Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

N-NET: POBI-LSTM: ENHANCED U-NET ARCHITECTURE WITH DUAL ENCODER MODEL FOR CT IMAGE SEGMENTATION WITH PARAMETER OPTIMIZED BI-LSTM

SATHYA SUNDARAM M [1], KARTHICK S [2], THIYAGARAJAN P [3]

[1] Department of Computational Intelligence, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur-603203, Tamil Nadu, India
[2] Department of Computational Intelligence, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur-603203, Tamil Nadu, India
[3]Department of Computer Science and Engineering, Paavai Engineering College, Pachal-637018, Tamil Nadu, India

E-mail: [1]sm0432@srmist.edu.in, [2]karthiks2@srmist.edu.in, [3]tvm210@gmail.com

ABSTRACT

Intracranial hemorrhage (ICH) is a condition where blood flows into the brain due to trauma or medical disorders, requiring rapid medical and surgiacal care due to its high mortality rate and potential life-threatening consequences. Accurate and automated organ segmentation methods are crucial for diagnosis and treatment planning. Previous research on ICH segmentation using U-Net models has shown inadequate feature transfer rates, high model size, and slower speed. This research proposes a parameter-optimized version of N-Net, called N-NET: POBi-LSTM, to diagnose ICH images of the human brain. The model uses Bi-Directional Long Short-Term Memory (Bi-LSTM) for classification, with hyperparameters optimized using Fuzzy Elite Opposition Social Spider Optimisation (FEOSSO). The research emphasizes the importance of assisting the encoder in extracting more detailed characteristics and proposes an N-Net for enhanced segmentation of CT images. The dual encoder model is proposed to increase the depth of the U-Net network and improve feature extraction capabilities. The method incorporates the Squeeze-and-Excitation (SE) module for channel-level global characteristics and full-scale skip links for hybrid of higher and lower-level details. The model is compared to UNet, UNet++, and Dense-Net201_U-Net in terms of quantitative assessment using DSC, IoU, Accuracy, and loss evaluation.

Keywords: *U-Net, FEOSSO, Bi-LSTM, CT image, Segmentation*

1. INTRODUCTION

Intracerebral hemorrhage, often known as ICH, is a disorder that occurs when blood vessels burst unexpectedly for causes other than external damage, resulting in bleeding in the ventricle region of the brain [1]. Ischemic stroke is the most common kind of stroke, also its frequently happens in individuals who are in their middle years. subarachnoid hemorrhage (SAH), intraparenchymal hemorrhage (IPH), Epidural hemorrhage (EDH), subdural hemorrhage (SDH), intraventricular hemorrhage (IVH), and are the five forms of intracranial hemorrhage (ICH) that may be classified according to the location of the bleeding inside the brain. As a result of the fact that ICH has evolved into a disease that poses a danger to the patient's life and places a burden on family members of those who are afflicted with the condition, it is of the utmost importance to create methods of diagnosis and treatment that are both precise and fast.

Intracerebral hemorrhage (ICH) is the term used to describe bleeding resulting from intracranial vascular rupture, one of the most prevalent disorders in the world. ICH is very harmful to human health and is often brought on by blood disorders, trauma, hypertension, and other conditions. While patients with severe bleeding and high intracranial pressure are more likely to have erratic breathing, respiratory failure, and even death, individuals with intracranial hemorrhage may experience disturbances in consciousness and may go into a coma. Now is the ideal time to take action and treat bleeding before it becomes worse. Consequently, early and precise ICH area segmentation may aid medical professionals in developing a more thorough and

quantitative understanding of the intracerebral hemorrhage region's condition and in formulating an optimal treatment strategy.

It is possible to accurately determine the position of a hematoma, the quantity of bleeding, the tumor's consequence, whether ventricle region consists of bleeding, also the damage volume by subarachnoid and other brain tissues through the use of a computed tomography (CT) images, which both quick and has a high-resolution detail for processing. Because of this, it is regarded to be the best option for the analysis and treatment of ICH detection [2]. As a general rule, medical professionals will first use CT scans to establish the existence of hemorrhage, and then they will proceed to determine the type and place of the bleeding. On the other hand, a diagnosis of this kind requires an extended period of time from an expert for the analysis, particularly when there is a potential that the diagnosis may be missed.

Segmentation has higher clinical relevance than classification in an ICH image analysis because it properly identifies the bleeding location quantity in the first phase of recognizing the presence of bleeding. With the exception of ICH images, segmentation methods are also employed to assess medical diagnostic images. The most popular segmentation models are those with a modified encoder-decoder architecture built on the basis of U-Net.

The process of identifying affected region by applying medical diagnostic technologies such as computed tomography (CT) or magnetic resonance imaging (MRI) is stated to as medical image segmentation. There were two issues that continued to exist in the ICH image segmentation using deep learning (hence referred to as "ICH segmentation") approaches that were already in existence and used the U-with encoder-decoder architecture achieved satisfactory outcomes. In the first issue, since there is a great deal of parameters, these networks need a significant amount of time for inference and training. The time required for inference increases when the input images have a high resolution. Secondly, whenever
low-resolution attributes taken from an encoder unit are transformed to high-attribute in the decoder, which leads to a considerable loss of sensitivity to final segmentation sensitivity. This is due to the decoder is responsible for transforming the features. When it comes to diagnosing ICH, the rapidity of the diagnostic model is very important, in addition to its performance, since it is required that the condition be clearly identified and treated within one hour of its beginning. As a result, the purpose of this research is to propose an N-NET: POBi-LSTM ICH segmentation networks as a means of overcoming such limitations in order to conduct an efficient finding and classification of ICH. An enhanced N-Net with encoder-decoder unit is segmentation parts of the N-NET: POBi-LSTM, which has the following abilities and characteristics: In order to get channel-level global features, our technique includes the Squeeze-and-Excitation (SE) modules [3] within the dual encoder architecture. This gives us the ability to acquire these characteristics. Furthermore, the inclusion of full-scale skipping links makes it possible to concurrently include high-level attribute data with lower-level attributes of the images.

The accuracy of diagnosis of the Improved N-Net model in emergency neuroradiology has not been evaluated in any studies that have been conducted by researchers. The main aim of this study is to find intracranial hemorrhage (ICH) from the following subtypes of images: intraparenchymal hemorrhage (IPH)/intraventricular hemorrhage (IVH),epidural hemorrhage (EDH)/subdural hemorrhage (SDH), subarachnoid hemorrhage (SAH), and. For this reason, the study aims to detect ICH from various region of brain. In addition, experiments were conducted using U-Net to evaluate the diagnostic efficiency of the proposed technique in comparison to that of other approaches.

2. RELATED WORKS

Although segmentation successfully identifies region of bleeding in the first stage of recognizing the existence of bleeding in an ICH image analysis, it is more clinically applicable than classification. This is because segmentation is able to accurately recognize the amount of bleeding. Additionally, segmentation methods are used in the analysis of medical diagnostic images, with the exception of the ICH. Within the context of classification, the Gaussian Mixture Models (GMM) represent a vital component. From the color of each pixel, it distinguishes the region of background image from the region in the foreground [20]. Then, following that, Deep learning is used for data classification [25]. Here, LSTM based data classification is carried out with maximum efficiency. Convolutional Neural Network is widely used in image processing for classification [31]. Most of the segmentation models employ a U-Net-with encoder-decoder architecture, which is the most used one in segmentation process. The segmentation model known as U-Net [4], which was published in 2015, makes use of a symmetric of encoder and decoder unit along with a skip connection. It demonstrates exceptional act in medical images

segmentation through converging multiscale attributes. Because of their better outcomes in the examination of clinical images, other network using U-Net structure, like U-Net++ [5], 3D U-Net [6], and Attention U-Net [7], was used by a large number of users. Additionally, models with an encoder-decoder architecture, such as SegNet [6] and PSPNet [7], have also been extensively deployed since their introduction. In their paper [8], Zhang et al. introduced a method with generator net to build an ICH image. This image has been synthesized together with typical ICH image that had an inadequate quantity of trained data by utlizing a U-Net-based network. Based on this finding, it was determined that the performance might be enhanced by concurrently training the ICH model for both the synthetic information and the real information. The use of U-Net in synthesis of clinical image in segmentation, was a novel instance in this specific case. After conducting research, Kushnure and Talbar [9] were able to properly extract the global data and local data attributes from CT images. This was accomplished through upgrading U-Net structure by Res2Net [10] also integrating it with a SE network. The 3D U-Net based SE structure has been employed to the U-Net in the segmentation model that was suggested by Abramova et al. [11].

