Survey of Diabetes Classification and Prediction based on Artificial Intelligence Techniques

[1]Chandan Kumar, [2]Amit Dalal
[1]Research Scholar, [2]Assistant Professor
Department of Electronics & Communication Engineering,
Ganga Institute of Technology and Management, Kablana, Jhajjar, India

***Abstract*— It is crucial to the profession of medicine to be able to anticipate illnesses early on in order to avoid them. Diabetes is widely recognized as one of the most lethal illnesses on a global scale. Because of the current lives we lead, sugar and fat are frequently included in our dietary habits; as a result, the risk of developing diabetes has grown. The metabolic condition known as diabetes mellitus (DM) is characterized by elevated levels of glucose in the blood. The pathology may show itself in a variety of disorders, the most prominent of which is neuropathy, which is brought on by diabetes illness. It is essential to have a thorough understanding of the disease's symptoms in order to make an accurate prognosis. In the field of illness diagnosis, machine learning (ML) techniques are now quite useful. The purpose of this study is to provide an overview of machine learning approaches for the categorization and prediction of diabetes.**

Keywords—AI, Machine Learning, Diabetes Classification, Disease.

I. INTRODUCTION

Diabetes mellitus, more frequently referred to as diabetes, is a collection of metabolic illnesses that are defined by a high level of blood sugar (hyperglycemia) over an extended period of time. Diabetes is a popular name for diabetes mellitus. Symptoms often consist of increased urine frequency, an increase in thirst, and an increase in hunger. Diabetes may result in a variety of health problems if it is not properly handled. Acute consequences might manifest as anything from diabetic ketoacidosis to a condition of hyperosmolar hyperglycemia or even death. Some serious long-term problems include cardiovascular disease, stroke, chronic renal disease, foot ulcers, damage to the nerves, damage to the eyes, and cognitive impairment. Other long-term concerns include cognitive impairment.

Diabetes may be caused by either an insufficient amount of insulin being generated by the pancreas or an improper response by the cells of the body to the insulin that is produced. Insulin is a hormone that helps glucose from meals enter cells so that it may be utilized for energy. Insulin is responsible for this process. There are three primary forms of diabetes mellitus, which are as follows:

Loss of beta cells in the pancreas is the root cause of type 1 diabetes, which occurs when the pancreas is unable to generate enough insulin. The term "insulin-dependent diabetes mellitus" or "juvenile diabetes" was more often used to refer to this kind in the past. An autoimmune reaction is the root cause of the decline in beta cell numbers. It is yet unclear what causes this autoimmune reaction to occur. Diabetes type 1 may also develop in adults, despite the fact that the disease often manifests itself in childhood or adolescence.

Insulin resistance is the first step in developing type 2 diabetes. Insulin resistance is a disease in which cells fail to react appropriately to insulin. If the illness worsens, there is a possibility that insulin production may decrease. This kind of diabetes was once known as "adult-onset diabetes" or "non insulin-dependent diabetes mellitus." A large rise in the frequency of obesity among children has led to an increase in the number of instances of type 2 diabetes in younger individuals. Nonetheless, type 2 diabetes is more frequent in older persons than in younger people. The combination of having an unhealthy amount of body fat and not getting enough exercise is the most typical reason.

The third primary kind of diabetes is known as gestational diabetes, and it takes place when pregnant women who have never been diagnosed with diabetes before see an increase in their blood sugar levels. In most cases, a woman's blood sugar will recover to normal levels shortly after birth, even if she had gestational diabetes. On the other hand, women who have experienced gestational diabetes during pregnancy have a greater chance of getting type 2 diabetes in later years of their lives.

Injections of insulin are required for the management of diabetes type 1. Maintaining a nutritious diet, frequent physical activity, a normal body weight, and avoiding the use of tobacco are all important components in the prevention and management of type 2 diabetes. Oral antidiabetic medicines, either in combination with insulin or on their own, may be used to treat diabetes type 2. Patients with the condition should be cautious to keep their blood pressure under control and give their feet and eyes the attention they need. Insulin and several other drugs taken orally may lower blood sugar levels (hypoglycemia). Surgical weight reduction may be an effective treatment option for type 2 diabetes in patients who are obese and meet the criteria for the condition. Diabetes that develops during pregnancy often disappears following delivery of the baby.

Individuals diagnosed with diabetes have an increased likelihood of acquiring a retinal condition known as proliferative diabetic retinopathy (PDR). Neovascularization, a condition in which aberrant blood vessels are created on the retina, is one of the most prominent features of PDR. Neovascularization is a disease in which abnormal blood vessels are generated on the retina. If this problem is not diagnosed and treated promptly, it may lead to permanent vision loss. Detecting neovascularization in fundus pictures has been the subject of a significant number of research, each of which has presented a unique image processing approach. Neovascularization, on the other hand, is difficult to identify due to the random development pattern and tiny size of the affected area. As a result, deep learning approaches are gaining greater momentum in the field of neovascularization detection as a result of their capacity to carry out automated feature extraction on objects that include a complex combination of information. In this study, a technique for the identification of neovascularization that is based on transfer learning is presented [1, 3]. The most prevalent and sneaky microvascular consequence of diabetes is called diabetic retinopathy (DR), and it may grow asymptomatically until a sudden loss of vision occurs. Although though DR is common in today's world, finding effective ways to avoid it is difficult [4]. The diagnosis of type 2 diabetes at an early stage is essential for preventing the condition.

