RESEARCH ARTICLE

Redefining Retinal Lesion Segmentation: A Quantum Leap With DL-UNet Enhanced Auto Encoder-Decoder for Fundus Image Analysis

B. NAVEEN KUMAR[1], T. R. MAHESH [1], (Senior Member, IEEE), G. GEETHA[1], (Member, IEEE), AND SURESH GULUWADI [2]

[1]Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru 562112, India
[2]Department of Mechanical Engineering, Adama Science and Technology University, Adama 302120, Ethiopia

Corresponding author: Suresh Guluwadi (suresh.guluwadi@astu.edu.et)

ABSTRACT The first diagnosis of diabetic retinopathy (DR) must include lesion segmentation. As it takes a lot of time and effort to label lesions, automatic segmentation methods have to be created manually. The degree of the retina's degenerative lesions determines how severe diabetic retinopathy is. A major influence is on the early detection of illness and treatment of DR. To reliably identify the sites of related lesions and identify various abnormalities in retinal fundus pictures, deep learning algorithms are crucial. Additionally, utilizing patch-based analysis, a deep convolutional neural network is constructed. In this study, encoder-decoder neural networks along with channel-wise spatial Attention Mechanisms are proposed. The IDRiD dataset, which includes hard exudate segmentations, is used to train and evaluate the architecture. In this method, image patches are created using the sliding window technique. To determine the effectiveness of the recommended strategy, a thorough experiment was conducted on IDRiD. In order to predict the various sorts of lesions, the trained network analyses the picture patches and creates a probability map. This technique's efficacy and supremacy are confirmed by the expected accuracy of 99.94 %. The findings of this experiment show significantly enhanced performance in terms of accuracy when compared to prior research on comparable tasks.

INDEX TERMS Diabetic retinopathy, patch generation, fundus images, lesion segmentation, encoder-decoder network, convolutional neural network, spatial attention mechanism, deep learning.

I. INTRODUCTION

Diabetic Retinopathy (DR) is the primary cause of lifelong blindness and visual loss in young and middle-aged persons. In 2010, it was estimated that 438 million people worldwide would have diabetes by 2025. But before five years, that estimate had already surpassed 25 million. According to the International Diabetes Federation (IDF), there will be ''578 million adults affected by 2030 and 700 million by 2045'' [1]. However, studies have shown that continuous diabetes therapy and blood glucose management can dramatically reduce the course of DR-induced blindness.

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang [ID].

In order to make the analysis more sophisticated and convenient for people in remote areas, numerous works have developed algorithms for examining fundus pictures and automatically detecting symptomatic indicators. There are differences in the texture, size, color, and form of those indicators, making this exceedingly complex. Because retinal lesions come in a variety of sizes, colors, and forms, an ophthalmologist must do a labor-intensive investigation of them. Due to considerable improvements in their accuracy, older techniques for DR segmentation are being replaced by machine learning methodologies.

The fundus picture is frequently employed in numerous medical diagnostics. The crucial element in determining the retinal pathology is picture segmentation in the fundus image. Ophthalmologists are able to diagnose retinal diseases thanks

to the examination of human retinas. Segmenting the retinal vessels first, then properly analyzing the retinal blood vessels, can help to identify the alterations in the blood vessel and the retinal disease.

To detect eye diseases like diabetic retinopathy early, automatic retinal vascular segmentation is essential. Several segmentation methods, including a piecewise threshold, can be used to separate images. Further segmentation techniques for the retinal arteries. Discover changes brought on by retinal disorders, such as diabetic retinopathy, in the vessels. The image is represented as a Max-Tree, and the vessels are separated Study of image formation is the primary application of mathematical morphology. The open and closed are built on top of these steps. Combining these algorithms makes it possible to eliminate the background. Vascular segmentation determines if a pixel is a vessel or not recovered after the algorithm has been applied and the image has been firstly smoothed.

An ophthalmologist may find and diagnose these disorders, as well as others that affect the heart and brain, utilizing aberrant anomalies in the retinal vascular structure. As a result, changes to the arterioles and veins' anatomical structure in the retina have important diagnostic implications. For photographing the retina, fundus cameras, an optical instrument, are typically utilized. The reference fundus camera may be compared to a low-power microscope that is specifically designed for photographing the retina. The linked camera takes pictures when the retina is illuminated.

The process of fundus photography may be conceived of as being used to record the inner retinal structure and retinal neurosensory tissues. The retinal neurosensory tissues transform the optical reflections we perceive foundation of retina photography, which may enter and depart through the eye pupil.

Oculists scan their patients' retinas using HD cameras. Therefore, diagnosing retinal diseases involves looking at the retina's blood vessel health. There are several cases where it is found that there is minimal distinction between the retinal vascular structure and the surroundings. As a result, diagnosing retinal problems is challenging. Applying the right image segmentation technique is essential for the very precise detection of retinal vascular structures, which in turn aids in diagnosis. The following list of challenges that might come up when locating and removing retinal vessels. First off, the retinal image's retinal vessels' widths can have color intensities ranging. An improved vessel picture is produced by controlled uneven lighting and noise levels. Figure 10 shows the final morphological picture for each channel, which makes it simpler to see the vessels. However, an imaging method cannot be regarded as the best instrument for development. After contrasting the two approaches, we considered homomorphic filtering before settling on the best operating base. Homomorphic filtering is discussed in the paragraph that follows. The effectiveness of our suggested retinal segmentation approach was assessed using the most used measurement parameters. A few of these measuring criteria include recall, specificity, and accuracy. Accuracy provides all of the pixels in the segmented vessels' information.

The experiment's findings, which make up our contribution, significantly surpass previous research on issues of a similar kind in terms of accuracy. Due to the time-consuming manual labeling of lesions by professionals, automatic segmentation methods have to be developed. The severity of diabetic retinopathy is based on the size of the degenerative lesions in the retina. Despite this, the surgery is difficult because of the tiny size and poor local contrast of the lesions shown in the imaging. Deep learning-based automated segmentation of retinal lesions has a significant impact on early disease identification and DR treatment. The trained network evaluates the picture patches with the aim of predicting the different types of lesions. Our purpose is to create a probability map. The method's efficacy and superiority are demonstrated. Additionally, a deep convolutional neural network is built using patch-based analysis of tiny retinal. The model was trained using images of the retinal fundus resulting from minute lesions. As a result, our proposed study improves the precise answer when assessing the effectiveness and superiority of this strategy.

The main objectives of the research on small lesions are

- Network analysis for lesion prediction: The study uses network analysis techniques to analyze picture patches in order to forecast different types of lesions. This approach can provide valuable insights into lesion detection and prediction in retinal fundus images.

- Patch-based analysis for lesion segmentation: The study proposes a deep convolutional neural network (CNN) including U-Net and Autoencoder along with channel-wise spatial Attention Mechanism, that uses patch-based analysis for segmenting tiny lesions in retinal fundus images. This approach can improve the accuracy of lesion segmentation, which is an important step in diagnosing and monitoring retinal diseases.

