Neural Computing and Applications
https://doi.org/10.1007/s00521-018-03974-0

INTELLIGENT BIOMEDICAL DATA ANALYSIS AND PROCESSING

An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network

D. Jude Hemanth[1] · Omer Deperlioglu[2] · Utku Kose[3]

Received: 25 September 2018 / Accepted: 20 December 2018

Abstract
The objective of this study is to propose an alternative, hybrid solution method for diagnosing diabetic retinopathy from retinal fundus images. In detail, the hybrid method is based on using both image processing and deep learning for improved results. In medical image processing, reliable diabetic retinopathy detection from digital fundus images is known as an open problem and needs alternative solutions to be developed. In this context, manual interpretation of retinal fundus images requires the magnitude of work, expertise, and over-processing time. So, doctors need support from imaging and computer vision systems and the next step is widely associated with use of intelligent diagnosis systems. The solution method proposed in this study includes employment of image processing with histogram equalization, and the contrast limited adaptive histogram equalization techniques. Next, the diagnosis is performed by the classification of a convolutional neural network. The method was validated using 400 retinal fundus images within the MESSIDOR database, and average values for different performance evaluation parameters were obtained as accuracy 97%, sensitivity (recall) 94%, specificity 98%, precision 94%, FScore 94%, and GMean 95%. In addition to those results, a general comparison of with some previously carried out studies has also shown that the introduced method is efficient and successful enough at diagnosing diabetic retinopathy from retinal fundus images. By employing the related image processing techniques and deep learning for diagnosing diabetic retinopathy, the proposed method and the research results are valuable contributions to the associated literature.

Keywords Diabetic retinopathy · Image processing · Deep learning · Convolutional neural network

1 Introduction

General guidelines for diagnosis and classification of diabetes have been published by the World Health Organization (WHO) for more than 50 years. The final version prepared as a guide for the identification, and classification of diabetes and its complications was published in 1998. The WHO and the technical advisory group from International Diabetes Federation (IDF) established a meeting in Geneva in order to review and update the existing WHO guidelines [1]. The IDF indicated in 2013 that there were around 385 million people worldwide with diabetes, and most of them are type-2 diabetes. It is estimated that the related amount will reach 592 million by 2035 [2]. Economic, social, and medical costs of diabetes are large public health problems, and diabetes is known as the fourth cause for deaths worldwide [3–5]. It is also too remarkable that disruptive macro- and micro-cardiovascular complications regarding the diabetes cause significant reduces in life quality and the life expectancy [5–8]. Because of that, definition and diagnosis–classification of diabetes are important research problems and many scientists have been working widely to find alternative solutions for them.

An important complication of the diabetes is called as diabetic retinopathy, and it causes visual impairment in long-term vision. Diabetic retinopathy is briefly a serious eye disease caused by diabetes mellitus, and that disease is

✉ Utku Kose
utkukose@sdu.edu.tr

[1] Department of ECE, Karunya Institute of Technology and Sciences, Coimbatore, India

[2] Department of Computer Technologies, Afyon Kocatepe University, Afyonkarahisar, Turkey

[3] Department of Computer Engineering, Suleyman Demirel University, Isparta, Turkey

Published online: 01 January 2019

a common cause of blindness, especially within developed countries [9–11]. At this point, early diagnosis and treatment are crucial to prevent patients from being affected by the blindness condition, or at least to slow down the progression of diabetic retinopathy toward that. Because of that, mass screening of diabetic patients is highly desirable. However, manual grading always gives the right results because it requires a good experience and expertise. In time, much effort has been put into establishing reliable computerized scanning systems based on color fundus images [12, 13]. Currently, diabetic retinopathy detection from digital fundus images is still an open problem and it is needed to develop alternative solutions because of that. In addition to imaging and computer vision-based systems, the next step is the employment of intelligent diagnosis systems.

Considering background of intelligent systems, it is clearly seen that artificial intelligence-based approaches, and especially machine learning, are widely used in almost all fields of modern life in order to reduce the effort or expenditure and achieve better results at the same time. Mathematical nature of the world has made it possible to adapt intelligent systems in different problems [14–23]. In the associated literature, it is possible to see applications of intelligence systems in the field of also medical [24–28] and the problem of disease diagnosis is a remarkable research interest in this manner [29–34]. As associated with the topic of this study, diagnosis of diabetes over computer-assisted intelligent system is performed for reducing the human effort required to provide basic diagnosis and discrimination between diabetic and healthy patients. Doctors generally look for some specific signs to diagnose diabetes, and a similar approach can be used by employing parameters in different machine learning techniques. At this point, a machine learning-oriented system can be trained to determine whether a person is diabetic or healthy based on this indication [35]. There can be different types of applications for that purpose, and the current trend is using deep learning. Furthermore, it is also an important approach to employ image processing for achieving diagnosis over medical image data [36–41].

The objective of this study is to provide an alternative, hybrid solution method for diagnosing diabetic retinopathy from retinal fundus images. By employing both image processing and deep learning, one remarkable motivation of the study is obtaining improved results of diagnosis and contributing to the associated literature with a hybrid solution for diagnosis of a critical disease. Additionally, another motivation is to demonstrate that classification can be done easily with deep learning assisted by classical and simple image processing techniques, which are different from the ones used previously in the literature. In this sense, the image processing includes both histogram equalization (HE) and the contrast limited adaptive histogram equalization (CLAHE) in order to fit the retinal fundus images for better classification, which leads the system to diagnosis. After the related image processing stages, the classification is done by a convolutional neural network (CNN). The hybrid solution model is easy to design for obtaining an alternative solution way for diagnosing diabetic retinopathy.

Considering the general topic of the study, some remarkable studies from the literature can be explained briefly as follows: Seoud et al. [42] introduced an automatic grading system for diabetic retinopathy using image processing and machine learning. In the related study, they first obtained 35 features as combining size and probability information for classification. It was achieved by detecting a red lesion to create a lesion probability map. They classified these features using a random forest and obtained a classification accuracy of 74.1%. In another study by Savarkar et al. [43], a three-stage system for early diagnosis of micro-aneurysms (MAs) using filter banks was employed. In the first stage of that system, all possible candidate regions were extracted for the MAs in the retinal image. These regions were then classified thanks to a hybrid classifier system formed by Gaussian mixture model (GMM) and also support vector machine (SVM). It was stated that the method allows recognition of circular or slightly elongated image structures and can distinguish vessel bifurcations and transitions [43]. Another approach for diagnosing diabetic retinopathy from retinal fundus images was based on an automated solution aiming to classify diabetic retinopathy with speeded up robust features [13]. In this study, DRIDB0, DRIDB1, HRF, STARE, and MESSIDOR databases were used and accuracies of 94.4%, sensitivity (recall) of 94%, precision of 94%, and FScore of 94% were obtained. Additionally, Sreejini and Govindan introduced a method containing image preprocessing, and also a hybrid system of particle swarm optimization (PSO) and fuzzy C-means (FCM) for optic disk elimination, exudates segmentation, macular region localization, fovea, and finally classification [44]. In another diagnostic method using image preprocessing and classification, Safitri and Juniati used fractal analysis and components of K-nearest neighbor (KNN) [45]. The accuracy of classification in this study was 98.17%. Akrametal et al. [46] employed an approach on identifying and classifying micro-aneurysms for realizing early detection of diabetic retinopathy. Acharya et al. [47] employed the texture features and SVM classifier for automatic mass screenings regarding diabetic retinopathy. As normal retina, macular edema, proliferative diabetic retinopathy, and non-proliferative diabetic retinopathy, a total of four classes were identified. In the analysis works, they focused on a total of 238 images of retinal fundus and five texture features such

as correlation, homogeneity, long run emphasis, short run emphasis, and run percentage were obtained from the target fundus images. After that stage, the related features were then fed into a SVM for achieving an automatic classification [47]. At this point, several methods were also employed in the literature to determine pathological and normal retinal images. One of them is an automated solution for classifying diabetic macular edema. In this method, the fovea is localized and the grading scale of Early Treatment Diabetic Retinopathy Studies (ETDRS) was employed for marking the regions of macula [48]. In the related study, an extraction method using marker-controlled watershed transformation was adopted and modified from previously done research. In order to classify diabetic macular edema into normal, stage 1, and stage 2 diabetic macular edema, the location of the extracted exudates on the marked macular regions was computed accordingly [48]. An alternative method in this manner has been the employment of multi-scale amplitude modulation–frequency modulation methods [49]. Antal and Hajdu have introduced a community-based framework to facilitate diagnosis of micro-aneurysm. They tried to employ a combination of internal components regarding micro-aneurysm detectors, i.e., pre-treatment methods and candidate extractors, instead of the most commonly used multiple classifiers. They used 180 images within the MESSIDOR database to evaluate the introduced method and obtained a value of 90% related to area under the ROC curve (AUC), as a classification performance criterion [12].

