Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Ejection Fraction Estimation using Deep Semantic Segmentation Neural Network on 2D Echocardiography Data

MD. GOLAM RABIUL ALAM[1], (Member, IEEE), ABDE MUSAVVIR KHAN[1], MYESHA FARID SHEJUTY[1], SYED IBNA ZUBAYEAR[1], MD. NAFIS SHARIAR TALUKDER.[1], (Member, IEEE), MOHAMMAD MEHEDI HASSAN[2],(Senior Member, IEEE), ABDU GUMAEI[2], AHMED ALSANAD[2]

[1]Department of Computer Science and Engineering, BRAC University, Mohakhali, Dhaka 1212, Bangladesh
[2]Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Ahmed Alsanad (e-mail: aasanad.ksu@gmail.com).

The authors are grateful to the Deanship of Scientific Research, King Saud University for funding through Vice Deanship of Scientific Research Chairs

ABSTRACT Ejection Fraction value denotes how much blood is pumped out of the heart to different parts of the body. It is a routine clinical procedure in heart function assessment, where the Left Ventricle of the heart has to be manually outlined by doctors in clinical settings to measure the Ejection Fraction value which is time-consuming and highly varies by the observer. However, deep learning methods can be used to automatically complete this type of outlining task with much ease and better efficiency. Most of the state-of-the-art automated Ejection Fraction estimation methods applied statistical or neural network models on generic and expensive clinical procedures like 3D Ultrasound, MRI and CT imaging. However, 2D echocardiography is a specialized diagnosis method that is inexpensive and routinely used in clinical diagnosis for heart diseases. This paper proposed an automated Ejection Fraction estimation system from 2D echocardiography images using deep semantic segmentation neural networks. The three different semantic segmentation neural networks namely UNet, ResUNet and Deep ResUNet have been applied to find their accuracy score based on the dice accuracy metric. The most accurate model among the three achieved a dice score of 82.6% was utilized to form the Left Ventricle segmentation network for End Systole and End Diastole images. The Ejection Fraction value is then determined by applying the volume measurement formula on the output of the Left Ventricle segmentation network. Therefore, the proposed automated Ejection Fraction system can be used in clinical settings to remove the eyeball estimation practice and reduce the inter-observer variability problem.

INDEX TERMS Semantic segmentation, deep learning, 2D echocardiography, ejection fraction, convolutional neural network, left ventricle, end systole, end diastole

I. INTRODUCTION

EJECTION FRACTION is a calculation of the volume of blood departing our heart every single time it goes through contractions and relaxations. The human heart is a complex, vital organ undergoing contractions and relaxations all through day and night to pump blood to each organ and cell of the body. When the heart is in a contraction state, it starts discharging blood from the pair of chambers (ventricles) which are for pushing out blood through the aorta and vena cava located at the lower part of the heart and when in a relaxation state, blood is refilled in the ventricles. Regardless of the power of contractions, the human heart is unable to pump out each and every drop of blood from the ventricle. "Ejection Fraction" or EF signifies the amount of blood that has been released from the ventricle during a single heartbeat. EF estimation is done on the chamber in the left-bottom side of the heart (Left Ventricle - LV) since the LV is the major chamber pumping blood all through the body after receiving oxygenated blood from the lungs. EF measurement is a routine clinical procedure in most hospitals, measured by 2-Dimensional Echocardiography. Determining EF is a cumbersome process since most ECHO ultrasound machines require an attending Cardiologist to outline the different chamber walls of the heart so that the software then can find out the EF value on the outlined heart chambers. As the process is time-consuming attending Cardiologists usually tend to note down an eyeball estimation of EF value on the report. This creates a problem in the EF measurement and comparison process since the eyeball estimation values are prone to inter-observer variability among different attending Cardiologists. EF measurement is one of the most common routine procedures done in hospitals for a variety of reasons like patient heart assessment before major surgery or clinical assessment of patients' heart condition. An autonomous system based on modern deep learning techniques could help solve the problem of manual eyeball estimation of different Cardiologists on the same report, removing the inter-observer variability problem and making EF measurement procedure less time consuming, saving both patient, doctor and hospitals valuable time.

Deep Learning (DL) is becoming more common day by day as processing power gets cheaper and dedicated Processing Units are integrated into everyday devices. While DL and Neural Network (NN) seem similar in nature it is not, DL incorporates NN within its architecture to solve most probabilistic predictions. Neural Network is an architectural structure consisting of many Machine Learning (ML) algorithms where artificial neurons are the main workforce of the system all collectively focusing on recognizing patterns, connections within a data set just like the human brain. Deep Learning develops on the idea of neural networks imitating the functions of a human brain going much deeper than simple neural networks all the while keeping into consideration feature transformation and extraction in multiple layers. Since the problem of measuring EF value needs to find out the volume of the LV of heart in both contracted (systolic) and expanded (diastolic) stage from 2D ECHO image, segmentation of the LV chamber of the heart [1], [2] would be the first problem to solve in implementing this autonomous EF estimation system.

Extracting the EF measurement of LV then required a precise depiction of the Left Ventricular Endocardium (LVE) also known as the left bottom-most heart wall in both contracted and expanded states. Not just any contractions or expansions rather those that occur in the seemingly largest expanded state of LV (End Diastole, ED) and the smallest contracted state of LV (End Systole, ES) during echocardiography procedure throughout which the ECHO machine keeps taking ultrasound video of the heart. This two-video frame is normally determined by doctors using ECG (Electrocardiogram) reading on the monitor of the ECHO machine, the largest positive spike in ECG reading denotes ES and the largest negative spike denotes ED. The difficulties commonly faced in segmenting ECHO views are as follows: i) the heart wall known as the myocardium and the blood pool shows up on video with weaker contrast; ii) vividness and exposure-related issues; iii) dissimilarity of dots (freckle pattern) along the heart wall because of the inclination of the ECHO machine ultrasound probe with reference of body substance (skin and muscle tissue vary from position to position); iv) the presence of fine muscles between heart muscle tissues (trabeculae) and thin muscles (papillary muscles) that escalates resembling the myocardium; v) noteworthy body substance pattern (echogenicity) variance in the denizens; and vi) formation, strength, progress variance of the shape of the heart over various subjects and clinical specimens.

The early studies on finding the measurements of the heart chamber particularly LV was conducted as early as 2011 which include work by Carneiro & Nascimento [1] where the measurements were estimated from ultrasound images using deep learning architectures and derivative-based search methods. Moreover, ejection fraction value is also estimated from MRI [3], [4] and CT imaging [5] using statistical approaches. Furthermore, [6], [7] utilized 3D ultrasound imaging with statistical and deep neural network models for EF estimation. All these studies have used expensive and generic diagnostic procedures for EF estimation purposes whereas 2D echocardiography is a specialized, inexpensive and routine procedure in the clinical setting for diagnosing heart diseases. Therefore, this paper focuses on EF estimation using 2D echocardiography.