A 3D Dissimilar-Siamese-U-Net had been proposed by You et al. [12], which consisted of two U-Nets that were linked to the encoder through a distance unit. Both left and right inputs were received by the 3D Dissimilar-Siamese-U-Net, which was then used to do analysis on the brain CT images. A review was carried out by Mizusawa ei al. [13], in which U-Net has been used for the purpose of reconstructing an image of X-ray. The field of natural language processing (NLP) has made extensive use of recurrent neural networks (RNN), which are a model created to the purpose of analyzing sequence data. As a result of the fact that CT or MRI images are created as linking slices within clinical image studies, RNNs are able to be used for the analysis of these images. The 3D PyraMiDLSTM model was used in the research that was carried out by Stollenga ei al. [14] for the purpose of segmenting MRI brain images. Due to the fact that the N-net was built to support parallel processing, the outcome of training process was greatly improved, thus resulted in strong segmentation outputs in the MRBrainS challenge [15]. Furthermore, In [16], they developed a spatial clockwork recurrent neural network (CW-RNN) for the purpose of segmenting images of muscle illness. This network used fewer parameters than RNNs, allowing for more accurate results. As a consequence of this, the CW-RNN was having the accuracy as 5% greater than the U-Net, also speed of execution was one hundred times faster than the Convolutional Neural Network model. In Poudel et al. [17], they developed recurrent fully convolutional networks for the purpose of real-time computation of heart segmentation. These networks were built on FCN and RNN data structures. Using TransUNet, which was created by merging U-Net with 12 transformer layers in Chen et al. [18], which were able to conduct CT image segmentation and acquire remarkable results. An MRI brain tumor segmentation model in three dimensions was constructed by Wang et al. [19] using a transformer structure that was developed by the basic concept of U-Net. Analysis of sequence data by transformer based merging a CNN model with an encoding and decoding unit is developed, [18] and [22] shown that the outcome of the segmentation algorithm has been enhanced. This was demonstrated by the incorporation of the transformer. On the other hand, In [19], they used a three-dimensional CNN layer that had a large number of constraints and displayed a slower time when previous architecture of U-Net was employed.

In Vamsi et al. [28], they introduced a lightweight convolution model that made use of a Random Forest method with a VGG-16 architecture. They were able to achieve a better DSC & accuracy by using this architecture. In [29], they recommended a semi-supervised system for ICH segmentation by proposing an inverse sigmoid-based learning technique. A technique known as White Matter Fuzzy c-Means (WMFCM) was employed in [30] to exclude region like the cranium. Even though the segmentation wasn't based on deep learning, they still managed to reach an average DSC score by using their model. Different existing models have been used in a number of studies that have been carried out on the subject of brain tumor segmentation utilizing magnetic resonance imaging (MRI), in addition to CT image segmentation.

UNet++ [35] further enhances conventional skip connections through incorporating nested with dense skip connections. This is done to reduce the fusion of semantically distinct attributes that occurs as a result of the ordinary skip connections in UNet. Additionally, UNet ++ has included a depth monitoring structure, which allows dense network topologies to be reduced. This results in the model having a greater degree of flexibility. Additionally, in order to compensate for the data that was lost, UNet3 + [36] rebuilt the interconnections that were among the encoder and the decoder. Full-scale in-depth training is furthermore implemented in

UNet3 + for the purpose to learn hierarchical representation. Results of segmentation were obtained using UNet3+ that were accurate, organs were highlighted, and coherent borders were established. For the purpose of quantitative testing on two datasets, it is important to point out that it outperforms all of the prior existing approaches, including PSPNet [33], DeepLab version [32-34], UNet + + [35], and Attention UNet [7]. Additionally, despite the fact that UNet3 + has attained outstanding performance, there remains a great deal of space for optimization.

The process of detecting disease-affected regions utilizing medical diagnostic technologies, such as computed tomography (CT) or magnetic resonance imaging (MRI), is known as medical image segmentation. Two issues remained even if the U-shaped encoder-decoder architecture-based deep learning-based ICH image segmentation techniques currently in use produced satisfactory results. First, since these networks have a lot of parameters, inference and training take a long time. When high resolution input images are used, the inference time increases significantly. Second, there is an important reduction of sensitivity to the final segmentation sensitivity when low-resolution features retrieved from the encoder are converted into high-resolution features in the decoder.

To overcome the above-mentioned limitations, we offer N-Net with the intention of achieving more precise medical segmentation images. This study makes a number of important contributions, which are summarized in the following order:

- We propose the use of N-Net to accomplish precise segmentation of medical images. Additionally, in contrast to UNet++ and UNet3+, we introduce a greater weight on the extraction of features from the encoder unit using dual encoder structure.
- Analysis the importance of the SE module in terms of improving the efficiency of the segmented network, and we include the enhanced SE module [37] within the dual encoder model.
- Additionally, a Bi-LSTM-based classification technique is carried out, which improves overall performance by optimizing the hyperparameters through the use of Fuzzy Elite Opposition Social Spider Optimization (FEOSSO), which will enhance the accuracy of image classification.
- We carry out a large number of tests using CT datasets, and the results reveal that N-Net consistently produces improvements across a wide range of baselines for comparison.

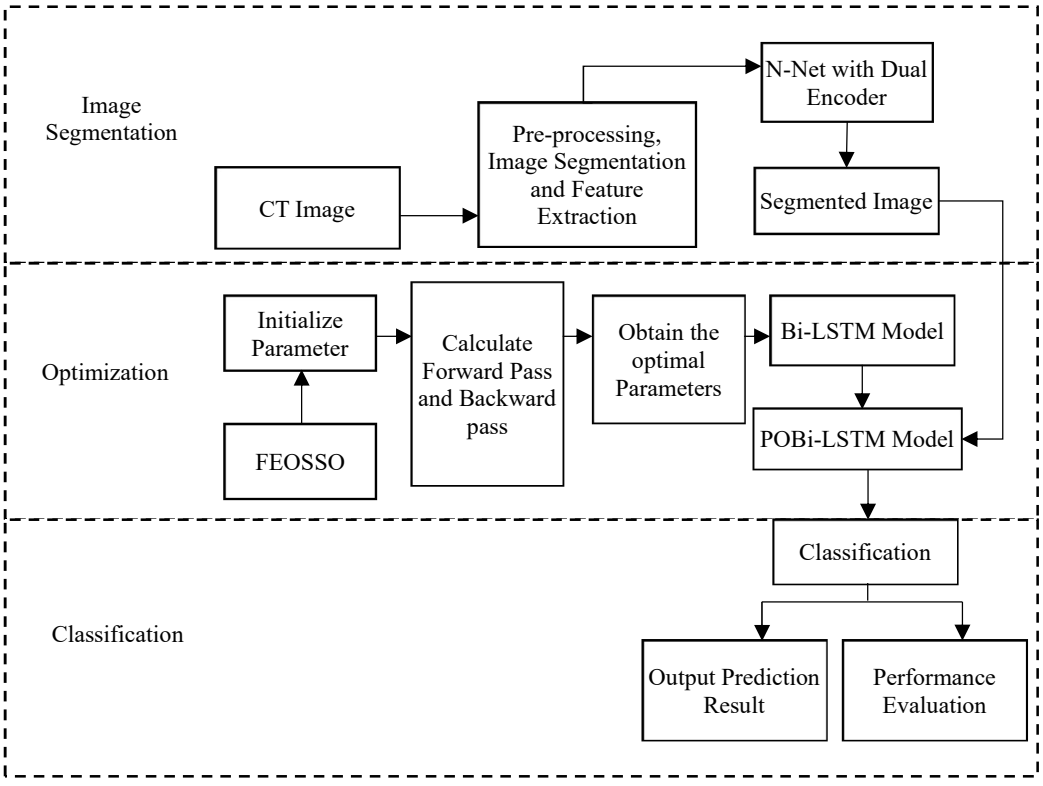

Figure 1: Flow Diagram of N-NET: POBi-LSTM

3. METHODOLOGY

Continuing in the footsteps of U-Net [23] and BiLSTM [24], our objective is to provide the N-NET: POBi-LSTM, which may be noticed in figure 1. Strongly linked convolutions and BiLSTM model are both used by the network in order to ensure that it receives the most benefit from its capabilities. In spite of the fact that the bleeding is only segmented from the images, it is still conceivable to consider this to be instance segmentation. This study used a modified U-Net architecture in order to segment the area of bleeding that was shown in a CT image [3]. This was done due to the U-Net architecture performed better than other deep learning networks due to its superior performance. We are going to go into the different components of the network in further detail below in the subsequent sub sections. There is an illustration of the image segmentation of the proposed work in Algorithm I.