As it can be understood, results on diagnosis performance may vary according to different preprocessing phases and structures of the employed solution methods. As associated with one of the mentioned motivations, it is aimed to improve diagnosis performance in general. Here, it is believed that the deep learning, which is an effective, advanced form of machine learning, will give the advantage of analyzing image data in detail for better results. Also called as deep learning network, typical network structure considered in this study, has a large number of abstract layers communicating with each other. Each layer is deeply connected to the previous layer and makes its decisions based on the output fed by the previous layer for dealing better with non-structured or unlabeled data [50, 51]. As different from the mentioned studies from the literature, the proposed method employs deep learning and uses its more advanced characteristics rather than using some already used techniques such as SVM, K-NN, FCM, or optimization techniques. By considering also use of image data, CNN is also a role-based choice to deal with diabetic retinopathy image data better. In addition to novelty and advantage, the method also employs an easy-to-design structure with even simple image processing techniques in preprocessing stage. Although the related studies try to get good results by combining and adapting several techniques, the proposed method in this study employs HE and CLAHE, which are effective techniques to preprocess data and make it ready to the CNN. Mixed methods with high processing load are used to increase the classification success of retinal fundus images, and the method proposed in this study is an on-target solution.

By employing HE and CLAHE to support deep learning (CNN) for diagnosing diabetic retinopathy, the proposed method of this study and the research results are considered as valuable contributions to the associated literature. In addition to technical contributions in this manner, the novelty is caught with the use of the related hybrid method for diagnosis of diabetic retinopathy. In theoretical meaning, the contribution of the research and the proposed approach is to provide more emphasis on role of image enhancement in intelligent diagnosis process and effectiveness of different techniques to use for that purpose and support advanced technologies such as deep learning. Of course, every research may have some gaps, which can be considered also as suggestions or future works. General gaps of this study are discussed under next, far paragraphs.

In the context of the study topic, the sections are organized as follows: The data and methods used along the study and the performed applications with the obtained results in this manner are explained in detail under the second and the third sections. In detail, the results were also compared with the ones from some other studies using the same database but different classification methods. After giving some results regarding real use case of the proposed method, some possible gaps are also discussed accordingly. With the last section, the content is ended by discussions on conclusions and some future work plans.

2 Data and methods

Based on the objective and motivations, this study is associated with some background ideas and research efforts as shown in Fig. 1. Briefly, especially using deep learning for diagnosis and supporting it with image processing has been remarkable ideas to follow. As general, widely followed diagnosis approach performed with artificial intelligence has been directed to the disease of diabetic retinopathy.

As it was mentioned before, this study followed an easy-to-design image processing–deep learning approach for diagnosing diabetic retinopathy, by considering retinal fundus images as input data. In this context, Fig. 2 represents the stages within the flow of the introduced hybrid approach. For retinal fundus images enhancement approach, a practical phase including HE and CLAHE was used accordingly. These are well-known image

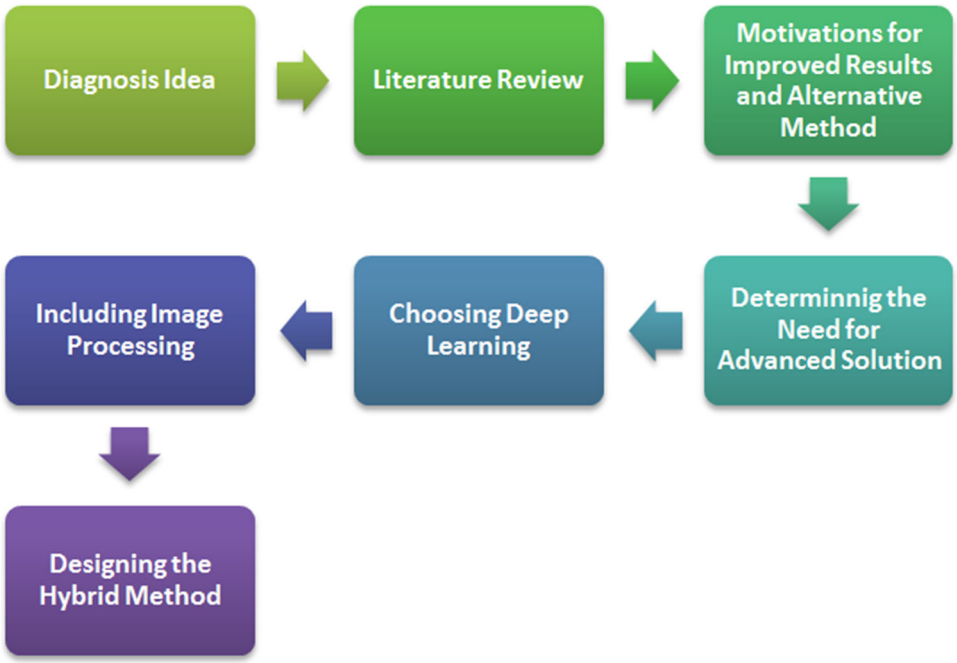

Fig. 1 Ideas and research efforts on the background of this study

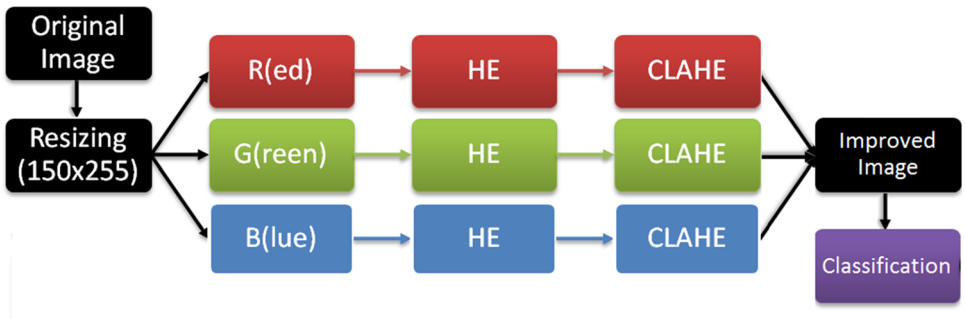

Fig. 2 Stages within the flow of the proposed approach

enhancement techniques, and implementation of them is easy. After the image processing-based enhancement, the classification was made by using a CNN. In next stages, the introduced method was aimed to be evaluated by using 400 retinal fundus images in the MESSIDOR database. Here, both image processing techniques are important for a good image enhancement, which will be effective for better diagnosis–classification at the end. The whole flow is an intelligent approach applied to target image data, which is essential for diagnosing from medical inputs in the form of visual elements.

2.1 The MESSIDOR database

In the context of medical diagnosis studies, the MESSIDOR database has been created during the MESSIDOR project for evaluating different methods of lesion segmentation, by considering color eye fundus images. The project approach is associated with the tasks regarding diabetic retinopathy screening and diagnosis. The database has been open to the public for ten years.

Image obtaining processes realized for creating the MESSIDOR database can be explained briefly as follows [52, 53]: Three ophthalmologic departments acquired a total of 1200 eye fundus color images of the posterior pole, thanks to a color video 3CCD camera (on a Topcon TRC NW6 non-mydriatic retinograph over a 45 degree field of view). In this context, the related images were captured via 8 bits per color plane at, respectively, 1440*960, 2240*1488 or 2304*1536 pixels. Four hundred images without dilation and 800 images with pupil dilation (done with one drop tropicamide at 0.5%) were acquired along the related process.

2.2 Image processing

This study introduces an alternative solution way for retinal fundus images enhancement approach by employing both HE and CLAHE techniques within image processing. HE

and CLAHE are well-known image enhancement methods, and implementation of them under a common approach is very fast and easy to do. In the study, the following stages were followed briefly: First, sizes of the target images are reduced to 150 × 225 pixels, in order to reduce the amount of required computer memory during learning process (deep learning) that will be done. At the second stage, the retinal fundus images are separated into three image components of R (red), G (green), and B (blue). Following that, HE and CLAHE are applied, respectively, to each R, G, and B components. At the final stage, also a concatenation procedure is applied to the R, G, and B components in order to obtain the colored image, which will be then used by the CNN within the deep learning process.

More details regarding the applied two image processing techniques (HE and CLAHE) are explained under the following sub-sections.

