ORIGINAL ARTICLE

Diabetic retinopathy detection and classification using capsule networks

G. Kalyani¹ ⬤ · **B. Janakiramaiah²** ⬤ · **A. Karuna³** ⬤ · **L. V. Narasimha Prasad⁴**

Abstract

Nowadays, diabetic retinopathy is a prominent reason for blindness among the people who suffer from diabetes. Early and timely detection of this problem is critical for a good prognosis. An automated system for this purpose contains several phases like identification and classification of lesions in fundus images. Machine learning techniques based on manual extraction of features and automatic extraction of features with convolution neural network have been presented for diabetic retinopathy detection. The recent developments like capsule networks in deep learning and their significant success over traditional machine learning methods for a variety of applications inspired the researchers to apply them for diabetic retinopathy diagnosis. In this paper, a reformed capsule network is developed for the detection and classification of diabetic retinopathy. Using the convolution and primary capsule layer, the features are extracted from the fundus images and then using the class capsule layer and softmax layer the probability that the image belongs to a specific class is estimated. The efficiency of the proposed reformed network is validated concerning four performance measures by considering the Messidor dataset. The constructed capsule network attains an accuracy of 97.98%, 97.65%, 97.65%, and 98.64% on the healthy retina, stage 1, stage 2, and stage 3 fundus images.

Keywords Diabetic retinopathy · Deep learning · Capsule networks · Classification

Introduction

Diabetes is a ailment that weakens the ability of the human body to regulate blood sugar. The International Diabetes Federation (IDF) have assessed that worldwide 451 million people suffers with diabetes in 2017 and they estimated that it increase to 693 million by 2045 [1]. A higher level of blood sugar damages the organs and tissues in the body. The higher the value of sugar and the duration of having diabetes, the more prominent is the danger for intricacies. Diabetes can lead to risky complications like a coronary episode, issues with neuropathy and nephropathy, vision misfortune, misery, and dementia.

Based on the parts that are influenced by high blood sugar levels the health issues are termed as diabetic nephropathy (problems identified with kidneys), diabetic neuropathy (problems identified with nerves), and diabetic retinopathy (problems in vision). Diabetic retinopathy is one of the major problems in the diabetic people. In earlier days, diabetes is the problem which attacks only elder people. But, in today's world, diabetic is the issue with which most of the people are exposed irrespective of their age and for those people as consequence problems may arise related to retinopathy. 8.5% of adults who have 18 years and more have reported diabetes in the year 2014. In 2016, diabetes was the direct cause of the fatality of 1.6 million and for 2.2 million people in 2012.

✉ G. Kalyani
  kalyanichandrak@gmail.com

  B. Janakiramaiah
  bjanakiramaiah@gmail.com

  A. Karuna
  karunagouthana@gmail.com

  L. V. Narasimha Prasad
  lvnprasad@iare.ac.in

¹ Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

² Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India

³ University College of Engineering Kakinada(A), Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh, India

⁴ Institute of Aeronautical Engineering, Hyderabad, Telangana, India

As indicated by the International Diabetes Federation, 463 million individuals have diabetes on the planet and 88 million individuals in the Southeast Asia area in 2020. Among 88 million individuals, 77 million people are from India. The quantity of individuals with diabetes in India has expanded from 26 million to 65 million in 1990 and 2016, respectively [2]. Most of the diabetes cases are of type-2 diabetes [3]. In 2010, the Government of India started National Program for Preclusion and Control of Cancer, Diabetes, Cardiovascular Diseases, and Stroke (NPCDCS) [4]. The goal of NPCDCS is to organize outside camps for deft testing at all levels in the medicinal services conveyance framework for early identification of diabetes, among different ailments [5]. Hence, in this paper, we are proposing a methodology for early detection of the vision problems arises with diabetes which can be treated without any loss of vision.

Individuals with diabetes in the long haul can suffer from problems related to vision. Eye issues that can happen with diabetes incorporate cataracts, glaucoma, and retinopathy. Among the three, the retinopathy issue that emerges because of diabetes is named diabetic retinopathy. It arises if high sugar levels ruin the blood veins in the retina. The veins can grow and spill or may be closed, stopping the flow of blood. In other cases, new blood veins may grow on the retina. These developments can cause problems for the vision. Figure 1 shows the retinal pictures of healthy and diabetic retinopathy.

In general, diabetic retinopathy is examined in four stages. Mild non-proliferative retinopathy is the early stage among the four stages. In this stage, microaneurysms will be formed. Microaneurysms are little regions of balloon-like growth in the retina's small veins. The subsequent stage to mild non-proliferative retinopathy is termed moderate non-proliferative retinopathy. Veins that support the retina are obstructed at this stage. The next stage is named as severe non-proliferative retinopathy. More veins are obstructed at this stage, denying a few regions of the retina with the blood gracefully. These zones of the retina impart signs to the body to develop fresh blood vessels for sustenance. The last phase is Proliferative Retinopathy. Here, the signs given by the retina for sustenance activate the development of a fresh artery. The arteries are anomalous and delicate. They develop beside the retina and outside of the reasonable, vitreous gel that fills within the eye. Without anyone else, these veins do not cause manifestations or vision misfortune because they have slim, delicate walls. In those blood vessels release blood, then serious vision misfortune and visual deficiency can occur.

As specified by the 2019 National Diabetes and Diabetic Retinopathy Survey report, the frequency of occurrence was seen as 11.8% in individuals beyond 50 years old [6]. The pervasiveness was the same in male and female populaces. It was higher in urban territories [7]. When studied for diabetic retinopathy, which compromises visual perception, 16.9% of
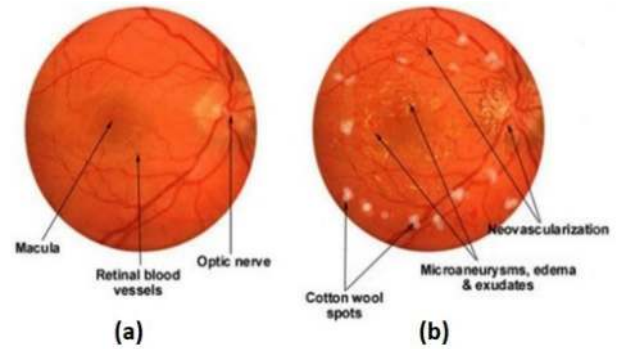

Fig. 1 Retinal images. **a** Healthy retina; **b** diabetic retinopathy

the diabetic populace up to 50 years were seen as influenced. As per the report, diabetic retinopathy in the 60- to 69-year age range was 18.6%, and in the 70- to 79-year age range was 18.3%, and for more than 80 years old, it was 18.4%. A lower pervasiveness of 14.3% was perceived in the 50- to 59-year age range [6]. High prevalence of diabetes is observed in economically and epidemiologically boosted states, like, Tamil Nadu and Kerala, in which many research centers are carried out their research on occurrence frequency are additionally present [4].

Review of literature

Diabetic retinopathy is the most noticeable problem in diabetic patients which leads to loss of vision which can be avoided if the problem is identified and treated at a prior stage. When the problem has been diagnosed, the patient must be checked at regular intervals to know the advancement of the ailment [8]. An efficient mechanism for detecting the problem based on the patient's reports will be useful for the ophthalmologist to avoid the loss of vision due to diabetes. Researchers have designed various algorithms for the investigation of scan reports thereby to do the exact diagnosis of diabetic retinopathy [9–11]. The human eye consists of optic nerves and discs. So the eye images are to be segmented into parts for identification and categorization of DR. It can also be possible by inspecting the fundus images for the existence of hemorrhages, injuries, smaller scale aneurysms, exudates, and so forth. Deep learning plays a vital role in making the scenarios smart, i.e., making the environment smart with automation [12] and making the medical domain also as smart with automated disease detection [13–15].

In the recent decade, number of automated systems have been created for the diagnosis of diabetic retinopathy [16–18]. Generally, in the manual diagnosis, the specialists will concentrate on lesions like microaneurysms, exudates, and hemorrhages in examining the fundus pictures. Hence, the majority of the researchers concentrated on automatically

identify and categorize the lesions [19]. In the paper [20], the author's Shahin et al. built up a framework to automatically categorize retinal fundus pictures as having and not having diabetic retinopathy. The authors considered morphological dispensation to retrieve necessary features like the area of the blood vessels and exudates region along with entropy and homogeneity indexes. Later the extracted features are given as inputs to the deep neural network and they recorded the performance parameters as 88% and 100% for sensitivity and specificity, respectively.

