



A unified technique for entropy enhancement based diabetic retinopathy detection using hybrid neural network

Fatima^a, Muhammad Imran^{b,d,*}, Anayat Ullah^{c,d}, Muhammad Arif^c, Rida Noor^a

^a Department of Computer Engineering, FICT, BUITEMS, Quetta, 87300, Pakistan

^b Department of Electrical Engineering, FICT, BUITEMS, Quetta, 87300, Pakistan

^c Department of Electronic Engineering, FICT, BUITEMS, Quetta, 87300, Pakistan

^d Control, Automotive, and Robotics Lab, National Center of Robotics and Automation, Rawalpindi, Pakistan



ARTICLE INFO

2000 MSC:

41A05
41A10
65D05
65D17

Keywords:

Diabetic retinopathy
Neural network
Discrete wavelet transform
Histogram
Classification
Medical data

ABSTRACT

In this paper, a unified technique for entropy enhancement-based diabetic retinopathy detection using a hybrid neural network is proposed for diagnosing diabetic retinopathy. Medical images play crucial roles in the diagnosis, but two images representing two different stages of a disease look alike. It, consequently, make the process of diagnosis extraneous and error-prone. Therefore, in this paper, a technique is proposed to address these issues. Firstly, a novel entropy enhancement technique is devised exploiting the discrete wavelet transforms to improve the visibility of the medical images by making the subtle features more prominent. Later, we designed a computationally efficient hybrid neural network that efficiently classifies diabetic retinopathy images. To examine the effectiveness of our technique, we have chosen three datasets: Ultra-Wide Filed (UWF) dataset, Asia Pacific Tele Ophthalmology Society (APOTOS) dataset, and MESSIDOR-2 dataset. In the end, we performed extensive experiments to validate the performance of our technique. In addition, the comparison of the proposed scheme – in terms of accuracy, specificity, sensitivity, precision and recall curve, and area under the curve – with some of the best contemporary schemes shows the significant improvement of our techniques in terms of diabetic retinopathy classification.

1. Introduction

Normally, humans have a glucose level between 80 mg/dl and 100 mg/dl, and it can result in the form of diabetes if it crosses this range (>100 mg/dl) [1]. Diabetes subsequently leads to several problems, such as kidney failure, bleeding teeth, lower limb confiscation, nerve failure, diabetic retinopathy (DR), and heart attack [2,3]. DR, a complication of diabetes, affects the back of the eye, called the retina, due to high blood sugar, and it can result in irreversible blindness if not timely diagnosed [4]. Fortunately, DR is a progressive disease, and it usually takes several years to reach a point where it can cause complete blindness. Therefore, the affected can be saved from permanent blindness if DR is timely diagnosed. Despite being a slow-paced disease, a study shows that in the year 2010, 3.7 million and 0.8 million, respectively, suffered from vision and irreversible vision loss due to DR, and according to an estimate, by 2030, approximately 191 million people will be affected by DR due to increasing number of diabetic patients [5,6].

A healthy retina, also called fundus, has three main parts: retina blood vessels, optic discs, and macula, as shown in Fig. 1a. Retina blood vessels are the central retina arteries, veins, and their branches [7]; an optic disc is a central and circular area located at the back of an eye, which represents the beginning of the optic nerve and is used to connect the human eye to the human brain [8]; and the macula is the portion of an eye at the center of the retina that processes sharp, clear, straight-ahead vision and also called photo-receptor cells [9]. The photo-receptors are very important neurons in the retina that detect the light and convert the light into an electrical signal and send signals to the brain, and then the brain converts those signals into images [10]. The eye performs its tasks normally until all these parts function properly; however, the vision suffers if any of these parts undergo any abnormality. Depending on the severity of the abnormality, ophthalmologists divided DR into three classes: Normal (no abnormalities), non-proliferative diabetic retinopathy (NPDR), and proliferative diabetic retinopathy (PDR). NPDR is further classified as mild, moderate, and severe [11–13], as illustrated in Table 1.

* Corresponding author. Department of Electrical Engineering, FICT, BUITEMS, Quetta, 87300, Pakistan.

E-mail addresses: mi14@my.fsu.edu, enrgr.imran@buitms.edu.pk (M. Imran).

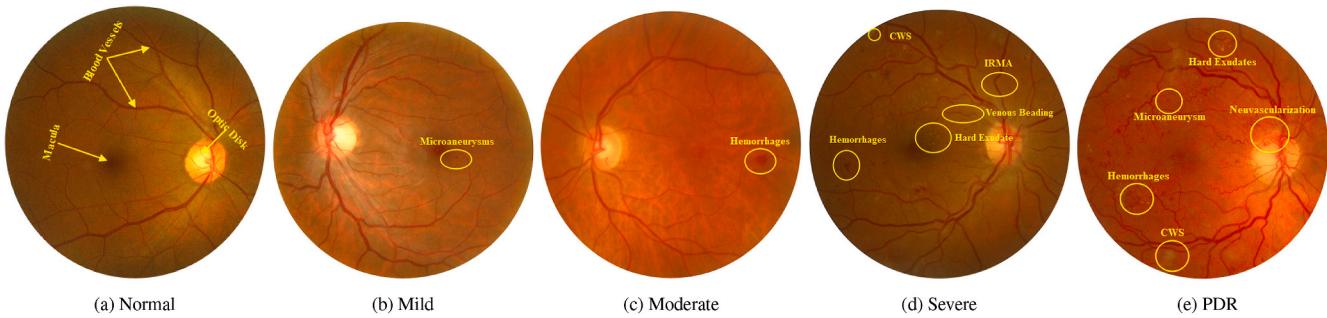


Fig. 1. Five stages of diabetic retinopathy.

Table 1
Classification of DR.

Classes	Sub-Classes	Symptoms
PDR	Normal	No abnormalities
NPDR	Mild	Microaneurysm; Hemorrhages; Intraretinal hemorrhages
	Moderate	and venous beading
	Severe	
PDR	Normal	Neovascularization and growth of new blood vessels

Normal eyes perform their functions accurately and have no symptoms of DR as shown in Fig. 1a. In the case of mild NPDR, the retina contains microaneurysm (MA), which are small circular red dots at the end of blood vessels, and, usually, MAs are not visible by the human naked eye, as shown in Fig. 1b. At the moderate NPDR stage, the MAs transform into deeper layers with flame-shaped hemorrhages inside the retina or hemorrhages, as shown in Fig. 1c. If it is not timely treated, it can progress into the severe stage, and in that case, the chances of irreversible loss increase significantly. Whereas, at the severe NPDR stage, there are more than 20 intra-retinal hemorrhages in each of the four quadrants with venous beading in two quadrants and intra-retinal microvascular abnormalities (IRMAs) in any of the four-quadrant [14, 15]], as shown in Fig. 1d. The 4 : 2 : 1 rule is applied for diagnosing the severe DR [14, 15]]. PDR is the advanced stage of DR, which leads to neovascularization that stimulates the growth of new retinal blood vessels to bypass the damaged vessels, and these new vessels are leaky, fragile, and often misdirected, as shown in Fig. 1e. When the DR reaches the PDR stages, human eyes show significant symptoms, such as floaters, leakages, blurred vision, distortion, and impaired vision loss [16] – these symptoms are discernible by the naked eye. Whereas, at the primitive stage of DR, the symptoms are insubstantial (usually less than 125 µm in diameter [17]) and are hard to be noticed by naked human eyes. Therefore, sometimes, they go unnoticed and consequently result in the form of PDR development. In addition, several patients fail to go for annual or timely eye examinations due to cost and time [14]. Researchers, therefore, propose Artificial Intelligence (AI) based techniques to detect and classify DR images.

