WILEY | Hindawi

Research Article

An Efficient Retinal Segmentation-Based Deep Learning Framework for Disease Prediction

R. Dhanagopal [iD],[1] P. T. Vasanth Raj,[1] R. Suresh Kumar,[1] R. Mohan Das,[2] K. Pradeep,[3] and Owusu-Ansah Kwadwo [iD][4]

[1]Centre for System Design, Chennai Institute of Technology, Chennai, Tamil Nadu, India
[2]Department of Electronics and Communication Engineering, Balaji Institute of Technology & Science Narsampet, Warangal, India
[3]Department of Biomedical Engineering, Chennai Institute of Technology, Chennai, Tamil Nadu, India
[4]Biomedical Engineering Technology, Koforidua Technical University, Koforidua, Eastern Region, Ghana

Correspondence should be addressed to Owusu-Ansah Kwadwo; doansah@anuc.edu.gh

Deep learning (DL) technology has shown to be the most effective method of completing class assignments in the last several years. Specifically, these approaches were used for segmentation, classification, and prediction of retinal blood vessels, which was previously unattainable. U-Net deep learning technology has been hailed as one of the most significant technological advances in recent history. In the proposed work, improved segmentation of retinal images using U-Net, bidirectional ConvLSTM U-Net (BiDCU-Net), and fully connected convolutional layers, such as absolute U-Net, BiConvLSTM preferences, and also the fully connected convolutional layer method are proposed. Three well-known datasets were subjected to the suggested technique's evaluation: the DRIVE, STARE, and CHASE DB1 databases. This suggested technique was tested using the required precise measures in percentage of accuracy, F1 score, sensitivity, and specificity in DRIVE, 97.32, 83.85, 82.56, and 98.68 in CHASE, 97.44, 81.94, 83.92, and 98.45 in STARE, 97.33, 82.3, 82.12, and 98.57 in STARE, respectively. Furthermore, we assert that the strategy outperforms three other similar strategies in terms of effectiveness.

1. Introduction

Deep learning is now being used for picture segmentation, classification, captioning, and prediction, among other applications. Deep convolutional neural network (DCNN) versions like Residual Net, AlexNet, VGG, DenseNet, and CapsuleNet have made significant strides ahead in recent years [1, 2]. For some applications, a DL-based (in particular, a CNN-based) solution provides cutting-edge execution for split assignments and order, among other things. For the time being, the enactment capacity will be enough to overcome the DL models' inadequacies. Second, the networks' dropout becomes more consistent. Third, the creation of CNN models [3] necessitates specific augmentation procedures, which are explained further down. It is common practice to check and estimate models for massive datasets such as the yields of the probability value categorization

are derived from ImageNet, and the jobs are single labels, by using arrangement techniques. Additionally, modest architectural version models [4] are employed for semantic picture segmentation tasks.

For example, vessel segmentation may be utilized to identify distinct disorders that involve blood vessels in different ways. A variety of disorders can result in changes in the breadth and curvature of the retinal veins. In order to detect many of these pathogens, including glaucoma, hypertension, and diabetic retinopathy, at an early stage [5], it is necessary to detect them at an early stage. Glaucoma, hypertension, and diabetic retinopathy are just a few of the pathogens that cause vision to deteriorate in working people. This section of the skin sore is responsible for detecting and investigating cancerous skin progress in its initial stages. Melanoma is the worst type of skin cancer, and it is caused by melanocytes that grow too fast. Dermoscopy is a

noninvasive imaging method that records the skin's surface and provides sensations of the skin's surface using light amplifiers and breathing fluids. Dermoscopic pictures of melanoma might lead dermatologists to believe that the tumor is incorrect or abstract. It is possible to achieve 92 percent relative durability over a five-year period if melanoma is discovered in its early stages [6, 7].

Deep learning networks produce outstanding results, and they are being used in clinical imaging to tackle control techniques that have been developed. Because of the large number of network characteristics, that network requires a considerable amount of information in order to train and execute high-quality speculative operations. Large (and well-clarified) datasets are difficult to get, which presents a significant challenge in the segmentation of medical adages [8, 9]. It is necessary to identify pixels in clinical images instead of using the name of the imaging stage when doing clinical image segmentation. The fully convolutional neural network (FCN) [10] has been shown to be one of the utmost powerful deep networks accessible for picture segmentation.

The ability to enhance this CNN to a U-Net allows for achieving the best results without the requirement to train on a large quantity of data. This framework is composed of individuals that code and analyze methods. Numerous maps with lesser dimensions are excluded from consideration by the encoding approach [11]. By running up convolutions of the same scale as input, this disengaging technique may be utilized to construct segmentation maps of units from input data. There have been a number of U-Net extensions proposed so far [12, 13]. The most significant disadvantage is each set of feature maps to be done independently the encoding procedure.

From an architectural standpoint, the CNN model for characterization tasks necessitates the use of an encoder and provides the possibility of a class as a yield. Activation functions and subsampling layers in network processes that decrease the dimensionality of feature maps allowed us to do convolution tasks. Due to the increasing number of input samples transiting network layers, the size of a feature map grows as the number of feature maps grows [14]. This is depicted in Figure 1 in the initial half of the model (in blue), which represents the model's start. Individually, the number of networks may increase as the overall number of feature maps increases across the network hierarchy's deeper tiers. SoftMax behaviors are unavoidable while processing goal group probability since they are essential for proper processing.

Segmentation tasks necessitate the use of both convolutional encoding and decoding units, rather than classifications, in the design of the job. The encoding unit converts the photos provided as input into a relatively small number of different maps. The decoding unit conducts an upconvolution (deconvolution) operation on the original image to construct segmentation feature maps with comparable dimensions. As a result, the structure responsible for segmentation often requires twofold many more network parameters as the classification process design.