- Significant accuracy improvement: Through experimental evaluation, we demonstrate that our proposed method achieves a substantial improvement in the accuracy of lesion segmentation compared to existing studies. This significant improvement in accuracy can contribute to the reliability and effectiveness of the proposed method for clinical applications, such as early detection and monitoring of retinal lesions.

- Value and superiority assessment: The study assesses the value and superiority of the proposed method by achieving a high accuracy improvement. This assessment can provide evidence of the effectiveness of the proposed method and its potential for outperforming existing methods in lesion segmentation in retinal fundus images.

In summary, this paper introduces the "channel-wise spatial Attention Mechanism" as a novel contribution to the field of small lesion analysis in retinal fundus images. Our objectives encompass network analysis for lesion prediction, patch-based analysis for lesion segmentation, achieving significant accuracy improvement, and assessing the value and superiority of our proposed method. By addressing these objectives, we aim to advance the field and contribute to the early detection and monitoring of retinal lesions, ultimately improving patient care and outcomes.

II. REVIEW OF THE RELATED WORK

In order to analyze retinal images, deep learning's theory, and methods are thoroughly reviewed in this work. Without a good clinical diagnosis and medical care, many eye illnesses frequently result in blindness. One such condition that damages the retinal blood vessels in human eyes is diabetic retinopathy (DR), for instance. Based on their expert knowledge, ophthalmologists perform a labor-intensive DR diagnosis. These techniques effectively distinguish abnormal blood vessels by visual analysis, outperforming manual operations. More than 80 publications on the identification of retinal vessels in the field of inquiry, from segmentation to classification, are included in the pertinent articles published since 1982.

DR, which develops when the blood capillaries that supply the retina of the eye deteriorate. The severity of diabetic retinopathy is based on the extent of the degenerative lesions in the retina. In order to stop the disease from progressing and to safeguard eyesight, essential assess symptoms. When compared to earlier research on analogous tasks, the proposed study's findings demonstrate much superior accuracy and recall performance.

One of the study topics that scientists and doctors are paying close attention to is the identification of exudates [1]. We suggest new approaches to exudates identification. The exudates are coarse and fine-segmented after pre-processing techniques. An assortment of properties, including texture, color, size, and edge, are retrieved from the retinal pictures to categorize them into exudates and non-exudates [2]. Additionally, the outcomes are shown to highlight the benefit of the suggested approach in a public database with a 92.14 accuracy rate. Recurrent Residual U-Net is a deep learning technique that is used [3].

Disorder eyes are brought on long-term that affect people all over the world [4]. A crucial and visible DR symptom that might aid in early identification is a hard exudate. In this research, a method for automatically segmenting hard exudates is suggested to help ophthalmologists make an early diagnosis of DR. We overcame the challenge of the small and unbalanced dataset by generating sample patches using the SLIC super pixel technique. Additionally, the loss function is focused loss. To assess the effectiveness of the suggested strategy, several tests have been run on the IDRiD dataset. The reported values for precision, recall, and accuracy are 98.47%, 98.36% and 99.94%, respectively. This shows how
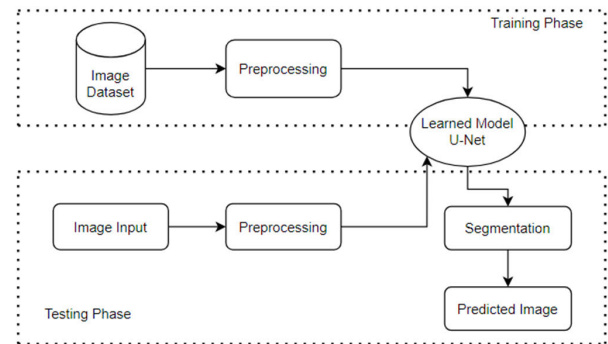

FIGURE 1. Architecture overview.

successful and excellent our approach is. The segmentation findings obtained demonstrate the method's potential for clinical diagnosis [5]. For better image map results, Opella and Hernandez [6] investigated fusing ConvNet, a feed-forward neural network with a focus on image processing and prediction, with SVM, a supervised machine learning for classification and regression analysis. U-net: Convolutional networks for biomedical image segmentation was proposed by Ronneberger et al. [7]. According to Pizer et al. [8], clipped adaptive histogram equalization should replace previous digital imaging. IDRiD dataset was suggested by Porwal et al. [9] to be integrated into a database for research on diabetic retinopathy screening Silva et al. [10]. The study on diagnostic assessment of deep learning algorithms for diabetic retinopathy screening, highlighting the effectiveness of these algorithms in improving screening accuracy and efficiency was conducted by Li et al. [28].

III. METHODOLOGY

In this study, we propose to segment retinal lesions in fundus images utilizing the encoder-decoder neural network in the suggested system due to the fact that the automated segmentation of retinal lesions employing a deep learning approach has a significant influence on the early identification and treatment of DR. The overview architecture of retinal lesions segmentation of fundus images is depicted in Figure 1.

With the given framework for lesion segmentation using CNN, a number of steps will be conducted. A collection of pre-processed images and, after training, examines the pre-processed patches using the sliding window method to create a probability map. With the use of the sliding window method, patches were extracted with a resolution of 512×512 (160 patches per image) for both the original and the ground truth images. These patches were then utilized as input for CNN's training. To improve understanding of the picture patches and reinforce the model, patches with random right, left, up, and down flips were included. The suggested CNN design is a 13-layer encoder-decoder network. 3×3 filters with Batch Normalization and Rectified Linear Unit (ReLU) activation functions make up the convolution layer. Batch
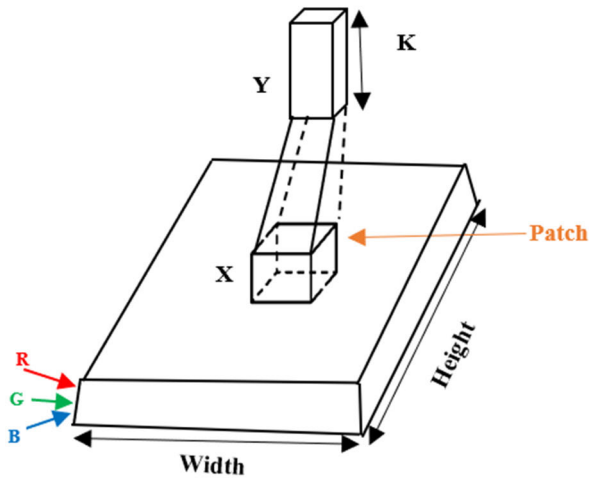
normalization significantly accelerates a network's training process.

A. CONVOLUTION NEURAL NETWORK (CNN)

The reader is presumed to be familiar with neural networks as a concept. Artificial neural networks do amazingly well when it comes to machine learning. Various categorization tasks, such as those involving text, sounds, and images, employ artificial neural networks for varied purposes, and multiple neural network types are used. We're going to create a foundational component.

1. *Input Layers:* Our data's total number of features is equal to this layer's total number of neurons.
2. *Hidden Layer:* There are often more neurons per hidden layer than there are features, however this is not always the case. Learnable biases are added to the network to make it nonlinear, and then an activation function is applied.
3. *Output Layer:* The output of each layer is then determined when the model has been fed the data. In the following stage, known as feed-forward, we use an error function to determine the error. Cross-entropy, square loss error, and other examples of standard error functions. The derivatives are then calculated and again propagated.
4. *CNN:* Convolution Networks that share their parameters are known as convnets or neural networks.

