Proceedings of the International Conference on Sustainable Computing and Smart Systems (ICSCSS 2023)
IEEE Xplore Part Number: CFP23DJ3-ART; ISBN: 979-8-3503-3360-2

Predictions of Diabetic Mellitus using ML Techniques: A Systematic Overview

Krishna Manaswini T
Dept. of IT
GLWEC
Hyderabad, India
manaswini077@gmail.com

Padmalaya Nayak
Dept. of CSE
GLWEC
Hyderabad, India
drpadmalaya2010@gmail.com

Vanam Sri Harshitha
Dept. of IT
GLWEC
Hyderabad, India
sriharshithavanam@gmail.com

Shreya Barlapudi
Dept. of IT
GLWEC
Hyderabad, India
shreyabarlapudi@gmail.com

2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS) | 979-8-3503-3360-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCSS57650.2023.10169244

***Abstract*—** Diabetes Mellitus is a severe, chronic disease that occurs when blood glucose levels cross certain limits. There are several factors such as age, obesity, hereditary diabetes, bad diet, high blood pressure, improper lifestyle, and lack of exercise that can cause diabetes mellitus. People suffering from diabetes have a high risk of heart disease, nerve problems, kidney diseases, strokes, eye problems, finger numbness, etc. Over the last few years, the healthcare industry has focused a lot to diminish and prevent the disease in different ways. But unfortunately, diabetes mellitus is increasing exponentially, and touching at any age people due to multiple factors like unhealthy food habits, environmental pollution, work pressure, lack of physical activity, ill sleeping habits, and so forth. So, there is a demand for researchers across the globe to put their effort and explore AI/ML techniques on real-time diabetes datasets to predict the pros. and cons. of human health. In this view, various ML techniques are applied in the healthcare sector on real-time datasets collected from various hospitals in order to predict diabetic Miletus and its complications. Based on this fact, the proposed objective is to present an overview of ML techniques that have been applied to the diabetic datasets that are discussed in isolation from each other.

Keywords— Diabetes Miletus; Diabetes Symptoms; Classification Model; ML Techniques

I. INTRODUCTION

Diabetes mellitus is a group of endocrinal diseases associated with reduced glucose acceptance levels developed due to deficiencies in insulin excretion, insulin action, or both [1]. Diabetes mellitus is a major health issue across the globe. As per American Diabetes Association (ADA) analysis report, these are the major criteria for diabetes; 1) level of glycated hemoglobin (HbA1c) > 6.5%; 2) blood glucose level with empty stomach >126 mg/dL, and; 3) blood glucose levels > 200 mg/dl, 2 hours after glucose tolerance test with 75g of glucose [1]. In 2019, the information provided by the IDF (International Diabetes Federation) shows that almost 463 million people around the globe are suffering from diabetes, and according to their analysis, it might outgrow to 51% by the year 2045. Furthermore, according to estimations, there is said to be one undiagnosed diabetes person for each diagnosed diabetes person [2]. Initial diagnosis and treatment of type 2 diabetes are the most applicable cautions which help in controlling further developments and complications in diagnosed diabetes [3]. Besides, Europe Simulation Model Study reveals that an early diagnosis reduces the risk of facing cardiac movements and also death [4]. Diabetes might be associated with severe diseases like strokes, heart attacks, eye diseases, kidney diseases, and many more. In the big data sector, large amounts of data hide various useful information that needs to be extracted and analyzed. Analyzing the risk value of patient health conditions using data mining techniques through artificial intelligence allows physicians/health experts to serve people physically as well as from remote locations in a better way. Detecting diseases at an early stage can help to be treated more easily and efficiently. Recent advancement in Artificial Intelligence is influencing all spheres of human life including the healthcare system. The increased cases of diabetes in India represent a huge burden to the healthcare authority and provide an opportunity in terms of the potential availability of data. This data can provide some insight by applying AI/ML techniques and finding some potential solutions for this problem. Keeping this view in mind, we have put our effort to summarize the ML techniques and propose a new model in the future which is not discussed in the paper. In summary, the main contributions of this research study are listed as follows:

- Studied the symptoms and categories of diabetics
- Studied the different types of datasets for prediction models
- Extracted ML techniques applied to diabetes datasets and presented them in a structured form (tabular form)

The remaining part of the study is designed as follows. Section 2 presents different types of diabetic patients and the symptoms of patients, Section 3 presents the state-of-art of current literature and Section 4 concludes the paper guiding toward future work.

