International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT2062141

A Machine Learning Approach for the Diagnosis of Diabetes : A Review

Pravin S. Rahate[1], Dr. Nikhat Raza[2]
[1]PhD Scholar, Department of CSE, MPU, Bhopal, Madhya Pradesh, India
[2]Associate Professor Department of CSE MPU, Bhopal, Madhya Pradesh, India

ABSTRACT

Diabetes mellitus (DM) is a chronic disease that affects 382 million patients' worldwide (2013 data) and is predicted to increase to as many as 592 million adults by 2035. DM is one of the major causes of blindness in young adults around the world. The most serious ocular complication of DM is diabetic retinopathy (DR).Diabetic retinopathy is the most common microvascular complication in diabetes1, for the screening of which the retinal imaging is the most widely used method due to its high sensitivity in detecting retinopathy. Prompt diagnosis is important through efficient screening. The evaluation of the severity and degree of retinopathy associated with a person having diabetes is currently performed by medical experts based on the fundus or retinal images of the patient's eyes As the number of patients with diabetes is rapidly increasing, the number of retinal images produced by the screening programmes will also increase, which in turn introduces a large labor-intensive burden on the medical experts as well as cost to the healthcare services. Manual grading of these images to determine the severity of DR is rather slow and resource demanding. This could be alleviated with an automated system either as support for medical experts' work or as full diagnosis tool. This labor-intensive task could greatly benefit from automatic detection using machine learning technique. Early detection and timely treatment have been shown to prevent visual loss and blindness in patients with retinal complications of diabetes. Machine learning in recent years has been the evolving, reliable and supporting tools in medical domain and has provided the greatest support for predicting disease with correct case of training and testing. The objective of this paper is to explore the work happening on the detection, progression and feature selection process for the prediction of DR and to establish the extent and depth of existing knowledge on RD prediction process.

Keywords : Machine Learning, Diabetic Macular, Diabetic Retinopathy, Image Screening.

I. INTRODUCTION

The prevalence of diabetes increased tenfold, from 1.2% to 12.1%, between 1971 and 2000. It is estimated that 61.3 million people aged 20-79 years live with diabetes in India (2011 estimates). This number is expected to increase to 101.2 million by 2030 and 77.2 million people in India are said to have pre-diabetes. Diabetes affects people both in urban and rural India though the impact on urban India is higher. A most disturbing trend is the shift in age of onset of diabetes to a younger age. Indians get diabetes on average 10 years earlier than their Western counterparts. The risk for coronary artery disease (CAD) is two to four times higher in diabetic subjects, and in Indians, CAD occurs prematurely, i.e., one to two decades earlier than in the West. Costs of diabetes care are alarmingly high. According to the WHO, if one adult in a low-income family has diabetes, "as much as 25% of family income may be devoted to diabetes care. [1]

Lifestyle changes have lead to decreased physical

activity, increased consumption of fat, sugar and calories, and higher stress levels, affecting insulin sensitivity and obesity. Also psychological factors are affecting the functioning of our body organs specially stress level and behavior of individual person may directly affect pancreas functioning. In other words there are so many factors (other than medico) which are affecting our health badly. Also their nature is of non linear way which makes the whole process more complicated as well as beyond the reach of human perception and analysis. The number of dimensions and nonlinear nature of parameters makes the diagnosis more challenging.

In all but the most trivial cases, insight or knowledge you're trying to get out of the raw data won't be obvious from looking at the data. Because now we really do have essentially free and ubiquitous data. So the complementary scarce factor is the ability to understand that data and extract value from it. With so much of the economic activity dependent on information, you can't afford to be lost in the data. The widening gap between data and information calls for the systematic development of tools that can turn data tombs into "golden nuggets" of knowledge. Machine learning will help you get through all the data and extract some information. Machine learning is turning data into information. Any field that needs to interpret and act on data can benefit from machine learning techniques. Machine learning uses statistics. In a nutshell, machine learning is making sense of data. So if you have data you want to understand, if you want to get data and make sense of it, then machine learning is useful.

A basic function of information is to reduce uncertainty. It is often valuable to reduce uncertainty because of how risk affects the things we all do. It is difficult to say, but in general we would like to narrow things down in a way that maximizes the chances of a good outcome and minimizes the chance of a bad one. To do this, we need to make better decisions and to make better decisions we need to reduce uncertainty. Summarily our attempt is to study, analyze and review the use of machine learning approach to get enhanced and pertinent attainment for the diagnosis of diabetics especially retinopathy and prediabetes accentuate psychological and behavioral symptoms.

