

Automated Recognition of Diabetic Retinopathy using Machine Learning Techniques

J. Pradeepkandhasamy, S. Balamurali, M. Arun, S. Gokulnath

Abstract: *Diabetic Retinopathy (DR) is the illness due to severe polygenic disorders that result in loss of vision for the patients. The development in computer science leads to the timely recognition of DR through an automatic system that is more advantageous than the diagnosis done by a doctor. This paper reviews the DR diagnosis technique that includes deep learning, machine learning and image processing based approaches and their performance. Among the machine learning approaches, the Artificial Neural Network (ANN) classification technique results in high accuracy. The green channel extraction based image contrast enhancement has high classification accuracy, which outperforms the image processing techniques. The performance of the model is estimated by the metrics including sensitivity, specificity and accuracy. This study presents depth insights of techniques for automated DR diagnosis.*

Keywords: *Artificial Neural Network, Artificial Intelligence, Diabetic Retinopathy, Image Processing, Machine Learning.*

I. INTRODUCTION

The retina may be a spherical shaped organ near the inner aspect in the rear of attention. It consists of rods and cones using which the process of photoreception takes place. In the middle of the membrane, a dark spherical space known as macula is present. The function of the macula is to provide a sharp vision. The retina in the eye receives blood through the vasculature similar to all the other parts of the body. The optic disk has macula, may be a part where the nerves optics exits the attention and is additionally the positioning for the veins and arteries of the eye. The symptoms of DR involve hemorrhages, swollen blood vessels, exudates and so on. Diabetes could be a quite common sickness everywhere around the globe. This sickness typically results in visual disorder in patients between twenty and seventy four years old because of a condition caused by uncontrolled polygenic

Revised Manuscript Received on December 05, 2019.

* Correspondence Author

J. Pradeepkandhasamy *, Department of Computer Applications, School of Computing, Kalasalingam Academy of Research and Education, India. Email: j.pradeepkandasamy@klu.ac.in

S. Balamurali, Department of Computer Applications, School of Computing, Kalasalingam Academy of Research and Education, India. Email: sbmurali@rediffmail.com

M. Arun, Department of Computer Applications, School of Computing, Kalasalingam Academy of Research and Education, India. Email: m.arunvincent@gmail.com

S. Gokulnath, Department of Computer Applications, School of Computing, Kalasalingam Academy of Research and Education, India. Email: svgokul3712@gmail.com

disorder referred to as Diabetic Retinopathy (DR) [1, 2]. Blood vessels carry blood to the retinal of the eye that need a optimum glucose intensity associate an unimpeded flow of blood. If great deal of sugar molecules, like aldohexose and levulose, accumulates within the blood and this leads to the rupture of blood vessels and its cells. The abnormal structures are formed as results of shriveled rate that results in DR. Here we get the input images from the publicly available databases such as DIARETDB1, Indian Diabetic Retinopathy Image Dataset (IDRiD), and Structured Analysis of Retina (STARE) database

Proliferative and non-Proliferative are the two stages of DR. Once the blood vessels within the eye gets eroded, the fluid begin to drip into the tissue layer, it results in non Proliferative DR. It leads to wet and swollen tissues in retina. Once the abnormal vessels are developed in numerous components of the tissue layer, it results in Proliferative DR. This stage of DR will cause loss of vision [3]. Recognition of DR at associated stages is extremely necessary so as to save lot of the patients from losing their eye sight. 90 % of the patients are saved from diabetes through timely recognition of DR. Diagnosing of DR may be done either by oculist or by an automatic system. It has both advantages and disadvantages to each of those strategies of DR recognition. The advantage of diagnosis by a doctor is that, it doesn't need laptop help for the recognition of DR; whereas it needs the oculist should be associated professional within the medical field.

Typically, the symptoms of DR are therefore low as even associated professional oculist is unable to acknowledge them in a proper way. Because of the improvements in Artificial Intelligence, timely diagnosis of the illness by an automatic system is additional possible and is better over the recognition of DR by a doctor. The advantages could involve belittled quantity of employment on the oculist and fewer probability of error. An automatic system could also be able to discover abnormal lesions with way larger potency. Hence, automated recognition of DR is crucial. Both the machine learning approaches and deep learning approaches are used to develop a machine-controlled system [4].

