International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

A SURVEY REPORT ON DETECTION OF DIABETICS IN RETINAL IMAGES USING MACHINE LEARNING ALGORITHMS

[1]S. Kalyan kumar, [2]K.Thaiyalnayaki

ABSTRACT— *Diabetic retinopathy is a pervasive eye disease in diabetic patients and it is the most widely known cause of diabetic visual blindness. Diabetic retinopathy requires early detection to prevent patients from losing their eyesight. This paper focuses on decisions regarding the existence of disease by the use of machine learning algorithms. The main reason is to immediately identify the degree of diabetic retinopathy that is non-proliferative in any retinal image . In this case, the initial eye treatment stage isolates necessary characteristics of blood vessels (microaneurysms, hardened exudates) to determine the retinopathy of any retinal image using the support vector machine and other algorithms such as Random Forest, Naïve Bayes, K Neighbor, Decision tree are used for classification .This paper thus focuses on the performance of disease detection using machine learning algorithms and also focuses on previous approaches for the development of efficient algorithms.*

Keywords-- *Diabetic Retinopathy, Pre-processing, algorithms, retinal images, Data sets, performance, classification.*

I. INTRODUCTION

Diabetic retinopathy is a common devastating eye disease. It is well-known that the estimated 2.2 billion people with diabetic disease suffer from vision loss in the population worldwide, with about 33 % DR symptoms, and another 33% of DR vision-threats. The diabetic retinography would otherwise produce full blindness. HEM, which is the hemorrhages, hard exudates, micro aneurysm that should be treated in the initial stage, are the main symptoms. When the person suffering from diabetics should undergo a proper eye test on a regular basis. It'll help avoid the problem of blindness. However, Diabetic Retinopathy is caused by Blood and fluid leakage in the surrounding tissue of eye that damages the veins of the eye. This spillage is triggering microaneurysms, hemorrhages, untreated exudates and patches of cotton wool. Diabetic retinopathy is a silent condition which patients are able to experience if changes in the retina have advanced to such a point that medication cannot be understood. The scheduled discovery of diabetic retinopathy would cause a lot of research time and energy funds. So many image pre-processing techniques were proposed to detect diabetic retinopathy. Diabetic retinopathy is still an area for development despite all the previous works. This paper suggests a new way to diagnose retinal images digitally, helping people recognize diabetic people in the eye early. The main objective is to structure the

[1] *Department of Electronics and Communication, Saveetha School of Engineering, SIMATSChennnai-602105, Tamil Nadu,India,saiempukalyan1@gmail.com*

[2] *Department of Electronics and Communication, Saveetha School of Engineering, SIMATSChennnai-602105, Tamil Nadu, India*

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

classification of diabetic non-proliferative retinopathy and the degree of disease progression in any retinal image, and to isolate blood vessels, microaneurysms and hard exudates in the original image processing process to extract features that may be used by a support vector machine for evaluating any grade of the retinal images. Many retinal images on a database have proven our approach.

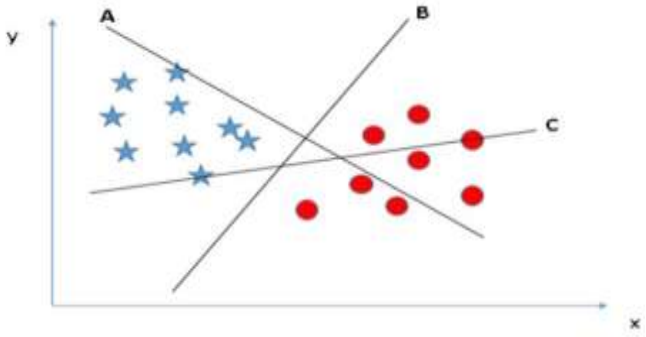

Fig 1 :- Advancement of disease among patients

1.1 DATASETS

The Medical Big Data Center collects patient information data, type of diagnosis, mode of illness and other picture configurations. It is essential to create proper data features in machine learning models to enhance the efficiency of infection prediction. That's because raw information with unstructured information is often noisy and scattered. Classification is based on wide spread models such as Decision tree, K neighbor, SVM, Random forest, Naïve bayes. Such models of classification are widely used in a wide number of areas and are known as specific role choices for classification. Datasets undergoes stages like pre-processing phase, segmentation, extraction of features and classification. Good functionalities for machine learning models need to be developed to guarantee the diagnostic efficacy of diseases. The raw data are often damaged, distorted and unstructured. For our study, we maintain for each case some general demographic data, statistical data and diagnostic marks. The experimental accuracy was excluded from both clinical reports and complicated text details. The data set also omits time columns. The loss of human factors or the lack of objective reasons may lead to an absence of data values. If the outcomes of the diagnostic tests are preserved, the diagnostic accuracy may be significantly decreased. Incoherent data formats also affect future study.