This paper presents their proposal for BiDCU-Net, a better-quality U-Net version that uses BiConvLSTM to reuse an FCL and bypass connection with the feature maps [15–43]. The ensuing encoding layer's features have a higher degree of resolution, whereas the feature maps acquired from prior upconvolutional layers have a lot of semantic information, indicating that the approach is working properly. It is preferable over merely concatenating these 2 types of feature maps using nonlinear functions since it will result in more dependable segmentation performance. As a result, we add BiConvLSTM to the skip connections in the U-Net architecture described in this article. During each block, the traits that have been obtained are carried over to the future block. As a result of this strategy, the process may learn a broad variety of attributes that are dependent on past amounts of relevant data while simultaneously minimizing the necessity for recurrent learning functions, which would otherwise be necessary.

2. Background

With the aid of multiple medical image datasets and computer vision, semantic segmentation is a very important field of study in which deep convolutional neural networks (DCNNs) are utilized to categorize an individual pixel in an image. Prior to the deep learning revolution, the usual machine learning approach depended mostly on the manual characteristics that were employed to individually identify pixels [18–19]. Several models have shown in recent years that deeper networks are the most successful for tasks such as detection and segmentation. The severe gradient problem solved by the introduction of modern activation functions like Exponential Linear Units (ELUs) and Rectified Linear Units (ReLUs), which makes training very deep models more difficult, makes it more difficult to train very deep models. Another alternative is to use a deep residual model to solve the problem, as discussed in [20–22], which uses identity mapping to speed up the training process.

Furthermore, when it comes to the segmentation of actual pictures, FCN-based CNN segmentation approaches outperform their counterparts. Random architecture is considered to be one of these image patch-based implementations, and it is extremely computationally costly due to the large number of network parameters it contains (about 134.5 million). The most important constraint would be that there is a considerable degree of overlap between pixels and that the same convolutions are done on a constant basis. With the introduction of recurrent neural networks (RNN), the efficiency of FCNs has increased and is being tuned with increasingly large datasets [23]. SegNet is made up of two primary components: the network and the server [24–26]. One such network is the VGG16 encoding network, which has 13 layers and uses pixel-wise classification layers for the equivalent decoding network. The most important component of this article is discussed about how the upsamples done in decoder with the input feature maps that are of lower resolution than the target resolution. In 2015, a new SegNet variation, known as Bayesian SegNet [27], was proposed as an upgrade to the original. Machine views are used to study the many types of architectures available. There are also numerous deep learning structures that have been
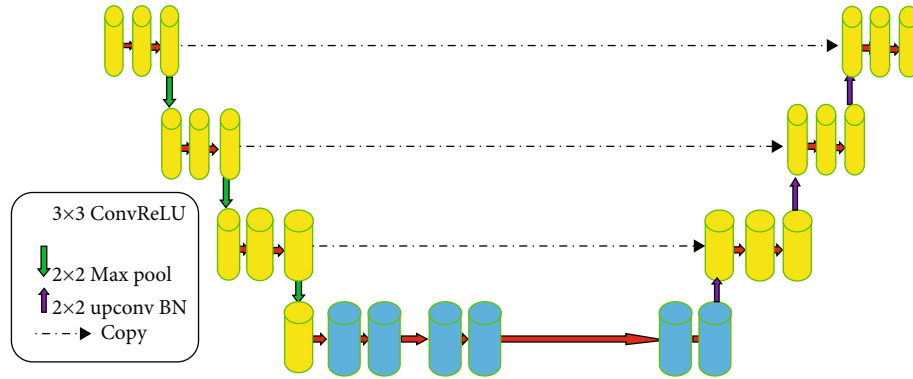
FIGURE 1: Basic architecture of U-Net.

expressly presented for the medical image segmentation because they detect the inadequacy of data as well as concerns with class imbalances in the medical picture.

"U-Net" was one of the earliest and most widely used methods for segmenting semantic medical pictures when they were initially introduced. The U-Net model's essential design is seen in Figure 1. According to its construction, the network is composed of two essential components: the convolutional encoder and the decoder [28]. The simple convolutional tasks are carried out in both regions of a network by activating ReLU in both directions. For the purpose of downsampling, the encoding unit will perform the 2×2 max-pooling operations. During the decoding process, conversion is carried out in order to sample the feature maps from the input data. The original version of U-Net was used to produce and replicate feature maps from the encoder to the decoding system. U-Net design offers various advantages when it comes to segmenting activities. For starters, such a paradigm allows for the utilization of global location and context at a similar time. Second, it is capable of operating with just a small number of samples while increasing the efficiency of segmentation jobs [29]. Third, the complete picture is processed end to end in the front route, and segmentation maps are created directly from the image data. Because of this, as compared to patch-based segmentation [30, 31], U-Net effectively covers the whole backdrop of the input photos.

In addition, U-Net is currently widely used in a wide range of applications. Since then, many different U-Net models have been developed, together with one specialized for CNN-based medical imaging data segmentation [32, 33]. The original U-Net architecture has been modified in two ways in this model. Multiple segmentation maps and forward feature maps are blended and merged across the whole network. It is normal practice to obtain and summarize maps from many levels of encoding and decoding systems. The authors indicate positive gains in their performance during preparation, but no effect was shown when summed features were incorporated during the test time [34], in contrast to U-Net. The definition, on the other hand, has shown that summing attributes have an effect on network performance. Residual networks and U-Net are used for biomedical image segmentation tasks [35]; it is pos-

sible to objectively analyze the effects of skipped connections.

It is proposed in this study that BiDCU-Net be used as an upgrade to U-Net and that it produces significantly more output than current segmentation jobs. Furthermore, the rate of convergence of a network of this type is significantly impacted by BN. In order to attain our objectives, we are experimenting with several methods of retinal imaging.

The primary contribution is the proposal [44] of a novel U-form deep learning architecture that makes use of lightweight convolution blocks. This is done with the goal of maintaining a better level of segmentation performance while simultaneously minimizing the amount of computational complexity. Preprocessing and data augmentation techniques are recommended with regard to the retinal picture and blood vessel properties as the second key contribution.