The scantiness of a large, public and significant 2D ECHO data set has been a hindrance to further studies, research and in-depth assessment of the possibility of deep learning methods in order to determine Ejection Fraction value. Moreover, the applications of these NNs are quite successful in other therapeutic fields and routine procedures. Even though the number of medical imaging-related issues compared to DL methods has increased vastly this last decade, very few focused on Echocardiographic image segmentation, feature extraction and contouring. The majority of the EF estimation data set and works were focused on heart MRI images which is a rare and expensive clinical procedure. Mostly MRI images of the heart are produced for Medical research or in complex heart disease cases. The shortfall of better annotated 2D ECHO data sets is because of the challenge of extracting data from ECHO machines and gaining a huge number of images thoughtfully interpreted by Cardiologists for the ultimate nature of ECHO where a Cardiologist is required to take a video of the heart and manually annotate the different chambers of the heart. Moreover, segmentation of the Left Ventricle (LV) of the heart required a data set collected from the same type of ECHO machine with contours on the LV chamber in any of the major ECHO views (A2c, A4c, Plax, Psax, etc). Therefore, data set has been collected on A4c view of 2D ECHO which is a common clinical procedure in detecting heart diseases.

This paper proposes an automated EF Estimation System of the human heart to aid healthcare services, patients & cardiologists. The system is built utilizing the deep semantic segmentation neural networks namely UNet, ResUNet, Deep ResUNet models. The EF value is then determined by applying the volume measurement formula on the output of the LV segmentation network. The key contributions of this research are as follows:

1) An Ejection Fraction estimation system for 2D echocardiography images has been presented for the automated measurement of Ejection Fraction value that can be used in clinical settings to remove the eye-ball estimation practice and reduce the inter-observer variability problem. The three different deep semantic segmentation neural networks namely UNet, ResUNet and DeepResUNet have been utilized to form the Left Ventricle segmentation network for both End Systole and End Diastole images.
2) Two parallel pipelines of deep semantic segmentation neural network models are used on End Systole and End Diastole data set for segmentation of the Left Ventricle of the heart in its Systolic (Contracted) and Diastolic (Expanded) state for better accuracy in the different states for much more accurate Ejection Fraction value.

This paper has been organized as follows. Section 2 presents the previous works. Section 3 discusses the data set. The proposed method for EF estimation using semantic segmentation networks and the whole workflow of the system have been discussed in Section 4. In Section 5, the results of the training of the segmentation network and its accuracy metrics have been discussed. Section 6 concludes the paper with future research directions.

II. PREVIOUS WORK

A. DATASETS

The Echocardiography data has been made publicly available for the MICCAI 2014 Challenge on Endocardial Three-dimensional Ultrasound Segmentation (CETUS) [8]. The CETUS data set composed of 45 3D echocardiographic views (15 for training, 15 for phase 1 testing, 15 for phase 2 testing) divided into three subcategories: i) 15 healthy patients; ii) 15 patients who had previous myocardial infarction history; iii) 15 patients with a history of dilated cardiomyopathy. Different automatic (deformable models, Hough random forest, Kalman filter, etc) and semi-automatic methods (graph-cut method, structured random forest, etc) were applied on the data set by different challengers other than deep semantic segmentation networks for segmentation of LV [9].

B. VOLUME ESTIMATION AND EJECTION FRACTION MEASUREMENT

In [10], Echocardiographic LV volumes had been calculated through LV dimension parameters (Dd and Ds) of the ECHO image using the volume formula of a prolate spheroid (ellipsoid of revolution around the significant diameter) as in (1).

$$V = \frac{\pi}{6} D_1 D_2 L \tag{1}$$

Here D_1 and D_2 represents the angiographic anteroposterior and lateral minor diameters respectively and L represents the longer major diameter from the two radiographic projections which can be clearly spotted by looking at the Fig. 1.

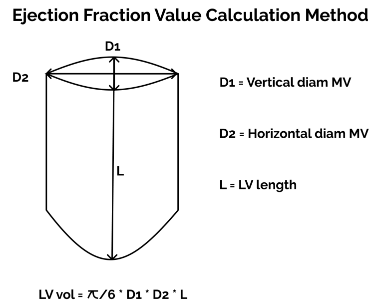


FIGURE 1. In a geometric model, the short-axis area multiplied by the long-axis length is used to estimate left ventricular (LV) volumes using two-dimensional echocardiography

However, respective minor diameters for the LV are approximately identical. Therefore, equation (1) can be written as (2).

$$V = \frac{\pi}{6} D^2 L \tag{2}$$

Furthermore, if the long diameter is considered to be two times the short diameter, the formula is further simplified to (3).

$$V = \frac{\pi}{6} D^2 (2D) \tag{3}$$

Similarly, volume in both systole and diastole states approximately equals the cube of the short diameter. This formula assumes the algebraic cancellation of the factor $\pi/3$. As a result, ECHO LV volume in both states as shown in Fig. 2 can easily utilize ultrasonic dimensions for calculation. Systolic dimension, $(D_s)^3$, could be used to find out the value of End Systolic Volume (ESV). Similarly, End Diastolic Volume (EDV) could be calculated from $(D_d)^3$. Therefore, EF can be calculated from equation (4).

$$EF = \frac{EDV - ESV}{EDV} \tag{4}$$

The detailed derivation of volume estimation (1), (2), (3) and (4) is presented in [10].

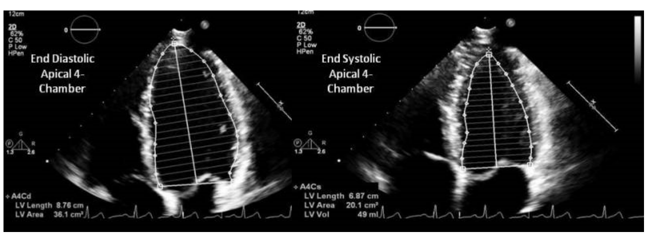


FIGURE 2. Ejection Fraction Estimation from Apical 4-Chamber (A4c) view by Simpson's method, source: [11]

C. IMAGE SEGMENTATION

The prospects and challenges of deep learning methods for cardiac image segmentation has been discussed thoroughly in [12]. While the convolutional neural networks (CNN's) [13] had been existing for a long time, its success was limited to the size of the available training sets and the size of the considered networks, also larger data set doesn't always mean better performance and accuracy since many networks after a certain portion of data input show diminishing returns which are not worth the extra computation time [14]. However, usually, a larger data set produces better validation accuracy in CNN as found by [14], [15]. The common and widespread use of the legacy CNN models are on image classification tasks, where pixel-wise consideration is not done and the output is a single class labeled image. However, in many cases like in biomedical image processing, the desired output must include localization, which means a class label must be assigned to each pixel of the image. Moreover, the availability of thousands of training images is usually beyond reach in biomedical tasks. The authors in [16] proposed the U-Net that is a novel neural network architecture and training strategy that relied on the strong use of data augmentation to use the available annotated samples more efficiently. This made U-Net architecture very popular for semantic segmentation problems even in cases as above. In [16], the U-Net model was trained in a sliding-window set up to predict the class label for each pixel by forming a patch around that pixel.