Let's now discuss some mathematics that is utilized throughout the convolution procedure.

Learnable filters make up the convolution layers. Figure 2 depicts the deep learning udacity.

As an illustration, as tiny as but not as large as the picture dimension.

B. CONVNETS CONSTRUCTION LAYERS

Every layer in a convnent's series performs a differentiable function to turn one volume into another.

Let's use an image with the dimensions 32 by 32 by 3 as an example and run a convnets on it.

- *Input Layer:* This layer contains the image's unprocessed input, which has the following dimensions: 32 wide, 32 high, and 3 deep.
- *Convolution Layer:* By calculating this layer controls the output volume and is sandwiched between all filters and image patches. The convolution will be subjected to this layer's element-wise activation function. Leaky RELU, Tanh, and RELU: $\max(0, x)$ are a few examples of frequent activation functions. The output volume will have dimensions of 32 by 32 by 12 because the volume is unaltered.
- *Pool Layer:* Injected into convnets on a regular basis, calculation. The terms "max pooling" & "average pooling" refer to types that will produce a volume with dimensions of 16 by 16 by 12.
- *Data Pre-processing:* A common format for passing the data to models, resizing, and scaling is applied to the data.
- *Dividing Data:* Separating data into training and test datasets.
- *Model training:* Using the pre-processed dataset to train the model and then storing it. H5 Extension
- *User interface:* The front end has a number of pages, including the homepage, the about page, the upload page, and the result page.

C. ATTENTION MECHANISM

The attention mechanism is a mechanism used in deep learning models to allow the model to focus on different parts of the input data with varying levels of importance. It is particularly useful in tasks where the input data has long sequences or multiple parts that need to be processed in parallel. The attention mechanism can be applied to biomedical image segmentation tasks where the goal is to accurately identify and outline specific structures or regions of interest within medical images. The use of attention mechanisms in this context helps the model to focus on the relevant areas of the image and refine the segmentation results.

There are several types of attention mechanisms, Including Self-Attention or Multi-head Attention, Global Attention or Soft Attention, Local Attention or Hard Attention etc. Our study focus and makes use of Spatial attention. Spatial attention mechanism is a type of attention mechanism used in computer vision tasks, particularly in convolutional neural networks (CNNs) for image processing. It allows the model to selectively focus on different spatial regions or regions of interest (ROIs) in an image, based on their relevance or importance for the task at hand. In this also there are sub classifications like channel wise spatial attention mechanism (CWSM), spatial attention with convolution filters and region-based attention. In this study, we have applied CWSM. The channel-wise spatial mechanism (CWSM) is
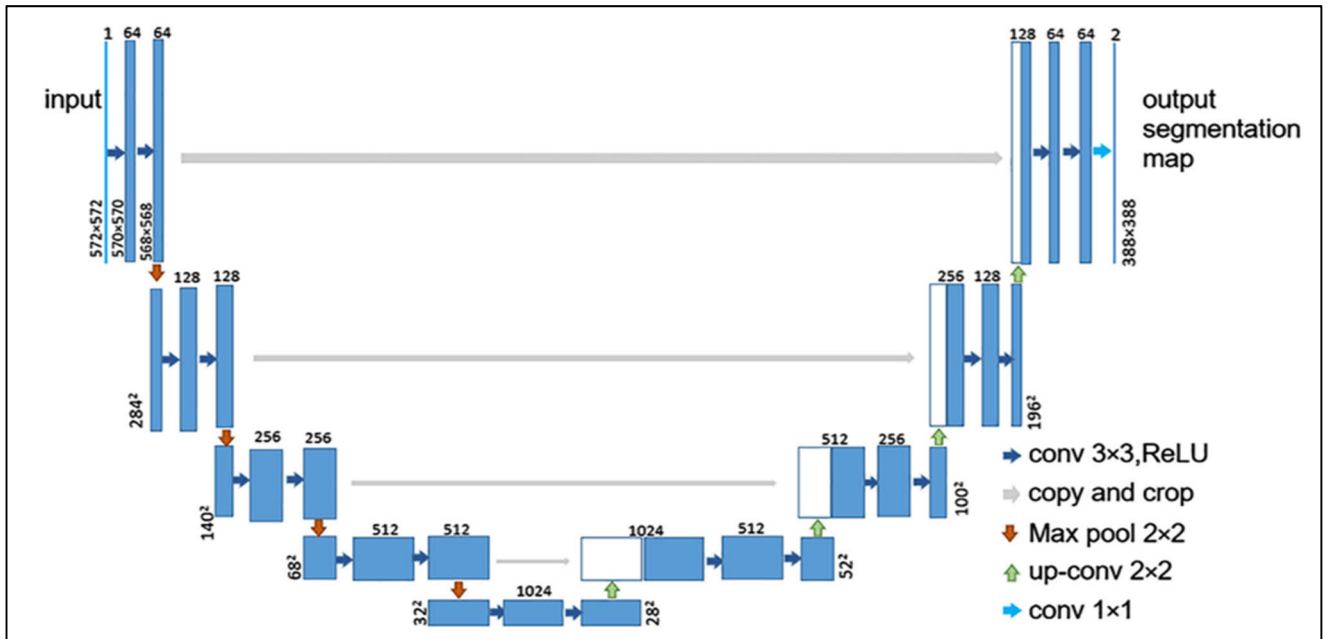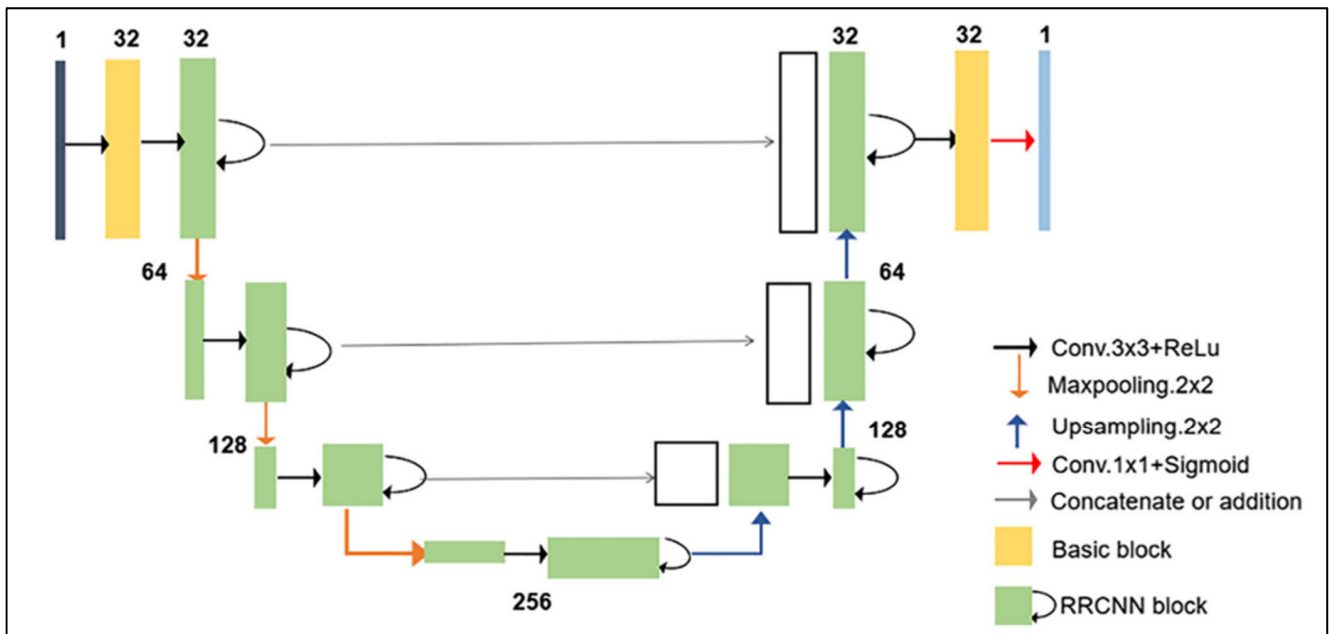
FIGURE 3. U-Net outperform [4].

