Computer Science • 25(4) 2024 https://doi.org/10.7494/csci.2024.25.4.5997

Le Akanksha Bali
Kuljeet Singh
Vibhakar Mansotra

EYE DISEASE SEGMENTATION USING HYBRID NEURAL ENCODER DECODER BASED U-NET HYBRID INCEPTION

Abstract *Diabetic retinopathy (DR) is one of the major causes of vision problems worldwide. With proper treatment, early diagnosis of DR can prevent the progression of the disease. In this paper, we present a combinative method using U-Net with a modified Inception architecture for the diagnosis of both the diseases. The proposed method is based on deep neural architecture formalising encoder decoder modelling with convolutional architectures namely Inception and Residual Connection. The performance of the proposed model was validated on the IDRid 2019 contest dataset. Experiments demonstrate that the modified Inception deep feature extractor improves DR classification with a classification accuracy of 99.34% in IDRid across classes with comparison to Resnet. The paper Benchmark tests the dataset with proposed model of Hybrid Dense-ED-UHI: Encoder Decoder based U-Net Hybrid Inception model with 15 fold cross validation. The paper in details discusses the various metrics of the proposed model with various visualisation and multifield validations.*

Keywords fundus images, UNET, deep learning, diabetic retinopathy (DR), Indian Diabetic Retinopathy Image Dataset (IDRID)

Citation Computer Science 25(4) 2024: 579–620

1. Introduction

Blindness and vision loss can result from diseases of the fundus. A few frequent fundus conditions that impair vision are cataract, diabetes retinopathy, and age-related macular degeneration [4]. Diabetes is a serious pandemic worldwide, especially in Indian culture and around the world. As a result, there is an excess of diabetes, which spreads disorders like DR.

Diabetes has affected more than 422 million people worldwide out of which India ranks in the top 3 nations in terms of the number of diabetics. Half of the population, which has increased from 108 million to 422 million in recent years, now resides in India, China, the USA, Brazil, and Indonesia [11]. Diabetes can cause an eye condition called diabetic retinopathy. It takes place when high blood sugar levels harm the blood vessels in the retina, causing visual issues. Although diabetes can result in several problems, this paper will concentrate on diabetic retinopathy. After having diabetes mellitus for a while, diabetic retinopathy, an eye condition, develops. It harms the tiny veins in the retina, resulting in obstruction, damage, and erratic vein growth. The narrowing of the blood vessels in the eyes is a secondary effect of diabetes brought on by high blood sugar levels. A reduction in the blood flow to the retina may cause blindness. Among the symptoms of DR are lesions that develop on the retinal surface [7]. The vessels that carry blood in the retina are harmed by sustained high blood sugar levels, which results in diabetic retinopathy. This results in weaker arteries, obstructions, and fluid leaks. The body may react by developing abnormal new blood vessels, which can bleed and cause more harm. Its advancement is aided by factors such as elevated blood sugar, high levels of blood pressure and cholesterol as well as chronic inflammation. For early identification and treatment, periodic vision tests and managing diabetes are essential. The small blood vessels in the retina are harmed by high blood glucose levels, which results in DR. The macula enlarges and thickens because of the retina producing a greater amount of the blood, cholesterol, the latter of other lipids [28]. As sample blood is given to the retina, IrMAs (Intraretinal Microvascular Abnormalities) start to form as new, abnormally delicate blood vessels [12]. The rising pressure within the eye may potentially cause damage to the optic nerve. The most efficient method for identifying early indications of anomalies in the present clinical diagnosis is routine screening of diabetes individuals using fundus examinations. Blindness can be prevented if DR is identified in its earliest stages and treated right away [37].

The figure beneath shows the different stages of DR. As shown in the figure below, there are various stages of diabetic retinopathy. Fluid leakage and microaneurysms are features of mild non-proliferative retinopathy. Blockages in the blood vessels, edoema, and distortion are symptoms of moderate non-proliferative retinopathy. There are more blocked vessels because of severe non-proliferative retinopathy. A characteristic of proliferative retinopathy is the development of aberrant blood vessels, which can result in hemorrhage and retinal detachment. Due to the many fundus imaging tools, accurately grading and identifying diseases might be difficult. It may

also be challenging to distinguish between normal and pathological regions due to occlusion, shadow, reflection, or inadequate illumination in DR fundus pictures (Fig. 1).

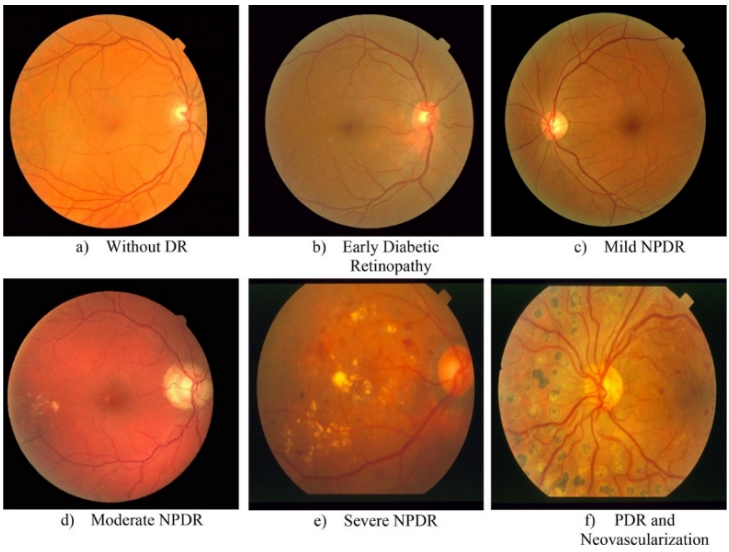


Figure 1. Dataset visualisation

Due to the many morphologies, sizes, and colours of lesions, manual retinal lesion assessment is laborious and demands great accuracy. Computer-assisted diagnostic (CAD) tools make it possible to diagnose diabetic retinopathy (DR) accurately and quickly, which helps with choosing the best course of action. To examine the Region of Interest (ROI) and determine the grade and severity of the disease, DR lesion segmentation is essential. The conditions that received the most research attention were glaucoma, AMD, and diabetic retinopathy [6]. Because ML techniques are more precise than traditional segmentation methods that rely on few indicators, they have taken their position. In comparison to conventional ML algorithms, DL, a contemporary technology, has demonstrated improved performance in DR lesion classification and segmentation [5].

In recent years, deep neural networks and particular machine learning models have shown promise in a variety of computer vision applications, particularly in medical picture analysis for diabetic retinopathy. Deep learning-based computer-aided diagnostic (CAD) systems that can categories anomalies can aid in medical decision-making and enhance patient care [25]. The rest of the paper is structured as follows. Related works on DR image classification are presented in Section 2. The information and suggested approach are explained in Section 3. The experimental analysis is presented in Section 4 and discussion are presented in Section 5. Section 6 presents the conclusions at the end.

2. Literature review

Depending on their study focus and area of interest, researchers in the subject of diabetic retinopathy have accomplished a lot of work. Researchers have suggested and applied a variety of machine learning techniques, as evidenced by the linked work in the fields of medical sciences and machine learning, however for diabetic retinopathy, a comparative analysis of these deep learning techniques is still absent. The work completed demonstrates that a fresh technique can be used when looking at the outcomes and comparing different machine learning algorithms for DR. In addition to several other analysis approaches, [22] used image processing for the automated and early detection of diabetic retinopathy. In their study, [40] suggested a method for improving images based on morphological operations, along with threshold-centered static wavelet transforms for the retinal fundus image and CLAHE (Contrast Limited Adaptive Histogram Equalization) for vessel enhancement. The method used by [50] is centered on a mixed classifier that detects retinal lesions through preprocessing, lesion extraction from candidates, feature formulation, and classification. The research leads to a further development of the m-Mediods-based modelling methodology, integrating it with the Gaussian Mixture Model to create a hybrid classifier that will increase classification accuracy. [2] investigated whether neural networks could recognize the signs of diabetes present in fundus photos and compared the network against a set of fundus images used for ophthalmologist screening. The research demonstrated the ability to find vessels, exudates, and hemorrhages. Their network achieved greater accuracy for the identification of diabetic retinopathy as compared to ophthalmologists. [15] contributed to the decrease in the number of features needed for the lesion categorization using feature ranking and Adaboost in their study. They suggested a novel two-step hierarchical classification method in which false positives or non-lesions are eliminated in the first stage.Bright lesions are further divided into two categories in the second stage: cotton wool patches and hard exudates. Additionally, red lesions are still categorized as haemorrhages and micro-aneurysms (MA).

Guo et al. [18] suggested a multi-scale feature fusion technique to address tiny lesion detection problems. To improve sensitivity, binary cross-entropy (BCE) loss alongside balancing coefficients was applied. The computational model was trained using full-resolution photos that were scaled to pixels without any additional preprocessing. Yan et al. [47] presented a mutually local-global U-Net to address the shortcomings of patchwise training, which fails to capture global context.

Other methods researchers have developed for segmentation purposes over the years like Fuzzy C-Means (FCM) clustering algorithms which are commonly used to divide image pixels into diverse cluster [30], and others like region growing methods are used to form distinct image regions based on some uniformity criteria such as grey level and colour [39], and mathematical morphology operations are performed by analysing the geometrical makeup of certain retina components [41,43]. However, most existing approaches are limited to acting on a single retinal image, making it impossible to extract spatial and spectral characteristics across many spectral slices at the same time.

Various researchers have advocated various segmentation strategies. In any event, these approaches only work on fundus pictures and have no pathological implications [14]. It is challenging to segment the vascular vessel treemap without discontinuities. Azzopardi et al. [3] created a filter called the Bar-selective Combination of Shifted Filter Responses (BCOSFIRE). Its parameters have an impact on the filter's performance. To detect retinal vessels, Wilfred Franklin and Edward Rajan [13] devised the Multilayer Perceptron Neural Network. For its representation, the weight of the feedforward network is altered using the backpropagation technique. Because it is dependent on pixel processing, it has a lower level of accuracy of 95.03%. Partovi et al. [31] proposed a model in which photos were classified using an error-based autonomous network. It was validated using a dataset of remote sensing photos. In the categorization of medical images, the Deep CNN (DCNN) model provided an extended feature extraction-classification technique. Gulshan et al. [17] trained the Deep Convolutional Neural Network (DCNN) to detect DR in retinal fundus images. Deep learning can be utilised in retinal fundus pictures to construct an algorithm that automatically detects DR and diabetic macular edoema. The specificity and sensitivity of the method for assessing DR expressed as moderate or worse DR or both were generated based on the key decision of the ophthalmology team. The method, which has 96.5% sensitivity and 92.4% specificity, was created using deep convolutional neural networks and a large quantity of data in various grades per image. By learning a deconvolution network, the authors of (Noh et al., 2015) [26] suggested an improved semantic segmentation technique. The convolutional layers are adopted from VGG16 as well, and the deconvolution network is made up of deconvolution and unpooling layers that identify pixel-wise class labels and forecast segmentation masks. On the same PASCAL VOC 2012 dataset, the proposed technique achieved 72.5% IoU. Shen et al. [38] introduced a multicrop pooling method that was used in DCNN to capture object salient characteristics in order to classify lung nodules on CT images.

In the literature [20], DR pictures were classified using a Gabor filtering and SVM classification model. Before employing the classifier, the Circular Hough Transform (CHT) and CLAHE models were fed with input images, yielding a detection rate of 91.4% on the STARE dataset. [36] made a contribution to the decrease in the amount of features needed for the lesion categorization using feature ranking and Adaboost in their study. They suggested a novel two-step hierarchical classification method in which false positives or non-lesions are eliminated in the first stage. Bright lesions are further separated into two categories in the second stage: cotton wool patches and hard exudates. Additionally, red lesions are still categorized as hemorrhages and micro-aneurysms (MA). For the diagnosis of various retinal abnormalities, the division of the blood vessel represents a crucial prerequisite. The accuracy of the cascaded U-Net architecture used by [36] on the DRIVE dataset was 96.92%, and its precision on the STARE dataset was 97.40%. [24] propose a deep CNN network that simultaneously slices arterioles and venous vessels from the DR pictures. The DRIVE dataset, on which the framework was tested, achieved 87% sensitivity and 98% specificity.

Data augmentation using the U-Net model was done over the DRIVE dataset by [46] which obtained an AUROC score of 0.97 for segmenting blood vessels.

On the MESSIDOR dataset [37] classified DR fundus images using a modified AlexNet architecture by applying the appropriate pooling, softmax, and ReLU layers. The MESSIDOR dataset showed 96.6% accuracy for the suggested model. Xiancheng et al. [45] demonstrated a DR model for classification using the InceptionV3 as architecture and a short dataset transfer learning approach. Utilizing an SGD optimizer including the cosine loss function for binary classification, Hagos' technique obtained 90.9% accuracy. In a study by [19], the referable lesion regions in DR images were examined using a generalization of the backpropagation approach and a poorly supervised model. On the Kaggle dataset and the E-Ophtha dataset, [33]'s suggested method has an area around the ROC curve (AUC) of 95.50% and 94.90%, respectively. An ensemble of models trained with deep learning was used in a study by [29] to detect red lesions in fundus images. In that technique, 3232 pixel patches were extracted initially and put into a Deep CNN. The random forest (RF) classifier also received hand-crafted features that were extracted in addition. Orlando's approach demonstrated how a hybrid feature vector with both hand-crafted and deep learning-based features might enhance the networks' performance and obtain 89.32% AUC. In order to classify eye illnesses, Bali et al. [8] suggested a DFex-hybrid strategy combining the BeeHive model, CGAN, and PSO. In the RFMiD and ODIR datasets, they obtained, respectively, accuracy of 98.79%, sensitivity of 95.99%, specificity of 99.79%, and accuracy of 97.16%, F1 score of 96.81%.