The U-Net model since its inception has utilized a narrowing path (also known as Encoder path) in capturing contextual information and a symmetric widening path (also known as Decoder path) that enabled precise localization (forming patches around the selected area) as shown in Fig. 3. Furthermore, U-Net model showed that it could be trained to outperform the previous best methods (a sliding-window convolutional network) even with very few pictures. An example application of this semantic segmentation architecture to segment the carotid vessel wall in 3D ultrasound images shown in [17].

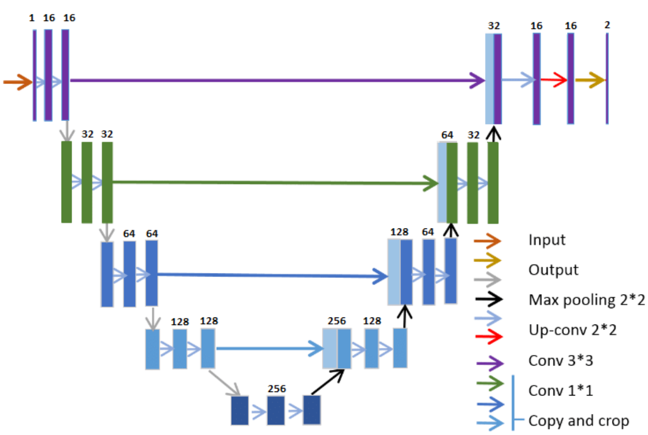


FIGURE 3. The U-Net architecture utilized in the proposed 2D ECHO image segmentation for EF estimation, the boxes represents multi-channel feature map. Number of channels denoted on top

Drozdzal et al. [18] studied the effect of both long and short skip connections on Fully Convolutional Networks (FCN's) for segmentation tasks in medical imaging. Even though it is related to FCN's the impact of skip connections both long and short might help in further increasing the accuracy of semantic segmentation in LV identification from the 2D Echocardiography data set. This is further demonstrated by their FCN arguments. CNN's are generally implemented using a narrowing path implemented using convolutional, pooling and Full Convolution layers, FCN's built for segmentation develops further on an expanding path built with upsampling layers or deconvolutional layers. The widening path recovers spatial information lost by the contracting path by keeping together features ignored from the various resolution levels or layers on the shortening path. Therefore, in standard FCN's, mostly lengthy skip connections are being utilized to skip features from the contracting path (Encoder-like function) to the expanding path (Decoder-like function) to recover spatial information lost during downsampling. The research reflects upon that by extending normal FCN's to include short skip connections, which works similarly to the ones introduced in ResNet which makes it easier to implement a very deep FCN of hundreds of layers. The research concluded by showing an implementation of a very deep FCN which achieved the best possible results on the EM data set without the need for extra post-processing. Similarly, there was an advent of the 3D U-Net architecture which worked similar to normal U-Net but also achieved good results on sparsely annotated training samples [19].

Kaiming et al. [20], presented a neural network model implementing residual learning making it easier for network training to go far deeper than common neural networks. The research observed that as deeper neural networks begin to converge the problems of degradation are exposed. With increasing depth, the accuracy is saturated and more training of the network becomes useless and degrades rapidly. Unfortunately, this type of deterioration is not due to over-fitting, which could be solved easily. Moreover, adding more NN layers to a very deep traditional NN model leads to higher training error, as reported in [21] and thoroughly proved by experiments. To overcome the obstacles with training deeper networks the layers had been transformed as learning residual functions referring to the layer inputs, instead of learning on disconnected functions. Furthermore, solid evidence was provided that residual networks had easier optimization, and accuracy gains even from increased depth. Accordingly, residual networks were implemented with depths of up to 152 layers which was 8 times deeper than VGG nets while at the same time had lower complexity. Furthermore, this model of residual nets achieved a 3.57% error on the ImageNet data set. The model was also able to train on CIFAR-10 data set with a very deep layer neural network. The study concluded that the depth of representations in Neural Networks for many visual recognition tasks was very important, which could be solved by incorporating Residual Blocks in most Neural Networks.

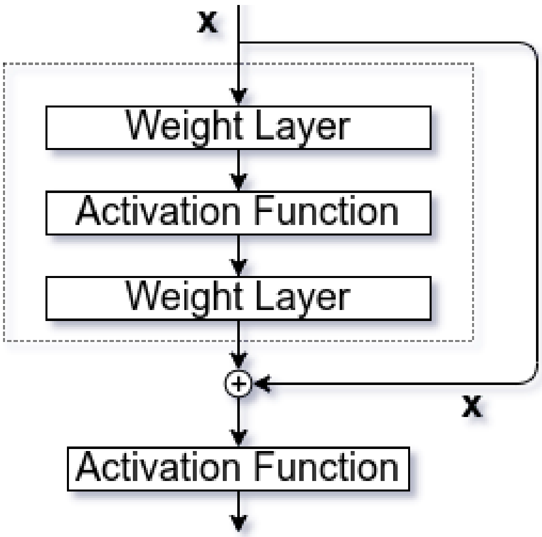


FIGURE 4. Residual Block utilizing skip connection

Zhengquan et al. [22], studied the segmentation of Sea-land as significant work on near seashore ship observation and coastal monitoring based on satellite view images. Even though semantic segmentation networks displayed good prospects in this task, the research pursued for improvement in accurately segmenting the fine details of the small ships near the coast viewed from the top using satellite images. The proposed method was based on U-Net utilizing some additional layers for sea-land segmentation. The contraction part of the U-Net architecture was modified to incorporate residual blocks (ResNet) which specialized in handling complicated scenes by helping to form a much deeper Neural Network that was able to extract even tiny details. As a result, the final semantic segmentation network structure was named Res-UNet.

Diakogiannis et al. [23], presented the understanding of the scene of first-rate satellite or aerial images had significant value for automated monitoring in numerous remote sensing applications. Because of the large within-class and small between-class variance in pixel values of objects of interest, this remained a complex task. At that time deep CNN's had gained popularity in remote sensing applications and demonstrated state-of-the-art performance in classifying objects at the pixel level. The research presented a novel DL architecture, ResUNet-a, which combined concepts from the state-of-the-art modules for semantic segmentation tasks used in computer vision. The presented model differed from normal U-Net or Res-UNet with the application of a PSP Pooling layer to connect the encoder and decoder portion of the neural network. The model was evaluated on the ISPRS 2D Potsdam data set which resulted in an average state-of-the-art performance Dice score of 92.9% overall classes.