FIGURE 4. DRU-Net architecture [4].

a technique used in deep learning for object detection and image classification tasks. Here are the general steps for implementing CWSM:

1. Input tensor: The first step is to define the input tensor, which typically has dimensions (batch_size, channels, height, width). The channels dimension represents the number of feature maps in the input tensor.

2. Convolution: Next, a convolution operation is applied to the input tensor using a set of learnable filters or kernels. The convolution operation produces a set of output feature maps,

which represent the response of each filter to different regions of the input image.

3. Spatial pooling: The output feature maps are then passed through a spatial pooling layer, such as max pooling or average pooling, to reduce the spatial dimensions of the feature maps. The pooling operation helps to make the network more robust to small variations in the input image.

4. Channel attention: A channel attention mechanism is then applied to the pooled feature maps to compute a set of channel weights for each feature map. This is typically
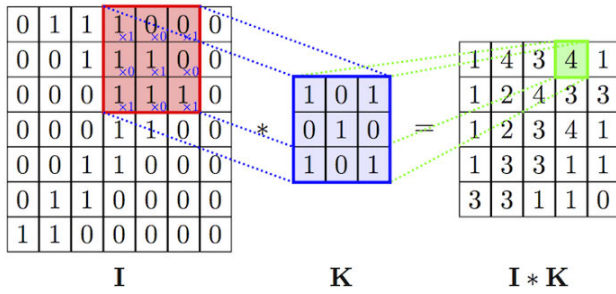
FIGURE 5. Convolution image [11].

done using a shared multi-layer perceptron (MLP) or a similar neural network architecture. The channel weights represent the relative importance of each feature map for the task at hand.

5. Channel-wise multiplication: Finally, the output feature maps are computed by element-wise multiplication of the pooled feature maps with the channel weights. This operation emphasizes the most important feature maps while suppressing the less important ones.

D. ARCHITECTURE OVERVIEW

1) BASIC ARCHITECTURE

It has since been discovered to have applications in a variety of fields. The model seeks output as an image segmentation tool as represented in figure 1. The task of "picture segmentation" has been attempted by several Neural Nets, however U-Net outperforms its forerunners by requiring less CPU power and causing the least amount of information loss shown in figure 3 [4]. To discover more about how U-Net does this, let's take a deeper look. The deep recurrent U-Net is shown in figure 4 [4].

Networks for high-resolution photos would be exceedingly computationally costly if only fully linked layers were used to build them. Convolution, a mathematical procedure, is a hero in the Computer Vision tale because of this. To lower the cost of computing, convolution preserves all of the input pixels' effect while keeping them only loosely related.

The steps below should be repeated for the complete input picture matrix to perform a convolution operation:

Step 1: Consider a filter matrices K that are smaller in size than the input image matrices I. To get a single value for the output matrix, multiply the components that are superimposed element-by-element and then add.

Step 2: In accordance with the specified stride, move the filter to the columns to the right and repeat step 1.

For instance, if stride is 3 and column 1 was the starting point for the operation, we would travel to column 4 and repeat Step 1 there. Figure 5 shows the convolution image [11].

2) POOLING (MAX POOL)

Convolution and pooling are both used to decrease the number of parameters and speed up calculation, respectively.
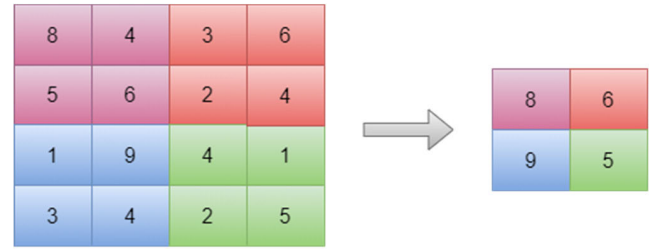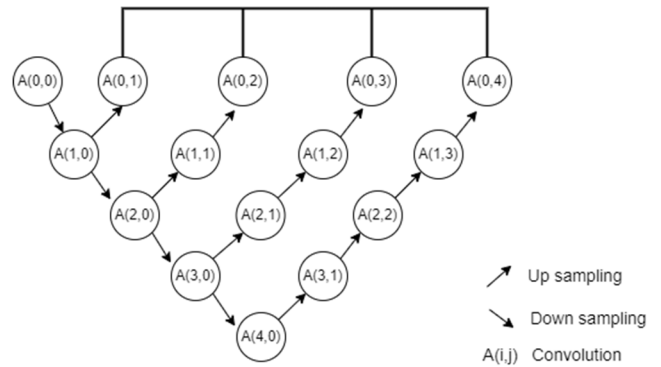

FIGURE 6. Max pool.

FIGURE 7. Image segmentation.

Inadvertently, the layer also permits some regularization. Average or maximum are the two procedures that are commonly carried out during pooling. To apply these functions (max or average), we first build subsets of the input depending on the filter size ('f') and stride ('s').

As shown in figure 6, using the input x and the max() function across the set of 0.0 from Equation below, we can theoretically describe this function $f()$ as follows:

$$f(x) = \max\{0.0, x\} \qquad (1)$$

Max As seen in figure 6, pooling layers has the goal of lowering the size of feature maps with a 22-kernel size for faster convergence. Loss is used as a loss function in this procedure because, with this Loss, Equation below provides the Cross Entropy formula:

$$CE = -\sum_{i=1}^{C'=2} t_i \log\left(f\left(s_i\right)\right) = -t_1 \log\left(f\left(s_1\right)\right)$$
$$- \left(1 - t_1\right) \log\left(1 - f\left(s_1\right)\right) \qquad (2)$$

The loss is calculated by Adam optimizer is used to reduce it while aiding in the updating. To accurately segment retinal fundus from the fundus image, the trained model's probability map is used.

3) EXACTION ARCHITECTURE

Exaction architecture is a practice that satisfies both functional and expressive needs, and as a result, it fulfils both utilitarian and aesthetic goals. Below is a presentation of the biomedical segmentation's architecture and its newly created step paths.