The authors Kumar et al. [21] have presented an improved methodology for hemorrhage recognition and microaneurysms that contributed to the improvement when compared to the existed methodology for identification of DR. In the methodology, an upgraded segmentation strategy was introduced for separating the blood veins and optic disc. Consequently, the task of classification was done based on a neural network system that was trained by considering the features of microaneurysms. Toward the end, the results have exhibited the advancement of the considered methodology with accuracy as the performance measure.

In the article [22], author Jaafar et al., projected an algorithm for automatic identification which comprises of majorly two sections: the top-down approach to divide the exudates region and a coordinate system focused at the fovea to review the seriousness of rigid exudates. The author tested it on a dataset of 236 fundus pictures and obtained 93.2% sensitivity. The author Casanova et al., in the paper [23] presented an algorithm based on random forest technique to categorize the individuals without and with diabetic retinopathy and got an accuracy of over 90%. The author has evaluated the chance of diabetic retinopathy based on fundamental information and reviewed fundus photos. But, the projected methodology relies upon clinical applicable factors, for example, count of microaneurysms and variation from the normal representation in the reviewed retinal pictures marked by human specialists.

The majority of the works discussed above either depend on the factors physically estimated by specialists or put a lot of exertion into separating the required highlights using the techniques related to image processing which adds additional time complexity. Hence, there is a need to have deep learning strategies with the capacity to gain necessary highlights reasonably from the fundus pictures that have excited the consideration of researchers in the latest years. The author Quellec et al. [24] developed a framework to recognize diabetic retinopathy by utilizing a convolutional neural system (CNN) which automatically categorizes the injuries by making heatmaps which have the indications for the possibility to find novel biomarkers in the fundus pictures. They trained CNN with a collaborative learning strategy which is positioned second place in Kaggle competition on Diabetic

Retinopathy [25] and acquired an outcome with the zone of 0.954 under the ROC curve on the Kaggle dataset.

In the article [26], the author Gulshan et al. used a CNN model called Inception-V3 to recognize referable diabetic retinopathy given a dataset that comprises over 128,000 fundus pictures. Because of huge training data and a very well separation of fundus pictures, the model accomplished a good performance with a metric area under curve as 0.991 and 0.990 and 97.5% and 96.1% sensitivity on two distinctive test sets correspondingly. The authors Gargeya and Leng [27] projected a technique that integrates deep CNN with customary algorithms of machine learning. In the projected technique, after the preprocessing of the fundus images, they are considered as inputs to a residual network. To the result of the fundus images from the last pooling layer, a few metadata factors are added and are considered as input to a classifier named decision tree classifier to have the separation among normal fundus images and a fundus image with Diabetic retinopathy. The technique accomplishes 0.94 as AUC, 0.93 as sensitivity, and 0.87 as specificity on a test set taken from the public.

Thomas et al. [28] have built up a system that resolute the achievement pace of creating gradable pictures for recognizing DR in kids age below 18 years. They considered the pictures for screening DR which was illuminating the unpretentious varieties like microaneurysms. The implementation results demonstrate that just 6of the underlying DR acknowledgment was ineffectual, accordingly uncovering its advancement in treating DR in youngsters. Gupta et al. [29] have presented a methodology by identifying the basic dissimilarities in vessel designs to achieve better scores. This forecast leads to the identification of the absence or presence of DR. The technique of random forest was considered to handle the issues related to dimensionality. The introduced methodology was assessed by considering both clinic and public datasets with measures sensitivity and specificity.

In the article [30], the authors introduced a methodology that considers several features, for example, occupied area by veins, foveal zone inconsistencies, and microaneurysms. The introduced strategy uses curvelet coefficients of fundus picture and angiograms. The categorization of images is performed in three categories as Normal, Mild NPDR, and PDR. The authors performed the study on 70 patients and got the sensitivity as 100%. The authors of [5] have grouped DR pictures based on the existence of microaneurysms. They extracted and considered the features such as circularity and zone of microaneurysms. The implementation has done by considering the datasets DRIVE [31], and DIARETDB1 [32]. The methodology introduced by the authors achieves 94.44% of sensitivity and 87.5% of specificity. The projected strategy utilizes principal component analysis to isolate the picture of the optic plate from the fundus picture. BY utilizing an enhanced MDD classifier, they have accomplished a

growth of 25–40% as the probability of detecting the exudates closer to the area encompassing the optic plate in the picture. The study considers 39 images out of which 4 are classified as normal images and 35 are classified as fundus images with exudates.

The authors of [33,34] have depicted the utilization of SVM, Bayesian strategy, and probabilistic neural network for distinguishing the phases of DR as NPDR and PDR after the investigation of fundus pictures. The investigation has done by considering a dataset named DIARETDB0 which consists of 130 pictures. As the first step the algorithm separated the parts of veins, exudates, and bleedings in the pictures of DR. In the second phase, the classification algorithms have been applied. The authors have achieved an accuracy of 87.69%, 95.5%, and 90.76% for probabilistic neural network, SVM, and Bayes, respectively.

In the article [35], author Georgios has presented an improved methodology for the exact examination of retinal highlights. The methodology concentrates on examining the high and low points of DR for perceiving the vascular variations. Likewise, Retinal fundus pictures were analyzed, and alongside that, different investigations were done. At last, the advancement of the considered methodology was demonstrated over different models. The researchers of the article [36] have built up a strategy, to identify the presence of hemorrhages and microaneurysms based on supervised classification models. The authors have assessed different patients who were suffering from diabetic retinopathy. The investigation results have offered a superior degree of affectability, which was seen as the one accomplished by the specialists.

The authors of the article [37] validate the approval of results got from utilizing a trained SVM classifier. The results are validated by considering three open datasets, DRIVE [31], DIARETDB1 [32], and MESSIDOR [49]. A precision of 93% has been accomplished in categorizing exudates and microaneurysms by dividing the pictures of blood veins. The article [38] clarifies the utilization of local binary pattern texture feature to recognize exudates and a precision of 96.73% has been acquired. In the article [39], a hybrid classification technique is presented. The strategy includes a bootstrapped decision tree to characterize the fundus pictures. The method creates two binary vessel maps by decreasing the number of dimensions. The proposed strategy has delivered a precision of 95%.

To classify the DR images, the authors of [40] have introduced an algorithm based on SVM classifier which utilizes a Gabor filtering approach. As part of preprocessing the authors have applied Circular Hough Transform (CHT) and CLAHE methods to the DR pictures before applying the classifier. The authors reported a precision of 91.4% with the STARE database [41]. An approach for DR detection based on Multi-Layer Perception Neural Network (MLPNN) was proposed by the authors of the article [42]. They have used the 64-point Discrete Cosine Transform (DCT) activity to retrieve nine measurable highlights. The extracted highlights are consequently given as input to the neural system. Morphological activity by utilizing the intensity of the picture as the limit to fragment is depicted in the article [43]. The way toward sectioning the segments of exudates in the HIS space is clarified. The use of the CNN framework combined with data augmentation techniques for classifying the DR images is discussed in the article [44] by categorizing the DR severity into five parts. The authors have used the Kaggle database for validation and obtained a precision of 75%.

The researchers Shanthi and Sabeenian [45] have set up a plan for the classification of fundus pictures with CNN structure. The adopted CNN has categorized the pictures by considering the seriousness of the problem, to attain higher precision. At last, the precision of the executed model was demonstrated from the implemented results. In the article [46], the authors have demonstrated an automatic way of characterizing a predefined set of fundus pictures. Also, the CNN structure was considered for identifying DR that included division, classification, and identification. From the examination, ideal precision was accomplished that depicted the upgrade of the introduced model.