The symptoms of the DR disease are gathered at the starting stage of machine learning (ML) approaches. Then preprocessing is performed using different

image preprocessing techniques. A master feature vector (MFV) is built by retrieving the discriminating options available in the preprocessed image. The machine learning algorithms which builds a classification model for DR diagnosis receives the input from the MFV. Finally, the performance is evaluated using test datasets.

II. REVIEW PROCESS

This study involves the following aspects that provide deep insights for DR diagnosis using computational techniques. The dimension in which the study has been undergone includes;

- Preprocessing strategies
- Machine learning algorithms
- Evaluation metrics

Further, the open issues and the upcoming trends are also discussed in addition to the survey.

A. Preprocessing Techniques

Preprocessing removes the inconsistencies and enhances the resolution and quality of the dataset. It results in highly accurate classification and hence considered as a vital step.

Preprocessing techniques reviewed for the study includes

- Green channel extraction
- Gaussian filtering
- Augmentation technique
- Contrast limited adaptive histogram equalization

An in experienced image band that provides a lot of information about the image is extracted from the color image using the Green channel extraction techniques. The contrast of the image is enhanced by distinction. Illumination correction is typically applied when distinction improvement, to extend the luminosity and image brightness. The noise from the image is eliminated with the help of a filter namely, Gaussian filter. A low resolution image is obtained by resizing the image which is consistent with the acceptable system. In general, the authors remove and hide storage device and vessels so that the misclassification of DR lesions can be avoided. Equal amount of objects for varied classes are augmenting the image in a manual way. The imbalance in images of varied classes can be removed by augmentation. The images are revolved, upturned, reflected or converted to provide other objects of the chosen images in a category wherever quantity of images is below the opposite categories.

Sometimes associate degree color image is transformed to grey scale image. The discussions used gray-scale images and every one of those strategies used ML. The blue and red image bands retain less information about the image and the green band has optimal information required. Most preprocessing techniques use inexperienced channel extraction from the RGB image, as the green band shows

maximum information as compared to the red and blue bands. Initially, the inexperienced image channel for recognizing micro aneurysms (MAs) is extracted, wherever the MAs reveal higher distinction. The authors increased the distinction on the inexperienced channel by adopting Contrast Limited Adaptive Histogram Equalization (CLAHE) methodology so the unseen options come into sight.

B. Machine learning based approaches

In ML based image processing, the retina images are collected as a dataset and the preprocessing step is followed to remove the inconsistency. The important and distinct features are extracted for training the model. A classification model is built to classify the healthy retina and abnormal retina images. The dataset is segregated in to training and testing set. The learning set is utilized to make the model learn from the features. The testing set is utilized for the performance evaluation of the proposed model. Fig. 1. shows the basic steps of ML

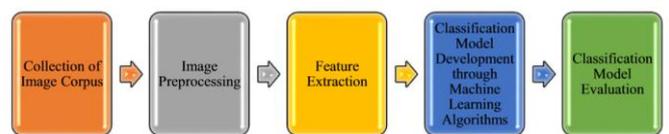


Fig. 1 Basic steps of machine learning.

In general, processing and segmentation of the images are studied within the survey. Medical image segmentation is done by identifying the similarities of neighbor pixels using Active Contour Model (ACM) [5]. In DR recognition algorithms, the process identifies whether lesions are situated in the image or not. The choice of options is a vital work, in which the training becomes economical. Totally the researchers have varied options to observe diabetic retinopathy. Moreover, the authors conjointly combined micro aneurysms and hemorrhages (HMA) as one form based on the shape. Fig. 2 shows the classification of features analysis based on the properties at which it concentrates.