Biomedical Image Segmentation

(*a*) Original Retina Image   (*b*) Ground Truth/Mask Image   (*c*) Predicted/Segmented Image

FIGURE 8. Images at different stages.

- Updated Skip Pathways
- Thorough Supervision
- Experimental Findings

The image segmentation is presented in the figure 7.

Re-designed Skip Pathways

The first picture depicts the skip pathway among nodes A(0,0) and A(1,3).

The output from the previous convolution layer of the same dense block is combined with the matching up-sampled output of the lower dense block in the concatenation layer that comes before each convolution layer, and we can formulate as follows.

$$a^{i,j} = \begin{cases} \mathcal{H}\left(a^{i-1,j}\right), & j = 0 \\ \mathcal{H}\left(\left[\left[a^{i,k}\right]_{k=0}^{j-1}, \mathcal{U}\left(a^{i+1,j-1}\right)\right]\right), & j > 0 \end{cases} \quad (3)$$

E. METRICS

Three distinct methods were used to assess the segmentation models' performance [26].

Pixel Level, Exudate (or Lesion) Level, and Picture Level were the three levels at which the metrics were computed.

1) PICTURE SIZE

The accepted method [27] for evaluating segmentation results is at the pixel level. To determine if a pixel is True Positive (TP), False Positive (FP), True Negative (TN), or False Negative, every pixel from the segmented output is compared with the appropriate ground truth pixel (FN). Mistakenly classified as positive or negative, respectively, is denoted by FP and FN, respectively.

2) SENSITIVITY (OR RECALL)

Recall measures how well a learning model categorizes events that are positive. The number of accurate positive cases that the model correctly detected is denoted by a term called True Positive Rate (TPR) or Recall. Recall is calculated as the proportion of properly detected vessels to all vessels, where

$$\begin{aligned} \text{TPR} &= \frac{\text{Number of correctly detected nuclei}}{\text{Total number of automatically detected nuclei}} \\ &= \frac{TP}{TP + FN} \end{aligned} \quad (4)$$

$$\begin{aligned} \text{FPR} &= \frac{\text{Number of detected nuclei}}{\text{Total number of patches without nuclei}} \\ &= \frac{\text{FP}}{\text{FP} + \text{TN}} \end{aligned} \quad (5)$$

3) PRECISION

The precision p(r) is a function of recall.

$$\begin{aligned} \text{Precision} &= \frac{\text{Number of correctly detected nuclei}}{\text{Number of manually identified nuclei}} \\ &= \frac{TP}{TP + FP} \end{aligned} \quad (6)$$

4) ACCURACY

The percentage of occurrences that successfully classifies the whole sample count that is computed. Accuracy is determined by dividing the total number of properly detected vessels and non-vessels by the total number of pixels.
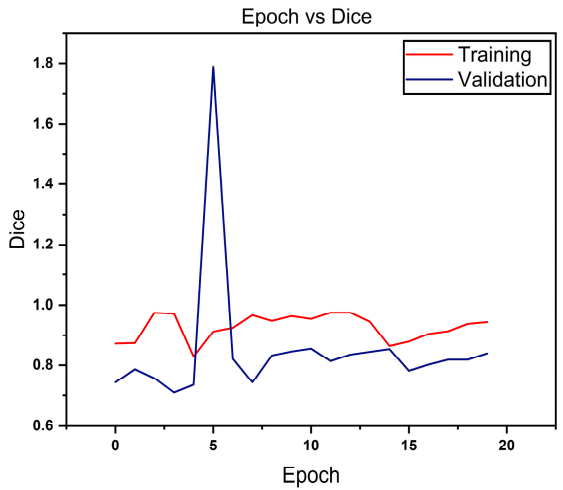
$$\text{Accuracy}(AC) = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

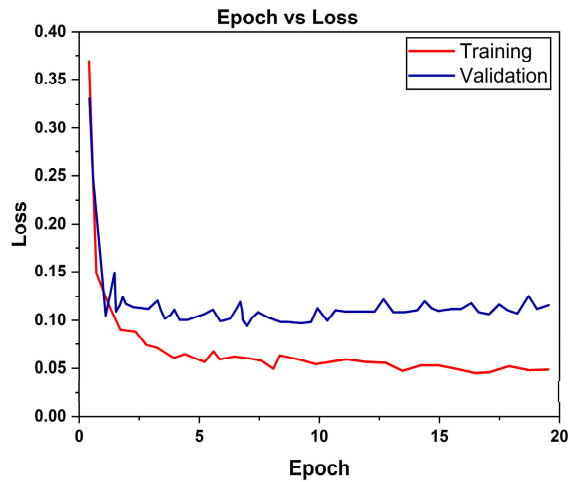#### 5) INTERSECTION OVER UNION (IoU)

Intersection over Union (IoU), a measure for segmentation problems, tries to increase the intersection between the segmented pixel and the ground truth and limit the union between them.

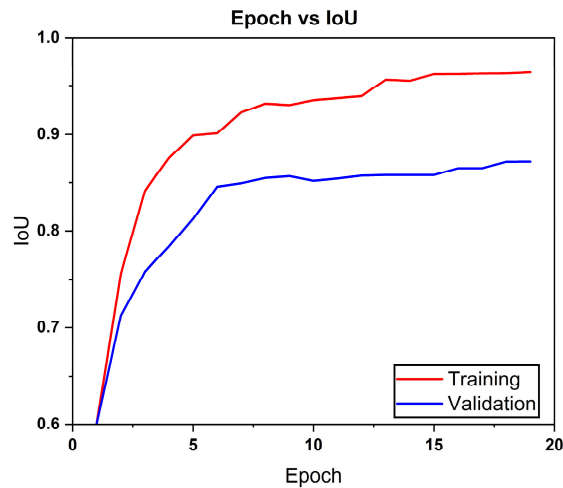$$\text{IoU} = \frac{|G \cap P|}{|G \cup P|} \quad (8)$$

where P is the set of predicted pixels and G is the set of ground truth pixels.
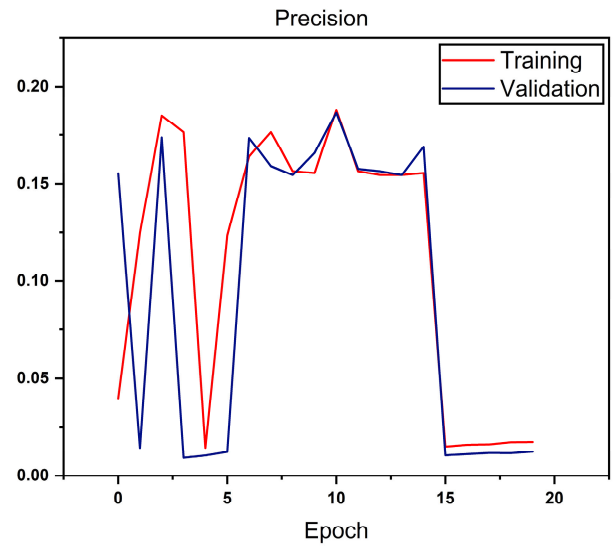
(a) Epoch vs Dice

