Volume 12, No.12, December 2025

Journal of Global Research in Mathematical Archives

ISSN 2320 – 5822

UGC Approved Journal

RESEARCH PAPER

Available online at http://www.jgrma.info

A REVIEW ON DIABETIC DISEASE PREDICTION USING MACHINE LEARNING IN HEALTHCARE SECTOR

Dr. Manish Saraswat[1]

[1] Associate Professor (CSE) and Controller of Examinations, Faculty of Science and Technology, The ICFAI University, Himachal Pradesh

manish.saraswat@iuhimachal.edu.in

Abstract: Diabetes is a long-term illness characterised by high blood sugar levels as a consequence of either a lack of insulin production or an ineffective action of insulin. The current research outlines in detail the diverse types of diabetes, their symptoms, associated complications, and main risk factors, while also mentioning machine learning (ML) methods that are proficient at diagnosis and prognosis. Autoimmune destruction of β-cells leads to Type 1 diabetes, while Type 2 develops from insulin resistance and ultimately results in β-cell death; both processes are accompanied by oxidative and reductive stress. The major symptoms observed through clinical examinations, including eyesight issues, reduction of the body, difficulties with urination, and delayed recovery of wounds, all indicate the disease's slow nature. A prolonged period of high blood sugar is a cause of severe complications, which can be liver cirrhosis, NAFLD, NASH, and cases of liver disease due to alcohol. The paper also discusses different ML methods, including Decision Trees, Random Forests, AdaBoost, XGBoost, K-means, DBSCAN, and Autoencoders, both supervised and unsupervised, for the purposes of early patient detection, stratification, and risk prediction. By merging AI with medical science, the study has shed light on the role that ML-based systems can play in diabetes diagnosis, enriching it, supporting individualised treatment, and progressively reducing the global burden of diabetes mellitus.

Keywords: Diabetes prediction, machine learning, healthcare informatics, early diagnosis, predictive modelling.

1 INTRODUCTION

The healthcare industry has experienced extraordinary changes in the last several decades, as technological changes have been implemented, which focus on the better outcomes of patients, more accurate diagnosis, and simplification of medical operations [1][2]. The transformation of healthcare systems worldwide has been shaped by the rise in neoliberal economic policies since the late 20th century. As the Keynesian models of welfare state have been supplanted by new models of privatization, privatization has come to be a prevailing mode in the reorganization of official services [3]. Whereas privatisation is mostly justified on the basis of enhancing efficiency and quality of service delivery, controversy over its impact on accessibility, equity, and healthcare delivery services remains. Contrary to the case of industrial privatization, in the healthcare field, privatization does not necessarily imply a direct ownership transfer; rather, it is organized in numerous ways, including the redistribution of resources, changes in regulations, and the growth of the number of the private service providers [4], who are brought into the framework of the state financing.

Diabetes is one of the most chronic diseases in the world in which the sugar level of blood becomes too high. It has become a fifth-ranked disease for disease related deaths [5][6]. In cases of insulin deficiency or cellular insulin resistance, glucose is retained in the blood. This chronically high level of glucose in the blood leads to serious health problems such as damage of nerves, kidney failure, loss of sight, cardiovascular diseases and in some worst-case scenarios may end up causing premature death. Early diagnosis of diabetes can prevent complications and increase the chances of saving lives and reducing costs [7].

In machine learning, a branch of artificial intelligence, algorithms are used to analyze data and attempt to mine potential patterns in the data to predict new information [8]. As a new data analysis and processing method, machine learning has been widely used in many fields owing to its high precision, flexible customization, and convenient extensibility [9]. Complex nonlinear relational data can be easily handled with machine learning, which facilitates the discovery of the underlying mechanisms [10]. The excellent adaptability of machine learning has demonstrated its potential as a tool in the fields of environmental science and engineering in recent years [11]. Therefore, more accurate evaluation results can be expected despite the complexity of applying machine learning to water quality analysis.

1.1 Organization of the Study

The organization of the paper goes like this: Section 2 delivers information about diabetes and related clinical factors. Section 3 explains how machine learning techniques were in diagnosing diabetes mellitus. Section 4 lists the major risk factors for the development of diabetes. Section 5 explains the reasons for the use of machine learning and deep learning approaches in related literature through diabetes prediction. Finally, Section 6 presents the research conclusion and future directions for the study.

2 INSIGHTS OF DIABETIC DISEASE AND CLINICAL FACTORS IN THE HEALTHCARE SECTOR

Diabetes is a long-lasting disease that happens when the pancreas fails to create enough insulin, or when the body cannot use the insulin produced efficiently. Insulin is a hormone that controls the level of sugar in the blood[12].

2.1 Types of Diabetes Mellitus

Diabetes refers to a metabolic condition that is associated with the elevated sugar level in the blood. Early detection of diabetes could help individuals to manage and delay the progression of this disorder effectively [13]. Diabetes mellitus represents a group of physiological dysfunctions characterized by hyperglycaemia resulting directly from insulin resistance (in the case of type 2 diabetes mellitus—T2DM) [14], inadequate insulin secretion/production, or excessive glucagon secretion (in type 1 diabetes mellitus—T1DM).

2.1.1 Type 1 diabetes mellitus (T1DM)

T1DM also known as insulin-dependent diabetes, is the consequence of insulin deficiency arising from the progressive destruction of pancreatic β-cells through an autoimmune response. Histologic analysis of pancreas in a patient with T1DM showed infiltration of various immune cells, including T and B lymphocytes, macrophages, dendritic cells, natural killer cells, as well as islet-reactive autoantibodies and islet-reactive T-cells in the islets of Langerhans. The possibility of developing T1DM is associated with β-cells turnover or damage which leads to the release of autoantigens. In turn, β-cells auto-antigens are presented by the antigen-presenting cell (APC) to T-helper cells. In conjunction with major histocompatibility complex (MHC), APC will then migrate to the pancreatic lymph node. Autoantibodies and autoreactive T-cells will become activated in the presence of APC and they will be directed against β-cells auto-antigen. These activated T cells re-encounter cognate β-cell antigens and reactivate again, thus killing the β-cells.

