
UNet3P	98.51±0.66	86.73±6.35	99.47±0.41	92.90±4.84	88.87±3.96	89.49±3.90	81.19±6.16
Fractal-UNet	98.41±0.74	85.68±6.71	99.45±0.43	92.61±5.35	88.13±4.58	88.78±4.50	80.10±6.92
Squeeze-UNet	98.53±0.63	86.29±6.01	99.51±0.39	93.40±4.95	88.91±4.09	89.52±4.01	81.25±6.30
Vision Trans	98.48±0.69	86.43±6.65	99.45±0.40	93.23±4.67	87.98±4.65	87.32±4.89	79.94±8.45
Attention-UNet	98.58±0.59	86.86±5.73	99.52±0.38	93.54±4.62	89.31±3.76	89.90±3.69	81.86±5.90

Table 7.3 Segmentation performance (in %) of all UNet models using CCA database DB2X.

Model	Acc	Sens	Spec	Prec	MCC	DSC	JI
UNet	98.97±0.53	81.58±8.58	99.72±0.27	92.20±7.39	85.97±5.77	86.07±6.17	76.01±8.59
UNet++	99.05±0.38	85.82±7.21	99.62±0.31	90.04±8.09	87.23±5.43	87.48±5.57	78.14±8.01
UNet3P	98.93±0.56	80.39±9.09	99.73±0.26	92.42±7.25	85.43±6.24	85.49±6.68	75.18±9.00
Fractal-UNet	98.91±0.54	80.08±8.67	99.72±0.27	85.17±6.56	91.96±7.61	85.06±6.29	74.68±8.92
Squeeze-UNet	98.90±0.63	79.67±10.45	99.73±0.27	84.94±7.56	92.33±7.09	84.94±7.60	74.48±9.98
Attention-UNet	99.01±0.42	81.75±7.94	99.74±0.26	92.71±7.20	86.38±5.70	86.50±5.94	76.65±8.36

Acc= Accuracy; Sens= Sensitivity; Spec= Specificity; Prec= Precision; MCC= Mathew's Correlation Coefficient; DSC= Dice Coefficient; JI= Jaccard Index

The segmentation performance of the six models with ICA DB1 and CCA DB2A databases are shown in Table 7.2 and 7.3, respectively. We have acquired accuracy, sensitivity, specificity, precision, Mathew's correlation coefficient (MCC), dice-similarity coefficient (DSC) and Jaccard Index (JI). These

truth masks. The attention-based UNet show Mean±SD values of all parameters as 98.58±0.59, 86.86±5.73, 99.52±0.38, 93.54±4.62, 89.31±3.76, 89.90±3.69, and 81.86±5.90 (all in %) respectively for ICA DB1 database. The same parameters acquire the Mean±SD value of 99.01±0.42, 81.75±7.94, 99.74±0.26, 92.71±7.20, 86.38±5.70, 86.50±5.94, and 76.65±8.36, (all in %) respectively for CCA DB2A database.

The visual results of UNet, UNet++, UNet3P, Fractal-UNet, Squeeze-UNet, and attention-UNet models are shown in Figure 7.5. The top row represents binary masks of all databases. The second-row shows overlays of the GT masks over the raw grayscale images in green colour. Third-row shows the overlay of the difference between estimated and GT masks on raw grayscale images. Similarly, fourth, fifth, sixth, seventh, and eighth rows also show the difference between estimated and GT mask on raw grayscale images for UNet++, UNet3P, Fractal-UNet, Squeeze-UNet, and attention-UNet models, respectively. The red colour indicates the estimated mask, and the green colour represents the difference of the two masks.

The attention mechanism states that the attention blocks modify the deep features by applying the attention weights. The same can be seen visually in some critical images shown in Figure 7.6, which are not successfully segmented using other models.

7.6 Performance Evaluation

The results of all models show almost equal segmentation indices. Hence we perform some more performance evaluation tests to validate our experiments. We performed a series of performance tests on ICA DB1 and DB2A databases, such as regression analysis, receiver operating characteristics (ROC) analysis, paired-t-test, and Bland-Altman's plot analysis.

7.6.1 Regression Analysis

Regression analysis is a powerful statistical tool to analyze the relation between two quantities. Regression analysis generates a correlation coefficient (CC) between the two variables, which occupies values between 0 and 1. A CC value close to '1' refers to a very high correlation, whereas a CC close to '0' refers to very little correlation between the two quantities. We used ground truth plaque area (GTPA) and predicted plaque area from the model for regression analysis. Using ICA DB1, we obtained CC values between GTPA and PA from UNet, UNet++, UNet3P, Fractal-UNet, Squeeze-UNet and Attention-UNet as 0.99, 0.98, 0.98, 0.96, 0.96 and 0.99. Similarly, using CCA DB2A, the CC values for same models are 0.93, 0.96, 0.92, 0.94, 0.89 and 0.96. All these numbers are summarized in Table 7.4. Also, the regression analysis curve is shown in Figure 7.7 and Figure 7.8. From the analysis of all CC numbers, it is clear that the attention-based UNet outperforms other models.