2.2.1 Histogram equalization

HE is one of the most widely used image processing techniques for improving visibility and quality of images. This is achieved by increasing the dynamic range of the histogram for the target image. The HE briefly maps gray levels of the input image into uniform gray levels of an output image. In this way, the output image has a uniform distribution of the gray levels. So, it is possible to indicate that the HE is employed for obtaining a uniform histogram. HE eventually provides a new intensity value based on its previous intensity level for each pixel. Because the histogram regarding low-contrast image is narrow and centered toward the middle of the gray scale, the quality of the image is improved and the histogram is distributed to a wider range [54, 55]. The HE flattens and stretches the dynamic range of the histogram regarding the input image so the contrast of the image is improved.

More details regarding HE can be understood by considering its mathematical approach as follows [56]: It is possible to consider a digital image $F(i, j)$ including a total of N pixels and the gray level within the range: '[0, $K-1$].' Moving from that, the probability density function of the related image can be calculated as:

$$p(k) = \frac{n_k}{N}, \quad \text{for} \quad k = 0, 1, \ldots, K-1 \tag{1}$$

where n_k is the total number of pixels with the number of grayscale k in the image. Next to that, the cumulative distribution function of the image $F(i,j)$ can be calculated as:

$$C(k) = \sum_{m=0}^{k} P_m, \quad \text{for} \quad k = 0, 1, \ldots, K-1 \tag{2}$$

By considering the cumulative distribution function values found via Eq. 2, HE matches an input level k to an output level H_k. This can be achieved by using the following equation:

$$H_k = (K-1) \cdot C_k \tag{3}$$

As a result, the gain H_k at the output level for the conventional HE can be obtained as:

$$\Delta H_k = H_k - (H_k - 1) = (K-1) \cdot P(k) \tag{4}$$

By considering the related equations, it is possible to indicate that the increase in the level of H_k is proportional to the probability of the associated level k in the context of the original image.

Figure 3 provides a brief explanation of the HE process seen over histogram data [57].

HE is very useful for the images having wide spread regions of a tone so that the image having a very light background and dark foreground is observed. Thanks to the HE, it is possible to figure out hidden details within an image thanks to stretching out the contrast of local regions, which allows making observable the differences within the regions processed [58].

2.2.2 Contrast limited adaptive histogram equalization (CLAHE)

CLAHE is a technique, which is used for enhancing, especially underwater images. CLAHE was actually developed for improving medical images but nowadays, it can be used for all types of images that are wanted to be enhanced. CLAHE briefly performs a very clear contrast

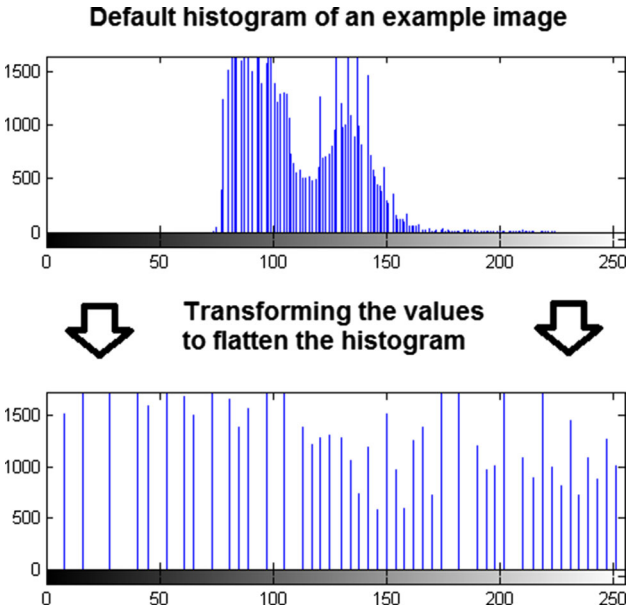

Fig. 3 A typical HE process seen over histogram data [57]

improvement in both medical and normal images. CLAHE is also very effective for condensation of low-contrast images such as underwater images.

Solution mechanism of the CLAHE is briefly as follows: It firstly separates the images into nominal partition and applies the histogram equalization to each partition considered. This process equalizes the distribution of the used gray values in the target image. In this way, the hidden properties of the image become more visible according to its first form. At this point, the whole gray spectrum is used to describe the image and also sharp area edges can be protected by selective condensation of the area boundaries. Selective condensation is done by detecting the area edge in an image first, and then only processing those regions of the image takes place inside the area edge. By employing a combination of the CLAHE, edge sharpening, and also median filtration, it is even possible to reduce noises while the high spatial frequency content of the image is maintained [58, 59]. Technically, CLAHE is used for limiting the amplification, which was done by clipping the histogram at a certain value. That value determines how much noise in the histogram will be smoothed, which means enhancement in contrast [49]. Generally, clipping the histogram is done with Rayleigh distribution as follows [60]:

$$g = g_{\min} + \left[2(\propto^2) ln\left(\frac{1}{1 - P(f)}\right)\right]^{0.5} \quad (5)$$

In Eq. 5, $P(f)$ corresponds to the cumulative probability distribution and also nonnegative real scalar (the distribution parameter), and $g_{\min}$ defines a minimum pixel value.

Figure 4 represents a brief explanation of the CLAHE process done over image data [61].

2.3 Classification with convolutional neural network

In this study, the classification regarding the diagnosis tasks was done by using deep learning and the technique of CNN in this manner. Performance shown by the whole proposed method was evaluated–validated by using 400 retinal fundus images from the MESSIDOR (database). A typical CNN has a multi-layered structure such as feed forward networks. As different from feed forward networks, a CNN can include several convolutional layers with a sub-sampling section. That multi-layer structure employs layers, which are completely interconnected. From a general perspective, CNN is modeled for gathering two-dimensional images and handling them easily in research operations. CNN makes that possible by employing local links with varying weights for easy image processing. Because of that, CNNs can produce parameters by requiring less training processes [62–64].

CNN models are similar to widely known multi-layered perceptron networks. CNNs employ local correlation with relative links by applying a local linking model between the nodes in neighboring layers. In detail, entries of hidden nodes in a layer are done with a subset of the nodes from the previous layer. In this way, a nonlinear spiral filtering is achieved by gathering so many layers in a common structure. In a typical CNN, each filter scans the input

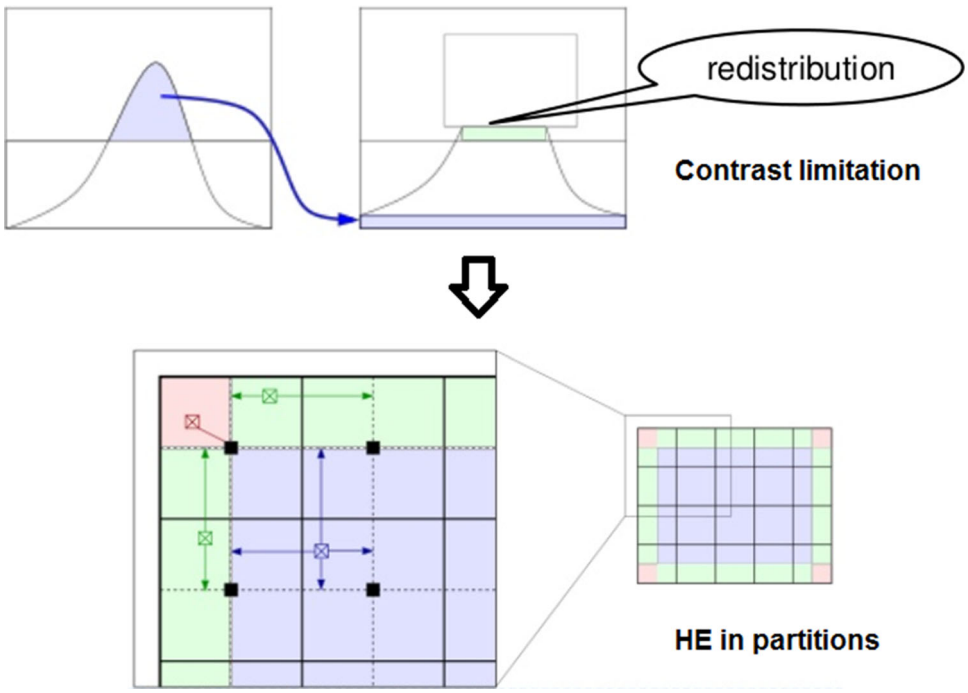

Fig. 4 A typical CLAHE process done over image data [61]

images and the nodes form the feature maps, by using their bias values and shared weights. Here, gradient methods can be used for learning bias values and the shared weights. In Fig. 5, a typical structure of a convolutional layer is represented [64].