III. DATA

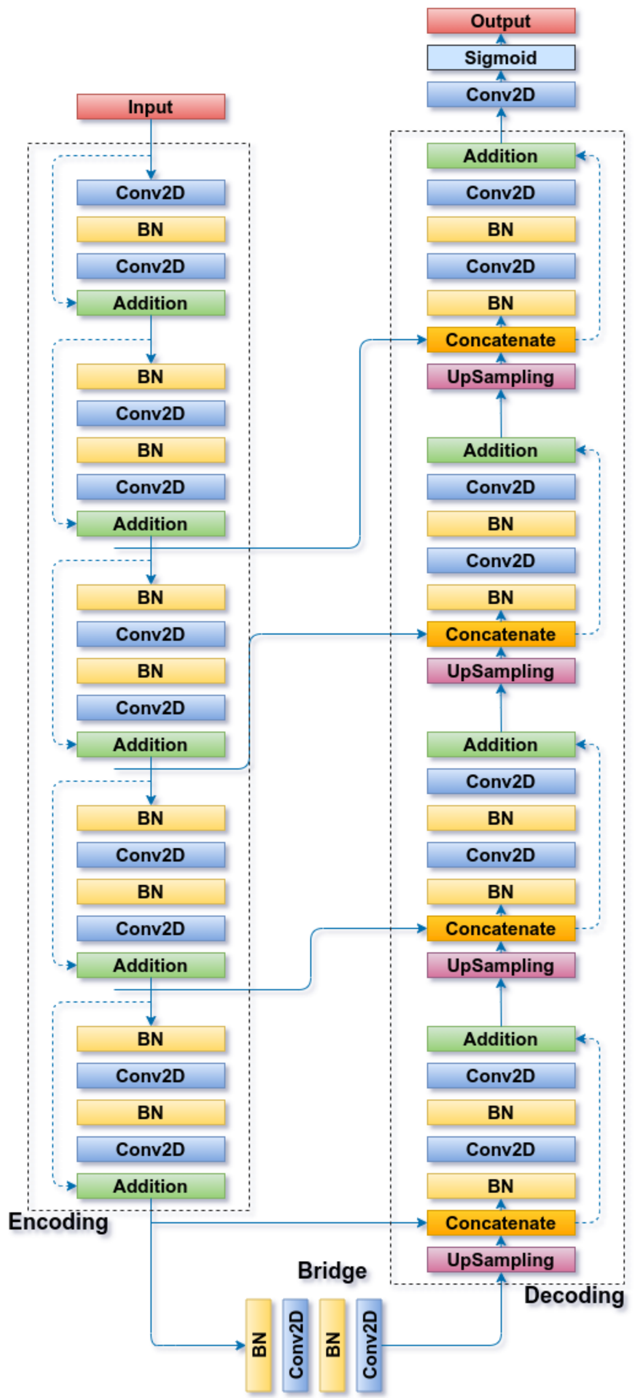


FIGURE 5. Overview of the ResUNet architecture. The left (downward) branch is the encoder part of the architecture. The right (upward) branch is the decoder. The last convolutional layer has as many channels as there are distinct classes

A. DATA SET ACQUISITION

As described in section II, publicly available Echocardiographic data sets are rare and CETUS data set [8] consisted of 45 3D echocardiographic views. However, this research focused on EF estimation from 2D echocardiography data. Therefore, this research required active data collection from health care providers. Echocardiographic data mainly A4C view images have been collected from National Heart Foundation Hospital Research Institute with consent from the academic council of the hospital after forwarding data request application from BRAC University. The full data set has been acquired from a GE Vivid E95 ultrasound Echocardiography machine. The only protocol followed in collecting the data set followed the usual clinical routine. This has been a very slow process since most patient data from routine procedures had abnormal heart conditions where the heart chamber walls have stretched too far out or the chambers are enlarged and the heart walls are very thin. Training the neural network model for segmenting the LV required normal heart chamber volume. Therefore, data of about 400 patients had been collected in MHD format containing RAW images of A4C views in both ES and ED stages (most expanded volume and mostly contracted volume), also image masks for these were well-annotated by an attending Cardiologist for ground truth reference of LV segment.

B. DATA PRESENTATION

Data set provided by the hospital consisted of .mhd format where .mhd file contained the header with different header information for the raw image files like object type of the image, dimensions of the image, binary pixel value flag, compressed or not, transformation matrix value, the center of rotation value, anatomical orientation parameter, pixel offset values, element spacing for each pixel, the original resolution of the image and the actual raw file name. Visual representation of the collected data has been presented in Fig. 6.

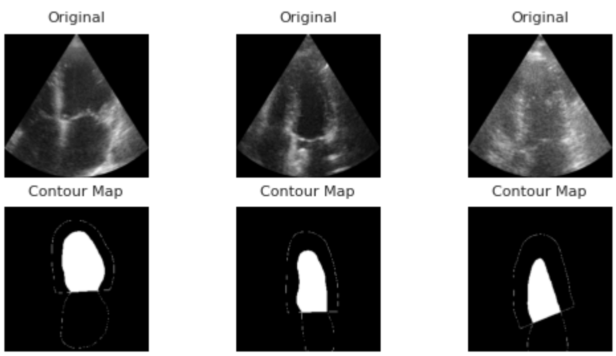


FIGURE 6. Images labeled Original, represent the normal A4C view of the heart of 3 different patients whereas the images labeled Contour Map, represent the Left Ventricle annotated by the attending Cardiologist.

C. DATA PREPROCESSING

The collected data were in raw format with all its header stored in mhd format. Therefore, no extra preprocessing was necessary. However, the 2D ECHO view images were of the dimension $549 * 778 * 1$ have been transformed into $256 * 256 * 1$. Furthermore, random transformations were also applied to the input data set while feeding to the semantic segmentation network to maximize learning efficiency. Moreover, images were re-scaled to crop out unnecessary blacked-out portions from the two sides of the 2D ECHO image.

IV. PROPOSED METHOD FOR EF ESTIMATION

This section progressively describes the proposed system model, the semantic segmentation network models and the automated EF estimation procedure.