The goal of this paper is to develop an automatic diagnosis method for DR. The paper discusses multiple deep learning techniques for the segmentation of the fundus image as shown in Table 1. The paper is benchmarked over main dataset of IDRID for respective classifications. The paper is capable of not only reducing the computation time for the enhanced Dense-ED-UHI: Encoder Decoder based UNet Hybrid Inception architectures but also is able to hybridize it with inception V3 to calculate complex features for better extractions.

Considering the aforementioned review, we developed a unique architecture to overcome the aforementioned restrictions. The following is a description of the main contributions of the suggested research efforts:

1) This study uses a customised U-Net model (UHI) to semantically segment retinal lesions. The encoder used in the model is a deep network encoder called "InceptionV3". The smaller convolutions performed by the InceptionV3 model allow for a quicker training procedure.
2) To resize the convolution nearest neighbour, the model applies a unique upsampling technique that involves pixel-wise periodic shuffling convolution. Compared to traditional approaches, our procedure speeds up network convergence and creates retinal images with excellent resolution that are free of checkerboard artefacts.

3) Comparing the suggested model to other existing works, it produced state-of-the-art results for the diagnosis of DR lesions.
4) A hybrid U-Net architecture with inception design and multiple kernel extractions is used in the study to infect deep encoding.
5) Finally, the study evaluates the model's performance across multiple classes.

Table 1
Literature review comparison

Study	Methodology	Advantages	Disadvantages	Specificity	Recall	AUC
[17]	DL: Convolutional Neural Network	High accuracy and precision	Large labelled datasets required	0.99	0.90	NA
[1]	DL: Ensemble of CNNs	Automated feature learning	Computationally intensive	0.87	0.96	0.98
[16]	DL: CNN	Handles diverse lesion morphologies	Prone to overfitting	0.98	0.94	0.97
[42]	DL: CNN	Robust to noise and variations	Data augmentation may be required	0.91	0.90	0.93
[49]	ML: Random Forest	Can handle small or imbalanced datasets	Parameter tuning may be required	0.70	0.80	NA
[34]	ADL (Active Deep Learning)	Can be used to define the level of severity of retinal images important patches with provision of (ROI).	Can be improved through data mining and data exchange methods	0.95	0.92	NA
[48]	Deep DR	It is useful for DR severity classification	The model is too complex	0.97	0.97	NA
[35]	Deep learning methods	It assigns pixel values for determining the importance and evaluation of forecasted category	Parameter tuning may be required	0.90	0.90	NA
Proposed Model	UHI Network	Described in section 3	After Hyperparameter Tunning	0.997	0.989	0.999

3. Methodology

The methodology, experiment design, and workflow employed in the proposed framework in Figure 2 were used in this section's full explanation of the dataset. The following shown below is the proposed architecture

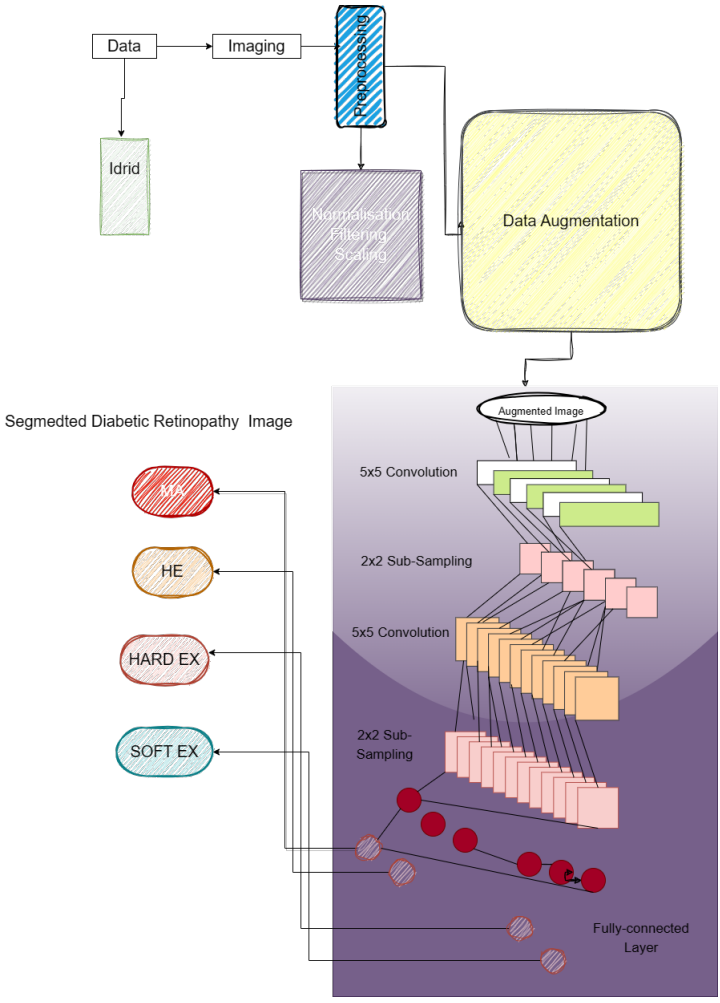


Figure 2. Flow diagram

3.1. Dataset rescription

The open-access Indian Diabetic Retinopathy Image Dataset (IDRid) focuses on diabetic retinopathy, an eye-related consequence of diabetes. The dataset is intended to support study and advancement in the area of retinal analysis of images, particularly for the detection and grading of diabetic retinopathy. There are 516 retinal pictures

in all, each with a resolution of 4288 × 2848. The pixel-level annotation has provided a binary mask for each distinct DR deformity, such as haemorrhages, hard-exudates, soft-exudates, and microaneurysms. Furthermore, every 516 photos have been graded and given the DR severity [21]. The data description in Table 2.

Table 2
Data description

S. No	Keypoints	Description
1	Content	The IDRid dataset includes fundus photographs of the retina. 516 high-resolution retinal pictures make up the entire collection
2	Annotation	Expert-verified ground truth labels describing the existence and severity of diabetic retinopathy are added to each image in the collection as annotations. The labels, which vary from 0 to 4, represent the International Clinical Diabetic Retinopathy severity levels from "No DR" (zero) to "Proliferative DR" (four)
3	Image formats	The uncompressed TIFF format used by the IDRid dataset preserves the high-resolution details required for analysis and diagnosis
4	Dataset division	A training set as well as a testing set were created from the dataset. There are 413 images in the training set and 103 images in the testing set. This category enables researchers to assess how well their algorithms function on hypothetical data
5	Terms of usage	In accordance with the rules of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence, the IDRid dataset is available for research purposes. This means that, as long as the original authors are properly credited, you are free to use the dataset, modify it, and redistribute it for any non-commercial reasons

3.2. Implementation details

For model training, a GPU with 8GB of RAM is used. In the papers of [9, 10], both attempted were some of the earliest papers to utilise U-Net for image segmentation. In these studies, it is discussed how to segment retinal lesions in images of diabetic retinopathy using the U-Net design. Since then, the U-Net architecture has been widely used for a variety of medical picture segmentation applications, including the identification of diabetic retinopathy. After being initially presented for broad biomedical image segmentation tasks, the U-Net architecture has now been adopted and used in a number of medical imaging areas, including diabetic retinopathy.

While segmenting retinal structures and lesions in images of diabetic retinopathy has shown the U-Net architecture to be useful, there are certain restrictions as well. Utilising U-Net for diabetic retinopathy has some drawbacks, including a restricted

capacity to capture global context. U-Net may not successfully consider greater spatial connections and contextual signals because of its encoder-decoder architecture, which places a heavy emphasis on local information. This constraint may reduce the model's capacity to incorporate global data necessary for precise categorization of diabetic retinopathy, which may have an adverse effect on the model's general efficacy. Additional architectural alterations or the incorporation of other methodologies might be required to improve the integration of global context and increase the categorization accuracy in order to counteract this limitation.

The study adopts concepts from the Inception design and inserts them into the neural network instead of using the conventional U-Net architecture. The detection becomes simpler as Inception design combines numerous kernel sizes to collect information on various scales and at various granular levels. Additionally, U-Net Inception improved feature representation, led to a more thorough knowledge of the retinal image, and may improve the performance of the suggested model as a whole. Additionally, the expansive to contract cross-linking helps produce the extremely precise segmented output image. There are four blocks of encoder units in the U-Net contract chain. Each encoder unit has two convolution layers that are followed by a max-pooling layer. Every time a pooling process is used, the feature elements are doubled. The bottleneck, which consists of two convolution layers with one at the top, is the essential component that separates contractile and expanded techniques. A U-Net expanding path is being performed by four decoder units and consists of two convolution layers, a de-convolution layer, and two comparable feature maps using the contract trail. The Table 3 below discusses the brief difference between the both.

The proposed work with the dataset split into the training, validation, and testing set for DR classification in retinal pictures. In the Figure 3 seen below, the paper proposes a detailed overflow of utilising four benchmark datasets to train the images, follow the methods and update if required. Following this, testing the dataset is done and if the model performs up to the expectations, it can be utilised in the field of medical imaging. The Figure 3 and Figure 4 describe the primary steps of data preprocessing the the following order

1) Training – Testing and Validation Split, as discussed, the training data 70% is used to train and reiterate the model. Testing and validation process used to analyse the fitting parameter and ensure precise modelling.
2) Data preprocessing, it includes various steps like Normalisation of the given input sequences, Filterining Scaling etc.
3) Mask Encoding or One Hot encoding is done to create a binary mask for multiple classes used.
4) Model creation, hyperparameters and other units are discussed below.
5) Post training cycle is completed the model goes through a vivid evaluation for Cross validation across all metrics such as accuracy, precision, recall, sensitivity, specificity.

Table 3
Difference between U-Net Inception and Standard U-Net

Aspect	UHI: U-Net Inception	Standard U-Net
Architecture	The U-Net framework incorporates concepts from the Inception architecture	Encoder-decoder as well as skip connections are part of the original U-Net architecture
Multi-scale feature extraction	Uses various kernel sizes to efficiently capture data at various scales	Focuses heavily on skip connections for the extraction of local features
Global context	Due to the incorporation of modules inspired by Inception, the global context may be captured more successfully	Capability to capture the global context is somewhat limited
Fine detail representation	Possibility of enhanced fine detail representation in retinal pictures	May rely more heavily on skip connections to express fine details
Performance potential	Possibility of improving categorization accuracy by using data from multiple scales	Established performance on tasks involving the classification of diabetic retinopathy
Complexity	May add to the complexity of the architecture and processing	Architecture that is less complex and contains fewer extra parts

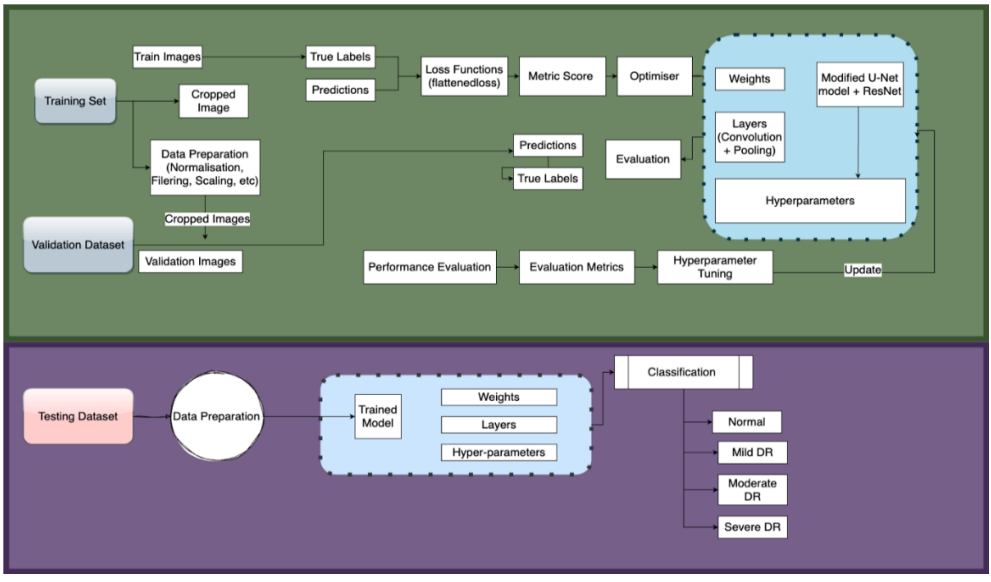


Figure 3. Data split and implementation

The figure below talks about the proposed U-Net Inception model for diabetic retinopathy classification. The primary aims the figure postulates to demonstrate is that training and testing sets are passed via various preprocessing pipelines after which the train images are segregated with loss function and back propagation utilities of the deep network inclusive of weights, biases, layers and hyperparameters as in Figure 4.