Conventional and manual approaches require specific knowledge, expertise, specialized techniques, exorbitant equipment to detect and diagnose any medical disease. United Nations in 2020 categorized 126 countries, out of a total of 195 countries, as developing countries. All these countries lack experts – who are adept at diagnosing medical diseases – and other state-of-the-art medical facilities. On the other hand, countries with sufficient human resources and necessary medical equipment suffer from outrageous costs. Contrarily, compared to the conventional diagnostic methods, artificial-intelligence-assisted methods can be cheap and accessible everywhere. Over the last few years, artificial intelligence has obtained outstanding results in detecting and recognizing different objects within an image. For instance, in 2012, a deep learning model called AlexNet [18], based on convolutional neural networks (CNN), achieved 84.7% accuracy. The model optimized approximately 62 million parameters and trained on 1.4 million images.

Several models, VGG16, VGG19, ResNet50, Inception, Xception, etcetera, came along the way to improve the accuracy and obtained more than 90% accuracy. The tremendous success of CNN-based models persuades researchers to use them in the medical field for diagnosing diseases using medical data, especially images. Over the few years, artificial-intelligence-based techniques have obtained tremendous results in diagnosing numerous diseases, such as Alzheimer's [19–21], arrhythmia [22–24], autism [25–27], breast cancer [28–30], dental cavities [31–33], gram stains [34–36], lung cancer [37–39], onychomycosis [40,41], skin cancer [42–44], etcetera. As we have formerly discussed, the CNN-based models are as successful for medical images as for general-purpose images. However, medical data are highly imbalanced; for example, healthier people are far more, fortunately, than sick ones, and, consequently, the data related to healthy people is far greater than those suffering from any disease. This great disproportion in data makes the network biased towards the majority class. In this paper, therefore, we have tried to solve this problem and to do so, we have divided the proposed approach into two parts. In the first part, the pre-processing part, we have devised a novel entropy enhancement technique to make the subtle and less comprehensible features of a medical image more prominent and visible. As the last step, we designed a computationally efficient hybrid neural network to train small and imbalanced datasets with satisfactory results.

1.1. Related work

A deep learning model, called Hyperparameter Tuning Inception-V4 (HPTI-V4) model, was proposed in Ref. [2] to detect and classify color fundus images. The researchers in Ref. [2], first used Contrast Limited Adaptive Histogram Equalization (CLAHE) technique to enhance the details of images. Later, they segmented the region of interest using the histogram-based segmentation algorithm. At the final stage, they fed the segmented images to their proposed model, and the model works well, especially for smaller datasets, like MESSIDOR. Authors of [5] proposed a deep learning network based on ResNet-32 (a residual network with 34 layers) to classify DR and used Ultra-Wide Field (UWF) dataset. In addition to that, they employed automatic segmentation of the Early Treatment of Diabetic Retinopathy Study (ETDRS) 7 standard fields (7SF) to remove undesirable components from the fundus images. Later, they fed the segmented images to their proposed architecture for DR classification. In Ref. [16], authors used five existing well-known convolutional neural network models – ResNet50, Inception-V3, Xception, Dense 121, and Dense 169 – to detect the five classes of DR. Contrary to previous techniques, researchers employed up-sampling – a technique to apply augmentation only to minority classes – and down-sampling – a technique to remove extra instances of the majority classes. To augment the minority class, they copied patches of sizes 512×512 , applied flipping and rotation up to 90° . In addition, they used the publicly available Kaggle dataset. In Ref. [45], authors aimed to propose a deep learning architecture to classify smaller datasets, and, therefore, they used the MESSIDOR dataset to train and test their architecture. They simply modified the existing AlexNet architecture by deploying the

convolutional and max-pooling layers for the first eight layers and fully connected layers for the last three layers. Their model works well for smaller datasets and gives acceptable accuracy. In Ref. [46], authors implemented a few well-known existing CNN architectures, such as AlexNet, VGG, GoogleNet, and ResNet. To reduce the overfitting problem, they deployed hyper-parameter tuning. In addition to transfer learning and hyper-parameter tuning, they segmented the Kaggle dataset to extract the region of interest and then trained models on segmented datasets. Their study reveals that among the aforementioned architectures, VGG works better than others. Authors in Ref. [12] uses semantic segmentation to find the region-of-interest, especially the presence of microaneurysm, and then feed to their devised deep learning model to detect and classify early-stage DR. Researchers proposed a technique in Ref. [47] to automatically classify the Kaggle dataset into referable DR and non-referable DR – binary classifications – and, for that purpose, they used only 1000 images from the dataset. As a pre-processing step, authors applied various augmentation techniques, including resizing (to a size of $224 \times 224 \times 3$), rotation, zooming in, zooming out, rescaling, flipping, shearing, and translation. In addition, they proposed a novel CNN model, comprising eight convolutional layers, four max-pooling layers, and lastly two fully connected layers. Their model works satisfactorily as far as binary classification is concerned. Unlike the techniques discussed so far, authors of [2] used two datasets: Kaggle and Keio, to validate their proposed architecture. To address the issue of imbalanced datasets, they used an under-sampling strategy. They achieved 81.5% and 71.9% training and validation accuracies, respectively, for the Kaggle dataset, and 90.8% and 80% training and testing accuracies, respectively, for the Keio dataset that indicates overfitting issue with their approach. Though we have briefly and precisely discussed a few of the recent and prominent DR detection techniques, readers are encouraged to read [48] for comprehensive details of research in the field of DR detection. Authors in Ref. [48], discuss the existing DR detection and classification techniques, their advantages and disadvantages, and DR datasets available.

We have seen that numerous techniques have been proposed to address the DR detection problems, but none of them provide a comprehensible solution. In addition, medical data exist in small quantities and are highly imbalanced. Moreover, medical images, representing different stages of a disease, vary from one another very subtly, and to see those differences requires special expertise. For instance, all the images shown in Fig. 1, denoting different stages of DR, look almost alike. Therefore, in this paper, we have proposed a novel technique that enhances the subtle and invisible features and makes them visibly conspicuous; in addition, we have also proposed a novel hybrid deep neural network that classifies fundus images accurately with low computational requirements. The paper contributes in many aspects, but the major ones are highlighted as follows:

- Proposed a novel entropy enhancement technique to make less prominent features visibly more noticeable, as shown in Fig. 5.
- Proposed a novel hybrid neural network that is computationally efficient, provides better results for classification, requires comparatively less memory and computational time.
- Tested the proposed approach using multiple datasets.
- The proposed technique is specifically designed for small and imbalanced datasets; results corroborated this claim as well.

The rest of the paper is organized as follows: Section 2 is devoted to the discussion of the proposed methodology for entropy enhancement and deep learning model creation; in Section 3, we discuss the implementation, datasets used, and experimental results in detail; finally, Section 4 concludes the overall research article.

