Diabetic Retinopathy Detection Using Machine Learning

Mohamed Sadiqh [1],Mrs. Roopa R [2]

[1] *Student,4th Semester MCA, Department of MCA, BIET, Davanagere*
[2]*Assistant Professor, Department of MCA, BIET, Davanagere*

ABSTRACT

Diabetic retinopathy (DR) is a prominent cause of vision loss and disability in people around the world. This progressive condition, due to long diabetes at requires early detection and intervention to protect viewing. Traditional diagnostic techniques rely primarily on the retina and manual. To address these issues, this project aims to create an intelligent system for recognizing diabetic retinopathy, excluding direct image processing, using the methodology of dataset control for machine learning. The system uses structured tabular patient information including retinal studies results, including demographic data, clinical indicators (such as blood glucose levels), duration of diabetes, and lifestyle factors. The advanced

-preliminal processing method is used to process these data records, such as normalization, coding of categorical variables, missing data, and functional scaling. Processed data is used to train and evaluate various algorithms of machine learning, including Support Vector Machines (SVM), Random Forest, Decision Trees, K-Nearest-Neighbors (KNN), and Logistics Regression.

Keywords: Sign language, information retrieval, computer vision, natural language processing, accessibility, deaf individuals.

I.INTRODUCTION

Diabetic retinopathy (DR) is a severe eye condition caused by diabetes and is the main cause of blindness in developed countries. Timely interventions can prevent patients from being affected by the disease or at least slowing down the progression of DR. A critical aspect of early detection lies in the timely identification of microuria (MAS) in the fundus of the eye. Therefore, extensive screening of diabetic patients is in great demand, while manual assessments are time-consuming and resource-intensive. MicroNeurysms (MAS) serves as an early indicator of DR and detects effective screening programs for critical importance that adheres to clinical guidelines. Immediate identification of microsymptoms can significantly reduce the risk of blindness, and its detection is the early stages of automated screening for diabetic retinopathy. A reliable screening system for identifying MAS in Digital Find images can greatly support ophthalmologists challenging their diagnostic situation. This project introduces machine learning methods such as Support Vector Machines (SVMs) to evaluate patient data attributes and are suitable for subsequent processing.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Diabetes is a widely recognized disease that can lead to retinal abnormalities (diabetic retinopathy) and problems with nervous system (diabetic neuropathy). Furthermore, diabetes is a serious risk of cardiovascular disease. Diabetic retinopathy is a microvascular complication caused by diabetes and can lead to blindness in a population of the acquired age. The blood vessels that supply the retina gradually become weaker due to diabetes, leading to swelling and sealing. Small, disordered, fragile blood vessels do not provide a sufficient blood supply, which means they break and penetrate the vitreous cavity. Blood flowing through the vitreous interferes with the pathways of light that affectvision. Diabetic retinopathy is one of the major disorders and is the global main cause of avoidable blindness.

II. LITERATURE SURVEY

[Farrikh Alzami, 2019] described a classification system for classification of diabetic retinopathy based on recording of measured data and fractal analysis using random forests. The system seized the photos and calculated the fractal dimensions as properties. They were unable to distinguish slight diabetic retinopathy from severe diabetic retinopathy [1].

[Qomariah 2019] created an automated system for classification of diabetic retinopathy and normal retinal images using simultaneous neuronal networks (CNNs) and support vector machines (SVMs). Characteristics of exudate, bleeding, and microbial urinary disease. The authors divided the proposed system into two parts. The first part was performed in the second part of neuronal network-based distinctive extraction and classification by SVM [2].
[Kuma, 2018] proposed a system that improves the perception of diabetic retinopathy by extracting the region and number of microbial diseases using color- fundus images of DIOTDB1 [3].

Preprocessing of Fundus images was performed using green canal extraction, histogram compensation, and morphological processes. Main Component Analysis (PCA), contrast- limited adaptive histogram equalization (CLAHE), morphological processes, and mean filtering for classification of microuria are performed by the Linear Support Vector Machine (SVM). [Mohamed Chetoui, 2018] created a system that records diabetic retinopathy using a variety of texture features and machine learning classification models. Two properties of bleeding and exudate are extracted using local three-component pattern (LTP) and local energy-based shape (resh). SVM is used to tilt and classify extracted histograms using the LTP and LEH properties [4].

[S Choudhury, 2016] proposed a system that deals with fuzzy C-based characterization extraction and classification of diabetic retinopathy using SVM [5].

Vesicle extraction is performed using a top hat filter and mathematical morphology. The density and exudate of the retinal vessel are selected as properties. Exudate extraction is performed by segmentation, meaning fuzzy c. The GaußscheRadial Basic function is used to allocate training data to the SVM core region.

