

A Hybrid Ensemble Feature Selection-based Segmentation and Deep Majority Voting Framework on Large Multi-class Diabetes Retinopathy Databases

Dr. Shaik Akbar ^a, Dr. Divya Midhunchakkaravarthy ^b

^a 1Department of Computer Science & Engineering, PDF Scholar, Lincoln University College, Malaysia

^b Professor, Department of Computer Science & Engineering, Head - Academic and Student Affairs, Centre of Post Graduate Studies, Lincoln University College, Malaysia

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 23 May 2021

Abstract: Diabetic retinopathy is a micro vascular disease that induces a number of changes in the retina. Micro aneurysms, haemorrhage exudates, and the development of new blood vessels all alter the diameter of the blood vessel. Most of the conventional multi-class diabetes retinopathy has different issues such as problem of over-segmentation, classification precision, recall and error rate on high dimensional features space. Ensemble feature selection measures are used to filter the essential features in the large feature space. In this work, a hybrid ensemble feature selection based multiple classification models are used to improve the classification accuracy on multi-class diabetes retinopathy databases. In this work, a novel image segmentation, ensemble feature extraction measures, and multiple classification approaches are used to find the majority voting in the classification problem. Experimental results show that the proposed ensemble feature extraction-based voting classification model has better efficiency compared to the state of art of conventional approaches.

Keywords: Ensemble feature selection, deep learning, diabetes retinopathy

1. Introduction

Diabetes is a relatively common disease, but it can be made much simpler by using modern technology to separate the needed insulin from the available insulin. The victims of the crisis have had to endure high levels of sugar in their blood for long periods of time. In the event of the body not being able to meet the daily insulin requirements, the individual will suffer from diabetes. It is important to keep your eyes open in order to take care of other important things in the body (**S. Stolte and R. Fang 2020**). The retina is the part of the eye that contains the photosensitive light-sensitive layer. Another important job of the retina is to convert light rays into electrical signals. These images are being constructed. Diabetic retinopathies assist in the recovery of retinal tissue. It also reduces the amount of fluid in the eye is able to show, which reveals the hidden vision. Another of the many risk factors for this disease is age. When a disease is discovered in the early stage, there is a good chance of being able to cure it. Diabetic retinopathy is equally distributed between the eyes (**Q. Xie Sep. 2020**). Cholesterol consumption and heavy smoking are major risk factors for this illness. Eventually, the entire veins in the eye are affected by this disease. In order to create new blood cells, the eyesight must be good. Almost all of the blood cells are in a resting state during this time. It is also known as diabetic retinopathy. Diabetic retinopathy shows up in the eyes as retinopathy. The best balance of sensitivity and specificity of all possible percentages, currently, was achieved here. Diabetic retinopathy is one of the many diabetes-related eye diseases, no fundus or retina images are used to identify the DR disease (**Y. Xie May 2020**). To predict the DR, the majority of the ophthalmologist's uses only rely on standard techniques. Automatic disease prediction is rather cost-effective, but the population in rural areas must receive the benefits. Sensitivity and specificity are also calculated at each level of the Random Forest classifier and neural network level. Diabetic retinopathy was studied extensively in relation to vision loss among older patients with diabetes. This condition impairs the retina's ability to perform its normal function. Retinal is the tissue that is extremely sensitive to light. Prediction and early detection were employed to avert the loss of the eyesight (**A. Samanta, A. Saha, S. C. Satapathy, S. L. Fernandes and Y.-D. Zhang Jul. 2020**). Most recent reports state that people all over the world have to contend with the new disease called 'diabetes.' In this study, artificial intelligence was used to carry out the disease diagnosis. Using the pre-processing concept on the input images suppresses noise. Using SVM, Navie Bayes, and LR Bayes are employed to classify the images before resizing. The result shows improved sensitivity, increased accuracy, and a decreased error rate. Pre-processing and segmentation were used to discover DR in the human eye. Initially, the optic nerve is made up of longitudinally segmented. The network was created using retinal micrographic imaging. If the retinal nerve field was small, the person would have either non-proliferative diabetic retinopathy or proliferative retinopathy (DR) (**W. Zhang Jul. 2019**). Suppose there was no such connection between the portion of the nervous system and this disease. The funduscopic images are used to diagnose the disease and to assess the extent of the problem. Detection of the disease from a non-processed image is challenging. A number of speakers volunteered their clinical and research data related to diabetic retinopathy, or Priya et al., who shared information about predicting retinopathy offered clinical and research data (DR). NPDR may be able to more accurately classify DLT: The study's authors may be able to divide NPDR into separate classifications for the sake of NPD: Non-proliferative

Diabetic Retinopathy and Proliferative Diabetic Retinopathy (PDR). PNN, Bayesian Classification, and Support Vector machine (SVM) were used to assess the probability of the disease. The above models are assessed to see how they perform against each other. After different classifications are done, traditional methods are applied and the classifications are studied (Y. Katada, N. Ozawa, K. Masayoshi, Y. Ofuji, K. Tsubota, and T. Kurihara Dec. 2020). These results show that multi-label architecture is employed for identifying retinopathy caused by diabetes. In this study, various methods such as the Local Ternary Pattern (LTP) and Local Energy-based Histogram (LESH) are used to separate characteristics are identified and presented. Both CNN and deep learning were combined to create the predictive model as shown by the applied mathematical model; it has also been used to correctly identify other conditions like kidney failure, cardiovascular disease, and cerebrovascular disease. Normal and abnormally marked images are handled using a set of classifiers (S. Wan, Y. Liang, and Y. Zhang Nov. 2018). A number of different studies have been done to identify potential lesions in retinal fundus images.