The contribution of this work is to propose the BiDCU-Net, which combines U-Net, BiConvLSTM, and entirely linked layered convolution. The network does this by utilizing its capabilities, which include totally linked layer convolutions of both BConvLSTM states. Encoding, decoding, batch normalization (BN), and Bi-ConvLSTM are the network's main components.

3. Proposed Framework: An Efficient Retinal Segmentation-Based Deep Learning Framework for Disease Prediction (EDLFDPRS)

As shown in Figure 2, The BiDCU-contracting Net's encoding procedure is separated into four phases. Two 3×3 convolutional filters, a 2×2 maximum pooling technique, and a ReLU are used in each phase. The no. of feature maps doubles with each level. The contracting method extracts pictures at a faster rate, layer by layer, and increases the layer size of those representations. To finish, the encoding path passes via the last layer of a multidimensional picture definition, which contains important semantic data. At the completion of the encoding procedure, primary U-Net had series of convolutional layers. The method works by feeding a network a series of convolutional layers to identify distinct
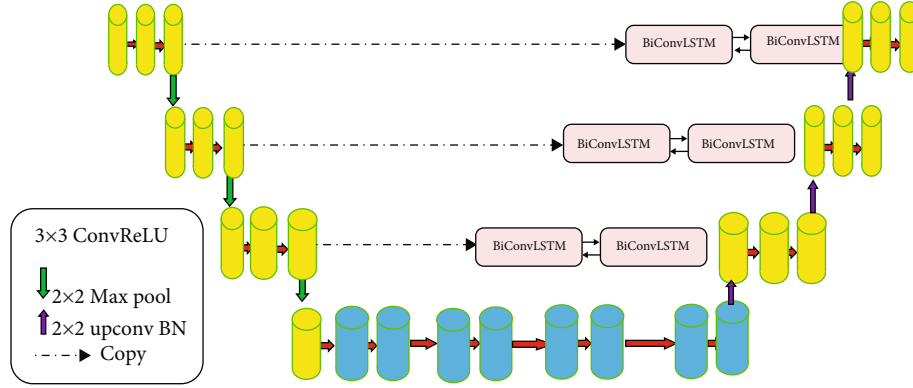
FIGURE 2: Proposed framework for disease prediction.

features. The network may, however, acquire duplicate properties as convolutions continue. Fully linked convolutions [36] are presented as a solution to this problem. This allows the network to improve the system's efficiency in recycling feature maps by using the concept of "collective awareness". It often entails integrating all previous convolutional layers' feature maps to the feature map acquired from current layer and utilized as input data for the subsequent convolution. The concept of entirely linked convolutions has a number of advantages over ordinary convolutions. It permits the network to use variety of feature maps instead of duplicates. By allowing information to flow throughout the network and reusing functionality, this notion improves the network's representational capacity even more. Prior to being used, fully coupled convolutions should take advantage of all available capabilities, since this will prevent the network from exploding or eliminating gradients [37]. In addition, the gradients will be returned to their original locations in the network at a faster rate. In the proposed network, the concept of entirely linked convolutions is used. As a result, a single block has two successive convolutions. The encoding route's last convolutional layer displays a succession of N blocks, as seen in Figure 3. The blocks become densely connected.

Consider Z_e^j is the output of every $j^{\text{th}}(j \in \{0, \cdots, N\})$ convolutional block.

input $j \longrightarrow$ convolutional block.

For example,

$$\left[Z_e^1, Z_e^2, Z_e^3, \cdots, Z_e^{j-1}\right] \in R^{(j-1)F_k * W_k * H_k}. \tag{1}$$

Equation (1) is the feature maps of all previous convolutional blocks connected to their input data, and the block's output is

$$\left[Z_e^j\right] \in R^{F_k * W_k * H_k}. \tag{2}$$

We use Z_e rather than Z_e^j.

In the decoding part, each stage of the decoding procedure begins with an examination of the previous layer of yield. In the decoding process, the components' important feature maps are clipped and replicated in the standard U-

Net. These feature maps may be connected to the result of the upsampling algorithm. In BiDCU-Net, BiConvLSTM is utilized to investigate these two types of feature maps in greater depth. Let be the set of sequential convolutional layer feature maps, where F_{kli} is the layer number and $W_{\text{ml}}H_{\text{ml}}$ is the feature map for each layer l and be the collection of copied feature maps from the encoding component. We should also pay attention to and. When an upsampling operation followed by a 2×2 doubles the size of a feature map while halves the number of feature channels, Z_d is initially transferred from an updated convolutional layer, resulting in. To put it another way, the extended route increases the image size by layering the feature maps.

Instead of BiDCU-Net, BiConvLSTM was used in the proposed method to deeply analyze the details of these 2 types of feature maps.

Let,

$$Z_e \in R^{F_{kl} * W_{kl} * H_{kl}}. \tag{3}$$

The copied encoding component's feature mappings are gathered in equation (3). And

$$Z_d \in R^{F_{kl+1} * W_{kl+1} * H_{kl}+1}. \tag{4}$$

Equation (4) is the just be set of feature maps of the subsequent convolutional layer in which the F_{kl} feature map is the no. of layer l and $W_{kl} \times H_{kl}$ of each layer l feature map.