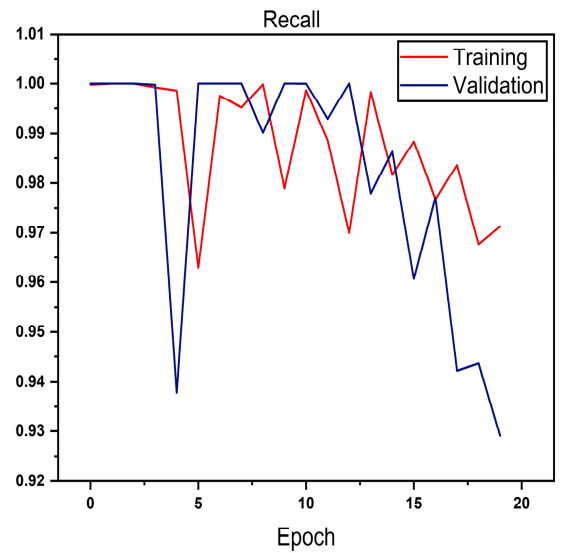(b) Epoch vs Loss

(c) Epoch vs IoU

FIGURE 9. Epoch vs dice, loss and IoU.

(a) Precision

(b) Recall

FIGURE 10. Precision and recall.

harmonic mean of these two metrics.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \qquad (9)$$

7) LOSS FUNCTION
We employ a binary cross-entropy loss function for training the U-Net model. The loss function's formula is constructed as follows:

$$L = -\frac{1}{n} \sum_{i=1}^{n} \hat{y}_i \log y_i + \left(1 - \hat{y}_i\right) \log \left(1 - y_i\right) \qquad (10)$$

They are as follows: $\hat{y}_i$ is the actual output, and its value is $\hat{y}_i \in \{0, 1\}$; y_i is the predicted output, or the genuine data label.

6) DICE COEFFICIENT
The DICE coefficient, which balances the trade-off between accuracy and recall by giving the quantity a lower score, is the

8) DICE LOSS

While backpropagation and DL minimize the metric and reduce the class-imbalance problem, respectively, they also ensure network convergence for efficient training. DL is represented mathematically as follows.

$$DL = 1 - \left(\frac{2 \times \sum_k^A P_{Pro-k} G_{Truth-k}}{\sum_k^A P_{Pro-k}^2 + \sum_k^A G_{True-k}^2} \right) \quad (11)$$

F. IDRID DATASET

The photos were shot with a Kowa VX-10. The photos are saved as jpg files and have a size of 4288 × 2848 pixels. Anomalies are related to assessing the efficacy of specific lesion segmentation techniques [9]. First, ADCIS created specialized software to execute annotation at the pixel level.

The IDRiD (Indian Diabetic Retinopathy Image Dataset) was explicitly developed to address the need for a publicly available dataset that focuses on diabetic retinopathy. There are several reasons why the IDRiD dataset was created and why it is valuable for diabetic retinopathy research:

1. Representation of Indian Population: The IDRiD dataset is unique in that it represents the Indian population, which is known to have a high prevalence of diabetes. By focusing on this specific population, the dataset provides a more relevant and representative sample of retinal images for studying diabetic retinopathy in an Indian context.

2. Diversity of Lesions: The IDRiD dataset covers a range of diabetic retinopathy-related lesions, including exudates, haemorrhages, and microaneurysms. This diversity allows researchers to develop and evaluate algorithms that can effectively detect and segment these lesions, which are important indicators for disease diagnosis and progression monitoring.

3. Annotated Ground Truth: Each image in the IDRiD dataset comes with expert annotations for various retinal structures and lesions, such as the optic disc, fovea, exudates, hemorrhages, and microaneurysms. These annotations serve as ground truth labels, enabling the training and evaluation of algorithms for automatic lesion detection and segmentation.

4. Benchmark for Comparison: The availability of a standardized dataset like IDRiD provides a common benchmark for researchers working on diabetic retinopathy. It allows for fair comparisons between different methods, facilitating the evaluation and advancement of algorithms for detection, classification, and segmentation tasks.

5. Advancing Diabetic Retinopathy Research: The IDRiD dataset plays a crucial role in advancing research on diabetic retinopathy. It provides researchers with the necessary data to develop and test novel algorithms and models for early detection, accurate diagnosis, and effective monitoring of diabetic retinopathy. This, in turn, contributes to the development of more efficient and reliable tools for healthcare professionals in managing the disease.

TABLE 1. Experimental outputs.

Epoch	dice_coef	val_dice	loss	val_loss	IoU	val_IoU
1	0.87437	0.78641	0.36923	0.33125	0.60125	0.60092
2	0.97513	0.75634	0.0902	0.11384	0.7557	0.71261
3	0.97183	0.70924	0.07121	0.11125	0.84169	0.75805
4	0.82958	0.73482	0.06442	0.10038	0.87529	0.78391
5	0.91234	1.7896	0.05637	0.10642	0.89889	0.8128
6	0.92467	0.82137	0.06178	0.09422	0.90124	0.84581
7	0.96725	0.74315	0.05763	0.10302	0.92308	0.84933
8	0.94783	0.83256	0.04958	0.09825	0.93189	0.85472
9	0.96409	0.84562	0.05901	0.11258	0.93037	0.85648
10	0.95455	0.85546	0.05637	0.11044	0.93542	0.85178
11	0.97569	0.81422	0.05901	0.10919	0.93745	0.85413
12	0.97564	0.83543	0.05562	0.12201	0.93953	0.85707
13	0.94532	0.84463	0.04757	0.10843	0.95608	0.85766
14	0.86430	0.85377	0.05298	0.11258	0.95491	0.85766
15	0.87932	0.78158	0.04958	0.11183	0.96196	0.85766
16	0.90470	0.80217	0.04543	0.10843	0.96203	0.8645
17	0.91326	0.81869	0.04619	0.11044	0.96255	0.86456
18	0.93788	0.81892	0.0482	0.10705	0.96273	0.87127
19	0.94384	0.83948	0.04883	0.11183	0.96393	0.87139

IV. RESULTS AND DISCUSSION

For both training and testing, the model has been applied to the Indian Diabetic Retinopathy Imaging Dataset (IDRiD). 4288*2848-pixel resolution and jpg file format was used to record all of the photos while centring them on the macula. 81 color fundus photos with 81 different types of DR signals make up collection (HE).

Our system was tested using a total of 40 photographs divided into training and test sets, each of which had 20 images of normal and abnormal cases with ages ranging from 31 to 86, to assess the vascular segmentation technique. The effectiveness of our system is assessed and confirmed using a variety of different methodologies. The true positive rate measures the proportion of pixels that were correctly classified as vascular pixels (TPR). The false positive rate, or FPR, is a measure of the proportion of pixels that were wrongly identified as vessel pixels. The total number of pixels in the digital images divided by the sum of true positive (pixels that are properly designated as vessels) and true negative pixels is the definition of accuracy (pixels that are accurately labeled as non-vessels but are truly non-vessel points). Compared to morphological processing approaches, the recommended method offers exceptionally great accuracy. As the dataset's number of images increases, accuracy increases with a very low misclassification rate. In both cases, the recommended models perform better.

