

Machine Learning Approaches on Diabetic Retinopathy Prediction

¹Gowri Prasad, ²Vrinda Raveendran, ³Sri Vidya B M, ⁴Tejavati Hedge

¹Assistant Professor, Information Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

²⁻⁴ Student, Information Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

Diabetic retinopathy is a eye disorder which is developed due to high blood sugar that affects the neurons in retina. A dangerous fact about this disease is that it can lead to blindness. The possible cure is through detection of disease at early age. This can be done using different machine learning algorithms. This paper does a comparative study on different machine learning algorithms that can be used for early detection of diabetic retinopathy. This study is done to find out the most efficient algorithm suitable for the process and to increase the efficiency of the particular algorithm.

Keywords :Diabetic Retinopathy, Machine Learning, Random Forest Algorithm, Support Vector Algorithm, K-Nearest Algorithm, Neural Networks.

I. INTRODUCTION

Diabetic Retinopathy is a serious eye disease associated with long diabetics those results in progressive damage to the retina, eventually leading to blindness. Recent studies have proved that diabetic's retinopathy affects up to 80 percent of those who have high blood sugar for 20 years or more. The World Health Organization estimates that 347 million people have the disease all around the world. Diabetic retinopathy is developed due to high blood sugar that affects the neurons in retina. A dangerous fact about this disease is that it can lead to blindness. The possible cure is through detection of disease at early age. This can be done using different machine learning algorithms.

Diabetic retinopathy which is also known as diabetic eye disease which damages the eye retina of patient due to diabetics mellitus or high blood sugar content. Diabetics' retinopathy can be developed in humans

with both types of diabetics. There are mainly two types of diabetics, type 1 and type 2. Generally body's immune system is responsible for fighting off foreign invaders, like viruses and bacteria. In type 1 diabetics the immune system mistakes our body's own healthy cells for harmful viruses and bacteria. The immune system attacks and destroys the beta cells which produce insulin in the pancreas. When these beta cells are destroyed, body is unable to produce insulin. This causes type 1 diabetic. A type 2 diabetic is not caused due to lowering of insulin resistance. People with type 2 diabetics produce insulin in their body but will be unable to use it efficiently. In a type 2 diabetic patients body pancreas will produce more insulin since body is unable to use insulin which is available in body. At first diabetic retinopathy shows no symptoms or mild vision problems may occur. Eventually it can lead to blindness. Current treatments available for diabetic retinopathy are laser treatment, eye injections and eye surgery. Other possible ways are prevention and early detection. In

order to prevent diabetic retinopathy we should maintain a proper balanced diet and maintain a healthy life style.

Diabetic retinopathy is still manually done by screening by ophthalmologist. This is a time consuming process which delays the result. This project is done to remove this process and make consultation easier and within less time. Disease will be classified into five different stages based on the severity such as no diabetic retinopathy, mild diabetic retinopathy, moderate retinopathy, severe retinopathy and proliferative retinopathy. We can build an artificial trained system which takes input as patient's fundus image of the eye. Diabetic retinopathy occurs when blood and other fluid start leaking from the tiny blood vessels of the eye causing retinal tissues to swell. As a result of this cloudy or blurred vision will happen. This condition will affect both the eyes.

Diabetic retinopathy will show no symptoms at early stage. So it is always recommended for patients with diabetics to consult doctor and check for diabetic retinopathy. Detecting and treating this disease at early stage can prevent the condition worsening. Convolution neural network is a branch of deep learning which can be used in medical imaging for the process of image analysis and interpretation. Medications for diabetic retinopathy depend on the severity of disease. Laser surgeries play an important role in preventing the leaking of fluid inside eyes.

II. ALGORITHMS

A. Random forest algorithm

Random forest algorithm is a supervised classification algorithm. From the name only we can figure out about the working of this algorithm. There is a direct relationship between the number of trees present in a forest and the result it can produce. The more number of trees, the accuracy of result will be more

and can prevent over fitting. Random forest algorithm is applicable for both regression and classification. Random forest work on the concept of ensemble learning which is a process of combining multiple classifiers to solve a complex problem. It also helps to improve the performance of the model. The major advantages of random forest algorithm is that it take less training time as compared to other algorithms. It predicts the result with high accuracy even if the dataset is comparatively larger. It can also maintain its high accuracy rate when major part of data is missing.