II. LITERATURE REVIEW

The purpose of a literature review is to gain an understanding of the existing research and debates relevant to an area of study, and to present that knowledge in the form of a written report. The objective is to explore the work happening on the detection, progression and feature selection process for the prediction of DR and to establish the extent and depth of existing knowledge on RD prediction process.

The Md. Faisal Faruque et al.[2] This research work mainly explores various risk factors related to diabetic using machine learning techniques. In this work, four popular machine learning algorithms are employed, namely Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN) and C4.5 Decision Tree (DT), on adult population data to predict diabetic mellitus. This experimental results show that C4.5 decision tree achieved higher accuracy compared to other machine learning techniques.

The Sajratul Yakin Rubaiat et al.[3] Improving the identification level of high-risk factors would help to reduce the rate of complications. The objective of this paper is to find which type of model that works best for predicting diabetes. Primarily it focuses on important feature selection & accuracy comparisons of different machine learning models for early diabetes detection. It includes Data Recovery followed by feature selection. First approach input these features to the MLP neural network classifier which achieved an accuracy of 85.15%. In second approach, they applied noise reduction based method using k-means followed by feature selection. The maximum accuracy

obtained among these classifiers is 77.08%.

The Micheal Dutt et al.[4] This paper highlights need of a promising method for automated diagnosis and use of artificial intelligence. This paper presents an application of Multi-Layer Feed Forward Neural Networks (MLFNN) in diagnosing diabetes on publicly available Pima Indian Diabetes (PID) data set. The achieved accuracy is 82.5% best of all related studies.
The HYPERLINK "https://ieeexplore.ieee.org /author/37086533301" Fikirte Girma Woldemichael et al.[5] This study proposed to predict diabetes using data mining techniques. Back propagation algorithm is used to predict whether the person has diabetic or not. And also J48, naive bayes and support vector machine were used to predict diabetes. These neural networks were having an input layer with having 8 parameters, one hidden layer having 6 neurons and produce one output layer.5 fold cross-validation technique and large value learning rate was used to improve the performance of the model. The performance of Back propagation algorithm is used to predict diabetes diseases with 86.53% sensitivity and the result shows improvement from previous work.

The Jiangxue Han et al.[6] This paper proposes a parametric optimized SVM classifier for diabetic retinopathy. Firstly, the classifier uses PCA and KPCA method to extract the prominent features of the image without artificial recognizing the features of the image, eliminates the specific feature extraction method, reduces the algorithm complexity, increases the generalization ability of the algorithm, and greatly improves the image processing speed. Secondly, grid search and genetic algorithm are used to optimize the parameters, avoid the problem of slow operation speed and low classification accuracy due to the large amount of data or the unsuitable selection of kernel parameters. Finally, a combinatorial optimization algorithm of KPCA and grid search is created. Meanwhile, the designed experiments verify that this combination optimization algorithm can make the classifier achieve the best classification state. The experimental results show that the classification accuracy of this combinatorial optimization algorithm reaches 98.33%, which can realize the automatic classification of diabetic retinopathy more accurately and rapidly.

The Mohamed Chetoui et al.[7] This paper focuses on Hemorrhages, hard Exudates, and Micro-aneurysms (HEM) features that appear in the retina which are the early signs of DR. In this work, they introduce the use of different texture features for DR, mainly Local Ternary Pattern (LTP) and Local Energy-based Shape Histogram (LESH) instead of LBP. A histogram binning scheme for features representation is proposed. The experimental results show that LESH is the best performing technique with an obtained accuracy of 0.904 using SVM with a Radial Basis Function kernel (SVM-RBF).

The NurselYalçin et al.[8] In this study, a deep learning-based approach is presented for the early detection of diabetic retinopathy from retinal images. The proposed approach consists of two steps. In the first stage, pretreatments were performed to remove retinal images from different data sets and standardize them to size. In the second stage, classification was made by Convolutional Neural Network which is a deep learning algorithm and 98.5% success was achieved. The most prominent difference of this study from similar studies is that instead of creating the feature set manually as in traditional methods, the deep learning network automatically constructs itself in a very short time by using the CPU and GPU in training phase.

The R. GeethaRamani et al.[9] In this work, a two-level classification is adopted to classify Diabetic Retinopathy. In this classification, first level classification is performed through ensemble of Best First Trees. The misclassified instances are considered

as noise and outlier and removed. The resultant clean data is then subjected to second level classification, performed through ensemble of J48Graft trees. The evolved rules are evaluated through 3 fold cross validation and the best rules are generated for detection of Diabetic Retinopathy. The method achieves an accuracy of 96.14% revealing a marked improvement when compared to the earlier works.