Unequivocally how reliable the suggested models are for comprehensive. In addition, throughout the training and validation phases, we also computed MSE errors. The

TABLE 2. Experimental outputs for precision, recall, validation precision, validation recall.

Epoch	Precision	val_precis	Recall	val_recall
0	0.03970	0.155487	0.999756	1
1	0.12470	0.014375	1	1
2	0.184937	0.17364	1	1
3	0.17634	0.009283	0.999203	0.999739
4	0.014423	0.010503	0.998554	0.93785
5	0.12348	0.012374	0.962943	1.
6	0.16385	0.17329	0.997522	1
7	0.176342	0.159136	0.995184	1
8	0.15655	0.15465	0.999823	0.990195
9	0.15565	0.16564	0.978964	1
10	0.18785	0.186487	0.998643	0.999967
11	0.15655	0.157686	0.988671	0.992964
12	0.15463	0.15653	0.969995	1
13	0.15463	0.15463	0.998289	0.97798
14	0.15566	0.16878	0.981619	0.986319
15	0.015099	0.010582	0.988405	0.960651
16	0.016013	0.011194	0.97681	0.977199
17	0.016217	0.011749	0.983596	0.94228
18	0.017355	0.011739	0.967693	0.943713
19	0.017499	0.012495	0.971293	0.929121

TABLE 3. Comparison of performance analysis.

References	Accuracy (%)	Precision	Recall
Zong et al. [12]	97.95	-	96.38
Lee and Ke [13]	97.21	-	70.21
Xu et al. [14]	78.83	88.84	82.85
Sudheer and Bindu [15]	99.70	-	92.81
Alom et al. [16]	99.18		98.32
Raza et al. [17]	99.91		90.40
Shan and Li [18]	91.38	91.57	91.16
Khojasteh et al. [19]	98.0	-	96.0
Dayana and Emmanuel [20]	99.0	-	97.0
Sambyal et al. [21]	99.88	-	**99.85**
Kou et al. [4]	99.9	-	93.0
Al-Hazaimeh et al. [22]	98.4	96	99
Gharaibeh et al. [23]	98.60	99	99
Al-N'awashi et al. [24]	98.80	96.40	99.20
Alshorman et al. [25]	93.50	91.70	89.20
Proposed study	**99.94**	**98.47**	**98.36**

validation loss, IoU, and validation IoU (val_IoU). Table 2 presents precision, recall, validation precision (val_precis), and validation recall (val_recall).

A. QUANTITATIVE RESULTS

Table 3 presents the attained Accuracy, Precision, and Recall. The quantitative measures of the proposed study are also contrasted with previously existing algorithms in this table. As shown in the table, the suggested network outperformed all prior research in terms of segmenting hard exudates using patch-based techniques.

V. CONCLUSION

In this study, we were able to successfully segment retinal lesions photographs with the use of the segmentation technique. Moreover, the veins that go to the retinal vasculature are becoming segmented. We evaluated the system by uploading the image and segmenting it after the training. Long-term diabetes causes diabetic retinopathy, which is the main factor in permanent vision loss globally. This makes it necessary to employ an artificial segmentation technique based on deep learning to correctly identify the severity level of DR patients. Using IDRiD datasets, a technique for segmenting retinal lesions such as microaneurysm and hard exudates at the pixel level has been developed in this study. An encoder that has been pre-trained to segregate hard exudates and micro aneurysms from the fundus dataset. To eventually forecast a group segmentation map, the model integrates geographic data from the down sampling. The main challenge is Autoencoder architectures can be complex Determining the optimal architecture, including the number of layers, layer sizes, and activation functions, requires extensive experimentation and parameter tuning.

training accuracy of the suggested models R2U-Net and RU-Net was compared with that of ResU-Net and U-Net. The outcome unequivocally demonstrates how reliable the suggested segmentation approach is. The subsequent images graphically depict the development of our proposed study as shown in figure 8.

This measure demonstrated outstanding performance on a sizable data set. Also, the system attained average scores of 98.36% recall. With the least amount of accuracy and recall trade-off, this value reflected the highest point on the receiver operating characteristic curve. The experimental outcomes of precision, recall, epoch vs dice, loss and IoU is shown in figures 9 and 10.

Our system correctly identified aberrant areas in the input fundus pictures in addition to image categorization, enabling clinical assessment and confirmation of the automated diagnosis. Our trials using held-out data show that deep learning systems have the capacity to effectively model and forecast illness in fundus pictures. We discovered that our results supported those of a recent deep learning study on automated DR assessment. Table 1 depicts the results of dice coefficient (dice_coef), validation coefficient (val_dice), loss,

The proposed architecture achieves microaneurysm and exudates segmentation accuracy of 99.94% and a dice score of 0.9998 using the IDRiD dataset. The proposed study outperforms the strategies currently recommended in the literature and gives cutting-edge results for the segmentation of both microaneurysms and hard exudates. On the clinically collected datasets, however, the work has not yet been tested. Architectures typically operate at the pixel level and do not explicitly consider the spatial relationships between pixels and also it can be sensitive to input variations, such as changes in lighting conditions, image quality, or artifacts present in fundus images. Retinal lesions are often rare compared to the background pixels in fundus images, resulting in a severe class imbalance. In retinal images, lesions often exhibit specific spatial patterns or structures. In light of this, the suggested study may be expanded in the future to include the detection of other retinal abnormalities, such as haemorrhage and soft exudates, using benchmark fundus picture sets obtained clinically.

Data Availability Statement: Data availability is mentioned in this manuscript.

Conflicts of Interest: There is no conflict of interest among the authors.

REFERENCES

[1] T. A. Soomro, A. J. Afifi, L. Zheng, S. Soomro, J. Gao, O. Hellwich, and M. Paul, "Deep learning models for retinal blood vessels segmentation: A review," *IEEE Access*, vol. 7, pp. 71696–71717, 2019.

[2] K. Wisaeng and W. Sa-Ngiamvibool, "Automatic detection of exudates in retinal images using region-based, neighborhood and block operation," *J. Comput. Sci.*, vol. 14, no. 4, pp. 438–452, Apr. 2018, doi: 10.3844/jcssp.2018.438.452.

[3] K. Vora, D. Mehta, D. Thakker, N. Mehendale, "A deep learning based approach to segment exudates in retinal fundus images using recurrent residual U-Net," *TechRxiv*, 2022, doi: 10.36227/techrxiv.21196657.v3.

[4] C. Kou, W. Li, W. Liang, Z. Yu, and J. Hao, "Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network," *J. Med. Imag.*, vol. 6, no. 2, 2019, Art. no. 025008, doi: 10.1117/1.JMI.6.2.025008.

[5] G. Kaur and A. Bala, "An efficient automated hybrid algorithm to predict floods in cloud environment," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, May 2019, pp. 1–4, doi: 10.1109/CCECE.2019.8861897.

[6] J. M. A. Opella and A. A. Hernandez, "Developing a flood risk assessment using support vector machine and convolutional neural network: A conceptual framework," in *Proc. IEEE 15th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2019, pp. 260–265, doi: 10.1109/CSPA.2019.8695980.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 18, 2015, pp. 234–241.