A. PROPOSED SYSTEM MODEL FOR EF ESTIMATION

The system model of the proposed automatic EF estimation from 2D echocardiography has been presented in Fig. 6. To develop the system, the 2D echocardiography data (A4c view) has been collected from NHFH&RI using GE Vivid E95 ultrasound Echocardiography machine. The detailed procedure of data set acquisition is discussed in Section III(A). The collected data has been preprocessed through transformation and re-scaling to make it suitable for training the semantic segmentation neural network. The detailed procedure of data preprocessing has been discussed in Section III(B)&(C). Afterward in the sampling phase, the data set has been split to prepare training, validation and testing samples. Of the total data samples, 80% have been chosen for training whereas 20% of the data samples have been taken for testing. Out of the 80% training samples, then 5% of the samples have been chosen for validation purposes. The training data set is then fed into two parallel pipelines of deep semantic segmentation neural network models for the segmentation of the LV of the heart in its systolic and diastolic states. Since the proposed system has been developed for use in clinical settings, the estimated EF value needed to be as accurate as possible. Therefore, the system has been designed for higher accuracy in segmenting the LV of the heart in both ES and ED states. Most of the state-of-the-art EF estimation systems used either a fused or single NN model to accomplish the task of EF estimation in both ES and ED states. However, the proposed model followed a different approach i.e., parallel pipelines of deep semantic segmentation neural network models are used on End Systole and End Diastole data set for segmentation of the Left Ventricle of the heart in its Systolic (Contracted) and Diastolic (Expanded) states. Therefore, the proposed model ensures maximum LV segmentation accuracy both in ES and ED states for optimizing EF estimation. The deep semantic segmentation neural network composed of training, optimization and post-processing modules utilizes residual blocks, long and short skip connections, psp pooling, upsamplings and downsamplings in encoder and decoder neural network architecture to ensure better LV segmentation accuracy. The detailed discussion of the deep semantic segmentation networks has been presented in Section IV(B).

After completion of LV segmentation, the End Systolic Volume and End Diastolic Volume of the heart chamber (LV) are then calculated using equation (5).

$$V = \frac{\pi}{6} D^2 L \tag{5}$$

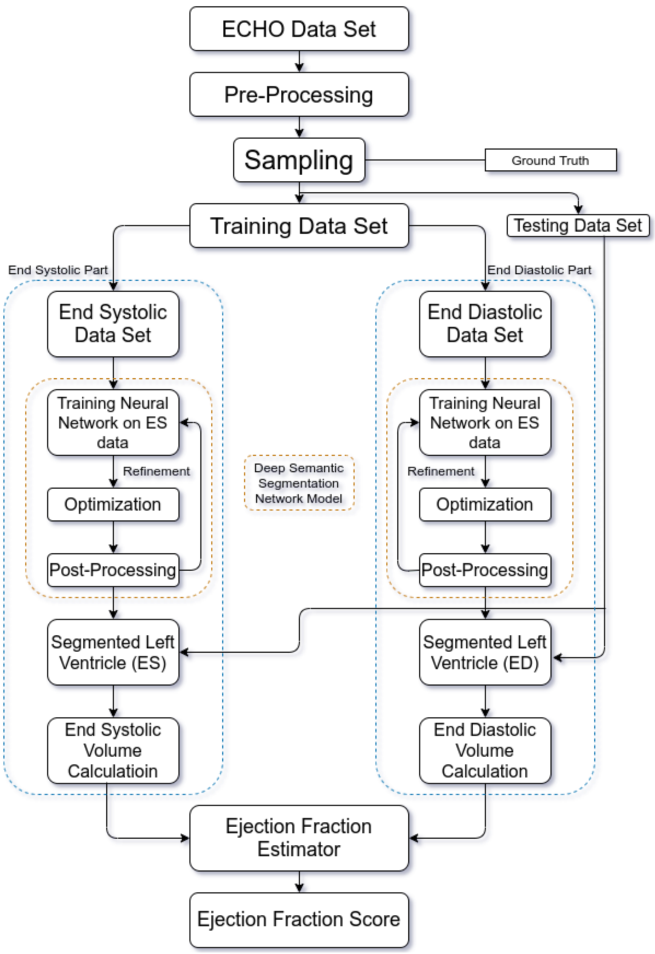


FIGURE 7. Proposed System model for EF estimation from 2D echocardiography data

Where V denotes either ESV or EDV depending on the LV segmented ECHO image chosen, D represents the maximum diameter of the LV segment laterally or horizontally and L represents the maximum vertical or longitudinal length of the LV chamber segment. The details of the volume estimation procedure have been discussed in Section II(B).

After finding out the value of ESV and EDV, the values are forwarded to the EF estimator block to determine the EF value. The details of the EF estimation block have been discussed in section IV(C).

B. SEMANTIC SEGMENTATION MODELS BASED ON ENCODER-DECODER ARCHITECTURE

The detailed overview of the utilized and evaluated deep semantic segmentation NN models are as follows:

1) U-Net Architecture: Different variations of the U-Net architecture are available based on the solution approach and the nature of a problem. For this LV segmentation from 2D Echocardiography image the U-Net architecture presented in Figure 3 has been used. The presented architecture is based on encoder and decoder path utilizing only long skip connections from the encoder path to the decoder path that have been evaluated on the acquired data set. The U-Net architecture is symmetric around the encoder and decoder path, the number of filters chosen in the downsampling layers of the encoder path are 32, 64, 128, 128, 128 whereas in the upsampling layers of decoder path 128, 128, 64, 32, 16.

2) ResUNet Architecture: The ResUNet architecture as shown in Figure 5 is based upon U-Net architecture and ResNet-32 architecture that has been used and evaluated for the proposed LV segmentation model from 2D Echocardiography image. ResUNet architecture utilizes Residual blocks (ResNet) where the block comprises of two Convolutional blocks, short skip connection with convolution or without convolution using binary normalization. The number of filters in the downsampling process in residual blocks are 16, 32, 64, 128, 256 in the bridging convolution layers filter size is constant at 256 whereas in the upsampling layers in the decoder path 256, 128, 64, 32, 16 respectively. The total number of layers in the whole network stood at 32.

3) Deep ResUNet Architecture: This architecture named Deep ResUNet is composed of U-Net architecture and ResNet-50 architecture. Similar to ResUNet architecture this network also utilized Residual blocks and encoder, decoder path but differed in the way that this network went much deeper by increasing the number of layers. of the whole network to 50 layers. The proposed EF estimation system used a deep semantic segmentation network (i.e., Deep ResUNet) of 50 layers.

All 3 of these Networks have been trained on ECHO data of 380 patients with contoured LV segments presenting the ground truth of segmentation. Of these 380 patients, 19 of them were chosen at random for validation of the Neural Network while the rest 20 patient data is for testing purposes to find out whether the model was able to perform well on unseen data.

All of the evaluated deep semantic segmentation neural network models have been trained for 200 epochs in batch sizes of 8 on the acquired data set to keep similarity in the training stage as much as possible which further helped in the evaluation of the models fairly. The decision on using 3 x 3 filter size for the convolutions was made upon observations and further studies on optimal filter size for CNN. For an odd-sized filter, all the previous layer pixels are placed symmetrically around the output pixel. Without this symmetry, distortions may happen across the layers which happens in the case of even-sized kernels. As a result, even-sized kernel filters are mostly skipped to promote implementation simplicity. Furthermore, if convolutions are thought of as interpolation from the given pixels to a center pixel, interpolation to a center pixel would not be possible using an even-sized filter. On observation of the impact of filter size, 3 x 3 has been found to be the optimum choice in detecting the edges of the segments than 5 x 5 [24] also since the data set was fed into the network after transformation to 256 * 256 * 1 image resolution 3 x 3 filter size performed much better in terms of validation accuracy.