2. Related Works

Extracted feature analysis gives common statistical measures, including the mean, maximum, and minimum, such as statistical analysis. feature vectors were supplied to SVM classifier to produce the classification result. The process begins with data pre-preparation, continues through segmentation, and ends with data classification. Segmentation and dark spots were applied with the help of the H- transformation and thresholding technique. For overall system performance, they evaluate on low resolution images and their system obtains 84.3% and 93.3% on sensitivity and specificity, but misses only 5.4% of Mas (G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard Jul. 2017). CAD allows determining the number of disease present in the eye, as well as separating the images of the diseased and healthy retinae. This will assist in increasing the capacity of the ophthalmologist's workload as well as screening the retina images. To improve the system's performance and accuracy, the proposed system is in use Preprocessing, segmentation, features, classification is part of the approach (N. Y. Gharaibeh Dec. 2020). To identify all regions in the retina image, we must scan the image multiple times and extract every possible candidate. The four main operations of feature types are shape, color, gray level, features, and texture, but there are other options for feature generation including region, which is nice for generating different kinds of features. For regions are selected and classified using support vector machines (SVM), kernel neural networks (kNN) and Support Vector Machines (SVM). In training mode, all normal and all suspicious image features are active. These four types of features are used to evaluate the classifier's performance. Total image was deconstructed using a local adaptive thresholding technique, and each region was handled by a different sub-sub region was dependent on artifact level. The workaround to this issue was a low pass with a Gaussian filter. A semi-glossy finish was produced by use of lighting effects was applied to the wall behind the object. The computer-aided method of classifying fundus images (T. Nazir, A. Irtaza, Z. Shabbir, A. Javed, U. Akram, and M. T. Mahmood Aug. 2019). The pre-process method comprises of pruning the images, followed by blood vessel segmentation and blood matching. Second, the local contrast technique was used to reduce the non-uniformity of the retinal image. Furthermore, the optic disc and fovea were determined to have a large local variation in the retina image and matching found to be correlated (G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. Teatini Salles Jan.2020). The arterioles were split using multilayer neural network segmentation. Using a recursive region growth algorithm, the exudates were found. Many candidate features were based on the candidate's color and shape. The multilayer perceptron (MLP), radial basis function (RBF), and majority voting schema were used in these algorithms. These characteristics were fed into the classifier to distinguish the two regions, which distinguished the normal from the abnormal (M. M. Abdelsalam Jan.2020). In order to evaluate the method, the model was tested on a series of 115 training and test image pairs. These tests gave them a 100% sensitivity, 56% of which was correct, and 83% was precise. Using a relative entropy-based threshold, the red lesions were easily distinguished from the rest of the image. It was also restricted by the use of the morphological top hat technique. Apply non-uniform contrast and color tone processing to the retina image (P. Cao, F. Ren, C. Wan, J. Yang, and O. Zaiane Nov.2018). Since there are so many holes in the retina in mathematic terms, the closure operator was utilized to bridge the gaps. The basic idea of the black hot-top contrast normalization was extracting the black against the retina. The CLAHE contrast enhancement algorithm was used to enhance the retina image. To allow the different forms of hemorrhage to be isolated, the normalized cross-correlation technique was used. They used an adaptive growth segmentation (AGS) algorithm to produce the exact shape of the hemorrhage. To obtain the hemorrhagic retina image, the visual mathematical tool of compactness, aspect ratio, and area ratio was employed. The conclusion of this approach indicates that 85% of the answers were found. There is a detailed list of these methods of checking for the possibility of hemorrhage. Morphological approach, neural network, and region-growing methods have been developed for spotting internal bleeding. With random filters and masks, the optic nerve was located. The tap-hanging filter and averaging technique was used to remove the noise and background in the retinal image. The image has been divided into separate segments and applied a random transform (RT) is applied to each one. Using a Circular Hough Transform (CHT) and a blood vessel saw, the disc was cut into segments and the pith was removed with a matched filter. To ensure that the reddish region