For image classification, the authors of [47] have introduced a technique for grouping error dependent networks. The framework is tried with a dataset that comprises remote sensing pictures of rural land in a town close to Feltwell (UK). The classification has been performed based on pixels. By considering the components of highlight vectors the pixels were described as one of the five distinctive locations of agriculture like sugar beets, carrots, potatoes, uncovered soil, and stubble. Three distinct sorts of outfits (E1, E2 and E3) were made. The precision got was 89.83%, 87.87%, and 90.46% for E1, E2, and E3 ensembles correspondingly. The authors in [33] have endeavored to perceive macular messes from SD-OCT pictures utilizing the recombined convolutional neural system. In the preprocessing stage, a filter called BM3D is used for the exclusion of noise from the pictures. Initially, the input image is divided into two-dimensional chunks in the BM3D approach. Afterward, the two-dimensional chunks are recombined to shape 3D chunks by considering the resemblance among the chunks. The low-level, medium-level, and high-level highlights are retrieved consecutively from a residual network consist of 18 layers. The network was tried for various sizes of kernel such as 3×3, 5×5, and 7×7. In the considered residual network, the kernel of size 3×3 accomplishes the highest accuracy of 90%.

In the article [48], the researchers have projected a novel methodology for identifying DR utilizing a deep patch-based model. The executed model has limited the complexities and it additionally improved the presentation of the framework utilizing CNN structure. Moreover, the computations related

Table 1 Summarizing the models for diabetic retinopathy with features and challenges

Author [citation]	Adopted scheme	Features	Challenges
Kumar et al. [21]	Neural network	(1) Increased accuracy. (2) Better specificity	Requires implementation on multi class DR classification
Thomas et al. [28]	Fundus images based screening	(1) Improved photographic outcome. (2) Better accuracy	Strict when evaluated over clinical setting requirements
Gupta et al. [29]	Random forest approach	(1) High sensitivity. (2) Improved specificity	Requires analysis on large systems for DR grading
Georgios et al. [35]	Random forest approach	(1) Improved discriminative probability 2. Better reliability	No contemplation on additional data
Manuel et al. [36]	CAD	(1) Improved sensitivity. (2) Better specificity	Occurrence of false negatives in mild cases
Welikala et al. [40]	Multi layered perceptron network	(1) CHT and CLAHE methods for preprocessing 2.used SVM Classifier	Takes more time if the dataset is large
Shanthi and Sabeenian [45]	Convolutional neural network	(1) Better sensitivity. (2) Improved accuracy	Analysis can be extended using large sized datasets
Wan et al. [46]	Convolutional neural network	(1) Enhanced efficiency. (2) Better accuracy	Increased overhead
Partovi et al. [47]	Residual networks	(1) BM3D filter for removal of noise. (2) Dividing the 3D blocks into 2D blocks for processing	Deciding the kernel size
Zago et al. [48]	Convolutional neural network	(1) Improved sensitivity. (2) Better precision	Need to consider clinical data

to the received recognizing strategy were implemented and the upgraded results were achieved for sensitivity and precision. Table 1 demonstrates the summary of the discussed literature.

To summarize, all the works done so far make use of the image processing techniques or convolution neural network (CNN) for detection purpose. But with the use pooling layer in CNN is a drawback which may not identify the retinopathy problems in diabetes people at early stages or with high accuracy. Hence, the objective of this article is detecting the diabetic retinopathy at early stages with high accuracy using a capsule network model. Due to pooling layer is not used in capsule networks, chances of missing the features of the image is less. Hence,the accuracy of the prediction may increases.

Materials and methods

Dataset description

Messidor [49] dataset comprises 1200 RGB fundus images with observations in an Excel file. The images were assimilated by 3 ophthalmologic departments. Images were taken with 8 bits per color and having the resolutions of 1440 × 960, 2240 × 1488, or 2304 × 1536 pixels. Among the 1200 images, 800 images were attained with pupil dilation and the remaining 400 images without dilation. All the

images are segregated as 3 sets, one for each ophthalmologic department. Each set comprises of four parts, with each part containing 100 images in TIFF format and for every image medical details are also available. For every image, two diagnoses were made available by medical experts which are retinopathy grade and macular edema risk. For the evaluation proposed methodology retinopathy grade is considered. The dataset has four grades 0, 1, 2, and 3 where 0 indicates normal or no diabetic retinopathy, and the remaining three grades indicate diabetic retinopathy. Grades 1, 2, and 3 indicate the severity levels with 1 as minimum and 3 as maximum. The grades are described based on the existence of microaneurysms and hemorrhages in the images.The images of four grades are shown in Fig. 2.

Method proposed

The proposed method in this paper aims to check the patients whether they are suffering from diabetic retinopathy or not and if they are having the diabetic retinopathy in which stage it is. Because most diabetic people are suffering from this and may also lead to loss of vision. Due to this reason, it is highly required to classify the persons in a short time with high accuracy. To get highly accurate results, capsule networks are trained for classifying the diabetic retinopathy by considering the fundus images as input.
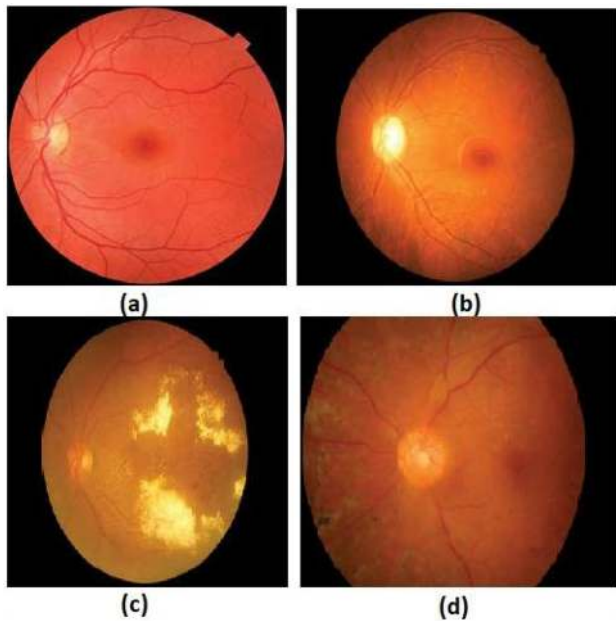
Fig. 2 Images of Messidor dataset **a** grade 0 **b** grade 1 **c** grade 2 **d** grade 3

Preprocessing

The accuracy or quality of the classification model depends on the quality of the input used model training. Hence, the training data in the datasets are applied to preprocessing before considering them as inputs. The retinal images available in the datasets are of different sizes. As a first step of preprocessing, all the images are scaled to the size 512×512. As the second step of preprocessing, the process has applied to the input images to increase the image contrast. The fundus images are generally available as RGB color images. Among the three red, blue, and green channels of the given image, only the green channel of the image is extracted initially in preprocessing. In general, the retinal fundus images are low in contrast. The microaneurysms which play a vital role in classifying diabetic retinopathy are visible more clearly in the green channel image because of the reason of high contrast in that channel image when compared to other channels of red and blue. Further preprocessing has been done for improving the green channel image contrast. The process of contrast limited adaptive histogram equalization (CLAHE) was implemented on the green channel image to enhance the contrast of the image at the tile level. Histogram equalization of CLAHE makes the images equalized all over the basis which improves the contrast of the images further. Later the noise in the images is removed using Gaussian filtering. As a result of the two-step preprocessing, the input images are converted to equally sized and illuminated with improved contrast.

Capsule networks

Convolution neural network has the great ability to automatically identify the feature maps in the given images. A wide variety of tasks that are related to images irrespective of whether the task is classification or detection are solved with CNN. The convolution layer of CNN performs the task of detecting the features from the pixels of the image. Deeper layers perceive the features like edges and color. The architecture of CNN has some drawbacks. As quoted by Hinton [37], the usage of the pooling layer in CNN is a big mistake, and the statement that it works so well is a disaster. Using pooling layers in CNN makes to lose some information related to the images. Another problem in CNN is training requires a huge amount of data. The output of CNN is independent of the small changes to the input image. To avoid these problems of the CNN, Capsule networks are designed.