Within the hidden layer m, there are two feature maps, which are, respectively, h_0 and h_1. The outputs of the nodes in h_0 and h_1 are obtained (calculated) from the outputs of the nodes entering the 2×2 receive area on the layer beneath the $m - 1$ layer. In this way, the weights: W_0 and W_1 and values: h_0 and h_1 are 3D weight tensors. The mean dimension summarizes the map entries summaries, by pointing to the two other neuron outputs. In order to gather all of them, with each pixel of minus k map of the layer, it focuses on the weight of the link between neuron outputs (i, j) in the coordinates in the context of the feature map: $m - 1$ [64].

2.4 Performance evaluation parameters

In medical classification studies, there are different widely used performance measure calculations. Accuracy, sensitivity, and specificity are some of them, and these measures are employed to evaluate the precision of a used method. Equations used for calculating these measures are as follows briefly [65]:

$$\text{Accuracy} = (TP + TN)/N \tag{6}$$

$$\text{Sensitivity} = TP/P \tag{7}$$

$$\text{Specificity} = TN/N \tag{8}$$

$$\text{Precision} = TP/(TP + FP) \tag{9}$$

$$\text{Recall} = \text{Sensitivity} \tag{10}$$

$$\text{Fscore} = 2 * [(\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall})] \tag{11}$$

$$\text{Gmean} = \text{sqrt}(\text{Sensitivity} * \text{Specificity}) \tag{12}$$

In Eqs. 6–12, TP and FP are used for defining the number of *T*rue *P*ositive and the number of *F*alse *P*ositive diagnoses. On the other hand, TN and FN are for the number of *T*rue *N*egative and the number of *F*alse *N*egative diagnoses. N is the number of total samples, which means also the sum of positive (P) and negative (N) samples in the classification process.

For a classifier, the ability to diagnose correct is calculated via ratio of accuracy (Eq. 6). Additionally, the rate of sensitivity (Eq. 7) defines the extent to which the classifier correctly defines the formation of the target class (that means also recall/Eq. 10). On the other hand, the rate of specificity (Eq. 8) is for defining the target class separation capability of the classifier, while precision rate (Eq. 9) is for understanding how the classifier gives certain class determination. Finally, FScore value (Eq. 11) is a harmonic mean of the precision and sensitivity (recall) to understand better about both abilities with one value and the GMean value (Eq. 12) does that for both sensitivity (recall) and specificity thanks to one value [65].

In this study, all of the related evaluation parameters were used for proving the performance of the solution approach via different perspectives required. All the parameters have been calculated for the CNN after three image processing stages as resizing the image, applying HE, and applying CLAHE. In this way, the performance has been evaluated in terms of also improvement level.

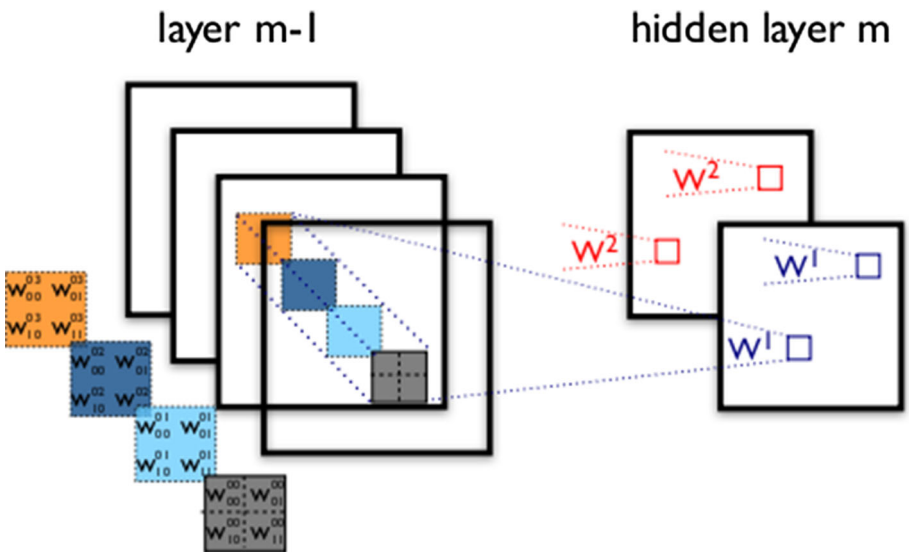

Fig. 5 Typical structure of a convolutional layer [64]

3 Application and evaluation

This section gives information regarding the application side of the proposed method. It is important to apply such a method by using an example database from the literature and run necessary evaluation parameters to understand better about its performance. A comprehensive enough evaluation in this manner should include also a comparative approach and consider real use case of the method, if it is applicable.

3.1 Application and evaluation on image processing

In the application side of this study, MATLAB r2017a and its related tools were used for all image processing and classification phases. A computer system with Intel R9 Core i5-3230 M, 2.60 GHz CPU, 8 GB RAM (running Windows 10 64 bit operating system) was employed for the application. As the data, 400 eye fundus color images regarding the posterior pole were used from the MESSIDOR DB1 database for evaluating the performance of the proposed method. The database contains four classes such as normal retina, macular edema, proliferative diabetic retinopathy, and non-proliferative diabetic retinopathy.

The application process has been done as follows: Image enhancement is applied to the related 400 images first. In such image processing studies, obtained results are often compared by using some known parameters such as entropy, peak signal-to-noise ratio (PSNR) and mean square error (MSE). These evaluation parameters have been used for a long time in order to compare different methods on image data. The best resultant image is known with a lower MSE, higher PSNR, and also higher entropy. Considering the different image processing stages (mentioned before), the obtained results in each are provided under next figures and tables. The sample images for (a) Original image, (b) resized image, (c) HE, (d) CLAHE are given in Fig. 6 for the 20051019_38557_0100_PP.tif image file from the MESSIDOR DB1. Also, sample histogram diagrams for (a) Original image, (b) Resized image, (c) HE, (d) CLAHE are represented accordingly in Fig. 7 (for the 20051019_38557_0100_PP.tif image file).

As shown in Table 1, the entropy value of the image is increased after the image is cured and the MSE value is reduced. This means a good enhancement for the image is considered.

3.2 Application and evaluation on diagnosis (Classification)

After the image processing, the diagnosis is performed via classification task done by the CNN. In this context, the CNN model has eight layers, which are called, respectively, as image input layer, convolutional layer, ReLU layer, cross-channel normalization layer, max pooling layer, fully connected layer, softmax layer, and classification layer. Characteristics of these layers are explained briefly as follows:

- Image input layer is used to specify the image values corresponding to $150 \times 225 \times 3$. The numbers here are, respectively, for height, width, and RGB format for the images. Data conversions such as data normalization or data augmentation in this layer are based on the idea of randomly flipping and truncating the related data. Data conversations are often used for preventing from the overfitting issue, and they are done automatically at the beginning of the training process.
- In the study, also one convolutional layer was used in the CNN and the parameters of the recursive layer correspond to the size of the network filtering here. These are the height and width of the filter used by the training function while scanning the images. The filter size was set to 4, and the second parameter, which is the number of neurons that determines the number of feature maps and bound to the same region of the output as the number of filters, was selected as 16.
- A nonlinear activation function, which is called as ReLU layer, was provided after the convolutional layer. In the diagnosis-oriented application considered in this study, the rectified linear unit function was used accordingly. At this point, one cross-channel normalization layer was employed and the size of the channel window was set to 2, which means also the channel window size for the normalization.
- Following the previous layers, the max pooling layer is employed for down-sampling as an alternative way of reducing the number of parameters and preventing from overfitting. With this layer, the maximum values of the regions (rectangular) of the inputs specified by the first argument pool size are returned. In this study, the size of the rectangle was set to 4:3. Stride function was also used to determine the size of step, while the image is scanned thoroughly by the training function.
- As the seventh layer, the fully connected layer gathers all the related features that previously located layers have learned about the image, for defining larger patterns. In order to perform classification over them, the last of the fully connected layer gathers all those data accordingly. Because of that, the output size parameter in the last fully connected layer is adjusted as the number of classes in the target table. For classification processes, the fully connected layer generally uses the softmax activation function.