C. Feature Analysis

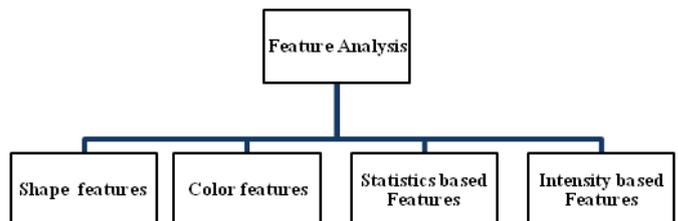


Fig. 2 Types of Feature Analysis

1) Shape features

The shape features embodied in the form and shape of varied DR lesions, which has arduous, exudates, hemorrhages and MAs. The



shape feature options employed in a paper to detect the micro aneurysms are based on space and perimeter of the lesion disk shape [5].

2) Color features

The color features supported the RGB image. The authors used four color feature options to detect the of arduous exudates mistreatment body structure images [6]. The histograms are created from the images based on the color features. The values of red and greed are extracted and normalized from the R-G channel.

3) Statistics based Features

The pixels are measured using mathematical approaches to identify the features in an image known as statistical approaches. Statistical and color features are integrated using math options such as variance, standard deviation and mean for retina images. [7].

4) Intensity based Features

The intensity based features works based on the values of red, green and blue and its intensity in the image. The authors used intensity options for the detection of cotton-wool spots in DR images [8]. The authors used intensity features by computing the most and least component intensity to detection the arduous and exudates [6].

D. ML algorithms

ML algorithms are utilized to train the classification model using features extracted from the preprocessed dataset.

The study also involves the learning algorithms such as

- Random forest classification
- kNN classification
- Linear discriminate analysis
- Artificial Neural Networks
- Ada Boost
- Unsupervised Classification

1) Support Vector Machine

A Support Vector Machine (SVM) is one of the classification algorithms that is used commonly in ML. It attracts a choice margin close to the acute points within the dataset. There are 2 categories, positive and negative, that represents the highest distance among the extreme points that lie in the positive or negative plane. Hyper plane is a plane that divides positive and negative categories. The positive values are grouped under positive class and the negative points are grouped under negative class.

2) Random Forest

Random Forest (RF) is a combination of several Decision Trees (DTs) that is used for classification in ML. In general, the trees combine and project the model in the form of forest. A vote for classification is offered to each tree and it is saved as a label to the tree, which supports classification process. The forest chooses the category having the foremost range of votes. In other terms, random forest classification

algorithmic rule is analogous to sacking approach. A decision tree is constructed for every subset in the RF. In RF, a class with highest number of votes is chosen where as in decision tree the entire set will be chosen.

3) k-Nearest Neighbor

The basic algorithm for ML based classification is k-Nearest Neighbor (kNN) algorithm. It considers the similar features in the learning dataset for object classification. The value of the neighboring nodes that are used for prediction is termed as k.

4) Linear Discrimination Analysis

Linear Discrimination Analysis(LDA) is suitable for reducing the spatial properties as well as classification and multiclass discrimination. The information is classified using the LLDA approach that preserves the changes in directions. The samples can be segregated from different class and can be projected to a line. Most of the prominent studies use LLDA approach in ML. The LLDA algorithm is compared with kNN and SVM classification algorithms using a large dataset of images, which resulted in the hindered performance of LLDA[9]. SVM outperformed both LLDA and kNN.

5) Artificial Neural Network

Artificial Neural Network (ANN) sometimes includes of 3 main layers, namely, input layer, hidden layer and output layer. The input and hidden layers contain several nodes and also the output layer contains only 1 node. A somatic cell in neural network is an activation node. The input layer forwards patterns to the hidden layer wherever actual process is finished. Random weights are assigned to the nodes within the hidden layer. The hidden layer is hooked up to the output layer that is liable for the ultimate output. It may be thought-about as a perceptron, that takes multiple inputs and produces one output [10].

6) Ada boost

Ada Boost adaptative Boosting (Ada Boost) can be applied to develop mathematical models that perform in an ordered manner so that the dataset gets fit in the tree structure. The process of fitting the tree in the dataset results in an accurate classification. In primary studies, Ada Boost classification formula was used just the once. The authors utilized the Ada Boost formula for the recognizing exudatess in DR images [11]. The DRiDB dataset that consists of 50 images is used for experimental purpose and the value of resulted sensitivity is 75%.