Let us consider

$$
\begin{aligned}
F_{kl+1} &= 2 \times F_{kl}, \\
W_{kl+1} &= \frac{1}{2} W_{nl}, \\
H_{kl+1} &= \frac{1}{2} H_{nl}.
\end{aligned}
\tag{5}
$$

According to Figure 3, Z_d is initially shifted from an enhanced convolutional layer when a 2×2 upsampling operation twice the size of a feature map while halves the number of feature channels, resulting in $Z_d^{\text{up}} \in R^{R^{F_{kl} * W_{kl} * H_{kl}}}$. The extended route, in other words, elevates the feature
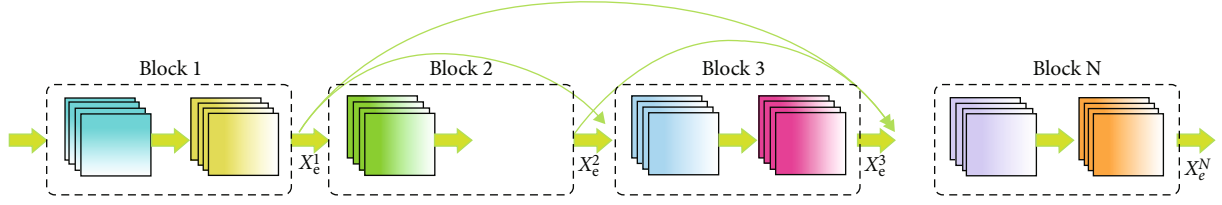
Figure 3: Block representation of the proposed method.

maps layer by layer to achieve the true image size from the last layer.

The output of the upconvolutional layer, Z_d^{up}, performs a BN function and creates $\widehat{Z}_d^{\text{up}}$ after upsampling the image. One difficulty encountered during preparation is that the distribution of activations varies between the concealed units. Because each layer must be modified to a different distribution in each training phase as a result of this difficulty, the training operation takes longer. A neural network is used in BN [38] to increase consistency by normalizing the input into a network layer by eliminating the batch means and splitting the batch standard deviations before feeding the standardized input into network layer. BN has an impact on the evolution of neural network training. Furthermore, model's efficiency is typically improved as a result of the small regulatory effect [39], which is a good thing.

The output of the BN phase $(\widehat{Z}_d^{\text{up}} \in R^{R^{F_{kl}*W_{kl}*H_{kl}}})$ is now being transmitted to a BConvLS layer through the BConvLS layer. Due to the fact that these models employ comprehensive relations between state-to-state and state transitions, the LSTM model has a significant shortcoming in that there is no spatial correlation of these networks. ConvLSTM, which turns operations into input-to-state and state-to-state transformations, was suggested as a solution to this problem [40] and should be implemented. An input gate $i(t)$, an output gate $o(t)$, a forgotten gate $f(t)$, and a memory cell C make up the circuit. It is broken down into four components (t). Gates that monitor access to, update, and clear memory cells are known as input, output, and forget gates. The following are some of ConvLSTM's characteristics.

$$
\begin{aligned}
i(t) &= \alpha\big(Wn_{xn}\, X\, en_t + Wn_{hi} XH_{t-1} + Wn_{ci}\, X\, C(t-1) + b_i, \\
f(t) &= \alpha\big(Wn_{xf}\, X\, e_t + Wn_{hf}\, X\, Hn_{t-1} + Wn_{cf}\, X\, C(t-1) + b_f, \\
C(t) &= f_t^{\circ}\, C(t-1) + i(t)\, \tanh\,(Wn_{xc} X\, e_t + Wn_{hc}\, XHn_{t-1} + b_c, \\
o(t) &= \alpha\big(Wn_{xo} Xe_t + Wn_{ho} XHn_{t-1} + Wm_{co} \circ C(t) + b_o, \\
Hm_t &= o(t)\, \tan\,(C(t)).
\end{aligned}
$$

$$(6)$$

X and $\circ$ are the function of the convolution and Hadamard.

e_t is the input tensor.

H_t is the hidden state tensor.

$C(t)$ is the tensor memory cell.

Wn_{xX} and Wn_{hX} are input and hidden kernel of 2D convolution.

b_i, b_f, b_g, b_o, and b are predominant function.

BConvLSTM used to encode $\widehat{y}_d^{\text{up}}$ and distribute information within this network. As a result, BConvLSTM analyzes the input data in both directions like forward and backward using two ConvLSTMs. Then, in either way, by addressing the data dependence, it decides which data input is now being used. Only dependency with a forward direction is taken into account in a traditional ConvLSTM. As a result, in order to effectively evaluate backward dependencies, all issues must be thoroughly addressed in this order [41]. Increasing statistical performance has been demonstrated through the investigation of both forward and backward temporal perspectives. Each ConvLSTM can be thought of as a standard in both the forward and backward direction. We have two sets of parameters since we have two different states: backward and forward. The BConvLSTM production is expected to be as follows:

$$
Z(t) = \tanh\left(Wn_y^{\vec{H}} \times \vec{H}n + Wn_y^H \times Hm + b\right), \quad (7)
$$

where $\vec{H}n$ and Hm denote hidden state tensors with forwarding and backward states, respectively, b denotes term for basic, and $Z(t)$ is output with regard to spatiotemporal bidirectional details. Aside from that, tanh is the hyperbolic tangent that is utilized to mix the front and rear output in a nonlinear fashion. To train the network, we employ the power approach, which is unique to the U-Net.

4. Experimental Results

Three prominent datasets, DRIVE, STARE, and CHASH_DB1, were used to investigate the segmentation of retinal blood vessels. A total of 40 color retinal images are included in this dataset, with 20 examples used for training and another 20 samples utilized for testing. The following is how the DRIVE dataset is organized. The image's size, according to the original, is 565×584 pixels. To create a square data collection, the photographs are only clipped to contain data from columns 9 to 574, resulting in a frame size of 565×565 pixels (in this case). We randomly picked 171,000 patches for training from a total of 190,000 patches discovered from 20 pictures in DRIVE dataset.

