Predicting Diabetic Retinopathy using AI Models: An Experimental Study

A S Mahalakshmi

PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

Abstract**—** *Diabetic retinopathy, a complication of diabetes that affects the eyes, can cause damage to the blood vessels in the retina. Artificial intelligence (AI) techniques play a crucial role in computer-aided diagnosis and have proven successful in identifying various diseases. This study aims to predict diabetic retinopathy by implementing feature extraction techniques to identify relevant factors. The dataset used in this study is sourced from the UCI Machine Learning Repository. Three machine learning algorithms, namely Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Naïve Bayes classifiers, are employed to analyze the dataset and determine the most effective performance and accuracy. Among these classifiers, the SVM algorithm demonstrates the highest performance with an accuracy of 96.56%.*

I. INTRODUCTION

Diabetic retinopathy is a complication of diabetes characterized by the damaging effects of elevated blood glucose on the retina, the surface of the eye. If left undetected and untreated, it can lead to vision loss and is a leading cause of new-onset blindness in adults. Diabetes affects the tiny blood vessels in various organs, including the retina [1]. Diabetic retinopathy is a result of diabetes mellitus and can lead to the formation of scars on the back of the eye. In severe cases, these scars can cause the retina to detach, a condition known as retinal traction detachment. Fortunately, there are several treatment options available that can help prevent or slow down the progression of the disease. Early detection and regular follow-up with a healthcare professional can significantly reduce the risk of blindness, up to 95%. The duration of diabetes is directly related to the likelihood of developing diabetic retinopathy, but maintaining well-controlled blood glucose levels can effectively delay its progression [3][4].

In recent literature, various AI algorithms have been employed for the detection of diabetic retinopathy. Diabetes mellitus is the leading cause of blindness among a significant age group in Western countries and its prevalence is also increasing in developing nations. Individuals with diabetes are at a significantly higher risk of developing blindness compared to those without diabetes. Moderate diabetic retinopathy and clinically significant macular edema can result in severe vision loss. Early detection through regular screening is crucial as it can be effectively treated in its initial stages. However, the cost and manual effort involved in screening are significant, making automated screening highly desirable. In diabetic retinopathy, the blood vessels that nourish the retina start leaking fluid and blood, leading to characteristic visual features such as microaneurysms, hemorrhages, hard exudates, cotton wool spots, and vein occlusion [9].

II. SUPERVISED LEARNING

Supervised learning is a subfield of artificial intelligence (AI) and machine learning that involves training a model using labeled data. It is called "supervised" because the training data provides explicit supervision or guidance to the model in the form of input-output pairs [2][10].

In supervised learning, the goal is to learn a mapping function that can accurately predict the output or label for new, unseen input data. The training data consists of examples where both the input (features) and the desired output (labels) are known. The model learns from these examples to generalize and make predictions on new, unseen data.

III. METHODOLOGY

In this way, the paper proposed Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Naïve Bayes calculations for productively finding the arrangement errands of the diabetic retinopathy information.

3.1 Support Vector Machine (SVM)

SVM is a strong regulated learning calculation utilized for grouping and relapse undertakings. It plans to find an ideal hyperplane that isolates various classes in the element space. The vital thought behind SVM is to expand the edge, which is the distance between the hyperplane and the closest data of interest of each class [7][8].

SVM can deal with both directly distinct and non-straightly detachable information by utilizing portion capabilities to plan the information into higher-layered highlight spaces. The most usually utilized pieces incorporate straight, polynomial, and spiral premise capability (RBF) parts. SVMs have strong hypothetical establishments and are viable in taking care of high-layered information.

SVMs are known for their capacity to actually catch complex choice limits and handle anomalies. They are additionally less inclined to overfitting. Be that as it may, SVMs can be delicate to the decision of hyperparameters, like the regularization boundary (C) and the bit boundaries. Preparing SVMs can be computationally serious, particularly for enormous datasets.