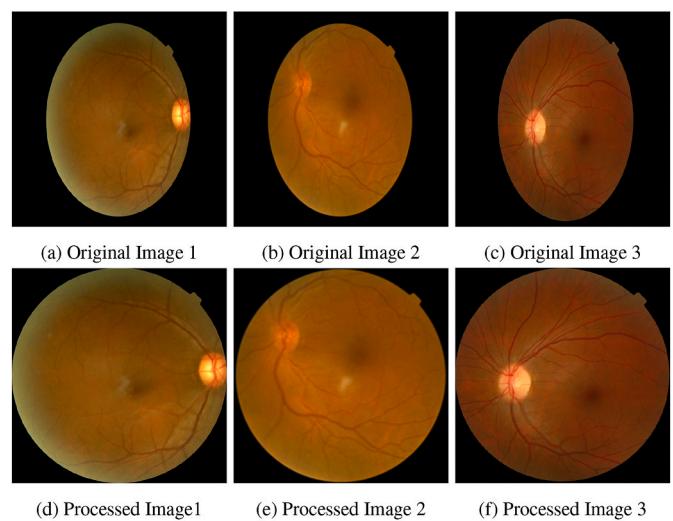


Fig. 2. Random images before and after unnecessary information removal.

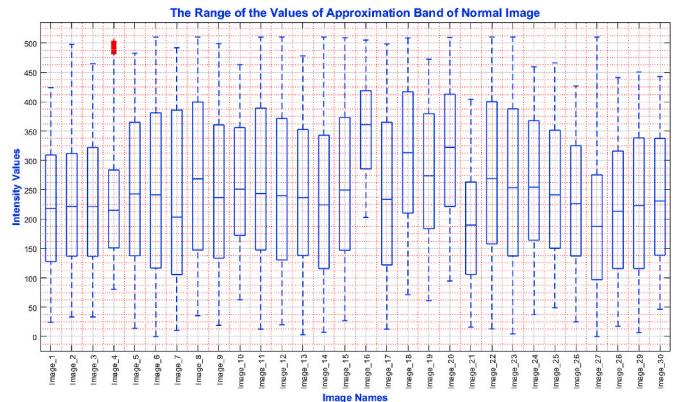


Fig. 3. Dynamic range of approximation band.

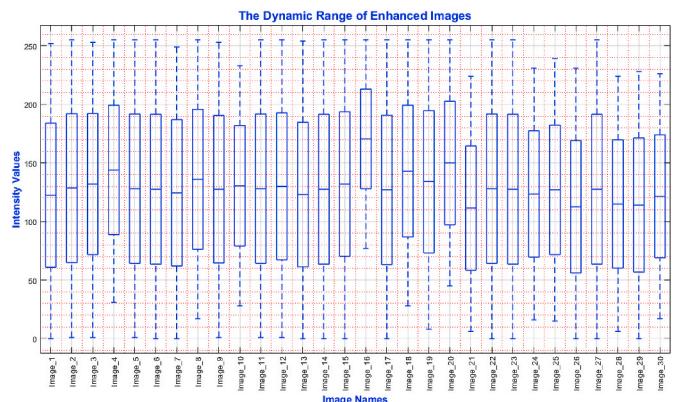


Fig. 4. Dynamic range of original images.

2. Unified technique for entropy enhancement based diabetic retinopathy detection using hybrid neural network

In this paper, we have proposed a technique to detect DR using small and highly imbalanced datasets. The proposed method can broadly be divided into three major steps:

1. Removal of Unnecessary Information
2. Increment of Images' Entropy
3. Designing of the Final Model for Training

In the subsequent subsections, we will discuss each of the aforementioned steps and their significance in detail.

2.1. Removal of Unnecessary Information

Different medical datasets contain different types of images; for instance, a few images contain only the region of interest – regions where the information to diagnose the disease is present –; whereas, others contain long borders. Neural networks learn to recognize images based on patterns; consequently, the same images with different borders sizes may look different to neural networks. Therefore, we, as a first step, remove all the unnecessary information in the form of unwanted borders using [Algorithm 1](#). As shown in [Fig. 2b, d, and 2f](#), our algorithm removes all the unnecessary information so that the neural network learns to detect based on the true information only.

Algorithm 1. Removal of Unnecessary Information

Algorithm 1: Removal of Unnecessary Information

```

Input:  $f(x, y, c) \in \mathbb{R}^{X \times Y \times C}$ : Image containing Unnecessary Information
Output:  $g(x, y, c) \in \mathbb{R}^{X \times Y \times C}$ : Image Without Unnecessary Information
1 [rows, cols, channels] = size(f)
2 for ind ← 1 to rows do
3   if  $\sum f(ind, :, :) > \tau$  then
4      $g(ind, :, :) = f(ind, :, :)$ 
      /* Rows having sum greater than a threshold ( $\tau$ )
       are retained and rest are discarded. */
5   end
6 end
7 for ind ← 1 to cols do
8   if  $\sum f(:, ind, :) > \tau$  then
9      $g(:, ind, :) = f(:, ind, :)$ 
      /* Columns having sum greater than a threshold
       ( $\tau$ ) are retained and rest are discarded. */
10 end
11 end

```

2.2. Entropy enhancement

Consider a comparatively darker image $I_{M \times N}$, with dynamic range $[0, L - 1]$, where L is the maximum value the image can represent. The discrete wavelet transform using Haar basis functions [\[49\]](#) is applied on the input image to get the approximation (\mathcal{A}), horizontal (\mathcal{H}), vertical (\mathcal{V}), and diagonal (\mathcal{D}) bands. The (\mathcal{H}), (\mathcal{V}), and (\mathcal{D}) contains textural and edge information. Keeping these bands intact and modifying the approximation (\mathcal{A}) band will result in higher entropy alternatively the modified image will contain more information and would perceptibly be more enhanced and vivid. The dynamic range of approximation band (\mathcal{A}) is not same as that of the original image, as shown in [Figs. 3 and 4](#). It can be seen that the dynamic range of the approximation band is wider than that of original images.

To find out the imperceptibility of the original image, we will calculate a term called Intensity-Exposure ($0 \leq \mathcal{E} \leq 1$) of the approximation band as follows:

$$\mathcal{E} = \frac{1}{\mathcal{K}} \frac{\sum_{i=k}^{\mathcal{K}-1} h(i) \times i}{\sum_{i=k}^{\mathcal{K}-1} h(i)} \quad (1)$$

where, $h(i)$ is the histogram of the approximation band, and k and \mathcal{K} represent the minimum and maximum values of the approximation band respectively, which are calculated as follows:

$$k = \min \min_{m=0}^{M/2} \mathcal{A}(m, n), \quad \mathcal{K} = \max \max_{m=0}^{M/2} \mathcal{A}(m, n).$$

The smaller exposure-value indicates bad imperceptibility; in other words, a darker image. Whereas, the exposure-value close to 1 indicates an overly saturated image. Exposure-values close to 0.5 are desired and indicate a well exposed image with good contrast and saturation. It is assumed that the input image is darker; therefore, the histogram for higher intensity values would be zero. To fill out those empty histogram values, a new histogram called optimal-histogram $h_o(i)$ is calculated as shown below.

$$h_o(i) = \begin{cases} h(i), & h(i) < \mathcal{T}, \\ \mathcal{T}, & h(i) \geq \mathcal{T}, \end{cases} \quad k \leq i \leq \mathcal{K}, \quad (2)$$

where, \mathcal{T} is the threshold representing an optimal value of hist counts, calculated as follows:

$$\mathcal{T} = \left\lceil \frac{1}{\mathcal{K}} \sum_{i=k}^{\mathcal{K}-1} h(i) \right\rceil. \quad (3)$$

$\lceil \cdot \rceil$ represent the “rounding to the nearest integer” operation. The threshold will set an upper bound on the histogram and thereby avoid over-enhancement. The optimal histogram is then divided into two parts: one for over-exposed region of the input image, denoted as h_{over} , and one for under-exposed region of the input image, represented as h_{under} . The normalized under and over exposed regions of the image are calculated using the exposure-intensity \mathcal{E} estimated in [eq. \(1\)](#). The under-exposed and over-exposed histograms are calculated using [\(4\)](#) and [\(5\)](#), respectively.