C. AUTOMATED EJECTION FRACTION ESTIMATION

The ejection fraction estimation algorithm is presented in Algorithm 1. As per the discussion in section IV(A), the two parallel deep semantic segmentation models have been trained to find out the LV segments from ED and ES ECHO views that are represented as ESLV and EDLV respectively in Algorithm 1. Usually, the LV volume measurement systems utilize different geometric and hand annotations to find estimations of EF score [10], [25], [26]. Therefore, the longer major diameters of the segmented ESLV and EDLV have been determined as ESlength and EDlength respectively. Moreover, the maximum lateral minor diameter of the segmented ESLV and EDLV have been measured as ESdepth and EDdepth respectively. Afterwards, the end systolic volume (ESV) and the end diastolic volume (EDV) have been determined through (3). Finally, Ejection Fraction (EF) value has been calculated through (4).

Algorithm 1: Ejection Fraction Estimation Algorithm

Input: Patient Data P; End Systolic image Left Ventricle segmentation model ES; End Diastolic image Left Ventricle segmentation model ED

Output: Ejection Fraction Value

$ESLV \leftarrow$ Segmented LV from ES image of patient P using segmentation network model of ES

$EDLV \leftarrow$ Segmented LV from ED image of patient P using segmentation network model ED

$ESlength \leftarrow$ Longer major diameter of the segmented $ESLV$

$ESdepth \leftarrow$ Maximum lateral minor diameter of the segmented $ESLV$

$EDlength \leftarrow$ Longer major diameter of the segmented $EDLV$

$EDdepth \leftarrow$ Maximum lateral minor diameter of the segmented $EDLV$

$ESV \leftarrow (\pi * ESdepth^2 * ESlength) / 6$

$EDV \leftarrow (\pi * EDdepth^2 * EDlength) / 6$

$EF \leftarrow ((EDV - ESV) * 100) / EDV$

return EF

V. RESULTS

For the evaluation of a deep semantic segmentation neural network model, predictions are usually classified into the following 4 categories: i) True Positive: for correctly predicted event values, ii) False Positive: for incorrectly predicted event values, iii) True Negative: for correctly predicted no-event values and iv) False Negative: for incorrectly predicted no-event values [27]. Based on these parameters Dice Coefficient (F1 Score) and Intersection Over Union (IoU, Jaccard Index) scores have been calculated. Semantic segmentation classifies each pixel in a picture therefore input and output spatial resolutions are matching where the depth of the channel is the same as the number of possible prediction states. For evaluating the segmentation accuracy of the different semantic segmentation neural networks dice similarity coefficient has been prioritized here as the dice coefficient score is much more reliable in terms of finding whether the predicted segment is as close as the given annotation [28].

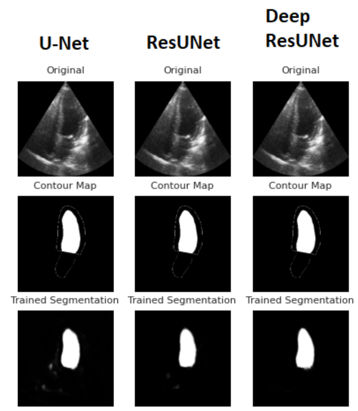


FIGURE 8. Results on testing or unseen data (ED).

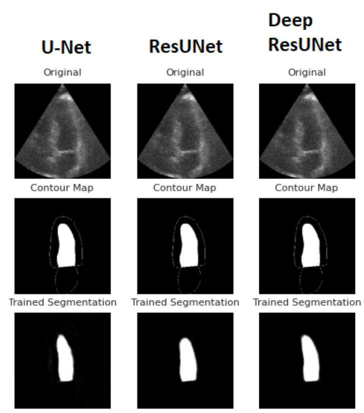


FIGURE 9. Results on testing or unseen data (ES).

Another evaluation method of semantic segmentation is the IoU score which evaluates pixel-wise similarity irrespective of how pixels are grouped in bounded regions. Pixel accuracy scores are another format of evaluation utilizing the simplest metric of computing a ratio between the amount of properly classified pixels and the total number of pixels [29]. This evaluation metric though popular gives higher accuracy scores in most segmentation tasks even when other evaluation metrics score much lower since the number of pixels in the non-segmented area may be much higher than the number of pixels in the segmented area [30].

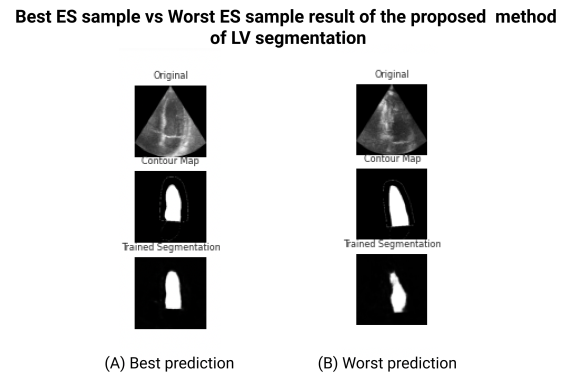


FIGURE 10. Deep ResUNet best and worst predictions of ES model

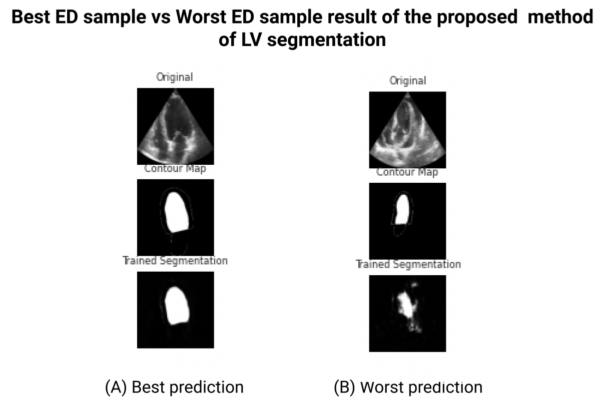


FIGURE 11. Deep ResUNet best and worst predictions of ED model

Apart from the above mentioned metrics, another evaluation metric named as the Hausdorff distance (HD) is used which is the largest distance that one may be compelled to travel by an opponent who selects a point in one of the two sets from which one must then go to the other set. In other words, it is the longest distance between a point in one set and the nearest point in the other set [31].The Hausdorff distance (HD) may be used in computer vision to locate a given template in an arbitrary target picture. After that, the algorithm tries to reduce the HD between the template and a portion of the target picture. The target picture area with the shortest Hausdorff distance to the template is the best option for locating the template in the target.