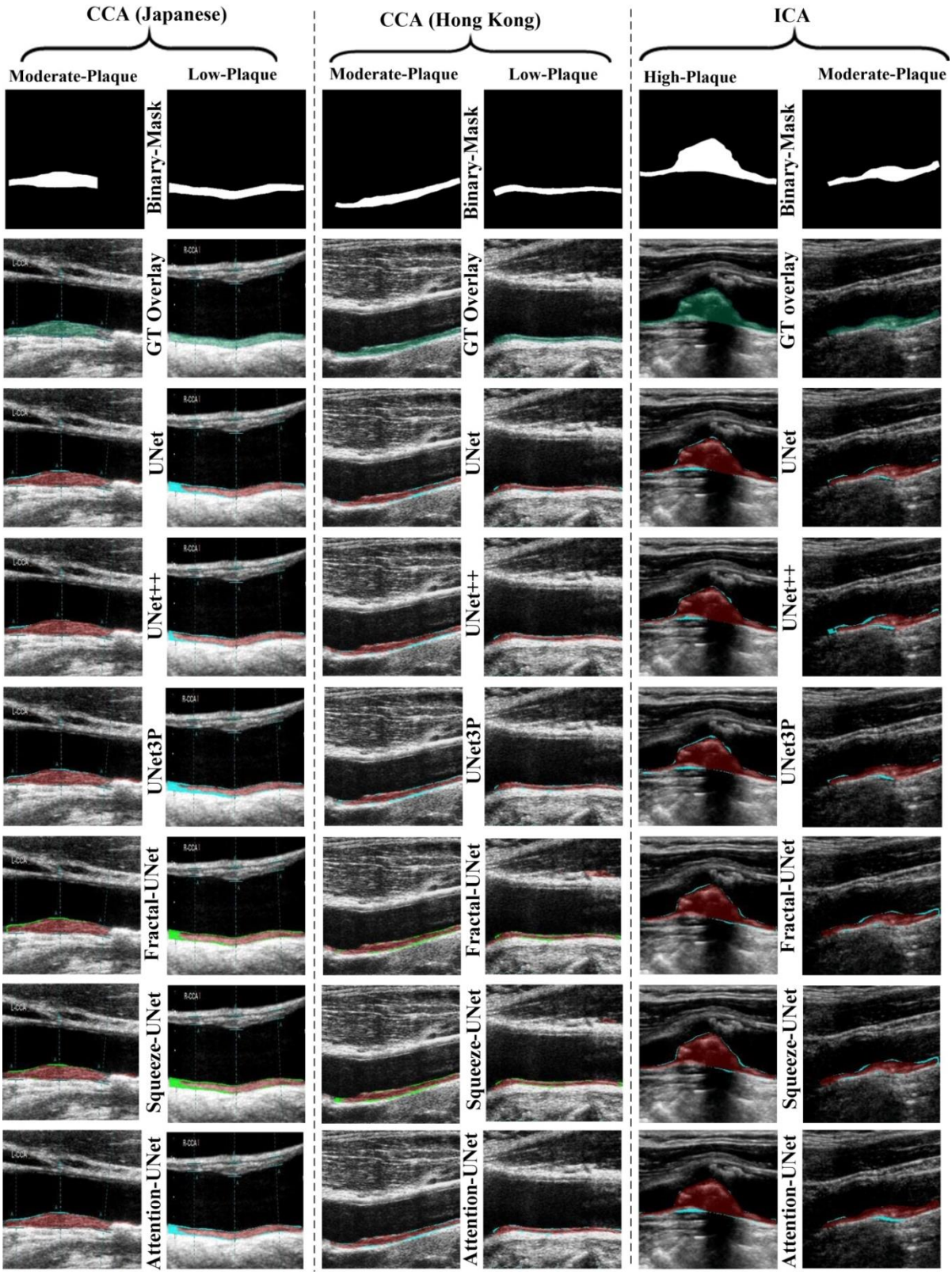

Figure 7.5 Visual results of the segmentation of Japanese, Hong Kong and UK (ICA) databases, performed by UNet, UNet++, UNet3P, Fractal-UNet, Squeeze-UNet, and Attention-UNet models.

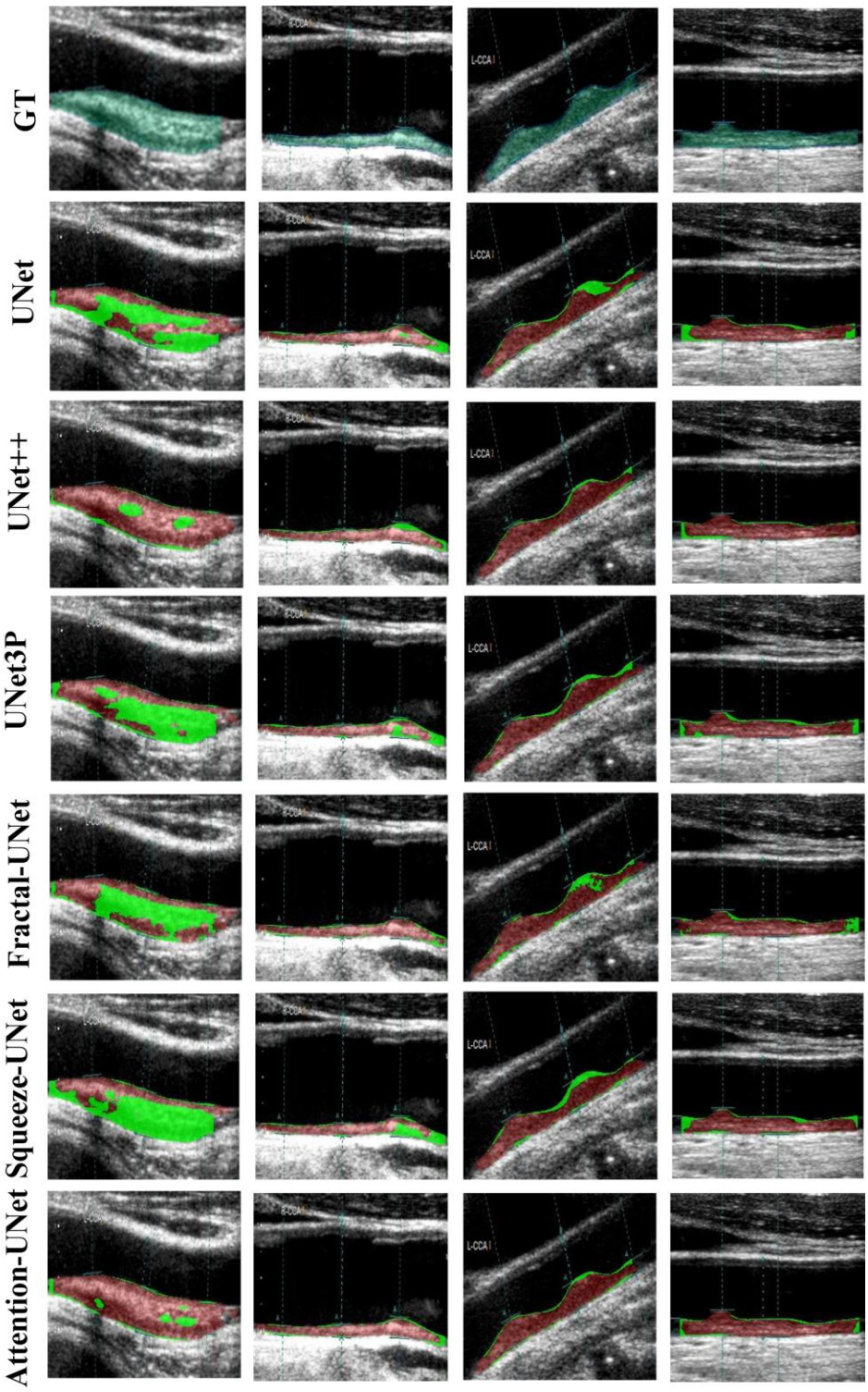

Figure 7.6 Attention channel effect on plaque segmentation on critical images of moderate and high plaque.