was preserved, the top-hat transform was applied on the image after the normalization. In order to distinguish the blood vessel MA from other red areas, the binarization of the blood vessel was required. Using the hit-or-or-miss technique, the blood vessel was extracted from the binary image (M. S. Ayhan, L. Kühlewein, G. Aliyeva, W. Inhoffen, F. Ziemssen, and P. Berens Aug.2020). At the start of training, a discriminative model was used to segment the vessels, whereas at the end of the validation, a mixture model was used to produce the vessels map. In the first step, contrast and brightness of the retina images are extracted. Secondly, an adaptive thresholding segmentation is used to isolate exudates from the retina image. A third step involves classifying the candidate exudates. These models used the logarithmic regression method to generate relevant features. After this, the system was done, the radial basis function network (RBF) was applied to identify exudates in the images. To reduce noise in the green channel, the median filter was applied to the image (P. Porwal Jan.2020). By means of a top-hat operation, the image contrast was enhanced. Subtraction produced a result which opened and closed the top hat. Using the Hough transform, which measures optical characteristics, will locate circular objects in the digital images, while edge-enhanced images will locate the candidate pixels. By dilation and abrasion, the artery was isolated. Finally, the exudates were found in an iterative manner by iterative process. In this technique, the green component of the fundus image is extracted and stored in memory. The averaging filter was used to remove the noise from the green channel. The average image was sent to Fuzzy C Means (FC) to separate the original image into three groups: normal retina, the more significant to lower, low concentration, and most significant. The two most important classes that get left out are participants in extraction (which accounts for the vast majority of all finance managers' company employees) and participants in manufacturing. For estimation of the background and removal of noise, the calculation of the mean and variance of each of the red, green, and blue channels was utilized. Preprocessing was performed on the picture using the morphological closing operator to remove the blood vessel from the retina (E. Saleh April.2018). Complementary contrast was used to normalize the retina image contrast. To improve the contrast in an image, the Gabor filter was used. Using the adaptive thresholding technique, an image was combined with a Gabor filter to generate a binary map. Gaussian Mixture model was applied to extract statistical features from the candidates and assign them to either wet or dry lesions. CLAHE's method produced increased the contrast of the retina' Improve the bright lesions by Gaussian or G-invariance kernel The final modification was applied to the Gabor image in order to detect exudates candidate region was adaptive threshold. Statistical features of the green channel and HSV color space were used Gaussian mixture models and support vector machine learning was employed to track down egress and ingress of bodily fluids. The Binary Contrast Enhancement map was employed in macular detection with low adaptive threshold efficient features was extracted from the green channel intensity, macular region area, and binary feature intensity map. There was a linear appearance of blood vessels in the retina scans. The limits of the short vessels are situated at the same horizontal location in the retinal image. Information on blood vessels that can be used to make a positive diagnosis is gained from a region-based Bayesian detection algorithm. The analysis was conducted using cross section and confidential information. The fundus imaging is considered a more promising alternative for noninvasive eye screening due to the eyes' hyper vascularity. These systems rely on precisely on the fundus imaging, and particularly on consistent and accurate algorithms for analyzing the image (N. Sambyal, P. Saini, R. Syal, and V. Gupta July.2020). To address this problem, they present a method they call MML-TH (maximum of multiple linear top hats). The second method used a threshold-based segmentation algorithm to identify local vessels. Lastly, the SVM was used to differentiate between blood blots. In the morphological method, they have utilized "Ball" shaped element to cut out the blood vessel from candidates to find hemorrhages. Another research group proposed the use of mathematical and artificial intelligence algorithms for finding vascular images in fundus images (Q.S.You Sep.2020). Like biological neurons, which contain adjustable tuning factors, neural networks are similar to one another, are made up of adaptive weights, and are capable of approximating nonlinear functions of their inputs. This technique used two-step pre- and post-processing as well as well as feature extraction and classification on the idea that. Before the processing stage, a binarized retinal image is obtained. After that, the large vessel extraction is done, the image is normalized. We estimated the background of the retina using the median filter. Our method utilized the normalized image to binarize the retinal image. As a series of transients, the photoreceptor pieces were cut with a wavelet and curve processing technique. To get rid of these problems and find suitable candidate lesions, an additional step is required before applying the processing must be done. This process is called both desaturation, and contrast enhancement, and emphasizes both channel selection and noise reduction. The color fundus image's red, green, and blue components were initially found to have higher contrast against the backdrop of the red blood vessels and retina, which was the darker one (S. R. Sadda Oct.2020).

3. Proposed Model

Figure 1 illustrates the technique of image thresholding, which can serve both segmentation and identification functions in a key part of the image. A hybrid image segmentation approach for separating the background and foreground is developed using the background and foreground regions. In this project, a mixture of supervised and unsupervised classification methods is used to partition a given image into an equi-dimensional space to achieve a

similar result to the kernel density estimate. Reaction time is an important in DR diagnostic systems. Threshold-based segmentation is also plays a key role in characterizing other characteristics. Without segmentation, a reference structure cannot be created, and all images must be dealt with unreferenced. Many people were working on finding ways to localize optical disks, all of which met with some success. Most of the techniques involving position detection did not solve the original disco focus alignment problem. The binary image was created by channelling all of the brightness, from the lighter to the darker areas. Deep learning-based model provides ways to discover and identify the disease trend in large numbers of training samples. Implementing the model as a feature segmentation deep learning Bayesian filter with two levels of severity.

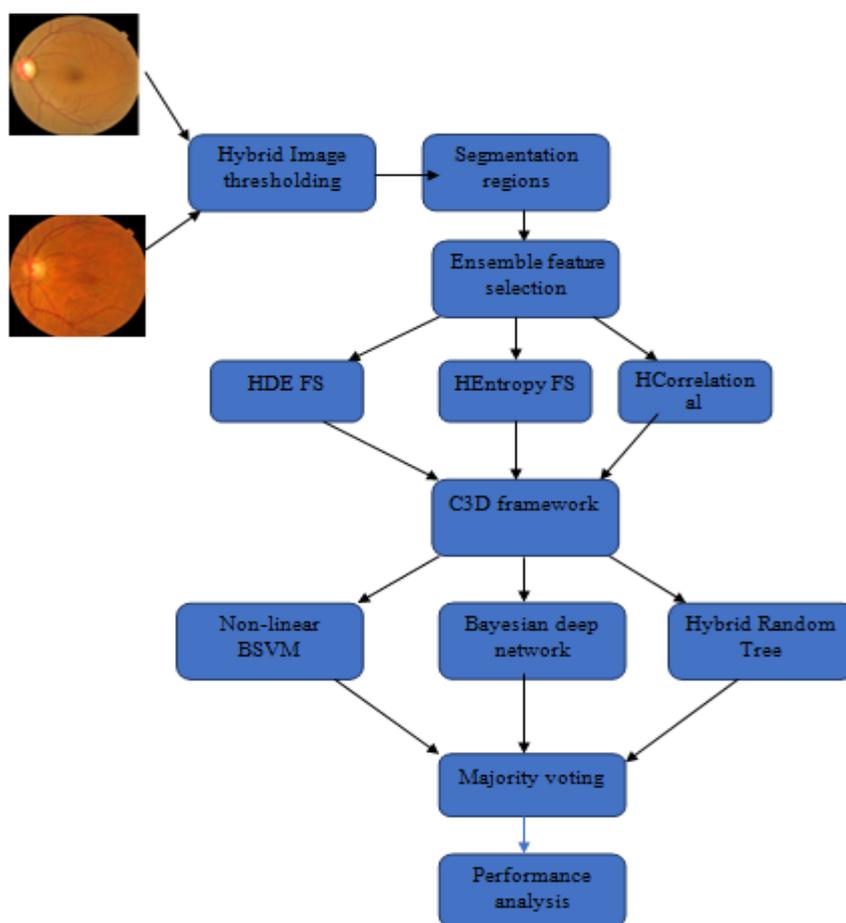


Figure 1. Proposed Ensemble feature selection based Multi-class ensemble classification learning

In the first step, the fundus and clinical parameters are used to detect the disease form and statistical assessment of severity. With this technique, the noise is eliminated, while simultaneously the spots and red lesions are smoothed to allow a simple identification of cellular structures to be made. The second step calls for applying a deep CNN involves searching for important features from the low-passed data. In the early phases of deep learning, pre-trained models are used to identify the salient components of each training picture. The final classification process also sees the implementation of a new classifier that is designed to identify the diseases. In this instance, a mix of non-linear and Bayesian support vector machines are used to aid in the classification of disease. In the previous work, we have developed non-linear SVM and hybrid bayesian model for classification process.