Algorithm I: Training and Classification of Steps

---

1. Input: Load_inputdata

// CT image dataset (EDH, IPH, IVH, SAH, SDH)

2. Pre-processing on CT Images
3. Splitting the data into Train_Test_Validation
4. Define model ()

a. //Initial the model with specified parameter

b. //Optimize the parameter using FEOSSO algorithm

c. Dual Encoder=add the weights

d. model.add(layers (Convolutional Layer, a ReLU activation function and batch normalization))

e. model.fit(X_training,Y_training,Validation_data

f. model.save

5. Training data

// Initialize the hyper parameter (FEOSSO, Batch size, SE module, Hybrid Loss function

6. Testing data

model.evaluate()

//Similarity between test image and ground truth image segmentation is evaluated

---

3.1 N-Net architecture: a brief overview

For the purpose of encoding the high-level semantic contextual data from feature maps, we propose utilizing an extra skip connection based dual encoder architecture. The added skip connection expands the receptive field and employs a bigger convolutional kernel than the regular convolutional layer, all without lowering the feature map's resolution. The deeper network's segmentation performance in the semantic segmentation network increases with the size of the final predicted pixels' receptive field. It should be noted that using added skip connection convolution does not result in an increase in calculations or parameters.

In figure 2, we present our N-Net architecture, which is an improvement on the U-Net design. Additionally, we offer the dual encoder structure in conjunction with the SE model. Because the link between the convolutional layers is formed like the letter N, we decided to call this design N-Net. In the dual encoder, the two parallel routes are joined to one another layer by layer using skip connections that are more conventional. From here on out, we will refer to the convolutional layers from the three branches of N-Net as X^{i}_{DC}, X^{i}_{EN}, and X^{i}_{DE} (where i may be any of the following: $(1, 2, 3, etc.)$. These layers are consists of the two separate dual encoder branches as well as the decoder branch. Both of these parallel convolutional layers are made up of two convolutional units. The convolution layer, the batch normalization layer, and the ReLU activation layer are the three components that make up each convolution unit. During each Xi DC and X^{i}_{EN} $(i = 1, 2, 3\ldots)$, the 2×2 max pooling technique is used to minimize the size of the attribute maps by an amount of 50%. A sequence of bilinear up-sampling procedures is used by the decoder to retrieve the spatial attributes that was lost in the pool layer. Following every upsampling operation, a X^{i}_{DE} $(i = 1, 2, 3\ldots)$, is attached. This X^{i}_{DE} is comprised of 320 filters with a size of 3 x 3, batch normalization, as well as a ReLU activation function. In final, the segmentation map is generated by using a 3×3 convolution layer and a sigmoid activation function. The map size is identical to size of the image.

Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

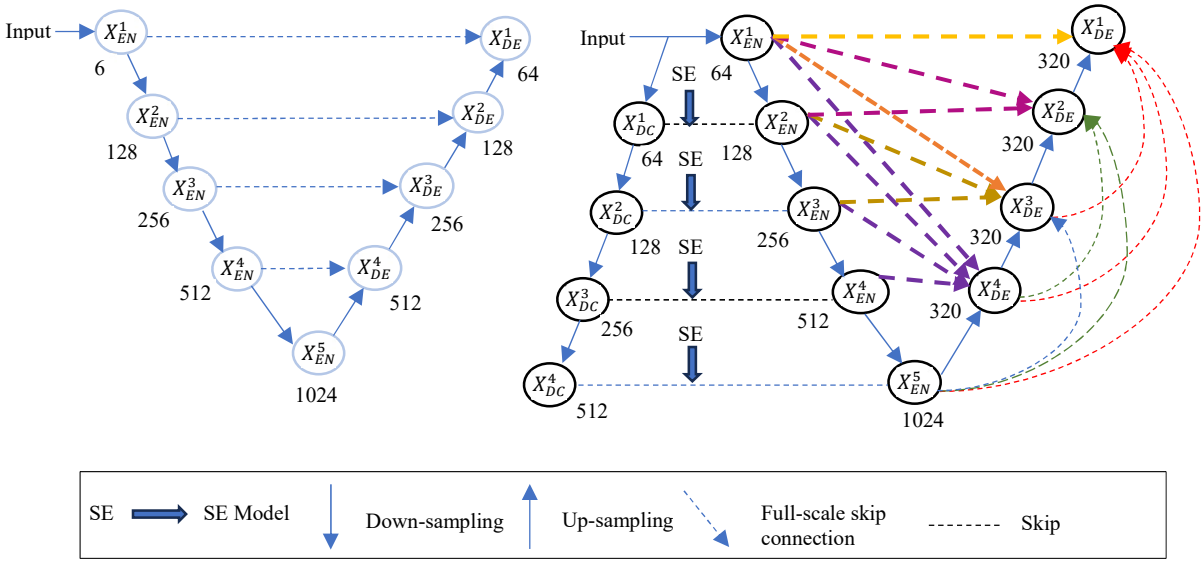

Figure 2: Proposed N-NET ARCHITECTURE

3.2 Dual Encoder

It is unavoidable that some data would be lost through the operation of encoder and decoder. We provide a dual encoder model consist of $X^{i}_{DC}\ (i = 1,2,3,4)$ and $X^{i}_{EN}\ (i = 1,2,3\ldots)$, with the intention of minimizing the amount of information that is lost. As seen in figure 3, the two parallel branches, each consisting of four maximum pooling layers, each have a dimension of two by two. After that, the two parallel branches are able to correspond to the similar resolution of the convolutional layer by use of the max pooling procedure. The $X^{i}_{DC}\ (i = 1,2,3,4)$ uses the dilated convolution for obtaining the attribute data, which is the difference between the $X^{i}_{EN}\ (i = 1,2,3\ldots)$ and the $X^{i}_{DC}\ (i = 1,2,3,4)$. It is necessary to regulate the quantity of channels in order to facilitate the merging of the feature maps of the X^{i}_{DC} and the $X^{i}_{EN}\ (i = 1,2,3\ldots)$ in order to enhance the subsequent layer of the $X^{i+1}_{EN}\ (i = 1,2,3\ldots)$ as input.

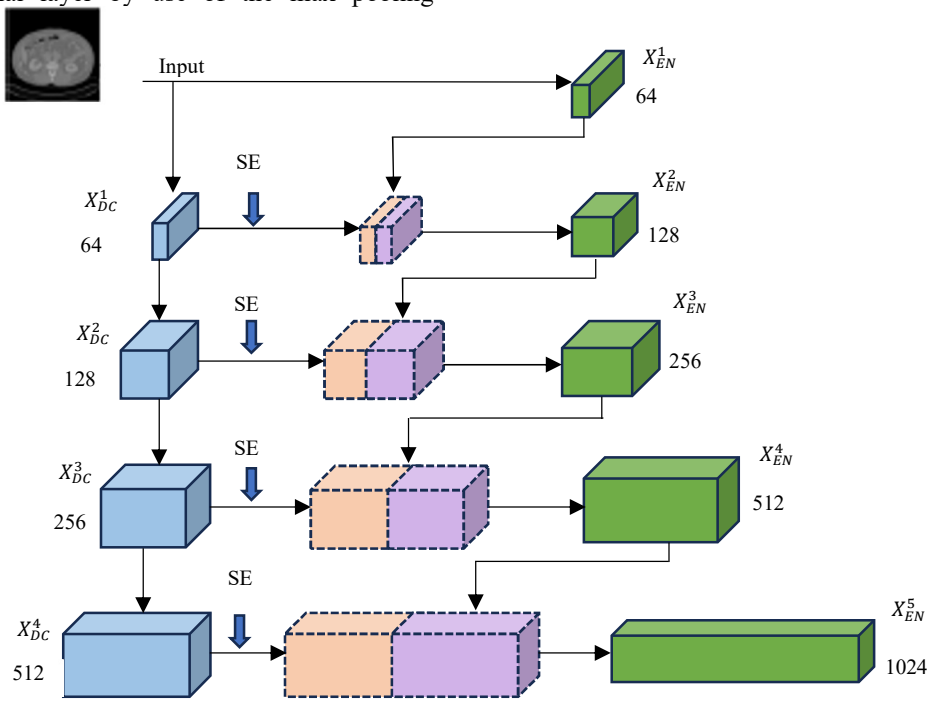

Figure 3: Proposed Dual Encoder Architecture

In addition to increasing the depth of the network, the dual encoder model also integrates all of the information. In order to compensate for the information that is lost as a result of max pooling operations, the encoder with dual branch and full-scale connections is used. These models incorporate full-scale data in order to get both fine-detailed features and coarse-detailed features on a full-scale basis. Through the use of several convolutions, the proposed encoder model is able to get multiscale attribute information, thus enriching semantic data. Obtaining a greater receptive field is possible via the use of dilated convolution. When applied to the similar feature mapping and the dilated convolution algorithm has the potential to enhance the performance of the task in terms of the identification and segmentation of small elements.