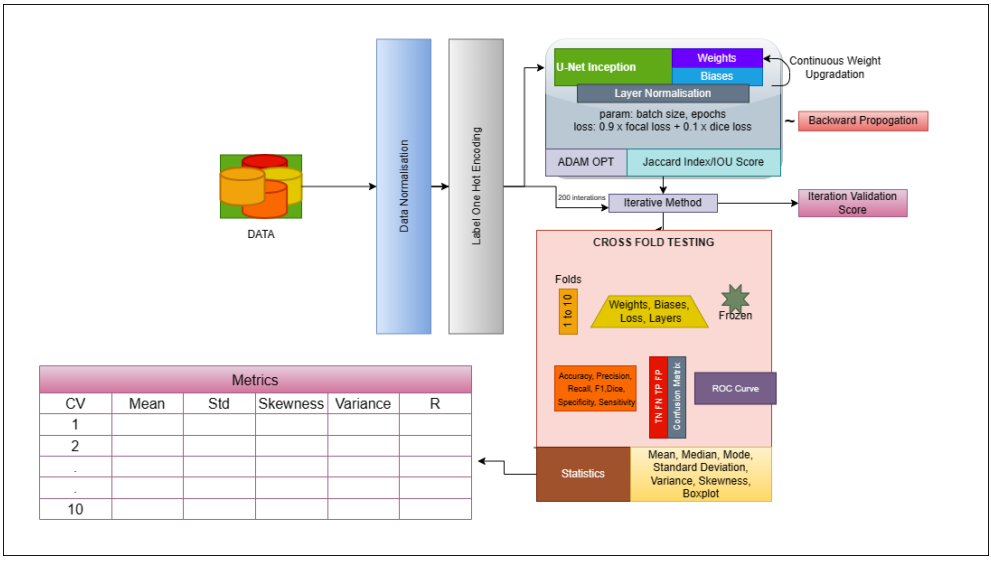


Figure 4. Data evaluation and cross validation set

The following is a description of the various building blocks' intended architectural details.

3.2.1. Encoder unit

The encoder is essential for capturing and extracting hierarchical characteristics from the input retinal pictures in the U-Net Inception Net structure for diabetic retinopathy classification. It is made up of convolutional layers which combine input data with filters to extract features. The encoding layers steadily increase the filter depth, enabling the neural network to capture intricate information important for categorising diabetic retinopathy.

The encoder uses down sampling techniques like max pooling or stride convolution to increase the number of feature channels while decreasing the spatial dimensions of the feature map. This down sampling aids in obtaining more advanced depictions of the input and summarizing the data. The incorporation of the Inception module into the encoder is one of the unique features of the U-Net Inception Net. The Inception module includes of a max pooling branch and parallel convolutional branches that use different kernel sizes, notably 1×1, 3×3, and 5×5. These branches catch

features at many scales and assist the network in simultaneously learning local and global knowledge. The encoder of the U-Net Inception Net improves the network's capacity to collect complex and multi-scale information that are necessary for precise diabetic retinopathy categorization by including the Inception module. Overall, the U-Net Inception Net encoder collects pertinent and discriminative characteristics from the input retinal images, creating a rich representation which is further used in the later phases of the architecture for classifying diabetic retinopathy.

3.2.2. Inception unit

The U-Net Inception Net design for classifying diabetic retinopathy includes the Inception module, which was first introduced in GoogLeNet. By obtaining multi-scale and multi-level features, it improves a network's feature extraction capabilities. The Inception module includes of a max pooling branch and parallel convolutional branches with various kernel sizes, such as 1×1, 3×3, and 5×5. The network may collect data at multiple scales since every branch captures characteristics at a distinct receptive field size. The 1×1 convolutions are in charge of dimensionality reduction and input channel reduction, which helps to reduce computational complexity.

The U-Net Inception Net is capable of recording local as well as global information by integrating the Inception module. The concurrent branches within the module allow the network to simultaneously learn complicated and varied information. This is especially helpful in the classification of diabetic retinopathy, because the network must collect information at numerous scales due to the presence of various lesions, including microaneurysms and exudates. The U-Net Inception Net's capacity to extract discriminative characteristics pertinent to diabetic retinopathy is improved by the Inception module. The network may collect both fine-grained minutiae and high-level semantic data by combining multi-scale features. This aids in increasing the classification task's accuracy and resilience, allowing for more precise detection and assessment of diabetic retinopathy (DR) in retinal pictures.

3.2.3. Skip connections

A key element of the U-Net design, which includes the U-Net Inception Net, for the classification of diabetic retinopathy is skip connections. These connections are essential for enhancing information transfer and maintaining spatial details across the network. Skip links create immediate connections among the encoding and decoding pathways in the U-Net Inception Net. These links make it possible to combine multi-scale along with multi-level features, which makes it easier to integrate low-level and high-level data. The feature maps from the respective encoder layers to the decoder layers are concatenated to create the skip connections.

The U-Net Inception Net successfully blends semantic detail from the decoder with low-level fine-grained features from the encoder by including skip connections. This gives the network a thorough comprehension of the linked features in the input retinal images. In addition to improving the network's capacity to precisely

classify diabetic retinopathy lesions, it aids in the preservation of spatial information. Additionally, skip connections help to solve the issue of a data bottleneck that might develop in deep networks. A network is able to access feature maps from various scales and levels by simply linking the encoding and decoding paths, which improves gradient flow and solves the vanishing gradient issue. The skip connections in the U-Net architecture Inception Net architecture, in general, promote efficient knowledge propagation, enable the integration of multi-scale data, and contribute to the precise identification of diabetic retinopathy lesions by fusing local information with global context.

3.2.4. Bottleneck

The bottleneck, which sits in the middle of the encoder and decoder paths, acts as a bridge to collect the input data in its most compressed and abstract form. The bottleneck in the U-Net Inception Net is often made up of several convolutional layers including Inception modules.These layers and modules are created to capture highly discriminative properties, which are essential for correctly classifying lesions associated with diabetic retinopathy.

The U-Net Inception Net gains the capacity to collect characteristics at various scales and levels of abstraction by including Inception modules within the bottleneck. The Inception modules' parallel convolutional branches make it easier to extract detailed information, allowing the network to pick up on intricate patterns and structures found in retinal images. The bottleneck's job is to compile the encoder's learnt representations and get them ready for the decoding path. It seeks to decrease the computational difficulty of the network while capturing the fundamental properties that are most pertinent to the classification task.

The bottleneck efficiently compresses the data, making the retinal pictures' representation more condensed and expressive. It allows the network to generate precise predictions throughout the succeeding decoding and classifying stages by collecting the most discriminative properties.

3.2.5. Activation function

In order to introduce non-linearity and improve the network's capacity to learn intricate correlations between the input data and the target labels, activation functions are crucial. The Rectified Linear Unit (ReLU), Leaky ReLU, and Exponential Linear Unit (ELU) are frequently used activation functions in the U-Net Inception Net. ReLU is a widely used activation function that maintains positive values while setting negative values to zero. Leaky ReLU avoids the dying ReLU problem by permitting a modest non-zero slope for negative values. ELU has a smooth curve for negative values, unlike ReLU, which does not. The non-linear mapping among labels for diabetic retinopathy (DR) and the input retinal pictures are simulated with the help of these activation functions. The U-Net Inception Net can capture complex patterns and data by incorporating non-linearity, which enhances its ability to distinguish between different types of diabetic retinopathy lesions and properly categorize them.

The selection of the activation function is influenced by the specific requirements for the classification task as well as the characteristics of the dataset.

3.2.6. Decoder unit

Upsampling the feature maps and regaining the lost spatial resolution are under the purview of the Decoder. The decoder for the U-Net Inception Net commonly consists of upsampling layers like bilinear interpolation or transposed convolutions. These layers expand the feature maps' spatial dimensions, enabling the network to recreate the retinal pictures' more minute details.

The decoder also includes skip connections that combine the feature maps from the associated encoder layers. In order to accurately classify data, these skip connections are essential for maintaining spatial data as well as combining low-level and high-level properties. The U-Net Inception Net's skip connections make it easier to combine multi-scale features, giving the network access to both fine-grained specifics and high-level semantic data. This helps gather pertinent data from many levels of abstraction and enhances the network's comprehension of diabetic retinopathy lesions. The decoder in the U-Net Inception Net successfully reconstructs the spatial details of the retinal pictures using upsampling operations and skip connections, providing the data required for precise diabetic retinopathy categorization.

3.2.7. Hyperparameter set

The performance of the U-Net Inception Net architecture for the categorization of diabetic retinopathy is strongly influenced by a number of hyperparameters. During optimisation, the learning rate defines the step size, which affects convergence speed and stability. The batch size regulates how many samples are handled during each training iteration, which has an impact on both the accuracy and efficiency of gradient estimation. The U-Net Inception Net's ability to learn complicated characteristics depends on how many layers it has. Deeper networks can capture more detailed patterns but need more computer power. Regularisation factors that avoid overfitting and regulate the amount of regularisation used include dropout and L1 or L2 regularisation. The model's non-linear behaviour showcased in Figure 5 and capacity to represent complicated relationships are significantly influenced by the selection of activation functions, such as ReLU, Leaky ReLU, or ELU.

Choosing the best hyperparameter values for diabetic retinopathy classification employing the U-Net Inception Net architecture typically requires experimentation and validation on a different validation set. Techniques like grid search or random search may be used to explore various configurations. The metrics used to evaluate the performance are iou_score, focal loss, dice loss, accuracy, binary accuracy, AUC and ROC, Specificity, Sensitivity.

a) b)

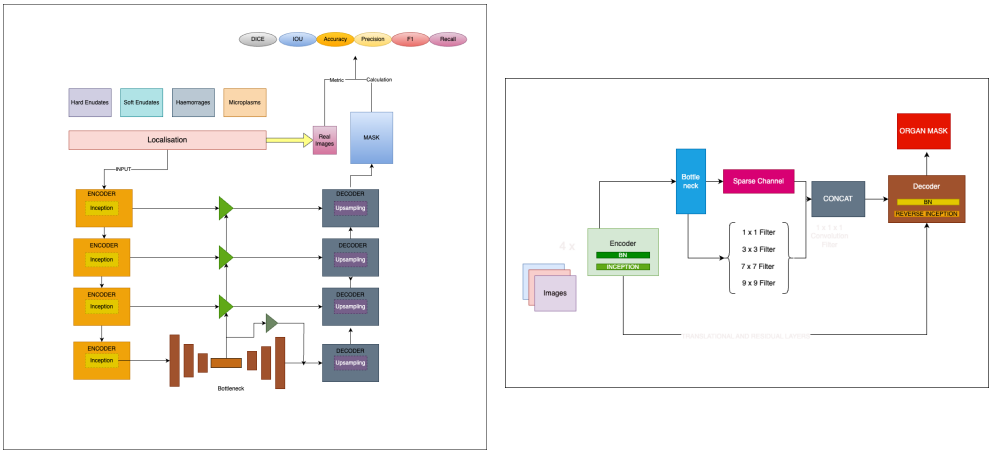


Figure 5. Data evaluation and cross validation set: a) Unet architecture; b) Inception V3 convolutional network

4. Results

4.1. IDRID

Specific kinds of diseases that are frequently seen in visual pictures include hard exudates, soft exudates, haemorrhages, and microaneurysms. These kinds of lesions are frequently a sign of several visual illnesses, such retinopathy caused by diabetes. An extensively used collection containing retinal pictures annotated for these abnormalities is called IDrid (Indian Diabetic Retinopathy Image collection). Let's talk about the traits associated with these tumours and the criteria needed for segmenting them.

4.1.1. Hard exudates

In the retina, triglycerides seep through injured veins, forming brownish or white plaques known to be hard discharges. They frequently take the form of tiny, rounded, or longitudinal tumours with distinct margins. As their buildup can cause blurred vision and act as a sign of the seriousness of diabetes-related retinopathy, these hard discharges must be segmented in order to be quantified and their development tracked.

4.1.2. Soft exudates

Puffy white or yellowish blemishes referred to as cotton fibre areas, soft discharges are the result of strokes of the ocular sensory fibre layer brought on by poor blood flow. These resemble amorphous tumours with hazy edges. Soft exudates must be segmented in order to determine their amount and location because they can be an indication of ocular hypoxia.

4.1.3. Haemorrhages

Blood leaks into the retinal cells during haemorrhages, which happen as veins burst. They resemble irregularly shaped, black lesions that are red or dark in colour. As they can reflect the extent of retinal harm and the evolution of diabetes-related retinopathy, haemorrhages must be segmented in order to determine their existence and pinpoint where they are.

4.1.4. Microaneurysms

These tiny, rounded expansions of the capillaries in the retina are believed to be the first symptoms of retinopathy caused by diabetes. They appear as tiny, rounded lesions that are brilliant red. Since the existence and growth of microaneurysms can suggest the possibility of progressing to more severe retinal stages, segmenting them is crucial for their measurement and management.

4.2. Retinopathy grade classification

From Table 4 and Figure 6 with a specificity of 0.909091 for type 0 retinal degeneration, cases who lack Retinopathy that can be identified with excellent precision. The F1-score value for third-degree retinopathy grade is 0.917197, which indicates that it is possible to forecast grade 3 retinopathy with an adequate level of recall as well as accuracy. Recall and sensitivity are used equally in this list. For example, the recall/sensitivity for retinopathy grade 2 is 0.972222, meaning that roughly 97.2% of cases of grade 2 retinopathy were effectively detected by the model. The detection rate for grade 1 retinopathy is 0.941176, meaning that the model correctly identified cases of grade 1 retinal with a 94.1% accuracy. Prediction of category wise classes is presented in Figure 7.