Capsule networks (CapsNet) are proficient in retrieving the spatial information as well as more significant features without losing any information. Capsules are designed with a set of neurons that works totally with vectors. The neurons of a capsule are designed to work individually for different properties of an object, like position, size, and hue, etc. With this property, capsules are to study the characteristics of an image along with its deformations and viewing conditions. This characteristic has the benefit of identifying the whole entity by identifying the parts initially. The features extracted from the convolution layer are taken as input to a capsule. Depending on the type of the capsule the features given as input to the capsule processed. The output of the capsule is represented as the probability of presenting that feature as a set of vector values named instantiation parameters. In the literature, the implementation of capsules was planned in three ways which are transforming auto-encoders [50], vector capsules based on dynamic routing [51], and matrix capsules based on expectation-maximization routing [52]. The proposed method for diabetic retinopathy of this paper is based on the second method which is capsules with dynamic routing. CapsNet consists of three layers named convolution layer, primary capsule layer, and Class capsule layer. The Primary capsule layer consists of many capsule layers till the last capsule layer also termed the Class capsule layer. The major components of CapsNet are Matrix Multiplication, Scalar Weighting of the Input vector values, dynamic routing, and Squashing Function. The squashing function converts the information into a vector and also the direction of the vector is maintained.

Proposed CapsNet architecture

Figure 3 portrays the Reformed CapsNet architecture for identification and classification of diabetic retinopathy by considering fundus images as input. The operations con-
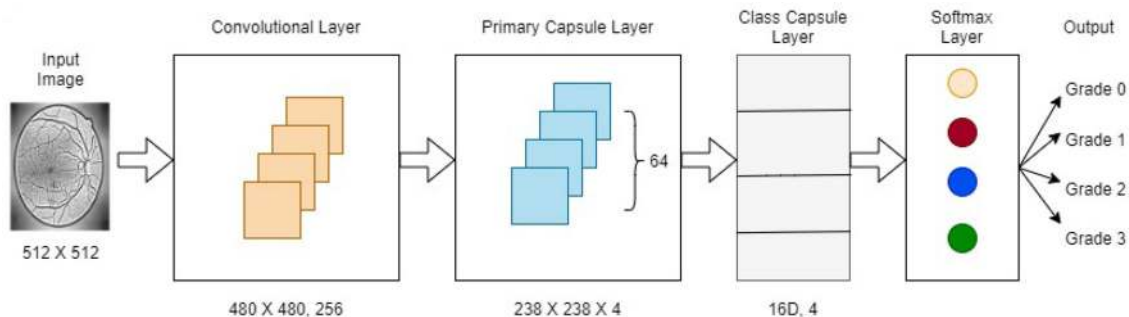
Fig. 3 Proposed architecture of the modified CapsNet for classifying diabetic retinopathy

voluted in the implementation of the reformed CapsNet architecture are as follows:

- Convolution layer for feature extraction from the input image
- Primary capsule layer to process the extracted features
- Class capsule layer on which dynamic routing has been performed to process the features for classification
- Softmax layer to convert class capsule layer result in probabilities corresponding to each class related to diabetic retinopathy

The fundus images are considered as the input image and the process of preprocessing which is mentioned in "Preprocessing" has been applied to elevate the contrast of the picture and to remove the noise in the images for more accurate results. The result of the preprocessing, which is 512 × 512 high-contrast smooth images are taken as input to the CapsNet architecture.

Convolution layer

The first layer in CapsNet is the convolution layer, in which every neuron is associated with a specific region of the area related to the input called a receptive field. The main objective of the convolution layer is to take out the features from the given image. The outcome of every convolution layer is a set of feature maps, created by a solitary kernel filter of the convolution layer. These maps can characterize as input to the following layer.

A convolution is an algebraic operation that slides a function over other space and measures the essentials of their location-based multiplication of value-wise. It has profound associations with the Fourier and Laplace transforms and is intensely utilized in processing the signals. Convolution layers utilize cross-relationships, which are fundamentally the same as convolutions. In terms of mathematics, convolution is an operation with two functions that deliver another function that is, the convoluted form of one of the input functions. The generated function gives integral of the value-wise prod-

uct of the two given functions as an element of the sum that one of the given input functions deciphered.

The design of the convolution layer enables to focus on low-level highlights in the first, and afterwards collect them into more significant level highlights in the following layers. This type of various level structure is regular in natural pictures. A convolution layer in Keras is with the following syntax.

Conv2D (filters, kernel size, strides, padding, activation='relu', input shape)

Arguments in the above syntax have the implications as follows:

- Filters: number of filters to be used.
- Kernel size: Size stipulating both the height and width of the convolution window. Additional optional arguments might also be tuned here.
- Strides: stride of the convolution. If the user does not specify anything, it is set to one.
- Padding: This is either valid or the same. If the user does not stipulate anything, it is set to valid.
- Activation: This is usually ReLu. If the user does not stipulate anything, no activation is used. It is intensely encouraged to add ReLu activation in convolution layers.

It is possible to denote both kernel size and stride as a number or a tuple. When using the convolution layer as a first layer (placed after the input layer) in a model, additional argument input shape. It is a tuple specifying the height, width, and depth of the input. Please make sure that the input shape argument is not included if the convolution layer is not the first layer in the network. The behavior of the convolution layer is controlled by specifying the number of filters and dimensions of each filter. Dimensions of the filters are to be enhanced to enlarge the size of the pattern. There are also a few other hyperparameters that can be tuned. One of them is the stride of the convolution. Stride is the amount by which the filter slides over the image. The stride of 1 moves the filter by 1 pixel horizontally and vertically.

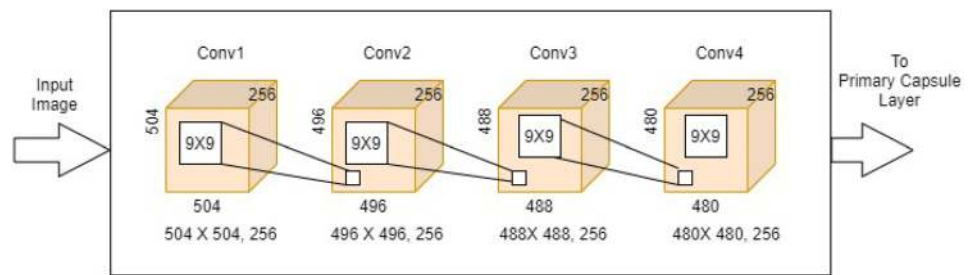**Fig. 4** Convolution layer of the modified CapsNet architecture

Figure 4 shows the design of the convolution layer used in the proposed CapsNet architecture. The preprocessed input image of size 512×512 is considered as input to the convolution layer. The convolution layer is designed with four internal layers of type convolution only named Conv1, Conv2, Conv3, and Conv4. In all the four layers, 256 filters are used with each filter of size 9×9 and stride as 1 for convolving the features.

For the first internal layer (Conv1), the input is of size 512×512 and a kernel of size 9×9 with stride 1 is applied. The resulting feature map is of size $[(512 - 9)/1 + 1]$, i.e., $[504 \times 504, 256]$ which is considered as input to the next internal layer (Conv2). The resulting feature map of Conv2 is of size $[(504 - 9)/1 + 1]$, i.e., $[496 \times 496, 256]$ which becomes an input to the next internal layer (Conv3). The resulting feature map of Conv3 is of size $[(496 - 9)/1 + 1]$, i.e., $[488 \times 488, 256]$ which is input to the fourth internal layer (Conv4). The result of Conv4 of size $[(488 - 9)/1 + 1]$, i.e., $[480 \times 480, 256]$ is the final output size of the convolution layer which is considered as input to the next layer of CapsNet which is the primary capsule layer. The number of parameters to be trained in every internal layer is $256 \times (9 \times 9 + 1)$, i.e., 20,737 parameters.

Primary capsule layer

The subsequent layer to the convolution layer is the primary capsule layer. It consists of three distinct processes: convolution, reshape function, and squash function. The input to the primary capsule layer is fed into a convolution layer. This result from this convolution layer is some array of feature maps, for example, consider that the output is an array of 36 feature maps. Then reshaping function is applied to these feature maps and for an instance, it is reshaped into four vectors of nine dimensions each ($36 = 4 \times 9$) for every location in the image. Now, the last process squashing is applied to guarantee that each vector length is at most one only because the length of every vector indicates the probability of either the object is located or not in the given location of the image. Hence, it should be between 1 and 0. For this purpose, the Squash function is used in the primary capsule layer. This function ensures that the length of the vector is in the range 1 and 0 without destroying the position information. Figure 5

displays the primary capsule layer of the proposed CapsNet Architecture. From the convolution layer, the feature map of size $[480 \times 480, 256]$ is the input to this layer. On the input feature map, a convolution is applied with 5×5 filter and stride as 2. The resulting feature map size is $[(480 - 5)/2 + 1]$, i.e., $[238 \times 238, 256]$.