Testing the trained models on unseen data for results of generalization the resulting segmentation images was found as that of Fig 8 & Fig. 9. Also, Fig. 10 & Fig. 11 depicts our suggested models best and worst forecasts for the ES and ED models, respectively.

Now the mean performance scores of the evaluation metrics on different models of the network for both ES and ED images have been depicted in Table. 1 where the statistics column roman value (I) and (II) represents mean ± standard deviation, confidence interval respectively. The table also shows that our recommended methodology is effective. In comparison to the other two approaches, the Hausdorff distance (HD) is comparatively lower in terms of considering both ES and ED models metric together. This metric helped us in determining the final image segmentation model to be used by the system.

TABLE 1. Mean, standard deviation, confidence interval (confidence level 95%) of HD, Dice Score, IoU metric scores for final evaluation of the best deep semantic segmentation neural network

Data	Models	Statistics	HD	Dice score	IoU
ED	UNet	(I)	24.9 ± 1.1	73.6 ± 0.9	44.2 ± 0.00
		(II)	26.2 - 23.5	74.6 - 72.6	44.24 - 44.20
	ResUNet	(I)	19.6 ± 0.1	86.2 ± 0.6	64.6 ± 6.1
		(II)	19.8 - 19.4	87.0 - 85.43	71.5 - 57.7
	Deep	(I)	19.7 ± 0.2	86.5 ± 1.1	63.7 ± 9.6
	ResUNet	(II)	19.96 - 19.95	87.62 - 85.51	72.85 - 5462
ES	UNet	(I)	30.4 ± 0.7	66.4 ± 0.7	46.2 ± 0.00
		(II)	31.2 - 29.6	67.3 - 65.5	46.24 - 46.20
	ResUNet	(I)	23.9 ± 0.3	79.7 ± 1.2	66.9 ± 6.4
		(II)	24.28 - 23.55	81.17 - 78.42	74.28 - 59.58
	Deep	(I)	23.8 ± 0.1	82.1 ± 0.8	60.0 ± 11.1
	ResUNet	(II)	24.0 - 23.7	82.96 - 81.36	70.68 - 49.48

VI. CONCLUSION

In most of the state-of-the-arts, the ejection fraction value had been estimated from the 3D Ultrasound, MRI and CT imaging. However, 2D echocardiography is a specialized diagnosis method which is inexpensive and routinely used in clinical diagnosis for heart diseases. Therefore, this research proposed an ejection fraction estimation system for 2D echocardiography using deep learning based semantic segmentation neural networks. Among the three utilized semantic segmentation neural networks the Deep ResUNet outperforms the other two models i.e., UNet and ResUnet with respect to dice score, IoU and pixel accuracy. However, the major problems with deep neural networks particularly deep CNN's like ResNet [32] and much deeper NN's, training becomes very slow if there is no GPU or dedicated parallel processors [33], [34]. Even with the availability of a wide range of GPU accelerated platforms for deep neural network training the speed varies with the architecture of the GPU system where server-grade hardware performs better than most low-cost consumer hardware [35]. As a solution to this problem, the end systole and end diastole datasets have been trained in two parallel pipelines of the deep semantic segmentation neural networks. The proposed solution gained network architectural efficiency and induced better EF accuracy due to the separated precise segmentation of the Left

Ventricle (LV) of the heart in its Systolic (Contracted) and Diastolic (Expanded) states.

Some bottlenecks faced during the study constrained the number of hidden layers in the Deep ResUNet model to only 50 layers whereas with a larger image resolution data set much deeper ResUNet models up to 152 layers [20] could be implemented. Furthermore, more studies could be done by modifying the residual block with squeeze and excitation block [36]. Also, atrous spatial pyramidal pooling could be a consideration [37]. Results obtained by different UNet-based methods using proper statistical tests can be compared with the results from this research as a comparative analysis to further enhance the scope of the work. Furthermore, future research could be done in detecting Global Longitudinal Strain (GLS) through the semantic segmentation models on the collected data set.

REFERENCES

[1] G. Carneiro, J. Nascimento, and A. Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," IEEE Transactions on Image Processing, vol. 21, pp. 968–982, 2012.

[2] E. Smistad, A. Østvik et al., "2d left ventricle segmentation using deep learning," in 2017 IEEE International Ultrasonics Symposium (IUS). IEEE, 2017, pp. 1–4.

[3] J. Lebenberg, I. Buvat, A. Lalande, P. Clarysse, C. Casta, A. Cochet, C. Constantinides, J. Cousty, A. de Cesare, S. Jehan-Besson, M. Lefort, L. Najman, E. Roullot, L. Sarry, C. Tilmant, M. Garreau, and F. Frouin, "Nonsupervised ranking of different segmentation approaches: Application to the estimation of the left ventricular ejection fraction from cardiac cine mri sequences," IEEE Transactions on Medical Imaging, vol. 31, no. 8, pp. 1651–1660, 2012.

[4] V. Ray and A. Goyal, "Image based sub-second fast fully automatic complete cardiac cycle left ventricle segmentation in multi frame cardiac mri images using pixel clustering and labelling," in 2015 Eighth International Conference on Contemporary Computing (IC3), 2015, pp. 248–252.

[5] Y. Liu, V. Sandoval, and A. J. Sinusas, "Potential impact of hybrid czt spect/ct imaging on estimation accuracy of left ventricular volumes and ejection fraction: A phantom study," in 2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC), 2013, pp. 1–5.

[6] E. Smistad, A. Østvik, I. M. Salte, D. Melichova, T. M. Nguyen, K. Haugaa, H. Brunvand, T. Edvardsen, S. Leclerc, O. Bernard, B. Grenne, and L. Løvstakken, "Real-time automatic ejection fraction and foreshortening detection using deep learning," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 67, no. 12, pp. 2595–2604, 2020.

[7] D. Krishnaswamy, A. R. Hareendranathan, T. Suwatanaviroj, H. Becher, M. Noga, and K. Punithakumar, "A semi-automated method for measurement of left ventricular volumes in 3d echocardiography," IEEE Access, vol. 6, pp. 16 336–16 344, 2018.

[8] "Challenge on endocardial three-dimensional ultrasound segmentation MICCAI challenge 2014," Mar 2014. [Online]. Available: https://www.creatis.insa-lyon.fr/Challenge/CETUS/databases.html

[9] O. Bernard, J. G. Bosch, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-Pop, F. Cervenansky, S. Valette, O. Mirea, M. Bernier et al., "Standardized evaluation system for left ventricular segmentation algorithms in 3d echocardiography," IEEE transactions on medical imaging, vol. 35, no. 4, pp. 967–977, 2015.

[10] J. F. Pombo, B. L. Troy, and R. O. RUSSELL JR, "Left ventricular volumes and ejection fraction by echocardiography," Circulation, vol. 43, no. 4, pp. 480–490, 1971.