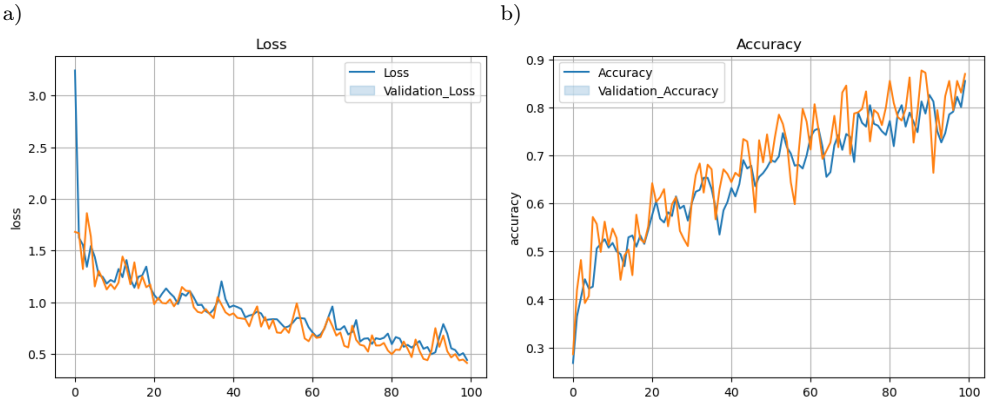


Figure 6. Loss and accuracy plot for Retinopathy Grade Classification: a) loss training validation plot; b) accuracy training validation plot

Table 4
Classwise metric distribution (testing)

Class	Accuracy	Precision	Recall or Sensitivity	F1-Score	Specificity
Retinopathy grade = 0	0.912621	0.955224	0.914286	0.934307	0.909091
Retinopathy grade = 1	0.941748	0.941176	1.000000	0.969697	0.842857
Retinopathy grade = 2	0.902913	0.897436	0.972222	0.933333	0.941935
Retinopathy grade = 3	0.873786	0.972973	0.867470	0.917197	0.900000
Retinopathy grade = 4	1.000000	1.000000	1.000000	1.000000	1.000000

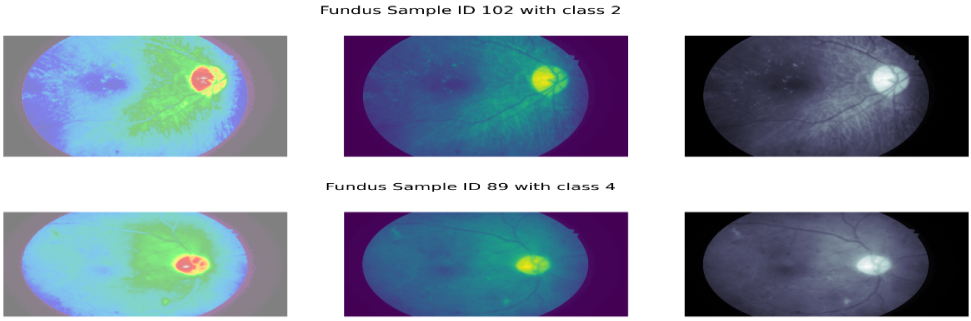


Figure 7. Class wise prediction

4.3. Classification risk of macular edema

The percentage of correctly predicted cases to all occurrences is what is meant by the term "accuracy", which measures overall competence. We can see the accuracy for every category in the following table. For example, the prediction accuracy for category 0 is 0.953995 (as per the training loss 8), meaning that in almost 95.4% of cases, the algorithm correctly predicted the likelihood of macular edema level 0. The model's prediction at classifying True Positives amongst the Predicted Classes that are positive is measured in the "Precision" row. For example, the precision for class 1 is 0.991736, indicating that the algorithm successfully identified cases of macula edema level 1 with a high precision of 99.2% (cf. Fig. 8).

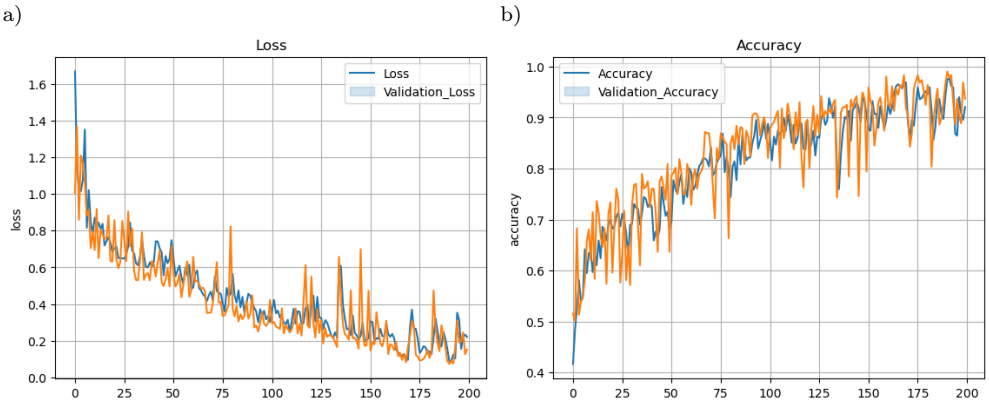


Figure 8. Loss and accuracy risk of macular edema: a) loss training validation plot; b) accuracy training validation plot

It is the same as sensitivities in Table 5 and 6. For example, the recall/sensitivity for category 2 is 1.000000, meaning that every incidence of macular edema level 2 was successfully detected by the algorithm, resulting in an ideal recall. Better performance is indicated by a higher F1 score. For instance, the F1 rating for class 1 is 0.979592, Indicating a strong equilibrium between recall and precision in classifying macular edema level 1. For instance, the specificity for class 0 is 0.971910, showing an elevated degree of reliability in detecting cases without macular edema at level 0.

Table 5
Class wise metric distribution (training)

Class	Accuracy	Precision	Recall or Sensitivity	F1-Score	Specificity
Retinopathy grade = 0	0.951456	0.950000	0.966102	0.957983	0.931818
Retinopathy grade = 1	0.961165	0.989011	0.967742	0.978261	0.900000
Retinopathy grade = 2	0.990291	0.981818	1.000000	0.990826	0.979592

Table 6
Class wise metric distribution (testing)

Class	Accuracy	Precision	Recall or Sensitivity	F1-Score	Specificity
Retinopathy grade = 0	0.953995	0.977876	0.940426	0.958785	0.971910
Retinopathy grade = 1	0.963680	0.991736	0.967742	0.979592	0.926829
Retinopathy grade = 2	0.956416	0.924051	1.000000	0.960526	0.907216

The class wise prediction is showcased in Figure 9.

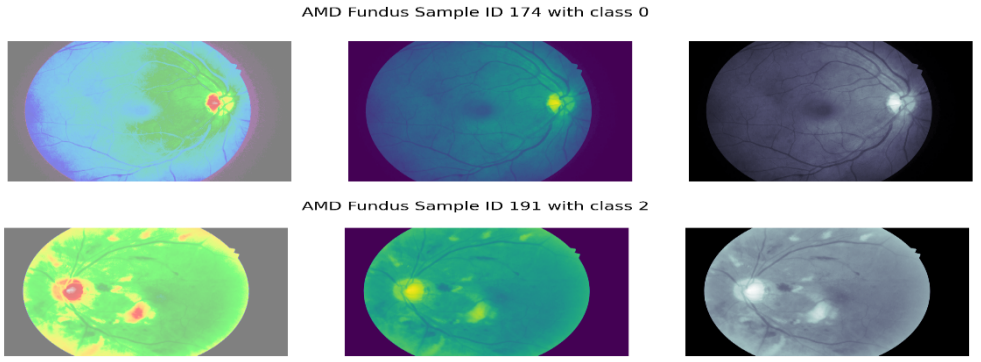


Figure 9. Edema class wise prediction

Table 7 highlights the distribution of samples of data throughout distinct categories of lesion types, possibly linked to medical imaging analysis. The table provides the total number of the samples for each given lesion type. "Hard Exudates" form the most prevalent class with 350 samples, next to "Soft Exudates" with 150 samples, "Hemorrhages" with 120 samples, and "Microaneurysms" with 110 samples. These values represent how often each lesion type occurs in the dataset, offering vital information into the dataset's makeup. Such information is critical for creating reliable artificial intelligence models, since the class distribution might affect the model's accuracy and generalization.

Table 7
Data samples per class

Lesion type	Number of samples
Hard exudates	350
Soft exudates	150
Hemorrhages	120
Microaneurysms	110

The provided table highlights the dataset's variety and identifies probable patterns or abnormalities in lesion incidence. In medical settings, this knowledge supports both scientists and clinicians in knowing the cause of certain problems, guiding diagnosis, treatment techniques, and future study. It is vital to remember that the application of these figures should be determined considering the unique clinical environment and the aims of the study, whether it's for boosting diagnostic accuracy or identifying disease features.

4.3.1. Hard exudates

The difference in the predicted and actual classes of hard exudates is measured using a loss operation, and a median loss of 0.057894 is the mean of the function throughout train. Better model scores in accurately representing the characteristics of hard exudates is shown by a smaller loss. The accuracy of the model appears to be stable throughout various training iterations based on Figure 10, Figure 11 and Figure 12, as indicated by a small variance of 0.014040. The amount of variance in the performance of the model is indicated by the least loss of 0.035471 to the largest loss of 0.090734. The model performs well overall in categorising pixels into either hard exudates or non-hard exudates, according to the mean accuracy of 0.999293. This shows that the model properly picks up on the characteristics that set hard exudates apart and can correctly distinguish these from various ocular components. The model's accuracy remains constant across assessments, as seen by the low standard deviation of 0.000162. The range of variation in the model's accuracy across several samples is highlighted by the least accurate value of 0.998942 and the highest accuracy of 0.999645.

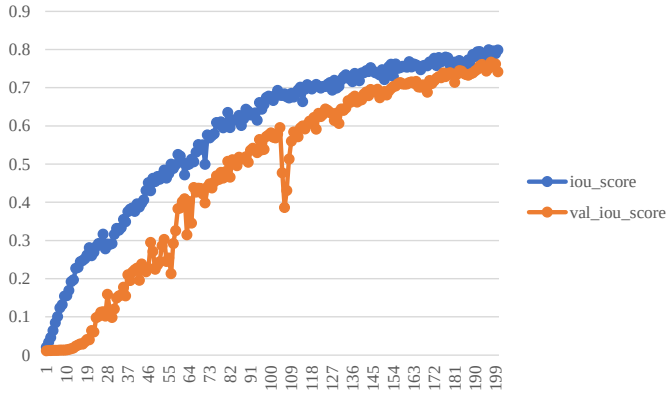


Figure 10. Training IOU_score

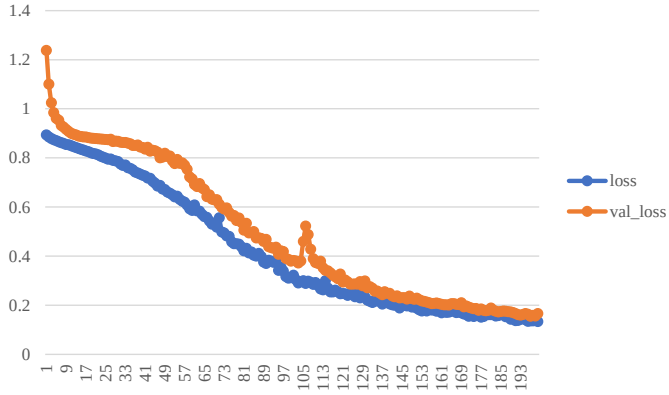
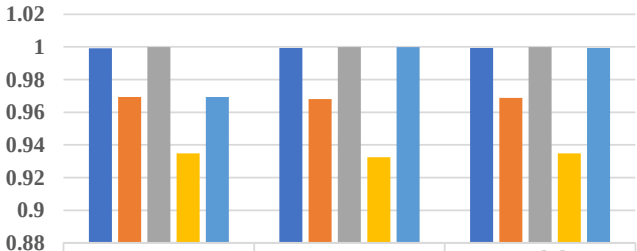


Figure 11. Focal+dice loss



	Train	Test	Valid
Accuracy	0.999257	0.999327	0.999326
Binary Accuracy	0.969272	0.968051	0.968795
AUC	0.999835	0.999873	0.999871
Specificity	0.934805	0.932559	0.934836
Sensitivity	0.969272	0.999873	0.999326

Figure 12. UHI model performance

Figure 13 and Figure 14 illustrates the prediction of Unet Inception and Resnet Models.