The ShreyaAliwadi et al.[10] The paper explores the hybrid of SVM and system of Artificial Neural Network as the finest binary classification system for calculating the diabetic nature of people in comparison to Support Vector Machine (SVM). The result of this research shows that this hybrid SVM and Artificial Neural Network (ANN) model is more precise than the SVM model. These results of the hybrid SVM and ANN model suggest that it is very effective for the classification of Diabetic and Non Diabetic nature of a person. This paper highlights the concept of Support Vector Machines and its integration with Artificial Neural Network, the two key characteristics with one being the generalization theory of the SVM model that best describes how to select a hypothesis and functions given by Kernel that introduces the idea of non-linearity without the inclusion of the actual algorithm.

The Maham Jahangir; HYPERLINK "https://ieeexplore.ieee.org/author/38580665300" et al.[11] The paper presents an application of automatic multilayer perceptron (Auto MLP) which is combined with an outlier detection method Enhanced Class Outlier Detection using distance based algorithm to create a novel prediction framework. Auto MLP is an auto-tunable and performs parameter optimization automatically on the run during training process, which otherwise requires human intervention. This framework performs outlier detection during pre-processing of data.

Ephzibah[12]has constructed a model for diabetes diagnosis. Proposed model joins the GA and fuzzy logic. It is used for the selection of best subset of features and also for the enhancement of classification accuracy.

An experimental work to predict diabetes disease is done by the Kumari andChitra[13]. Machine learning technique that is used by the scientist in this experimentis SVM. RBF kernel is used in SVM for the purpose of classification.

Meta learning algorithms for diabetes disease diagnosis has been discussed by Sen and Dash [14].The employed data set is Pima Indians diabetes that is received from UCI Machine Learning laboratory. WEKA is used for analysis. CART, Adaboost, Logiboost and grading learning algorithms are used to predict that patient has diabetes or not. Experimental results are compared on the behalf of correct or incorrect classification. CART offers 78.646% accuracy. The Adaboost obtains 77.864% exactness. Logiboost offers the correctness of 77.479%. Grading has correct classification rate of 66.406%. CART offers highest accuracy of 78.646% and misclassification Rate of 21.354%, which is smaller as compared to other techniques.

The Imran QureshiJun Maand Qaisar Abbas et al.[15]This paper provides an extensive study of computational segmentation and classification algorithms implemented for the detection of retinal anatomic landmarks with DR lesions using fundus images. Although deep learning algorithms have proven to be important in today's image and video processing applications, some possible directions remain that should be addressed for the development of powerful DL networks. First one is the design of DL-based methods that can learn from a limited dataset of fundus images is necessary. Also, in general, there is another class imbalance problem that can occur in the case of developing an automatic solution for DR retinal fundus images. It is important to explore

the class imbalance problem for the development of DL-based methods in the case of learning biases for a specific class. In the future, existing methods should be evaluated on much larger datasets of high resolution images for the recognition of the five severity levels of DR.

The Pedro Romero-Aroca et al.[16] Even though we have a robust system available for screening patients, only 30% of them are screened within the current limitations. It is difficult to achieve the levels recommended by different organizations involved in DM treatment, some of whom recommend biannual screening if patients are well controlled. Specificity has been linked to good prediction for those patients who are not at risk of developing DR in the near future, which is indeed what people are looking for. When building a system for DR screening, they were interested in lengthening the time of frequency of screening with as much confidence as possible that no DR will develop between screenings.Other current studies have focused on the automatic reading of fundus retinographies, which does increase the number of patients screened but the systems need to be validated and their cost effectiveness be demonstrated. This system also has some limitations, such as not including other risk factors like glomerular filtration rate or dyslipidemia, which would affect prognosis.

The Jaakko Sahlsten et al.[17] In this study they have demonstrated that a deep learning AI-system applied to a relatively small retinal image dataset could accurately identify the severity grades of diabetic retinopathy and macular edema and that its accuracy was improved by using high resolution and quality images. They also acknowledge the limitations of our present deep learning AI-system. The first one concerns the image grading reference, that can result in decreased generalization performance of the model. In addition, deep learning neural networks have an inherent limitation of possibly learning features that are unknown or ignored by medical experts, when the network is only fed in an image and its grading without defining diagnostically important features such as microaneurysms and exudates as well as their numbers that are important biomarkers of diabetic retinopathy.