[11] A. Kosaraju, A. Goyal, and Y. Grigorova, Left Ventricular Ejection Fraction. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-, May 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK459131/figure/article-24159.image.f1/

[12] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O'Regan et al., "Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation," IEEE TRANSACTIONS ON MEDICAL IMAGING, vol. 37, no. 2, 2018.

[13] M. G. R. Alam, S. F. Abedin, S. I. Moon, A. Talukder, and C. S. Hong, "Healthcare iot-based affective state mining using a deep convolutional neural network," IEEE Access, vol. 7, pp. 75 189–75 202, 2019.

[14] J. Barry-Straume, A. Tschannen, D. W. Engels, and E. Fine, "An evaluation of training size impact on validation accuracy for optimized convolutional neural networks," SMU Data Science Review: Vol. 1 : No. 4 , Article 12., vol. 1, no. 4. [Online]. Available: https://scholar.smu.edu/datasciencereview/vol1/iss4/12/

[15] S. Uchida, S. Ide, B. K. Iwana, and A. Zhu, "A further step to perfect accuracy by training cnn with larger data," in 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016, pp. 405–410.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[17] M. Jiang, J. D. Spence, and B. Chiu, "Segmentation of 3d ultrasound carotid vessel wall using u-net and segmentation average network," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 2043–2046.

[18] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in Deep Learning and Data Labeling for Medical Applications. Springer, 2016, pp. 179–187.

[19] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in International conference on medical image computing and computer-assisted intervention. Springer, 2016, pp. 424–432.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[21] K. He and J. Sun, "Convolutional neural networks at constrained time cost," 2014.

[22] Z. Chu, T. Tian, R. Feng, and L. Wang, "Sea-land segmentation with resunet and fully connected crf," in IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019, pp. 3840–3843.

[23] F. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. arxiv 2019," arXiv preprint arXiv:1904.00592.

[24] W. S. Ahmed and A. a. A. Karim, "The impact of filter size and number of filters on classification accuracy in cnn," in 2020 International Conference on Computer Science and Software Engineering (CSASE), 2020, pp. 88–93.

[25] D. Bamira and M. Picard, "Imaging: Echocardiology—assessment of cardiac structure and function," in Encyclopedia of Cardiovascular Research and Medicine, R. S. Vasan and D. B. Sawyer, Eds. Oxford: Elsevier, 2018, pp. 35–54. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128096574109536

[26] N. Benyounes, C. Van Der Vynckt, T. Tibi, A. Iglesias, O. Gout, S. Lang, and L. Salomon, "Left ventricular end diastolic volume and ejection fraction calculation: Correlation between three echocardiographic methods," Feb 2020. [Online]. Available: https://doi.org/10.1155/2020/8076582

[27] M. G. R. Alam, S. F. Abedin, M. Al Ameen, and C. S. Hong, "Web of objects based ambient assisted living framework for emergency psychiatric state prediction," Sensors, vol. 16, no. 9, p. 1431, 2016.

[28] T. Eelbode, J. Bertels, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimization for medical image segmentation: Theory and practice when evaluating with dice score or jaccard index," IEEE Transactions on Medical Imaging, vol. 39, no. 11, pp. 3679–3690, 2020.

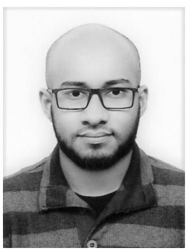
[29] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017.

[30] Y. Zhang, S. Mehta, and A. Caspi, "Rethinking semantic segmentation evaluation for explainability and model selection," 2021.

[31] W. J. Rucklidge, "Efficiently locating objects using the hausdorff distance," International Journal of computer vision, vol. 24, no. 3, pp. 251–270, 1997.

[32] Y. Lan and X. Zhang, "Real-time ultrasound image despeckling using mixed-attention mechanism based residual unet," IEEE Access, vol. 8, pp. 195 327–195 340, 2020.
[33] L. Cai, J. Gao, and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation," Annals of Translational Medicine, vol. 8, no. 11, 2020. [Online]. Available: https://atm.amegroups.com/article/view/36944
[34] S. Shi, Q. Wang, P. Xu, and X. Chu, "Benchmarking state-of-the-art deep learning software tools," in 2016 7th International Conference on Cloud Computing and Big Data (CCBD), 2016, pp. 99–104.
[35] J. Jo, S. Jeong, and P. Kang, "Benchmarking gpu-accelerated edge devices," in 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), 2020, pp. 117–120.
[36] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, 2018.
[37] J. Wang, P. Lv, H. Wang, and C. Shi, "Sar-u-net: squeeze-and-excitation block and atrous spatial pyramid pooling based residual u-net for automatic liver ct segmentation," 2021.

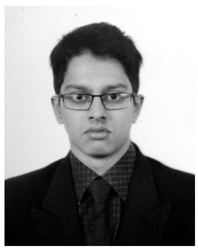
MD. GOLAM RABIUL ALAM (S'15–M'17) received the B.S. and M.S. degrees in computer science and engineering, and information technology, respectively, and the Ph.D. degree in computer engineering from Kyung Hee University, South Korea, in 2017. He also served as a Postdoctoral Researcher with Computer Science and Engineering Department, Kyung Hee University, South Korea, from March 2017 to February 2018. He is currently an Associate Professor with the Computer Science and Engineering Department, BRAC University, Bangladesh. His research interests include health-care informatics, mobile cloud and edge computing, ambient intelligence, and persuasive technology. He is a member of the IEEE IES, CES, CS, SPS, CIS, KIISE, and the IEEE ComSoc. He also received several Best Paper Awards in prestigious conferences.



ABDE MUSAVVIR KHAN completed his B.Sc degree in Computer Science And Engineering from BRAC University, Dhaka, Bangladesh in the year 2020. During his undergraduate studies, he was particularly interested in programming, neural networks and artificial intelligence. Currently, he works as a full stack developer and is interested to work in a professional environment in machine learning, neural network fields.

MYESHA FARID SHEJUTY completed her B.Sc degree in Computer Science And Engineering from BRAC University, Dhaka, Bangladesh in 2020. Her research interests include Deep Learning, Artificial Intelligence, Web Development and Business Intelligence Analytics.

SYED IBNA ZUBAYEAR received a B.Sc degree in Computer Science And Engineering from BRAC University, Dhaka, Bangladesh in the year 2020. He is currently working as a software developer at RedDot Digital Limited. His research interests include Deep Learning, Reinforcement learning, Cybersecurity, Blockchain technology, and Back end development.



MD. NAFIS SHARIAR TALUKDER completed his B.Sc degree in Computer Science And Engineering from BRAC University, Dhaka, Bangladesh in the year 2020. While studying he excelled in Computer Networking and Software Engineering. His research interests include Artificial Intelligence, Software Testing, and Web Development.