



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 1, January 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Review of Machine Learning Techniques for Diabetes Classification and Prediction

Aakash Shewani¹, Dr. Navneet Kaur²

Research Scholar, Department of Electronics and Communication Engineering, Sagar Institute of Research & Technology, Bhopal, India¹

Associate Professor, Department of Electronics and Communication Engineering, Sagar Institute of Research & Technology, Bhopal, India²

ABSTRACT: In the medical field, it is essential to predict diseases early to prevent them. Diabetes is one of the most dangerous diseases all over the world. In modern lifestyles, sugar and fat are typically present in our dietary habits, which have increased the risk of diabetes. Diabetes mellitus (DM) is a metabolic disorder characterized by increased blood glucose. The pathology can manifest itself with different conditions, including neuropathy, the main consequence of diabetic disease. To predict the disease, it is extremely important to understand its symptoms. Currently, machine-learning (ML) algorithms are valuable for disease detection. This paper presents review of machine learning techniques for diabetes classification and prediction.

KEYWORDS: Machine Learning, Diabetes Classification, Disease.

I. INTRODUCTION

Diabetes mellitus, commonly known as diabetes, is a group of metabolic disorders characterized by a high blood sugar level (hyperglycemia) over a prolonged period of time. Symptoms often include frequent urination, increased thirst and increased appetite. If left untreated, diabetes can cause many health complications. Acute complications can include diabetic ketoacidosis, hyperosmolar hyperglycemic state, or death. Serious long-term complications include cardiovascular disease, stroke, chronic kidney disease, foot ulcers, damage to the nerves, damage to the eyes and cognitive impairment.

Diabetes is due to either the pancreas not producing enough insulin, or the cells of the body not responding properly to the insulin produced. Insulin is a hormone which is responsible for helping glucose from food get into cells to be used for energy. There are three main types of diabetes mellitus:

Type 1 diabetes results from failure of the pancreas to produce enough insulin due to loss of beta cells. This form was previously referred to as "insulin-dependent diabetes mellitus" or "juvenile diabetes". The loss of beta cells is caused by an autoimmune response. The cause of this autoimmune response is unknown. Although Type 1 diabetes usually appears during childhood or adolescence, it can also develop in adults.

Type 2 diabetes begins with insulin resistance, a condition in which cells fail to respond to insulin properly. As the disease progresses, a lack of insulin may also develop. This form was previously referred to as "non insulin-dependent diabetes mellitus" or "adult-onset diabetes". Type 2 diabetes is more common in older adults, but a significant increase in the prevalence of obesity among children has led to more cases of type 2 diabetes in younger people. The most common cause is a combination of excessive body weight and insufficient exercise.

Gestational diabetes is the third main form, and occurs when pregnant women without a previous history of diabetes develop high blood sugar levels. In women with gestational diabetes, blood sugar usually returns to normal soon after delivery. However, women who had gestational diabetes during pregnancy have a higher risk of developing type 2 diabetes later in life.

Type 1 diabetes must be managed with insulin injections. Prevention and treatment of type 2 diabetes involves maintaining a healthy diet, regular physical exercise, a normal body weight, and avoiding use of tobacco. type 2 diabetes may be treated with oral antidiabetic medications, with or without insulin. Control of blood pressure and maintaining proper foot and eye care are important for people with the disease. Insulin and some oral medications can cause low blood sugar (hypoglycemia). Weight loss surgery in those with obesity is sometimes an effective measure in those with type 2 diabetes. Gestational diabetes usually resolves after the birth of the baby.

Patients with diabetes are at risk of developing a retinal disorder called Proliferative Diabetic Retinopathy (PDR). One of the main characteristics of PDR is the development of neovascularization, a condition in which abnormal blood vessels are formed on the retina. This condition can cause blindness if it is not detected and treated early. Numerous studies have proposed different image processing techniques for detecting neovascularization in fundus images. However, because of its random growth pattern and small size, neovascularization remains challenging to detect. Hence, deep learning techniques are becoming more prevalent in neovascularization identification because of their ability to perform automatic feature extraction on objects with complex features. In this work, a method of neovascularization detection based on transfer learning is proposed [1][3].