[31]:

	accuracy_all	accuracy_selection	diff_accuracy	cvs_all	cvs_selection	diff_cvs
SGD	0.798246	0.842105	0.043860	0.727449	0.834824	0.107375
SVC	0.921053	0.921053	0.000000	0.885841	0.885841	0.000000
NuSVC	0.921053	0.921053	0.000000	0.852445	0.854200	0.001754
LinearSVC	0.868421	0.745614	-0.122807	0.845304	0.824360	-0.020944
KNeighbors	0.938596	0.921053	-0.017544	0.885779	0.882270	-0.003509
GaussianNB	0.947368	0.947368	0.000000	0.913895	0.908663	-0.005232
RandomForest	0.947368	0.938596	-0.008772	0.947306	0.922698	-0.024608
ExtraTrees	0.947368	0.938596	-0.008772	0.940273	0.915681	-0.024592
DecisionTree	0.921053	0.877193	-0.043860	0.906878	0.901553	-0.005325

Table 1:- Machine learning algorithms table by applying dataset.

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

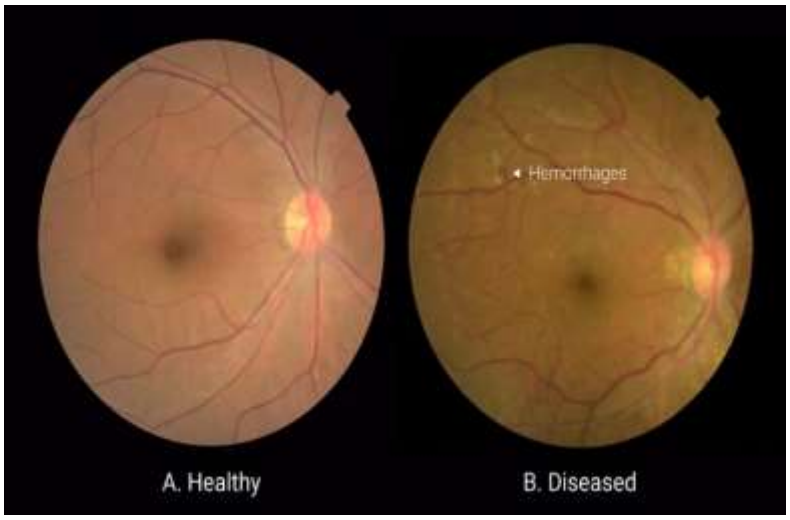

Figure 2:- retinal images taken for screening. The left has a healthy retina, and the right retina has lesions which indicate bleeding and fluid leakage in the eye.

II. PROPOSED ALGORITHMS

1. SUPPORT VECTOR MACHINE

Support vector machine (SVM) include supervised learning modelling with associated learning algorithms that analyse the clashes and regression analysis data. A SVM training algorithm creates a model that provides new examples to one category or another, with each training model marked for one or the other of two categories. A SVM model describes examples as space points, mapped so that the examples of individual categories are separated by the clearest possible distance. New instances are instead mapped into the same space and should be categorized based on the side of the void. Using linear classification, SVMs can effectively carry out a nonlinear classification by mapping their inputs in high-dimensional spaces using called a kernel trick. When data is not classified, supervised learning is not feasible and an unmonitored learning approach is necessary to try to find the normal grouping of data, and then map new data to those classes. The supporting vector algorithm uses the statistics of supporting vector machines.

2.RANDOM FOREST

Random forests is an ensemble learning method in order to score, regress and other tasks, which operate by creating a host of decision-making structures at the time of training and providing the classes for classification or regression of individual trees. Random forests are correct for the behaviour of decision-making trees which overfit their training. The random forest consists of many decision-making trees and is a classification algorithm. When constructing a single tree, it is using grabbing and random features to try to create an uncorrelated forest of trees

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

whose prediction is by committee more assured than that of each individual tree. Training is simpler than decision-making, since we focus on a subset of features in this model, so that hundreds of features are managed easily.

3.KNN

A non-Parametric Classification and Regression Method is the k-nearest Neighbor Algorithm(KNN). The input is the closest examples of workout in the function space k in both cases. The efficiency of your classification or regression depends on the use of k-NN. The development is a member of a class in K-NN classification. An object is marked as a majority of its nearest k (K is a positive integer, typically small) object and assigns the object to its common class. If k=1, the object is automatically allocated to the next class of the neighbor. The value of the property of the object is the product in the KNN regression. This is the sum of the values of the nearest k neighbors. KNN is a type of instance-based learning, or lazy learning that approximates the function only locally and delays all calculations until classification. The value assignment of neighbouring inputs may be useful for both grouping and regression, so that the neighbors contribute more to the average than the distant ones.