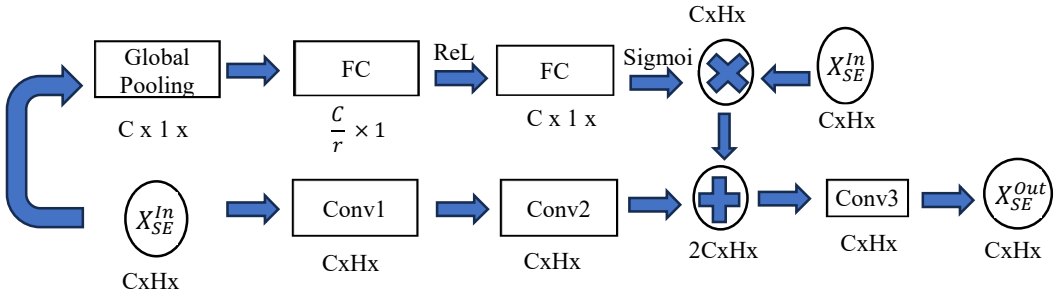

Figure 4. Process of SE Model for Attribute Extraction

To improve the capacity of N-Net attribute mining process, SE model in the skip connections of the proposed encoder model is implemented, which can be seen in figure 4. On the basis of the side outputs that are produced by each $X_{DC}^{i}(i = 1,2,3,4)$ the SE module is able to acquire knowledge on the significance of various channel characteristics. The first phase in the SE model process is the global average pooling (GAP) of the attributes with size of $H \times W$ for each channel. This is done in order to get a scalar, which is referred to as Squeeze. Given the input feature mappings F, which belong to the set $R^{C\times H\times W}$, the GAP algorithm creates a feature vector Z, which belongs to the set $R^{C\times 1\times 1}$. A formal definition of GAP can be expressed in equation (1):

$$Z_n = \frac{1}{H \times W} \sum_{x=0}^{H} \sum_{y=0}^{W} F_n(x,y) \qquad (1)$$

Where F_n is the nth channel feature map of $F, n \in \{1,2\ldots,C\}$ and (x,y) is a pixel in F_n.

We make use of two layers that are completely interconnected. The first completely linked layer brings the total channel countdown to C/r, here r is the scaling factor that is decided through the outcomes from subsequent tests. Additionally, the number of channels is returned to C when the second layer that is entirely joined together. In order to ensure that numerous channels may be emphasized, the ReLU algorithm is used after the first completely linked layer. Following the second layer, the sigmoid function is executed in order to produce the non-linear interactions that exist between the various channels. We assign a weight value between 0 and 1 to each of the two FC. In the end, by the process of multiplying every element of H × W attributes by the weight of the associated channel, the new feature maps referred as excitation are created. For the purpose of recalibration of channels of input features, the weight value ω is determined as in equation (2):

$$\omega = \sigma\left(W_2\delta\left(W_1\rho(X_{SE}^{In})\right)\right) \qquad (2)$$

The sigmoid function is represented by the symbol $\sigma(\cdot)$, whereas the ReLU activation function is represented by the symbol $\delta(\cdot)$. $\rho(\cdot)$ is a symbol that represents the global average pooling layer. In order to derive the "excitation" weight ω, a fully connected (FC) layer are parameterized by $W1 \in R^{\frac{C}{r}\times C}$ and a dimensionality-increasing FC layer that is parameterized by $W1 \in R^{C\times\frac{C}{r}}$ are used.

An entire SE block consists of a transformation that maps the input feature map $X_{SE}^{In} \in R^{C\times H\times W}$ into the output feature map $X_{SE}^{Out} \in R^{C\times H\times W}$. The expression of X_{SE}^{Out} is defined in equation (3).

$$X_{SE}^{Out} = H\left(F_{scale}(\omega, X_{SE}^{In})\right) + H_1(X_{SE}^{In}) \qquad (3)$$

In this case, the transformation $H1(\cdot)$ is used to represent two convolution processes, namely batch normalization as well as ReLU activation. Additionally, the transformation $H(\cdot)$ is used to represent a convolution operation that changes the

dimension in a similar manner. Through the channel-wise multiplication of the weight ω with the input feature map X_{SE}^{In}, the term "$F_{scale}(\omega, X_{SE}^{In})$" is associated with the process.

With the use of this attention mechanism, the model is able to prioritize the channel characteristics that contain the highest amount of information, while simultaneously suppressing the aspects that are not of significant importance. It is possible to see in each channel learning process of the weight coefficients, which ultimately results in the model being more discriminative with respect to the attributes specific to each channel. It is possible that this will assist us in obtaining more precise information on the position and boundaries of the organ.

3.3 Hybrid Loss Function

When attempting to measure the disparity between the estimated state of N- network and ground truth data, the loss function is used. There exists a particular connection between the loss function and the level of accuracy of the training model.

The following is a proposal that we have made to combine the two loss functions to produce a more accurate segmentation result, train our tasks easier, and retain the best aspects of both loss functions:

$$loss_{final} = \beta * loss_{IoU} + (1 - \beta) * loss_{FL}$$

The parameter in the equation is denoted by β. In accordance with a significant number of tests, we consistently make adjustments to the β in order to compare the outcomes of the tests, and we experimentally begin β to be equal to 0.15.

$loss_{FL}$ and supporting elements (ρ_t, α_t) are given in equation (4), (5) and (6)

$$loss_{FL}(\rho_t) = -\alpha_t(1 - \rho_t)^{\gamma} \log(\rho_t) \quad (4)$$

$$\rho_t \begin{cases} p & if\ y = 1 \\ 1 - p & Otherwise \end{cases} \quad (5)$$

$$\alpha_t \begin{cases} \alpha & if\ y = 1 \\ 1 - \alpha & Otherwise \end{cases} \quad (6)$$

Where p is the estimated probability of the model for a class with the label $y = 1, \alpha$ is a weighting factor that belongs to the interval [0, 1], and γ is a variable focusing parameter that is greater than or equal to zero. The regulating factor for focal loss is found to be $(1 - \rho_t)^{\gamma}$. The objective is to decrease the weight of the data that are readily categorized so that the model may concentrate more on the examples that are much more difficult to classify while it is being trained. According to reference [21], the parameters of focal loss are established as $\gamma = 2$ and $\alpha = 0.25$.

An innovative soft IoU loss is presented by Luo et al. [22] in order to get a more precise representation of road topology in aerial images. Back propagation is possible due to the fact that this loss function is differentiable. Also, it is well-defined by substituting the indicator functions with the softmax outputs, and This loss function could be described in the in equation (7):

$$L_{IoU} = 1 - \frac{1}{2}\left(\frac{\sum_i pi1 * p^*i1}{\sum_i pi1 + p^*i1 - pi1 * p^*i1} + \frac{\sum_i pi0 * p^*i0}{\sum_i pi0 + p^*i0 - pi0 * p^*i0}\right) \quad (7)$$

pix is the prediction score at position i for class x, whereas p^*ix represents the ground truth distribution. The ground truth distribution is a delta function at y_i^*, which represents the right label value.

3.4 PO-BiLSTM Model Establishment

There is a significant influence on the results depending on the parameters that are chosen for both the long-term and short-term memory neural network models. For the purpose of selecting the parameters of the LSTM, Fuzzy Elite Opposition Social Spider Optimization (FEOSSO) is presented, and the PO-BiLSTM traffic flow prediction model is developed. The initial parameter setting may have a significant influence on the prediction outcomes; however, improved settings may reduce that impact. FEOSSO includes the number of hidden units, training durations, gradient threshold, and learning rate that are all taken into consideration. To create the PO-BiLSTM model, the weight of the structure is adaptively altered via the process of the LSTM iteration procedure. The segmentation characteristics are then entered into the PO-BiLSTM model, and the resultant value is the ICH STROKE prediction value. This occurs after the best parameter combination has been found. The training data set is fed into the model for operation via a series of multiple tests. Algorithm II illustrates the proposed model implementation flow.