2.1 EXISTING SYSTEM

Diabetic retinopathy can generally be classified as nonnutrient diabetic retinopathy, and it is important to examine the retinal retina of diabetic patients. Additionally, the use of automated or computer-aided analysis of the retina of diabetic patients can help visual experts check on a larger patient population. Given the large number of patients, the workload of local ophthalmologists becomes extremely important. Therefore, automated detection systems should be designed to alleviate the severity of the disease and to support ophthalmologists in both diagnosis and effective treatment of the disease [1].

Developing such automated systems requires various modules to analyze anatomical retinal functions such as fovea, optical discs, blood vessels, bleeding, microbial systems, and general diabetic pathology such as excretion. A key step in automated detection of diabetic retinopathy is to identify microuria, one of the earliest detectable indicators of diabetic retinopathy. Due to the overwhelming number of patients, existing ophthalmologists are not sufficient to treat all cases, especially in rural areas [2].

2.2 PROBLEM STATEMENT

These studies use traditional models of machine learning, including support vector machines and k-nearest neighbours, to predict diabetes. These models often encounter difficulties when involved in large, complex data records and are not suitable for a wide range of patient populations. Additionally, they usually focus on and ignore important metrics such as accuracy and recalls, focusing only on accuracy. This study used a hybrid deep learning ensemble method to manage these challenges. This improves the predicament and improves a more comprehensive assessment of various metrics that improve the actual use of the model in a clinical setting [2].

2.3 PROPOSED SYSTEM

The proposed application, titled "Recognizing Diabetic Retinopathy through Machine Learning," is highly effective and efficient in identifying stages of retinopathy. In this context, we will develop Dr. Retina's early automated detection system. This can reduce the workload of your ophthalmologist. Furthermore, automatic detection by DR- retina recognition ophthalmologists can more effectively support disease in the testing and treatment of disease. To recognize this, you need to direct your automated screening program to invest heavily. For this diabetic retinopathy program, algorithms for machine learning need to be developed. Functional selection and exudate classification use naive Bayes and support vector machine (SVM) classifiers. Additionally, we successfully implemented a scaling process called standardization in our dataset. We also used a variety of techniques for machine learning, including K-Nearest Neighbor, Naive Bayes, Vector Machine support, decision trees, accidental forests, and logistics regression ideal for achieving improved results [1].

By experimenting with a variety of methods combinations, we created distinctive predictive models. After comparing the results with traditional algorithms for machine learning, the final combination of MLP, TabNet and XGBoost gave us the cheapest results. Previous studies focusing on diabetes prediction used several machine learning algorithms, such as DT, SVM, and LR. This correlates with the model used in this study.

3.1 DATA COLLECTION

Data is collected from structured sources containing numerical and categorical values instead of direct retinal images. These sources include:

- Demographic data: age, gender, etc.
- Clinical indicators: blood glucose, blood pressure, duration of diabetes.
- Retinal test results: encoded as structured data (not images) [1].

3.2 DATA PREPROCESSING

The collected data is cleaned and transformed using:

- Missing value handling: replacing or removing incomplete records.
- Normalization: applying scaling to ensure uniformity.
- Categorical Encoding: converting categorical data into numerical format.
- Feature Scaling: improving convergence speed of algorithms [3].

3.3 FEATURE SELECTION

Important features influencing diabetic retinopathy are identified using:

- Correlation analysis
- Statistical tests
- Domain knowledge

This step improves model accuracy and reduces overfitting [3].

-

III. METHODOLOGY

The proposed system follows a structured methodology to detect diabetic retinopathy using machine learning techniques. The workflow includes the following stages:

MODEL TRAINING

Multiple machine learning algorithms are trained and compared, including:

- Support Vector Machine (SVM)
- Random Forest
- Decision Tree
- K-Nearest Neighbors (KNN)
- Logistic Regression

The goal is to find the best-performing model for classifying DR stages [1].

3.4 MODEL EVALUATION

The models are evaluated using:

- Accuracy
- Precision, Recall, F1-score
- Confusion Matrix
- Cross-validation techniques

This ensures robust and reliable results [3].

3.5 PREDICTION AND CLASSIFICATION

The best model classifies the stage of DR into categories:

- No DR
- Mild DR
- Moderate DR
- Severe DR
- Proliferative DR

This output can help healthcare professionals take timely actions [1].

3.6 DEPLOYMENT

The model is deployed as a simple software interface that:

- Accepts user input
- Processes and classifies patient risk
- Displays DR stage with recommendations [2].

IV. SYSTEM REQUIREMENT SPECIFICATION

4.1 FUNCTIONAL REQUIREMENTS

Creating a machine learning software model aimed at effectively identifying and classifying details of diabetic patients by utilizing a substantial dataset and detecting diabetic stages based on retinal criteria. A machine learning technique has been suggested with respect to accuracy and detection rates across five disease categories: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR [1].

Our developed system is required to execute the following functions:

• Designing a GUI for the collection and processing of real-world retina-tested datasets.