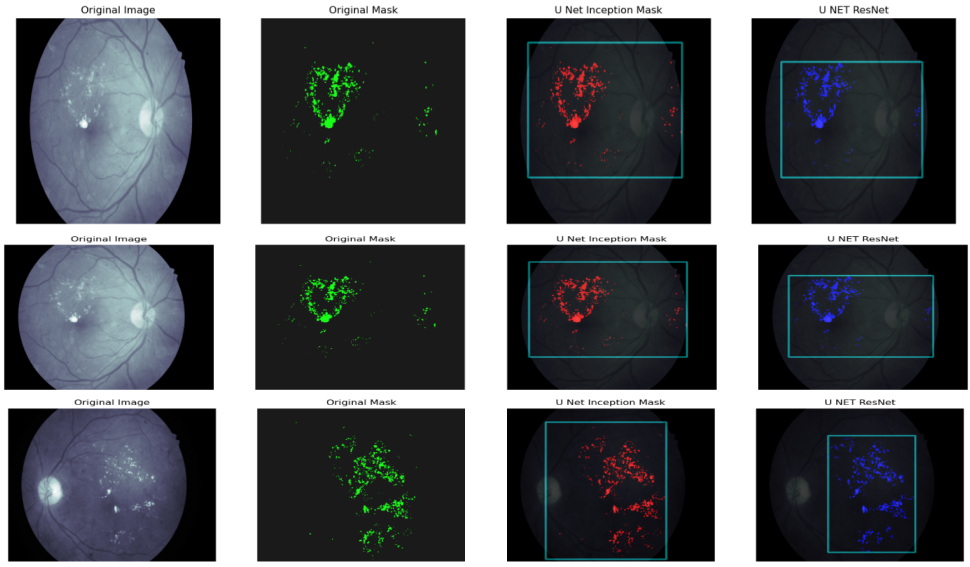


Figure 13. Mask prediction and BBox using Dense-ED-UHI:
Encoder Decoder based Unet Hybrid Inception (proposed model)

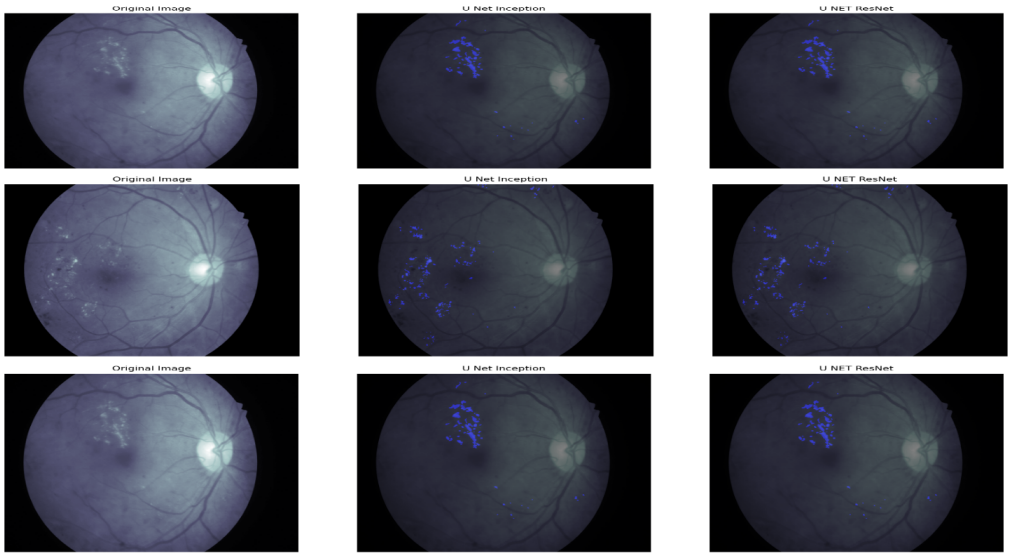


Figure 14. Comparison with other state of the art backbone (Resent50)

As per Table 8 the median amount of correctly anticipated hard exudates is represented by the median true positive value, which is 34,116.49. This score sheds light on the algorithm's precision in hard exudate detection and segmentation. The variance in true positives among various evaluations is shown by the standard variation of 13,782.728251. The range of variance in the model's ability to recognise hard exudates is shown by the smallest value of 11,497 and the highest number of 71,688.

Table 8
Statistical analysis

Statistical test	Loss	Accuracy	True positives	AUC	Specificty	Sensitivity
mean	0.057894	0.999293	34,116.490000	0.966475	0.999851	0.929403
std	0.014040	0.000162	13,782.728251	0.009651	0.000046	0.019548
min	0.035471	0.998942	11,497.000000	0.942636	0.999746	0.881022
25%	0.046062	0.999188	23,017.000000	0.959663	0.999817	0.915989
50%	0.057263	0.999304	31,707.500000	0.967367	0.999855	0.930924
75%	0.069872	0.999411	43,709.000000	0.974321	0.999885	0.945510
max	0.090734	0.999645	71,688.000000	0.982697	0.999944	0.962151

AUC (Area Under the Curve) whose value is 0.966475 represents the algorithm's overall efficacy in identifying hard and non-hard exudates. Stronger discriminating skills are indicated by a larger AUC, with numbers near 1 signifying outstanding

results. The model's capacity to accurately recognise non-hard secretions is demonstrated by its average specific of 0.999851. The algorithm's ability for correctly recognising hard exudate regions is indicated by a median sensitivity value of 0.929403. For assessing the model's effectiveness in accurately identifying hard exudates, the two metrics are crucial.

Overall, the statistical assessment of rigid exudates offers insightful information about the effectiveness of the categorization approach. The model is efficient in capturing the features of hard exudates, as evidenced by its minimal loss and excellent accuracy. The simulation is able to differentiate among hard and non-hard exudates, as evidenced by its elevated AUC and accuracy values. Yet, the model could at times overlook some hard exudates given the significantly reduced sensitivities. These data points act as crucial indicators for assessing and enhancing the categorization algorithm's efficacy for hard exudates in ocular pictures.

4.3.2. Haemorhages

Valuable information about its efficacy in identifying and categorising haemorrhages in retinal pictures may be learned from the statistical examination of the haemorrhages dataset. The model properly classifies pixels as haemorrhages or non-hemorrhages with an average accuracy and absolute accuracy of 0.999264, respectively. This shows the the simulation accurately depicts haemorrhages' distinctive characteristics and can separate these from various other ocular formations. The durability of the model is further supported by the low standard deviation of 0.000230 (as per Table 9), which shows that the correctness of the model is constant across several evaluations. The range of variance in the model's accuracy across several samples is highlighted by the least precision of 0.998607 and the highest confidence of 0.999657. Figure 15, Figure 16 and Figure 17 describe the training and UHI network on Haemorrhages dataset.

Table 9
Statistical analysis

Statistical test	Accuracy	Binary accuracy	AUC	Specificty	Sensitivity
mean	0.999264	0.999264	0.974184	0.999812	0.939016
std	0.000230	0.000230	0.008255	0.000072	0.018514
min	0.998607	0.998607	0.956402	0.999574	0.901336
25%	0.999184	0.999184	0.968939	0.999759	0.927588
50%	0.999331	0.999331	0.973982	0.999836	0.938238
75%	0.999403	0.999403	0.979575	0.999862	0.948270
max	0.999657	0.999657	0.990752	0.999934	0.976834

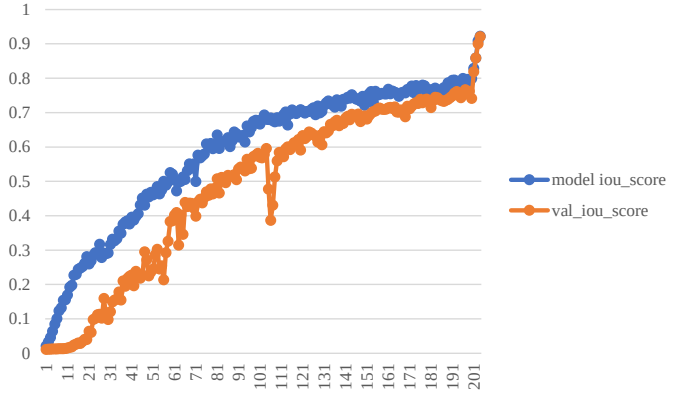


Figure 15. Training IOU_score

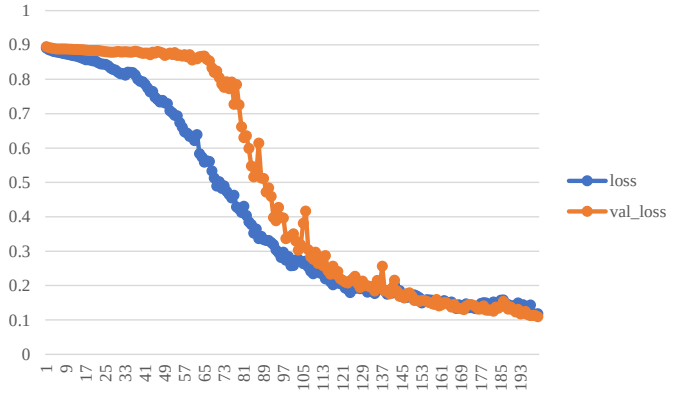
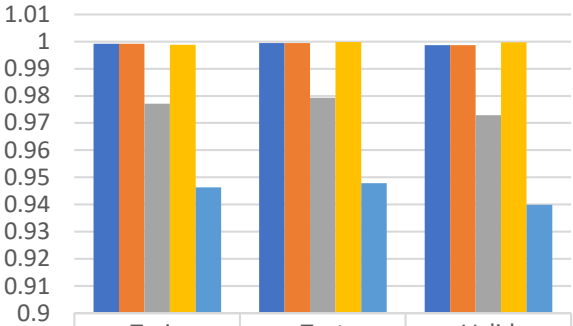


Figure 16. Focal+dice loss

The average value of 0.974184 for the AUC (Area Under the Curve) indicates that the algorithm has a high capacity to discriminate among haemorrhages and non-hemorrhages. Higher accuracy is indicated by a larger AUC, with numbers near 1 suggesting good discriminating. Consistent achievement with regard to of AUC is indicated by a small variance of 0.008255. The AUC values of 0.956402, 0.990752, and 0.956402 respectively show the range of variance in the model's capacity to discriminate between various samples. In addition, the algorithm's ability to precisely detect non-hemorrhage pixel is demonstrated by its average specific of 0.999812, which adds to its total accuracy. The model's capacity to correctly detect haemorrhage pixels is shown by its average sensibility, which is 0.939016.



	Train	Test	Valid
Accuracy	0.999244	0.999469	0.998732
Binary Accuracy	0.999244	0.999469	0.998732
AUC	0.977122	0.979326	0.972912
Specificity	0.998805	0.999862	0.999685
Sensitivity	0.946274	0.947909	0.939895

Figure 17. UHI model performance

For assessing the algorithm's effectiveness in accurately identifying haemorrhages, both metrics are essential. The ranges of quartiles (25%, 50%, and 75%) give information about the metrics' dispersion and show the variation of the model's effectiveness in various samples. The algorithm's excellent precision, good discrimination capacity, and even effectiveness in properly recognising haemorrhage and non-hemorrhage pixels are all demonstrated by the statistical evaluation of the haemorrhages dataset. The findings show how well the algorithm performs haemorrhage detection and segmentation in ocular pictures, which can be useful for identifying and tracking retinal disorders.

The initial picture in this illustration is a macular the fundus picture with retinopathy caused by diabetes symptoms. Numerous anomalies, such as exudates and microaneurysms, are visible in the picture. Specialists create the appropriate mask, which shows the areas that are important where these aberrations are found. The mask assists in emphasising the regions that require focus for additional investigation and evaluation. The network processes the picture for classification utilising the Dense-ED-UHI: Encoder Decoder based U-Net Hybrid Inception model. Intricate characteristics may be captured and accurate segmentation operations can be carried out using the U-Net Fusion framework. The aneurysm and exudate zones are correctly identified and highlighted on the segments map produced by U-Net Inception, which closely matches the surface truth mask. Figure 18 and 19 illustrates the prediction of U-Net Inception and Resnet Models.

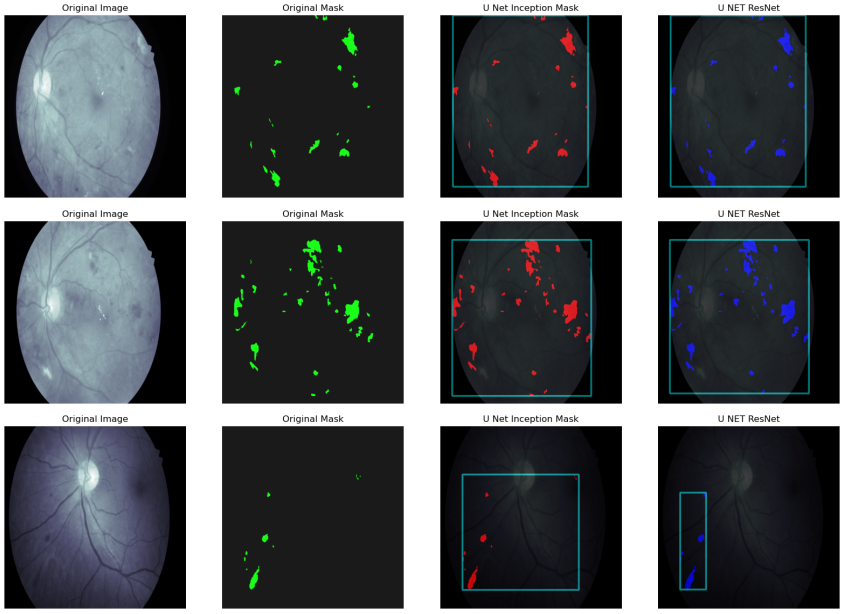


Figure 18. Mask prediction and BBox using Dense-ED-UHI: encoder decoder based unet hybrid inception (proposed model)

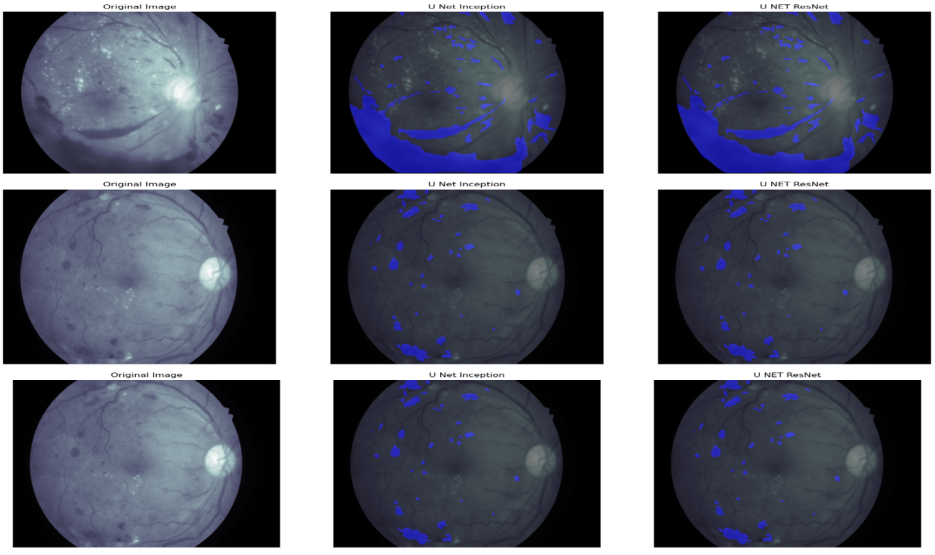


Figure 19. Comparison with other state of the art backbone (Resent50)

4.3.3. Microaneurysms

Figure 20, Figure 21 and Figure 22 describe the training and UHI network on Microaneurysms dataset. As per Table 10 and Figure 22 the algorithm has an excellent degree of general precision in detecting microaneurysms, as indicated by a median accuracy of 0.999264. It indicates that a large number of pixels in the picture, comprising all genuine positives and true negatives, are classified properly by the model.