Algorithm II : Proposed N-NET: POBi-LSTM

Parameters: input x(512,1), output y_t(512,5)

1: **For** each epoch **do**:

#N − Net segmentation

2: **For** each convolutional layer **do**:

3: **for** each sample in X **do**:

4: Calculate a_{ij}^{DC} from X_{DC}^{i} by the convolutional process

5: Calculate a_{ij}^{EN} from X_{EN}^{i} by the convolutional process

6: Calculate a_{ij}^{DE} from X_{DE}^{i} by the convolutional process

7: **End for**

Dimention appending

8: **IF** a_{ij}^{DC} and a_{ij}^{DE} and a_{ij}^{DE} length < 512 **do**:

9: ReLU activation function and a batch normalization and , is added to a_{ij}^{DC}, a_{ij}^{EN}, a_{ij}^{DE}

10: # Dimension of a is (512, Sigmoid activation function)

11: **End if**

12: **End For**

#Classification PO − BiLSTM

13: **For** each sample of a **do**:

14: Initilize FEOSSO parameter

15: Calculate Forward Pass of a

16: Calculate Backward pass of a

17: Calculate optimum result Y_t

18: **End for**

19: **End for**

The Process of Optimizing BiLSTM with FEOSSO

When working with the BiLSTM network, the choice of parameters within its structure significantly impacts the model's performance. This includes factors like the number of hidden layers, weights, number of hidden layer components, and learning rate. Researchers often establish these parameters through trial and error or based on their experience, leading to potential unreliability in the model's robustness and accuracy. Hence, this paper chooses for the FEOSSO algorithm due to its simplicity in principle, low complexity, fast convergence, and suitability for addressing real value problems in optimizing the structural variables of the BiLSTM network.

The core principle of FEOSSO is to determine the optimal solution for each particle simultaneously and then pass that information on to the next epochs. A best parameter is considered to be the greatest when it comes to fitness. As part of the fuzzy method, the SSO parameters served as inputs to the inference system. For the purpose of generating the best parameter for the BiLSTM Classification process, a fuzzy inference engine makes approximations between these indices using a rule basis. The approach increases population variety and broadens the algorithm's search field during global optimization. Therefore, it is possible to improve the global searching capability and get a

more ideal solution. Figure 5 provides a comprehensive overview of the computing process.

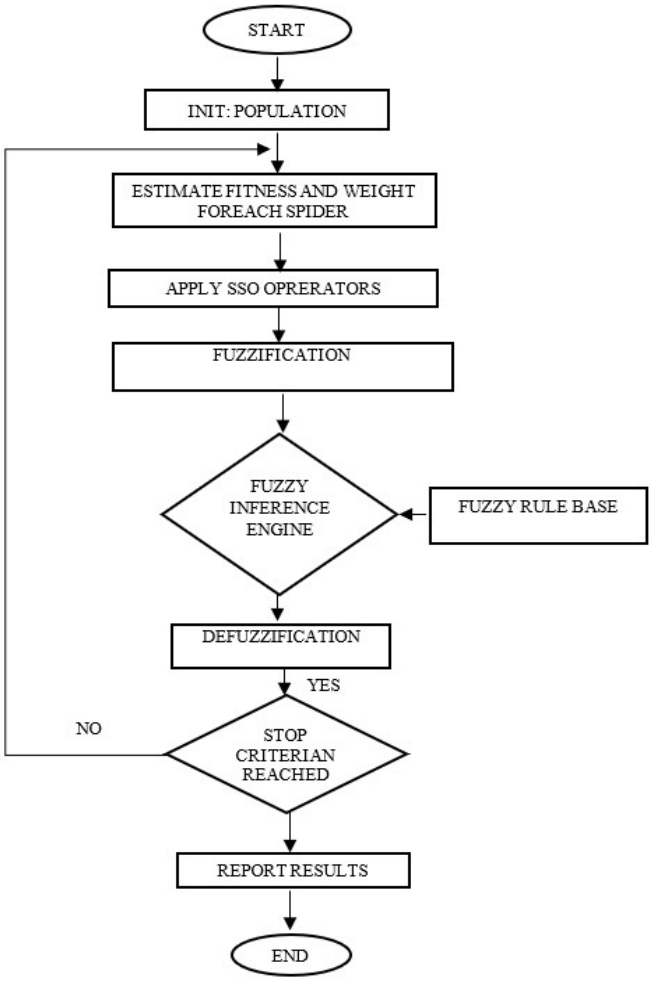

Figure 5: Flow Diagram for FEOSSO

4. RESULT & DISCUSSIONS

During the process of this research, ICH segmentation from input CT images was done by using an innovative architecture. The segmentation algorithm of U-Net and U-Net++ were utlized in order to assess the effectiveness of the techniques. On the basis of the DSC score, the IoU, the accuracy, and the loss coefficient, these models have been evaluated. In the next subsection, the outcomes of various approaches are discussed. A description of the experimental setting for the implementation is provided in table 1.

Table 1: Experimental Setup

Device	Specifications
OS	Windows 11
CPU	Interl Core I7
GPU	NVIDIA Ge Force RTX 2080T
RAM(Memory)	18GB
Storage	I TB SSD + 4TB HDD
Language	Matlab 2020a

Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.1 Dataset

A total of 2000 CT scan images of intracranial hypertension were employed as datasets in this research. These datasets were obtained from St. Mary's Hospital of University of Korea Seoul. On the basis of the data, the medical professionals manually located the region affected by the illness and showed the ground truth. Based on the bleeding position, each image was classified as either EDH, IVH, IPH, SDH or SAH [38]. The resolution of all the images was 512 ± 512, and they were very high. Sample input CT images from each category are shown in figure 6.a–f. These categories include epidural hemorrhage (EDH), intraventricular hemorrhage (IVH), intraparenchymal hemorrhage (IPH), subdural hemorrhage (SDH), subarachnoid hemorrhage (SAH) and at least one form of bleeding (multiple), respectively. The test data set included of 600 images, 100 of which were ICHs chosen from several categories such as EDH, SDH, IPH, SAH, IVH, and multiple. For the purpose of forming training and validation datasets, the remaining 1400 images were split in a ratio of 7:3 according to the disease classification. Figure 7. shows that the segmented image of our proposed model. Table 2 describes the more details about the dataset.

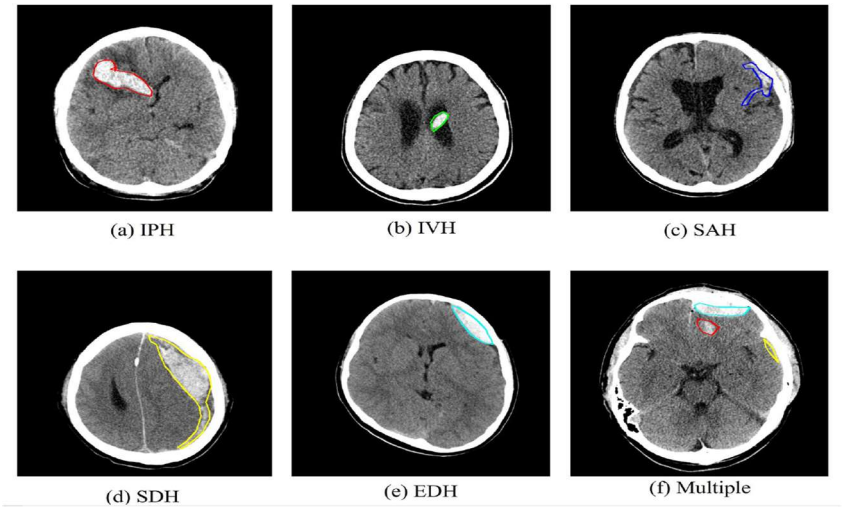

Figure 6: Various Categories of an Input images

Table 2: CT Dataset Details

	EDH	IVH	SDH	SAH	IPH	Multiple	Sum
Train	2286	5352	27413	13421	17605	15359	81436
Test	100	100	100	100	100	100	600

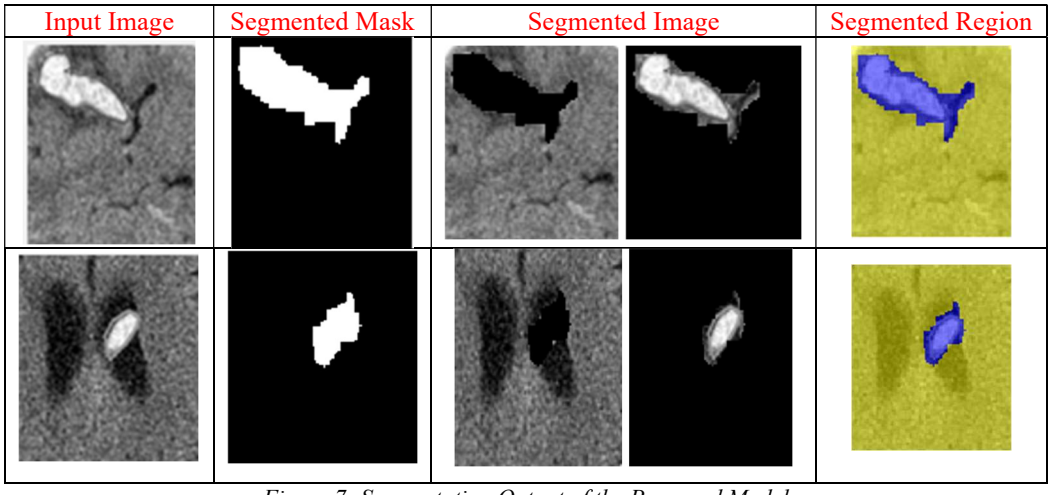

Figure 7: Segmentation Output of the Proposed Model

4.2 Evaluation Metrics

In this research, a number of factors, including the Dice similarity coefficient (DSC), intersection over union (IoU) and accuracy, have been examined. These parameters influence the degree to which the model is able to make accurate predictions. The accuracy, DSC, and IoU may be presented in the following manner:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Intersection\ over\ Union(IoU) = \frac{TP}{TP + FP + FN}$$

$$Dice\ Similarity\ Coefficient(DSC) = \frac{2TP}{2TP + FP + FN}$$

TP, TN, FP, and FN are abbreviations that stand for true positive, true negative, false positive, and false negative, respectively. TP represents the number of pixels that have a label of 1 and a predicted value of 1, TN represents the number of pixels that have a label of 0 and a projected value of 0, FP represents the number of pixels that have a label of 0 and a forecasted value of 1, and FN represents the number of pixels that have a label of 1 and a predicted value of 0. When it comes to these measures, the segmentation impact is improved when the value is higher.