3.2 Multilayer Perceptron (MLP)

A MLP is a legend among the most generally saw Frontal cortex Affiliation plan that has been utilized for different applications. The MLP coordinate is overall created utilizing various fixations or supervising units, and it is sorted out into an improvement of something like two layers [7]. The focal layer (or the most decreased layer) is named as a data layer where it gets the outside data while the last layer (or the most overwhelming layer) is a yield layer where the reaction for the issue is gotten. The mystery layer is the all around enthralling layer in the information layer and the yield layer, and may outline with some spot just about one layers. The plan of MLP could be conveyed as a nonlinear improvement issue. The target of MLP learning is to track down the best loads that limit the parcel between the data and the yield. The most transcendent preparing assessment utilized in NN is Back causing (BP), and it has been utilized in supervising different issues in model validation and depiction. This assessment relies on a couple of limits, for example, novel covered focus fixations at the concealed layers learning rate, energy rate, demand work and how much expecting to occur. Additionally, these limits could change the show on the getting from stunning to remarkable precision [8].

3.3 Naïve Bayes

Naive Bayes classification is a popular machine learning algorithm that is based on Bayes' theorem with an assumption of independence between the features. It is a simple yet effective probabilistic model used for classification tasks.

The algorithm is called "naive" because it assumes that the presence or absence of a particular feature is unrelated to the presence or absence of other features. In other words, it assumes that all features are independent of each other, which is not always true in real-world scenarios. Despite this simplifying assumption, Naive Bayes often performs well in practice and can provide reliable results [7].

The Naive Bayes algorithm works by calculating the probabilities of a sample belonging to each possible class based on the observed feature values. It then assigns the sample to the class with the highest probability. The calculation of these probabilities involves estimating the likelihood of each feature given each class and the prior probability of each class.

The algorithm is particularly useful when working with high-dimensional datasets and when the assumption of feature independence is reasonable. It is known for its computational efficiency and is often used in text classification, spam filtering, sentiment analysis, and other similar tasks[5][6].

One key advantage of Naive Bayes is its ability to handle both numerical and categorical data. It can handle continuous features by assuming a specific distribution, such as Gaussian (for continuous variables) or multinomial (for discrete variables).

IV. EXPERIMENTAL RESULTS

The experiments were conducted using the Python programming language, utilizing the powerful Scikit-learn library for data representation, manipulation, and analysis. For this study, the Diabetes Retinopathy Debrecen dataset from the University of California, Irvine (UCI) library of AI datasets was employed [11]. This dataset consists of 1151 instances, each containing 19 features, along with a binary target variable indicating the presence or absence of diabetic retinopathy. It was curated by a team of researchers from the University of Debrecen, Hungary, who extracted features from the test images to predict the presence of diabetic retinopathy.

In this study, three machine learning algorithms, namely Naïve Bayes, Multilayer Perceptron (MLP), and Support Vector Machine (SVM), were applied to the Diabetes Retinopathy Debrecen dataset. The performance of each algorithm was evaluated using accuracy, precision, and recall as evaluation metrics. The experimental results are summarized in the table-1 and same shown in the figure-1:

Table-1
Classifier Performance

Algorithm	Accuracy	Precision	Recall
Naïve Bayes	93.47	93.5	93.6
Multilayer Perceptron	95.86	95.8	95.7
Support Vector Machine	96.56	96.7	96

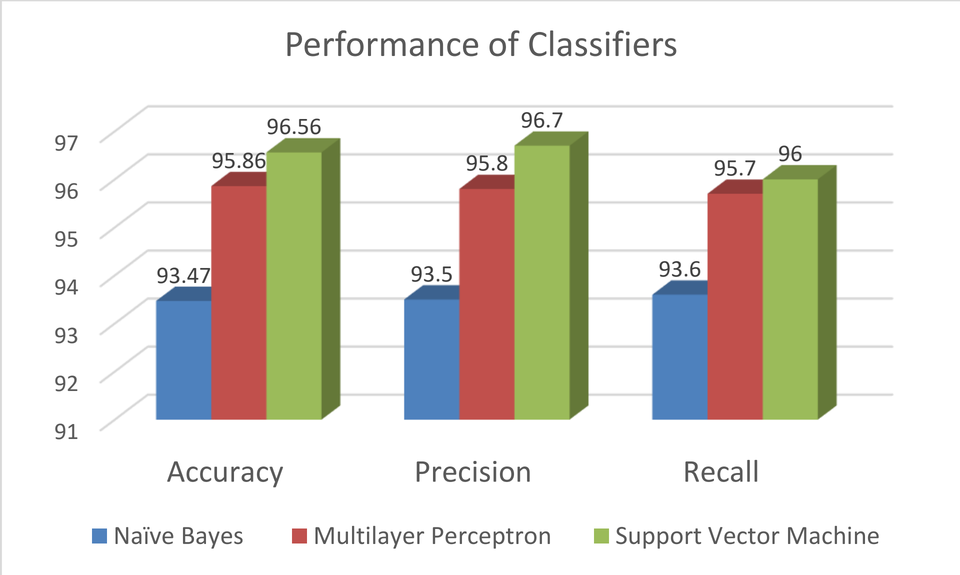

Figure 1: Classifier Results

We observe in the figure-1, Naïve Bayes demonstrated a respectable performance in predicting diabetic retinopathy with an accuracy of 93.47%. The precision and recall values were also high, indicating the model's ability to correctly identify instances of diabetic retinopathy. However, compared to the other two algorithms, Naïve Bayes showed slightly lower accuracy.