The patch size for each of the three datasets shown in Figure 4 is 48×48 pixels on each side. STARE consists of 20 color pictures, each of which is 700×605 pixels in width and height. The small sample size necessitates the employment of two methodologies for training and testing on this
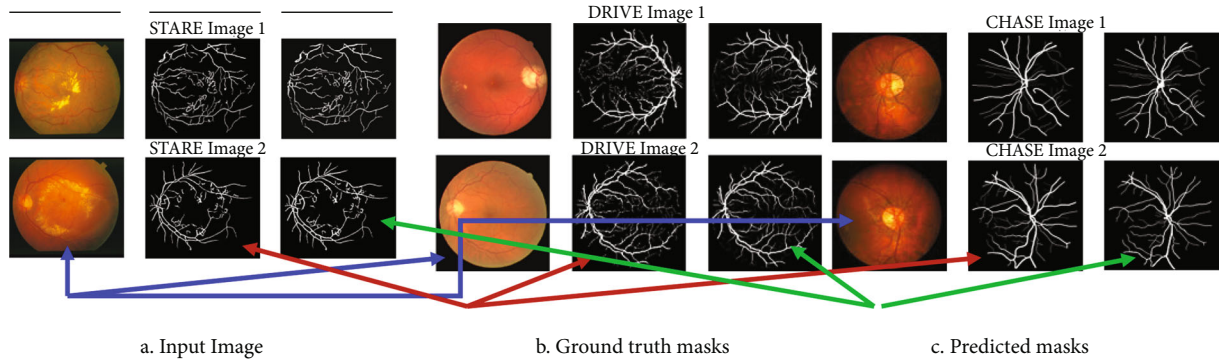
a. Input Image   b. Ground truth masks   c. Predicted masks

FIGURE 4: Output of EDLFDPRS method.

TABLE 1: DRIVE dataset performance with various architecture.

Network	U-Net [28]	RU-Net [4]	Dense U-net [18]	Proposed EDLFDPRS
Accuracy	95.08	95.65	95.53	97.32
$F1$ score	81.87	81.55	81.71	83.85
Recall	82.65	84.62	84.81	82.56
Precision	85.14	84.72	83.94	85.19
Sensitivity	82.65	83.74	79.86	83.96
Specificity	97.97	97.83	97.36	98.68
AUC	96.81	97.84	97.82	98.79

TABLE 2: STARE dataset performance with various architecture.

Network	U-Net [28]	RU-Net [4]	Dense U-net [18]	Proposed EDLFDPRS
Accuracy	96.33	97.6	97.21	97.33
$F1$ score	81.22	82.63	81.47	82.3
Recall	84.27	82.61	83.17	82.12
Precision	80.07	82.16	81.92	82.52
Sensitivity	82.12	82.69	79.37	84.27
Specificity	97.97	97.28	97.47	98.57
AUC	97.82	97.15	97.45	98.84

TABLE 3: CHASE_DB1 dataset performance with various architecture.

Network	U-Net [28]	RU-Net [4]	Dense U-net [18]	Proposed EDLFDPRS
Accuracy	96.12	96.12	96.43	97.44
$F1$ score	78.23	79.81	79.98	81.94
Recall	82.45	82.74	82.58	83.92
Precision	82.21	83.23	83.05	80.07
Sensitivity	82.45	85.55	78.86	83.92
Specificity	97.98	97.17	97.55	98.45
AUC	98.05	98.11	98.17	98.79

dataset, both of which are employed rather often. First and foremost, some of the training examples were chosen at random from each group of 20 photos. A sample is evaluated and remaining 19 specimens are trained on the results using a method known as the "leave-one-out" procedure. There is never any overlap between the training and testing samples. The "leave-one-out" method for STARE datasets is used in this implementation, which is detailed below. CHASH_DB1 is a dataset that consists of 28 images, each with a resolution of 999×960 pixels. The dataset would be divided into two groups, with samples chosen at random in each category. For training, a total of 20 samples are used, with the remaining 8 samples being used for testing.

Tamim et al. [42] test data from the DRIVE, STARE, and CHASE DB1 databases, among other places, was used to val-

idate the framework. Tables 1–3 as well as Figures 5–7 show the $F1$ score, accuracy, sensitivity, specificity, and AUC. As indicated in the table, the efficiency of BIDCU-Net is higher than that of other updated models. According to the proposed approach, the average BIDCU-Net $F1$ score for the DRIVE, STARE, and CHASE DB1 is 83.85, 82.3, and 81.94, respectively.

The area under the curve (AUC) is the measure of the capacity of a classifier to discriminate between classes and is used as a summary of the ROC curve. The greater the AUC, the better the performance of the model in distinguishing between the positive and negative groups.

$$AUC = \left(\text{percentage of similarities in the total available pairs} + \frac{1}{2} \& * \text{percentage of tied in the total available pairs} \right).$$

(8)

