



Machine learning for diabetes clinical decision support: a review

Ashwini Tuppad¹ · Shantala Devi Patil¹

Received: 17 October 2021 / Revised: 27 February 2022 / Accepted: 3 March 2022 / Published online: 13 April 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Type 2 diabetes has recently acquired the status of an epidemic silent killer, though it is non-communicable. There are two main reasons behind this perception of the disease. First, a gradual but exponential growth in the disease prevalence has been witnessed irrespective of age groups, geography or gender. Second, the disease dynamics are very complex in terms of multifactorial risks involved, initial asymptomatic period, different short-term and long-term complications posing serious health threat and related co-morbidities. Majority of its risk factors are lifestyle habits like physical inactivity, lack of exercise, high body mass index (BMI), poor diet, smoking except some inevitable ones like family history of diabetes, ethnic predisposition, ageing etc. Nowadays, machine learning (ML) is increasingly being applied for alleviation of diabetes health burden and many research works have been proposed in the literature to offer clinical decision support in different application areas as well. In this paper, we present a review of such efforts for the prevention and management of type 2 diabetes. Firstly, we present the medical gaps in diabetes knowledge base, guidelines and medical practice identified from relevant articles and highlight those that can be addressed by ML. Further, we review the ML research works in three different application areas namely—(1) risk assessment (statistical risk scores and ML-based risk models), (2) diagnosis (using non-invasive and invasive features), (3) prognosis (from normoglycemia/prior morbidity to incident diabetes and prognosis of incident diabetes to related complications). We discuss and summarize the shortcomings or gaps in the existing ML methodologies for diabetes to be addressed in future. This review provides the breadth of ML predictive modeling applications for diabetes while highlighting the medical and technological gaps as well as various aspects involved in ML-based diabetes clinical decision support.

Keywords Diabetes prediction · Machine learning · Clinical decision support · Risk assessment · Pre-diabetes

1 Introduction

Diabetes is a chronic metabolic disease marked by an alarming rise in the patient's blood glucose levels for long periods of time resulting in severe, life-threatening complications and premature death. It is primarily classified under two types as follows: (1) type 1 diabetes (T1D) occurs due to acute deficiency of insulin, a hormone secreted by pancreas that regulates the blood glucose levels. Known as an autoimmune disease, T1D involves the patient's own immune system destroying the insulin secretory β -cells of the pancreas thus requiring Insulin administration into the patient's

body for survival. (2) type 2 diabetes (T2D) is caused due to the ineffective use of insulin stemming from insulin resistance/inaction, which nowadays is found to be primarily linked with unhealthy lifestyle practice. T2D is the major contributor of the diabetes burden accounting for 90–95% of all diabetic cases. Earlier, age was believed to be the most prominent risk factor of T2D due to its high prevalence among the aged people. However, today T2D cases are reported in all age groups including adolescents, young adults and the elderly. The age adjusted prevalence rate of diabetes for adult population has doubled from 4.7% in 1980 to 8.5% in 2014 globally and by 2045, an estimated 629 million people would be diabetic, if no interventions to prevent the explosive growth of diabetes are implemented (World Health Organization 2016). Apart from its high prevalence, the adverse effects of diabetes on human health are many. The most common complications from persistent high blood glucose include diabetic retinopathy, lower limb amputation, kidney failure, cardiovascular illness and premature death.

✉ Ashwini Tuppad
tuppadashwini@gmail.com
Shantala Devi Patil
shantaladevipatil@reva.edu.in

¹ School of Computer Science and Engineering, REVA University, Rukmini Knowledge Park, Kattigenahalli, Bangalore, Karnataka, India

Out of 3.7 million deaths reported in 2012 due to high blood glucose, only 1.5 million deaths were directly caused by diabetes while remaining 2.2 million deaths resulted from its associated complications leading to all-cause mortality. Such widespread growth coupled with gradual but devastating health degradation have changed the perception of T2D to that of an epidemic silent killer, even though it is non-communicable. The clinical diagnosis of T2D is particularly challenging since the disease presents itself with very less or no obvious symptoms at all. The prevention and management of diabetes has emerged to be a pressing global healthcare concern. Thus, in WHO's Sustainable Development Agenda, global leaders have vowed to minimize at least one third of premature deaths resulting from four major chronic diseases including diabetes by the year 2030 (World Health Organization 2016, 2019).

Machine learning (ML) has emerged to be a promising field of computer science with wide variety of applications in domains like banking, aerospace, robotics, industry, education, enterprise, astronomy, agriculture, healthcare and so on. A sub-domain of Artificial Intelligence, it relies on learning from data by discovering inherent patterns and application of the newly acquired knowledge to solve problems over previously unseen data (Kavakiotis et al. 2017). The real-world utility of ML algorithms is very vast and diverse, capable of achieving highly complex tasks that may require human intelligence and expertise by intelligent data analysis. D'Angelo and Palmieri (2020a, b) proposed a genetic programming-based ML approach to automatically detect structural defects found in aerospace systems and further build reliable mathematical models for each defect. The models incorporated domain specific knowledge representing the interrelationships in aerospace structures for defect tracking that minimize the need for skilled human inspectors while providing reliable results. Another notable application of ML in challenging real-world premises is proposed by Rampone and Valente (2012), where the authors have used artificial neural networks to assess and forecast landslide hazard using hillslope features. The timely and accurate landslide prediction requires sound expertise to evaluate the degree of hill-slope instability. The proposed model produced promising results with less than 4.3% prediction error. Recently, ML models have been used for detection and understanding of SARS CoV-2 virus genomic pattern, responsible for the ongoing pandemic. In another study by D'Angelo and Palmieri (2020) a dynamic programming methodology to identify the nucleotide subsequences from genome data to recognize Spike glycoprotein pattern is presented. The extraction of nucleotide subsequences involved analysis of 5000 SARS-CoV-2 genomes to understand the spike protein, essential for drug/vaccine development and the results showed 99.35% recognition accuracy.