$$h_{under}(i) = \frac{h_o(i)}{\mathcal{U}}, \quad \text{for } k \leq i \leq \mathcal{X}, \quad (4)$$

$$h_{over}(i) = \frac{h_o(i)}{\mathcal{O}}, \quad \text{for } \mathcal{X} + 1 \leq i \leq \mathcal{K}. \quad (5)$$

Where, \mathcal{X} is a parameters that calculates the intensity limit using \mathcal{E} to divide the optimal-histogram into two parts. The value of \mathcal{X} is calculated as shown below,

$$\mathcal{X} = \lfloor \mathcal{K}(1 - \mathcal{E}) \rfloor ..$$

The terms \mathcal{U} and \mathcal{O} , used for normalization, are calculated as follows:

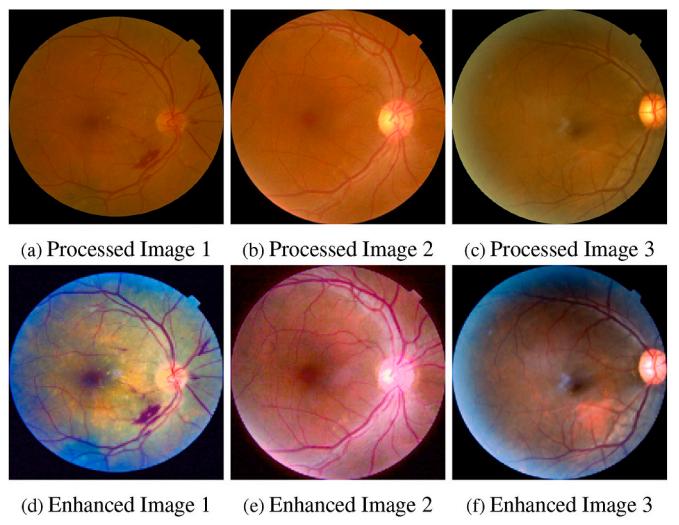


Fig. 5. Processed and corresponding enhanced images.

$$\mathcal{U} = \sum_{i=k}^{\mathcal{X}} h_o(i) \quad \text{and} \quad \mathcal{O} = \sum_{i=\mathcal{X}+1}^{\mathcal{K}} h_o(i).$$

Now, the Cumulative Distribution Function (CDF) of the under and over exposed histograms are estimated using (6) and (7), respectively.

$$\mathcal{C}_{under}(i) = \sum_{i=0}^{\mathcal{X}} h_{under}(i), \quad (6)$$

$$\mathcal{C}_{over}(i) = \sum_{i=\mathcal{X}+1}^{\mathcal{K}-1} h_{over}(i). \quad (7)$$

Finally, the transformation map is calculated as follows:

$$\mathcal{F} = \begin{cases} \mathcal{F}_{under}(i), & \text{for } k \leq i \leq \mathcal{X}, \\ \mathcal{F}_{over}(i), & \text{for } \mathcal{X} + 1 \leq i \leq \mathcal{K}, \end{cases} \quad (8)$$

where,

$$\begin{aligned} \mathcal{F}_{under}(i) &= \mathcal{X} \times \mathcal{C}_{under}(i) \quad \text{for } k \leq i \leq \mathcal{X}, \\ \mathcal{F}_{over}(i) &= (\mathcal{X}+1) + (\mathcal{K}-\mathcal{X}-1) \times \mathcal{C}_{over}(i) \quad \text{for } \mathcal{X} + 1 \leq i \leq \mathcal{K}. \end{aligned}$$

Once the transformation is obtained, apply it on the approximation band (\mathcal{A}) to get the enhanced approximation band ($\mathcal{A}_{enhanced}$) with comparatively large dynamic range. The last step is to apply the inverse discrete wavelet transform to get the imperceptibility better, more enhanced and more vivid image. It is evident from Fig. 5, which contains three original and their corresponding enhanced images, that the visual quality of the enhanced images is significantly improved. It can be seen that the entropy of enhanced images is more than that of original images; the processed images contain more textural information in terms of edges, wide dynamic range, and brightness, and, consequently, the subtle features are more prominent.

2.3. Hybrid neural network

The basic purpose of this paper is to provide a solution for the classification of medical data that is not only highly imbalanced but also contain a very limited number of images. For that, we have proposed an entropy enhancement technique in Section 2.2 to make subtle features conspicuous. In addition to that, in this paper, we have also proposed a novel deep learning model, called hybrid network, as shown in Fig. 6, consisting of two sub-networks; main and auxiliary networks. The

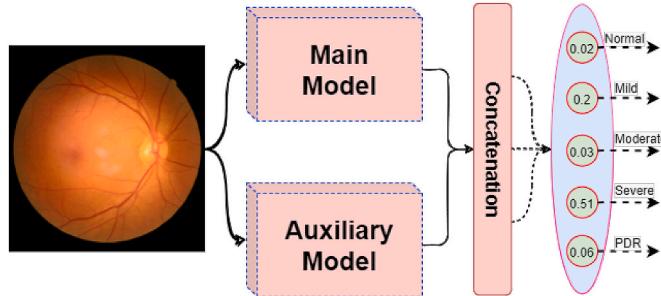


Fig. 6. The diagram of the proposed hybrid network.

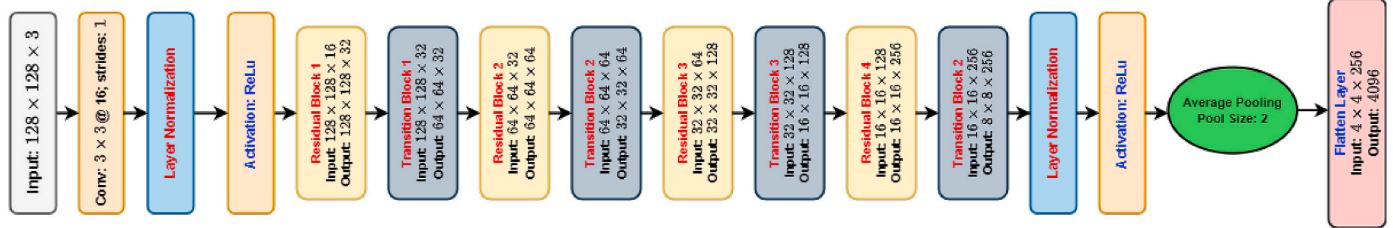


Fig. 7. The diagram of the main network.

results – Section 3 – shows that due to the entropy enhancement technique, our results outperform existing techniques for small and imbalanced datasets' classification, and our proposed model complements those results too. Moreover, we have increased the width of the network instead of the depth to enhance the performance of image classification task and for computational effectiveness, and results prove the worthiness of our approach.

The main network, primarily, consists of four residual and four transition blocks, in addition to convolution, normalization, activation, and max pooling layers, as shown in Fig. 7. The complete diagram of each of the residual blocks is shown in Fig. 8, and it can be seen that each of the residual blocks consist of three sub-residual networks. Each sub-residual network is made up of convolution, normalization, and activation layers. It should be noted that, in this paper, we have used the ReLU [50] as activation function, and layer normalization [51,52]. We have used layer normalization and preferred it over batch normalization to minimize the internal covariate shift, to prevent vanishing gradients problem, and to accelerate the convergence of training [53]. Results show that using layer normalization improves the performance as well. Each of the three sub-residual networks of the first, second, third, and fourth residual blocks uses (16, 16, 32), (32, 32, 64), (64, 64, 128), and (128, 128, 256) filters of sizes 3×3 . The skip connection used in the sub-residual networks of first, second, third, and fourth residual blocks uses 32, 64, 128, and 256 filters, and the dimension of each filter is 3×3 . Additionally, all the convolutional layers in all residual blocks use stride sizes of (1, 1).