2.1.2 Type 2 diabetes mellitus (T2DM)

T2DM is mainly linked to insulin resistance. The latter is majorly attributed to obesity, caused by poor dietary and lifestyle habits. The sensitivity of insulin fluctuates with intake of carbohydrate-rich foods, amount of physical activity, as well as stress signals. Obese individuals contain more adipose tissues, relating to the higher secretion of hormones and other substances that may increase the fluctuation of insulin sensitivity. Circulating non-esterified fatty acids (NEFA) in obese individuals is also associated with insulin resistance. A high-fat, hyperglycemic environment can lead to reduced insulin gene expression. Impaired function of cholesterol transporter destroys β-cells through sterol accumulation and islet inflammation [15]. Lipoprotein fractions and cholesterol metabolism contribute to β-cell failure. The low-density lipoprotein (LDL) that undergoes oxidation can decrease the pre-proinsulin expression in isolated β-cells, while very low-density lipoprotein can induce β-cells apoptosis. On the other hand, there is experimental evidence revealing the protective effects of high-density lipoprotein (HDL) on β-cells. They impact beneficially on glucose homeostasis by increasing pancreatic β-cells function and plasma glucose disposal.

Diabetes mellitus represents a group of physiological dysfunctions characterized by hyperglycemia resulting directly from insulin resistance (in the case of type 2 diabetes mellitus—T2DM) [16], inadequate insulin secretion/production, or excessive glucagon secretion (in type 1 diabetes mellitus—T1DM). Thus, similarly to both types of diabetes, particularly in the long term, insulin resistance and β-cell dysfunction/death may be present, impairing several tissues and cell function and metabolism. For instance, toll-like receptors (TLR) 7, TLR8, MyD88 and NLPR3 are responsible for T1DM disease predisposing (Figure 1).

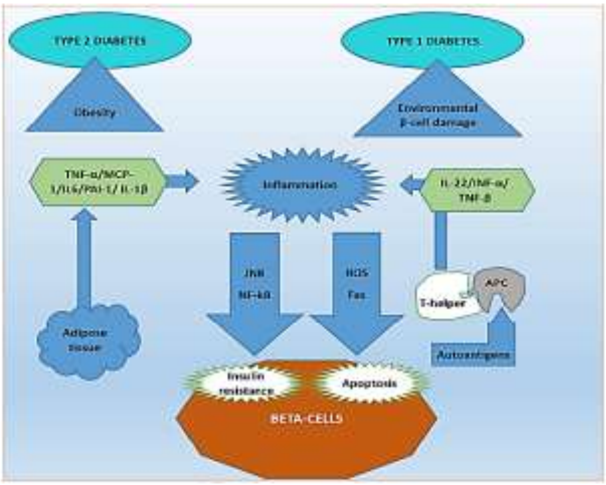

Figure 1: Inflammation as a common factor between T1DM and T2DM, which leads to βcells destruction

2.2 Symptoms of Diabetes Mellitus

The signs and symptoms of diabetes are disregarded by many because of the chronic progression of the disease. People do not consider this as a serious problem because unlike many other diseases the consequences of hyperglycaemia are not manifested immediately[17]. People are not aware that damage can start several years before symptoms become noticeable. This is unfortunate because recognition of early symptoms can help to get the disease under control immediately and to prevent vascular complications. The symptoms are given below:

- **Blurred Vision:** The blurry vision, seeing spots or dark objects, difficulty differentiating colors, the appearance of a dark spot or empty space in your vision, and having partial or total vision loss are all possible worrying signs that can be the manifestation of a hidden eye or health issue.
- **Urinary Tract:** An extremely prevalent and undesirable lower urinary tract symptom is nocturia. Age is a factor in nocturia frequency. Nocturia can lead to falls, fractures, and a rise in elderly people mortality in addition to disrupting sleep and lowering quality of life. According to one meta-analysis [18], nocturia increases the incidence of fractures by 32% and falls by about 20%. Additionally, a different meta-analysis has shown that nocturia is linked to a 1.27-fold increased risk of death.
- **Weight Loss:** It is also imperative to note that most drugs used to enhance glycemic control have a clinically significant effect on the body weight [19]. A network meta-analysis of glucose-lowering medications showed that, as a group, thiazolidinediones (TZDs) caused the most weight gain. verage increases of between 2.5 and 3.5 kg were induced by TZDs, in descending order, as follows; premixed insulins, basal/bolus regimes, sulphonylureas, prandial insulins, basal insulins, and meglitinides.
- **Slow wound healing:** There are various pathological alterations which are linked with diabetes and which lead to the poor healing of the wounds. Persistent hyperglycemia impairs vascularity and impairs adequate blood perfusion. Diabetic patients also often exhibit peripheral vascular disease and neuropathy, making wound detection difficult. Diabetic wounds are characterized by excessive inflammation, decreased angiogenesis, disrupted keratinocyte migration, and decreased fibroblast proliferation [20]. A combination of these changes leads to higher wound complications among diabetic patients such as infection, dehiscence of the wound and nonhealing wounds which are persistent.