Diabetic Retinopathy (DR) is the most common and insidious microvascular complication of diabetes, and can progress asymptotically until a sudden loss of vision occurs. Although DR is prevalent nowadays, its prevention remains challenging [4]. Early detection of Type 2 diabetes is necessary for its prevention. The prediction models for detection systems normally employ common factors that may not properly fit all persons having different health conditions. Therefore, this study proposes a method for type 2 diabetes prediction with factors representing personal health conditions. More specifically, this study proposes a novel prediction method named Average Weighted Objective Distance (AWOD) based on the assumption that the individual has diverse health conditions resulting from different individual factors, a requirement for an effective prediction model. AWOD is a modification of Weighted Objective Distance (WOD) by applying information gain to reveal significant and insignificant individual factors having different priorities, which are represented by different weights [5].

II. LITERATURE SURVEY

U. Ahmed et al.,[1] presents a model using a fused machine learning approach for diabetes prediction. The conceptual framework consists of two types of models: Support Vector Machine (SVM) and Artificial Neural Network (ANN) models. These models analyze the dataset to determine whether a diabetes diagnosis is positive or negative. The dataset used in this research is divided into training data and testing data with a ratio of 70:30 respectively. The output of these models becomes the input membership function for the fuzzy model, whereas the fuzzy logic finally determines whether a diabetes diagnosis is positive or negative. A cloud storage system stores the fused models for future use. Based on the patient's real-time medical record, the fused model predicts whether the patient is diabetic or not. The proposed fused ML model has a prediction accuracy of 94.87, which is higher than the previously published methods.

A. Anaya-Isaza et al.,[2] presents the deep convolutional neural networks paradigm, implementing 12 different data augmentation methods, of which four are conventional, and 8 are newly proposed methods. The results showed that the proposed and the conventional methods increased the network's performance, where a 100% detection was achieved by weighting the DM probability percentages for both images of the feet. Finally, it was also possible to demonstrate the importance of transfer learning, which does not depend on the type of database, but on the data corpus with which the transfer was trained.

M. C. S. Tang et al.,[3] The performance of the transfer learning method is investigated using four pre-trained Convolutional Neural Network (CNN) models, which include AlexNet, GoogLeNet, ResNet18, and ResNet50. In addition, an improved network based on the combination of ResNet18 and GoogLeNet is proposed. Evaluation on 1174 retinal image patches showed that the proposed network could achieve 91.57%, 85.69%, 97.44%, and 97.10% of accuracy, sensitivity, specificity, and precision, respectively. We demonstrated that the proposed method outperforms each individual CNN for neovascularization detection. It also shows better performance compared to another method that utilized deep learning models for feature extraction and Support Vector Machine (SVM) for classification.

M. Bernardini et al.,[4] The multiple aim of this study was to predict the risk of developing DR as diabetic complication (task 1) and, subsequently, temporally stratify the DR risk (task 2) using electronic health records data. To perform these objectives, a novel preprocessing procedure was designed to select both control and pathological patients, and moreover, a novel fully annotated/standardized 120K dataset from multiple diabetologic centers was provided. Globally, although the Extreme Gradient Boosting model offers satisfying predictive performance, the Random Forest model obtained the best predictive performance to solve task 1 and task 2, reaching the best Area Under the Precision-Recall Curve of 72.43 % and 84.38 %, respectively. Also the features importance extracted from the best Machine Learning (ML) models is provided.

P. Nuankaew et al.,[5] presents two datasets from open sources, Pima Indians Diabetes (Dataset 1) and Mendeley Data for Diabetes (Dataset 2) each containing 392 records, were studied. The prediction performance for both datasets is compared with the machine learning-based prediction methods, including K-Nearest Neighbors, Support Vector Machines, Random Forest, and Deep Learning. The comparison results showed that the proposed method provided 93.22% and 98.95% accuracy for Dataset 1 and Dataset 2, respectively, which are higher than those provided by other machine learning-based methods.