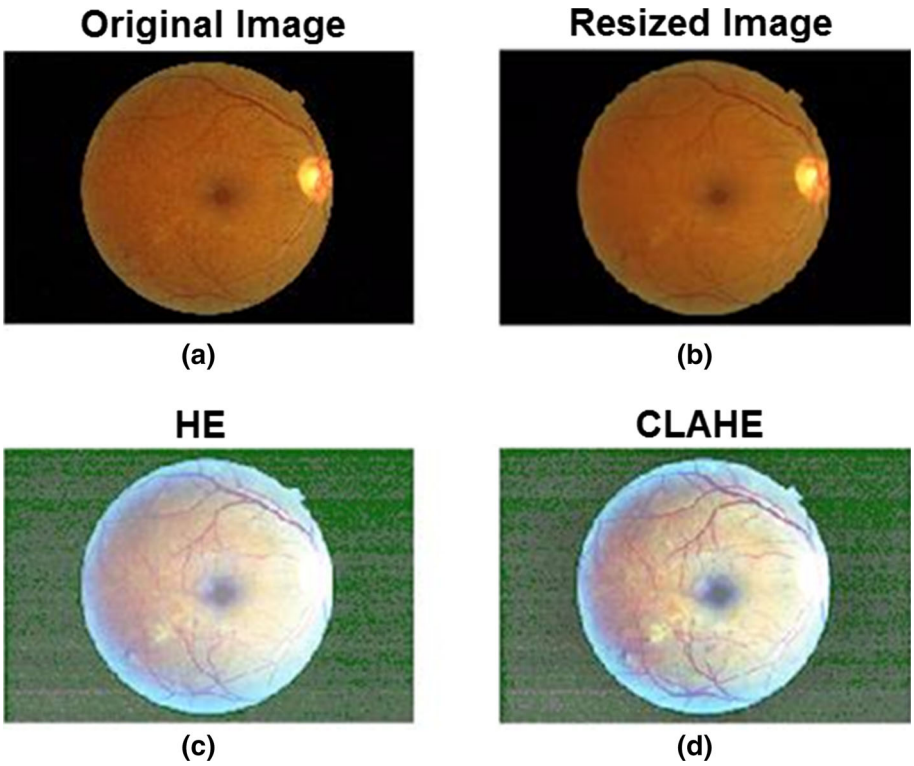

Fig. 6 **a** Original image, **b** resized image, **c** HE, **d** CLAHE

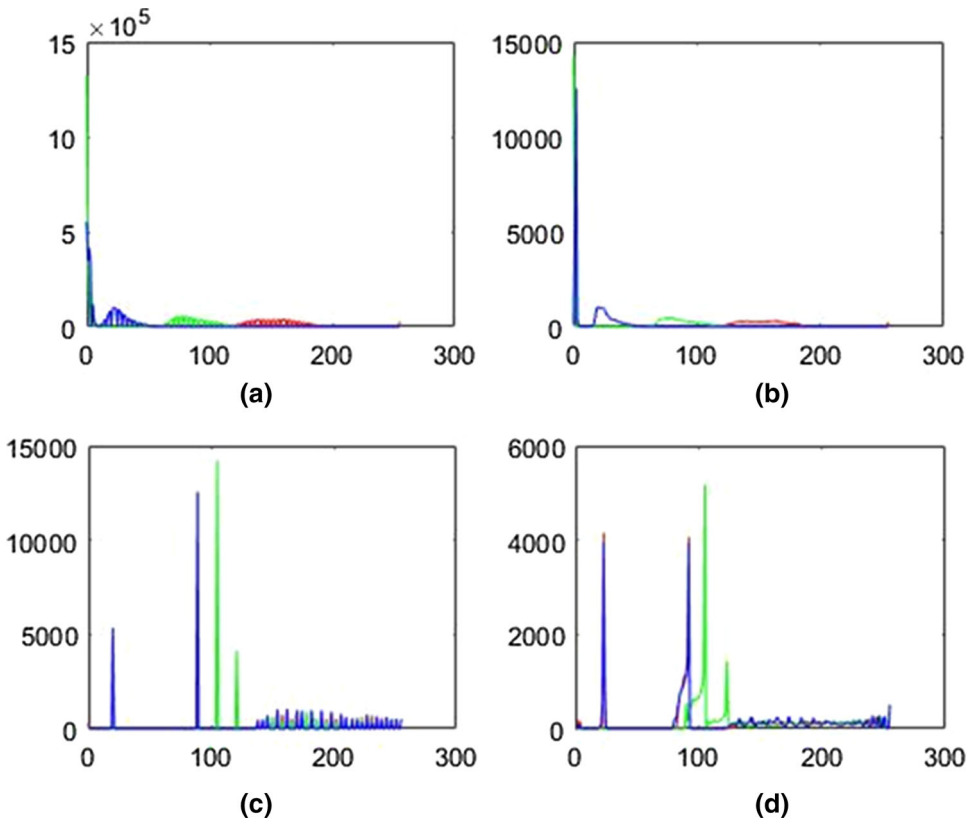

Fig. 7 Histogram diagrams for **a** Original image, **b** resized image, **c** HE, **d** CLAHE

Table 1 Results for different image evaluation parameters over each completed stage

Stage name	Entropy	PSNR	MSE
Original image	5.1425	–	–
Resized image	4.2245	–	–
HE	5.2742	7.7124	11,011.22
CLAHE	6.6517	7.9611	10,398.59

- The classification layer is the final layer of the employed CNN in this study. The layer briefly uses the possibilities returned by the softmax activation function for each input, in order to assign mutually one of the related special classes.

In the application–classification process, 400 images from MESSIDOR DB1 data were used. Here, 300 of the related images were chosen for learning, while the remaining 100 images were for testing–evaluation. Classification tasks were done for each stage to see the contributions of the image processing just before exact diagnosis (in other words, application–classification). A total of twenty independent runs were performed for each stage. In detail, a fine-tuning approach was done by considering stochastic gradient descent with momentum with the starting learning rate of 0.00001 and a limitation of maximum 100 epochs. After an average of 1 h and 50 min, each independent run was ended at the maximum epoch value. Table 2 briefly provides the average values (best values are shown in bold style) for the related evaluation parameters. Next, Fig. 8 represents a graphics for those values.

In the next phase, diagnosis–classification work has been done at after each stage of the image processing. When the target image resized first, the average values for the related evaluation criterions were as accuracy 91%, sensitivity (recall) 82%, specificity 94%, precision 82%, FScore 82%, and GMean 87%. Following that, applying HE caused the accuracy to increase by 3%, the sensitivity to increase by 7%, the specificity to increase by 2%, the sensitivity (recall) to increase by 7%, the FScore to increase by 7%, and finally, the GMean to increase by 5%. After the final CLAHE process, the average values for the classification were as the accuracy 97%, the sensitivity 94%, the specificity 98%, the sensitivity (recall) 94%, the FScore 94%, and the GMean 95%. As it can be understood from the results, the best classification performance was obtained after the CLAHE method. In terms of accuracy as example, there is a difference of around 6% between the resized image and the CLAHE classification. All the related evaluation parameters point that the image processing stage improves the classification success.

3.3 Comparative evaluation

In addition to the obtained results, it is also important to see whether the introduced approach is effective enough in solving the problem. In this context, performance of the method was compared with some previous studies considering the same MESSIDOR database (with same image data) to evaluate its performance well enough. While choosing the related studies to compare, the most remarkable ones (mentioned also under the literature view provided in the first section) with specific techniques such as PSO, FCM, SVM, and K-NN [12, 13, 44, 45, 47, 48] have been included to see their effects according to a possible deep learning-based approach. Additionally, one study with an alternative image processing supported CNN [66] and also three more studies with deep learning [67–69] have been included to have a wide touch with the literature. Firstly, different methods have been evaluated in terms of accuracy (Eq. 6: the ability to diagnose correct), sensitivity (Eq. 7: the extent to which the classifier correctly defines the formation of the target class/recall: Eq. 10), and specificity (Eq. 8: the target class separation capability of the classifier). Table 3 provides the obtained results in this manner. (Best values are shown in bold style.)

Considering Table 3, results can be explained briefly as follows (Fig. 9 also shows graphics for each performance evaluation parameter results):

- Diagnosis–classification of diabetic retinopathy with the hybrid method of this study (image processing and CNN) provided a classification accuracy rate at 97%.
- Additionally, the method had the average rate of sensitivity (recall) as 94%.
- Finally, the average rate of specificity by the proposed method was 98%.

