IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Diabetes Classification and Prediction using Machine Learning Approach

Aakash Shewani [1], Dr. Navneet Kaur [2]

[1]Research Scholar, [2]Associate Professor

Department of Electronics and Communication Engineering

Sagar Institute of Research & Technology, Bhopal, India

Abstract: 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. 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 diabetes classification and prediction using machine learning approach. Proposed system is increased the accuracy of the results by prediction of diabetes by using artificial neural network (ANN) deep learning algorithm. It enhances the performance of the overall classification results.

Index Terms **- ANN, Machine Learning, Diabetes Classification, Disease.**

I. INTRODUCTION

Diabetes is a metabolic disorder that impairs an individual's body to process blood glucose, known as blood sugar. This disease is characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both [1]. An absolute deficiency of insulin secretion causes type 1 diabetes (T1D). Diabetes drastically spreads due to the patient's inability to use the produced insulin. It is called type 2 diabetes (T2D) [2]. Both types are increasing rapidly, but the ratio of increase in T2D is higher than T1D. 90 to 95% of cases of diabetes are of T2D.

Both inadequate insulin production by the pancreas and improper insulin response by body cells contribute to the development of diabetes. Insulin is a hormone that facilitates glucose entry into cells, where it is used for energy. Three major subtypes of diabetes mellitus have been identified. Loss of beta cells in the pancreas causes insulin insufficiency and, ultimately, type 1 diabetes. This kind of diabetes was once known as juvenile diabetes or insulin-dependent diabetes mellitus. An autoimmune reaction is to blame for the death of beta cells. It is unclear what sets off this autoimmune reaction. While most people with Type 1 diabetes are diagnosed in infancy or adolescence, the condition may also strike adults.

Insulin resistance, a state in which cells do not react normally to insulin, is the underlying cause of type 2 diabetes. A deficiency in insulin may emerge later in the course of the illness. This kind of diabetes was once known as adult-onset diabetes or non-insulin-dependent diabetes mellitus. Although older persons make up the majority of those diagnosed with type 2 diabetes, the rising rates of childhood obesity have led to a rise in instances of type 2 diabetes among younger people. Excessive body fat and a lack of physical activity are the most prevalent risk factors. Thirdly, pregnant women who have no prior history of diabetes might acquire gestational diabetes, which causes them to have abnormally high blood sugar levels throughout their pregnancies. Women with gestational diabetes often see a recovery to normal blood sugar levels shortly after giving birth. Women who have experienced gestational diabetes are more likely to acquire type 2 diabetes in the future.

Insulin injections are the only way to control type 1 diabetes. Type 2 diabetes may be prevented and managed with a healthy lifestyle that emphasises eating well, exercising regularly, keeping weight in check, and not smoking. In addition to insulin, oral antidiabetic medicines may be used to treat type 2 diabetes. Treatment focuses on managing symptoms, including blood pressure control, foot care, and eye health. Low blood sugar may be caused by various oral medicines and insulin (hypoglycemia). Those who are morbidly obese and suffer from type 2 diabetes may find success with bariatric surgery. In most cases, gestational diabetes disappears after giving delivery.

Proliferative diabetic retinopathy is a disease of the retina that may affect diabetic patients (PDR). Neovascularization, a disorder in which aberrant blood vessels grow on the retina, is a hallmark of PDR. If this problem is not caught and addressed in time, it might lead to permanent blindness. Multiple research have suggested various image processing strategies for identifying neovascularization in fundus photographs. However, neovascularization is still difficult to identify because of its erratic development pattern and tiny size. Therefore, deep learning algorithms are gaining popularity in neovascularization recognition due to their capacity for autonomous feature extraction on objects with complicated properties. In this study, we offer a technique for detecting neovascularization via transfer learning.

II. LITERATURE SURVEY

U. Ahmed et al.,[1] presents 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] 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] 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] perform these objectives, a novel preprocessing procedure was designed to select both control and pathological patients, and moreover, a novel fully annotated/standardized.

P. Nuankaew et al.,[5] 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.

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. M. Shokrekhodaei 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.

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. 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.