2.3 Complications Associated with DM

Diabetes mellitus is the reason for a wide spectrum of complications both acute and chronic because of prolonged hyperglycemia. Diabetic ketoacidosis and hyperosmolar hyperglycemic state are the major acute complications, both of which can be life-threatening. Chronic complications are the results of long-term damage to blood vessels and nerves, which further leads to conditions like retinopathy [21], nephropathy, and neuropathy plus the higher risk of heart and brain attack due to the cardiovascular diseases. Incapable blood circulation could also lead to ulcers and infections in the feet, which untreated can result in amputation. However, if glycemic control is good and monitoring is regular, then these risks will be significantly reduced. A number of complications that are recorded in DM are listed below:

- **Liver Cirrhosis:** Liver plays a vital part in the metabolism of carbohydrate since it controls the level of blood glucose by following the glycogenesis and glycogenolysis pathways. Nonetheless, the metabolic processes of glucose in the liver can be disrupted by such diseases as insulin resistance and glucose intolerance and diabete0073. In cases of liver disease, both muscle and adipose tissue exhibit insulin resistance [22], and along with hyperinsulinemia, these factors are believed to be key underlying causes of diabetes. Additionally, as hepatitis C virus (HCV), alcohol, hemochromatosis, and NAFLD are frequently linked to DM, the liver disease etiology is significant in determining the occurrence of the condition. Since FBG levels may be normal in people with compensated liver cirrhosis, DM may be subclinical in these cases.
- **NAFLD:** NAFLD is a term used to describe a number of liver disorders, including cirrhosis, fibrosis, and simple steatosis. Fatty liver, considered to be the most benign form of NAFLD, is thought to affect one-third of adult Americans. The main fatty liver brought on by the buildup of fat is predominantly composed of triglycerides because of a metabolic syndrome caused by T2D, obesity, and dyslipidemia. Nonalcoholic steatohepatitis (NASH) is considered one of the severe manifestations of NAFLD that also causes tissue inflammation, cell death, and fibrosis. NASH is the most frequent make of cryptogenic cirrhosis currently. It is categorized as a condition that can lead to liver failure and cirrhosis. Figure 2 shows the diagram of NAFLD cause process.
- **NASH:** NASH is a complex manifestation of NAFLD. NASH is predisposed by obesity in visceral, hypertriglyceridemia, and insulin resistance. It is known that enlarged adipose tissue in a state of chronic inflammation linked to obesity secretes more adipokines. Cytokines released from the liver can induce some systemic effects such as hyperglycemia, hyperinsulinemia, and insulin resistance. These abnormalities prevent the liver from properly metabolizing lipids [23]. The most studied cytokine (TNF) directly stimulates the liver stellate cells leading to fibrosis of the liver. Losing body weight aids in the treatment of metabolic syndrome disorders like hyperlipidemia and fatty liver.
- **Alcohol-related Liver Disease:** In persons with alcoholic liver disease, diabetes is most likely to occur. This risk is inversely correlated with the amount of alcohol ingested, rising by a factor of two in patients who consume more than 270 g of alcohol per week as opposed to those who consume lower than 120 g/wk [24]. Following acute alcohol intake, an

increase in the insulin mediated glucose uptake is reduced significantly. Contrarily, those who are chronic alcoholics frequently get chronic pancreatitis and lose their pancreatic islet cells, which leads to DM.

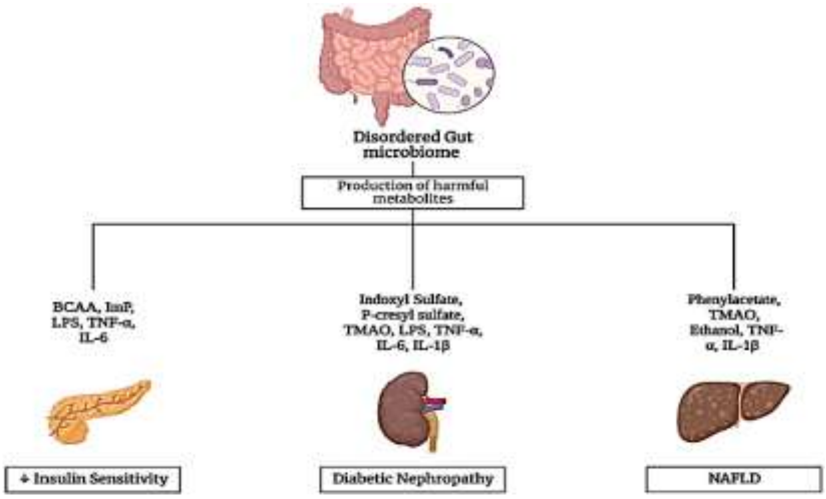

Figure 2: A schematic illustration shows how having more bacteria in the intestines causes NAFLD to proceed more quickly and cause more liver damage

3 MACHINE LEARNING IN DIAGNOSING DIABETES MELLITUS

High glucose levels in the blood are referred to as diabetes, and it has become a worldwide chronic medical condition in the past few decades. Machine learning is a branch of AI that pays attention to the development of computer systems [25], which can identify patterns in large amounts of data, which allows categorizing and predicting tasks. By utilizing the knowledge gained from big data analysis, these technologies help to develop more efficient and individualized methods of diagnosis and treatment [26]. This can subsequently allow medical practitioners to make better, timely decisions regarding a patient's care. The raised blood sugar prediction models are implemented using a set of various powerful machine learning algorithms.