Table 2 Results regarding performance evaluation parameters

Stage name	Accuracy	Sensitivity	Specificity	Precision	Recall	FScore	GMean
Resized image	0.9075	0.8150	0.9383	0.8150	0.8150	0.8150	0.8745
HE	0.9433	0.8867	0.9622	0.8867	0.8867	0.8867	0.9237
CLAHE	**0.9700**	**0.9400**	**0.9800**	**0.9400**	**0.9400**	**0.9400**	**0.9598**

Best values are shown in bold style

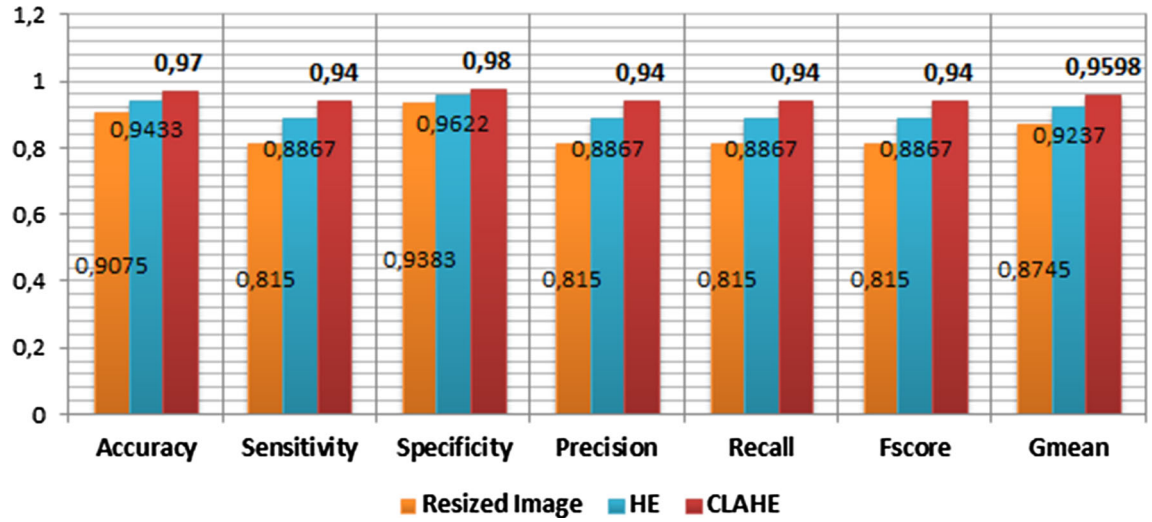

Fig. 8 Graphic of the results for performance evaluation parameters

Table 3 Comparative evaluation in terms of accuracy, sensitivity, and specificity

Method	Accuracy	Sensitivity	Specificity
Image processing and CNN classification (this study)	**0.9700**	**0.9400**	**0.9800**
Ensemble-based framework [12]	0.8200	0.7600	0.8800
Bag of words model [13]	0.9440	0.9400	–
Particle swarm optimization and the fuzzy C-means [44]	0.9450	0.9100	0.9800
Fractal analysis and K-nearest neighbor (KNN) [45]	0.8917	–	–
Texture features and support vector machine (SVM) [47]	0.8520	0.8950	0.9720
Marker-controlled watershed transformation [48]	0.8520	0.8090	0.9020
Alternative image processing and CNN classification [66]	0.9600	1.0000	0.9000
Deep neural network (DNN) [67]	(ROC[a]) 0.98	0.9000	0.9600
Entropy CNN classification [68]	0.8610	0.7324	0.9381
U-net deep learning [69]	(ROC[a]) 0.98	–	–

Best values are shown in bold style

[a]Receiver operating characteristic curve

- It is seen from these results that the proposed method gives better results than some other methods from the literature.
- In terms of accuracy, only [67] and [69] provided better performance (considering receiver operating characteristic curve ROC) with only a difference of %1.
- The alternative method of image processing and CNN [66] provided good performance as like the method proposed in this study. This means use of image processing to support CNN allows having effective enough results.
- Considering all three performance evaluation parameters, the proposed method has a good place among deep learning-based methods compared.
- The worst performance was shown by ensemble-based framework [12].
- All results show that use of deep learning is important to have improved results in terms of diagnosis, and the proposed method provides very good performance according to other methods not including deep learning.

In the context of comparative evaluation, a second evaluation was done based on a ranking approach. At this point, rank of each method considering accuracy value for each image data (a total of 400 images) from the MESSIDOR database was analyzed and a general ranking approach was realized by considering average rank of each method. Table 4 provides information regarding worst rank, best rank, average rank, and standard deviation value considering ranks of the methods. Next, Fig. 10 represents ranking of the compared methods, according to average rank values.

As it can be seen from Table 4, and Fig. 10, the proposed study takes the first place when the comparative

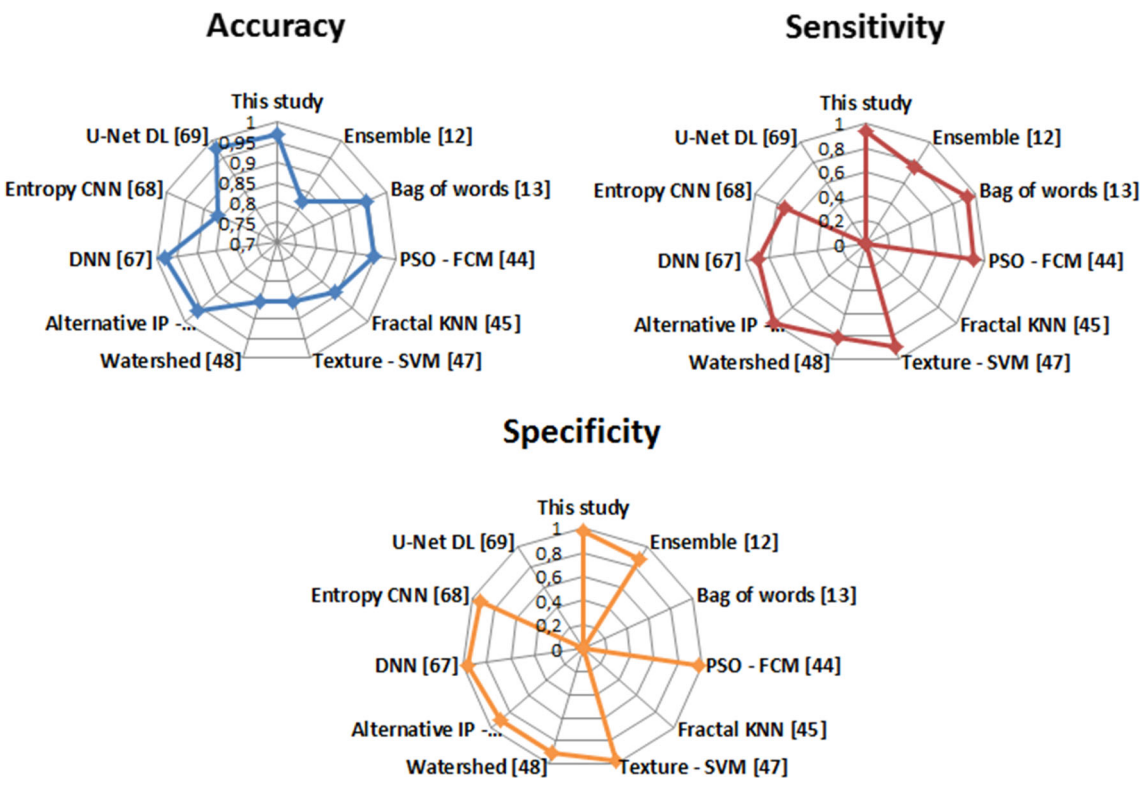

Fig. 9 Graphic of the results for performance evaluation parameters regarding comparative evaluation

Table 4 Comparative evaluation in terms of ranking according to accuracy for each image data

Method	Worst rank[a]	Best rank[a]	Average rank[a]	SD[a]
Image processing and CNN classification (this study)	4	1	1.57	0.367
Ensemble-based framework [12]	11	8	9.69	0.533
Bag of words model [13]	10	8	8.55	0.519
Particle swarm optimization and the fuzzy C-means [44]	8	6	7.92	0.379
Fractal analysis and K-nearest neighbor (KNN) [45]	9	6	6.88	0.283
Texture features and support vector machine (SVM) [47]	8	5	6.19	0.347
Marker-controlled watershed Transformation [48]	8	6	5.70	0.256
Alternative image processing and CNN classification [66]	5	2	2.88	0.311
Deep neural network (DNN) [67]	4	1	1.74	0.464
Entropy CNN classification [68]	6	4	4.06	0.378
U-net deep learning [69]	3	1	1.68	0.432

[a]All values were obtained with 400 images within 20 independent runs

approach was done in terms of ranking for all images and independent runs. The method is followed by [69], [67], and [66], respectively. Top places are taken again by the methods using deep learning.

3.4 Remarkable points and possible gaps

Experiences and the obtained results cause the following remarkable points that can be figured out from that study:

- Deep learning has a great impact on improving solution value of machine learning-oriented systems, which was understood from different perspectives in this study.
- It is important to note that developing hybrid systems is a trendy topic and that will probably be still alive in the future. Currently, deep learning is the strongest member of such systems.
- Medical diagnosis from image data is a critical research interest as it leads to accurate diagnosis, which cannot be done by doctors sometimes.