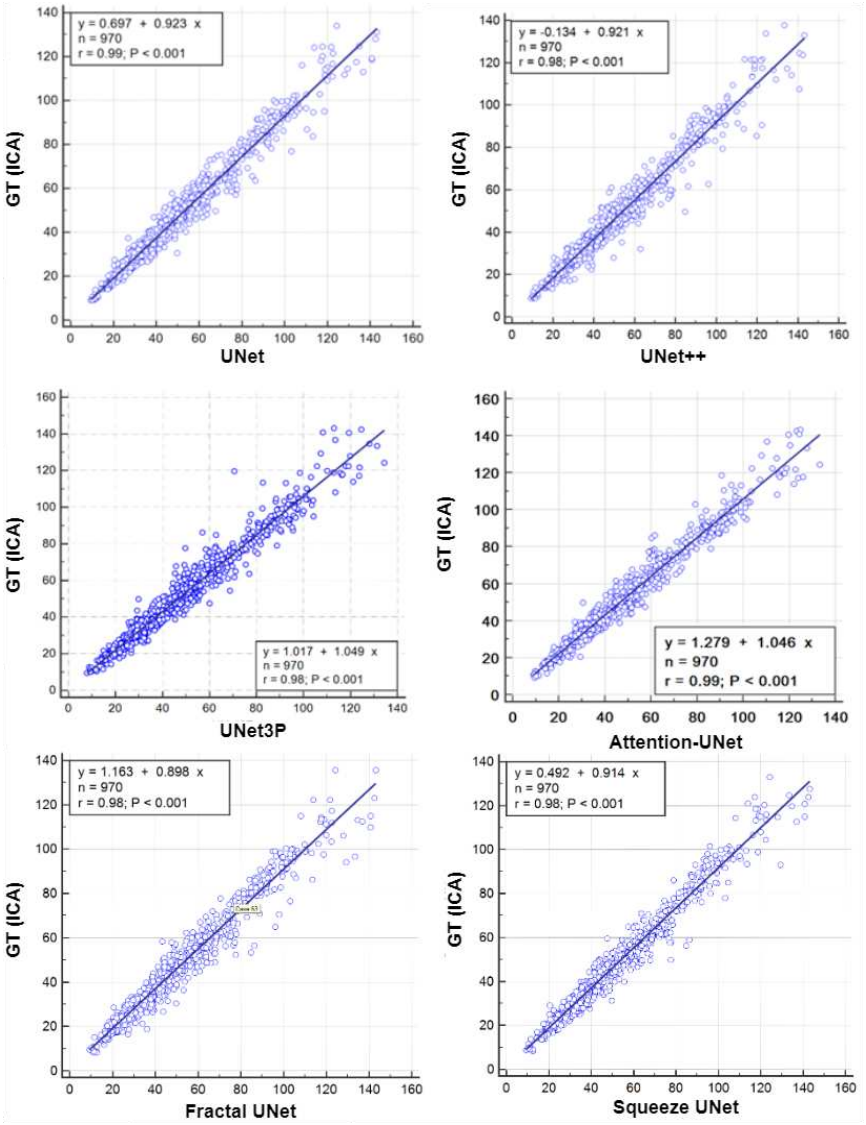

Figure 7.7 Regression plot for all UNet models for ICA DB1 databases.

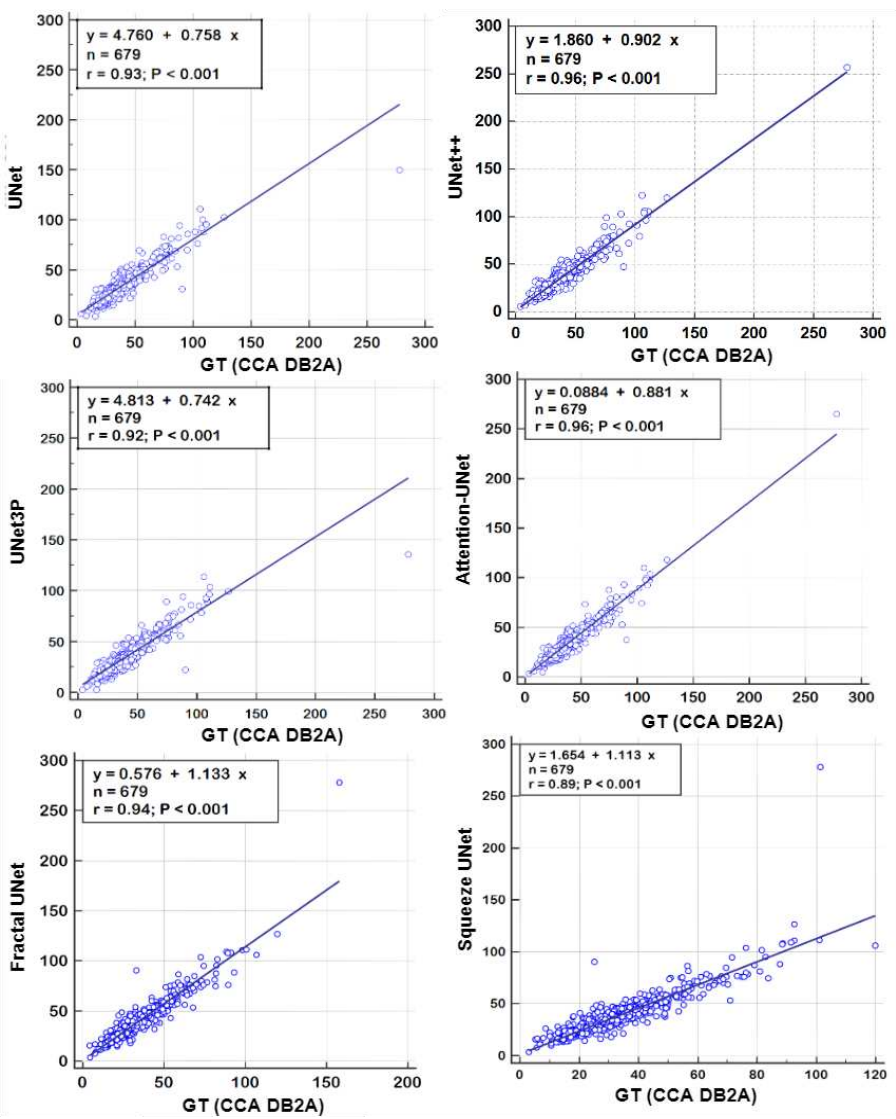

Figure 7.8 Regression Plot for all 6 types of UNet models for CCA DB2A database

Table 7.4 Performance parameters of all UNet models for ICA and CCA database experiments.

	ICA			CCA		
Model	**CC**	**AUC**	***p*-values**	CC	AUC	*p*-values
UNet	0.99	0.99	$P<0.001$	0.93	0.964	$P<0.001$
UNet++	0.98	0.988	$P<0.001$	0.96	0.966	$P<0.001$
UNet3P	0.98	0.988	$P<0.001$	0.92	0.965	$P<0.001$
Fractal-UNet	0.96	0.962	$P<0.001$	0.94	0.959	$P<0.001$
Squeeze-UNet	0.96	0.969	$P<0.001$	0.89	0.956	$P<0.001$
Attention-UNet	0.99	0.988	$P<0.001$	0.96	0.97	$P<0.001$

7.6.2 Receiver Operating Characteristics

Receiver operating characteristics (ROC) analysis is another performance evaluation tool to assess the classification performance. We cross-examined the literatures and found plaque area threshold value of 40 mm^2 used by researchers to classify the low and high risk plaque. Using this threshold value we generated GT labels '1' and '0' for GTPA>40 mm2 and GTPA<40 mm^2 respectively. Using predicted area as variable and GT labels as classification variable we plotted the ROC curve for UNet, UNet++, UNet3P, Fractal-UNet, Squeeze-UNet, and attention-based UNet models for ICA and CCA databases. These ROC curves are shown in Figure 7.9 and Figure 7.10 along with the area under the ROC curve (AUC) and p-values. Also, the AUC values are compared in Table 7.4. Again, it is clear from the AUC numbers and the ROC curves, that the attention-based UNet model is outperforming other UNet models.