[8] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., Image Process.*, vol. 39, pp. 355–368, Sep. 1987, doi: 10.1016/S0734-189X(87)80186-X.

[9] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, Sep. 2018, doi: 10.21227/H25W98.

[10] C. Silva, A. Colomer, and V. Naranjo, "Deep learning-based approach for the semantic segmentation of bright retinal damage," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.*, vol. 19. Cham, Switzerland: Springer, 2018, pp. 164–173.

[11] I. S. Mohamed, "Detection and tracking of pallets using a laser rangefinder and machine learning techniques," Ph.D. dissertation, Eur. Master Adv. Robot., Univ. Genoa, Genoa, Italy, 2017.

[12] Y. Zong, J. Chen, L. Yang, S. Tao, C. Aoma, J. Zhao, and S. Wang, "U-Net based method for automatic hard exudates segmentation in fundus images using inception module and residual connection," *IEEE Access*, vol. 8, pp. 167225–167235, 2020, doi: 10.1109/ACCESS.2020.3023273.

[13] C.-H. Lee and Y.-H. Ke, "Fundus images classification for diabetic retinopathy using deep learning," in *Proc. 13th Int. Conf. Comput. Modeling Simulation*, Jun. 2021, pp. 264–270.

[14] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[15] E. S. Kumar and C. S. Bindu, "Segmentation of retinal lesions in fundus images: A patch based approach using encoder–decoder neural network," in *Proc. 7th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, Mar. 2021, pp. 1247–1253.

[16] M. Zahangir 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, *arXiv:1802.06955*.

[17] A. Raza, S. Adnan, M. Ishaq, H. S. Kim, R. A. Naqvi, and S.-W. Lee, "Assisting glaucoma screening process using feature excitation and information aggregation techniques in retinal fundus images," *Mathematics*, vol. 11, no. 2, p. 257, Jan. 2023.

[18] J. Shan and L. Li, "A deep learning method for microaneurysm detection in fundus images," in *Proc. IEEE 1st Int. Conf. Connected Health, Appl., Syst. Eng. Technol. (CHASE)*, Jun. 2016, pp. 357–358.

[19] P. Khojasteh, B. Aliahmad, and D. K. Kumar, "Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms," *BMC Ophthalmol.*, vol. 18, no. 1, pp. 1–13, Dec. 2018, doi: 10.1186/s12886-018-0954-4.

[20] A. Mary Dayana and W. R. Sam Emmanuel, "A patch-based analysis for retinal lesion segmentation with deep neural networks," in *Proc. Int. Conf. Comput. Netw., Big Data IoT*. Cham, Switzerland: Springer, 2020, pp. 677–685, doi: 10.1007/978-3-030-43192-1_75.

[21] N. Sambyal, P. Saini, R. Syal, and V. Gupta, "Modified U-Net architecture for semantic segmentation of diabetic retinopathy images," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 1094–1109, Jul. 2020.

[22] O. M. Al Hazaimeh, K. M. O. Nahar, B. Al Naami, and N. Gharaibeh, "An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images," *Int. J. Signal Imag. Syst. Eng.*, vol. 11, no. 4, p. 206, 2018, doi: 10.1504/IJSISE.2018.10015063.

[23] N. Gharaibeh, O. M. Al-hazaimeh, A. Abu-Ein, and K. M. O. Nahar, "A hybrid SVM Naïve-Bayes classifier for bright lesions recognition in eye fundus images," *Int. J. Electr. Eng. Informat.*, vol. 13, no. 3, pp. 530–545, Sep. 2021.

[24] O. M. Al-hazaimeh, A. Abu-Ein, N. Tahat, M. Al-Smadi, and M. Al-Nawashi, "Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images," *Int. J. Online Biomed. Eng.*, vol. 18, no. 13, pp. 131–151, Oct. 2022.

[25] N. Y. Gharaibeh and A. A. Alshorman, "An effective diagnosis of diabetic retinopathy with aid of soft computing approaches," *J. Energy Power Eng.*, vol. 10, no. 8, pp. 1–10, Aug. 2016, doi: 10.17265/1934-8975/2016.08.003.

[26] M. Ahmad, Y. Ding, S. F. Qadri, and J. Yang, "Convolutional-neural-network-based feature extraction for liver segmentation from CT images," *Proc. SPIE*, vol. 11179, pp. 829–835, Sep. 2019, doi: 10.1117/12.2540175.

[27] S. F. Qadri, L. Shen, M. Ahmad, S. Qadri, S. S. Zareen, and S. Khan, "OP-ConvNet: A patch classification-based framework for CT vertebrae segmentation," *IEEE Access*, vol. 9, pp. 158227–158240, 2021, doi: 10.1109/ACCESS.2021.3131216.

[28] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Inf. Sci.*, vol. 501, pp. 511–522, Oct. 2019, doi: 10.1016/j.ins.2019.06.011.

B. NAVEEN KUMAR received the Bachelor of Technology (B.Tech.) degree from PES University, Bengaluru, India, in 2018. He is currently pursuing the Master of Technology (M.Tech.) degree with JAIN (Deemed-to-be University), Bengaluru. His research interests include the intersection of artificial intelligence (AI), machine learning (ML), the Internet of Things (IoT), and natural language processing (NLP). He has published two research papers in international conferences, showcasing his expertise in these areas. With over four years of work experience in SAP product design and development, he brings practical industry knowledge to his research pursuits. His academic excellence and professional experience make him a valuable contributor to the field of image segmentation using deep learning.

T. R. MAHESH (Senior Member, IEEE) is currently an Associate Professor and the Program Head of the Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, India. He has to his credit more than 50 research articles in Scopus/WoS and SCIE indexed journals of high repute. His research interests include image processing, machine learning, deep learning, artificial intelligence, the IoT, and data science. He has been the editor for books on emerging and new age technologies with publishers like Springer, IGI Global, and Wiley. He has served as a reviewer and a technical committee member for multiple conferences and journals of high reputation.

G. GEETHA (Member, IEEE) is currently the Director of the School of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, India. She has published several research articles and has graduated six Ph.D. students. Her research interests include security and cryptography, cyber-physical systems, and software engineering. She has served as a reviewer and a technical committee member for multiple conferences and journals of high reputation.
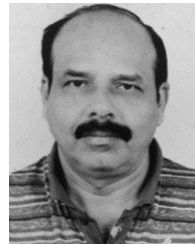
SURESH GULUWADI is currently an Associate Professor with Adama Science and Technology University, Adama, Ethiopia. With nearly two decades of experience in teaching, his areas of specialization include pervasive computing, artificial intelligence, the IoT, data science, and WSN. He has five patents in IPR and has published approximately more than ten articles in reputed international journals. He has authored Response Time Optimization in Wireless Sensor Networks. He has published numerous papers in national and international conferences. He is an active member of IEEE, ACM, I.E.(I), IACSIT, and IAENG. He has organized several national workshops and technical events. He has served as the editor/reviewer for Springer, Elsevier, Wiley, IGI Global, Emerald, ACM, and many more.

● ● ●