4.3 Performance Analysis of SEGMENTATION Process

Segmenting the disease region for ICH was the basis for the detection techniques that were carried out so successfully. This method was used in order to find out the affected area of the images that included the bleeding by using 2D segmentation pipelines for a number of pre-trained structures [26].

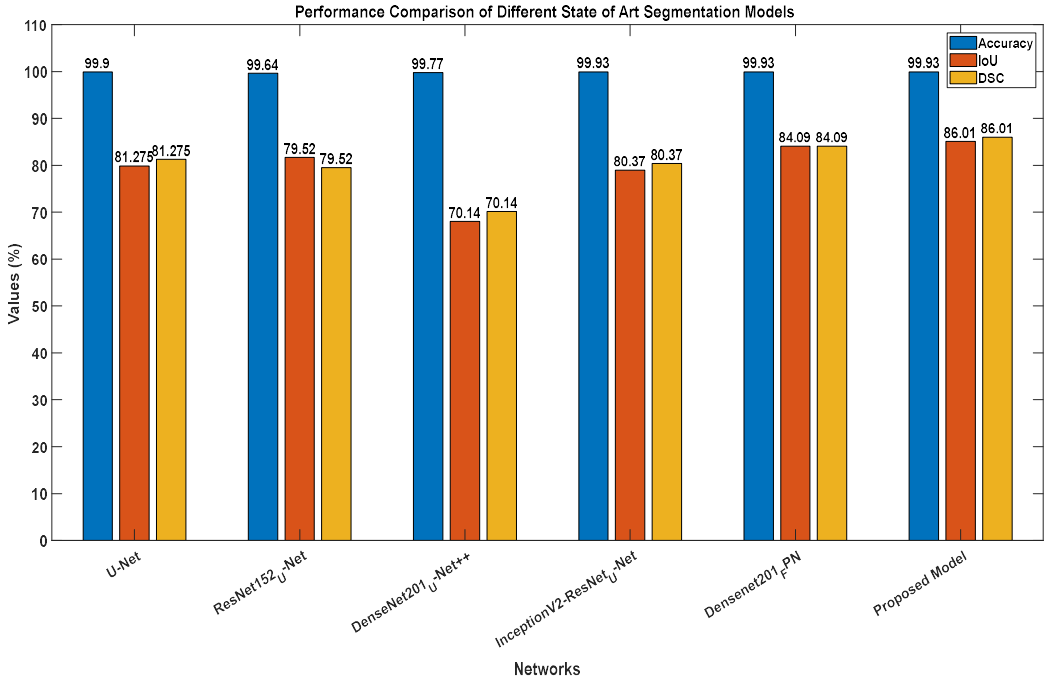

Figure 8: Analysis of Various Network Models

The performance matrices of several segmentation models are shown in table 3 and figure 8. These matrices include loss, accuracy, DSC, and IoU measurement. The N-NET: POBi-LSTM model was used in order to attain the maximum possible DSC score, which was 86.071%. It is also important to note that the IoU of 85.1% for N-NET: POBi-LSTM was the highest among the other models that were evaluated in this research. In the process of bleeding segmentation, this predicted efficiency of the N-NET: POBi-LSTM model was identified. Table 3 indicates that N-Net has a comparable level of accuracy to our proposed model. However, the proposed model's higher feature extraction section will bring the loss value down to a reasonable amount.

Table 3: Performance Comparison of Various Segmentation Models

Various U-Net Model	Accuracy (%)	IoU (%)	Loss	DSC (%)

U-Net	99.9	79.87	0.733385	81.275
ResNet152_U-Net	99.64	81.69	0.0321	79.52
DenseNet201_U-Net++	99.77	68.01	0.7036	70.14
InceptionV2-ResNet_U-Net	99.93	78.99	0.0029	80.37
Densenet201_FPN	99.93	84.09	0.8018	84.09
Proposed Model	99.93	85.10	0.0008	86.01

4.4 Evaluation of the SE model by varying "r" Value

In accordance with the information presented in Section III, hyperparameter of *r* is facilitates the modification of the capacity and computing cost of the SE module. It is essential that this parameter be adjusted with careful consideration. Experimenting with a wide variety of various values of r allows us to study the trade-off that exists between the performance that is controlled by r and the amount of computational load.

As can be seen in table 4 and figure 9, the extrusion of features increases in proportion to the amount of r, and the number of parameters (Params) required for N-Net decreases as the value of r increases. It is important that the values of F1, Dice, Precision and Recall, reach their highest possible levels when r equals 8. Nevertheless, there was a 0.05 M increase in the total number of parameters. To make a balance between performance and computational weight, we decided to make r equal to 8.

Table 4: Performance Evaluation by Changing "r" Values

r	Precision	Dice	Recall	F1	Params(M)	Mean ± std
2	0.9788	0.9354	0.9102	0.9451	41.32	0.8895 ± 0.0547
4	0.9791	0.9387	0.9154	0.9462	41.48	0.8917 ± 0.0654
8	**0.9798**	**0.9467**	**0.9184**	**0.9476**	**42.04**	**0.8957 ± 0.1256**

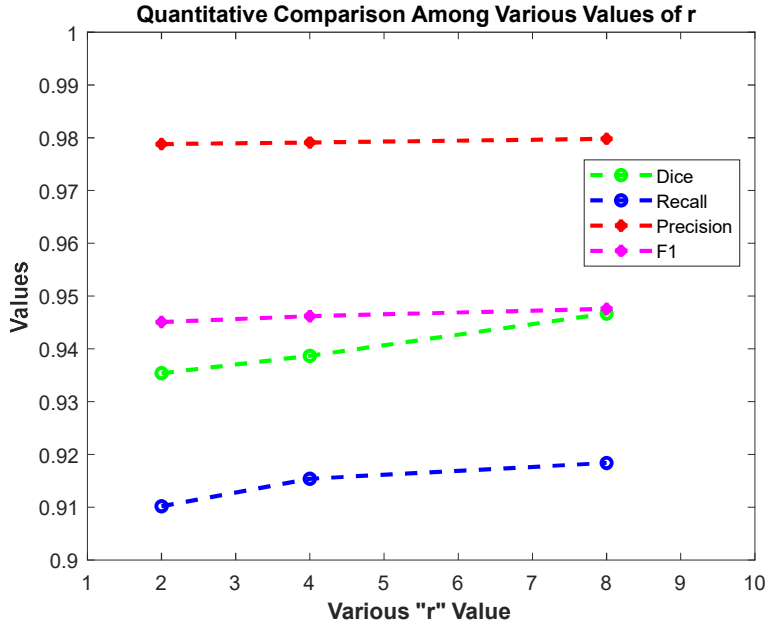

Figure 9: Performance Analysis of Proposed Model for Various Values of "r"

4.5 Comparison of proposed model Vs Existing Models

In order to determine whether or not the suggested approach is effective, we carried out a comparative evaluation by assessing the segmentation outcomes of the N-NET: POBi-LSTM model that was proposed in this work. In addition, we measured the performance of standard segmentation techniques such as U-Net[4], U-Net++[5], SegNet[15], PSPNet[16], and HarDNet[27]. Additionally, in order to guarantee uniformity throughout the training procedure, we

used completely similar hyper-parameters and the DiceCE loss function. A listing of the findings of the experiment can be found in table 5 and figure 10.