S. Samreen et al.,[6] the classification is performed on the preprocessed dataset using a wide range of heterogeneous classifiers like Naive Bayes', Logistic Regression, K-Nearest Neighbor, Decision Trees, Support Vector Machine, Random Forest, AdaBoost, and GradientBoost as base learners followed by their stacking ensemble. The performance evaluation of each machine learning pipeline is done through Repeated Stratified K-fold Cross Validation using the metrics of accuracy, precision, recall, F1 Score and area under Receiver Operating Characteristic curve. For each pipeline, the number of features in the preprocessed dataset varies and the highest accuracy of 98.4% is achieved with Crow Search algorithm through a stacking ensemble of multiple heterogeneous classifiers.

N. Fazakis et al.,[7] designed system concerns diabetes risk prediction in which specific components of the Knowledge Discovery in Database (KDD) process are applied, evaluated and incorporated. Specifically, dataset creation, features selection and classification, using different Supervised Machine Learning (ML) models are considered. The ensemble WeightedVotingLRRFs ML model is proposed to improve the prediction of diabetes, scoring an Area Under the ROC Curve (AUC) of 0.884. Concerning the weighted voting, the optimal weights are estimated by their corresponding Sensitivity and AUC of the ML model based on a bi-objective genetic algorithm. Also, a comparative study is presented among the Finnish Diabetes Risk Score (FINDRISC) and Leicester risk score systems and several ML models, using inductive and transductive learning. The experiments were conducted using data extracted from the English Longitudinal Study of Ageing (ELSA) database.

M. Shokrehodaie et al.,[8] use light sources with multiple wavelengths to enhance the sensitivity and selectivity of glucose detection in an aqueous solution. Multiple wavelength measurements have the potential to compensate for errors associated with inter- and intra-individual differences in blood and tissue components. In this study, the transmission measurements of a custom built optical sensor are examined using 18 different wavelengths between 410 and 940 nm. Results show a high correlation value (0.98) between glucose concentration and transmission intensity for four wavelengths (485, 645, 860 and 940 nm). Five machine learning methods are investigated for glucose predictions.

M. T. Islam et al.,[9] determine whether a person has diabetes or has the risk of developing diabetes are primarily reliant upon clinical biomarkers. In this article, we propose a novel deep learning architecture to predict if a person has diabetes or not from a photograph of his/her retina. Using a relatively small-sized dataset, we develop a multi-stage convolutional neural network (CNN)-based model DiaNet that can reach an accuracy level of over 84% on this task, and in doing so, successfully identifies the regions on the retina images that contribute to its decision-making process, as corroborated by the medical experts in the field.

J. Tulloch et al.,[10] provide a reference for areas of future research. PubMed, Google Scholar, Web of Science and Scopus were searched using the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines for works involving ML and DFUs. In order to be included, studies needed to mention ML, DFUs, and report relevant outcome measures regarding ML algorithm accuracy.

A. H. Syed et al.,[11] validated our proposed model with the existing models built using the National Health and Nutrition Examination Survey (NHANES) dataset and Pima Indian Diabetes (PID) dataset. The results of the Chi-squared test and binary logistic regression showed that the exposures, namely Smoking, Healthy diet, Blood-Pressure (BP), Body Mass Index (BMI), Gender, and Region, contributed significantly ($p < .05$) to the prediction of the Response variable (subjects at high risk of diabetes).

R. Sarki et al.,[12] presents a systematic survey of automated approaches to diabetic eye disease detection from several aspects, namely: i) available datasets, ii) image preprocessing techniques, iii) deep learning models and iv) performance evaluation metrics. The survey provides a comprehensive synopsis of diabetic eye disease detection approaches, including state of the art field approaches, which aim to provide valuable insight into research communities, healthcare professionals and patients with diabetes.