The Sajib Kumar Saha et al.[18] This manuscript review those automated methodsto analyse DR related changes in the retina using color fundus photographs. In order to accurately compare the evolution of DR over time, retinal images that are typically collected on an annual or biennial basis must be perfectly superimposed. However, in reality, for two separate photographic-eye examinations the patient is never in exactly the same position and also the camera may vary. Therefore, a registration method is applied prior to evolution computation. Knowing registration as a fundamental preprocessing step for longitudinal (over time) analysis, they also reviewed state-of-the art methods for the registration of color fundus images.

III. ANALYSIS & OBSERVATIONS

Initially, algorithms of ML were designed and employed to observe medical data sets. Today, for efficient analysis of data, ML recommended various tools. Technologies of ML are very effective for the analysis of medical data and great work is done regarding diagnostic problems.

Feature Selection:

One of the important steps in data preprocessing is feature selection, by this unnecessary features can be removed and improve the performance to build a better classification model. The feature selection is performed on the dataset to select a subset of relevant features for model building that aims to improve model accuracy. Feature Selection is an efficient data preprocessing technique for reducing dimensionality of data . In medical diagnosis, it is very important to identify most significant risk factors related to disease.

Relevant feature identification helps in the removal of unnecessary, redundant attributes from the disease dataset which, in turn, gives quick and better results. The goodness of these results can most likely be attributed on one hand to regularizing image preprocessing and on the other hand to the features in the dataset.

Segmentation & classification Process:

Even though we have a robust system available for screening patients, only 30% of them are screened within the current limitations. In this phase we want automatic reading of fundus retinographies, which does increase the number of patients screened but the systems need to be validated and their cost effectiveness be demonstrated. This process include segmentation and classification of images.

Currently, diabetes screening is carried out using nonmydriatic fundus cameras, and family doctors. It is difficult to achieve the levels recommended by different organizations involved in DM treatment, some of whom recommend biannual screening if patients are well controlled. We also observe that with higher specificity values we obtain lower sensitivity values. Only with random forest and fuzzy rules were we able to achieve a balance between specificity (85.96%) and sensitivity (80.67%), so we believe that is the most suitable choice, with both good sensitivity and good specificity values.

Automated Grading:

Automatic and correct screening of images will give the improved detection with more accuracy. Study reveals that a deep learning AI-system applied to a relatively small retinal image dataset could accurately identify the severity grades of diabetic retinopathy and macular edema and that its accuracy was improved by using high resolution and quality images. Therefore, the design of DL-based methods that can learn from a limited dataset of fundus images is necessary. Also, in general, there is another class imbalance problem that can occur in the case of developing an automatic solution for DR retinal fundus images. It is important to explore the class imbalance problem for the development of DL-based methods in the case of learning biases for a specific class.

Progression Analysis:

This is most crucial part of the system and till less focused. Progression analysis gives the progress of disease infection. As detection of disease requires more advanced techniques progression of infection will rise to high complexity.

IV. CONCLUSION & FUTURE SCOPE

Based on the above study, it can be concluded that there is a huge scope for machine learning algorithms in predicting medical data. Each of the above-mentioned algorithms has performed extremely well in some cases but poorly in some other cases. We divide our study as- feature extraction, segmentation & detection, automated grading and progressive analysis of diabetic disease.

There is a need to find out most preferable attribute as per the experiment to be performed. This helps us to make decision system more robust & accurate for the prediction of diabetes.

The screening of fundus images is the crucial and complex phase while prediction of DR. It includes segmentation of fundus images. It demands to develop an automatic solution for screening of DR retinal fundus images.

Then there is a need of DL-based methods that can learn from a limited dataset of fundus images where Specificity has been linked to good prediction for those patients who are not at risk of developing DR in the near future.

It is difficult to achieve the levels recommended by different organizations involved in DM treatment. The

accurate identification of the severity grades of diabetic retinopathy and macular edema is still hot research topic. We expect to achieve automatic recognition of the five severity levels of DR.

We also expect to analyze DR related changes in the retina using color fundus photographs automatically for the progression analysis of diabetic retinopathy.
In the future, these methods should be evaluated on much larger datasets of high resolution images for the recognition of the five severity levels of DR.