3.1 Supervised Learning

Supervised learning is one of the methods used in machine learning where a model is first trained with labeled data, that is, inputs are linked with their respective correct outputs, which gives the model the chance to learn how to predict the output for new and unseen inputs as well [27][28]. The various supervised learning approaches are given below:

- **Decision Tree:** Decision tree is a supervised machine learning algorithm that can be utilized to build classification and regression models. It is marked as one of the simplest straightforward machine learning algorithms, and is based on arranging the features in a tree structure, and recursively splitting them based on chosen impurity criteria, such as entropy measure [29], Gini index, or information gain value.
- **Random Forest:** Random forest is also one of the most powerful supervised machine learning algorithms that are typically applied in classification problems. It is based on constructing a forest of multiple decision trees for a subset of the variables that are selected randomly in each tree. The prediction results of all the generated decision trees are aggregated to obtain the final refined output [30]. By using this ensemble technique, the random forest makes an improved performance by mitigating the high variance issue that is known as a common issue in the decision tree.
- **AdaBoost:** The AdaBoost algorithm which stands for adaptive boosting is a supervised machine learning algorithm used for building classification models, based on combining multiple weak classifiers to obtain more improved performance, the AdaBoost classifier commonly uses one level decision tree classifier. In AdaBoost, overfitting problems can be less likely to occur with it compared to other learning algorithms; however [31], it is not a suitable choice for datasets containing outliers and noisy data.
- **XGBoost:** The XGBOOST algorithm, short for extreme gradient boosting, is a supervised machine learning tree-based algorithm that is an improved version of its earlier gradient boost algorithm [32]. It can be applied for regression and classification problems, in particular for large datasets due to its high efficiency in generating accurate results and fast execution time [33]. The working approach of the XGBOOST is based on passing the outcome of a processed tree into the next tree sequentially.

3.2 Unsupervised Learning

This algorithm uses an unlabeled training set; therefore, it aims at identifying patterns and relationships in the data without any background information about the structure and categories [34][35]. The various approaches are explained below:

- **K-Means Clustering**: The K-means clustering is very significant in medical studies as it contributes to the discovery of specific patient subgroups, which would result in more individual treatment plans and better clinical results [36]. Recently, K-means clustering was used to evaluate the immune signature of juvenile-onset systemic lupus erythematosus (SLE) with the aim of improving the stratification of patients [37] employed K-means clustering to assess the immune signature of juvenile-onset systemic lupus erythematosus (SLE), aiming to enhance patient stratification.
- **Hierarchical Clustering:** One of the common methods of data analysis is hierarchical clustering. The tools used to perform this procedure normally work with data in its original [38], readable format which poses a question of privacy when a clustering exercise of sensitive data that [39] should not be disclosed is outsourced to another server.
- **DBSCAN:** DBSCAN algorithm has a high I/O overhead when the amount of data is large, which leads to low clustering efficiency. Many scholars have proposed corresponding improved algorithms to improve its efficiency. In order to extend to large-scale data clustering, a common method is to prune the search space of neighborhood queries in DBSCAN using spatial indexing technology (such as R-Tree) to improve clustering efficiency [40], but the efficiency of optimization algorithm still depends on data size.
- **Autoencoders (AE):** The basic AE is an auto-associative neural network, and it derives from the multi-layer perceptron, which attempts to reproduce its input, i.e., the target output is the input. An AE network can convert an input vector into a code vector using a set of recognition weights. Then, a set of generative weights are used to convert the code vector into an approximate reconstruction of the input vector. The basic AE can be used as a building block to train deep networks. Being associated with a basic AE [41], each level of a deep network can be trained separately.

4 PRIMARY RISK FACTORS IN DIABETES MELLITUS

Diabetes mellitus (DM) has emerged as an international health epidemic due to its rapid rise in prevalence. Consequently, scientists and or researchers will continue to find novel, safe, effective, and affordable anti-diabetic medications [42]. DM is now being treated as a disease which is manifesting an iceberg effect, i.e., the figures recorded only constitute a tiny fraction of the large number of unrecognized cases [43]. It is a systemic disease implying either the insufficient production of insulin in the pancreas or the inability of the body to utilize the production.

The primary risk factors of the DM are as explained below:

- **NADH and Reductive Stress:** Most of the electrons generated in the aerobic metabolism of glucose are stored in the NADH to form the adenosine triphosphate and oxygen reduction product. Consequently, Reductive Stress (RS) might result from an excess of NADH, which is a reducing molecule. RS is not the reverse of oxidative stress, but it leads to oxidative stress. Therefore, RS followed by oxidative stress comprises the main mechanism of hyperglycaemia-induced metabolic syndrome. Excessive generation of reactive oxygen species (ROS) and NADH is brought on by prolonged hyperglycemia. This inhibits glyceraldehyde-3-phosphate dehydrogenase (GAPDH) activity resulting formation of excessive ROS that further exacerbates oxidative stress. When complex-I in the mitochondria is overloaded with NADH, the mitochondrial electron transport chain produces more ROS. Additionally, excessive NADH can inhibit the glycolytic pathway, the pyruvate dehydrogenase complex, and the Krebs cycle, leading to greater glucose passage via the polyol pathway. As a result, glucose is reduced by the enzyme aldose reductase to form sorbitol, and the formed sorbitol is then converted to fructose by another enzyme sorbitol dehydrogenase.
- **Oxidative Stress:** OS plays a crucial role in the development of diabetes. More than half of diseases are secondary to excessive production of free radicals. OS condition arose owing to the high level of NADH, achieving the transition from RS to OS. The oxidation of cellular machinery is caused by free radicals, reactive entities constantly present in the human body. High level of nutrients (glucose and or fatty acid) in the blood leads to the generation of ROS. An increase in ROS levels ultimately leads to increased OS across a variety of tissues. Any imbalance between ROS and antioxidants leads to the production of a condition known as OS that leads to DM incidence. Intracellular stress-associated pathways are triggered when a cell and/or tissue is overtaken by OS. OS may itself potentiate the generation of ROS along with other pro-inflammatory cytokines and chemokines around the b-cells that disrupt the blood flow into the b-cells and abolish its function. b-cells dysfunction is induced by multiple risk factors.
- **Genetics and Insulin Resistance:** Genetics is among the major risk factors in the case of T2DM patients who have inherited genes from parents that enable their tissues resistant to insulin. Aside from that, stress response pathways inside cells are induced and inter-organ communication networks mediated by certain peptide hormones and cytokines cause insulin resistance (IR). Likewise, contribute to the risk of T2DM. IR is associated with defects in glucose uptake and oxidation, reduced glycogen synthesis, and a lesser fall in the capacity to suppress lipid oxidation. IR has a significant impact on skeletal muscle, adipocytes, and liver tissue, as they have a high metabolic demand.
- **Shared Risk Factors for TIDM and T2DM:** Despite the fact that the risk factors of T1DM and T2DM are not the same. An overlapped risk factor contributes to their likelihood. Figure 3 shows common risk factors for T1DM and T2DM. Inflammation is a risk factor that both T1DM and T2DM share because it kills b-cells. T1DM patients with damaged b-cells emit auto-antigens, to which the T-helper was exposed via antigen-presenting cells (APC) [44]. Active Thelper cells create cytokines that increase inflammation, which in turn causes ROS and Fas to be released, which cause b-cell death. On the same note, the adipose tissues in T2DM produce cytokines, which activate JNK and NFkB pathways that ultimately exacerbate inflammation and affect the insulin signalling in b-cells. APCs that deliver antigens, TNF-a and b (tumor