After this, it consists of 64 primary capsules. The task of primary capsules is to take the main features identified by the convolution and generates the different combinations of the identified features. The layer has 64 "primary capsules" that work in the same manner as the convolutional layer in their characteristics. Each capsule applies four $5 \times 5 \times 256$ kernels (with stride 2) to the $480 \times 480 \times 256$ input volume and hence produces $238 \times 238 \times 4$ output volume. The final output volume size is $238 \times 238 \times 4 \times 64$ because 64 such capsules are used in the proposed architecture. The number of trainable parameters in this layer is $238 \times 238 \times 64$, i.e. 3,625,216 parameters of this layer.

Class capsule layer

The class capsule layer in CapsNet is the replacement for the max-pooling layer of CNN with dynamic routing-by-agreement [36]. This layer of the proposed methodology contains three different types of capsules of dimensions 4, 8, and 16 for every class label. The out of primary capsules is considered as input for the class capsule layer. The routing protocol called dynamic routing is applied between the primary capsules and class capsules. The class capsule layer of the proposed method is shown in Fig. 6.

The initial sub-layer consists of 4D capsules for which the input is from the primary capsule layer. As the second sub-layer 8D capsule is considered for which 4D capsules output is the input. The result of 8D capsules is next considered as input for 16D capsules. The result of 16D is considered as the output of the class capsule layer. The rioting between all the sub-layers of the class capsule layer is the dynamic routing which is described as follows.

The ith previous layer (l) capsule output denoted as v_i is considered as input for subsequent layer ($l + 1$) capsules. The jth capsule of that layer takes the input v_i applies the product with corresponding weight (w_{ij}) between the ith capsule and jth capsule. The resultant is denoted as $v_{j|i}$ which contributes

Fig. 5 Primary capsule layer of the modified CapsNet architecture
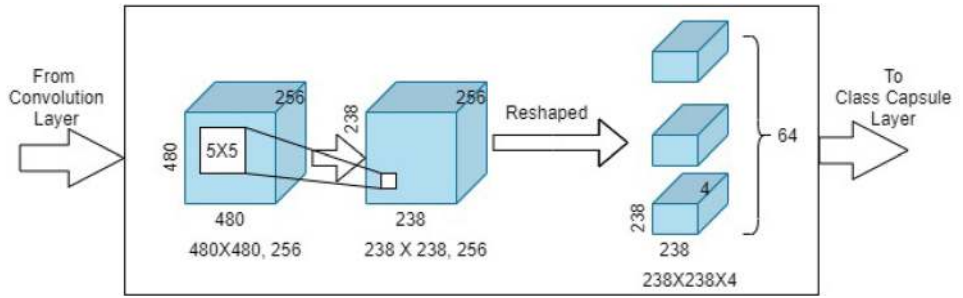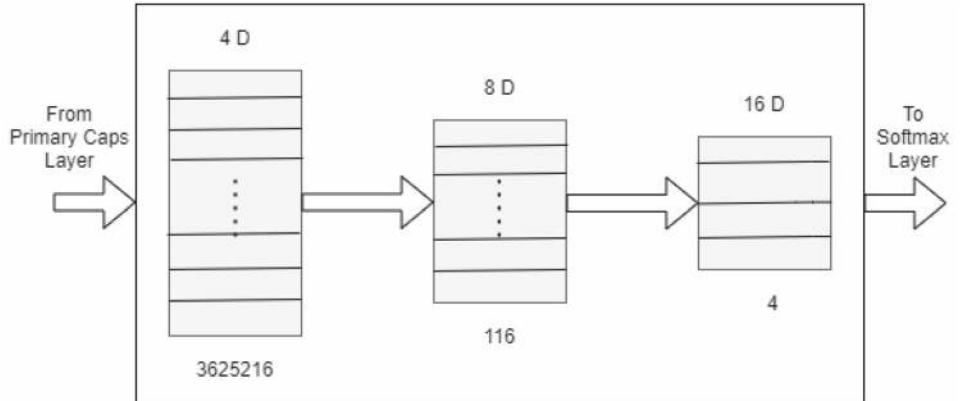

Fig. 6 Class capsule layer of the reformed CapsNet architecture

the contribution of ith capsule of l layer to the jth capsule of $(l + 1)$ layer.

$$v_{j|i} = W_{ij} * v_j, \tag{1}$$

where W_{ij} is the weight between the ith node of the l layer and jth node of the $l + 1$ layer, v_i is the input value for the ith node and $v_{j|i}$ is the output from the ith capsule.

A weighted sum of all the primary capsule predictions for the class capsule j is measured with the squash function as

$$S_j = \sum_{i=1}^{N} C_{ij} * v_{j|i}, \tag{2}$$

where C_{ij} is the result of softmax function in the capsule, $v_{j|i}$ is the weighted input and S_j is the weighted sum of all nodes in the capsule.

To ensure that this result is between 0 and 1, squashing function is applied which is computed as follows:

$$SQ_j = \frac{||S_j||^2 * ||S_j||}{1 + ||S_j||^2 * ||S_j||}. \tag{3}$$

The process of dynamic routing by agreement is illustrated in the following Algorithm 1.

The class capsule layer requires training of more number of parameters. The trainable parameters as C_{ij} are computed as no.of vectors received from the Primary capsule layer multiplied with no.of vectors required as output.

Algorithm 1 :Dynamic Routing

1: **procedure** ROUTING($v_{j|i}, m, l$)
2:     **for** every l layer capsule i to $(l + 1)$ layer capsule j **do**
3:         $w_{ij} \leftarrow 0$;
4:     **for** iteration 1 to m **do**
5:         **for** every capsule i $\in$ layer l **do**
6:             c $\leftarrow$ softmax(w_i);
7:         **for** every capsule j $\in$ layer $(l + 1)$ **do**
8:             $S_j \leftarrow \sum_i C_{ij} * v_{j|i}$;
9:         **for** every capsule j $\in$ layer $(l + 1)$ **do**
10:            $SQ_j \leftarrow squash(S_j)$;
11:        **for** every capsule i $\in$ layer l and capsule j $\in$ layer $(l + 1)$ **do**
12:            $w_{ij} \leftarrow w_{ij} + v_{j|i} * SQ_j$;
13:    **return** SQ_j

between 4D to 8D + between 8D to 16D

$$= 3,625,216 \times 116 + 116 \times 5$$
$$= 420,525,056 + 580$$
$$= 420,525,636.$$

Parameter training is also required at conversion of scalar values into vector which happens between the capsules, i.e., W_{ij}. The trainable parameters here are computed as follows:

W_{ij} between 4D to 8D capsules + W_{ij} between 8D to 16D capsules

$$= 420,525,056 \times 4 \times 8 + 580 \times 8 \times 16$$
$$= 13,456,801,792 + 72,240$$
$$= 13,456,874,032.$$

Hence, the total trainable parameters are

$$= 420, 525, 636 + 13, 456, 874, 032$$
$$= 13, 877, 399, 668.$$

SoftMax layer

Softmax is included generally as the output layer. The nodes in the softmax layer is same as the output labels. Hence, in the proposed method, the numbers of nodes in the softmax layer are taken as five nodes. The softmax is implemented as full softmax which indicates the calculation of probability for every possible class.

Performance metrics

When constructing a classification model, estimating how precisely it predicts the correct result is significant. But, this estimation alone is not sufficient as it conveys deluding results in some cases. That is the situation where the additional measures become an integral factor to conclude the more significant estimations of the constructed model.