II. LITERATURE SURVEY

U. Ahmed et al.,[1] proposes a model for the prediction of diabetes that makes use of a combined machine learning technique. Two distinct kinds of models make up the conceptual framework: support vector machine (SVM) models and artificial neural network models (ANN models). These models do an analysis on the dataset in order to evaluate whether or not a diabetes diagnosis is positive. The dataset that was used for this study was segmented into training data and testing data with a proportion of 70:30 correspondingly between the two. The output of these models is used as the input membership function for the fuzzy model, and the ultimate determination of whether a diabetes diagnosis is positive or negative is made by the fuzzy logic. The combined models are saved in a cloud storage system for usage at a later time. The fused model determines whether or not the patient has diabetes based on the patient's real-time medical information. The fused machine learning model that was developed has a prediction accuracy of 94.87, which is greater than the approaches that were previously published.

A. Anaya-Isaza et al., [2], who implemented a total of twelve distinct data augmentation techniques, four of which are considered to be standard while the other eight are considered to be newly developed approaches. According to the findings, both the suggested approach and the traditional way improved the performance of the network, to the point where a detection rate of one hundred percent was accomplished by weighting the DM probability percentages for both pictures of the foot. In conclusion, it was feasible to illustrate the significance of transfer learning, which does not rely on the kind of database but rather on the data corpus with which the transfer was trained. This was a successful endeavor.

M. C. S. Tang et al.,[3] The effectiveness of the transfer learning approach is evaluated with the assistance of four pre-trained CNN models. These models are named AlexNet, GoogLeNet, ResNet18, and ResNet50. In addition to this, a better network that is built on the combination of ResNet18 and GoogLeNet has been suggested. The suggested network was evaluated using 1174 retinal image patches, and the results revealed that it was capable of achieving accuracy levels of 91.57%, sensitivity levels of 85.69%, specificity levels of 97.44%, and precision levels of 97.10% accordingly. We provided evidence that the suggested strategy is superior than the performance of each individual CNN when it comes to neovascularization detection. In addition to this, its performance is superior than that of an alternative approach, which made use of deep learning models for feature extraction and a Support Vector Machine (SVM) for classification.

M. Bernardini et al.,[4] This research had many goals, the first of which was to predict the probability of acquiring DR as a diabetic consequence (task 1) and the second of which was to temporally stratify the DR risk (task 2) using data from electronic health records. In order to accomplish these goals, a novel preprocessing procedure was developed to select both control patients and patients with pathological conditions. In addition, a novel fully annotated and standardized 120K dataset containing information from multiple diabetologic centers was made available. The Extreme Gradient Boosting model does offer satisfactory predictive performance; however, the Random Forest model obtained the best predictive performance to solve tasks 1 and 2, reaching the best Area Under the Precision-Recall Curve of 72.43% and 84.38%, respectively. This was the case even though the Extreme Gradient Boosting model offers satisfactory predictive performance. In addition to that, the relevance of the characteristics as determined by the top Machine Learning (ML) models is presented.

P. Nuankaew et al.,[5]. The performance of several machine learning-based prediction algorithms, such as K-Nearest Neighbors, Support Vector Machines, Random Forest, and Deep Learning, is evaluated and compared with regard to both datasets' ability to make accurate forecasts. The results of the comparison indicated that the suggested technique produced an accuracy of 93.22% for Dataset 1 and 98.95% for Dataset 2, respectively. These percentages are greater than those offered by existing machine learning-based methods.

S. Samreen et al.,[6] the classification is carried out on the preprocessed dataset using a wide range of different heterogeneous classifiers. These include 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. Using the metrics of accuracy, precision, recall, F1 Score, and area under Receiver Operating Characteristic curve, the performance of each machine learning pipeline is evaluated by the process of Repeated Stratified K-fold Cross Validation. The amount of features in the preprocessed dataset differs for each pipeline, and the Crowd Search method, which uses a stacking ensemble of many heterogeneous classifiers to attain the greatest accuracy of 98.4%, is responsible for this achievement.

N. Fazakis et al.,[7] to forecast the risk of diabetes includes particular applications, evaluations, and incorporations of the Knowledge Discovery in Database (KDD) approach. In particular, the development of datasets, the selection of features, and classification via the use of a variety of Supervised Machine Learning (ML) models are taken into consideration. It has been suggested that the ensemble WeightedVotingLRRFs ML model, which currently has an Area Under the ROC Curve (AUC) score of 0.884, might enhance diabetes prediction. When it comes to the system of weighted voting, the ideal weights are determined by their respective corresponding The sensitivity and area under the curve (AUC) of the machine learning model that was built on a genetic algorithm with two objectives. In addition, a comparison analysis is offered between the Finnish Diabetes Risk Score (FINDRISC), which uses inductive learning, and the Leicester risk score system, which uses transductive learning. Many machine learning models are also included in this research. The studies were carried out with the use of information obtained from the database of the English Longitudinal Study of Aging (ELSA).