The proposed N-NET: POBi-LSTM showed better results with Dice coefficients, IoU, of 0.784 and 0.614, respectively in comparison to the four standard semantic approaches and demonstrated outstanding results. In addition, the N-NET: POBi-LSTM, which included the incorporation of a SE into the N-Net, resulted in a 1.7% increase in the accuracy of the model when compared to the U-Net alone by the use of a simple convolution computation.

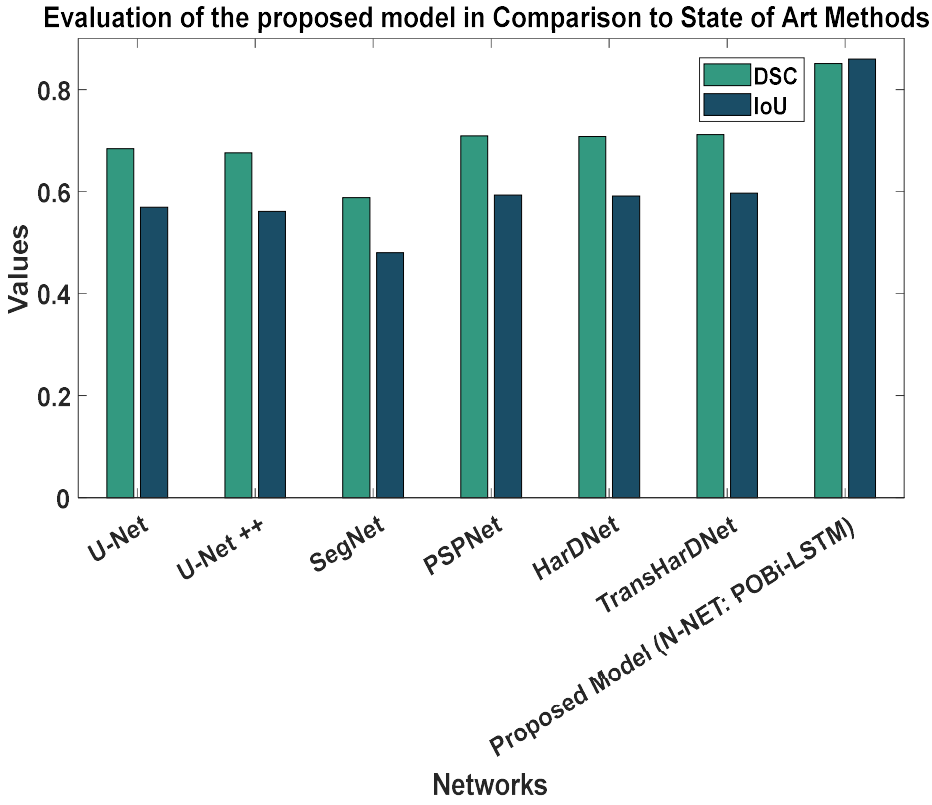

Figure 10: Performance Comparison with Various U-Net Models

Table 5: Comparative Analysis of Various Model

Model	Dice	IoU
U-Net [4]	0.684	0.569
U-Net ++ [5]	0.676	0.561
SegNet[15]	0.588	0.480
PSPNet[16]	0.709	0.593
HarDNet[27]	0.708	0.591
TransHarDNet[39]	0.712	0.597
Swin-HGI-SAM	0.403±0.014	0.283±0.011
Fully supervised U-Net	0.388±0.019	0.297±0.016
Fully supervised Swin-UNETR	0.428±0.018	0.330±0.011
YOLO-Clustering	0.625±0.020	0.506±0.019
YOLO-SAM-BBox	0.562±0.020	0.445±0.018

YOLO-SAM-Point	0.627±0.018	0.506±0.017
YOLO-SAM-PointBBox	0.629±0.018	0.508±0.017
Proposed Model (N-NET: POBi-LSTM)	0.851	0.860

4.6 Performance comparison on Time Factor

The amount of time that was spent training for 100 epochs and the amount of time that was spent testing images on the CT dataset are both stated in table 6. The fact that all of the findings were obtained straight from a single-model test without the use of any postprocessing techniques is something that should be mentioned. According to that, the segmentation accuracy of N-Net on the CT image datasets is higher to that of UNet++, UNet3 +, and U-net.

Table 6: Time Analysis of Proposed Model

Various U-Net Methods	Training Time(s)	Testing Time(s)
U-Net	635	1.567
U-Net3+	2609	1.042
U-Net++	1500	1.321
N-Net	2241	1.021

4.7 Performance comparison of N-NET: POBi-LSTM with Existing Literatures

A performance comparison of the proposed approach with the literature is illustrated in table 7. Comparisons were made between the results of our proposed research and those found in the current literature. As an example, in [29] they used CT images of 254 for training purpose and achieved a DSC score of 67% throughout their study. One of the sources from which they obtained their dataset was the Radiological Society of North America (RSNA). Alternatively, Vamsi et al. [28] acquired a DSC score of 72.92% after including CT images of size 578 including 463 stroke images. These images were included in their study. Through the use of White Matter Fuzzy c-Means (WMFCM) clustering and wavelet-based thresholding, Gautam et al. [30] discovered that there was an average DSC similarity of 82% when employing 20 brain CT images. In the course of this study, the model that we proposed was able to attain a DSC score of 85.757% and an IoU score of 84.3%.

Table 7: Performance Comparison with Our Literature Review

Authors	Methodology	Metric (%)
[28]	Lightweight Deep learning based NN	DSC=72.92
[29]	Semi-Supervised multitask attention-based U-Net	DSC=67
[30]	Automatic segmentation using WMFCMclustering	Average DSC=82
Proposed	N-Net with FEOSSO based BiLSTM	DSC = 86.54
	Swin-HGI-SAM	Accuracy= 0.950
	U-Net	Accuracy= 0.647
	Swin-UNETR	Accuracy= 0.655
	YOLOv8-m	Accuracy= 0.933

4.8 Comparative Analysis for The Model Speed

According to this study, the model proposed had N-Net features, which allow for quick and accurate ICH segmentation. Using the test dataset of 600 images, we were able to determine the total inference time, FPS, and parameter count for every semantic segmentation network. Table 8 shows that the proposed approach outperforms the majority of encoder-decoder-based segmentation networks in terms of inference time. In addition, there was a 43% improvement in inference speed when compared to PSPNet, the third most impressive segmentation model. The proposed framework is third simpler than all of the other semantic segmentation models with the exception of HarDNet in terms of model size.

Table 8: Performance Comparison of Model Speed

Model	Inference time (s)	FPS	Model Size(MB)

U-Net [4]	49.0	24.49/s	69.07
U-Net ++ [5]	51.5	23.30/s	36.65
SegNet[15]	46.9	25.58/s	117.78
PSPNet[16]	56.4	21.28/s	186.83
HarDNet[27]	38.6	31.09/s	16.47
TransHarDNet[39]	39.0	30.78/s	108.09
Proposed Model (N-NET: POBi-LSTM)	40.2	29.57/s	102.84

4.9 Ablation Study

We quickly perform ablation study of proposed model with the additional encoder, skip connection in order to illustrate the efficacy of the additional encoder function proposed in this technique. Table 9 presents the quantitative findings, which indicate that the dual encoder and skip connection works better.

Table 9: Gain comparison of Proposed Model

Architecture	Gain
N-Net	0.9245
N-Net (Dual Encoder)	0.9684
N-Net (Dual Encoder+ Extra Skip Connection)	0.9857

5. CONCLUSION

In this research, we present a technique for the segmentation of CT images that is named N-Net and is based on an upgraded version of the UNet model. A unique dual encoder model is presented to collect additional context information. This model is based on the encoder and decoder architecture of the UNet and was developed to do this research. There is an additional introduction of the SE module as well as the hybrid loss function in order to provide segmentation that is more accurate. In addition, the accuracy of classification has been enhanced by the use of parameter optimized BiLSTM when combined with the FEOSSO model. In order to evaluate the effectiveness of the proposed approach, a large number of experiments are carried out. Furthermore, we evaluate our model in comparison to more complex encoder–decoder structure networks. In terms of both quantitative and qualitative aspects, the results of the experiments demonstrate that the proposed framework is superior to the approaches that were examined. The fact that our technology is able to provide accurate and dependable automated segmentation of medical images is shown by this research.

For the purpose of verifying the proposed model of N-NET: POBi-LSTM, 2000 CT scan images donated by the Catholic University of Korea Seoul St. Mary's Hospital was used. These images included five distinct forms of ICH. A comparison was made between the N-NET: POBi-LSTM and standard segmentation models such as U-Net, U-Net++, SegNet, PSPNet, HarDNet, and TransHarDNet. The N-NET: POBi-LSTM demonstrated the greatest performance in all performance assessment metrics, with a Dice coefficient and IoU of 0.784 and 0.614, respectively. We are able to build an accurate ICH and brain segmentation image for CT imaging by using the model that offers the highest level of performance. This will enable us to detect the bleeding that occurs inside the brain. For the purpose of evaluating this method, more research with large data sets would be required. Additionally, we did not consider CT images that had surgical devices like clips or VP shunts, nor did we include medical records from individuals who had undergone postoperative procedures like craniectomy. In spite of the fact that this situation could have an effect on the outcomes, future research is required in order to verify our methodology for use in medical settings.