Other major advantages can be pointed out as it can be used for both classification and regression. It also avoids the major problem of over fitting. it also have methods to balance the errors in dataset. Now coming to its disadvantages, the main issue is about its complexity. This particular algorithm is pretty hard and time consuming in terms of construction of decision trees. This also requires more computational resources and is also less intuitive. In addition to all these, the prediction process using random forests is time-consuming than other algorithms. [1]

B. K-nearest Neighbor Algorithm

K-nearest Neighbors is a simple algorithm that saves all available cases and classifies new cases based on a measure of similarity. This algorithm assumes that similar things are present closely. This algorithm makes prediction for a new instance by searching in the entire training set for the most similar instance and will summarize an output variable for the new instance. K-nearest algorithm is based on feature similarity. The main two properties of KNN algorithm are that it is known as lazy learning algorithm since it is not having any specialized training phase and it is a non parametric learning algorithm because it does not assume anything about underlying data.

The main advantage of this algorithm is that it is simple, easy to understand and to implement. In order to use this algorithm we don't have to build a model or to make many assumptions [2].

C. Support Vector Machine Algorithm

Support vector machine is a supervised machine learning algorithm which is used for regression or classification. It is mostly used in classification problems. In this algorithm, we plot each data item like a point in n-dimensional space where n will be the number of features we will be having. We will consider the value of each feature as the value of particular coordinate. Now in this we will perform classification by finding a hyper plane which will differentiate the two classes very well.

Advantage of support vector machine algorithm is that it works well when there is clear margin of separation between the classes we make. This algorithm is more effective in higher dimensional spaces. Support vector machine is effective in cases where number of dimensions is greater than the number of samples available. This is relatively memory efficient. When coming to disadvantages of support vector machine, it is not well suited with larger number of datasets. When data set is noisy that is when they are overlapping, it can't perform efficiently. This algorithm will underperform when number of features for each data point exceeds the number of training data sample. [3]

D. Neural Networks Algorithm

In the field of machine learning neural networks are a subset of algorithms built around a model of artificial neurons spread across three or more layers. It can be also stated as a set of algorithms which are used to understand patterns. Neural networks help us to group unlabeled data according to similarities in the example inputs, and they classify the data when

they have a labeled dataset for training. Neural network will be having at least three layers, outer layer hidden layer and the input layer. Hidden layer can be multiple in number and they will be having connections in between these layers.

Ability to work with inadequate knowledge is one of the major advantages of neural networks. The entire information will be stored in the networks. Neural network have fault tolerance that is stopping of one or two cells won't effect from generating output. The major disadvantage of neural networks is its dependence on hardware. It also does not give us any clue about how and why this solution also the duration of the network is not known. [4]

III. Related Works

In a paper titled Application of Machine Learning Algorithms on Diabetic Retinopathy, Machine Learning algorithms have been used to predict the occurrence of Diabetic Retinopathy. Here among many possible classification algorithms, they have used four algorithms namely, Naive Bayes, Decision Tree, K-Nearest Neighbor and Support Vector Machine. Each of the algorithms have certain advantages and disadvantages specific to the type of applications. However, the results indicated the support vector machine had the highest accuracy among the rest, closely followed by KNN. The performances of the rest of the algorithms were average on the test data set with a few selected parameters. There were a few limitations in the approach in terms of the number of attributes used and the way the attributes were represented. Further studies can be conducted using an Image dataset with Deep Neural Networks rather than extracting features and representing them in numeric form. [5]

Automatic Prediction of Diabetic Retinopathy and Glaucoma through Retinal Image Analysis and Data Mining Techniques, Data mining is another paper in

which techniques have been generously applied in the field of medical diagnosis and prognosis in the past. This work mainly focus on the feature relevance and classification techniques to accurately categorize the disease associated with the retina based on the features extracted from retinal images through image processing techniques. Other than this a detailed comparison of feature relevance and classification algorithms has been performed to justify the statement of reporting C4.5 and Random tree algorithm to be the most accurate classifiers for predicting disease of the retinal image dataset.[6]

In a paper titled Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy they proposed diabetic retinopathy prediction system based on some algorithms and diabetic patient eye image dataset. They utilized local binary patterns for feature extraction from images, naive bayes and support vector machine algorithms to classify patient data because today medical data growing very vastly and that needs to process existed data for predicting exact disease based on symptoms They got accurate DR detected or not prediction as output, by giving the input as patients eye retina images which helps to understand the stage of DR. They compared the results between naive bayes and SVM algorithm in terms of accuracy, memory and time. The accuracy of SVM algorithm which is more than Naive bayes algorithm and time required for classification for SVM is less than naïve bayes also memory required for naïve bayes is greater than the SVM.[7]