1. Hybrid Gaussian Thresholding based feature segmentation

- (1) Read input dataset with disease and non-disease images.
- (2) Apply multi-level thresholding images using the computed threshold limits.
- (3) Compute the inner variance of the image by using the following equations.

The initial thresholding limit of the disease is given as is

$$G_{\min}(p, q) = (\log(\eta) / \lambda) \cdot \min \left\{ \frac{e^{-\sqrt{p^2+q^2}/2 \cdot \sigma_p^2}}{\sigma_p \sqrt{2 \cdot \pi}}, \frac{e^{-\sqrt{p^2+q^2}/2 \cdot \sigma_q^2}}{\sigma_q \sqrt{2 \cdot \pi}} \right\}$$

$$G_{\max}(p, q) = (\log(\eta) / \lambda) \cdot \max \left\{ \frac{e^{-\sqrt{p^2+q^2}/2 \cdot \sigma_p^2}}{\sigma_p \sqrt{2 \cdot \pi}}, \frac{e^{-\sqrt{p^2+q^2}/2 \cdot \sigma_q^2}}{\sigma_q \sqrt{2 \cdot \pi}} \right\}$$

$$T_1 = \prod_{i=1}^n \min \{ p \cdot \Pr((p, q) / c_1) / q \cdot \text{prob}((p, q) / c_1) \dots, p \cdot \Pr((p, q) / c_n) / q \cdot \text{prob}((p, q) / c_n) \} / |N|$$

$$T_2 = \prod_{i=1}^n \max \{ p \cdot \Pr((p, q) / c_1) / q \cdot \text{prob}((p, q) / c_1) \dots, p \cdot \Pr((p, q) / c_n) / q \cdot \text{prob}((p, q) / c_n) \} / |N|$$

- (4) Computing the Gaussian kernel estimation using the Gaussian distribution as

$$G(p, q) = (\max\{\lambda, \eta\}) \cdot e^{\left\{ -\left(\frac{p_1^2}{\sigma_p^2} + \frac{q_1^2}{\sigma_q^2} \right) \right\}} \log \left(\frac{2\pi p_1}{\lambda} \right)$$

where $p_1 = q_1 = p \cos \phi + q \sin \phi$ and $\sigma_p = \sigma_q = \sigma$

$$GD(p, q) = (\max\{\lambda, \eta\}) \exp \left(\frac{2p_1^2}{\sigma^2} \right) \cos \left(\max\{T_1 \cdot \sigma_p, T_2 \cdot \sigma_q\} \cdot \left(\frac{2\pi q_1}{\lambda} \right) \right)$$

$$FS(GD(p, q), C_m) = \text{Max} \left\{ \frac{\text{Pr ob}(GD(p, q) / C_m)}{|GD(p, q)| \cdot \text{Pr ob}(C_m)} \cdot (\max\{\lambda, \eta\}) \right\};, m = 1, 2 \dots |G|$$

Each image is classified into a variety of partitions to find the nearest features. The means and covariance matrices of the blocks from B_i to B_j represent the block partitions of the i^{th} and j^{th} blocks in frame. In this step, a block is used to find and identify disease patterns in each image, and each of its adjoining blocks is used to mark their locations.

2. Ensemble feature selection

1. Hybrid Distribution Entropy Feature selection measure (HDEFS)

In the hybrid distribution entropy based feature selection measure, a nonlinear probability estimation is performed on the block by block processing on the input image. The hybrid distribution entropy measure is given as

$$HDE(B, p[i]) = - \frac{p[i]}{\log(\sum B)} \sum p[i] \cdot \log(p[i])$$

2. Hybrid Correlation based feature selection measure (HCFS)

Proposed hybrid correlation based feature selection measure is used to find the mutual occurrence of block of pixels in the given input image. Here, the occurrence of the block in the image represents the relationship of the mutual blocks in the given image for disease prediction process.

$$HCFS(B(i, j)) = \frac{(B_i - \mu_{B_i}) \cdot (B_j - \mu_{B_j})}{\sigma(B(i, j))} \text{Prob}(B(i, j) / C_m)$$

3. Hybrid Entropy Feature selection measure (HEFS)

Hybrid Entropy Feature selection measure is used to find the essential key information among the block of pixel values. Here, the hybrid entropy measure is used to find the key pixel wise relational entropy information in each block.

$$HE(B(i, j)) = \max \left\{ \frac{(B_i - \mu_{B_i}) \cdot (B_j - \mu_{B_j})}{\sigma(B(i, j))}, -\frac{p[i]}{\log(\sum B)} \sum p[i] \cdot \log(p[i]) \right\}$$

4. Hybrid Bayesian deep networks

In the proposed Bayesian deep networks, a novel Gaussian estimator is used to find the correlation between the regions in the classification process. The proposed cosine metric used in the Gaussian estimation model is given below.