7.6.3 Paired-t-Test Analysis

Paired-t-test is mostly used in biostatistics to analyze the mean difference between the two measurements. The requirement for this test is paired quantity of measurement for the same subject. In our case we have GTPA and the predicted PA for the same arteries (CCA and ICA both). Thus this method analyses the mean difference between the PA pair is zero or not. The distribution of GTPA and predicted PA is shown using the box and whiskers plot in Figure 7.11 and Figure 7.12 for ICA and CCA databases, respectively. Paired t-test results such as Mean±SD, standard error of mean, mean difference, student t-value and p-values are shown in Tables 7.5 and 7.6 for ICA and CCA databases, respectively.

Table 7.5 Paired-t-Test metrics for GTPA and estimated area by all UNet models for ICA database.

Models	Mean±SD	Std Error of mean	Mean Difference	SD of differences	Std error of mean difference	95% CI	Test statistic -t	p-value
UNet	44.5496±24.2130	0.7774	-2.9449	4.5716	0.1468	-3.2330 to -2.6568	-20.063	<0.0001
UNet++	43.5908±24.2386	0.7783	-3.9037	5.0739	0.1629	-4.2234 to -3.5840	-23.962	<0.0001
UNet3P	44.2960±24.1313	0.7748	-3.1985	5.3020	0.1702	-3.5326 to -2.8644	-18.789	<0.0001
Fractal-UNet	43.8306±23.6977	0.7609	-3.6639	5.4349	0.1745	-4.0063 to -3.3314	-20.996	<0.0001
Squeeze-UNet	43.8814±23.9853	0.7701	-36131	4.7915	0.1538	-3.9150 to -3.3112	-23.485	<0.0001
Attention-UNet	44.1855±24.3427	0.7816	-3.3090	4.5570	0.1463	-3.5961 to -3.0219	-22.615	<0.0001

GTPA: Mean = 47.4945 mm^2 ; SD = 25.8415 mm^2.

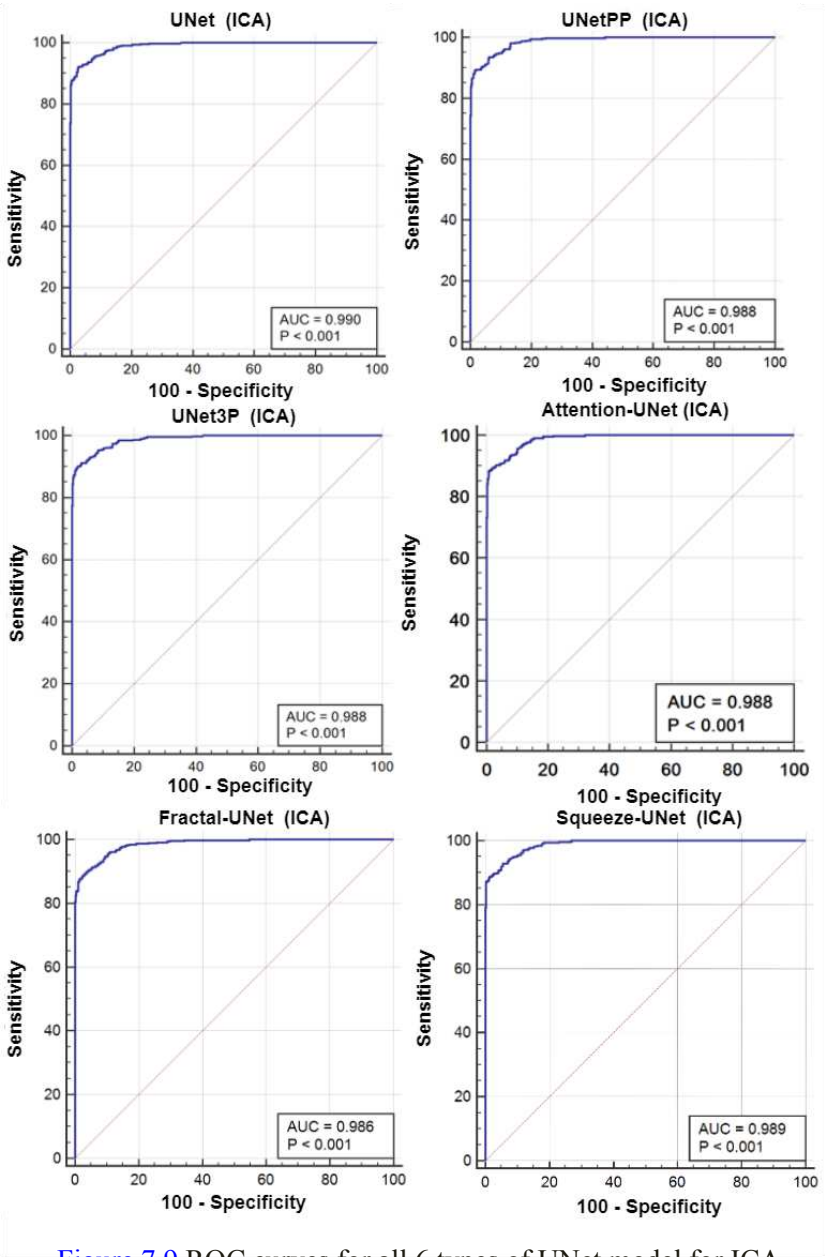

Figure 7.9 ROC curves for all 6 types of UNet model for ICA DB1 database.