III. CHALLENGES

The Diabetes can be classified into the following general categories:

Type 1 diabetes (due to autoimmune β -cell destruction, usually leading to absolute insulin deficiency, including latent autoimmune diabetes of adulthood)

Type 2 diabetes (due to a progressive loss of adequate β -cell insulin secretion frequently on the background of insulin resistance)

Specific types of diabetes due to other causes, e.g., monogenic diabetes syndromes (such as neonatal diabetes and maturity-onset diabetes of the young), diseases of the exocrine pancreas (such as cystic fibrosis and pancreatitis), and drug- or chemical-induced diabetes (such as with glucocorticoid use, in the treatment of HIV/AIDS, or after organ transplantation)

Gestational diabetes mellitus (diabetes diagnosed in the second or third trimester of pregnancy that was not clearly overt diabetes prior to gestation).

Type 1 diabetes and type 2 diabetes are heterogeneous diseases in which clinical presentation and disease progression may vary considerably. Classification is important for determining therapy, but some individuals cannot be clearly classified as having type 1 or type 2 diabetes at the time of diagnosis. The traditional paradigms of type 2 diabetes occurring only in adults and type 1 diabetes only in children are no longer accurate, as both diseases occur in both age-groups. Children with type 1 diabetes typically present with the hallmark symptoms of polyuria/polydipsia, and approximately one-third present with diabetic ketoacidosis (DKA) (2).

The onset of type 1 diabetes may be more variable in adults; they may not present with the classic symptoms seen in children and may experience temporary remission from the need for insulin (3–5). Occasionally, patients with type 2 diabetes may present with DKA (6), particularly ethnic and racial minorities (7). It is important for the provider to realize that classification of diabetes type is not always straightforward at presentation and that misdiagnosis is common (e.g., adults with type 1 diabetes misdiagnosed as having type 2 diabetes; individuals with maturity-onset diabetes of the young misdiagnosed as having type 1 diabetes, etc.). Although difficulties in distinguishing diabetes type may occur in all age-groups at onset, the diagnosis becomes more obvious over time in people with β -cell deficiency.

IV. PROPOSED METHODOLOGY

- The Load the diabetes prediction dataset from the Kaggle

In this step, the diabetes prediction dataset will be downloaded from kaggle source. It is a large dataset providing company. Then load this dataset into the python environment.

- Visualizing the Dataset

Now open the dataset files and view the various data in term of image features.

- Pre-process the Dataset

Now the data preprocess step applied, here data is finalize for processing. Missing data is either removal or replace form constant one or zero in this step.

- Splitting the Dataset into training and testing

In this step, the final preprocessed of dataset is divided into the training and the testing dataset. In the machine learning, firstly the machine is trained through given dataset then it comes in tested period for remaining dataset.



- Classification Using Machine Learning Algorithm

Now apply the machine or deep learning technique to find the performance parameters. The existing work applied several techniques and find Fused ML Decision is better method then others. In proposed method, we apply ANN method and optimize the better results than other approach; According to the researchers the deep learning method is good for optimization to enhance the accuracy.

- Performance Metrics

Now the performance parameters are calculated in terms of precision, recall, f-1 measure, accuracy etc by using the following formulas-

True Positive (TP): predicted true and event are positive.

True Negative (TN): Predicted true and event are negative.

False Positive (FP): predicted false and event are positive.

False Negative (FN): Predicted false and event are negative.

$$\begin{aligned}
 Precision &= \frac{|TP|}{|TP| + |FP|} \\
 Recall &= \frac{|TP|}{|TP| + |FN|} \\
 F1 &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \\
 Accuracy &= \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}
 \end{aligned}$$

V. CONCLUSION

Diabetes mellitus is a chronic disease and a major public health challenge worldwide. According to the International Diabetes Federation, there are currently 246 million diabetic people worldwide, and this number is expected to rise to 380 million by 2025. Furthermore, 3.8 million deaths are attributable to diabetes complications each year. It has been shown that 80% of type 2 diabetes complications can be prevented or delayed by early identification of people at risk. Machine learning techniques can use to find the detection and prediction of the diabetes classification. This paper presents review of machine learning techniques for diabetes classification and prediction. In future we develop efficient machine or deep learning technique to more accurate prediction model to detection of diabetes diseases.