The main model of the hybrid network, in addition to residual blocks, contains four transition networks. Each transition block consists of three convolution, normalization, and activation layers, as shown in Fig. 9. The first, second, third, and fourth transition blocks uses (16, 16, 32), (32, 32, 64), (64, 64, 128), and (128, 128, 256) filters of sizes 3×3 . The last convolution layer of each transition block and the skip connection uses a stride of size 2; the rest layers use strides of size 1. We have used strides of size 2 to reduce the computational complexity and receptive field.

The other part of the model, the auxiliary model, shown in Fig. 10, is used to complement the performance of the main model. The auxiliary model contains convolutional, max pooling, and flatten layers. For every convolutional layer, we have used the ReLU activation functions and strides of size 1. All the other required details are visible from Fig. 10.

In the end, the outputs of the main and the auxiliary models are concatenated and fed to the classification layer.

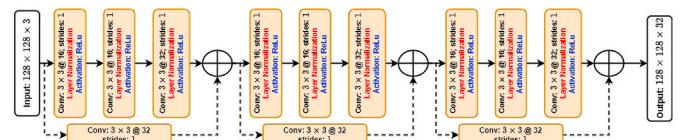


Fig. 8. The diagram of a generic residual network.

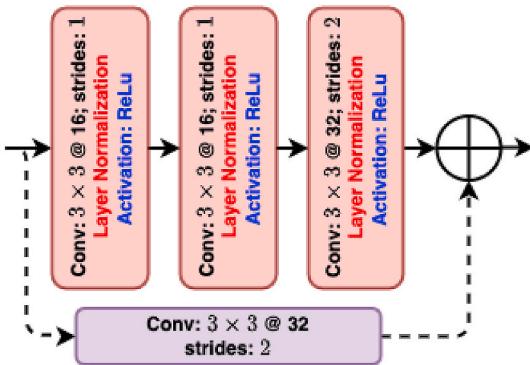


Fig. 9. The diagram of a generic transition network.

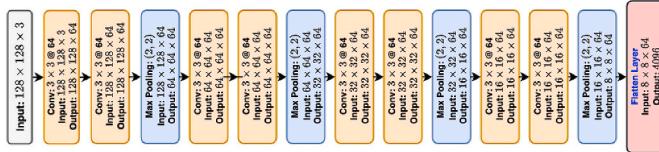


Fig. 10. The diagram of the auxiliary network.

3. Experiments and results

Medical images play crucial roles in the diagnosis of various diseases; however, interpreting them for diagnoses requires extreme expertise and knowledge. Moreover, medical images vary very subtly from one stage of a disease to another; for instance, to a layman, all the images shown in Fig. 1 look alike – even experts need to take a lot of care in the correct diagnosis. In addition, medical data mostly exist in small quantities and are highly imbalanced. Therefore, in this paper, we have proposed a novel technique that increases the entropy of medical images, and, thereby, makes the subtle features more prominent. In addition to that, we have also proposed a hybrid neural network for classification. To examine the effectiveness of our technique, we have chosen three datasets: Ultra Wide Filed (UWF), Asia Pacific Tele Ophthalmology Society (APOTOS), and MESSIDOR-2 – the details are mentioned in Section 3.1.

3.1. Datasets

The Ultra Wild Field (UWF) dataset, also known as the refuge dataset, contains 1600 images with five classes, and it is not publicly available. However, it can be accessed by requesting the owners [5]. We have reserved 80%, 10%, and 10%, respectively, for training, validation, and testing; the complete distribution is shown in Table 2.

The Asia Pacific Tele-Ophthalmology Society (APOTOS) dataset is publicly available on Kaggle [54], containing approximately 3662 images with CSV labeled file. Similarly, we distributed the dataset 80%, 10%, and 10% for training, validation, and testing, respectively, as shown in Table 3.

MESSIDOR-2 [55, 56]] is the collection of DR color fundus images used for DR detection and classifications. To collect the datasets, they examined 874 people and produced 1748 color fundus images with different image resolutions. The images have been categorized into five classes: normal, mild, moderate, severe, and PDR as shown in Table 4. The distribution of the MESSIDOR dataset is done the same way we did for the other two aforementioned datasets, as shown in Table 4.

Table 2

The distribution of UWF dataset for training, testing, and validation.

Classes	Total Images	Reserved For		
		Training	Validation	Testing
Normal	480(30%)	384	48	48
Mild	320(20%)	256	32	32
Moderate	320(20%)	256	32	32
Severe	320(20%)	256	32	32
PDR	160(10%)	128	16	16

Table 3

The distribution of APOTOS dataset for training, testing, and validation.

Classes	Total Images	Reserved For		
		Training	Validation	Testing
Normal	1805(49%)	1443	181	181
Mild	370(10%)	296	37	37
Moderate	999(27%)	799	100	100
Severe	193(5%)	153	20	20
PDR	295(8%)	235	30	30

Table 4

The distribution of MESSIDOR-2 for training, testing, and validation.

Classes	Total Images	Reserved For		
		Training	Validation	Testing
Normal	1020(58%)	816	102	102
Mild	264(15%)	212	26	26
Moderate	347(20%)	277	35	35
Severe	77(4%)	61	8	8
PDR	36(2%)	28	4	4

3.2. Performance measures

To assess the effectiveness of the entropy enhancement approach for the classification of DR images, we have proposed a hybrid neural network, as discussed in Section 2.3. We used Adam optimizer [57] with learning rate of 0.001. We kept decreasing the learning rate by a factor of 0.9 if the validation specificity does not improve for consecutive five epochs until it reaches its lower limit, which is $1.1e^{-5}$. Since medical images are taken under a controlled environment and are not very different as far as orientation is considered; therefore, for augmentation, we used horizontal and vertical flips only. Keeping the same settings, we first trained our network using the original images, and then we trained the same network afresh using the enhanced dataset.

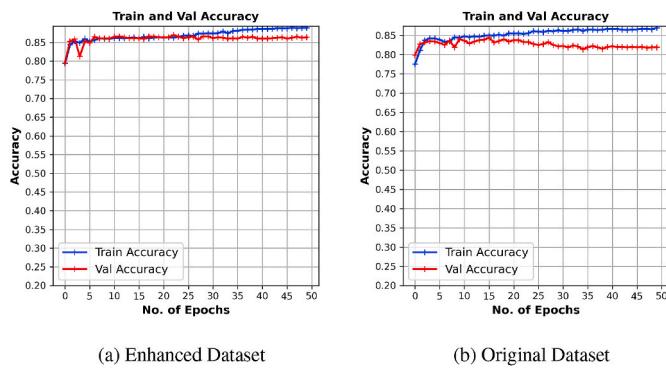
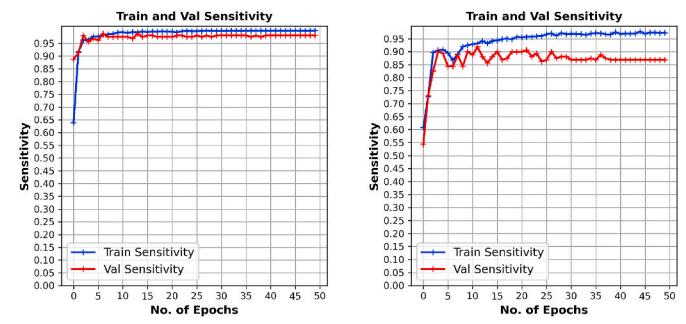
3.2.1. Classification accuracy

First metric that we used for comparison and performance measure is accuracy. Classification accuracy [2, 5, 16, 45], defined in Eq. (9), is the ratio of correct number of prediction to the total number of predictions.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions Made}}, \quad (9)$$

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

In the above equation, TP, FP, TN, and FN, represent, respectively, true positive, false positive, true negative, and false negative. It is evident from Fig. 11 that using the enhanced dataset – which is obtained using our entropy enhancement algorithm, mentioned in Section 2.2 – not only reduces the high variance but improves the training and validation accuracy significantly.