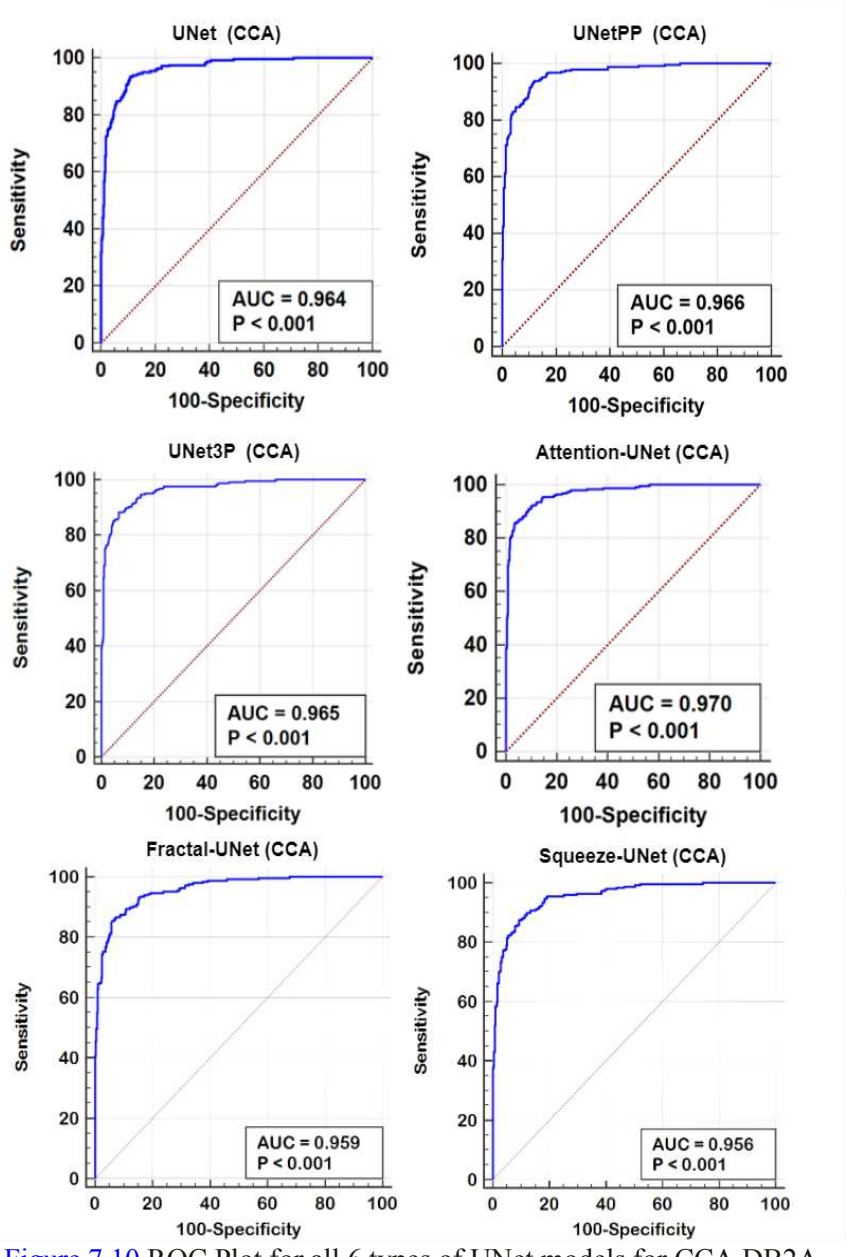

Figure 7.10 ROC Plot for all 6 types of UNet models for CCA DB2A database.

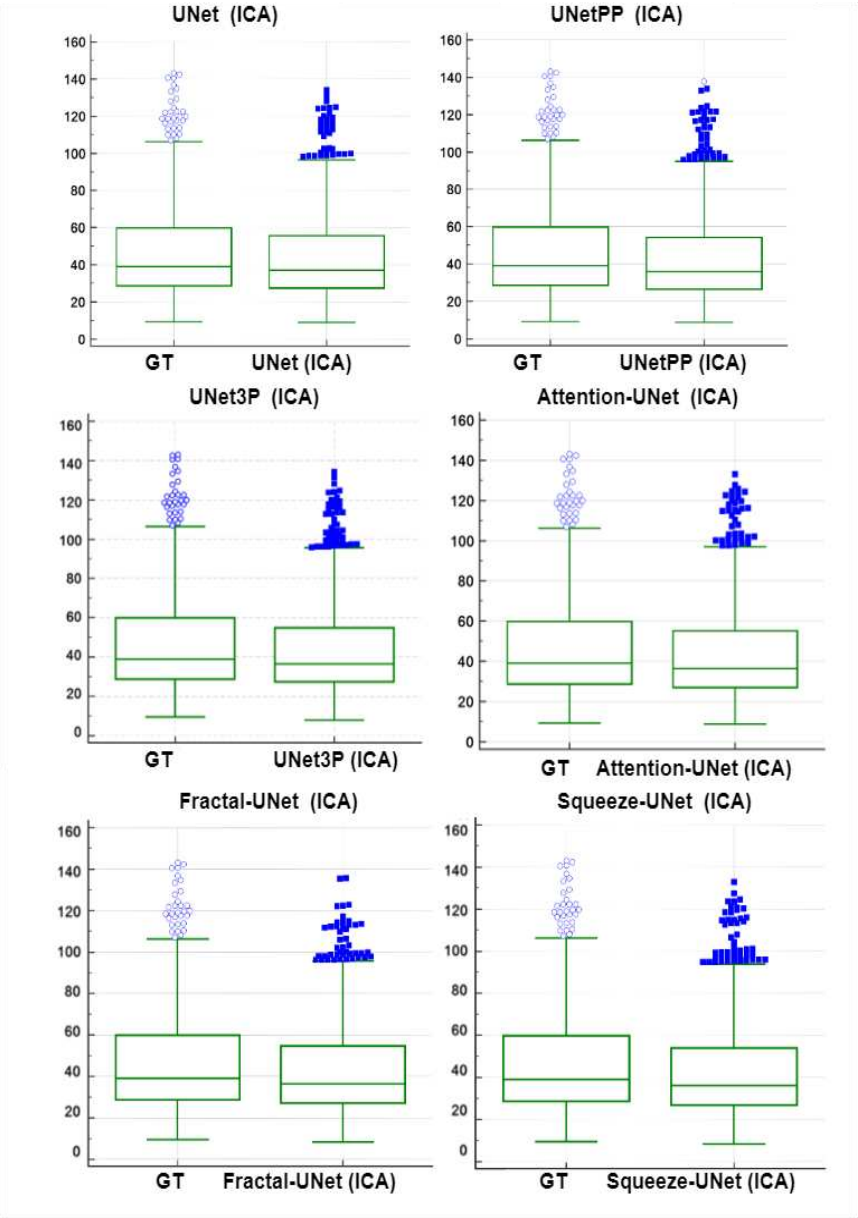

Figure 7.11 Paired samples t-Test Plot for all 6 types of UNet models for ICA DB1 databases.