4.DECISION TREE

Decision Tree algorithm belongs to the supervised algorithm family. The ultimate theme of the use of the Decision Tree is the construction of a training model that can be used to predict the class or meaning of the target variables by the learning decision rules derived from prior data (training data). The Decision Tree Algorithm comprehension rates is so simple to co-define. Decision tree is a way of approximating the discrete and valued target functions and a decision tree represents the learned function. The tree obtained is a collection of if-then statements that identify cases by sorting the tree from the root node to the leaf, for example by assigning a classification or mark. Every node in a Decision Tree defines a check for a specific instance attribute and each branch coming down from the node is a potential attribute value. Based on information gain, the attribute to be evaluated is selected. By using tree representation, the Decision Tree algorithm attempts to solve the problem.

5.Naive Bayes

Naive Bayes is a binary (two-class) and multi-class classification problem algorithm. When represented using binary or normative input values, the technique is most easy to understand. The estimation of the probability is made easier for each hypothesis, which makes their calculation traceable. These are considered to be conditionally independent in terms of the destination value and to be determined as P(d1)* P(d2) etc., rather than calculating the values for each attribute value P(d1,d2,d3). Naive Bayes is a simple but surprisingly effective predictive modeling algorithm.

III. PROPOSED METHOD

This paper predominantly identifies diabetics during early stages of diabetic retinopathy. When observed, it cannot lead to full blindness at first stage. In the retina there is a HEM film. The removal of HEM therefore implies a diagnosis of the local ternary pattern and the local energy system. The textural characteristics can be defined by the support vector machine. The process is approximately 0.92 % accurate. We also used additional learning

machine algorithms such as random forest, gaussian neighbor (NB) and Decision tree, k neighbor their accuracy is approximately 0.94%, 0.93% and 0.94%,0.93%.

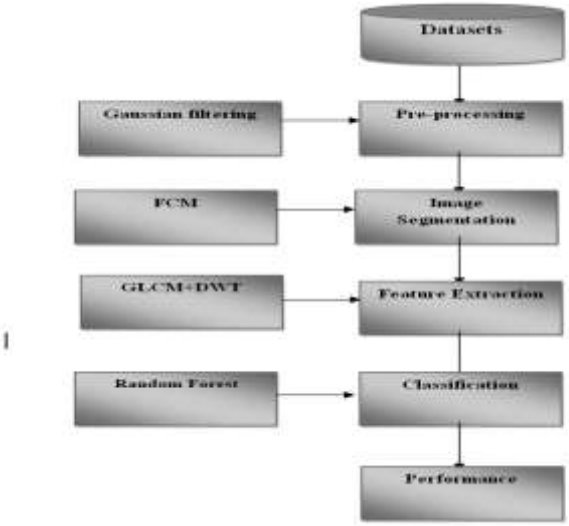

Figure 3:- Block Diagram

IV. REVIEW OF METHODS

Karan Bhatia et., al., [1] Diabetics were indicated in retinal images, such as the Decision Tree, Naive Bayes, Random Forest and SVM, using various algorithms. It is also equipped with an integrated diabetes surveillance system. It is used to test training data so that in different circumstances, images can be categorized as mild or severe. They suggested various diabetic detection technologies for eye disease.

P.R. Asha et., al., [2] A theoretical method for detection of machine-learning techniques was developed to determine retinal exudate presence or absence. In this phase, a total of 100 images are used, 80 for training and 20 for investigation. Extreme Learning Machine (ELM) is used to conduct the diabetic retinal exudate detection test.

Mohamed et., al., [3] In this paper, Various textures, Local Ternary and Local Energy Histogram , have been applied to this report. The histogram observed was described by SVM. The technique of LESH with SVM is approximately 0.904.

Jaykumar et., al., [4] It was suggested that the retinal microaneurysms and automed DR screening exudates be identified with the SVM (support vector machine) and KNN classification. The fundus images of Dataset were collected to classify microaneurysms. They found that SVM performs better than a KNN classification.

Ridam pal et., al., [5] This paper described a process by which machine learning algorithms could be applied to a specific dataset and an algorithm used on a particular data set could be used to define the exact algorithm. They used the Decision Tree, KNN, Naïve Bayes and SVM and ultimately found that SVM is a more accurate algorithm and KNN.