**Fig. 11.** Accuracies of original and enhanced datasets.**Fig. 13.** Sensitivities of original and enhanced datasets.

3.2.2. Specificity

The second metric, called **specificity** [2,5,16,45], examines a model's ability to predict negative classes effectively. Specificity, calculated using Eq. (10), has values between 0 and 1, and its value close to one is desired and represents better a model.

$$\text{Specificity} = \frac{TN}{TN + FN} \quad (10)$$

In terms of specificity – also called True Negative Rate – the results, shown in Fig. 12a, obtained using the enhanced dataset are a lot better than that of the original dataset, which means that the proposed enhancement technique makes the subtle features conspicuous and make it easier for a deep learning model to classify effectively.

3.2.3. Sensitivity

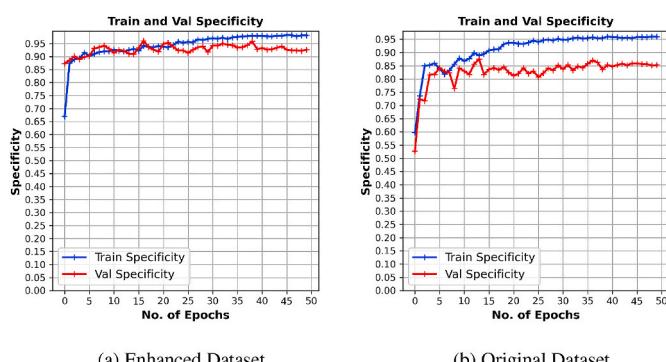
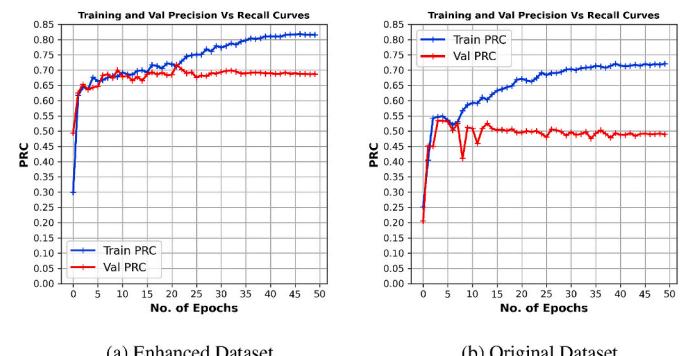
Another metric, called **sensitivity** [2,5,16,45], measures the capability of a deep learning model to classify true positive classes. In medical images, it is of utmost importance that a model correctly classify positive classes successfully. Sensitivity, calculated using Eq. (11), ranges from 0 to 1, and higher the sensitivity, better the model.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (11)$$

It is visible from Fig. 13 that for sensitivity – also known as True Positive Rate – enhanced dataset outperforms the original dataset.

3.2.4. Precision-recall curves

Precision and Recall Curves (PRC) is another useful metric to measure the quality of a classification model [2,5,45], especially for imbalanced dataset. PRC is a graph of precision, defined in Eq. (12), versus recall, defined in Eq. (13), as shown in Fig. 14.

**Fig. 12.** Specificities of original and enhanced datasets.**Fig. 14.** Precision and recall curves of original and enhanced datasets.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

Precision, has values between 0 and 1, measure how successfully and correctly classify true positive cases out of all positive predictions a model makes.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

Recall, same as sensitivity, measures the ability of a model to classify true positive classes effectively. It can be seen from Fig. 14 that the model using the enhanced dataset can give better PRC than that of using the original dataset that shows the success of the entropy enhancement technique.

3.2.5. Area under the curve

The last metric is the Area Under the Curve (AUC), a synopsis of the Receiver Characteristic Operator (ROC), which measures the ability of a deep learning model to distinguish between different classes [16]. It is a graph of sensitivity versus specificity, and its values lie in the range of 0 and 1. A model with $AUC = 1$ is desirable and able to distinguish among different classes with a hundred percent accuracy; on the other hand, a model with $AUC = 0$ is unable to differentiate different classes and is not worthwhile. In practical scenarios, AUC close to one is achievable and acceptable.

The graph, shown in Fig. 15, shows the AUC for original and enhanced datasets, and it is clearly visible that the enhanced dataset achieves better AUC compared to the original dataset.

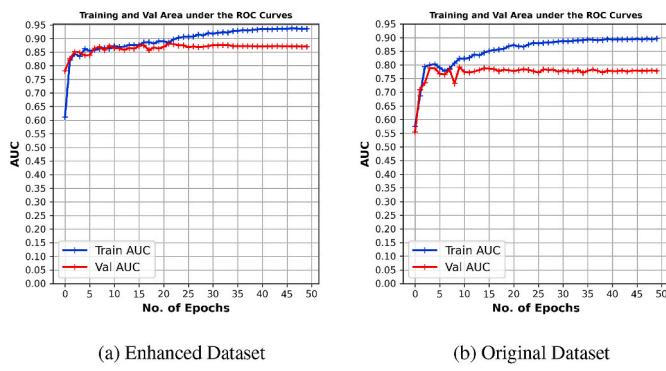


Fig. 15. Area under the RoC curves of original and enhanced datasets.

3.3. Comparison with contemporary techniques

In the aforementioned sections, we showed improved results when using enhanced datasets in comparison to when we used the original dataset. The basic purpose of this paper is to prove that using the enhancement technique proposed in Section 2.2, can improve the performance of a deep learning model, reduce the bias and variance, especially for small and imbalanced datasets. However, we have proposed a novel hybrid neural network; therefore, we have also compared our results with other state-of-the-art schemes for DR detection. To compare the performance, we have used three datasets: UWF, APTOS, and MESSIDOR-2.

It is clear from Table 5, Table 6, and Table 7 that the proposed scheme outperforms existing technique in terms of majority metrics used to measure a classification model.

Both UWF and MESSIDOR-2 are extremely small – as we know that deep learning models need a huge amount of data for satisfactory performance – and highly imbalanced. Even though, the proposed technique – due to the novel entropy enhancement method and the novel proposed hybrid neural network – achieved reasonable results. However, for APTOS, which is a relatively larger dataset, our technique exceptionally performed well, which indicates that for larger datasets our model along with the entropy enhancement technique, will give more promising results.

Table 5
Comparison using UWF dataset.

Metrics	Results for UWF Dataset				
	Proposed Technique	Techniques Presented In			
		[2]	[5]	[16]	[45]
Accuracy	0.84	0.81	0.87	0.83	0.76
Specificity	0.92	0.88	0.89	0.87	0.81
Sensitivity	0.97	0.81	0.93	0.88	0.80
Precision	0.67	0.64	0.65	0.61	0.52
PRC	0.63	0.55	0.61	0.62	0.50
AUC	0.84	0.78	0.82	0.80	0.75

Table 6
Comparison using APTOS dataset.