7) Unsupervised classification

In unsupervised classification, technique is created to seek out the feature area that is often through with cluster approaches. Most of the researches utilized unsupervised classification algorithms. A dataset that consists of 100 images is used to exploit

the ROC curve to classify micro aneurysms [5, 12]. On comparison with the supervised learning algorithms, the unsupervised learning has low performance for dataset of DR images. [13].

III. PERFORMANCE COMPARISON

This section compares the performance of the models studied in the review. In general, the performance of the classification model is evaluated using true positive, true negatives, false positives and false negatives from which parameters such as sensitivity, accuracy are derived.

Table I Performance Comparison of ML algorithms

S.No	Technique	Performance
1	SVM [14]	95% Accuracy
2	SVM vs kNN, LLDA [15]	0.286 Sensitivity
3	SVM[16]	0.973 AUC
4	SVM [17]	1030 Mean Value
5	kNN[18]	95% Accuracy
6	kNN vs SVM [19]	91% Accuracy
7	ANN [8]	88.3% Sensitivity
8	ANN [4]	94.4% Accuracy
9	AB [20]	75% Sensitivity
10	Unsupervised [21]	89% Sensitivity

The DR images are classified with the help of both unsupervised and supervised ML algorithms. On comparison with unsupervised learning approaches, most of the authors utilized supervised learning approaches. This is often as a result of, the supervised machine learning algorithms are of correct than unsupervised algorithms [22]. In supervised approaches, the model is trained using the labeled images based on the classification rules. The researches concludes that the SVM classifier offers high accuracy and hence its performance is outstanding. When compared with SVM, the performance of ANN is higher and it outperformed the accuracy of SVM. The master feature vector technique is used to extract the features and the extracted features are used to train the supervised classification model. In this study, the most of the classification models utilized only one ML algorithm.

IV. CONCLUSIONS

This paper gives idea about the ML based approaches to diagnose the deadly disease known as diabetic retinopathy. The medical images are fed to the ML model to diagnose the disease. The model classifies the images as healthy or abnormal image. Several feature extraction and ML algorithms are in practice. Out of which SVM classification outperforms the other models. In addition, ANN based approaches provides better results. This has been concluded from the other researches on DR diagnosis. Further, deep learning and machine learning can integrated to obtain better results

REFERENCES

- 1 Q. Abbas, I. Fondon, A. Sarmiento, S. Jiménez, and P. Alemany, "Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features," *Medical & biological engineering & computing*, vol. 55, pp. 1959-1974, 2017.
- 2 A. J. Barber, "A new view of diabetic retinopathy: a neurodegenerative disease of the eye," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 27, pp. 283-290, 2003.
- 3 M. A. Al-Jarrah and H. Shatnawi, "Non-proliferative diabetic retinopathy symptoms detection and classification using neural network," *Journal of medical engineering & technology*, vol. 41, pp. 498-505, 2017.
- 4 B. Antal and A. Hajdu, "An ensemble-based system for automatic screening of diabetic retinopathy," *Knowledge-based systems*, vol. 60, pp. 20-27, 2014.
- 5 W. Zhou, C. Wu, D. Chen, Y. Yi, and W. Du, "Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method," *IEEE Access*, vol. 5, pp. 2563-2572, 2017.
- 6 T. Jaya, J. Dheeba, and N. A. Singh, "Detection of hard exudates in colour fundus images using fuzzy support vector machine-based expert system," *Journal of digital imaging*, vol. 28, pp. 761-768, 2015.
- 7 D. Xiao, S. Yu, J. Vignarajan, D. An, M.-L. Tay-Kearney, and Y. Kanagasigam, "Retinal hemorrhage detection by rule-based and machine learning approach," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 660-663.
- 8 T. Bui, N. Maneerat, and U. Watchareeruetai, "Detection of cotton wool for diabetic retinopathy analysis using neural network," in *2017 IEEE 10th International Workshop on Computational Intelligence and Applications (IWCIA)*, 2017, pp. 203-206.
- 9 C. R. Chand and J. Dheeba, "Automatic detection of exudates in color fundus retinopathy images," *Indian Journal of Science and Technology*, vol. 8, 2015.
- 10 A. Dasgupta and S. Singh, "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 248-251.
- 11 P. Prentasic and S. Loncaric, "Weighted ensemble based automatic detection of exudates in fundus photographs," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 138-141.
- 12 M. Ponnibala and S. Vijayachitra, "A sequential learning method for detection and classification of exudates in retinal images to assess diabetic retinopathy," *Journal of Biological Systems*, vol. 22, pp. 413-428, 2014.
- 13 R. Biyani and B. Patre, "A clustering approach for exudates detection in screening of diabetic retinopathy," in *2016 International Conference on Signal and Information Processing (IconSIP)*, 2016, pp. 1-5.
- 14 Z. Xiao, X. Zhang, L. Geng, F. Zhang, J. Wu, J. Tong, et al., "Automatic non-proliferative diabetic retinopathy screening system based on color fundus image," *Biomedical engineering online*, vol. 16, p. 122, 2017.
- 15 J. Wu, J. Xin, L. Hong, J. You, and N. Zheng, "New hierarchical approach for microaneurysms detection with matched filter and machine learning," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 4322-4325.
- 16 M. J. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE transactions on medical imaging*, vol. 35, pp. 1273-1284, 2016.