In another paper Deep learning algorithm predicts diabetic retinopathy progression in individual patients, they describe an algorithm to predict DR progression by means of deep learning (DL), using as input color fundus photographs acquired at a single visit from a patient with DR. The proposed deep learning models were designed to predict future progression in diabetic retinopathy, defined as a two

step worsening on the Early Treatment Diabetic Retinopathy Diabetic Retinopathy Severity Scale, and were trained against severity scores assessed after 6, 12, and 24 months from the baseline visit by masked of diabetic retinopathy, well-trained, human reading center graders. The performance of one of these models (prediction at month 12) resulted in an area under the curve which is equal to 0.79. Interestingly, their results highlight the importance of the predictive signal located in the peripheral retinal fields, not routinely collected for DR assessments. Their findings show the feasibility of predicting future DR progression. Further development on very large and more diverse image datasets, an algorithm like this will be enable early diagnosis and referral to a retina specialist for more frequent monitoring and even consideration of early intervention. Moreover, it could also improve patient recruitment for clinical trials targeting DR.[8]

In another paper titled Artificial intelligence in diabetic retinopathy: A natural step to the future The AI DR tool can assist the clinician with fundus image analysis, which in turn helps to inform the next steps in the patient's treatment quickly. Also, attention by doctors to more patients that need attention can be given. Emerging healthcare technologies emphasize on reducing visits to eye specialists, curtailing the overall cost of treatment and optimizing the number of patients seen by each doctor. AI can help the health care professional in achieving the goal. Though it assists in health care sector but should not substitute a clinician at its current level. Novel developments in the sector of artificial intelligence are opening up new promises for running DR detection and grading algorithms.[9]

Artificial Intelligence for the Detection of Diabetic Retinopathy in Primary Care: Protocol for Algorithm Development is another paper which is discussing about diabetics retinopathy. Here it aimed at exploring the best classification algorithms and

attempting to create a large dataset in order to reduce errors. In this study the best algorithms are chosen for the detection of diabetic complications such as heart disease, retinopathy, diabetic foot, neuropathy, and nephropathy. Nine algorithms were compared, and the results indicated that the best one is the Random Forest for accurately predicting complications in Type 2 diabetes in most cases and the worst is Naive Bayes[10]

In one paper titled Diagnosis of Diabetic Retinopathy Using Machine Learning Classification Algorithm, methods to develop an automated system to detect the case of diabetic retinopathy among the diabetic patients and is aimed at helping ophthalmologists to detect early symptoms of diabetic retinopathy with ease was proposed. The ideas proposed for the intelligent system can be understood by this paper. Also this paper highlights various technologies used for diagnosis and detection of diabetic eye disease[11]

IV. CONCLUSION

Diabetic retinopathy is a serious disease which should be taken into consideration with providing a suitable cure. As of now early detection is the only technique to prevent the disease. From this work we understood that a neural network is the algorithm on which maximum efforts can be put on to achieve our target. It is an emerging area where we can see a huge possibility of success. Neural networks consist of many algorithms like convolution neural network which can be used in this field. Conclusion content.

Our proposed system aims on maximum accuracy with a proper algorithm. When we look into studies we can see that diabetic retinopathy is one of the leading causes of blindness. With early detection we can possibly bring a change in the current situation.

V. REFERENCES

- [1]. A Comparative Study on Decision Tree and Random Forest Using R Tool Prajwala T R Department of Information Science and Engineering, CMRIT, Bangalore.
- [2]. Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background Sadegh Bafandeh Imandoust And Mohammad Bolandraftar Department of Economics, Payame Noor University, Tehran, Iran.
- [3]. Support Vector Machines (SVM) as a Technique for Solvency Analysis by Laura Auria and Rouslan A. Moro.
- [4]. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes by Jack V. Tu.
- [5]. Yosinski J., Clune J., Nguyen A., Fuchs T., Lipson H. Understanding neural networks through deep visualization. arXiv preprint arXiv:1506.06579, 2015.
- [6]. Abramoff M. D., Reinhardt J. M., Russell S. R., Folk J. C., Mahajan V. B., Niemeijer M., Quilley G. Automated early detection of diabetic retinopathy. *Ophthalmology*, 2010.
- [7]. Gardner G., Keating D., Williamson T., Elliott A. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British journal of Ophthalmology*. 1996.
- [8]. Antal B., Hajdu A. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE transactions on biomedical engineering*, 2012.
- [9]. Quilley G., Lamard M., Josselin P. M., Cazuguel G., Cochener B., Roux C. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Transactions on Medical Imaging*, 2008.
- [10]. UR A. Decision support system for diabetic retinopathy using discrete wavelet transform. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 2013.
- [11]. Gulshan V., Peng L., Coram M., Stumpe M. C., Wu D., Narayanaswamy A., Venugopalan S., Widner K., Madams T., Cuadros J., et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 2016.

Cite this article as :

Gowri Prasad, Vrinda Raveendran, Sri Vidya BM, Tejavati Hedge, "Machine Learning Approaches on Diabetic Retinopathy Prediction", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 6, Issue 3, pp.341-345, May-June-2020. Available at
doi : <https://doi.org/10.32628/CSEIT206377>
Journal URL : <http://ijsrcseit.com/CSEIT206377>