REFERENCES

[1] K. Yang, Y. Feng, J. Mu, N. Fu, S. Chen, and Y. Fu, "The presence of previous cerebral microbleeds has a negative effect on hypertensive intracerebral hemorrhage recovery," *Front. Aging Neurosci.*, vol. 9, p. 49, 2017, DOI: 10.3389/FNAGI.2017.00049

[2] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[3] B. Liang, C. Tang, W. Zhang, M. Xu, and T. Wu, "N-Net: An UNet architecture with dual encoder for medical image segmentation," *Signal Image Video Process.*, vol. 17, no. 6, pp. 3073–3081, 2023, DOI: 10.1007/S11760-023-02528-9

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Cham,

Switzerland: Springer, 2015, pp. 234–241, DOI: 10.1007/978-3-319-24574-4_28

[5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imaging*, vol. 39, no. 6, pp. 1856–1867, 2019, DOI: 10.1109/TMI.2019.2959609

[6] T. Brox, O. Ronneberger, Ö. Çiçek, A. Abdulkadir, and S. S. Lienkamp, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," *Lect. Notes Comput. Sci.*, vol. 9901, pp. 424–432, 2016.

[7] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018, DOI: 10.48550/ARXIV.1804.03999

[8] H. Zhang et al., "Intra-domain task-adaptive transfer learning to determine acute ischemic stroke onset time," *Comput. Med. Imaging Graph.*, vol. 90, p. 101926, 2021, DOI: 10.1016/J.COMPMEDIMAG.2021.101926

[9] G. Xu, H. Cao, J. K. Udupa, Y. Tong, and D. A. Torigian, "DiSegNet: A deep dilated convolutional encoder-decoder architecture for lymph node segmentation on PET/CT images," *Comput. Med. Imaging Graph.*, vol. 88, p. 101851, 2021, DOI: 10.1016/J.COMPMEDIMAG.2020.101851

[10] S. H. Gao, M. M. Cheng, K. Zhao, X. Y. Zhang, M. H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, 2019, DOI: 10.1109/TPAMI.2019.2938758

[11] V. Abramova et al., "Hemorrhagic stroke lesion segmentation using a 3D U-Net with squeeze-and-excitation blocks," *Comput. Med. Imaging Graph.*, vol. 90, p. 101908, 2021, DOI: 10.1016/J.COMPMEDIMAG.2021.101908

[12] J. You et al., "3D dissimilar-siamese-U-Net for hyperdense middle cerebral artery sign segmentation," *Comput. Med. Imaging Graph.*, vol. 90, p. 101898, 2021, DOI: 10.1016/J.COMPMEDIMAG.2021.101898

[13] S. Mizusawa, Y. Sei, R. Orihara, and A. Ohsuga, "Computed tomography image reconstruction using stacked U-Net," *Comput. Med. Imaging Graph.*, vol. 90, p. 101920, 2021, DOI: 10.1016/J.COMPMEDIMAG.2021.101920

[14] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation," *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.

[15] A. M. Mendrik et al., "MRBrainS challenge: Online evaluation framework for brain image segmentation in 3T MRI scans," *Comput. Intell. Neurosci.*, vol. 2015, no. 1, p. 813696, 2015, DOI: 10.1155/2015/813696

[16] J. Koutník, K. Greff, F. Gomez, and J. Schmidhuber, "A clockwork RNN," in *Proc. 31st Int. Conf. Mach. Learn. (ICML)*, vol. 32, 2014, pp. 1863–1871.

[17] R. P. Poudel, P. Lamata, and G. Montana, "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation," in *Int. Workshop Reconstr. Anal. Moving Body Organs*, Cham, Switzerland: Springer, 2016, pp. 83–94, DOI: 10.1007/978-3-319-52280-7_8

[18] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021, DOI: 10.48550/ARXIV.2102.04306

[19] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "TransBTS: Multimodal brain tumor segmentation using transformer," in *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Cham, Switzerland: Springer, 2021, pp. 109–119, DOI: 10.1007/978-3-030-87193-2_11

[20] R. Rosnelly, B. S. Riza, L. Wahyuni, S. E. V. Haryanto, and A. Prasetio, "Vehicle detection using machine learning model with the Gaussian mixture model (GMM)," *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl.*, vol. 13, no. 4, pp. 233–243, 2022, DOI: 10.58346/JOWUA.2022.I4.016

[21] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.

[22] G. Máttyus, W. Luo, and R. Urtasun, "DeepRoadMapper: Extracting road topology from aerial images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 3458–3466, DOI: 10.1109/ICCV.2017.372

[23] H. R. Roth, L. Lu, A. Farag, H. C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Cham, Switzerland: Springer, 2015, pp. 556–564, DOI: 10.1007/978-3-319-24553-9_68

[24] J. Sengupta, R. Alzbutas, P. Falkowski-Gilski, and B. Falkowska-Gilska, "Intracranial

hemorrhage detection in 3D computed tomography images using a bi-directional long short-term memory network-based modified genetic algorithm," *Front. Neurosci.*, vol. 17, p. 1200630, 2023, DOI: 10.3389/FNINS.2023.1200630

[25] J. Choi and X. Zhang, "Classifications of restricted web streaming contents based on convolutional neural network and long short-term memory (CNN-LSTM)," *J. Internet Serv. Inf. Secur.*, vol. 12, no. 3, pp. 49–62, 2022, DOI: 10.22667/JISIS.2022.08.31.049

[26] M. M. Khan et al., "A deep learning-based automatic segmentation and 3D visualization technique for intracranial hemorrhage detection using computed tomography images," *Diagnostics*, vol. 13, no. 15, p. 2537, 2023, DOI: 10.3390/DIAGNOSTICS13152537

[27] P. Chao, C. Y. Kao, Y. S. Ruan, C. H. Huang, and Y. L. Lin, "HarDNet: A low memory traffic network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 3552–3561, DOI: 10.1109/ICCV.2019.00365

[28] B. Vamsi, D. Bhattacharyya, D. Midhunchakkravarthy, and J. Y. Kim, "Early detection of hemorrhagic stroke using a lightweight deep learning neural network model," *Trait. Signal*, vol. 38, no. 6, pp. 1727–1736, 2021, DOI: 10.18280/TS.380616

[29] J. L. Wang, H. Farooq, H. Zhuang, and A. K. Ibrahim, "Segmentation of intracranial hemorrhage using semi-supervised multi-task attention-based U-Net," *Appl. Sci.*, vol. 10, no. 9, p. 3297, 2020, DOI: 10.3390/APP10093297

[30] A. Gautam and B. Raman, "Automatic segmentation of intracerebral hemorrhage from brain CT images," in *Mach. Intell. Signal Anal.*, Singapore: Springer, 2018, pp. 753–764, DOI: 10.1007/978-981-13-0923-6_64

[31] Y. Camgözlü and Y. Kutlu, "Leaf image classification based on pre-trained convolutional neural network models," *Nat. Eng. Sci.*, vol. 8, no. 3, pp. 214–232, 2023, DOI: 10.28978/NESCIENCES.1405175

[32] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017, DOI: 10.48550/ARXIV.1706.05587

[33] L. C. 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 Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2017, DOI: 10.1109/TPAMI.2017.2699184

[34] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Eur. Conf. Comput. Vis. (ECCV)*, Cham, Switzerland: Springer, 2018, pp. 833–851, DOI: 10.1007/978-3-030-01234-2_49

[35] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Int. Workshop Deep Learn. Med. Image Anal.*, Cham, Switzerland: Springer, 2018, pp. 3–11, DOI: 10.1007/978-3-030-00889-5_1

[36] H. Huang et al., "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2020, pp. 1055–1059, DOI: 10.1109/ICASSP40776.2020.9053405

[37] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141, DOI: 10.1109/TPAMI.2019.2913372

[38] AIHub, "Dataset provider site," 2021. [Online]. Available: AIHub Dataset Provider Site. Accessed: Aug. 10, 2021.

[39] W. Li, Y. Sun, G. Zhang, Q. Yang, B. Wang, X. Ma, and H. Zhang, "Automated segmentation and volume prediction in pediatric Wilms' tumor CT using nnU-Net," *BMC Pediatr.*, vol. 24, no. 1, p. 321, 2024, DOI: 10.1186/S12887-024-04775-2