- 17 D. S. Sisodia, S. Nair, and P. Khobragade, "Diabetic retinal fundus images: preprocessing and feature extraction for early detection of diabetic retinopathy," *Biomedical and Pharmacology Journal*, vol. 10, pp. 615-626, 2017.
- 18 I. Qureshi, J. Ma, and Q. Abbas, "Recent Development on Detection Methods for the Diagnosis of Diabetic Retinopathy," *Symmetry*, vol. 11, p. 749, 2019.
- 19 S. S. Rahim, C. Jayne, V. Palade, and J. Shuttleworth, "Automatic detection of microaneurysms in colour fundus images for diabetic retinopathy screening," *Neural computing and applications*, vol. 27, pp. 1149-1164, 2016.
- 20 P. Prentašić and S. Lončarić, "Detection of exudates in fundus photographs using convolutional neural networks," in *2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2015, pp. 188-192.
- 21 W. Kusakunniran, Q. Wu, P. Ritthipravat, and J. Zhang, "Hard exudates segmentation based on learned initial seeds and iterative graph cut," *Computer methods and programs in biomedicine*, vol. 158, pp. 173-183, 2018.
- 22 A. Tsiola, F. Hamzei-Sichani, Z. Peterlin, and R. Yuste, "Quantitative morphologic classification of layer 5 neurons from mouse primary visual cortex," *Journal of Comparative Neurology*, vol. 461, pp. 415-428, 2003.

AUTHORS PROFILE



Mr. J. Pradeepkandhasamy, Assistant Professor, Department of Computer Applications, Kalasalingam Academy of Research and Education have more than 7 years of Teaching Experience. He completed his Under Graduation in Computer Science at Bharathidasan University and Post Graduation in Master of Computer Applications in Kalasalingam University, His Major research area is prediction and classification in Data Mining



Dr. S. Balamurali, Senior Professor, Department of Computer Applications, Kalasalingam Academy of Research and Education, he has more than 30 years of Teaching Experience, he has completed his Under graduate in Mathematics in PSG College of Arts and Science, Post Graduate in Statistics in PSG in College of Arts and Science and Doctoral Degree from Bharathiar University in the field of Statistics, His current area of Research is Statistics, Data Mining, Network Security. He has more than 200 publications in International Journals. He is acting as a reviewer for more than 20 peer reviewed journals.



Mr. M. Arun, Assistant Professor, Department of Computer Applications, Kalasalingam Academy of Research and Education have more than 12 years of Teaching Experience. He completed his Under Graduation in Computer Applications at Madurai Kamarajar University and Post Garduation in Computer Applications at Anna University, His Major research area is Network Security.



S. Gokulnath is a student doing his masters in Computer Applications at Kalasalingam Academy of Research and Education; he had done his Bachelor's degree in Computer Applications at same university. His research area is data mining and image processing.