Metrics	Results for APTOS Dataset				
	Proposed Technique	Techniques Presented In			
		[2]	[5]	[16]	[45]
Accuracy	0.92	0.86	0.91	0.89	0.80
Specificity	0.99	0.92	0.93	0.87	0.84
Sensitivity	0.99	0.80	0.90	0.88	0.78
Precision	0.85	0.79	0.78	0.76	0.63
PRC	0.87	0.73	0.80	0.84	0.66
AUC	0.95	0.85	0.88	0.89	0.82

Table 7
Comparison using MESSIDOR-2 dataset.

Metrics	Results for MESSIDOR-2 Dataset			
	Proposed Technique	Techniques Presented In		
		[2]	[5]	[16]
Accuracy	0.80	0.76	0.83	0.78
Specificity	0.90	0.85	0.83	0.83
Sensitivity	0.89	0.73	0.82	0.79
Precision	0.85	0.80	0.79	0.76
PRC	0.61	0.53	0.57	0.59
AUC	0.82	0.75	0.78	0.77

4. Conclusion

In this research article, we have addressed two issues: the subtle nature of features in medical images and imbalance datasets. To do so, first, we enhanced the imperceptible features using our proposed novel entropy enhancement technique. Later, we devised a computationally efficient hybrid neural network that classifies DR images efficiently, also evident from the results. In the future, we intend to extend our approach to solve the other medical classification datasets.

Declaration of competing interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgment

This research is conducted at the Control Automotive and Robotics Lab (CARL), BUITEMS, Quetta, Pakistan; funded by the National Center of Robotics and Automation (NCRA) with the collaboration of Higher Education Commission (HEC) of Pakistan.

References

- Kyong Hye Joung1, Sang Hyun Ju1, Ji Min Kim1, Sorim Choung, Jae Min Lee, Kang Seo Park, Hyun Jin Kim1, Bon Jeong Ku1, "Clinical Implications of Using Post-Challenge Plasma Glucose Levels for Early Diagnosis of Type 2 Diabetes Mellitus in Older Individuals," in Diabetes & Metabolism Journal, vol. 42, no. 2, pp. 147 – 154. <https://doi.org/10.4093/dmj.2018.42.2.147>.
- K. Shankar, Y. Zhang, Y. Liu, L. Wu, C.-H. Chen, Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification, in: IEEE Access 8, 2020, pp. 118164–118173, <https://doi.org/10.1109/ACCESS.2020.3005152>.
- K. Shankar, E. Perumal, R.M. Vidhyavathi, Deep neural network with moth search optimization algorithm based detection and classification of diabetic retinopathy images, Springer Appl. Sci. 2 (4) (2020) 1–10, <https://doi.org/10.1007/s42452-020-2568-8>.
- Overview diabetic retinopathy [Online]. Available: <https://www.nhs.uk/conditions/s/diabetic-retinopathy/>, October 30, 2018. (Accessed 18 August 2021).
- K. Oh, H.M. Kang, D. Leem, et al., Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images, Sci. Rep. 11 (2021) 1897, <https://doi.org/10.1038/s41598-021-81539-3>.
- X. Ai, P. Yu, Y. Hou, X. Song, J. Luo, N. Li, X. Lai, X. Wang, X. Meng, A review of traditional Chinese medicine on treatment of diabetic retinopathy and involved mechanisms, Biomed. Pharmacother. 132 (2020) 1–17, <https://doi.org/10.1016/j.biopha.2020.110852>.
- E. Singh, R.E. Redgrave, H.M. Phillips, H.M. Arthur, Arterial endoglin does not protect against arteriovenous malformations, Angiogenesis 23 (2020) 559–566.
- M. Park, J.S. Jin, S. Luo, Locating the optic disc in retinal images, in: International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06), IEEE, 2006, July, pp. 141–145.
- H.H. Ma, R. Liutkevičienė, Age-related macular degeneration: what do we know so far? Acta Med. Litu. 28 (1) (2021), 7–7.