Accuracy in Binary: the accuracy in basic measure similarly exhibits a high average value of 0.999264. This statistic, which solely takes into account true positives and true negatives, assesses the precision with which microaneurysms are classified. The elevated score shows how well the representation can separate microaneurysms from other areas of the image.

Table 10
Statistical analysis

Statistical test	Accuracy	Binary accuracy	AUC	Specificty	Sensitivity
mean	0.999264	0.999264	0.974184	0.999812	0.939016
std	0.000230	0.000230	0.008255	0.000072	0.018514
min	0.998607	0.998607	0.956402	0.999574	0.901336
25%	0.999184	0.999184	0.968939	0.999759	0.927588
50%	0.999331	0.999331	0.973982	0.999836	0.938238
75%	0.999403	0.999403	0.979575	0.999862	0.948270
max	0.999657	0.999657	0.990752	0.999934	0.976834

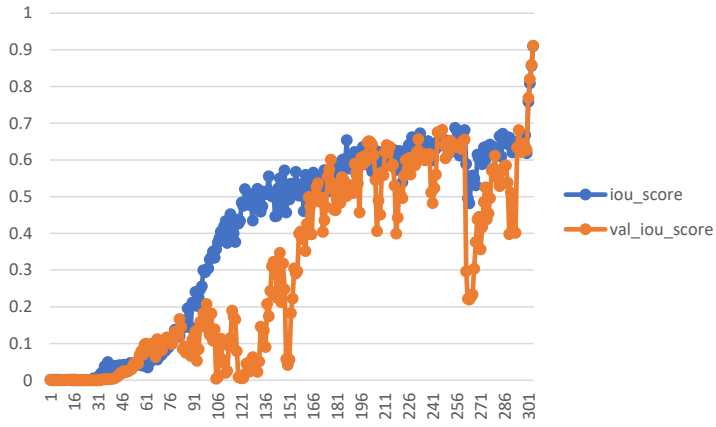


Figure 20. Training IOU_score

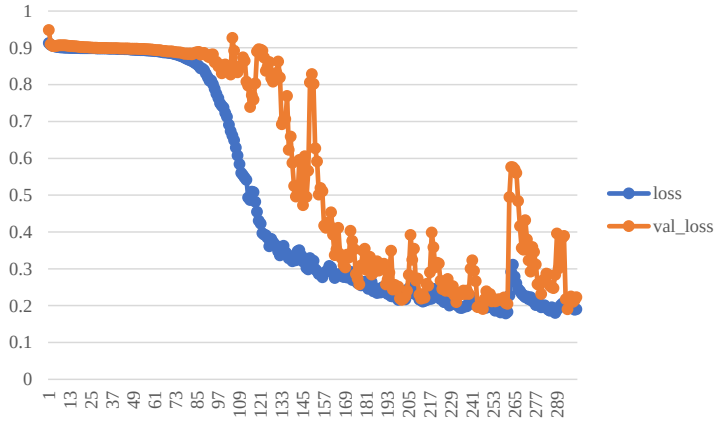
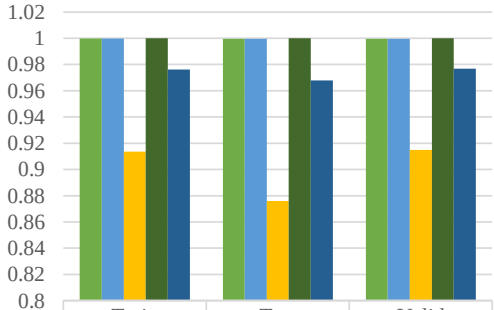


Figure 21. Focal+dice loss

The system's capacity to accurately detect non-microaneurysm areas in the picture is demonstrated by its average specificity of 0.999812 for the dataset. This means that the algorithm has a small percentage of fake positives and an elevated genuine negative rate. AUC (Area Under the Curve): the region underneath the Receiver Operating Characteristics (ROC) curve is represented by the AUC value of 0.974184. It gives an indication of how well the simulation can distinguish among microaneurysm-prone and non-prone locations. The statistical model performs better at differentiating among the two groups based on the area under the curve (AUC).



	Train	Test	Valid
Accuracy	0.99974	0.999681	0.99964
Binary Accuracy	0.99974	0.999681	0.99964
AUC	0.913566	0.875894	0.914899
Specificity	0.99997	0.999965	0.999958
Sensitivity	0.9761484	0.9677475	0.9768875

Figure 22. UHI model performance

A median rating of 0.939016 indicates how well the algorithm can identify microaneurysms. It displays the real negative percentage, which shows how well the algorithm can detect all of the microaneurysms visible in the picture. According to the statistical results, the algorithm does a good job of detecting microaneurysms generally. The model's efficiency in reliably classifying microaneurysm areas is demonstrated by its outstanding accuracy, dichotomous accuracy, and AUC scores. Furthermore, the model's excellent sensitivity and specificity scores show that it can reliably detect most aneurysm and distinguish non-microaneurysm locations. It's significant to keep in mind that the usual deviation numbers reveal information about the model's efficacy and inconsistency across different examples or datasets. The accuracy, categorical precision, AUC, particularity, and sensitivities have relatively small variances, which shows that the algorithm performs consistently throughout the assessed data.

In conclusion, the statistical evaluation shows that the algorithm consistently performs well throughout various data sets and reaches a high level of precision and efficacy in recognising microaneurysms. These findings point to the algorithm's possibility of helping doctors identify and diagnose microaneurysms promptly, and this is essential for treating patients with diabetes along with different ocular illnesses. Figure 23 and Figure 24 illustrates the prediction of U-Net Inception and Resnet Models.

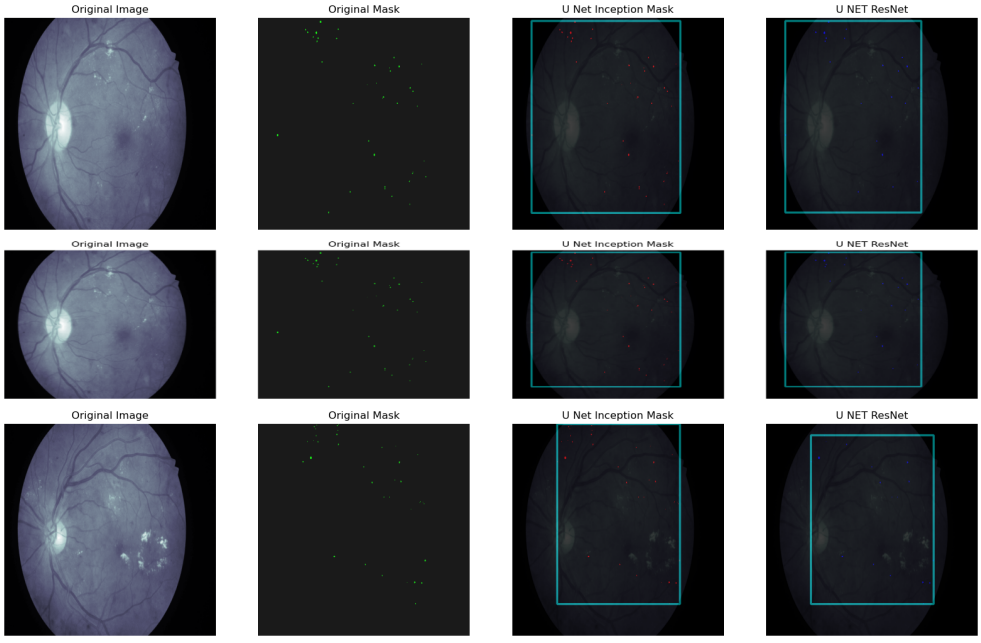


Figure 23. Mask prediction and BBox using Dense-ED-UHI: encoder decoder based U-Net hybrid inception (proposed model)

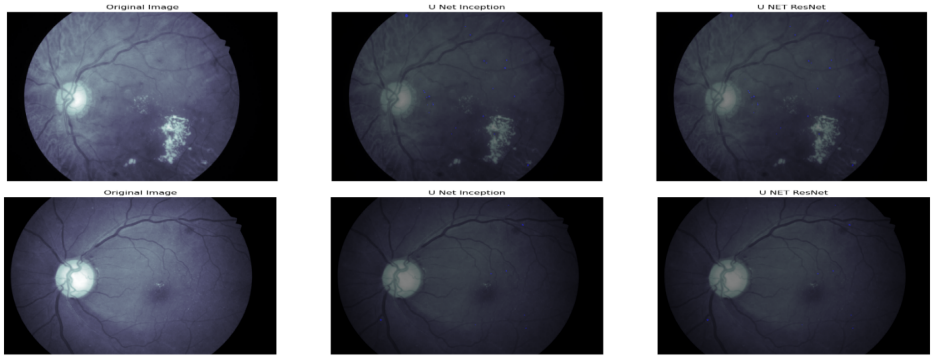


Figure 24. Comparison with other state of the art backbone (Resent50)

4.3.4. Soft exudates

The training and UHI networks on the hemorrhages dataset are shown in Figure 25, and Figure 26. The method has an outstanding level of general precision in recognizing soft exudates, as shown by a median accuracy of 0.999521, as shown in Table 11 and Figure 27. The efficiency of the Hybrid U-Net Inception model on the Soft Exudates IRID datasets seems nothing short of remarkable. The algorithm demonstrates its exceptional capacity to precisely categorise soft exudates with absolute accuracy scores of roughly 99.95% on the initial training, testing, and validation sets. Additionally, the AUC values are excellent, especially on the test and validation sets, where they are 99.90% and 99.99%, respectively, demonstrating the model's great ability to distinguish between positive and negative instances. In order to reduce the number of false positives in medical imaging analysis, the model also excels in specificity, scoring over 99.97% on all datasets. Furthermore metrics, sensitivity – which measures the model's capacity to identify positive cases – remains quite high, particularly for the test and validation sets, where it is 96.08% and 96.31%, respectively. In clinical settings, when avoiding false positives is crucial, this minor trade-off between specificity and sensitivity may be acceptable. After further validation and testing, the Hybrid U-Net Inception model exhibits strong and encouraging performance, demonstrating its potential value in real-world clinical applications.

Table 11
Statistical analysis

Statistical test	Accuracy	Binary accuracy	AUC	Specificty	Sensitivity
mean	0.999264	0.999264	0.974184	0.999812	0.939016
std	0.000230	0.000230	0.008255	0.000072	0.018514
min	0.998607	0.998607	0.956402	0.999574	0.901336
25%	0.999184	0.999184	0.968939	0.999759	0.927588
50%	0.999331	0.999331	0.973982	0.999836	0.938238
75%	0.999403	0.999403	0.979575	0.999862	0.948270
max	0.999657	0.999657	0.990752	0.999934	0.976834

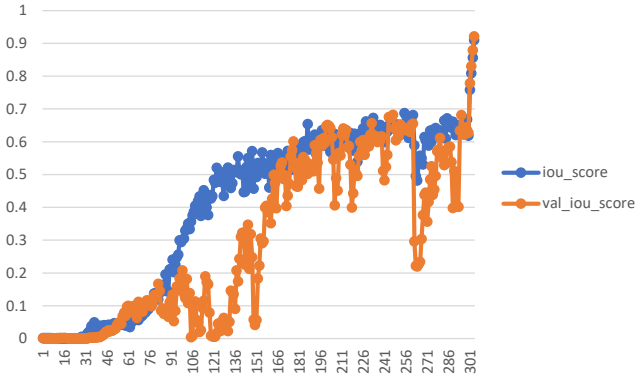


Figure 25. Training IOU_score

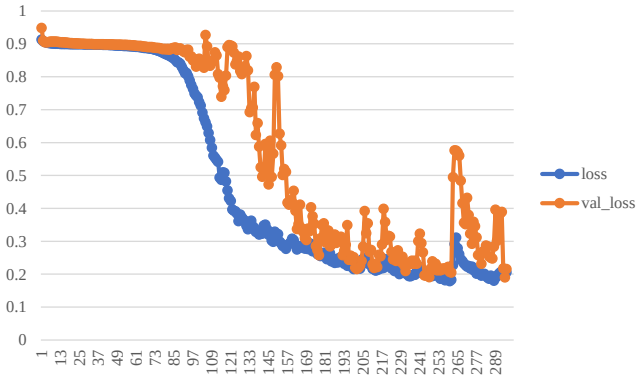
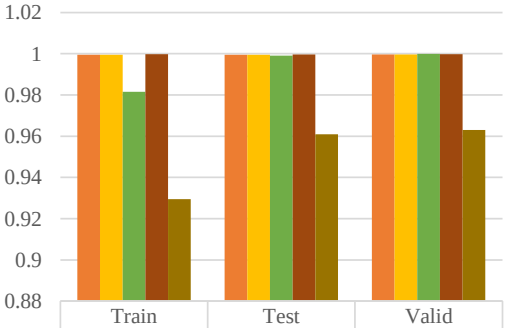


Figure 26. Focal+Dice Loss



	Train	Test	Valid
Accuracy	0.999545	0.999485	0.999671
Binary Accuracy	0.999545	0.999485	0.999671
AUC	0.981519	0.999047	0.999875
Specificity	0.999788	0.999668	0.999818
Sensitivity	0.929511	0.960885	0.963076

Figure 27. UHI model performance

Figures 28 and Figure 29 show how the Unet Inception and Resnet Models predict.