Confusion Matrix

Estimating the efficiency of the model developed for classification depends on the number of test records accurately and erroneously anticipated by the model. A confusion matrix is a way of presenting counts of the records correctly and wrongly estimated in the form of a table. The value in the confusion matrix indicates the number of estimations given by the model which includes the predictions done accurately or mistakenly. The entries of the confusion matrix are generally termed as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP indicates the count of the records for which the model properly predicts the positive class as positive. TN indicates the count of the records for which the model properly identifies the negative class as negative. FP indicates the count of the records for which the model erroneously estimates the negative class as positive. FN indicates the count of the records for which the model erroneously estimates the positive class as negative. In general, the confusion matrix is used in the case of binary classification. Due to the application of this paper is a multiclass classification problem, the concept of a binary class confusion matrix is also extended to the multiclass classification task. In diabetic retinopathy mostly the datasets contain either 4 or 5 classes. Hence, we are demonstrating the confusion matrix with 5 classes here. Table 2 shows a confusion matrix for the five-class classification model.

In multi-class classification problems TP, TN, FP, and FN are evaluated for each category of class separately. For

Table 2 Confusion matrix for multi-class classification

		Actual class				
		C1	C2	C3	C4	C5
Predicted class	C1	a	b	c	d	e
	C2	f	g	h	i	j
	C3	k	l	m	n	o
	C4	p	q	r	s	t
	C5	u	v	w	x	y

example, the estimation of TP, TN, FP, and FN for class C1 is demonstrated here based on the confusion matrix in Table 2. The TP of class C1 is 'a' because that is the count of the records for which the actual class is C1 and the predicted class is also C1. The TN of class C1 is '(g+h+i+j+l+m+n+o+q+r+s+t+v+w+x+y)' because that is the count of records which belongs to non C1 class and predicted also as non C1 class. The FP of class C1 is '(b+c+d+e)' because for these records the predicted class is C1 whereas the actual class is non C1. The FN is '(f+k+p+u)' because for these records the predicted class is non C1 but the actual class is C1. Similarly, TP, TN, FP, and FN can be estimated for the other classes also based on the confusion matrix. The performance outcomes that can be evaluated based on the confusion matrix are accuracy, precision, specificity, recall or sensitivity, and F1-score. For every class, the measures are evaluated separately. The average of all classes is considered as the final value for that measure.

Measures based on confusion matrix

Accuracy is an essential metric for classification models. It is easy to understand and simple to apply for binary and also for multiclass classification problems. Accuracy indicates the proportion of true results in the total number of records tested. Accuracy is effective for assessing the classification model which is constructed from balanced datasets only. Accuracy may interpret wrong results if the given dataset for classification is skewed or imbalance. Precision indicates the proportion of the true positives in predicted positives. Another important measure is recall which conveys more information in case if capturing all possible positives is important. Recall indicates the fraction of total positive samples were correctly predicted as positive. The recall is 1 if all positive samples are predicted as positive. If an optimal blend of precision and recall is required then these two measures can be combined as a new measure called F1 score. F1 score is the harmonic mean of the precision and recall which lies in the range of 0 and 1. The formulas to evaluate all these measures are shown in Eqs. (1)–(4).

Table 3 Confusion matrix of Messidor dataset

		Predicted class				Total images
		Grade 0	Grade 1	Grade 2	Grade 3	
Actual class	Grade 0	104	1	1	1	107
	Grade 1	2	54	1	1	58
	Grade 2	1	2	81	2	86
	Grade 3	0	0	0	53	53
Total images		107	57	83	57	304

Table 4 Observations based on the confusion matrix

	Grade 0	Grade 1	Grade 2	Grade 3
TP	104	54	81	53
TN	188	238	211	239
FP	3	3	2	4
FN	3	4	5	0

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$\text{Precition} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{F1-Score} = \frac{2 \times \text{precition} \times \text{recall}}{\text{precition} + \text{recall}} \tag{7}$$

In practical, a model is to be constructed with precision and recall as 1 which in turn gives F1 score as 1, i.e. a 100% accuracy which is not feasible in a classification task. Hence, the constructed classification model should have higher precision with a higher recall value.

Results and discussion

The implementation and training of the proposed capsule network is done on a system which is with nVIDIA CUDA architecture. The system is equipped with AMD Ryzen 2700x processor having eight cores and with a RAM of 32 GB. The system uses nVIDIA Graphics Processing Unit with DDR6 Memory of 8GB. Keras API running on top of tensor flow in python was used for training and evaluating the proposed network. The efficiency of the reformed CapsNet is assessed by taking the Messidor dataset described in section 3.1. In the evaluation, 75% of the images are considered as a training set and 25% of the images are taken for testing. 75% of the training data is again partitioned into two parts for training and validation. 90% of images in the training dataset are used to train the network and 10% of the images are used for validation. In Messidor dataset, the fundus images are classified as four levels.

Table 5 Performance measures evaluated from the observations of the confusion matrix

	Accuracy	Precision	Recall	F1-Score
Grade 0 (Healthy retina)	97.98	97.19	97.19	97.19
Grade 1 (DR stage 1)	97.65	94.73	93.10	93.90
Grade 2 (DR stage 2)	97.65	97.59	94.18	95.85
Grade 3 (DR stage 3)	98.64	92.98	100	96.36

In Messidor dataset, among the 1200 images, 806 images are considered to train the network, 90 images are used for validation and 304 images are for testing the performance of the trained network. Among the 304 images that are considered for testing, 107 belongs to grade 0 (healthy retina), 58 belongs to grade 1 (DR stage 1), 86 are with grade 2 (DR stage 2) and 53 are with grade 3 (DR stage 3). Table 3 shows the testing results of the reformed CapsNet with the Messidor dataset in the form of a confusion matrix. 104 out of 107 grade 0 images are correctly classified by the network. 54, 81, and 53 are the correctly predicted images out of 58, 86, and 53 images of grades 1, 2 and 3, respectively.

The confusion matrix of multi-class classification problem is considered with number of rows and columns equal to the number of different classes in the dataset. The major diagonal elements indicate the correctly classified instances. Based on the confusion matrix of Table 3, the observations are done concerning TP, TN, FP, and FN for each grade 0, 1, 2 and 3. Table 4 shows the observations made from Table 3. Based on the observations made in Table 4, different performance measures like accuracy, precision, recall, and F1 score are evaluated. The performance measures for every grade are shown in Table 5. For grade 0, 104 images were classified which are true positives and 188 images which belongs to the other grades are also correctly classified. Only 6 images are incorrectly classified. Hence, for grade 0, the accuracy of the model will be 97.98% ((104 + 188)/ (104 + 188 + 6)). Similarly, for other classes also accuracy is shown Table 5. Out of 107 images of grade 0, only 104 are correctly classified which indicated the true positive rate or precision as 91.19% (104/107). Similarly, the precision is also calculated for other grades also which are shown in Table 5. Based on the preci-
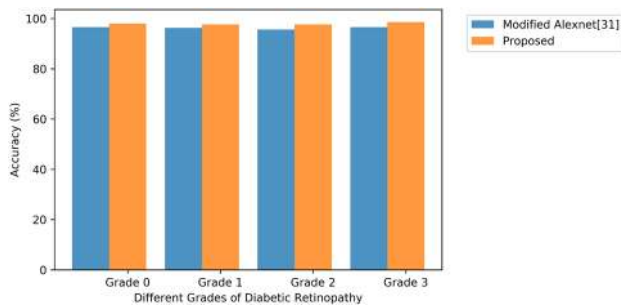
Fig. 7 Comparison of accuracy measure of proposed method with modified AlexNet
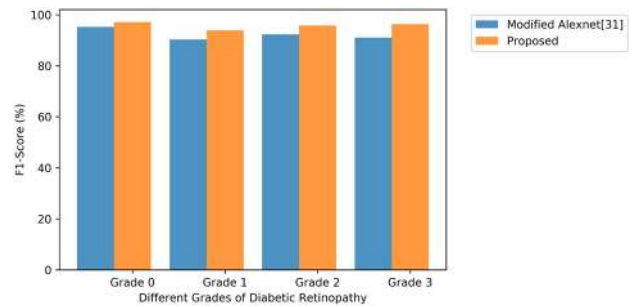