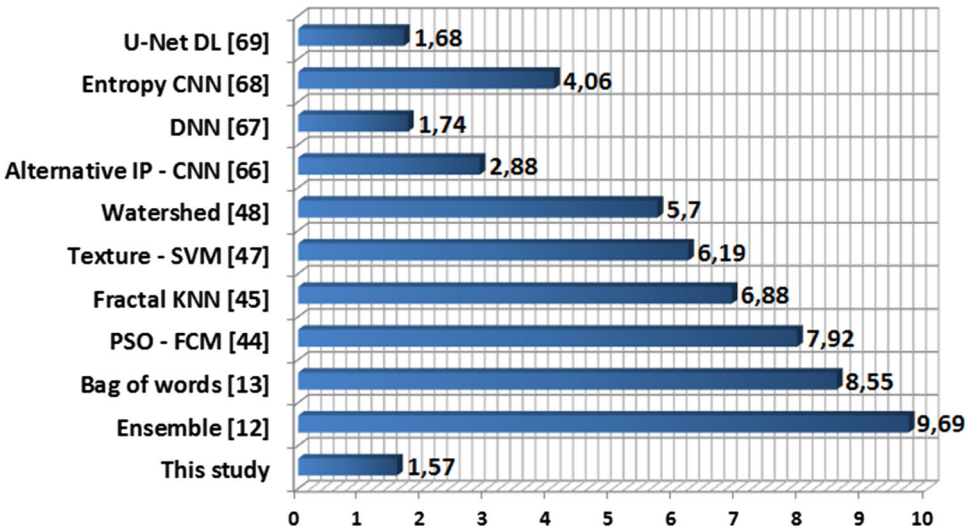

Fig. 10 Total ranking points of each method compared

- Artificial intelligence always needs additional support from alternative techniques to improve its performance in terms of time or complexity. In this study, the image data were supported by image processing to give better inputs to the CNN model.
- It is too remarkable that the proposed method in this study is currently used by some doctors at the Hospital of Suleyman Demirel University, Turkey, and the Hospital of Afyon Kocatepe University, Turkey. So far, experiences seem positive and more results supported with exact recorded responses to surveys or interviews are planned to be provided in the future.

There are also some gaps–limitations that should be expressed as contributive suggestions for further studies by interested researchers. In this sense, possible gaps should be expressed briefly as follows:

- The method has been designed for an easy-to-use approach. Although improved results have been obtained, the method may be edited in detail (in terms of additional techniques, different parameter values, etc.) for more improvements.
- This study was focused on only MESSIDOR database from the literature. The database is a widely used and important data infrastructure, but for more evaluation, additional databases may be used.

4 Conclusions and future work

In this study, using both image processing and deep learning for diagnosing diabetic retinopathy from retinal fundus images was proposed as an alternative solution way. For retinal fundus images enhancement approach, a practical method, which contains both HE, and CLAHE, was used. Both HE and CLAHE are well-known image enhancement solutions, and they are easily implemented to the target image data. After the image processing-based enhancement, the classification was made using the CNN. The performance of the introduced method was evaluated by using 400 retinal fundus images in the MESSIDOR database. In experiments, classification work has been done for each stage of the image processor. In the classification study performed after image processing, the accuracy was 97%, the sensitivity was 94%, the specificity was 98%, the precision was 94%, the FScore was 94%, and the GMean was 95%. The obtained results and two comparative evaluation approaches done with previous studies show that the proposed method is efficient and successful enough at diagnosing diabetic retinopathy from retinal fundus images. As compared with traditional techniques, the advantage of deep learning has been directly seen in this study. Also, the use of practical image processing techniques has supported the deep learning technique: CNN in terms of improving diagnosis–classification performance. Some gaps–limitations reported here can be accepted as suggestions to be followed by interested researchers.

Obtained positive results in this study have encouraged the authors to continue further studies. The future scope has been shaped by the authors to include use of alternative solution mechanisms to see whether results can be improved more and perform more applications. In this context, it is aimed to employ different image processing techniques to see whether the performance and solution quality can be taken to some steps ahead. Additionally, alternative medical databases will also be used for different diagnosis operations. Finally, responses from the hospitals where the method is actively used will be received and more consideration will be given to use of the method in hospitals for diagnosing different diseases as it is believed

that the solution can be very useful for doctors (and medical staff) in terms of general diagnosis processes.

References

1. World Health Organization and International Diabetes Federation (2005) Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia, a report of world health organization and international diabetes federation
2. International diabetes federation (2013) Diabetes atlas, 6th edn. Brussels, Belgium https://www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf. Accessed 23 Jan 2018
3. International diabetes federation (2009) Diabetes atlas, 4th edn. Brussels, Belgium. https://www.idf.org/sites/default/files/IDF-Diabetes-Atlas-4th-edition.pdf. Accessed 23 June 2018
4. Bloom DE, Cafiero E, Jané-Llopis E, Abrahams-Gessel S, Bloom LR, Fathima S et al (2012) The global economic burden of noncommunicable diseases (No. 8712). Program on the global demography of aging
5. Scully T (2012) Diabetes in numbers. Nature 485(7398):S2
6. Chase HP, Maahs DM (2011) A first book for understanding diabetes. Children's Diabetes Foundation at Denver, Colorado
7. Johnstone MT, Veves A (eds) (2005) Diabetes and cardiovascular disease. Springer Science & Business Media, New York
8. Veresiu AI, Bondor CI, Florea B, Vinik EJ, Vinik AI, Gâvan NA (2015) Detection of undisclosed neuropathy and assessment of its impact on quality of life: a survey in 25,000 Romanian patients with diabetes. J Diabetes Complications 29(5):644–649
9. Congdon NG, Friedman DS, Lietman T (2003) Important causes of visual impairment in the world today. JAMA 290(15):2057–2060
10. Taylor HR, Keeffe JE (2001) World blindness: a 21st century perspective. Br J Ophthalmol 85(3):261–266
11. Kaji Y (2018) Diabetic eye disease. diabetes and aging-related complications. Springer, Singapore, pp 19–29
12. Antal B, Hajdu A (2012) An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. IEEE Trans Biomed Eng 59(6):1720
13. Islam M, Dinh AV, Wahid KA (2017) Automated diabetic retinopathy detection using bag of words approach. J Biomed Sci Eng 10:86–96
14. Sutton J, Mahajan R, Akbilgic O, Kamaleswaran R (2018) PhysOnline: an online feature extraction and machine learning pipeline for real-time analysis of streaming physiological data. IEEE J Biomed Health Inf 10:11–12. https://doi.org/10.1109/jbhi.2018.2832610
15. Gencer C, Coskun A (2005) Robust speed control of permanent magnet synchronous motors using adaptive neuro fuzzy inference system controllers. Asian J Inf Technol 4(10):918–919
16. Kose U, Arslan A (2017) Optimization of self-learning in Computer Engineering courses: an intelligent software system supported by artificial neural network and vortex optimization algorithm. Comput Appl Eng Educ 25(1):142–156
17. Coskun A (2011) Optimization of a mini-golf game using the genetic algorithm. Electron Electr Eng 3(109):97–100
18. Coskun A, Horat B (2014) Mobese placement optimization within ankara by using genetic algorithms. Sci Res Essays 9(16):716–721
19. Coskun A (2016) Edge detection using ant colony optimization with specific images. Int J Res Advent Technol 4(12):2016
20. Coskun A, Yilmaz Y (2017) Bone age assessment with fuzzy logic. RA J Appl Res 3(12):2017
21. Coskun A, Arici N (2011) Defining the possible molecular structure of the drug to be penetrated through skin layers using genetic algorithm. Gazi Univ J Sci 24(2):2011
22. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN et al (2018) Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell 173(2):338–354
23. Coskun A (2011) Simulated annealing algorithm and layout optimization for the contents of a web page. In: 3rd international conference on electronics computer technology
24. Hussein AF, ArunKumar N, Ramirez-Gonzalez G, Abdulhay E, Tavares JMR, de Albuquerque VHC (2018) A medical records managing and securing blockchain based system supported by a genetic algorithm and discrete wavelet transform. Cogn Syst Res 52:1–11
25. Atanasov P, Gauthier A, Lopes R (2018) Applications of artificial intelligence technologies in healthcare: a systematic literature review. Value Health 21:S84
26. Wartman SA, Combs CD (2017) Medical education must move from the information age to the age of artificial intelligence. Acad Med 93:1107–1109
27. Hamet P, Tremblay J (2017) Artificial intelligence in medicine. Metabolism 69:S36–S40
28. Xing L, Krupinski EA, Cai J (2018) Artificial intelligence will soon change the landscape of medical physics research and practice. Med Phys 45(5):1791–1793
29. Gupta D, Sundaram S, Khanna A, Hassanien AE, de Albuquerque VHC (2018) Improved diagnosis of Parkinson's disease using optimized crow search algorithm. Comput Electr Eng 68:412–424
30. Gupta D, Julka A, Jain S, Aggarwal T, Khanna A, Arunkumar N, de Albuquerque VHC (2018) Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. Cogn Syst Res 52:36–48
31. Hemanth JD, Kose U, Deperlioglu O, de Albuquerque VHC (2018) An augmented reality-supported mobile application for diagnosis of heart diseases. J Supercomput. https://doi.org/10.1007/s11227-018-2483-6
32. Moreira MW, Rodrigues JJ, Al-Muhtadi J, Korotaev VV, de Albuquerque VHC (2018) Neuro-fuzzy model for HELLP syndrome prediction in mobile cloud computing environments. Concurr Comput. https://doi.org/10.1002/cpe.4651
33. Peixoto SA, Rebouças Filho PP, Kumar NA, de Albuquerque VHC (2018) Automatic classification of pulmonary diseases using a structural co-occurrence matrix. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3736-2
34. Pereira CR, Pereira DR, Rosa GH, Albuquerque VH, Weber SA, Hook C, Papa JP (2018) Handwritten dynamics assessment through convolutional neural networks: an application to Parkinson's disease identification. Artif Intell Med 87:67–77
35. Kakillioglu B, Sharma R, Jindal V (2015) Diabetes diagnosis by machine learning using acquired data set, project report, Syracuse University
36. Rebouças Filho PP, Peixoto SA, da Nóbrega RVM, Hemanth DJ, Medeiros AG, Sangaiah AK, de Albuquerque VHC (2018) Automatic histologically-closer classification of skin lesions. Comput Med Imaging Graph 68:40–54
37. Rebouças EDS, Marques RC, Braga AM, Oliveira SA, de Albuquerque VHC, Rebouças Filho PP (2018) New level set approach based on Parzen estimation for stroke segmentation in skull CT images. Soft Comput. https://doi.org/10.1007/s00500-018-3491-4
38. Reboucas Filho PP, Reboucas EDS, Marinho LB, Sarmento RM, Tavares JMR, de Albuquerque VHC (2017) Analysis of human tissue densities: a new approach to extract features from medical images. Pattern Recogn Lett 94:211–218