A. H. Syed et al.,[11] validated 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 ($p < 0.05$) to the prediction of the Response variable (subjects at high risk of diabetes). R. Sarki et al.,[12] 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. PROPOSED METHODOLOGY

The proposed methodology is explained using following sub modules-

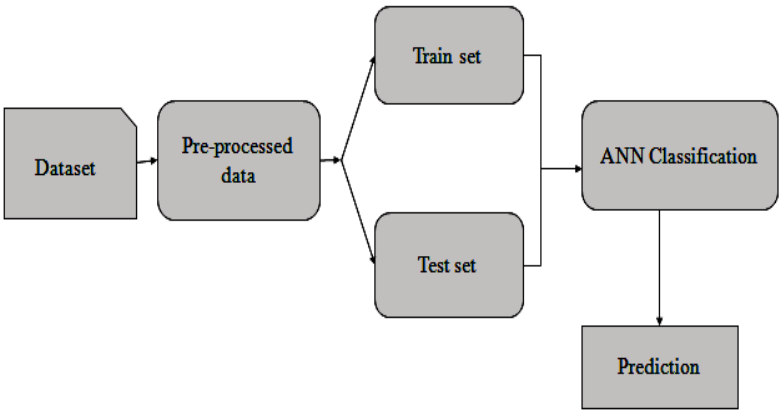

Figure 1: Flow Chart

1. Firstly, download the diabetes prediction dataset from online public available dataset.
2. Now apply the pre-processing of the data, here handing the missing data, removal null values.
3. Now extract the data features and evaluate in dependent and independent variable.
4. Now apply the ANN classification method based on the deep learning.
5. Now generate confusion matrix and show all predicted class like true positive, false positive, true negative and false negative.
6. Now calculate the performance parameters by using the standard formulas in terms of the precision, recall, f_measure, accuracy and error rate.

The methodology step is discussed based on the following steps-

- Data selection and loading
- Data preprocessing
- Splitting dataset
- Classification
- Result generation

Data Selection and Loading

- Data selection is the process of determining the appropriate data type and source, as well as suitable source to collect data.

Data Pre-processing

- Data pre-processing is the process of removing the unwanted data from the dataset.
- It handles missing data removal and encoding categorical data.

Splitting Dataset

- Data splitting is the act of partitioning available data into. Two portions, usually for cross- valedictory purposes.
- One portion of the data is used to develop a predictive model. And the other to evaluate the model's performance.

Classification

ANN- Artificial Neural Network can be best represented as a weighted directed graph, where the artificial neurons form the nodes. The association between the neurons outputs and neuron inputs can be viewed as the directed edges with weights. The Artificial Neural Network receives the input signal from the external source in the form of a pattern and image in the form of a vector. These inputs are then mathematically assigned by the notations x(n) for every n number of inputs.

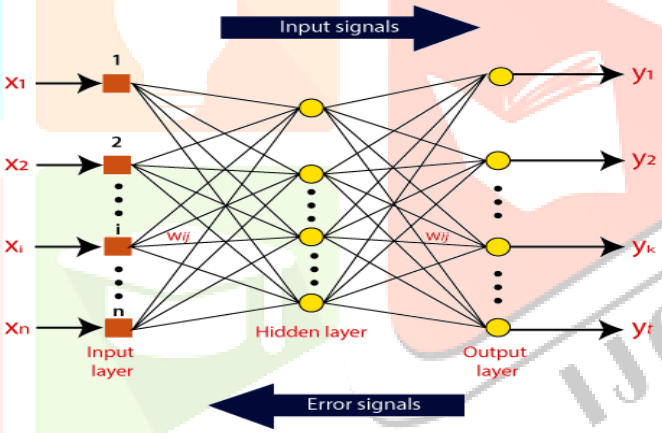

Figure 2: Layers representation of ANN.

Afterward, each of the input is multiplied by its corresponding weights (these weights are the details utilized by the artificial neural networks to solve a specific problem). In general terms, these weights normally represent the strength of the interconnection between neurons inside the artificial neural network. All the weighted inputs are summarized inside the computing unit.