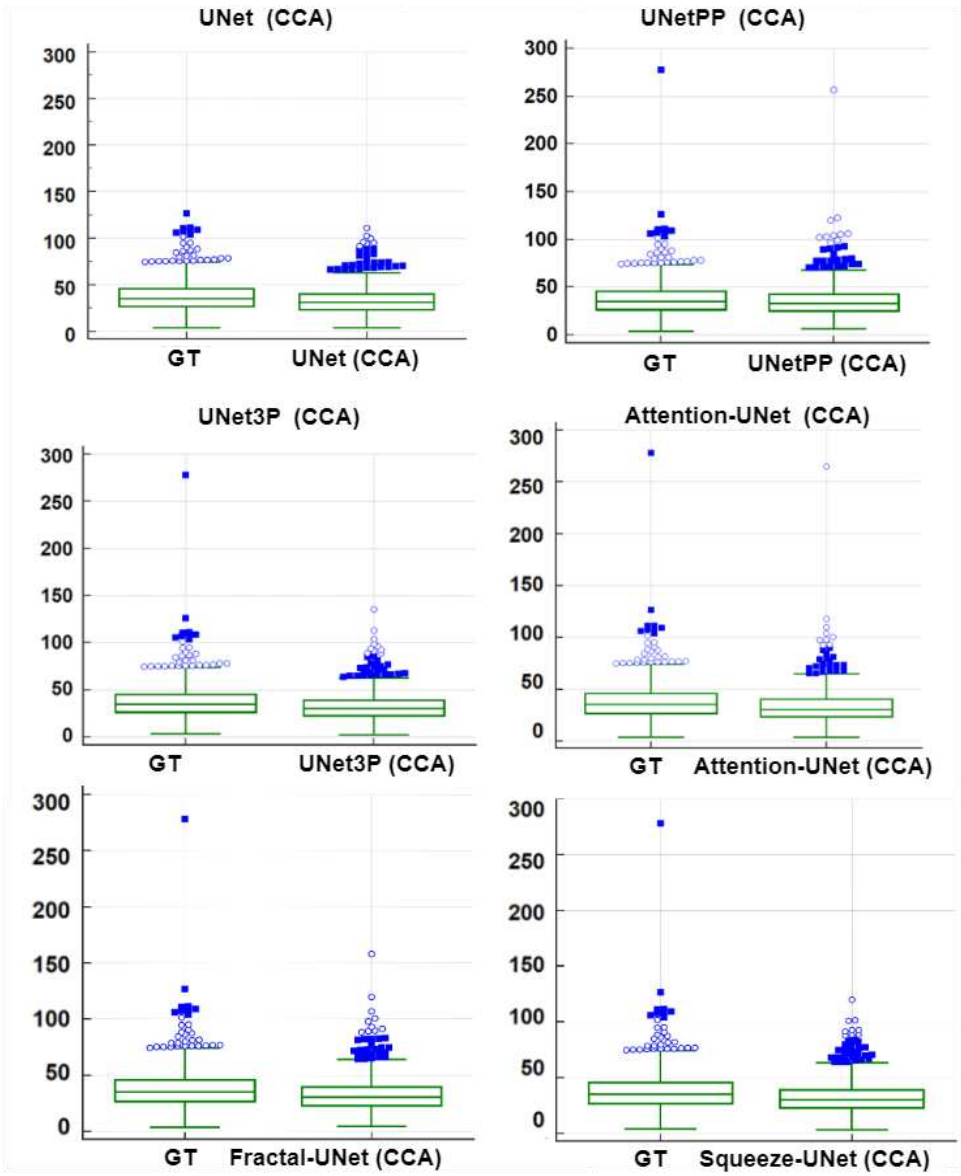

Figure 7.12 Paired samples t-Test Plot for all 6 types of UNet models for CCA DB2A database.