39. Rodrigues MB, Da Nóbrega RVM, Alves SSA, Rebouças Filho PP, Duarte JBF, Sangaiah AK, De Albuquerque VHC (2018) Health of things algorithms for malignancy level classification of lung nodules. IEEE Access 6:18592–18601
40. Rebouças Filho PP, Cortez PC, da Silva Barros AC, Albuquerque VHC, Tavares JMR (2017) Novel and powerful 3D adaptive crisp active contour method applied in the segmentation of CT lung images. Med Image Anal 35:503–516
41. Rebouças Filho PP, da Silva Barros AC, Ramalho GL, Pereira CR, Papa JP, de Albuquerque VHC, Tavares JMR (2017) Automated recognition of lung diseases in CT images based on the optimum-path forest classifier. Neural Comput Appl. https://doi.org/10.1007/s00521-017-3048-y
42. Seoud L, Chelbi J, Cheriet F (2015) Automatic grading of diabetic retinopathy on a public database. In: Chen X, Garvin MK, Liu JJ, Trusso E, Xu Y (eds) Proceedings of the ophthalmic medical image analysis second international workshop, OMIA 2015, pp 97–104
43. Savarkar SP, Kalkar N, Tade SL (2013) Diabetic retinopathy using image processing detection, classification and analysis. Int J Adv Comput Res 3(3):285
44. Sreejini KS, Govindan VK (2013) Severity grading of DME from retina images: a combination of PSO and FCM with bayes classifier. Int J Comput Appl 81(16):11–17
45. Safitri DW, Juniati D (2017) Classification of diabetic retinopathy using fractal dimension analysis of eye fundus image. In: AIP conference proceedings, vol 1867, No. 1, p 020011. AIP Publishing
46. Akram MU, Khalid S, Khan SA (2013) Identification and classification of microaneurysms for early detection of diabetic retinopathy. Pattern Recogn 46(1):107–116
47. Acharya UR, Ng EYK, Tan JH, Sree SV, Ng KH (2012) An integrated index for the identification of diabetic retinopathy stages using texture parameters. J Med Syst 36(3):2011–2020
48. Lim ST, Zaki WMDW, Hussain A, Lim SL, Kusalavan S (2011) Automatic classification of diabetic macular edema in digital fundus images. In: IEEE colloquium on humanities, science and engineering (CHUSER), 2011, pp 265–269
49. Agurto C, Murray V, Barriga E, Murillo S, Pattichis M, Davis H et al (2010) Multiscale AM-FM methods for diabetic retinopathy lesion detection. IEEE Trans Med Imaging 29(2):502–512
50. Chandrayan P (2017) Deep learning: autoencoders fundamentals and types. CodeBurst.IO. https://codeburst.io/deep-learning-types-and-autoencoders-a40ee6754663. Accessed 27 June 2018
51. Chicco D, Sadowski P, Baldi P (2014) Deep autoencoder neural networks for gene ontology annotation predictions. In: Proceedings of the 5th ACM conference on bioinformatics, computational biology, and health informatics, pp 533–540. ACM
52. Klein JC, Menard M, Cazuguel G, Fernandez-Maloigne C, Shaefer G, Gain P et al (2016) Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology (MESSIDOR). Ecole des Mines de Paris
53. Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C et al (2014) Feedback on a publicly distributed image database: the Messidor database. Image Anal Stereol 33(3):231–234
54. Singh B, Mishra RS, Gour P (2011) Analysis of contrast enhancement techniques for underwater image. Int J Comput Technol Electron Eng 1(2):190–194
55. Kassab A (2012) Image enhancement methods and implementation in MATLAB, Bachelor of Science Thesis, University of West Bohemia
56. Deperlioglu O, Kose U, Guraksin GE (2018) Underwater image enhancement with HSV and histogram equalization. In: International conference on advanced technologies 2018 (ICAT 2018), Antalya, Turkey
57. MathWorks (2018) Image enhancement by histogram equalization. MathWorks documentation. https://it.mathworks.com/help/hdlcoder/examples/image-enhancement-by-histogram-equalization.html. Accessed 1 Dec 2018
58. Kaur T, Sidhu RK (2015) Performance evaluation of fuzzy and histogram based color image enhancement. Procedia Comput Sci 58:470–477
59. Beohar R, Sahu P (2013) Performance analysis of underwater image enhancement with CLAHE 2D median filtering technique on the basis of SNR, RMS error, mean brightness. Int J Eng Innov Technol 3(2):525–528
60. Yussof WNJHW, Hitam MS, Awalludin EA, Bachok Z (2013) Performing contrast limited adaptive histogram equalization technique on combined color models for underwater image enhancement. Int J Interact Digit Media 1(1):1–6
61. Huang Y (2013) Image color correction and contrast enhancement. SlideShare.Net. https://www.slideshare.net/yuhuang/image-color-correction-contrast-adjustment. Accessed 30 Nov 2018
62. Goodfellow I, Bengio Y, Courville A, Bengio Y (2016) Deep learning. MIT Press, Cambridge
63. Sze V, Chen YH, Yang TJ, Emer JS (2017) Efficient processing of deep neural networks: a tutorial and survey. Proc IEEE 105(12):2295–2329
64. LISA Lab (2018) Convolutional neural networks (LeNet). DeepLearning.Net Documentation. http://deeplearning.net/tutorial/lenet.html. Accessed 27 June 2018
65. Deperlioglu O (2018) Classification of phonocardiograms with convolutional neural networks. Brain 9(2):22–33
66. Sudha LR, Thirupurasundari S (2014) Analysis and detection of haemorrhages and exudates in retinal images. Int J Sci Res Publ 4(3):1
67. Ramachandran N, Hong SC, Sime MJ, Wilson GA (2018) Diabetic retinopathy screening using deep neural network. Clin Exp Ophthalmol 46(4):412–416
68. Lin GM, Chen MJ, Yeh CH, Lin YY, Kuo HY, Lin MH et al (2018) Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy. J Ophthalmol. 2018:2159702. https://doi.org/10.1155/2018/2159702
69. Aujih AB, Izhar LI, Mériaudeau F, Shapiai MI (2018) Analysis of retinal vessel segmentation with deep learning and its effect on diabetic retinopathy classification. In: 2018 International conference on intelligent and advanced system (ICIAS), pp 1–6. IEEE