REFERENCES

- [1]. U. Ahmed et al., "Prediction of Diabetes Empowered With Fused Machine Learning," in IEEE Access, vol. 10, pp. 8529-8538, 2022, doi: 10.1109/ACCESS.2022.3142097.
- [2]. A. Anaya-Isaza and M. Zequera-Diaz, "Detection of Diabetes Mellitus With Deep Learning and Data Augmentation Techniques on Foot Thermography," in IEEE Access, vol. 10, pp. 59564-59591, 2022, doi: 10.1109/ACCESS.2022.3180036.
- [3]. M. C. S. Tang, S. S. Teoh, H. Ibrahim and Z. Embong, "A Deep Learning Approach for the Detection of Neovascularization in Fundus Images Using Transfer Learning," in IEEE Access, vol. 10, pp. 20247-20258, 2022, doi: 10.1109/ACCESS.2022.3151644.
- [4]. M. Bernardini, L. Romeo, A. Mancini and E. Frontoni, "A Clinical Decision Support System to Stratify the Temporal Risk of Diabetic Retinopathy," in IEEE Access, vol. 9, pp. 151864-151872, 2021, doi: 10.1109/ACCESS.2021.3127274.



- [5]. P. Nuankaew, S. Chaising and P. Temdee, "Average Weighted Objective Distance-Based Method for Type 2 Diabetes Prediction," in IEEE Access, vol. 9, pp. 137015-137028, 2021, doi: 10.1109/ACCESS.2021.3117269.
- [6]. S. Samreen, "Memory-Efficient, Accurate and Early Diagnosis of Diabetes Through a Machine Learning Pipeline Employing Crow Search-Based Feature Engineering and a Stacking Ensemble," in IEEE Access, vol. 9, pp. 134335-134354, 2021, doi: 10.1109/ACCESS.2021.3116383.
- [7]. N. Fazakis, O. Kocsis, E. Dritsas, S. Alexiou, N. Fakotakis and K. Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction," in IEEE Access, vol. 9, pp. 103737-103757, 2021, doi: 10.1109/ACCESS.2021.3098691.
- [8]. M. Shokrehodaie, D. P. Cistola, R. C. Roberts and S. Quinones, "Non-Invasive Glucose Monitoring Using Optical Sensor and Machine Learning Techniques for Diabetes Applications," in IEEE Access, vol. 9, pp. 73029-73045, 2021, doi: 10.1109/ACCESS.2021.3079182.
- [9]. M. T. Islam, H. R. H. Al-Absi, E. A. Ruagh and T. Alam, "DiaNet: A Deep Learning Based Architecture to Diagnose Diabetes Using Retinal Images Only," in IEEE Access, vol. 9, pp. 15686-15695, 2021, doi: 10.1109/ACCESS.2021.3052477.
- [10]. J. Tulloch, R. Zamani and M. Akrami, "Machine Learning in the Prevention, Diagnosis and Management of Diabetic Foot Ulcers: A Systematic Review," in IEEE Access, vol. 8, pp. 198977-199000, 2020, doi: 10.1109/ACCESS.2020.3035327.
- [11]. A. H. Syed and T. Khan, "Machine Learning-Based Application for Predicting Risk of Type 2 Diabetes Mellitus (T2DM) in Saudi Arabia: A Retrospective Cross-Sectional Study," in IEEE Access, vol. 8, pp. 199539-199561, 2020, doi: 10.1109/ACCESS.2020.3035026.
- [12]. R. Sarki, K. Ahmed, H. Wang and Y. Zhang, "Automatic Detection of Diabetic Eye Disease Through Deep Learning Using Fundus Images: A Survey," in IEEE Access, vol. 8, pp. 151133-151149, 2020, doi: 10.1109/ACCESS.2020.3015258.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details