V. REFERENCES

[1]. www.arogyaworld.org

[2]. Md. Faisal Faruque ; Asaduzzaman ; Iqbal H. Sarker ,"Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus", International Conference on Electrical, Computer and Communication Engineering (ECCE),2019

[3]. Sajratul Yakin Rubaiat ; MdMoniborRahman ; Md. KamrulHasan , "Important Feature Selection & Accuracy Comparisons of Different Machine Learning Models for Early Diabetes Detection",International Conference on Innovation in Engineering and Technology (ICIET),2018

[4]. MichealDutt ; VimalaNunavath ;" Morten Goodwin,A Multi-layer Feed Forward Neural Network Approach for Diagnosing Diabetes", 11th International Conference on Developments in eSystems Engineering (DeSE),2018 5.FikirteGirmaWoldemichael ; SumitraMenaria ,"Prediction of Diabetes Using Data Mining Techniques", 2nd International Conference on Trends in Electronics and Informatics (ICOEI),2018

[5]. Jiangxue Han ; Wenping Jiang ; Cuixia Dai ; Hongyan Ma, "The Design of Diabetic Retinopathy Classifier Based on Parameter Optimization SVM", International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS),2018

[6]. Mohamed Chetoui ; Moulav A. Akhloufi ; MustanhaKardouchi, "Diabetic Retinopathy Detection Using Machine Learning and Texture Features" , IEEE Canadian Conference on Electrical & Computer Engineering (CCECE),2018 8.NurselYalçin ; SeyfullahAlver ; NeclaUluhatun, "Classification of retinal images with deep learning for early detection of diabetic retinopathy disease",26th Signal Processing and Communications Applications Conference (SIU),2018

[7]. R. GeethaRamani ; JeslinShanthamalar J ; Lakshmi B,"Automatic Diabetic Retinopathy Detection Through Ensemble Classification Techniques Automated Diabetic Retionapthy Classification", IEEE International Conference on Computational Intelligence and Computing Research (ICCIC),2017 10.ShreyaAliwadi ; VrindaShandila ; TanishaGahlawat ; ParulKalra ; DeeptiMehrotra, "Diagnosis of diabetic nature of a person using SVM and ANN approach", 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO),2017

[8]. Maham Jahangir ; HammadAfzal ; Mehreen Ahmed ; KhawarKhurshid ; RaheelNawaz,"An expert system for diabetes prediction using auto tuned multi-layer perceptron", Intelligent Systems Conference (IntelliSys),2017

[9]. Ephzibah, E.P. "Cost Effective Approach on Feature Selection using Genetic Algorithms and Fuzzy Logic for Diabetes Diagnosis". International Journal on Soft Computing (IJSC), 2011, 2, 1-10.

[10]. Kumari, V.A. and Chitra, R."Classification of Diabetes Disease Using Support Vector Machine.", International Journal of Engineering Research and Applications (IJERA), 2013, 3, 1797-1801.

[11]. Sen, S.K. and Dash, S. "Application of Meta Learning Algorithms for the Prediction of Diabetes Disease.", International Journal of Advance Research in Computer Science and Management Studies, 2014, 2, 396-401.

[12]. Imran Qureshi , Jun Ma, and Qaisar Abbas," Recent Development on Detection Methods for the Diagnosis of Diabetic Retinopathy", Symmetry 2019, 11, 749; doi:10.3390/sym11060749

[13]. Pedro Romero-Aroca, Aida Valls, Antonio Moreno,RamonSagarra-Alamo, JosepBasora-Gallisa, EmranSaleh, Marc Baget-Bernaldiz, DomenecPuig, "A Clinical Decision Support System for Diabetic Retinopathy Screening: Creating a Clinical Support Application ",Telemed J E Health. 2019 Jan 1; 25(1): 31–40.

[14]. JaakkoSahlsten, Joel Jaskari, JyriKivinen, LauriTurunen, EsaJaanio, KustaaHietala&KimmoKaski, "Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading", Scientific Reports | (2019) 9:10750 | https://doi.org/10.1038/s41598-019-47181-w

[15]. Sajib Kumar Saha , Di Xiao , AlauddinBhuiyan , Tien Y. Wong , YogesanKanagasingam "Color fundus image registration techniques and applications for automated analysis of diabetic retinopathy progression: A review", Biomedical Signal Processing and ControlVolume 47, January 2019, Pages 288-302

Cite this article as :

Pravin S. Rahate, Dr. Nikhat Raza, "A Machine Learning Approach for the Diagnosis of Diabetes : A Review ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 509-516, March-April 2020. Available at doi : https://doi.org/10.32628/CSEIT2062141
Journal URL : http://ijsrcseit.com/CSEIT2062141