II. TYPES OF DIEBETES AND SYMPTOMS

The current study presents that diabetes diseases can be categorized into 3 types. These are Type 1, Type 2, and Type 3. Type 1 (T1DM) is also known as autoimmune diabetes and is mostly a chronic disease. The under-30 age group (young adults) and children are generally diagnosed with this type of diabetes, although it can affect any group of people. It is commonly known as insulin-dependent diabetes, which

occurs because of insulin deficiency caused by the loss of pancreatic beta cells. The symptoms of Type 1 diabetes are as follows:

- Frequent Urination
- Increased Hunger
- Blurred Vision
- Unexpected Weight Loss

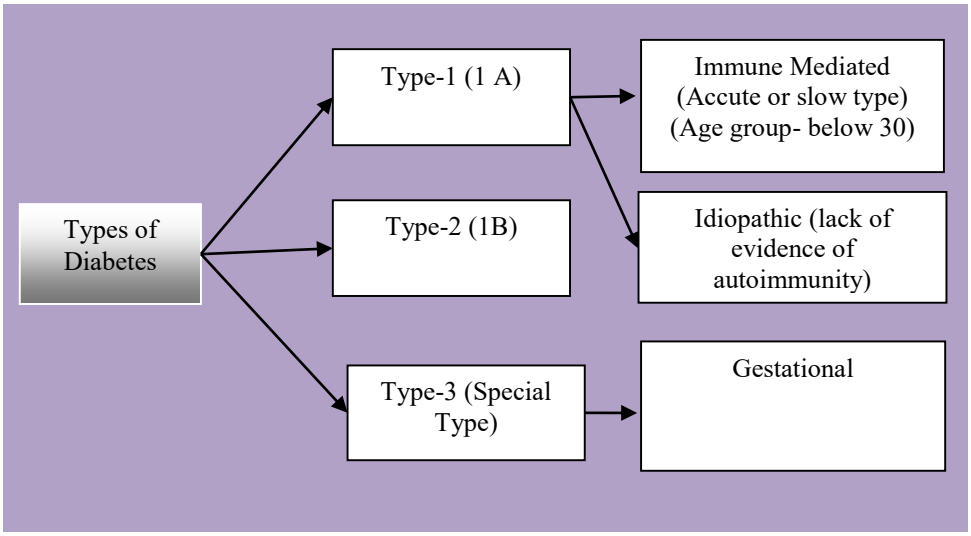

Fig. 1: Classification of diabetes

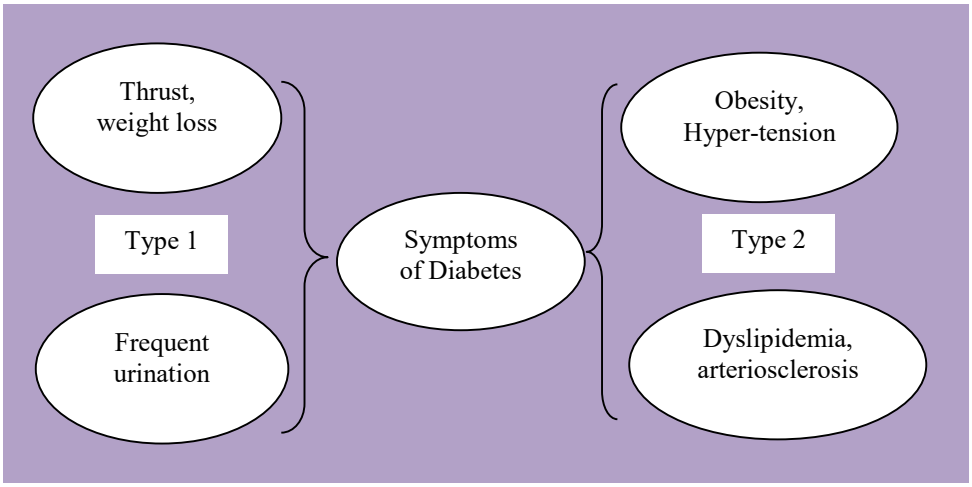

Fig. 2: Symptoms of Diabetes

The main reason for diabetes is the inability of the body to control blood sugar. Type 2 diabetes is recognized as non-insulin-dependent diabetes because body cells are not able to respond to insulin. Type 2 diabetes is also related to insulin resistance (broken insulin action) which causes body cells not to respond to insulin or to respond poorly. It affects directly their glucose consumption [5]. It occurs in both childhood as well as adolescence and has long complications associated with it. The early symptoms may not affect much but long-term high blood sugar can impose chronic injury and dysfunction of various tissues, especially the eyes, heart, blood vessels, kidneys, and nerves. Therefore, premature diabetes prediction is very much essential. The etiology and pathogenesis of diabetes are extremely multifaceted, and each one has a specific cause. Environmental factors like viral infections, chemical poisons, and dietary factors play an important role in pathogenesis. Other typical symptoms of Type 2 diabetes are as follows:

- Blurred Visualization
- Shortness of breath
- Slow wound healing
- Numbness and Skin Itching
- Sudden Misperception
- COMA, Periodontal disease
- Sexual dysfunction

Type-3 diabetes is not an officially recognized health disease. Some researchers have given their opinion on insulin resistance or insulin-like growth factor dysfunction in the brain causes Alzheimer's disease. More research has to be done on diabetes and its association with Alzheimer's disease. Few research studies convey that Gestational Diabetes can be considered a Type 3 diabetes that causes a sudden increase in blood sugar level in a pregnant woman which was detected earlier. These symptoms disappear automatically over the period once the delivery is over. A detailed description of diabetes millets is shown in Fig. 1. The

major symptoms of Type 1 and Type 2 diabetes are shown in Fig. 2.

III. PREVENTION OF DIABETES

Once a person is affected by diabetes, no one can be cured completely. However, this can be prevented in two ways. Usually, the first prevention involves the physical activity and control strategy suggested by health experts to patients with diabetes or prediabetes. As per the reference [5], physical exercise helps to increase the immune system of a person and other benefits to the human cell which is discussed in Fig. 3. The researchers of [6] surveyed the importance of exercises required for diabetes patients in Japan. In, [7], regular exercise has been suggested to control and prevent type 2 diabetes. In specific, the metabolic effect on tissues of diabetic patients has been studied and concluded significantly progresses in individuals who perform regular exercise. The second prevention is through periodical medical check-ups so that early precautionary measures related to medications can be taken to prevent diabetes.