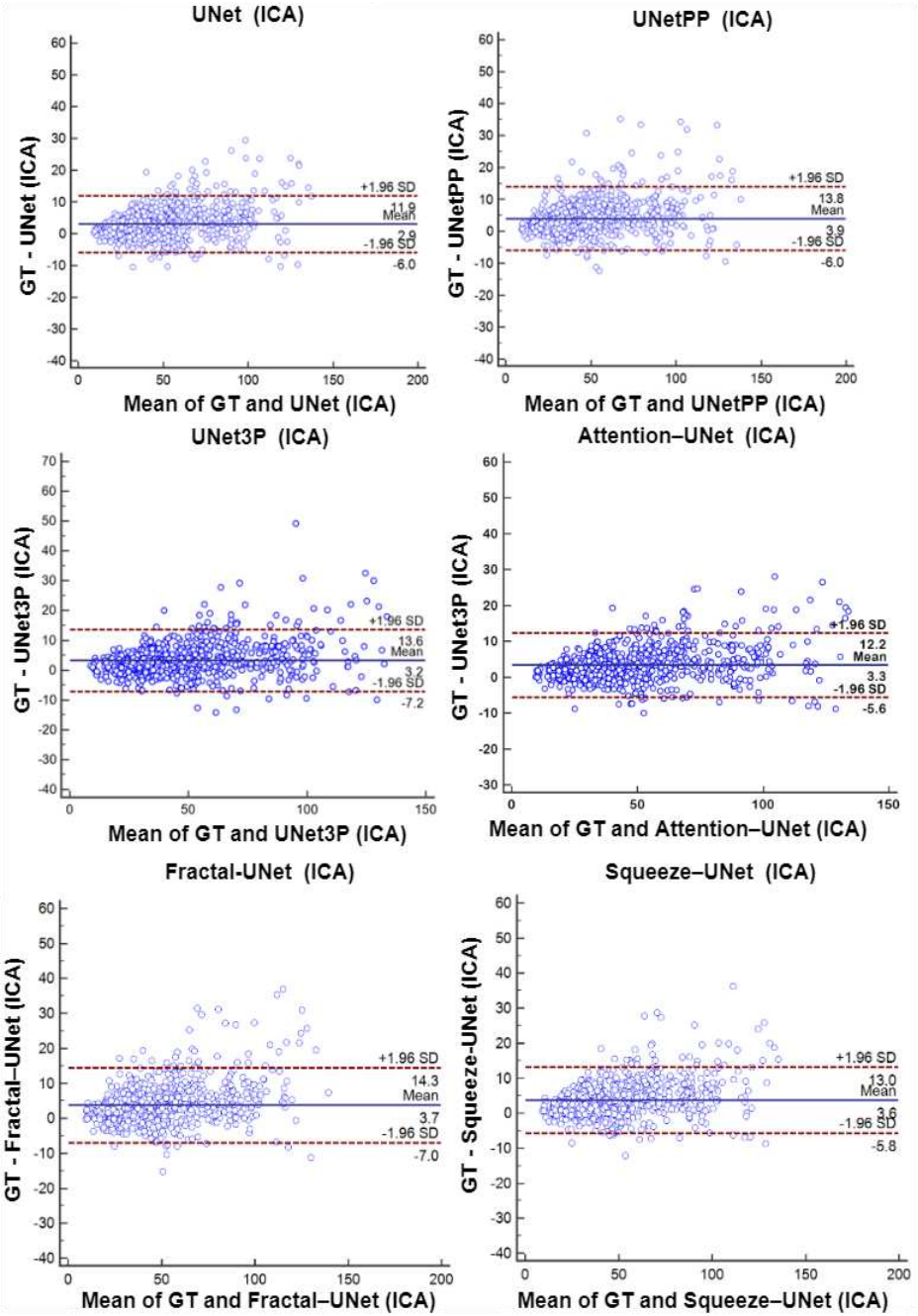

Figure 7. 13 Bland-Altman Plot of all 6 types of models for ICA DB1

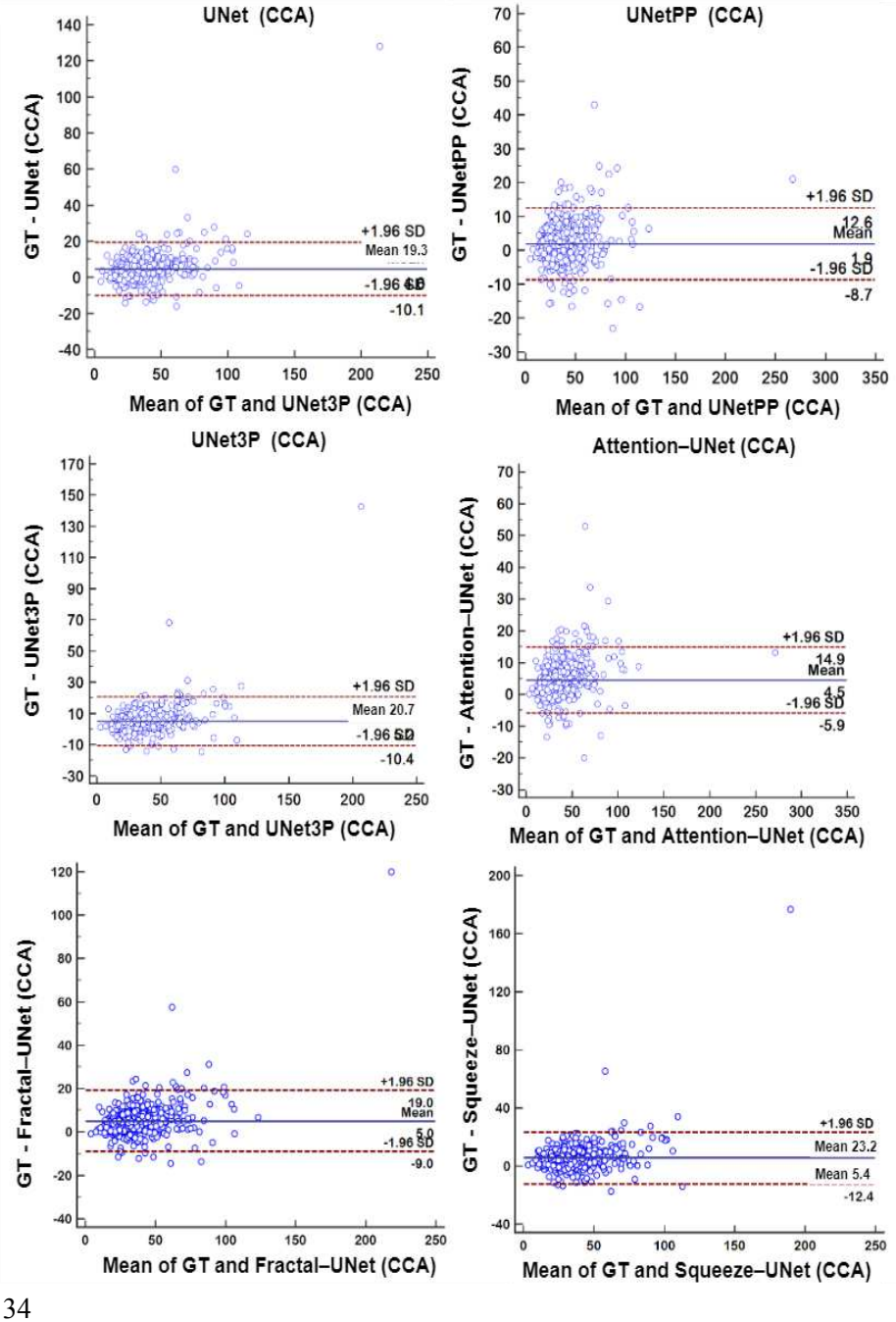

Figure 7. 14 Bland-Altman Plot of all 6 types of models for CCA DB2

Table 7.6 Paired-t-Test metrics for GTPA and estimated area by all UNet models for CCA database.

	Mean±SD	Std Error of mean	Mean Difference	SD of differences	Std error of mean difference	95% CI	Test statistic - t	p-value <
UNet	34.0097±15.8775	0.6093	-4.5858	7.5137	0.2883	-5.1520 to -4.0197	-15.904	0.0001
UNet++	36.6595±18.2781	0.7014	-1.9361	5.4177	0.2079	-2.3443 to -1.5278	-9.312	0.0001
UNet3P	33.4363±15.6883	0.6031	-5.1592	7.9247	0.3041	-5.7564 to -4.5621	-16.964	0.0001
Fractal-UNet	33.5501±16.1022	0.6179	-5.0454	7.1421	0.2741	-5.5836 to -4.5073	-18.408	0.0001
Squeeze-UNet	33.1997±15.5589	0.5971	-5.3958	9.0953	0.349	-6.0811 to -4.7105	-15.459	0.0001
Attention-UNet	34.0731±17.8005	0.6831	-4.5224	5.3039	0.2035	-4.9221 to -4.1228	-22.218	0.0001

GTPA: Mean = 38.5955 mm^2 ; SD = 19.4775 mm^2.

7.6.4 Bland-Altman's plot

Bio statistical analysis frequently uses Bland-Altman's plot or simply a difference plot when two different methods or instruments measure a parameter. The BA plot is used to show the bias between the mean differences between the two methods. Also, it offers an agreement interval of 95% (confidence interval) in which the difference between the second and first method fall. In our experiments plaque area is the quantity which is measured by the expert sonographer by manual method i.e. GTPA and the predicted area measured by all UNet models. Thus the difference between GTPA and the predicted PA is plotted on Y-axis and the mean of the both quantity is plotted on X-axis. BA-plot between GTPA and predicted PA for all UNet models using ICA and CCA databases are shown in Figure 7.13 and Figure 7.14, respectively.

7.7 Discussion

Current research work is a novel application of attention-based UNet models for carotid plaque segmentation. We hypothesized that the many images are very fuzzy and bright to identify the plaque constituents. Thus attention-based UNet model enhances the deep features and provide better segmentation [285, 286]. Also, the model shows excellent segmentation performance for low-to-moderate CCA and moderate-to-high ICA plaque images. Further, we compared the attention-based model against UNet, UNet++, and UNet3P models for a multicenter, multi-ethnic database. We put an effort to present a bias free stroke risk assessment system.

7.7.1 Bias in Medical Imaging Models

Deep learning-based clinical models have been gaining much attraction in recent times. The models use clinical data available from hospitals and medical research centers. Further, these clinical data are processed using established algorithms or new algorithms. These AI-based clinical decision support systems have shown promising results in diagnosis. However, the biases in these systems are not reported in many studies [267, 268]. These models may suffer in data selection bias (from single source), observer bias, data labelling bias, data source bias (data device selection bias), validation bias, racial bias (multi-

ethnic data selection), measurement bias, bias due to variabilities in the data sets [295], and many other types of biases which may affect the clinical results in some way. These biases must be discussed in detail and considered while designing the clinical support systems. There have been recent methods for computing the bias which can be extended for UNet-based systems [265, 266, 296].

7.7.2 Supervised and unsupervised learning based DL models

The current algorithm uses binary masks of ICA and CCA images for model training. For successful training of the model error free mask preparation is a must, failing which the model may result in false recognition. Binary mask preparation is tedious and time-consuming, and expert sonographers are required to accomplish this task. The current study involves the number of images in hundreds due to the unavailability of the datasets. However, considering the case where multiple thousands of images are available for analysis, generating such masks is nearly impossible. Therefore supervised learning loses its significance in big data analysis. Thus unsupervised models may replace such scenarios where no binary mask or labelled information is required for training. However, such unsupervised DL models have not gained much attraction in medical image processing till now, and much scope is available in this area. We may see many Unsup-DL models in near future.

7.7.3 Benchmarking

We have presented an attention-channel-based UNet model for atherosclerotic plaque segmentation. Many efforts have been done in ICA and CCA plaque area segmentation; still, these methods are not perfect in one way or another. Previous methods suffer in some biases discussed in the above section, resulting in poor image segmentation from other databases. Zhou *et al.* presented the UNet++ model for ICA and CCA plaque segmentation from multi-ethnic databases [98]. However, they trained their model with only 33, 33, and 34 images and tested on 44 images. Using such a low number of images for the test doesn't infer proper justification to the clinical setting. Further, they didn't benchmark their system with any other established system. In another study, Jain *et al.* presented hybrid deep learning models for ICA plaque segmentation [97]. They proposed SegNet-UNet, SegNet-UNet+ HDL models for plaque area segmentation. Although their model used only one segment of the artery, i.e. ICA, they enhanced the image database by applying the rotation transform augmentation technique. Thus, their models also suffer some biases such as data selection bias, source bias and racial bias. However, their models are a milestone in HDL model studies.