The Multilayer Perceptron algorithm exhibited promising results in diabetic retinopathy prediction. With an accuracy of 95.86%, it outperformed Naïve Bayes. The precision and recall values were also high, indicating the model's ability to accurately classify diabetic retinopathy cases.

SVM demonstrated the highest performance among the three algorithms, achieving an accuracy of 96.56%. It showcased excellent precision and recall values, indicating its effectiveness in accurately identifying instances of diabetic retinopathy. SVM outperformed both Naïve Bayes and MLP in terms of accuracy and overall performance.

Overall, all three machine learning algorithms yielded promising results in predicting diabetic retinopathy using the Diabetes Retinopathy Debrecen dataset. SVM emerged as the top-performing algorithm, followed by MLP and Naïve Bayes. These findings suggest that these algorithms have the potential to assist in the early detection and diagnosis of diabetic retinopathy, allowing for timely interventions to prevent vision loss in diabetic patients.

V. CONCLUSION

Based on the experimental results, it can be observed that all three machine learning algorithms achieved high accuracy in predicting diabetic retinopathy. The Support Vector Machine algorithm showed the highest overall performance, with an accuracy of 96.56%. The Multilayer Perceptron algorithm also performed well, with an accuracy of 95.86%. Naïve Bayes demonstrated good performance, albeit slightly lower than the other two algorithms, with an accuracy of 93.47%.

These results indicate that machine learning models have the potential to effectively predict diabetic retinopathy using the Diabetes Retinopathy Debrecen dataset. The high accuracy, precision, and recall values achieved by the algorithms highlight their ability to accurately classify instances of diabetic retinopathy. The findings of this study contribute to the development of AI-based models for the early detection and diagnosis of diabetic retinopathy, which can aid in timely interventions and prevent vision loss in patients with diabetes. Further research and refinement of these algorithms can potentially improve their performance and expand their applications in the field of diabetic retinopathy prediction.

REFERENCES

[1] Agurto, Carla, et al. (2010) "Multiscale AM-FM methods for diabetic retinopathy lesion detection." IEEE transactions on medical imaging 29(2): 502-512.

[2] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)

[3] Ferris, F. L. (1993). How effective are treatments for diabetic retinopathy?. Jama, 269(10), 1290-1291

[4] Fong, D. S., Aiello, L., Gardner, T. W., King, G. L., Blankenship, G., Cavallerano, J. D., & Klein, R. (2004). Retinopathy in diabetes. Diabetes care, 27(suppl 1), s84-s87.

[5] G. Ravi Kumar, S. Rahamat Basha, Surya Bhupal Rao, "A Summarization on Text Mining Techniques for Information Extracting from Applications and Issues", Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 324-332, ISSN:0973-8975.

[6] G. Ravi Kumar, K. Tirupathaiah and B. Krishna Reddy, "Client Churn Prediction of Banking and fund industry utilizing Machine Learning Techniques", International Journal of Computer Sciences and Engineering, Volume-7, Issue-6, e-ISSN: 2347 — 2693, PP: 871-875, June 2019

[7] J.Han and M.Kamber,"Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2nd ed.San Mateo, CA; Morgan Kaufmann, 2006.

[8] N.Michael, "Artificial Intelligence – A Guide to Intelligent Systems", 2nd Edition, Addison Wesley 2005

[9] Ong, Gek L., et al. (2004) "Screening for sight-threatening diabetic retinopathy: comparison of fundus photography with automated color contrast threshold test." American journal of ophthalmology 137(3): 445-452.

[10] Surya Bhupal Rao, S.Rahamat Basha, G. Ravi Kumar, "A Comparative approach of Text Mining: Classification, Clustering and Extraction Techniques", Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 120-131, ISSN: 0973-8975.

[11] UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/.