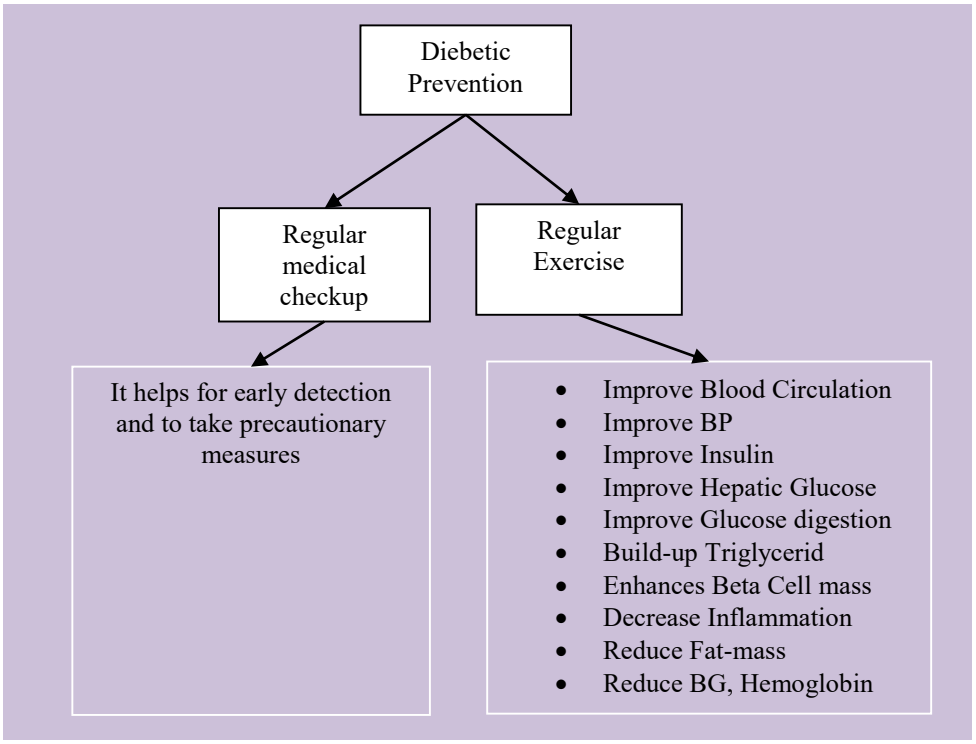

Fig. 3: Prevention of Diabetes

IV. RELATED WORKS ON THE DIABETES DATASET

Diabetes Mellitus is a global concern. The development of AI and data mining techniques has brought a global concern among researchers and practitioners to apply ML techniques to healthcare datasets and to develop prediction models in order to help healthcare experts to diagnose the disease at an earlier stage and take precautionary measures. In [8], the researchers implemented a system using Hadoop and Map Reduce techniques to analyze Diabetic data, to predict the type of diabetes and also the risks associated with it. In [9], classification techniques are used to study hidden patterns in the diabetes dataset. Two ML techniques such as Naïve Bayes and Decision Trees are used and their performance measures are compared. In [10], a data mining strategy is discussed to define the risk of T2DM using a set of machine learning models. The model performance is compared using a fixed number of factors versus a dynamic set of variables. The findings demonstrated the superior capacity of ML techniques to detect risk variables and forecast results across a variety of data and a growing number of factors which results in increased insights on illness risk variables without making a cautious assumption. In [11], the proposed approach employs a variety of classifiers including SVM, KNN, Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), and Gradient Boosting (GB) classifiers. Their results attained 77% accuracy. In [12], the prediction results of different algorithms were compared. The findings show that RF performs well when compared to other algorithms with a maximum accuracy of 97.5%, however, the J48 algorithm takes less time to build the model. In [13], eight machine learning models are applied for early risk prediction on the diabetes dataset. The 1-dimensional convolution neural network is the most successful one by providing 99.04% accuracy. The proposed model used a 5-fold cross-validation mechanism. In [14], the systematic review for type 2 diabetes is presented which is worked out based on machine learning and deep learning predictive models. This survey covers 18 different machine learning models applied in the healthcare system and is summarized in Table 1. In [15], classification and prediction methods are discussed for diabetes prediction in healthcare. The importance of BLE-based sensors and machine learning techniques for self-monitoring of diabetes mellitus in healthcare is emphasized

in particular. In [16], the authors have looked into the models that can predict more correctly on diabetes datasets. To achieve this, they followed some steps like the removal of missing values, data splitting, etc. They have selected a few classification and ensemble techniques like SVM, KNN, DT, LR, RF, and GB among to predict diabetes datasets. In [17], for diabetes prediction. In [18], a dataset of Diabetes Diseases Classification (DDC) has been introduced. A DDC pipeline has also been recommended by proposing a weighted ensemble classifier using various ML frameworks for classifying this DDC dataset. The overall summary is presented in Table 1.

Table 1. Summary of ML Techniques Applied on Diabetes Dataset

Sl.no	References and Year	Datasets Obtained from	Machine learning technique	Tools Used	Performance Measures
1	2018 [9]	The real-time dataset in Luzhou, China)	Decision tree, RF, NN, MCC	MRMR, PCA	Accuracy
2	2019 [10]	Clustering (ICRTAC)	LR, AdaBoost, RF, LDA, KNN	NA	Accuracy
3	2020 [11]	Real-time dataset (Henan Rural Cohort, China)	AUC, CART, LR, GBM, ANN, SVM, RF	NA	Accuracy, Confusion Matrix
4	2020 [13]	Pima Indian (from UCI repository)	SVM, KNN, Decision tree, LR, RF, GB	Python	Accuracy
5	2020 [14]	Training data (Sylhet Diabetes hospital in sylhet, Bangladesh)	DNN, SVM, KNN, DT, GBT, Naïve Bayes, LR, RF, MPL, LDA, CHR, LSR, etc.	Weka	Precision, Accuracy, F-Measure, Recall
6	2021 [15]	Real time datasReal-timet Diabetes Hospital in Sylhet, Bangladesh)	Decision tree, RF, SVM, XGBoost, KNN, Naïve Bayes, ANN	CNN	Accuracy, Precision, F-Score
7	2021 [16]	Real-time dataset (Pima Indian diabetes dataset)	LR, RF, Multiplayer Perception, Moving Averages, LR, LSTM		Accuracy
8	2022 [17]	Pima Indian (from UCI repository)	SVM, KNN, DT, LR, RF, GB	Python	Accuracy
9	2022 [18]	Pima Indian (from UCI repository)	SVM, Predicting diabetes	Python	Estimates expectation precision
10	2022 [12]	the Northeastern part of South Asia (Bangladesh)	AIML, ANN, ANOVA, LR, AdaBoost, RF, Extra Trees, LDA, KNN	DDC	Accuracy

SVM along with several other algorithms has been utilized

V. CONCLUSION

This study proves that diabetes patients are increasing exponentially on a daily basis. Any group of people is affected by this disease and it can affect any part of the human body system. So, there is a global alarm in the healthcare sector about how to minimize the diabetes millet. Along with healthcare experts and practitioners, data scientists from educational institutes as well as an industry putting their effort to eradicate the diabetes millet. This study has presented an outlook about the nature of diabetes disease and given some guidelines to prevent and detect the disease. As the applications of Artificial Intelligence (AI) are booming in all fields of engineering, this study is focused on the application of Machine Learning (ML) techniques in the healthcare field, which have been applied to the diabetes datasets during the last five years. This study will provide an overview of diabetes diseases as well as it will provide some guidelines to the young researchers to carry forward their research work in future.