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

S M Asiful Huda et., al., [6] suggested a method of detecting diabetics in the eye using computer algorithms like Decision Tree and Logistic Regression. They have been using SVMs predictive. The proposed method obtained 88% accurate results compared to the existing procedure. Their accuracy is roughly 97% and 92% respectively.

Dinial Utami Nurul Qomariah et., al., [7] In this paper, they used a detailed study approach to extract features and classification using the supporting vector machine and used CNN for classification with SVM as an input function. This approach is done by using 77 and 70 photos in the database of bases 12 and 13. For base 12 and base 13, the accuracy values are around 95.83% and 95.24%.

Keerthi ram et., al., [8] The clustering methods in this paper were proposed to divide the exudates into functions in the fundus picture. This paper focuses primarily on the transmission element. This approach is therefore roughly 89.7 percent more reliable and 87 percent more predictive.

Alireza et., al., [9] In this paper they identified features in fundus pictures. They established an approach. For process segmentation, they used Fuzzy. For extraction of features and the C mean, they used the genetically based algorithm for grouping. As for other approaches, 90% accuracy and 92.1% algorithm sensitivity are higher than the exactness of the proposed technique.

Benzamin et., al., [10] Using image processing and machine learning methods, a system was proposed to identify difficult exudates. The model employed profound learning algorithms, with sensitivity 98.29 percent and predicted specificities 41.35 percent. The model is 96.6 percent accurate over image inspection patches. Therefore, the model tests the accuracy of rough exudates about 98.6%.

V. CONCLUSION

Different types of methods and application of different algorithms are studied to classify diabetics in retinal images. There are a number of different algorithms used that work successfully. The paper provide an outline for this field to help researchers with an interest in diabetic retinopathy to develop more efficient algorithms. A computer-learning model is used in this paper to detect the possible DR in a data set. We have optimized the dataset to eliminate and transform several unstructured, irregular and noisy elements. In order to predict patient samples respectively, machine learning models were then used. By examining its underlying assumptions, we can see that random forests produce better results than other classification methods for this problem. The random forest model is the most accurate algorithm used for detecting the diabetics in retinal images. To boost prediction accuracy, other factors such as patient demographics, past history, family history etc. must be evaluated. To order to filter images of poor quality, parameters must be implemented. In order to help ophthalmologists, lesions in fundal images should be identified. The identification of the DR-staging and laterality of the eye from funduscopic retinal images would be better if the improvement areas were implemented.

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020

REFERENCES

1. Diabetic Retinopathy Diagnosis Using Machine Learning Classification Algorithm by Karan Bhatia, Shikar Arora, Ravi Tomar, Petroleum and Energy Studies University, 2016 , Dehradun, India.
2. Detection of Diabetic Retinal Exudates using Machine Learning Techniques by P.R.Asha, S.Karapagavalli, International Conference on Advanced Computing and Communication System, Coimbatore, India
3. Diabetic retinopathy detection with machine learning and texture functionality, The Department of Computer Science, Université de Moncton, NB. Mohamed Chetoui, Moulay A. Akhloufi, Moustapha Kardouchi, Percentage Robotics and Intelligence Group
4. Diabetic Retinopathy Using machine Learning and Morphological Operations, Jaykumar, Deorankar, Sagar Lachure, Swati gupta, Romit, 2015 IEEE International Advance Computing Conference
5. Application of Diabetic Retinopathy Machine Learning Algorithms by Ridam pal, Jayanta , Mainek Sen. Department of Computer Science and Engineering, Techno India University, West Bengal, 2017 2nd International IEEE Conference on Recent Trends in Information and Communication Technology, 2017, India.
6. Improved Approach for Diabetic Retinopathy Detection Using feature importance and Machine Learning Algorithms by S M Asiful Huda, Ishrat Jahan Ila, Shahrier Sarder, Shamsuja, Nawab Yousuf Ali, 7th International Smart Computing & Communications Conference , 2019.
7. Classification of diabetic retinopathy and normal retinal images using CNN and SVM by Dinial Utami Nurul Qomariah, Handayani Tjandrasa, and Chastine Fatichah, 12thInternational Conference on Information & Communication Technology and System (lCTS) 2019.
8. Ram and JayanthiSivaswamy, Multi-space clustering for segmentation of Exudates in Retinal Colour Photographs, Annual International Conference of the IEEE EMBS, USA, September 2009.
9. Osare et al., "A Computational – Intelligence-Based Approach for Detection of Exudates in Diabetic Retinopathy Images," IEEE Transactions on Information Technology in Biomedicine, July 2009.
10. Hard Exudates observed by using deep learning in retinal images by Avula Benzamin, Chandan Chakraborty, School of Medical Science and Technology, Indian Institute of Technology,2018 IEEE.

Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020