necrosis factor a and b), MCP-1 (monocyte chemo-attractant protein1), IL-6 (interleukin-6), IL-1b (interleukin-1b), and plasminogenactivator inhibitor-1 (PAI-1).

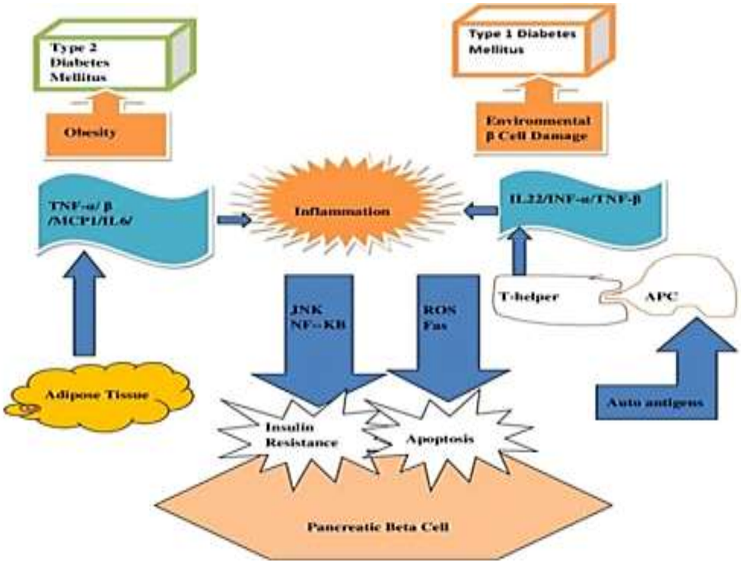

Figure 3: A diagram showing the common risk factor of inflammation for diabetes mellitus of both Types 1 and Type 2

5 LITERARTURE REVIEW

This section provides the comparative studies on the diabetes mellitus in healthcare sectors. It will show how the existing studies worked in detecting the diabetes at the earlier stage.

Sivakumar (2025) proposes a Multi-Task Learning (MTL) model to simultaneously predict multiple comorbidities on diabetic patients in order to enhance the overall capabilities in the earlier detection and individual healthcare treatment plan. By leveraging MTL, the proposed model will train to predict multiple comorbid conditions such as cardiovascular-related disease, kidney-affecting disease, and even neuropathy in a single unified methodology [45].

Nithvika *et al.* (2025) focused on the extent to which machine learning systems can help determine the risk of cardiovascular disease (CVD) based on diabetes status. So far, they have used extensive data to identify people with diabetes. After that, diagnosis of presence of CVD on separate diabetic-positive and diabetic-negative populations was done using Support Vector Classifier (SVC), Naive Bayes (NB), Decision Tree (DT), and Random Forest algorithms [46].

G *et al.* (2025) identified several drawbacks in existing methods for predicting Diabetic Kidney Disease (DKD) that limit their effectiveness and accuracy. Existing models are usually based on classic biomarkers which include blood glucose levels and urine albumin levels that are sometimes not enough to detect the disease at an early stage since they only show signs of kidney damage when the disease is at an advanced stage. Proposed a novel technique by utilizing machine learning algorithm such as navie Bayes to predict the DKD in early stage leads to decrease the mortality rate [47].

Apoorva, Krishna and Swamy (2024) note that complications such as diabetic retinopathy and nephropathy associated with diabetes mellitus have risen globally. These diseases however require early identification and intervention in the management process. In order to create a web-based system for prediction concerning diabetic complications, especially the first complication regarding diabetic retinopathy, CNNs were used [48].

Sharma and Sandhu (2024) Diabetic Eye Disease (DED) notably affects the retina of human eyes. It is an ophthalmic complication generally found in diabetic patients that may cause blindness incase not diagnosed at an early stage. For experimentation, the researchers use large-sized datasets consisting of retinal images. The careful training of these retinal images is very significant for training deep learning models so that signs of diabetic eye disease can be identified accurately [49].