M. Shokrekhodaei et al.,[8] employ light sources that emit light at several wavelengths. It is possible for several wavelength measurements to adjust for inaccuracies caused by inter- and intra-individual variances in the components of blood and tissue. In this investigation, the transmission measurements of an optical sensor that was created specifically for this purpose are analyzed using 18 distinct wavelengths ranging from 410 to 940 nm. The findings demonstrate a strong association between glucose content and transmission intensity across all four wavelengths, with a value of 0.98. Five different machine learning approaches are being examined for their ability to predict glucose levels.

M. T. Islam et al., [9] research, clinical biomarkers are the most important factor in determining if a person has diabetes or is at risk of getting the condition. In this piece, we explore the possibility of using a unique deep learning architecture to determine whether or not a person has diabetes based just on an image of their retina. We develop a multi-stage CNN-based model called DiaNet with the help of a relatively small-sized dataset. This model is able to reach an accuracy level of over 84% on this task, and as a result, it is able to successfully identify the regions on the retina images that contribute to its decision-making process. This finding is corroborated by the medical professionals who are experts in the field.

J. Tulloch et al.,[10] provide a reference for potential topics of investigation in the future. The Recommended Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) standards were used to search PubMed, Google Scholar, Web of Science, and Scopus for publications incorporating ML and DFUs. In order for research to be considered for inclusion, they required to make reference to machine learning, DFUs, and disclose pertinent outcome metrics indicating the accuracy of ML algorithms.

A. H. Syed et al.,[11] who compared it to other models that had been constructed using data from the National Health and Nutrition Examination Survey (NHANES) and the Pima Indian Diabetes (PID) datasets. The results of the Chi-squared test and the 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. This was shown by the fact that the results showed that the exposures contributed significantly to the prediction of the Response variable (subjects at high risk of diabetes).

R. Sarki et al.,[12] gives a thorough examination of automated methods to the identification of diabetic eye illness from a variety of perspectives, including the following: I accessible datasets, ii) picture preprocessing techniques, iii) deep learning models, and iv) performance evaluation metrics. This survey presents a thorough description of diabetic eye disease detection methodologies, including state-of-the-art field approaches, with the intention of providing significant information into research communities, healthcare professionals, and people who have diabetes.

III. CLASSIFICATION

The condition known as diabetes may be broken down into the following broad categories:

Diabetes type 1 is characterized by autoimmune beta-cell death, which in most cases results in an extreme insulin shortage and may also manifest as latent autoimmune diabetes in adults.

Diabetes type 2 is characterized by a gradual decline in sufficient insulin release from beta cells, which typically occurs against a backdrop of insulin resistance.

Certain forms of diabetes that are brought on by other factors, such as 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 the use of glucocorticoids, in the treatment of HIV/AIDS, or after organ transplantation), are examples of these specific types of diabetes.

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

Both type 1 and type 2 diabetes are considered to be heterogeneous illnesses since the clinical presentation and development of the disease may vary greatly from patient to patient. At the time of diagnosis, it may not be possible to definitively categorize certain people as having type 1 or type 2 diabetes. Classification is essential for identifying the most effective course of treatment. Both types of diabetes affect people of both young and old ages, disproving the long-held belief that type 2 diabetes is a condition that only affects adults and that type 1 diabetes is something that only affects youngsters. The classic signs of diabetes type 1 in children include polyuria and polydipsia, and roughly one-third of these children also present with diabetic ketoacidosis (DKA) (2).

The start of type 1 diabetes in adults may be more unpredictable; they may not exhibit the traditional signs that are observed in children, and they may experience brief remission from the need to take insulin (3–5). Patients with type 2 diabetes may sometimes manifest with diabetic ketoacidosis (DKA), especially those of ethnic and racial minorities (7). It is essential for the provider to be aware of the fact that the categorization of diabetes type is not always simple when the patient presents, and that it is frequent for a wrong diagnosis to be made (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.). While challenges in recognizing the type of diabetes may arise in all age groups at the outset of the disease, the diagnosis becomes increasingly clear with time in those who have a shortage in beta cells.

IV. CONCLUSION

Diabetes mellitus is a condition that lasts for a long time and poses a significant threat to people's health all over the globe. The World Diabetes Federation estimates that there are presently 246 million individuals in the globe who suffer from diabetes, but that figure is predicted to more than double by the year 2025, reaching 380 million people. In addition, diabetes-related complications are responsible for 3.8 million fatalities worldwide each year. It has been shown that early detection of those at risk may avoid or postpone the onset of 80 percent of the problems that are associated with type 2 diabetes. Techniques from the field of machine learning may be used to investigate and make predictions about the diabetes categorization. The purpose of this study is to provide an overview of machine learning approaches for the categorization and prediction of diabetes. In the near future, we will build an effective machine learning or deep learning approach to create a more accurate prediction model for the diagnosis of diabetic disorders.

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.