Table 2 compares and contrasts each technique with a variety of additional state-of-the-art methodologies, as indicated in the headers, in comparison to the DRIVE, STARE, and CHASE datasets. We discover that the backdrop and blood vessels can be segmented more reliably than the remainder of the image, based on the BiDCU-highest Net's $F1$ score and highest AUC. The $F1$ score indicates a high level of recall and precision, as well as a successful implementation of our strategy. BiDCU-Net has the highest levels
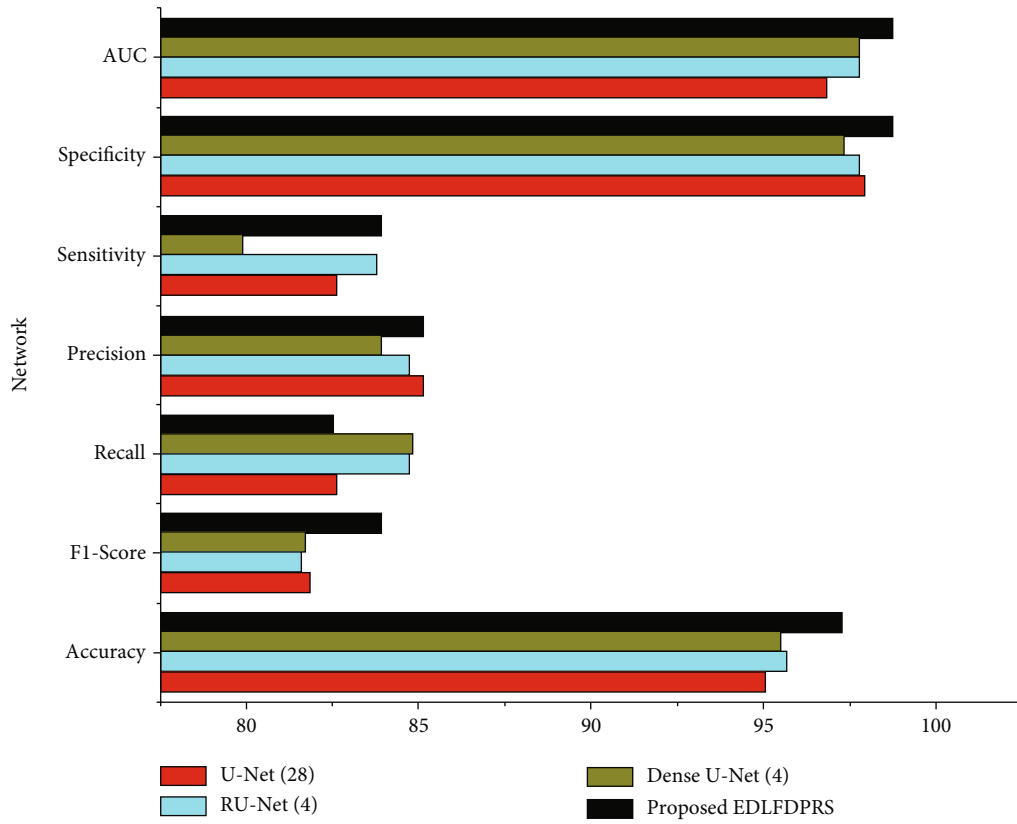
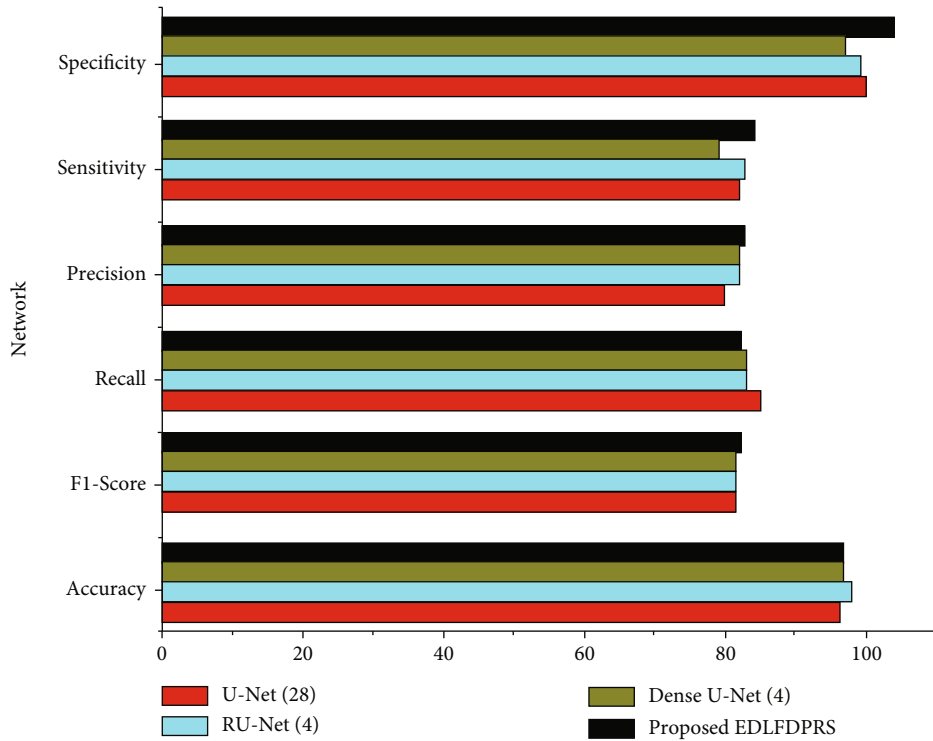FIGURE 5: DRIVE dataset performance with various architecture.

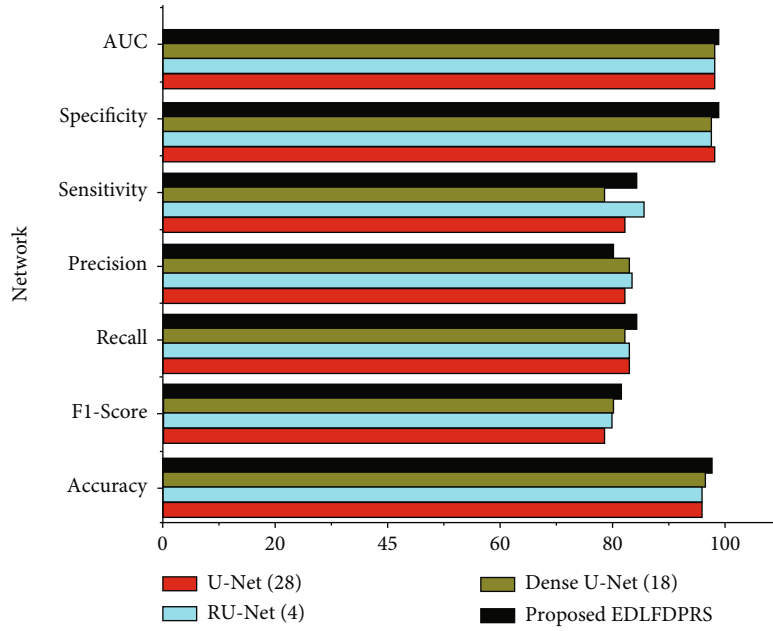FIGURE 6: STARE dataset performance with various architecture.