In the proposed work, T et al. (2023) apply a variety of machine learning techniques to the patient health care dataset. The quantity of datasets and clinical factors utilized to train the algorithm determines the range of prediction accuracy levels. The goal of this study is to increase overall accuracy by combining machine learning algorithms. The suggested model uses the K-Nearest Neighbour, SVM, and Hybrid algorithm -GBRF -Random Forest mixed with Gradient Boosting that allows for hyper parameter fine tuning [50].

Table 1 provides the focus, findings, limitations, and future directions of the existing literature reviews for early prediction of the Diabetes Mellitus disease

Table 1: Summary of recent studies of machine learning techniques for diabetes prediction

Reference	Study Focus	Key Findings	Limitations	Future Work
Sivakumar (2025)	Multi-Task Learning (MTL) for predicting multiple comorbidities in diabetic patients	MTL enables simultaneous prediction of cardiovascular disease, kidney disease, and neuropathy; improves early detection and personalized healthcare strategies	Performance depends on balanced training data and feature correlations	Expand to additional comorbidities; integrate real-time clinical decision support
Nithvika et al. (2025)	Cardiovascular Disease (CVD) risk prediction across diabetic vs non-diabetic populations using ML	SVC, NB, DT, and RF were used; clear distinction observed in CVD risk patterns between diabetic-positive and diabetic-negative groups	Limited generalizability due to dataset constraints	Incorporate deep learning; explore larger and more diverse datasets
G et al. (2025)	Early prediction of Diabetic Kidney Disease (DKD) using ML (Naive Bayes)	Proposed ML-based early DKD prediction technique reduces mortality likelihood	Traditional markers detect damage late; model may underperform with sparse clinical data	Include imaging, genetic, or lifestyle features; compare more advanced ML methods
Apoorva, Krishna & Swamy (2024)	Web-based prediction system for diabetic complications using CNNs	CNNs effectively detect diabetic retinopathy for early intervention	Focuses mainly on DR; may not cover other complications deeply	Extend system to multi-complication prediction and telemedicine integration
Sharma & Sandhu (2024)	Detection of Diabetic Eye Disease (DED) from retinal images using deep learning	Deep learning accurately identifies DED signs when trained on large retinal datasets	Generalization limited if training images lack diversity	Improve model robustness; use multimodal imaging and ensemble DL models
T et al. (2023)	ML techniques for healthcare datasets using KNN, SVM, and hybrid GBRF	Hybrid GBRF model improves accuracy through hyperparameter tuning and model fusion	Performance varies with dataset size and clinical features	Implement automated tuning; test on real-world clinical environments

6 CONCLUSION AND FUTURE WORK

Machine learning and deep learning methods of diabetes prediction have demonstrated great potential in enhancing the diagnosis and management of diabetes at its earlier stages. This review work identifies diabetes mellitus as a multifaceted and progressive disorder of metabolism, primarily caused by the defects in insulin action closely associated with hyperglycemia of prolonged duration, and resulting in major systemic complications. The study, through the analysis of its types, symptoms and risk factors, indicates that there is a necessity for early detection and effective clinical interventions of the disease. The blending of machine learning methods—playing both supervised and unsupervised roles—opens up a channel for the great possibilities to be realized in the areas of diagnosing, predicting and treating patients' conditions with individualized plans. The latter investor-friendly and prompt doctor's to patients' interventions have been made possible through the usage of state-of-the-art computational techniques which in turn resulted in lessened suffering and better health care results. In a nutshell, the research emphasizes the need for collaboration between the medical experts and the artificial intelligence in order to cope with the global burden of diabetes more effectively.

The future work should be directed towards combining extensive, more heterogeneous patient data sets with cutting-edge machine learning approaches, thus improving the detection of complications at an early stage, as well as creating real-time clinical decision-support systems that will lead to more accurate, personalized, and scalable diabetes management solutions.

REFERENCES

[1] M. A. Mostafiz, “Machine Learning for Early Cancer Detection and Classification : AI- Based Medical Imaging Analysis in Healthcare,” *Int. J. Curr. Eng. Technol.*, vol. 15, no. 3, pp. 251–260, 2025, doi: https://doi.org/10.14741/ijcet/v.15.3.7.

[2] A. B. Singh, “Transforming Healthcare: A Review of Additive Manufacturing Applications in the Healthcare Sector,” *Eng. Proc.*, vol. 72, no. 1, 2024, doi: 10.3390/engproc2024072002.

[3] Y. Macha, “Cloud-Based CRM for Healthcare Data Management : A Review of HIPAA Compliance and Validation Rules,” *TIJER - Int. Res. J.*, vol. 10, no. 11, pp. 118–123, 2023.

[4] G. Sarraf, “Privacy-Preserving Blockchain for Healthcare : Addressing Security Challenges Through Decentralized Architecture,” *Int. Res. J.*, vol. 10, no. 11, pp. 111–117, 2023.

[5] R. P. Mahajan, “Early Prediction of Disease Using Machine Learning : Leveraging Medical Data for Accurate Classification,” *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 3, 2025.