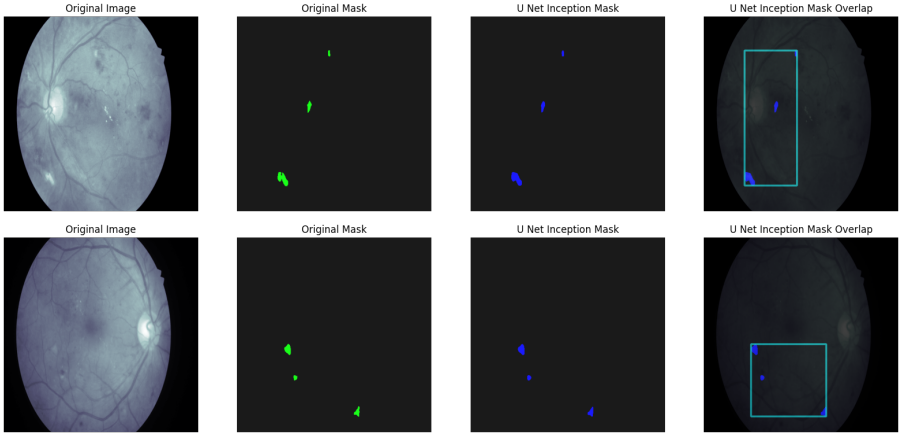


Figure 28. Mask prediction and BBox using Dense-ED-UHI: encoder decoder based U-Net hybrid inception (proposed model)

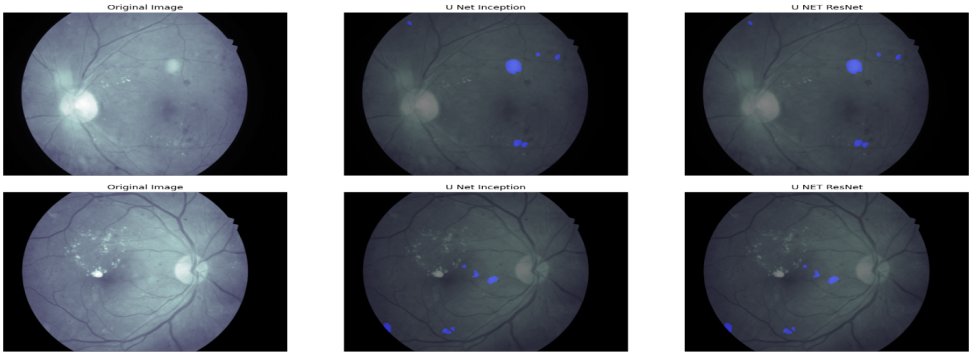


Figure 29. Comparison with other state of the art backbone (Resent50)

4.4. Comparative analysis

The comparison Table 12 illustrates how the Dense-ED-UHI model fared better than other evaluated models when specificity and accuracy were taken into account.

The model put out by [21] outperformed earlier research, achieving a sensitivity of 80.32% and a specificity of 99.83% across many datasets. In their analysis on the e-ophtha dataset for exudates identification, Chudzik et al. [9] found that specificities were higher than 99.97% and sensitivities ranged from 84.58% to 86.66%. Li et al.'s [23] work on the DDR dataset revealed an accuracy of 82.84%, whilst Playout et al.'s [32] research attained a sensitivity of 80.02% and a specificity of 82%.

Xia et al. [44], in comparison, obtained 82% sensitivity and 98.3% specificity on the CHASE dataset and 81.2% sensitivity and 98% specificity on the DRIVE dataset. Sensitivities and specificities were found to be between 80 and 83% in Oh et al.'s [27] research using ETDRS datasets.

Table 12
Comparative study

Study	Dataset	Sensitivity [%]	Specificity [%]	Accuracy [%]
[21]	HEIMED, e-ophtha, DIARETDB	80.32	99.83	–
[9]	e-ophtha (exudates)	86.66	99.98	–
[10]	e-ophtha (exudates)	84.58	99.97	–
[32]	e-ophtha (exudates)	80.02	–	82
[23]	DDR	–	–	82.84
[44]	DRIVE	81.2	98	95.4
	CHASE	82	98.3	97
[27]	ETDRS 7SF	83.38	83.41	83.3
	ETDRS F1-F2	80.60	80.61	80.6

5. Discussion

The presented research focuses on addressing the challenges of diabetic retinopathy (DR) through a novel approach employing a combinative method utilizing U-Net with a modified Inception architecture. One of the leading causes of visual impairment worldwide is diabetic retinopathy, for which prompt identification is essential to successful treatment. Deep neural architecture, notably encoder-decoder modelling using convolutional architectures like Inception and Residual Connection, is included into the suggested model. To improve performance in image processing tasks like semantic segmentation or image-to-image translation, the U-Net and Inception architectures are combined. Prominent for its effectiveness in semantic segmentation, the encoder-decoder structure of the U-Net architecture allows for accurate localization. Conversely, GoogLeNet's Inception design, which minimises computational cost by using convolutional filters of different widths inside a single layer, is excellent at collecting multi-scale information. By combining these designs, Inception modules are added to the U-Net encoder (3 and 4), improving feature extraction at various scales. Similar changes are made to the decoder, where layers of upsampling combine with feature concatenation from scale-matched encoder features. Most importantly, skip connection retention guarantees gradients and spatial information are preserved during training. The last layers are customized for each job; for pixel-wise predictions, they usually consist of convolutional layers followed by sigmoid or softmax

activations. This combination combines the multi-scale feature extraction power of Inception with the accurate localization expertise of U-Net to provide better results, especially in applications where collecting data at several sizes is critical.

The efficiency of the modified Inception deep feature extractor is shown by the study's validation on the IDRid APTOS 2019 contest dataset, which produced an exceptional classification accuracy of 94.21% across classes, outperforming the comparison with Resnet. The suggested model attains a segmentation test accuracy of 99.90% on the IDRid dataset across various classes, surpassing Resnet. Specifically, it achieves testing accuracies of 92.62% for retinopathy grade classification and 95.80% for DME classification. Moreover, the model demonstrates mean accuracies of 99.89%, 99.84%, 99.96%, and 99.97% for Hard Exudates, hemorrhage, Microaneurysms, and Soft Exudates respectively. A hybrid dense-ED-UHI model, an encoder-decoder-based U-Net with an inception architecture, is used in the article. It is cross-validated 15 times across four classes of IDRID dataset.

The model performs very well; the haemorrhages dataset examination confirms this, with mean accuracy and binary accuracy reaching 99.84%. The model is successful in identifying true positives and negatives, as seen by the AUC and sensitivity values, which highlight the model's accuracy in differentiating between advantageous and harmful scenarios. In the context of DR classification systems, the research emphasizes the relevance of the custom U-Net model with Inception architecture and a novel up-sampling technique employing pixel-wise periodic shuffling convolution. The comprehensive validation of the proposed framework against extant literature presents it as a promising advancement that might lead to better outcomes and therapies for people with visual impairment worldwide. When it comes to particularity and accuracy, the Dense-ED-UHI model performs better than the other models that were assessed. The dataset's ability to properly identify and distinguish retinal disorders is supported by its excellent accuracy scores. Because of the model's proven specificity, false positives are less likely to occur and unfavourable instances are accurately recognised. Together, these results highlight the Dense-ED-UHI model's effectiveness and dependability, highlighting its potential for accurately classifying retinal disorders.

Lastly, with regard to microaneurysms, the mathematical models show remarkable precision and binary accuracy. With a 99.96% median accuracy, AUC values, sensitivity values, and correct classification of unfavourable situations, the model demonstrates its ability to discriminate between advantageous and harmful cases while effectively identifying true positives. The high sensitivity values demonstrate even more how well the model categorises unfavourable situations. The study report concludes by presenting a unique strategy for treating diabetic retinopathy using an advanced combinative technique. The suggested approach, verified on many datasets, has exceptional precision and efficacy in categorising retinal disorders, presenting encouraging opportunities for enhanced identification and management in the worldwide context of visual impairment.

6. Conclusion

Although one is caused by ageing and the other by varied reasons, diabetic retinopathy is complicated, multifaceted diseases, and our understanding of these diseases is continually changing. Our systems of categorising these have undergone numerous revisions over time, and they have had to be revised and updated to keep pace with medical knowledge and technological developments. A new classification system must consider the tremendous advancements made in the last few decades in disease pathophysiology, imaging technologies, artificial intelligence, and treatment. In the training set as well as the test set, the accuracy of the model was 92.5%, demonstrating its capacity to make accurate assumptions about unobserved data. At 99.66% on the test set, the level of sensitivity, or real positive rate, is also incredibly high. The simulations exhibit remarkable accuracy for the examination of hard exudates, with an average precision of 99.93%. The sensitiveness scores, with a mean sensitivity of 92.94%, show how well the models can detect true positives. The models' high ability to differentiate among both positive and negative situations is reflected in the AUC values, which range from 0.9426 to 0.9827. The precise classification of adverse cases by the models is highlighted by their elevated specific values, which range from 0.9996 to 0.9999. The data analysis indicates that the models are effective in recognising and categorising exudates that are hard overall.

The hypotheses performed exceptionally well in the evaluation of the haemorrhages dataset, with a mean accuracy and binary accuracy of 99.93%. The AUC values, which range from 0.9564 to 0.9908, show how well the models can distinguish between both benign and detrimental cases. The algorithms' extremely sensitive values, which range from 0.9013 to 0.9768, indicate their accuracy in detecting genuine positives. The precise classification of negative cases by the models is highlighted by their elevated sensitivity values, which range from 0.9996 to 0.9999. These findings show that the algorithms are capable of identifying and categorising haemorrhages The new DR classification systems make use of a customised UHI or U-Net model to conceptually segment using Inception as a spatial up-sampling method that uses pixel-wise periodic shuffling convolution. It has been thoroughly validated by comparison to the body of literature and should lead to better treatments and results for the millions of individuals who suffer from visual loss globally. In terms of correctness and accuracy, the Dense-ED-UHI: Encoder Decoder based Unet Hybrid Inception models scored better than other models that were evaluated. High accuracy ratings were attained, demonstrating the datasets' capacity to appropriately diagnose and separate the retinal diseases. The algorithm furthermore showed good specificity, demonstrating its capacity to accurately recognise the adverse cases (non-pathological areas) and prevent false positives. These findings demonstrate the efficiency and dependability of the Dense-ED-UHI: Encoder Decoder based Unet Hybrid Inception framework for segmenting retinal disease.

The mathematical models are highly accurate and binary accurate when it comes to microaneurysms, with a median accuracy of 99.93%. The AUC numbers, which

range from 0.9564 to 0.9908, show how well the models can distinguish among both beneficial and detrimental cases. The sensitive values, which range from 0.9013 to 0.9768, show how well the models are able to detect true positives. The precise classification of unfavourable cases by the models is highlighted by their elevated sensitivity values, which range from 0.9996 to 0.9999. The Dense-ED-UHI: Encoder Decoder based U-Net Hybrid Inception structure's ability to use the structure known as Inception and gather global as well as local pigment data within the image which constitutes one of its main advantages. This enables the model to effectively record fine details in retina pictures by utilising multi-scale characteristics.

In conclusion, segmenting retinal disease can greatly benefit from the application of deep learning models, particularly the UHI system and the possible integration of the Classification network for broad category classification. These simulators have proven to be highly accurate, durable, and capable of automating and assisting in the early identification and recognition of retinal disorders. Continuing this field's study and development could result in better healthcare.

Authors contribution

AB and VM created the idea for the study. AB led the writing of this paper. AB analyzed and interpreted the data. AB contributed to the implementation of the paper as well as to the identification of theoretical problems. AB contributed to the writing of the paper as well as participated in revising this manuscript. VM encouraged AB to investigate and supervised the findings of this work. All authors contributed significantly to the writing of the paper, discussed the results, reviewed the final draft, and approved the final version of this manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

The data URL supporting this article has shown in Table 13.

Table 13
Data availability statement

Dataset name	Availability of data	URL
IDRID	Openly accessible data stored in a public database.	The data that support the findings of this study are available in Grand-Challenge at https://idrid.grand-challenge.org.

Conflict of interest

No authors have disclosed any conflicts of interest. The authors affirm that they have no known financial or interpersonal conflicts that would have appeared to have an impact on the research presented in this study.