REFERENCES

[1]. AD Association. Classification and diagnosis of diabetes: standards of medical care in diabetes-2020. Diabetes Care. 2019. {HYPERLINK "https://doi.org/10.2337/dc20-S002"}.

[2]. International Diabetes Federation. Diabetes. Brussels: International Diabetes Federation; 2019.

[3]. Gregg EW, Sattar N, Ali MK (2016). "The changing face of diabetes complications. Lancet Diabetes Endocrinol", ;4(6):537–47. https://doi.org/10. 1016/s2213-8587(16)30010-9.

[4]. Herman WH, Ye W, Griffin SJ, Simmons RK, Davies MJ, Khunti K, Rutten GEhm, Sandbaek A, Lauritzen T, Borch-Johnsen K, et al. (2015) Early detection and treatment of type 2 diabetes reduce cardiovascular morbidity and mortality: a simulation of the results of the Anglo-Danish-Dutch study of intensive treatment in people with screen-detected diabetes in primary care (addition-Europe). Diabetes Care, 38(8):1449–55, {HYPERLINK "https://doi.org/10.2337/dc14-2459"}.

[5]. Quan Zou, Kaiyang Qu, Yamei Luo, et al. (2018) "Predicting Diabetes Mellitus USING Machine Learning Techniques", Frontier in Genetics, Volume 9, Article 515.

[6]. D. Yang, Y. Yang, Y. Li, and R. Han, (2019)"Physical exercise as therapy for type 2 diabetes mellitus: from mechanism to orientation," Annals of Nutrition and Metabolism, vol. 74, no. 4, pp. 313–321.

[7]. J. Y. Kim and J. Y. Jeon,(2020) "Role of exercise on insulin sensitivity and beta-cell function: is exercise sufficient for the prevention of youth-onset type 2 diabetes?" Annals of Pediatric Endocrinology & Metabolism, vol. 25, no. 4, pp. 208–216.

[8]. Gauri D. Kalyankar, Shivananda R. Poojara and Nagaraj V. Dharwadkar,(2017) " Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop", International Conference On I-SMAC,978-1-5090-3243-3.

[9]. Aishwarya Mujumdara, Dr. Vaidehi Vb(2019) "Diabetes Prediction Using Machine Learning Algorithms" ICRTAC, Procedia Computer Science 165, Procedia computer science, pp. 1-8.

[10]. Liying Zhang et al.(2020) "Machine learning for characterizing the risk of type 2 diabetes in a rural Chinese population: the Henan Rural Cohort Study", scientific reports nature research, pp. 1-10.

[11]. Mitushi Soni,Dr.Sunita Varma(2020), "Diabetes Prediction using Machine Learning Techniques", IJERT, Vol. 9 Issue 09, pp.1-10.

[12]. P. Parameswari and N. Rajathi(2020) "Comparative Study of Machine Learning Approaches in Diabetes Prediction" Biosc. Biotech. Res. Comm. Special Issue Vol 13 No 11 ,pp. 1-5.

[13]. Özge Nur Ergün, Hamza Osman İlhan (2021) "Early Stage Diabetes Prediction Using Machine Learning Methods ", EJOSAT 29, pp. 1-6.

[14]. Fregoso-Aparicio et al. (2021), "Machine learning and deep learning predictive models for type 2 diabetes: a systematic review", Diabetology & Metabolic Syndrome (2021) 13:148.

[15]. Umair Muneer Butt, Sukumar Letchmunan, Mubashir Ali, et al.(2021) "Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications", Hindawi, Volume 2021, Article ID 9930985 pp. 1-17.

[16]. Mr.A.Hemantha Kumar1, Ms.R.Swetha2 (2022) "Diabetes prediction using machine learning techniques", Journal of Engineering Sciences, Volume 13, pp. 1-12.

[17]. Senthil Kumar M, Sundar Naveen Kumar C, Sivanarayanan M, Vijaya Kumar S (2022) "Diabetes prediction using machine learning" IJCRT, Volume 10, Issue 5 pp. 1-5.

[18]. Aishwarya Dutta, Md. Kamrul Hasan et al. (2022) "Early Prediction of Diabetes Using an Ensemble of Machine Learning Models" Int. J. Environ. Res. Public Health, Public Health,19, 12378. pp. 1-25.