FIGURE 7: CHASE_DB1 dataset performance with various architecture.

TABLE 4: Comparison results of existing with proposed method.

Existing methods	Alom [4]						Wang [18]		Jiang [11]			Guo [34]			Proposed		
Methods	RU-Net			R2U-Net			Dense U-net		FCN			BSCN			EDLFDPRS		
Dataset	D	S	C	D	S	C	D	S	D	S	C	D	S	C	D	S	C
ACC	95.56	97.06	96.22	95.56	97.12	96.34	95.11	95.38	96.46	96.33	97.7	98.46	98.72	98.89	97.32	97.33	97.44
$F1$ score	91.55	93.96	78.1	91.71	84.75	92.8	—	—	76.75	77.55	82	82.36	85.47	75.6	83.85	92.3	81.94
Sensitivity	77.51	81.08	74.59	77.92	82.98	77.56	79.86	79.14	66.63	83.32	83.23	81.79	87.51	79.72	82.56	82.12	83.92
Specificity	98.16	98.71	98.36	98.13	98.62	98.2	97.36	97.22	99.33	97.4	98.67	98.79	98.94	98.96	98.68	98.57	98.45
AUC	97.82	99.09	98.03	97.84	99.14	98.15	97.4	97.04	97.8	97.9	99.12	98.74	99.41	98.74	98.79	98.84	98.79

Note: D: DRIVE; S: STARE; C: CHASE; ACC: accuracy; sensitivity; specificity; AUC: area under the ROC curve.
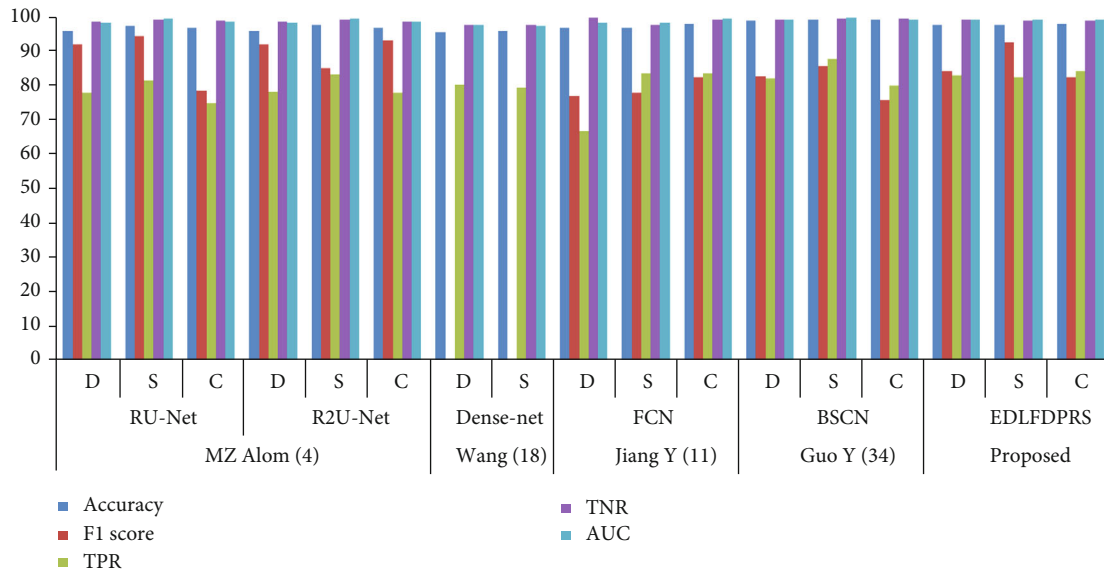

FIGURE 8: Comparison results of existing with proposed method.

of specificity and sensitivity, implying that more pixels in vessels can be better labelled in the future. Vascular pixels make up a modest percentage of all visual pixels most of the time. Because the unbalanced type is utilized in the description, defining the segmentation of the retinal blood vessels is more challenging. As a result, our technique's high sensitivity is critical, and a computer-aided demagnetization device may detect blood vessels without using false instances. The BiDCU-Net displays maximal accuracy over a similar period of time.

Table 4 and Figure 8 show the proposed method comparison results of various performance metrics with existing methods, Except $F1$ score, all the remaining parameters of the EDLFDPRS method are improved than those of the existing methods for the DRIVE dataset. The proposed method performance metric of $F1$ score is improved than the existing methods for the CHASE dataset. Finally, the proposed method proves that it has better accuracy when compared with existing methods.

5. Conclusion

To construct an effective framework, a mix of bidirectional LSTM network, U-Net, and fully connected layer is used in this research work. The network also includes a tightly linked convolutional layer block to make more biased data available, leading to more reliable segmentation results. This work was also able to boost the network's speed by roughly 6 times by introducing BN after the upconvolutional layer. The proposed method is given the accuracy as 97.32%, 97.33%, and 97.44% for the datasets DRIVE, STARE, and CHASE, respectively. When compared with existing method, the proposed method accuracy is improved 2.24%. Also, the proposed work is compared with the different architectures such as U-Net, RU-Net, and Dense U-net models in segmentation activities for all three datasets utilizing the same number of network parameters.

Data Availability

The data are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.

[2] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[3] M. Z. Alom, T. M. Taha, C. Yakopcic et al., "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, 2019.

[4] H. C. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[5] A. D. Dick, N. Tundia, R. Sorg et al., "Risk of ocular complications in patients with noninfectious intermediate uveitis, posterior uveitis, or panuveitis," *Ophthalmology*, vol. 123, no. 3, pp. 655–662, 2016.

[6] L. F. di Ruffano, Y. Takwoingi, J. Dinnes et al., "Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults," *Cochrane Database of Systematic Reviews*, vol. 12, 2018.

[7] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm," *Diagnostics*, vol. 9, no. 3, p. 72, 2019.

[8] S. Leclerc, E. Smistad, J. Pedrosa et al., "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2198–2210, 2019.