1. To each feature in the segmented regions B(i)
2. Do
3. Perform the Gaussian estimator measure on the segmented regions B(i) as
4. $G(B(i), B(j), \sigma) = \frac{1}{2\pi\sigma_B^2} \exp\left(-\frac{\cos(i, j)}{2\sigma_B^2}\right) \cdot \min\{B(i), B(j)\}$
5. The angular similarity of the segmented regions is given as
6. $ADG = \tan^{-1}\left(\frac{B_j}{B_i}\right) \in (0, 2\pi]$
- Apply normalized kernel estimator on the gaussian estimator as
7. $NK(B) := \frac{(G(B(i), B(j), \sigma) - ADG(B(i), B(j)))}{\sigma_{B(i), B(j)}}$
8. End

5. Hybrid Random Tree Approach

In this hybrid random tree approach, a novel feature selection measure is designed and implemented on the features space in order to improve the

For each randomized samples S_i

Do

$$\text{GaussianKernel} : GK(\phi, \theta) = e^{-\theta^2} / (2 * \log(\phi))$$

$$\psi = gkv = GK(\sum HB_f, 1 / 2 \sum B_f);$$

$$\text{ExponentialGaussian Probability} = KP(D) = |HB_f / (\sum \log(\psi) * HB_f)|$$

$$\text{GaussianEntropy} : GE(d_i) = -GK(\sum d_i \cdot \exp(d_i), \sigma_d)$$

Done

In the above equations, the Gaussian entropy is used to check the feature entropy value based on the Gaussian estimator.

4. Experimental Results

Using the Python and Java programming, training and testing are simulated in the cloud environment. Every image dataset is used to locate the multi-class prediction process. A novel approach IS used in this work to remove the over classification process. We use multiple Amazon AWS GPU instances to approximate the image result dataset results in our design process.

$$Sen = \frac{TP}{(TP+FN)}$$

$$Spec = \frac{TN}{(TN+FP)}$$

$$PPV = \frac{TP}{(TP+FP)}$$

$$Acc = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

Where TP: True positive DR rate, FP: false positive of DR rate, TN: True negative of DR rate, FN: false negative of DR rate.

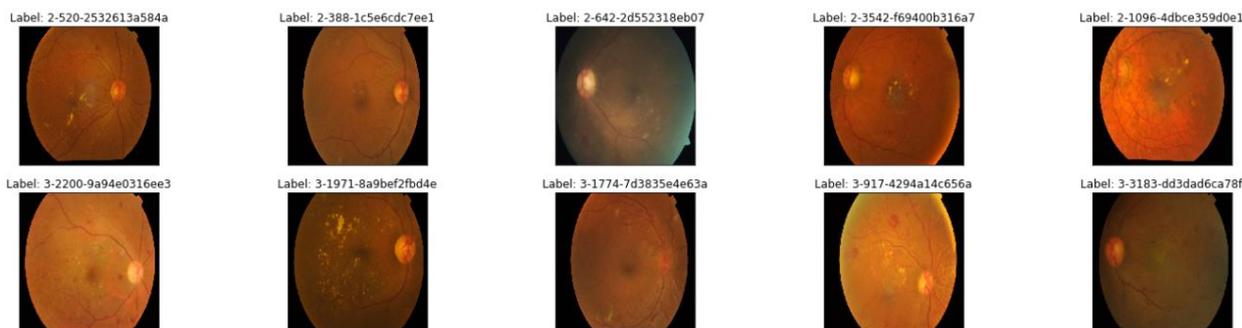


Figure 2. Multi-class image training samples

Figure 2, represents the training data of the multi-class DR images. These images are used to filter the essential block of regions for segmentation and classification process.

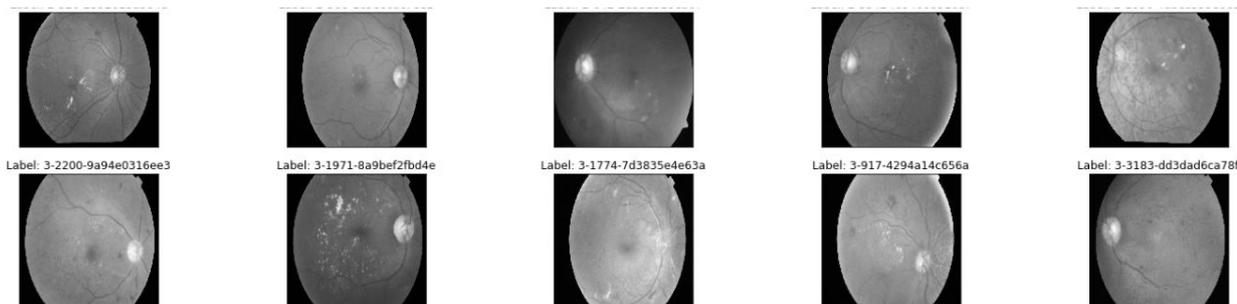


Figure 3. Image thresholding of input training image

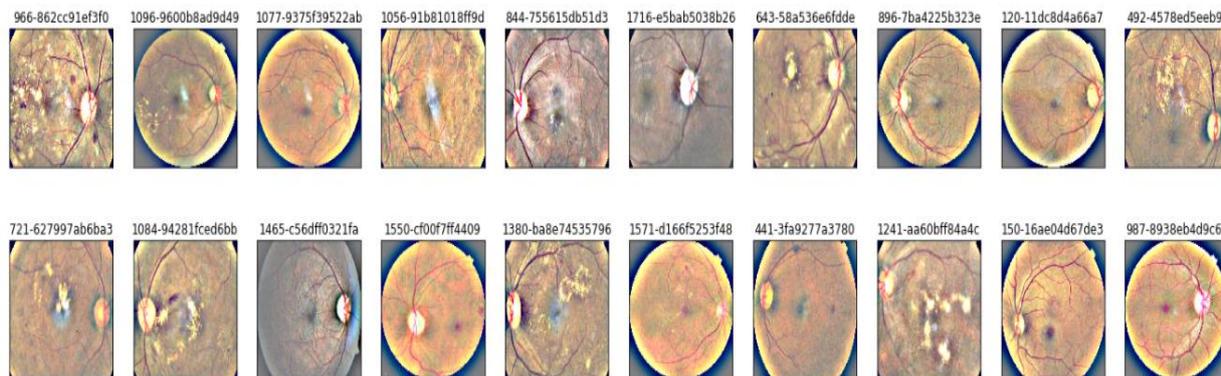


Figure 4. Image segmentation on the filtered DR images

Table 1. Comparative results of present ensemble feature selection based CNN classifier to the conventional models by using number of extracted features