• Creating an algorithm to convert unstructured data into a structured format.

• Developing a framework for predicting stages of DR by analyzing factors such as age and gender.

• Establishing a framework for classifying normal and abnormal DR retinal data [3].

4.2 ARCHITECTURE DIAGRAM

1. Dataset Preparation and Input

• The system initiates with a pre- assembled dataset that comprises a substantial number of images. These images are frequently labeled and categorized to provide context for the deep learning algorithm. The preparation of the dataset may involve preprocessing steps such as resizing, normalization, noise reduction, or augmentation to guarantee consistent quality and usability [1].

• Utilizing datasets as input presents a significant advantage—it enables greater control over the type and diversity of data. This ensures that the system is trained on pertinent samples tailored to its specific application, such as medical imaging, facial recognition, or autonomous driving [5].

2. Deep Learning Algorithm

The foundation of the system is the deep learning algorithm, which processes the dataset through advanced neural network architectures, such as convolutional neural networks (CNNs). These networks extract features, recognize patterns, and derive insights from the dataset, adapting to the specific goals of the system [1].

For instance, the algorithm may classify objects, identify anomalies, or produce predictive analyses based on the content of the dataset. The utilization of pre-collected datasets enhances the algorithm's capacity to generalize across]. various conditions and environments [3].

3. Online Training with Dataset

The system integrates an online training process to continuously enhance its performance. With datasets supplying the algorithm, the system iteratively refines its parameters, learns from mistakes, and adjusts to new data. This methodology ensures that the algorithm remains robust and precise, even as the dataset expands or evolves [3].

Online training is especially beneficial in situations where data is updated periodically, such as in medical diagnostics or industrial quality control. The variety and scope of the dataset are crucial in determining how effectively the system can generalize its learning [3].

4. Big Data Collection and Management

The system is based on the principles of big data to effectively manage and handle the dataset. Adequate organization, labeling, and storage methods are essential for guaranteeing the system's scalability and dependability. The dataset may encompass thousands or even millions of samples, necessitating efficient frameworks for access and analysis [4].

Big data collection is particularly crucial for training deep learning algorithms, as larger datasets typically enhance performance and minimize overfitting [5].

5. Classification Levels

Following processing and analysis, the system categorizes the dataset into predefined levels—designated from 0 to 4 in this instance. These levels may signify categories, conditions, or actions pertinent to the dataset's application. For instance: Classification guarantees that actionable insights are derived from the dataset, rendering the system valuable for decision- making [4].

6. Modular, Adaptive System

The amalgamation of datasets with cloud computing and deep learning results in a modular, adaptive system. Each element— the dataset, the cloud infrastructure, and the algorithm—functions in unison, providing scalability, efficiency, and flexibility. This architecture facilitates seamless upgrades and customization according to the application's requirements [2].

4.3 IMPLEMENTATION

The recommendation system was implemented using Python, leveraging its extensive libraries for data processing and machine learning. Initially, the dataset containing user preferences and item details was collected and thoroughly preprocessed by removing duplicates, handling missing values, and encoding categorical variables where necessary. Once the data was cleaned, feature extraction techniques were applied to identify meaningful patterns. For building the recommendation model, both content- based filtering and collaborative filtering approaches were explored. Content-based filtering analyzed the item attributes to suggest similar products, while collaborative filtering utilized user-item interaction data to predict preferences by identifying similarities between users or items. The model's performance was assessed using evaluation metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), ensuring its accuracy and reliability. Finally, the system was integrated into a simple user interface, allowing users to receive personalized and accurate recommendations based on their historical interactions and preferences [5].

V. REFERENCES

[1] Farrikh Alzami, Abdussalam, Rama Arya Megantara and Ahmad Zainul Fanani, "Diabetic Retinopathy Grade Classification based on Fractal Analysis and Random Forest", International Seminar on Application for Technology of Information and Communication, 2019.

[2]. Dinial Utami Nurul Qomariah, Handayani Tjandrasa and Chastine Fatichah, "Classification of Diabetic Retinopathy and Normal Retinal Images using CNN and SVM", 12th International Conference on Information and Communication Technology and System, 2019.

[3]. Shailesh Kumar and Basant Kumar "Diabetic Retinopathy Detection by Extracting Area and Number of Microaneurysms from Colour Fundus Images", 5th International Conference on Signal Processing and Integrated Networks, 2018.

[4]. Mohamed Chetoui, Moulay A Akhloufi, Mustapha Kardoucha , "Diabetic Retinopathy Detection using Machine Learning and Texture Features", IEEE Canadian Conference on Electrical and Computer Engineering, 2018.

[5]. S Choudhury, S Bandyopadhyay, SK Latib, DK Kaole, C Giri, "Fuzzy C Means based Feature Extraction and Classifiaction of Diabetic Retinopathy using Support Vector Machines", International Conference on Communication and Signal Processing, April 2016.