[6] R. Ahuja, S. C. Sharma, and M. Ali, "A diabetic disease prediction model based on classification algorithms," *Ann. Emerg. Technol. Comput.*, vol. 3, no. 3, pp. 44–52, 2019, doi: 10.33166/AETiC.2019.03.005.
[7] M. M. Chowdhury, R. S. Ayon, and M. S. Hossain, "Diabetes Diagnosis through Machine Learning: Investigating Algorithms and Data Augmentation for Class Imbalanced BRFSS Dataset," *medRxiv*, 2023, doi: 10.1101/2023.10.18.23292250.
[8] S. B. Shah, "Evaluating the Effectiveness of Machine Learning in Forecasting Financial Market Trends: A Fintech Perspective," in *2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2025, pp. 1–6. doi: 10.1109/ICICACS65178.2025.10968297.
[9] V. Panchal, "Mobile SoC Power Optimization : Redefining Performance with Machine Learning Techniques," *IJIRSET*, vol. 13, no. 12, pp. 1–17, 2024, doi: 10.15680/IJIRSET.2024.1312117.
[10] S. Garg, "Predictive Analytics and Auto Remediation using Artificial Inteligence and Machine learning in Cloud Computing Operations," *Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.*, vol. 7, no. 2, 2019.
[11] M. Zhu *et al.*, "A review of the application of machine learning in water quality evaluation," *Eco-Environment Heal.*, vol. 1, no. 2, pp. 107–116, 2022, doi: https://doi.org/10.1016/j.eehl.2022.06.001.
[12] N. Pradhan, G. Rani, V. S. Dhaka, and R. C. Poonia, "Diabetes prediction using artificial neural network," *Deep Learn. Tech. Biomed. Heal. Informatics*, vol. 121, pp. 327–339, 2020, doi: 10.1016/B978-0-12-819061-6.00014-8.
[13] A. Ahmed *et al.*, "Machine Learning Algorithm-Based Prediction of Diabetes Among Female Population Using PIMA Dataset," *Healthc.*, vol. 13, no. 1, pp. 1–18, 2025, doi: 10.3390/healthcare13010037.
[14] P. G. Vinay Singh, DhirajKumar Pathak, "Integrating Artificial Intelligence and Machine Learning into Healthcare ERP Systems: A Framework for Oracle Cloud and Beyond," *ESP J. Eng. Technol. Adv.*, vol. 3, no. 2, pp. 171–178, 2023.
[15] S. Y. Tan *et al.*, "Type 1 and 2 diabetes mellitus: A review on current treatment approach and gene therapy as potential intervention," *Diabetes Metab. Syndr. Clin. Res. Rev.*, 2018, doi: 10.1016/j.dsx.2018.10.008.
[16] M. Krause and G. De Vito, "Type 1 and Type 2 Diabetes Mellitus: Commonalities, Differences and the Importance of Exercise and Nutrition.," Oct. 2023, *Switzerland*. doi: 10.3390/nu15194279.
[17] A. Ramachandran, "Know the signs and symptoms of diabetes.," Nov. 2014, *United States*.
[18] F. Yousuf, W. Elahi, A. Z. Syed, S. Kumar, I. Sharif, and A. Ali, "Urinary Clinical Manifestation In Type I And II Diabetes ; An Observational Study," *J. Pharm. Negat. Results*, vol. 14, no. 04, pp. 149–156, 2023, doi: 10.47750/pnr.2023.14.04.20.
[19] F. Losada-Díaz, S. Lizarazo-Bocanegra, J. J. Perdomo-Lugo, S. A. Gutiérrez-Romero, I. Correa-Osio, and C. O. Mendivil, "Differential Efficacy of Weight Loss Interventions in Patients with Versus Without Diabetes.," *Diabetes Ther. Res. Treat. Educ. diabetes Relat. Disord.*, vol. 15, no. 11, pp. 2279–2291, Nov. 2024, doi: 10.1007/s13300-024-01646-y.
[20] N. Dasari *et al.*, "Updates in Diabetic Wound Healing, Inflammation, and Scarring.," *Semin. Plast. Surg.*, vol. 35, no. 3, pp. 153–158, Aug. 2021, doi: 10.1055/s-0041-1731460.
[21] R. P. Mahajan, "Improvised Diabetic Retinopathy Detection Accuracy in Retinal Images Using Machine Learning Algorithms," *TIJER*, vol. 12, no. 3, pp. 155–161, 2025.
[22] A. Elkhalek *et al.*, "Managing diabetes and liver disease association," *Arab J. Gastroenterol.*, vol. 2018, 2018, doi: 10.1016/j.ajg.2018.08.003.
[23] E. Hashimoto, M. Taniai, and K. Tokushige, "Characteristics and diagnosis of NAFLD / NASH," vol. 28, pp. 64–70, 2013, doi: 10.1111/jgh.12271.
[24] J. Y. Kim *et al.*, "Chronic alcohol consumption potentiates the development of diabetes through pancreatic b-cell dysfunction," vol. 6, no. 1, 2015, doi: 10.4331/wjbc.v6.i1.1.
[25] P. B. Patel, "Predictive Maintenance in HVAC Systems Using Machine Learning Algorithms: A Comparative Study," *Int. J. Eng. Sci. Math.*, vol. 13, no. 12, pp. 106–120, 2024.
[26] G. Dharmarathne, T. N. Jayasinghe, M. Bogahawaththa, D. P. P. Meddage, and U. Rathnayake, "A novel machine learning approach for diagnosing diabetes with a self-explainable interface," *Healthc. Anal.*, vol. 5, p. 100301, 2024, doi: https://doi.org/10.1016/j.health.2024.100301.
[27] A. Gogineni, S. K. Malaraju, S. K. Madishetty, and S. Narang, "Enhancing Telemedicine Services Through AI, Blockchain, and Cloud Computing Integration," in *Data Science and Big Data Analytics*, D. Mishra, X. S. Yang, A. Unal, and D. S. Jat, Eds., Cham: Springer Nature Switzerland, 2025, pp. 420–431.
[28] D. Patel, "Enhancing Banking Security: A Blockchain and Machine Learning- Based Fraud Prevention Model," *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, Dec. 2023, doi: 10.14741/ijcet/v.13.6.10.
[29] H. Sharma and S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 2094–2097, 2016.
[30] A. Ziegler and I. R. König, "Mining data with random forests: current options for real-world applications," *WIREs Data Min. Knowl. Discov.*, vol. 4, no. 1, pp. 55–63, 2014, doi: https://doi.org/10.1002/widm.1114.
[31] G. L. Cao Ying, Miao Qiguang, , Liu Jiachen, "Research Progress and Prospects of AdaBoost Algorithm," *J. Autom.*, vol. 39, no. 6, pp. 745–758, 2013.
[32] M. R. R. Deva and N. Jain, "Utilizing Azure Automated Machine Learning and XGBoost for Predicting Cloud Resource Utilization in Enterprise Environments," in *2025 International Conference on Networks and Cryptology (NETCRYPT)*, IEEE, May 2025, pp. 535–540. doi: 10.1109/NETCRYPT65877.2025.11102235.
[33] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. doi: 10.1145/2939672.2939785.
[34] V. Verma, "LSTM-based to Predicting Stock Market Trends Using Machine Learning and Sentiment Analysis," *Int. J. Res. Anal. Rev.*, vol. 8, no. 3, 2021.