TestData	Gaussian	Naïve Threshold	Median Filter	Ensemble Filter
TestData(1)	40	42	46	67
TestData(2)	48	46	36	62
TestData(3)	36	51	37	59
TestData(4)	44	43	36	60
TestData(5)	49	43	36	65
TestData(6)	41	44	43	65
TestData(7)	44	42	41	61
TestData(8)	35	43	41	66
TestData(9)	48	42	49	69
TestData(10)	36	43	39	61
TestData(11)	47	47	43	60
TestData(12)	39	42	39	61
TestData(13)	46	38	49	68
TestData(14)	35	41	51	68
TestData(15)	43	45	42	68
TestData(16)	43	49	39	63
TestData(17)	45	38	39	62
TestData(18)	50	42	44	68
TestData(19)	49	36	43	63
TestData(20)	44	47	38	60

Table 1, describes the comparative analysis of present ensemble feature selection based deep learning framework by using variable size features. From the table, it is analysed that the present model has better ensemble feature extraction than the conventional approaches using deep learning framework.

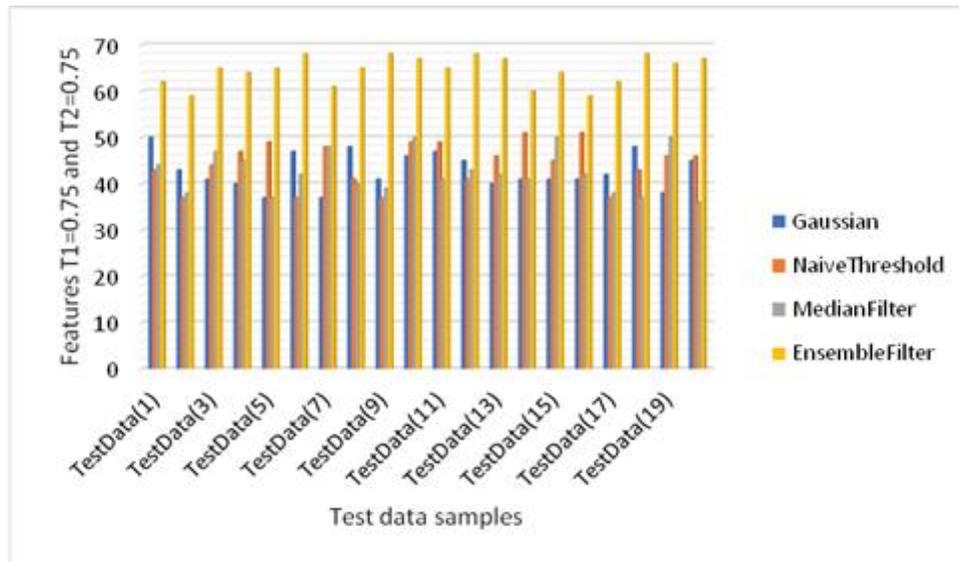


Figure 5. Comparative results of present ensemble feature selection based CNN classifier to the conventional models by using threshold based segmentation values

Figure 5, describes the comparative analysis of present ensemble feature selection based deep learning framework by using variable size features. From the figure, it is analysed that the present model has better ensemble feature extraction than the conventional approaches using deep learning framework.

Table 2. Comparative results of present ensemble feature selection based CNN classifier to the conventional models by using runtime (ms)

TestData	Gaussian	Naïve Threshold	Median Filter	Ensemble Filter
TestData(1)	3441	3553	3355	2878
TestData(2)	3512	3277	3631	2798
TestData(3)	3563	3490	3865	2871
TestData(4)	3378	3504	3454	2720
TestData(5)	3482	3472	3573	2971
TestData(6)	3385	3743	3709	2747
TestData(7)	3786	3718	3382	2835
TestData(8)	3836	3620	3659	2897
TestData(9)	3752	3607	3329	2842
TestData(10)	3441	3666	3759	3038
TestData(11)	3571	3799	3731	3044
TestData(12)	3808	3612	3273	2699
TestData(13)	3398	3496	3809	2839
TestData(14)	3575	3494	3518	2708
TestData(15)	3825	3528	3391	2973
TestData(16)	3261	3854	3467	2685
TestData(17)	3708	3393	3330	2774
TestData(18)	3589	3328	3613	2963
TestData(19)	3389	3266	3592	2814
TestData(20)	3545	3439	3274	2771

Table 2, describes the comparative analysis of present ensemble feature selection based deep learning framework by using variable size features and runtime (ms). From the table, it is analysed that the present model

has better ensemble feature extraction runtime(ms) than the conventional approaches using deep learning framework.