[9] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für MedizinischePhysik*, vol. 29, no. 2, pp. 102–127, 2019.

[10] Y. Jiang, H. Zhang, N. Tan, and L. Chen, "Automatic retinal blood vessel segmentation based on fully convolutional neural networks," *Symmetry*, vol. 11, no. 9, p. 1112, 2019.

[11] J. Schlemper, O. Oktay, M. Schaap et al., "Attention gated networks: learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, 2019.

[12] P. Poudel, A. Illanes, D. Sheet, and M. Friebe, "Evaluation of commonly used algorithms for thyroid ultrasound images segmentation and improvement using machine learning approaches," *Journal of healthcare engineering*, vol. 2018, 13 pages, 2018.

[13] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: a persistent memory network for image restoration," in *Proceedings of the IEEE international conference on computer vision*, pp. 4539–4547, Venice, Italy, 2017.

[14] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Advanced Engineering Informatics*, vol. 32, pp. 139–151, 2017.

[15] R. Shang, J. He, J. Wang, K. Xu, L. Jiao, and R. Stolkin, "Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification," *Knowledge-Based Systems*, vol. 194, p. 105542, 2020.

[16] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. B. Ayed, "HyperDense-net: a hyper-densely connected CNN for multi-modal image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 5, pp. 1116–1126, 2019.

[17] C. Wang, Z. Zhao, Q. Ren, Y. Xu, and Y. Yu, "Dense u-net based on patch-based learning for retinal vessel segmentation," *Entropy*, vol. 21, no. 2, p. 168, 2019.

[18] B. Sahiner, A. Pezeshk, L. M. Hadjiiski et al., "Deep learning in medical imaging and radiation therapy," *Medical Physics*, vol. 46, no. 1, pp. e1–e36, 2019.

[19] B. Zhao, J. Feng, X. Wu, and S. Yan, "A survey on deep learning-based fine-grained object classification and semantic segmentation," *International Journal of Automation and Computing*, vol. 14, no. 2, pp. 119–135, 2017.

[20] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul, "Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields," *Remote Sensing*, vol. 9, no. 7, p. 680, 2017.

[21] C. Cui and T. Fearn, "Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 182, pp. 9–20, 2018.

[22] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, article 014006, 2019.

[23] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," 2017, https://arxiv.org/abs/1707.04912.

[24] S. Shrestha and L. Vanneschi, "Improved fully convolutional network with conditional random fields for building extraction," *Remote Sensing*, vol. 10, no. 7, p. 1135, 2018.

[25] S. Stolte and R. Fang, "A survey on medical image analysis in diabetic retinopathy," *Medical Image Analysis*, vol. 64, p. 101742, 2020.

[26] S. Ahlawat, A. Choudhary, A. Nayyar, S. Singh, and B. Yoon, "Improved handwritten digit recognition using convolutional neural networks (CNN)," *Sensors*, vol. 20, no. 12, p. 3344, 2020.

[27] Y. Zhu and N. Zabaras, "Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification," *Journal of Computational Physics*, vol. 366, pp. 415–447, 2018.

[28] N. Ibtehaz and M. S. Rahman, "MultiResUNet : rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.

[29] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[30] M. H. Yap, G. Pons, J. Martí et al., "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2018.

[31] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: a deformable network for retinal vessel segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149–162, 2019.

[32] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Annual conference on medical image understanding and analysis*, pp. 506–517, Springer, Cham, 2017.

[33] Y. Guo and Y. Peng, "BSCN: bidirectional symmetric cascade network for retinal vessel segmentation," *BMC Medical Imaging*, vol. 20, no. 1, pp. 1–22, 2020.

[34] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," 2018, https://arxiv.org/abs/1802.06955.

[35] 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*, pp. 179–187, Springer, Cham, 2016.

[36] D. Müller, U. Netzelmann, and B. Valeske, "Defect shape detection and defect reconstruction in active thermography by means of two-dimensional convolutional neural network as well as spatiotemporal convolutional LSTM network," *Quantitative InfraRed Thermography Journal*, vol. 19, no. 2, pp. 126–144, 2022.

[37] Y. Wu, Y. Lin, Z. Zhou, D. C. Bolton, J. Liu, and P. Johnson, "DeepDetect: a cascaded region-based densely connected network for seismic event detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 62–75, 2019.

[38] I. Gitman and B. Ginsburg, "Comparison of batch normalization and weight normalization algorithms for the large-scale image classification," 2017, https://arxiv.org/abs/1709.08145.

[39] M. Asadi-Aghbolaghi, R. Azad, M. Fathy, and S. Escalera, "Multi-level context gating of embedded collective knowledge for medical image segmentation," 2020, https://arxiv.org/abs/2003.05056.

[40] Z. Li, K. Gavrilyuk, E. Gavves, M. Jain, and C. G. Snoek, "Videolstm convolves, attends and flows for action recognition," *Computer Vision and Image Understanding*, vol. 166, pp. 41–50, 2018.

[41] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction," 2018, https://arxiv.org/abs/1801.02143.

[42] N. Tamim, M. Elshrkawey, G. Abdel Azim, and H. Nassar, "Retinal blood vessel segmentation using hybrid features and multi-layer perceptron neural networks," *Symmetry*, vol. 12, no. 6, p. 894, 2020.

[43] B. M. S. Rani, V. R. Ratna, V. P. Srinivasan, S. Thenmalar, and R. Kanimozhi, "Disease prediction based retinal segmentation using bi-directional ConvLSTMU-Net," *Journal of Ambient Intelligence and Humanized Computing*, 2021.

[44] B. Henda, E. Yaroub, A. Mohamed, M. H. Bedoui, K. Rostom, and B. A. Asma, "Fast and efficient retinal blood vessel segmentation method based on deep learning network," *Computerized Medical Imaging and Graphics*, vol. 90., article 101902, 2021.