- [10] R. Jerath, S.M. Cearley, V.A. Barnes, E. Nixon-Shapiro, How lateral inhibition and fast retinogeniculo-cortical oscillations create vision: a new hypothesis, *Med. Hypotheses* 96 (2016) 20–29.
- [11] M. Ahmad, N. Kasukurthi, H. Pande, Deep learning for weak supervision of diabetic retinopathy abnormalities, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, April, pp. 573–577.
- [12] L. Qiao, Y. Zhu, H. Zhou, Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms, *IEEE Access* 8 (2020) 104292–104302.
- [13] M. Ahmad, N. Kasukurthi, H. Pande, Deep learning for weak supervision of diabetic retinopathy abnormalities, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 573–577, <https://doi.org/10.1109/ISBI.2019.8759417>.
- [14] Y. Katada, N. Ozawa, K. Masayoshi, Y. Ofuji, K. Tsubota, T. Kurihara, Automatic screening for diabetic retinopathy in interracial fundus images using artificial intelligence, *Intell. Based Med.* 3 – 4 (2020) 1–8, <https://doi.org/10.1016/j.ibmed.2020.100024>.
- [15] S. Qummar, F.G. Khan, S. Shah, A. Khan, S. Shamshirband, Z.U. Rehman, W. Jadoon, A deep learning ensemble approach for diabetic retinopathy detection, *IEEE Access* 7 (2019) 150530–150539.
- [16] N. Asiri, M. Hussain, F.A. Adel, N. Alzaidi, Deep learning based computer-aided diagnosis systems for diabetic retinopathy: a survey, *Artif. Intell. Med.* 99 (2019) 1–10, <https://doi.org/10.1016/j.artmed.2019.07.009>.
- [17] A. Krizhevsky, I. Sutskever, G.E. Hinton, G. Geoffrey, ImageNet classification with deep convolutional neural networks, in: Conference on Neural Information Processing Systems, 2012, pp. 1106–1114, in: <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>.
- [18] M. Raza, M. Awais, W. Ellahi, N. Aslam, H.X. Nguyen, H. Le-Minh, Diagnosis and monitoring of Alzheimer's patients using classical and deep learning techniques, *Expert Syst. Appl.* 136 (2019) 353–364, <https://doi.org/10.1016/j.eswa.2019.06.038>.
- [19] A. Ebrahimpaghavieh, S. Luo, R. Chiong, Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review, *Comput. Methods Progr. Biomed.* 187 (2020), <https://doi.org/10.1016/jcmpb.2019.105242>. Article No. 105242.
- [20] J. Venugopalan, L. Tong, H.R. Hassanzadeh, M.D. Wang, Multimodal deep learning models for early detection of Alzheimer's disease stage, *Sci. Rep.* 11 (2021), <https://doi.org/10.1038/s41598-020-74399-w>. Article No. 3254.
- [21] G. Sannino, G. De Pietro, A deep learning approach for ECG-based heartbeat classification for arrhythmia detection, *Future Generat. Comput. Syst.* 86 (2018) 446–455, <https://doi.org/10.1016/j.future.2018.03.057>.
- [22] M. Hammad, A.M. Ilyasut, A. Subasi, E.S.L. Ho, A.A.A. El-Latif, A multitier deep learning model for arrhythmia detection, in: *IEEE Trans. Instrum. Meas.* 70, 2021, pp. 1–9, <https://doi.org/10.1109/TIM.2020.3033072>.
- [23] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, *Nat. Med.* 25 (2019) 65–69, <https://doi.org/10.1038/s41591-018-0268-3>.
- [24] A.S. Heinsfeld, A.R. Franco, R.C. Craddock, A. Buchweitz, F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the ABIDE dataset, *Neuroimage: Clin.* 17 (2018) 16–23, <https://doi.org/10.1016/j.nic.2017.08.017>.
- [25] J. Zhou, C.Y. Park, C.L. Theesfeld, A.K. Wong, Y. Yuan, C. Scheckel, J.J. Fak, J. Funk, K. Yao, Y. Tajima, A. Packer, R.B. Darnell, O.G. Troyanskaya, Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk, *Nat. Genet.* 51 (2019) 973–980, <https://doi.org/10.1038/s41588-019-0420-0>.
- [26] F. Ke, S. Choi, Y.H. Kang, K.-A. Cheon, S.W. Lee, Exploring the structural and strategic bases of autism spectrum disorders with deep learning, in: *IEEE Access* 8, 2020, pp. 153341–153352, <https://doi.org/10.1109/ACCESS.2020.3016734>.
- [27] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, S. Li, Breast cancer multi-classification from histopathological images with structured deep learning model, *Sci. Rep.* 7 (2017), <https://doi.org/10.1038/s41598-017-04075-z>. Article No. 4172.
- [28] L. Shen, L.R. Margolies, J.H. Rothstein, E. Fluder, R. McBride, W. Sieh, Deep learning to improve breast cancer detection on screening mammography, *Sci. Rep.* 9 (2019), <https://doi.org/10.1038/s41598-019-48995-4>. Article No. 12495.
- [29] Y. Yari, T.V. Nguyen, H.T. Nguyen, Deep learning applied for histological diagnosis of breast cancer, in: *IEEE Access* 8, 2020, pp. 162432–162448, <https://doi.org/10.1109/ACCESS.2020.3021557>.
- [30] M. Aubreville, C. Knipfer, N. Oetter, C. Jaremenko, E. Rodner, J. Denzler, C. Bohr, H. Neumann, F. Stelzle, A. Maier, Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning, *Sci. Rep.* 7 (2017), <https://doi.org/10.1038/s41598-017-12320-8>. Article No. 11979.
- [31] P. Singh, G.V. black dental caries classification and preparation technique using optimal CNN-LSTM classifier, *Multimed. Tool. Appl.* 80 (2021) 5255–5272, <https://doi.org/10.1007/s11042-020-09891-6>.
- [32] K.P. Smitha, A.D. Kanga, J.E. Kirby, Automated interpretation of blood culture gram stains by use of a deep convolutional neural network, *J. Clin. Microbiol.* 56 (3) (2018) 1–12, <https://doi.org/10.1128/JCM.01521-17>.
- [33] Y. Zhang, H. Jiang, T. Ye, M. Juhas, Deep learning for imaging and detection of microorganisms, *Trends Microbiol.* 29 (7) (2021) 569–572, <https://doi.org/10.1016/j.tim.2021.01.006>.
- [34] K.P. Smitha, H. Wang, T.J.S. Durant, B.A. Mathison, S.E. Sharp, J.E. Kirby, S. W. Long, D.D. Rhoads, Applications of artificial intelligence in clinical microbiology diagnostic testing, *Clin. Microbiol. Newsl.* 42 (8) (2020) 61–70, <https://doi.org/10.1016/j.clinmicnews.2020.03.006>.
- [35] D. Ardila, A.P. Kiraly, S. Bharadwaj, Bo Choi, J.J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D.P. Naidich, S. Shetty, End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, *Nat. Med.* 25 (2019) 954–961, <https://doi.org/10.1038/s41591-019-0447-x>.
- [36] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyo, A. L. Moreira, N. Razavian, A. Tsirigos, Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning, *Nat. Med.* 24 (2018) 1559–1567, <https://doi.org/10.1038/s41591-018-0177-5>.
- [37] S.K. Lakshmanaprabu, S.N. Mohanty, K. Shankar, N. Arunkumar, G. Ramirez, Optimal deep learning model for classification of lung cancer on CT images, *Future Generat. Comput. Syst.* 92 (2019) 374–382, <https://doi.org/10.1016/j.future.2018.10.009>.
- [38] S.S. Han, G.H. Park, W. Lim, M.S. Kim, J.I. Na, I. Park, S.E. Chang, Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network, *PLoS One* 13 (1) (2018) 1–14, <https://doi.org/10.1371/journal.pone.0191493>.
- [39] Y.J. Kim, S.S. Han, H.J. Yang, S.E. Chang, Prospective, comparative evaluation of a deep neural network and dermoscopy in the diagnosis of onychomycosis, *PLoS One* 15 (6) (2020) 1–9, <https://doi.org/10.1371/journal.pone.0234334>.
- [40] T.Y. Tan, L. Zhang, C.P. Lim, Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models, *Appl. Soft Comput.* 84 (2019), <https://doi.org/10.1016/j.asoc.2019.105725>. Article No. 105725.
- [41] A. Dascalu, E.O. David, Skin cancer detection by deep learning and sound analysis algorithms: a prospective clinical study of an elementary dermatoscope, *EBioMedicine* 43 (2019) 107–113, <https://doi.org/10.1016/j.ebiom.2019.04.055>.
- [42] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (2017) 115–118, <https://doi.org/10.1038/nature21056>.
- [43] T. Shanthi, R.S. Sabreenian, Modified Alexnet architecture for classification of diabetic retinopathy images, *Comput. Electr. Eng.* 76 (2019) 56–64, <https://doi.org/10.1016/j.compeleceng.2019.03.004>.
- [44] S. Wan, Y. Liang, Y. Zhang, Deep convolutional neural networks for diabetic retinopathy detection by image classification, *Comput. Electr. Eng.* 72 (2018) 274–282.
- [45] K. Xu, D. Feng, H. Mi, Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image, *Molecules* 22 (12) (2017) 2054.
- [46] W.L. Alyoubi, W.M. Shalash, M.F. Abulkhair, Diabetic retinopathy detection through deep learning techniques: a review, *Inform. Med. Unlocked* 20 (2020), 100377.
- [47] Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, third ed., Prentice-Hall, Inc., USA, 2006.
- [48] Konstantin Eckle, Johannes Schmidt-Hieber, A comparison of deep networks with ReLU activation function and linear spline-type methods, *Neural Netw.* 110 (2019) 232–242, <https://doi.org/10.1016/j.neunet.2018.11.005>.
- [49] Jimmy Lei Ba, Jamie Ryan Kiros, E. Hinton Geoffrey, Layer Normalization, 2016 arXiv preprint arXiv:1607.06450.
- [50] Jingjing Xu, Sun1 Xu, Zhiyuan Zhang, Guangxiang Zhao, Junyang Lin, Understanding and improving layer normalization, in: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, 2019, pp. 1–9.
- [51] Chunwei Tian, Yong Xu, Wangmeng Zuo, Image denoising using deep CNN with batch renormalization, *Neural Netw.* 121 (2020) 461–473, <https://doi.org/10.1016/j.neunet.2019.08.022>.
- [52] Available on, <https://www.kaggle.com/c/aptos2019-blindness-detection>.
- [53] Decencière, et al., Feedback on a publicly distributed database: the Messidor database, *Image Anal. Stereol.* 33 (3) (aug. 2014) 231–234, <https://doi.org/10.5566/ias.1155>. ISSN 1854-5165.
- [54] Diederik P. Kingma, Jimmy Ba, Adam: A Method for Stochastic Optimization, 2014 arXiv preprint arXiv:1412.6980.