Fig. 8 Comparison of precision measure of proposed method with modified AlexNet

Fig. 9 Comparison of recall measure of proposed method with modified AlexNet

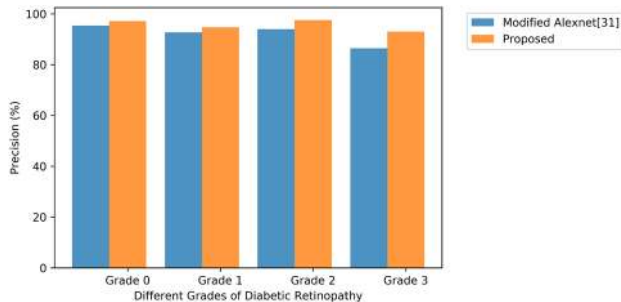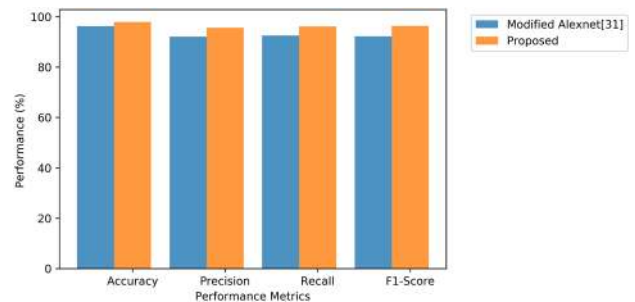**Fig. 10** Comparison of F1 score measure of proposed method with modified AlexNet

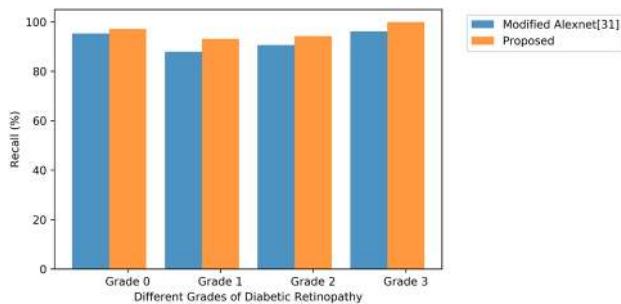**Fig. 11** Comparison of the all the performance metrics

sion and recall values F1 score is calculated in Table 5. The proposed reformed CapsNet achieves a maximum of 98.64% accuracy on Grade 3 images and 97.19% F1 score on grade 0 images. The proposed reformed CapsNet achieves 100% recall, i.e. true-positive rate on grade 3 images.

The evaluated performed measures of proposed reformed CapsNet are compared with the results of Modified AlexNet architecture introduced by Shanthi et al [31]. Figures 7, 8, 9 and 10 demonstrates the comparison of the performance measures accuracy, precision, recall, and F1-score. Figure 7 shows the comparison of accuracy. For all the four classes of the Messidor dataset, the proposed method has attained 97.98%, 97.65%, 97.65%, 98.64% accuracy whereas the existing Modified AlexNet [31] attained 96.24%, 95.59%, 96.57%, 96.57% accuracy corresponding to each

class. Figure 8 shows the comparison of precision for proposed and existing networks. The existing Modified AlexNet got 95.32%, 92.72%, 93.97% and 86.44% precision for grades 0, 1, 2 and 3 accordingly. The proposed reformed CapsNet got 97.19%, 94.73%, 93.97% and 86.44% precision for the grades 0, 1, 2 and 3 respectively. Figure 9 shows the comparison of recall for proposed and existing networks. The existing Modified AlexNet got 95.32%, 87.93%, 90.69% and 96.22% recall for grades 0, 1, 2 and 3 accordingly. Reformed CapsNet got 97.19%, 93.1%, 94.18% and 100% recall for the grades 0, 1, 2 and 3, respectively. Figure 10 shows the comparison of F1 score for proposed and existing networks. The existing Modified AlexNet got 95.32%, 90.26%, 92.30% and 91.06% recall for grades 0, 1,2 and 3 accordingly. Reformed CapsNet got 97.19%, 93.9%, 95.85% and 96.36% recall for the grades 0, 1, 2 and 3, respectively.

The performance of the proposed reformed CapsNet is calculated by averaging the four grades measures. The reformed CapsNet finally achieves 97.98% accuracy, 95.62% precision, 96.11% recall, and 96.36% F1 score. The comparison of the all the four performance measures for the model is done with Modified AlexNet [31]and is shown in Fig. 11. For all the four performance measures by Concerning all the four classes, the proposed network achieves good results when compared to the existing network suggested by Shanthi et al. [31].

Conclusion

One of the main challenges for the people who are suffering from diabetes is diabetic retinopathy. Due to this, the blood vessels in the retina get damaged which leads to consequences related to the vision and in some cases vision loss also. Diabetic retinopathy can be avoided with early detection. The traditional methods available for this purpose are taking more time and the prediction accuracy is low, i.e. it cannot be detectable in the early stages. If the problem is identified in advanced stages, a chance of efficient treatment for recovery is low. Hence, early detection with high accuracy plays a major role in diabetic retinopathy. Deep learning is one technology that is efficient results in the health domain. CNN is one of the well-known networks for image analysis. Capsule networks are designed to avoid the problem in CNN with pooling layers. Hence, in this paper, we constructed a capsule network for diabetic retinopathy identification with the convolution layer, the primary capsule layer, and class capsule layer. The first two layers are used for feature extraction and class capsule layer for identifying the probability of a particular class. The formed CapsNet identifies the problem accurately in all four stages. The performance of the proposed CapsNet is compared with one of the existing methods called modified AlexNet which belongs to the CNN approach. When compared to that method, the proposed CapsNet achieves good results for early detection of retinopathy problems in diabetic patients. The proposed model got 97.98% accuracy in identifying the problem. The proposed capsule network is trained with only Messidor dataset in which diabetic retinopathy is identified in four stage grades 0 to 3. There are other datasets also in which the classification of the diabetic retinopathy is done in five stages. For that datasets, the network has not trained. The future extension of this article was to train the proposed capsule network for all the possible classes of the diabetic retinopathy. Hence, the conclusion is with CapsNet the retinal problems in diabetic patients can be detected early for better diagnosis to avoid vision loss.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Lin X, Yufeng X, Pan X, Jingya X, Ding Y, Sun Xue, Song Xiaoxiao, Ren Yuezhong, Shan Peng-Fei (2020) Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. Sci Rep 10(1):1–11
2. Tandon N, Anjana RM, Mohan V, Kaur T, Afshin A, Ong K, Mukhopadhyay S, Thomas N, Bhatia E, Krishnan A, Mathur P (2018) The increasing burden of diabetes and variations among the states of India: the Global Burden of Disease Study 1990–2016. Lancet Glob Health 6(12):1352–1362
3. Atre S (2015) Addressing policy needs for prevention and control of type 2 diabetes in India. Perspect Public Health 135(5):257–263. https://doi.org/10.1177/1757913914565197 (ISSN 1757-9147, PMID 25585513)
4. Atre S (2019) The burden of diabetes in India. Lancet Glob Health 7(4):e418. https://doi.org/10.1016/S2214-109X(18)30556-4 (ISSN 2214-109X)
5. National Programme for Prevention and Control of Cancer, Diabetes,Cardiovascular Diseases and Stroke. Directorate General Of Health Services. Retrieved 29 Apr 2020
6. Sharma NC (2019) Government survey found 11.8% prevalence of diabetes in India. Livemint. Retrieved 29 Apr 2020
7. Diabetes in India. Cadi Research. Retrieved 29 Apr 2020
8. Mo W, Xiaoshu L, Yexiu Z, Wenjie J (2019) Image recognition using convolutional neural network combined with ensemble learning algorithm. J Phys Conf Ser 1237(2):022026
9. Samanta A, AheliSaha SCS, Steven LF, Yo-Dong Z (2020) Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset. Pattern Recognit Lett 2020(04):026. https://doi.org/10.1016/j.patrec
10. Shiva SR, NilambarSethi RR, Gadiraju M (2020) Extensive analysis of machine learning algorithms to early detection of diabetic retinopathy. Mater Today Proc. https://doi.org/10.1016/j.matpr.2020.10.894 (ISSN 2214-7853)
11. Gaurav S, Dhirendra KV, Amit P, Alpana R, Anil R (2020) Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public datasets. Intell Based Med 3–4:100022. https://doi.org/10.1016/j.ibmed.2020.100022 (ISSN 2666-5212)
12. Janakiramaiah B, Kalyani G, Jayalakshmi A (2020) Automatic alert generation in a surveillance systems for smart city environment using deep learning algorithm. Evol Intell. https://doi.org/10.1007/s12065-020-00353-4
13. Sajid S, Saddam H, Amna S (2019) Brain tumor detection and segmentation in MR images using deep learning. Arab J Sci Eng 44(11):9249–9261
14. Alyoubi WL, Shalash WM, Abulkhair MF (2020) Diabetic retinopathy detection through deep learning techniques: a review. Inf Med Unlocked 20:00377. https://doi.org/10.1016/j.imu.2020.100377 (ISSN 2352–9148)
15. Janakiramaiah B, Kalyani G (2020) Dementia detection using the deep convolution neural network method. Trends Deep Learn Method Algor Appl Syst 157
16. Mohsin Butt M, Ghazanfar L, Awang Iskandar DNF, Jaafar A, Adil HK (2019) Multi-channel convolutions neural network based diabetic retinopathy detection from fundus images. Procedia Comput Sci 163:283–291. https://doi.org/10.1016/j.procs.2019.12.110 (ISSN 1877-0509)