In another study, same team of researchers, Jain *et al.*, used the above HDL models for plaque segmentation from multi-ethnic CCA databases [110]. They used two kinds of databases for the experiments, one from Japan and another from Hong Kong, and performed some unseen experiments. Thus, they tried to avoid data selection and racial bias using these HDL models for unseen experiments.

they failed to validate their experiments against any established system; therefore, their models suffer from validation bias. Further, the same team of researchers, Jain *et al.* tried to avoid the validation bias by comparing their SDL and HDL models against a commercially available state-of-the-art plaque segmentation system, AtheroEdge 2.0 developed by AtheroPoint LLC, CA, USA [96]. Their HDL model shows a plaque area error of 8 mm^2 compared to 9.9 mm^2 for SDL and 9.6 mm^2 for AtheroEdge 2.0 models for 90% of the image database. The proposed method tried to overcome previous bias by intensive exercises. By using ICA and CCA databases, we tried to avoid data selection bias. We also care for racial biases by incorporating multi-ethnic, multi-center ICA and CCA databases. DB1, DB2, and DB3 are from the UK, Japan and Hon Kong, respectively. Further, we validated our Attention-UNet model against previous UNet, UNet++, UNet3P, and Vision Transformer (see results in Fig 7.15) models to avoid model selection and validation bias. Table 7.7 summarizes the comparisons of present study with some benchmark studies.

Table 7.7 Benchmarking current research against previous studies.

Authors	Artery Segment	DL Model	#Patients/#Images	Results	Bias Identified
Zhou *et al.* [98]	ICA, CCA	UNet++	N1=144/510 N2 =497/638	TPA error: 5.55±4.34 mm2	Data selection, model selection, validation bias
Jain *et al.* [97]	ICA	UNet, UNet+, SegNet, SegNet-UNet, SegNet-UNet+,	N = 97/970	PA error 3.49 mm2 for SDL 4.21 mm2 for HDL	Data selection bias
Jain *et al.* [110]	CCA	UNet	N1 = 379 N2 = 300	FoM of 70.96 and 91.14 (unseen) against 97.57, 88.89, and 99.14 (seen)	Validation bias
Jain *et al.* [96]	CCA	UNet, SegNet-UNet, AtheroEdge 2.0	N1 =379	PA error HDL = 8 mm2 SDL = 9.9 mm2 AtheroEdge 2.0 = 9.6 mm2	Data selection bias, racial bias
Proposed method	ICA, CCA	UNet, UNet++, UNet3P, Attention-UNet	N1= 970; N2 =379, N3 = 300	CC: 0.99 and 0.96 for ICA and CCA experiments	Free from data selection, racial, validation biases

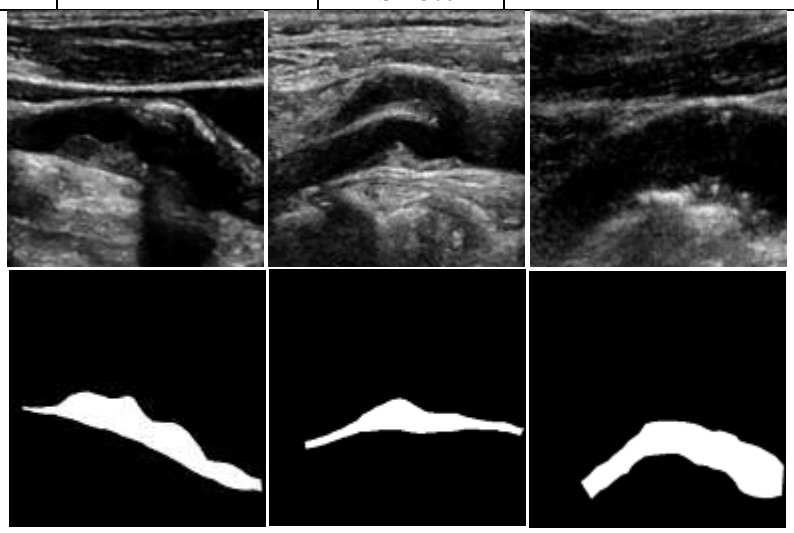

Figure 7.15 Vision Transformer based segmentation of ICA ultrasound images.

7.7.4 Strength, Weakness, and Extension

The present work shows a bias-free study of the plaque segmentation from ICA and CCA images from multi-ethnic, multicenter databases. We presented a powerful attention mechanism to modify the shallow and deep features of the carotid plaque images, which can capture those plaque areas which other models do not detect. We compared attention-based UNet segmentation results with other models used in previous studies such as UNet, UNet++, and UNet3P and the results are comparable or superior to such models. Also, the visual results show promising improvement in many images. Further, we validated our experiments using a series of performance evaluation tests such as regression, ROC, Bland-Altman's, and paired-t-tests. The results of such performance test are also comparable or superior to other models.

Further, we tried to fill the gap in data selection and racial biases from previous studies using multicenter, multi-ethnic and augmented databases. We feel that these biases are not sufficient to overcome, and there is still scope for improvement in such biases, which can be overcome in future studies as attempted here [194]. Further, the attention mechanism can be employed in other variants of the UNet to see its effect on other HDLs, such integration of advanced image processing [148] methods with UNet.

Since training models are large in size, one can adapt weight pruning techniques using algorithms like genetic algorithms and whale optimization [233, 264]. Also, the UNet-based segmentation method can be used for plaque tissue characterization due to its vital feature extraction and modification capability [118, 174, 253]. Carotid segmented lesions and plaque needs to be correlated to coronary SYNTAX score as part of the clinical validation [246]. Segmented plaque can be tried on different clinical groups such as Rheumatology patients for understanding cardiovascular risk [120]. Finally, since coronary plaque has been observed in COVID-19 patients [199], one can extend the UNet-based solution for plaque segmentation and measurement in carotid scans on COVID-19 patients.

7.8. Conclusion

This work presents a novel concept of attention mechanism incorporated with UNet as an attention-based UNet. The attention-based UNet model has successfully demonstrated plaque segmentation in complex images with fuzzy and bright plaque. The results of attention-UNet models are benchmarked against UNet, UNet++, and UNet3P models. The CC value of attention-based UNet model for CCA database is 0.96 compared to 0.93, 0.96, and 0.92 for UNet, UNet++, and UNet3P. The AUC value for attention-based UNet was 0.97 compared to 0.964, 0.966, and 0.965 for other models. The attention gate weight modifies the shallow and deep features to identify the complex plaque images; therefore, the attention mechanism is vital in plaque feature extraction and tissue characterization. The system can be adopted in clinical settings for cardiovascular disease risk stratification.

7.9 Proposed Extension for Next Chapter

In this chapter, we have covered all aspects of the plaque segmentation models. Attention based UNet model is capable of segmenting the critical images as well. However, we have incorporated attention-module only in the basic UNet model. Therefore, there is still scope for adding the same in other UNet models.