[35] A. Trezza, A. Visibelli, B. Roncaglia, O. Spiga, and A. Santucci, “Unsupervised Learning in Precision Medicine: Unlocking Personalized Healthcare through AI,” *Appl. Sci.*, vol. 14, no. 20, 2024, doi: 10.3390/app14209305.

[36] S. B. Shah, “Advanced Framework for Loan Approval Predictions Using Artificial Intelligence-Powered Financial Inclusion Models,” in *2025 IEEE Integrated STEM Education Conference (ISEC)*, 2025, pp. 1–10. doi: 10.1109/ISEC64801.2025.11147327.

[37] G. A. Robinson *et al.*, “Disease-associated and patient-specific immune cell signatures in juvenile-onset systemic lupus erythematosus: patient stratification using a machine-learning approach,” *Lancet Rheumatol.*, vol. 2, no. 8, pp. e485–e496, Aug. 2020, doi: 10.1016/S2665-9913(20)30168-5.

[38] N. K. Pavan Nutalapati, Jayapal Reddy Vummadi, Suresh Dodda, “Advancing Network Intrusion Detection: A Comparative Study of Clustering and Classification on NSL-KDD Data,” *Int. Conf. Data Sci. Its Appl.*, pp. 880–885, 2025.

[39] L. Sokhonn, Y.-S. Park, and M.-K. Lee, “Hierarchical Clustering via Single and Complete Linkage Using Fully Homomorphic Encryption,” *Sensors*, vol. 24, no. 15, 2024, doi: 10.3390/s24154826.

[40] X. Li, P. Zhang, and G. Zhu, “DBSCAN Clustering Algorithms for Non-Uniform Density Data and Its Application in Urban Rail Passenger Aggregation Distribution,” *Energies*, vol. 12, no. 19, 2019, doi: 10.3390/en12193722.

[41] S. Chen and W. Guo, “Auto-Encoders in Deep Learning—A Review with New Perspectives,” *Mathematics*, vol. 11, no. 8, 2023, doi: 10.3390/math11081777.

[42] V. Shah, “Managing Security and Privacy in Cloud Frameworks : A Risk with Compliance Perspective for Enterprises,” *Int. J. Curr. Eng. Technol.*, vol. 12, no. 6, pp. 606–618, 2022.

[43] N. Malali, “Predictive AI for Identifying Lapse Risk in Life Insurance Policies: Using Machine Learning to Foresee and Mitigate Policyholder Attrition,” *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 5, pp. 1–11, 2025.

[44] B. A. Tegegne, A. Adugna, A. Yenet, Z. H. Teffera, G. A. Amare, and T. K. Zeleke, “A critical review on diabetes mellitus type 1 and type 2 management approaches : from lifestyle modi fi cation to current and novel targets and therapeutic agents,” no. October, pp. 1–20, 2024, doi: 10.3389/fendo.2024.1440456.

[45] N. R. Sivakumar, “Multi-Task Learning Approach for Simultaneous Prediction of Multiple Comorbidities in Diabetic Patients,” in *2024 International Conference on IT Innovation and Knowledge Discovery (ITIKD)*, 2025, pp. 1–5. doi: 10.1109/ITIKD63574.2025.11005343.

[46] P. Nithvika, G. Revathy, C. Nithiya, P. G. Palanimani, and S. Gokulraj, “Machine Learning for Comorbidity Prediction: Diabetes and Cardiovascular Disease,” in *2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 2025, pp. 336–340. doi: 10.1109/ICSCDS65426.2025.11167592.

[47] V. G, B. Rambabu, S. K. P. J, S. R, and K. S. Balamurugan, “Optimizing Diabetic Kidney Disease Diagnosis with Machine Learning: A Novel Approach,” in *2025 8th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, 2025, pp. 1–4. doi: 10.1109/IEMENTech65115.2025.10959447.

[48] A. Apoorva, C. N. Krishna, and E. S. C. Swamy, “Diabetic Complication Detection and Prevention using Convolutional Neural Networks,” in *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2024, pp. 1251–1256. doi: 10.1109/I-SMAC61858.2024.10714699.

[49] P. Sharma and A. K. Sandhu, “Prediction of Diabetic Eye Disease in Type 2 Diabetes Using Hybridization of Deep Neural Network-Based Approaches,” in *2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)*, 2024, pp. 48–53. doi: 10.1109/GlobalAISummit62156.2024.10947818.

[50] M. T, N. J. D, V. V. S, and K. V, “Machine Learning Based System for Nephrologist to Classify Diabetic Kidney Disease,” in *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, 2023, pp. 1–7. doi: 10.1109/ICCEBS58601.2023.10448895.