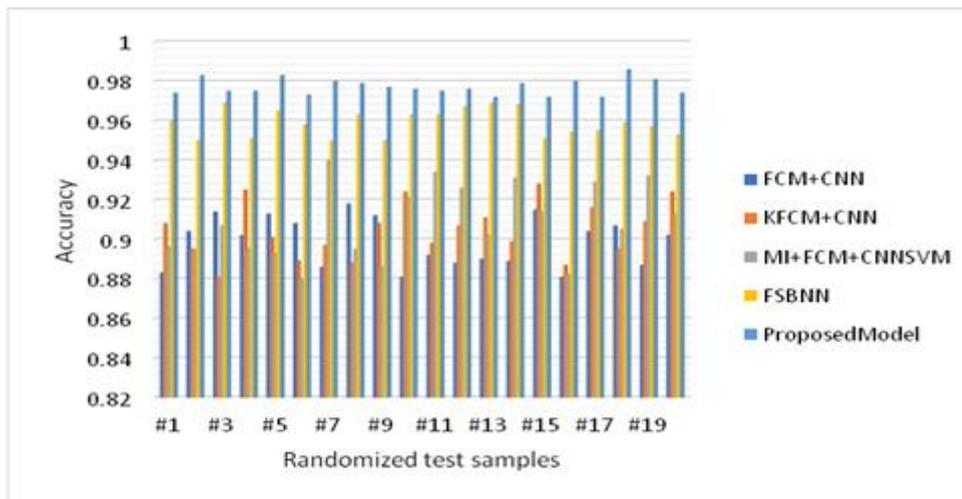


Figure 6. Comparative results of present ensemble feature selection based CNN classifier to the conventional models by using classification accuracy

Figure 6, describes the comparative analysis of present ensemble feature selection based deep learning framework by using variable size features and accuracy measure. From the figure, it is analysed that the present model has better ensemble feature extraction accuracy than the conventional approaches using deep learning framework.

Table 3. Comparative results of present ensemble feature selection based CNN classifier to the conventional models by using classification recall

TestData	FCM+CNN	KFCM+CNN	MI+FCM+CNNSVM	FSBNN	Proposed Model
#1	0.899	0.889	0.925	0.952	0.983
#2	0.882	0.911	0.91	0.956	0.98
#3	0.905	0.924	0.913	0.962	0.974
#4	0.89	0.914	0.909	0.962	0.979
#5	0.888	0.891	0.905	0.968	0.983
#6	0.892	0.911	0.894	0.967	0.98
#7	0.894	0.889	0.91	0.959	0.986
#8	0.891	0.907	0.933	0.956	0.976
#9	0.895	0.887	0.932	0.967	0.976
#10	0.907	0.898	0.899	0.968	0.975
#11	0.904	0.885	0.883	0.958	0.985
#12	0.889	0.929	0.898	0.964	0.978
#13	0.909	0.925	0.903	0.959	0.973
#14	0.918	0.883	0.923	0.961	0.972
#15	0.883	0.887	0.934	0.955	0.98
#16	0.899	0.908	0.932	0.954	0.972
#17	0.884	0.906	0.937	0.96	0.984
#18	0.889	0.895	0.885	0.954	0.982
#19	0.903	0.902	0.922	0.964	0.976
#20	0.891	0.921	0.918	0.953	0.979

Table 3, describes the comparative analysis of present ensemble feature selection based deep learning framework by using variable size features and accuracy measure. From the table, it is analysed that the present

model has better ensemble feature extraction recall than the conventional approaches using deep learning framework.

Table 4. Comparative results of present ensemble feature selection based CNN classifier to the conventional models by using classification precision

TestData	FCM+CNN	KFCM+CNN	MI+FCM+CNNSVM	FSBNN	Proposed Model
#1	0.885	0.922	0.893	0.963	0.982
#2	0.896	0.895	0.902	0.955	0.977
#3	0.912	0.886	0.904	0.958	0.982
#4	0.91	0.899	0.887	0.958	0.981
#5	0.919	0.929	0.922	0.956	0.972
#6	0.919	0.898	0.939	0.955	0.976
#7	0.901	0.929	0.933	0.966	0.986
#8	0.889	0.899	0.902	0.955	0.978
#9	0.918	0.906	0.921	0.952	0.974
#10	0.882	0.92	0.939	0.968	0.977
#11	0.918	0.915	0.907	0.965	0.982
#12	0.906	0.895	0.91	0.954	0.985
#13	0.896	0.913	0.938	0.952	0.974
#14	0.912	0.919	0.915	0.964	0.986
#15	0.899	0.883	0.919	0.959	0.981
#16	0.883	0.912	0.935	0.957	0.976
#17	0.912	0.919	0.94	0.959	0.972
#18	0.916	0.88	0.94	0.962	0.982
#19	0.92	0.897	0.93	0.968	0.981
#20	0.901	0.93	0.914	0.956	0.975

Table 4, describes the comparative analysis of present ensemble feature selection based deep learning framework by using variable size features and precision measure. From the table, it is analysed that the present model has better ensemble feature extraction precision than the conventional approaches using deep learning framework.

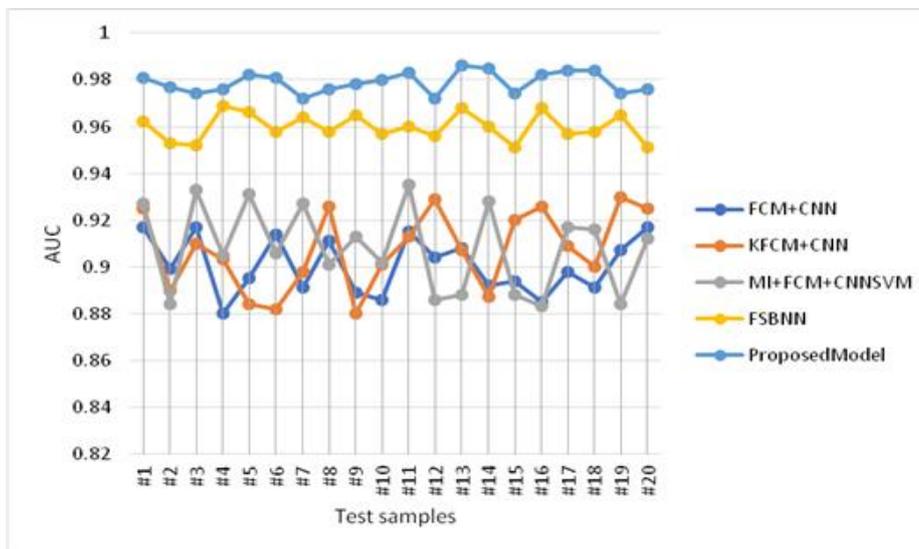


Figure 7. Comparative results of present ensemble feature selection based CNN classifier to the conventional models by using classification AUC

Figure 7, describes the comparative analysis of present ensemble feature selection based deep learning framework by using variable size features and AUC measure. From the figure, it is analysed that the present model has better ensemble feature extraction AUC than the conventional approaches using deep learning framework.