If the weighted sum is equal to zero, then bias is added to make the output non-zero or something else to scale up to the system's response. Bias has the same input, and weight equals to 1. Here the total of weighted inputs can be in the range of 0 to positive infinity. Here, to keep the response in the limits of the desired value, a certain maximum value is benchmarked, and the total of weighted inputs is passed through the activation function.

Prediction

- This study successfully forecasted the data from the dataset by improving the overall performance of the prediction findings, and it does so by using a technique for predicting diabetes prediction.

Algorithm

Input: Diabetes Prediction Dataset.

Consider the basic information characteristics, such as Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes, Pedigree, Function, Age, Outcome etc.

Filtering the null value

Sort the data set according to the characteristics chosen.

Output: Best values for F-measure, Precision, Accuracy, Recall, and Classification Error

Step: 1. now dataset is divided into 2 part train and test dataset like train of y and x and test of y and x
2. Extractions of features, features = {} for diabetes count: features [diabetes count] = True
3. Model selection and split

Y train
Y test
4. Use a classifier based on deep learning's artificial neural network.
5. Confusion matrix with TP, FP, TN, and FN values shown.
6. Determine the percentage of correct answers, standard error, recall, and f-measure.
7. Create a ROC graph.

Evaluation

Accuracy, precision, and recall are the main metrics used to assess a classification model.

- Accuracy is defined as the ratio of true positives to total positives, while recall is defined as the ratio of positives to negatives.
- Accuracy = [TP +TN] / [TP+TN+FP+FN];
- F1-Score = 2x (Precision x Recall)/(Precision + Recall)
- Classification Error = 100- Accuracy.

Result Generation

The total categorization and prediction will be used to create the final result. Accuracy, error rate, and other similar metrics are used to assess the effectiveness of the suggested method.

IV. SIMULATION RESULTS

The execution of the proposed calculation is done over python spyder 3.7. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library assists us with utilizing the capacities accessible in spyder climate for different strategies-

Figure 3: Dataset

Figure 3 is showing the dataset in the python environment. The dataset have 1547 numbers of rows and 8 no of column. The features name is mention in each column.

Figure 4: Y test

Figure 4 is showing the y test of the given dataset. The given dataset is divided into 1044 data for training and 115 data for testing.

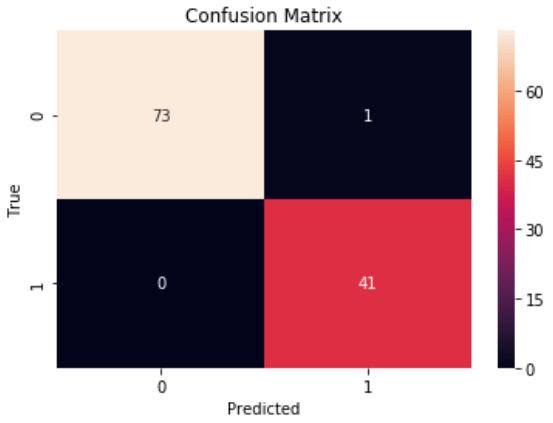

Figure 5: Confusion matrix of SVM

Figure 5 is presents the confusion matrix of the proposed ANN approach. The value of true positive is 73, false positive is 1, false negative is 0 and true negative is 41.

Table 1: Result Comparison

Sr. No.	Parameters	Previous Work [1]	Proposed Work
1	Method	Fused ML Decision	ANN
2	Accuracy	94.87 %	99.13 %
3	Error Rate	5.13 %	0.87 %
4	Sensitivity	95.52 %	100 %
5	Specificity	94.38 %	97.61 %

V. CONCLUSION

Diabetes is a very familiar word in the present world and crucial challenges in both developed and developing countries. This paper presents diabetes classification and prediction using machine learning approach. The simulation is performed using python sypder 3.7 software. Simulated results show that the overall accuracy achieved by the proposed work is 99.13 % while previous it is achieved 94.87%. The classification error of proposed technique is 0.87 % while 5.13 % in existing work. Therefore it is clear from the simulation results; the proposed work achieves significant better results than existing work.

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. Shokrekhodaei, 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