17. Islam MM, Yang H-C, Poly TN, Jian W-S, Li Y-C (2020) Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: a systematic review and meta-analysis. Comput Methods Prog Biomed 191:105320. https://doi.org/10.1016/j.cmpb.2020.105320 (ISSN 0169–2607)

18. Hamid S, Sare S, Ali H-M, Hamid A (2018) Early detection of diabetic retinopathy. Surv Ophthalmol 63(5):601–608. https://doi.org/10.1016/j.survophthal.2018.04.003 (ISSN 0039-6257)

19. Ahmad A, Mansoor AB, Mumtaz R, Khan M, Mirza SH (2015) Image processing and classification in diabetic retinopathy: a review. In: Proceedings of European workshop visual and information process

20. Shahin EM, Taha TE, Al-Nuaimy W, El Rabaie S, Zahran OF, El-Samie FEA (2013) Automated detection of diabetic retinopathy in blurred digital fundus images. In: Proceedings of 8th international computer engineering conference

21. Shailesh K, Abhinav A, Basant K, Amit KS (2020) An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation. Opt Laser Technol 121:66–70

22. Jaafar HF, Nandi AK, Al-Nuaimy W (2011) Automated detection and grading of hard exudates from retinal fundus images. In: Proceedings of 19th European signal processing conference

23. Casanova R, Saldana S, Chew EY, Danis RP, Greven CM, Ambrosius WT (2014) Application of random forests methods to diabetic retinopathy classification analyses. PLoS One 9(6):178–193

24. Quellec G, Charrière K, Boudi Y, Cochener B, Lamard M (2017) Deep image mining for diabetic retinopathy screening. Med Image Anal 39:178–193

25. Kaggle (2015) Diabetic retinopathy detection. https://www.kaggle.com/c/diabeticretinopathy-detection/. Accessed 7 May 2018

26. Gulshan V et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22):2402–2410

27. Gargeya R, Leng T (2017) Automated identification of diabetic retinopathy using deep learning. Ophthalmology 124(7):962–969

28. Gräsbeck TC, Gräsbeck SV, Miettinen PJ, Summanen PA (2016) Fundus photography as a screening method for diabetic retinopathy in children with type 1 diabetes: outcome of the initial photography. Am J Ophthalmol 169:227–234

29. Garima Gupta S (2017) Kulasekaran, Keerthi Ram, Niranjan Joshi, Rashmin Gandhi, Local characterization of neovascularization and identification of proliferative diabetic retinopathy in retinal fundus images. Comput Med Imaging Graphics 55:124–132

30. Singh N, Tripathi RC (2010) Automated early detection of diabetic retinopathy using image analysis techniques. Int J Comput 8:18–23

31. Qureshi TA, Habib M, Hunter A, Al-Diri B (2013) A manually-labeled, artery/vein classified benchmark for the DRIVE dataset. Proc IEEE Symp Comput Based Med Syst. https://doi.org/10.1109/CBMS.2013.6627847

32. Kauppi T, Kalesnykiene V, Kamarainen J-K, Lensu L, Sorri I, Raninen A, Voutilainen R, Uusitalo H, Kälviäinen H, Pietilä J (2007) DIARETDB1 diabetic retinopathy database and evaluation protocol. In: Proceedings of medical image understanding and analysis (MIUA). 10.5244/C.21.15

33. Bhatia K, Arora S, Tomar R (2016) Diagnosis of diabetic retinopathy using machine learning classification algorithm. In: 2016 2nd international conference on next generation computing technologies (NGCT)

34. Maher RS, Kayte SN, Meldhe ST, Dhopeshwarkar M (2015) Automated diagnosis non-proliferative diabetic retinopathy in fundus images using support vector machine. Int J Comput Appl 125(15):7–10. https://doi.org/10.5120/ijca2015905968

35. Leontidis G, Al-Diri B, Hunter A (2017) A new unified framework for the early detection of the progression to diabetic retinopathy from fundus images. Comput Biol Med 90:98–115

36. Gegundez-Arias ME, Diego M, Beatriz P, Fatima A, Jose MB (2017) A tool for automated diabetic retinopathy pre-screening based on retinal image computer analysis. Comput Biol Med 88:100–109

37. Cunha-Vaz BSPJG (2002) Measurement and mapping of retinal leakage and retinal thickness—surrogate outcomes for the initial stages of diabetic retinopathy. Curr Med Chem Immunol Endocr Metab Agents 2:91–108

38. Anandakumar H, Umamaheswari K (2017) Supervised machine learning techniques in cognitive radio networks during cooperative spectrum handovers. Clust Comput 20:1–11. https://doi.org/10.1007/s10586-017-0798-3

39. Omar M, KhelifiF, Tahir MA (2016) Detection and classification of retinal fundus images exudates using region based multiscale LBP texture approach. In: 2016 international conference on control, decision and information technologies (CoDIT)

40. Welikala RA, Fraz MM, Williamson TH, Barman SA (2015) The automated detection of proliferative diabetic retinopathy using dual ensemble classification. Int J Diagn Imaging. https://doi.org/10.5430/ijdi.v2n2p72

41. Goldbaum MDM (1975) STARE dataset website. Clemson University, Clemson. http://www.ces.clemson.edu

42. Haldorai A, Ramu A, Chow C-O (2019) Editorial: big data innovation for sustainable cognitive computing. Mob Netw Appl 24:221–223

43. Purandare M, Noronha K (2016) Hybrid system for automatic classification of Diabetic Retinopathy using fundus images. In: 2016 online International conference on green engineering and technologies (IC-GET)

44. Bhatkar AP, Kharat GU (2015) Detection of diabetic retinopathy in retinal images using MLP classifier. In: 2015 IEEE international symposium on nanoelectronic and information systems

45. Shanthi T, Sabeenian RS (2019) Modified Alexnet architecture for classification of diabetic retinopathy images. Comput Electric Eng 76:56–64

46. Wan S, Liang Y, Zhang Y (2018) Deep convolutional neural networks for diabetic retinopathy detection by image classification. Comput Electric Eng 72:274–282

47. Partovi M, Rasta SH, Javadzadeh A (2016) Automatic detection of retinal exudates in fundus images of diabetic retinopathy patients. J Anal Res Clin Med 4(2):104–109. https://doi.org/10.15171/jarcm.2016.017

48. Zago GT, Andreão RV, Dorizzi B, Teatini Salles EO (2020) Diabetic retinopathy detection using red lesion localization and convolutional neural networks. Comput Biol Med 116:103537

49. Decenciére E, Zhang X, Cazuguel G, Lay B, Trone C, Gain P, Ordonez R, Massin P, Erginay A, Charton B, Klein J-C (2014) Feedback on a publicly distributed database: the Messidor database. Image Anal Stereol 33(3):231–234

50. Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. Adv Neural Inf Process Syst

51. Hinton GE, Alex K, Sida DW (2011) Transforming auto-encoders. In: International conference on artificial neural networks. Springer, Berlin

52. Hinton G, Sabour S, Frosst N (2018) Matrix capsules with EM routing. In: 6th international conference on learning representations, ICLR 2018—conference track proceedings

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.