5. Conclusion

In this paper, a hybrid ensemble feature selection measures are designed and implemented on the large multi-class diabetes retinopathy databases. Since, most of the conventional multi-class deep learning frameworks are independent of ensemble feature selection and ensemble classification models due to large data features and noise. In order to solve these issues, ensemble feature selection measures are used to filter the essential features in the large feature space. In this work, a hybrid ensemble feature selection based multiple classification models are used to improve the classification accuracy on multi-class diabetes retinopathy databases. In this work, a novel image segmentation, ensemble feature extraction measures, and multiple classification approaches are used to find the majority voting in the classification problem. Experimental results show that the proposed ensemble feature extraction-based voting classification model has better efficiency compared to the state of art of conventional approaches. In the future work, a parallel deep learning framework is required to optimize the runtime of the high dimensional and large data with variable multiple classes. To implement parallel deep learning framework, a cloud based Hadoop model is required to perform improve segmentation based classification methods in the deep learning framework.

References

1. S. Stolte and R. Fang (2020). *A survey on medical image analysis in diabetic retinopathy*, Medical Image Analysis, vol. 64, p. 101742.
2. Q. Xie et al., (2020). *An innovative method for screening and evaluating the degree of diabetic retinopathy and drug treatment based on artificial intelligence algorithms*, Pharmacological Research, vol. 159, p. 104986.
3. Y. Xie et al., (2020) *Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study*, The Lancet Digital Health, vol. 2, no. 5, pp. e240–e249.
4. Samanta, A. Saha, S. C. Satapathy, S. L. Fernandes, and Y.-D. Zhang (2020). *Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset*, Pattern Recognition Letters, vol. 135, pp. 293–298.
5. W. Zhang et al., (2019). *Automated identification and grading system of diabetic retinopathy using deep neural networks*, Knowledge-Based Systems, vol. 175, pp. 12–25.
6. Y. Katada, N. Ozawa, K. Masayoshi, Y. Ofuji, K. Tsubota, and T. Kurihara (2020). *Automatic screening for diabetic retinopathy in interracial fundus images using artificial intelligence*, Intelligence-Based Medicine, vol. 3–4, p. 100024.
7. S. Wan, Y. Liang, and Y. Zhang (2018). *Deep convolutional neural networks for diabetic retinopathy detection by image classification*,” Computers & Electrical Engineering, vol. 72, pp. 274–282, Nov. 2018.
8. G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard (2017) *Deep image mining for diabetic retinopathy screening*, Medical Image Analysis, vol. 39, pp. 178–193.
9. N. Y. Gharaibeh (2020) *Detection of diabetic retinopathy using partial swarm optimization (PSO) and Gaussian interval type-2 fuzzy membership functions (GIT2FMFS)*, Materials Today: Proceedings.
10. T. Nazir, A. Irtaza, Z. Shabbir, A. Javed, U. Akram, and M. T. Mahmood (2019). *Diabetic retinopathy detection through novel tetragonal local octa patterns and extreme learning machines*, Artificial Intelligence in Medicine, vol. 99, p. 101695.
11. S. Sridhar, J. PradeepKandhasamy, M. Sinthuja, and T. N. Sterlin Minish (2021). *Diabetic retinopathy detection using convolutional neural networks algorithm*, Materials Today: Proceedings.
12. G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. Teatini Salles (2020). *Diabetic retinopathy detection using red lesion localization and convolutional neural networks*, Computers in Biology and Medicine, vol. 116, p. 103537.
13. M. M. Abdelsalam (2020). *Effective blood vessels reconstruction methodology for early detection and classification of diabetic retinopathy using OCTA images by artificial neural network*, Informatics in Medicine Unlocked, vol. 20, p. 100390.

14. P. Cao, F. Ren, C. Wan, J. Yang, and O. Zaiane (2018). *Efficient multi-kernel multi-instance learning using weakly supervised and imbalanced data for diabetic retinopathy diagnosis*, Computerized Medical Imaging and Graphics, vol. 69, pp. 112–124.
15. M. S. Ayhan, L. Kühlewein, G. Aliyeva, W. Inhoffen, F. Ziemssen, and P. Berens (2020). *Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection*, Medical Image Analysis, vol. 64, p. 101724.
16. P. Porwal et al., (2020). *IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge*, Medical Image Analysis, vol. 59, p. 101561.
17. E. Saleh et al., (2018). *Learning ensemble classifiers for diabetic retinopathy assessment*, Artificial Intelligence in Medicine, vol. 85, pp. 50–63.
18. N. Sambyal, P. Saini, R. Syal, and V. Gupta (2020). *Modified U-Net architecture for semantic segmentation of diabetic retinopathy images*, Biocybernetics and Biomedical Engineering, vol. 40, no. 3, pp. 1094–1109.
19. Q. S. You et al., (2020). *Optical Coherence Tomography Angiography Avascular Area Association With 1-Year Treatment Requirement and Disease Progression in Diabetic Retinopathy*, American Journal of Ophthalmology, vol. 217, pp. 268–277.
20. S. R. Sadda et al., (2020). *Quantitative Assessment of the Severity of Diabetic Retinopathy*, American Journal of Ophthalmology, vol. 218, pp. 342–352.