References

[1] Abràmoff M.D., Lavin P.T., Birch M., Shah N., Folk J.C.: Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices, *npj Digital Medicine*, vol. 1(1), 39, 2018. doi: 10.1038/s41746-018-0040-6.

[2] Akram M., Khalid S., Khan S.: Identification and classification of microaneurysms for early detection of diabetic retinopathy, *Pattern Recognition*, vol. 46(1), pp. 107–116, 2013. doi: 10.1016/j.patcog.2012.07.002.

[3] Azzopardi G., Strisciuglio N., Vento M., Petkov N.: Trainable COSFIRE filters for vessel delineation with application to retinal images, *Medical Image Analysis*, vol. 19(1), pp. 46–57, 2015. doi: 10.1016/j.media.2014.08.002.

[4] Bali A., Mansotra V.: Deep Learning-based Techniques for the Automatic Classification of Fundus Images: A Comparative Study. In: *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 351–359, 2021. doi: 10.1109/ICAC3N53548.2021.9725464.

[5] Bali A., Mansotra V.: An Overview of Retinal Image Classification-Techniques and Challenges. In: *2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT)*, pp. 91–97, 2021. doi: 10.1109/ICACFCT53978.2021.9837371.

[6] Bali A., Mansotra V.: FUNDUS and OCT Image Classification Using DL Techniques. In: V.S. Rathore, S.C. Sharma, J.M.R.S. Tavares, C. Moreira, B. Surendiran (eds.), *Rising Threats in Expert Applications and Solutions. Proceedings of FICR-TEAS 2022*, Lecture Notes in Networks and Systems, vol. 434, Springer, 2022. doi: 10.1007/978-981-19-1122-4_8.

[7] Bali A., Mansotra V.: Analysis of Deep Learning Techniques for Prediction of Eye Diseases: A Systematic Review, *Archives of Computational Methods in Engineering*, vol. 31, pp. 487–520, 2024. doi: 10.1007/s11831-023-09989-8.

[8] Bali A., Mansotra V.: Multiclass multilabel ophthalmological fundus image classification based on optimised deep feature space evolutionary model, *Multimedia Tools and Applicatios*, vol. 83, pp. 49813–49843, 2024. doi: 10.1007/s11042-023-17530-z.

[9] Chudzik P., Al-Diri B., Caliva F., Ometto G., Hunter A.: Exudates segmentation using fully convolutional neural network and auxiliary codebook. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 770–773, 2018. doi: 10.1109/embc.2018.8512354.

[10] Chudzik P., Majumdar S., Caliva F., Al-Diri B., Hunter A.: Exudate segmentation using fully convolutional neural networks and inception modules. In: E.D. Angelini, B.A. Landman (eds.), *Medical Imaging 2018: Image Processing*, vol. 10574, 1057430, International Society for Optics and Photonics, SPIE, 2018. doi: 10.1117/12.2293549.

[11] Di Cesare M.: Global trends of chronic non-communicable diseases risk factors, *European Journal of Public Health*, vol. 29(Supplement_4), ckz185-196, 2019. doi: 10.1093/eurpub/ckz185.196.

[12] Early Treatment Diabetic Retinopathy Study Research Group: Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12, *Ophthalmology*, vol. 98(5), pp. 823–833, 1991.

[13] Franklin S.W., Rajan S.E.: Computerized screening of diabetic retinopathy employing blood vessel segmentation in retinal images, *Biocybernetics and Biomedical Engineering*, vol. 34(2), pp. 117–124, 2014. doi: 10.1016/j.bbe.2014.01.004.

[14] Fraz M.M., Barman S.A., Remagnino P., Hoppe A., Basit A., Uyyanonvara B., Rudnicka A.R., Owen C.G.: An approach to localize the retinal blood vessels using bit planes and centerline detection, *Computer Methods and Programs in Biomedicine*, vol. 108(2), pp. 600–616, 2012. doi: 10.1016/j.cmpb.2011.08.009.

[15] Gardner G.G., Keating D., Williamson T.H., Elliott A.T.: Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool, *British Journal of Ophthalmology*, vol. 80(11), pp. 940–944, 1996. doi: 10.1136/bjo.80.11.940.

[16] Gargeya R., Leng T.: Automated identification of diabetic retinopathy using deep learning, *Ophthalmology*, vol. 124(7), pp. 962–969, 2017. doi: 10.1016/j.ophtha.2017.02.008.

[17] Gulshan V., Peng L., Coram M., Stumpe M.C., Wu D., Narayanaswamy A., Venugopalan S., *et al.*: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *JAMA*, vol. 316(22), pp. 2402–2410, 2016. doi: 10.1001/jama.2016.17216.

[18] Guo S., Li T., Kang H., Li N., Zhang Y., Wang K.: L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images, *Neurocomputing*, vol. 349, pp. 52–63, 2019. doi: 10.1016/j.neucom.2019.04.019.

[19] Hagos M.T., Kant S.: Transfer learning based detection of diabetic retinopathy from small dataset, *arXiv preprint arXiv:190507203*, 2019.

[20] Haldorai A., Ramu A., Chow C.O.: Editorial: Big Data Innovation for Sustainable Cognitive Computing, *Mobile Networks and Applications*, vol. 24, pp. 221–223, 2019. doi: 10.1007/s11036-018-1198-5.

[21] Imani E., Pourreza H.R.: A novel method for retinal exudate segmentation using signal separation algorithm, *Computer Methods and Programs in Biomedicine*, vol. 133, pp. 195–205, 2016. doi: 10.1016/j.cmpb.2016.05.016.

[22] Kassani S.H., Kassani P.H., Khazaeinezhad R., Wesolowski M.J., Schneider K.A., Deters R.: Diabetic Retinopathy Classification Using a Modified Xception Architecture. In: *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 1–6, IEEE, 2019. doi: 10.1109/isspit47144.2019.9001846.

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

[24] Lian S., Li L., Lian G., Xiao X., Luo Z., Li S.: A Global and Local Enhanced Residual U-Net for Accurate Retinal Vessel Segmentation, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18(3), pp. 852–862, 2019. doi: 10.1109/tcbb.2019.2917188.

[25] Mansour R.F.: Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy, *Biomedical Engineering Letters*, vol. 8, pp. 41–57, 2018.

[26] Noh H., Hong S., Han B.: Learning deconvolution network for semantic segmentation. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528, 2015. doi: 10.1109/iccv.2015.178.

[27] Oh K., Kang H.M., Leem D., Lee H., Seo K.Y.: Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images, *Scientific Reports*, 2021. doi: 10.1038/s41598-021-81539-3.

[28] Orlando J., Prokofyeva E., Del Fresno M., Blaschko M.: An ensemble deep learning based approach for red lesion detection in fundus images, *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 115–127, 2018. doi: 10.1016/j.cmpb.2017.10.017.

[29] Orlando J.I., Prokofyeva E., Del Fresno M., Blaschko M.B.: An ensemble deep learning based approach for red lesion detection in fundus images, *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 115–127, 2018. doi: 10.1016/j.cmpb.2017.10.017.

[30] Osareh A., Mirmehdi M., Thomas B., Markham R.: Automated identification of diabetic retinal exudates in digital colour images, *The British Journal of Ophthalmology*, vol. 87(10), pp. 1220–1223, 2003. doi: 10.1136/bjo.87.10.1220.

[31] Partovi M., Rasta S.H., Javadzadeh A.: Automatic detection of retinal exudates in fundus images of diabetic retinopathy patients, *Journal of Research in Clinical Medicine*, vol. 4(2), pp. 104–109, 2016. doi: 10.15171/jarcm.2016.017.

[32] Playout C., Duval R., Cheriet F.: A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images, *IEEE Transactions on Medical Imaging*, vol. 38(10), pp. 2434–2444, 2019. doi: 10.1109/tmi.2019.2906319.

[33] Quellec G., Charrière K., Boudi Y., Cochener B., Lamard M.: Deep image mining for diabetic retinopathy screening, *Medical Image Analysis*, vol. 39, pp. 178–193, 2017. doi: 10.1016/j.media.2017.04.012.

[34] Qureshi I., Ma J., Abbas Q.: Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning, *Multimedia Tools and Applications*, vol. 80, pp. 11691–11721, 2021. doi: 10.1007/s11042-020-10238-4.

[35] Romero-Aroca P., Verges-Puig R., de la Torre J., Valls A., Relaño-Barambio N., Puig D., Baget-Bernaldiz M.: Validation of a deep learning algorithm for diabetic retinopathy, *Telemedicine and e-Health*, vol. 26(8), pp. 1001–1009, 2020. doi: 10.1089/tmj.2019.0137.

[36] Roychowdhury S., Koozekanani D.D., Parhi K.K.: DREAM: diabetic retinopathy analysis using machine learning, *IEEE Journal of Biomedical and Health Informatics*, vol. 18(5), pp. 1717–1728, 2013. doi: 10.1109/jbhi.2013.2294635.

[37] Shanthi T., Sabeenian R.S.: Modified Alexnet architecture for classification of diabetic retinopathy images, *Computers & Electrical Engineering*, vol. 76, pp. 56–64, 2019. doi: 10.1016/j.compeleceng.2019.03.004.

[38] Shen W., Zhou M., Yang F., Yu D., Dong D., Yang C., Zang Y., Tian J.: Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification, *Pattern Recognition*, vol. 61, pp. 663–673, 2017. doi: 10.1016/j.patcog.2016.05.029.

[39] Sinthanayothin C., Boyce J.F., Williamson T.H., Cook H.L., Mensah E., Lal S., Usher D.: Automated detection of diabetic retinopathy on digital fundus images, *Diabetic Medicine*, vol. 19(2), pp. 105–112, 2002. doi: 10.1046/j.1464-5491.2002.00613.x.

[40] Soomro T.A., Gao J., Khan T., Hani A.F.M., Khan M.A., Paul M.: Computerised approaches for the detection of diabetic retinopathy using retinal fundus images: a survey, *Pattern Analysis and Applications*, vol. 20, pp. 927–961, 2017. doi: 10.1007/s10044-017-0630-y.

[41] Sopharak A., Uyyanonvara B., Barman S., Williamson T.H.: Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods, *Computerized Medical Imaging and Graphics*, vol. 32(8), pp. 720–727, 2008. doi: 10.1016/j.compmedimag.2008.08.009.

[42] Ting D.S.W., Tan G.S.W., Agrawal R., Yanagi Y., Sie N.M., Wong C.W., San Yeo I.Y., Lee S.Y., Cheung C.M.G., Wong T.Y.: Optical coherence tomographic angiography in type 2 diabetes and diabetic retinopathy, *JAMA Ophthalmology*, vol. 135(4), pp. 306–312, 2017. doi: 10.1001/jamaophthalmol.2016.5877.

[43] Walter T., Massin P., Erginay A., Ordonez R., Jeulin C., Klein J.C.: Automatic detection of microaneurysms in color fundus images, *Medical Image Analysis*, vol. 11(6), pp. 555–566, 2007. doi: 10.1016/j.media.2007.05.001.

[44] Xia Z., Liu Q., Zhang Z., Wang Y.: Segmentation of exudates in fundus images using fully convolutional neural networks, *Journal of Medical Imaging and Health Informatics*, 2020.

[45] Xiancheng W., Wei L., Bingyi M., He J., Jiang Z., Xu W., Ji Z., Hong G., Zhaomeng S.: Retina blood vessel segmentation using a U-Net based Convolutional neural network. In: *Procedia Computer Science: International Conference on Data Science (ICDS 2018)*, pp. 8–9, 2018.

[46] Xu X., Tan T., Xu F.: An improved U-Net architecture for simultaneous arteriole and venule segmentation in fundus image. In: M. Nixon, S. Mahmoodi, R. Zwiggelaar (eds.), *Medical Image Understanding and Analysis: 22nd Conference, MIUA 2018, Southampton, UK, July 9–11, 2018, Proceedings*, pp. 333–340, Springer International Publishing, 2018. doi: 10.1007/978-3-319-95921-4_31.

[47] Yan Z., Han X., Wang C., Qiu Y., Xiong Z., Cui S.: Learning mutually local-global u-nets for high-resolution retinal lesion segmentation in fundus images. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 597–600, IEEE, 2019. doi: 10.1109/isbi.2019.8759579.

[48] Zhang W., Zhong J., Yang S., Gao Z., Hu J., Chen Y., Yi Z.: Automated identification and grading system of diabetic retinopathy using deep neural networks, *Knowledge-Based Systems*, vol. 175, pp. 12–25, 2019. doi: 10.1016/j.knosys.2019.03.016.

[49] Zhang X., Thibault G., Decencière E., Marcotegui B., Laû B., Danno R., Cazuguel G., *et al.*: Exudate detection in color retinal images for mass screening of diabetic retinopathy, *Medical Image Analysis*, vol. 18(7), pp. 1026–1043, 2014. doi: 10.1016/j.media.2014.05.004.

[50] Zhao Y.Q., Wang X.H., Wang X.F., Shih F.Y.: Retinal vessels segmentation based on level set and region growing, *Pattern Recognition*, vol. 47(7), pp. 2437–2446, 2014. doi: 10.1016/j.patcog.2014.01.006.

Affiliations

Le Akanksha Bali
University of Jammu, Department of Computer Science and IT, akankshabali5@gmail.com

Kuljeet Singh
University of Jammu, Department of Computer Science and IT, khalsakuljeet004@gmail.com

Vibhakar Mansotra
University of Jammu, Department of Computer Science and IT, vibhakar20@yahoo.co.in

Received: 17.12.2023
Revised: 9.06.2024
Accepted: 9.06.2024