Healthcare industry today has witnessed data revolution, not only in the amount and variety of medical data generated, stored, analyzed and accessed at high speeds but also in terms of the number of applications built on top of such huge data, the utility and insights derived from it and the impact of such data driven applications to the society. From the past decade, ML is being increasingly applied to offer diabetes clinical decision support as well as enhance its self-management. Amongst them, the majority of research works focused on developing ML based predictive models aimed at its diagnosis, risk assessment, early prediction and prognosis.

In this paper, we present a review of ML research efforts aimed at providing T2D clinical decision support, particularly in the areas of its prevention and management by identifying the gaps with respect to medical and technological background from the literature review of recent, relevant papers. Section 2 provides a brief background of clinical diagnosis and manifestation of T2D. A summary of existing review studies and novelties of proposed study are also presented. Section 3 describes the Materials and Methods adopted for the review; Sect. 4 briefly presents the background of machine learning and general architecture of ML-based clinical decision support system; Sect. 5 discusses the gaps in diabetes knowledge base, guidelines and practice; Sect. 6 describes diabetes risk assessment in terms of traditional statistical risk scoring techniques and ML based risk assessment; Sect. 7 is devoted to ML-based diabetes diagnosis systems based on invasive and non-invasive features; Sect. 8 is related to ML prognostic modeling in two distinct areas- prognosis of diabetes incidence v/s prognosis to diabetic complications. Section 9 is a discussion of the body of research works referred in our review and the findings therein. Section 10 is a conclusion with future research directions. This article provides the breadth of ML applications and their utility in T2D care while reflecting on the potential gap areas to target for building accurate as well as clinically usable T2D predictive models.

2 Background

The clinical characteristic of T2D i.e., persistent increase in blood glucose levels beyond normal range is known as hyperglycemia. Currently, there are three laboratory tests to measure the blood glucose levels namely—fasting plasma glucose (FPG) test, oral glucose tolerance test (OGTT) or HbA1C test.

While FPG and OGTT tests measure blood glucose levels in fasting and post intake of glucose solution respectively, HbA1C indicates a summary measure of blood glucose levels from past 2–3 months. The normal blood glucose ranges in a healthy individual involve fasting blood glucose below 100 mg/dL, OGTT value below 140 mg/dL and HbA1C

Table 1 American Diabetes Association (ADA) diagnostic criteria for pre-diabetes and diabetes based on FPG, OGTT and HbA1C blood tests

Diagnostic test	Pre-diabetes	Diabetes
Fasting plasma glucose (FPG)	100 mg/dl \leq FPG < 126 mg/dl (impaired fasting glucose)	FPG \geq 126 mg/dl
Oral glucose tolerance test (OGTT)	140 mg/dl \leq 2 h BG < 200 mg/dl (impaired glucose tolerance)	2 h BG \geq 200 mg/dl
HbA1C	5.7% \leq HbA1C < 6.5%	HbA1C \geq 6.5%

below 5.7%. Table 1 below shows the American Diabetes Association (ADA) defined diagnostic criteria for diabetes as well as pre-diabetes measured in terms of FPG, OGTT and HbA1C tests. Pre-diabetes (also called Intermediate hyperglycemia) is a condition where blood glucose levels are above normal ranges but below diabetic thresholds and it is comprised of two components namely—impaired fasting glucose (IFG) and/or impaired glucose tolerance (IGT) (World Health Organization 2016, 2019).

The common symptoms of diabetes include frequent thirst, urination, blurred eyesight, unexplainable weight loss and slow healing of wounds. However, these symptoms in T2D are not easily noticeable or may be completely absent due to the gradual progression of disease. Recently many studies have reported that pre-diabetes in itself is marked by serious micro-vascular and macro-vascular complications similar to those of diabetes (Bowen et al. 2018; Yudkin 2016; Yokota 2017). The risk composition of T2D consists of various parameters, which can be broadly categorized into modifiable and non-modifiable risk factors. While age, ethnic predisposition, genetics, comorbid conditions like gestational diabetes fall under non-modifiable risks, dietary habits, physical activity, exercise, BMI, smoking etc. can be modified to reverse the risks. Some clinical parameters like elevated blood glucose, glycated hemoglobin (HbA1C), lipids, blood pressure, triglycerides, urinary protein indicators like urinary albumin and creatinine values are considered not only as strong evidence-based indicators of future diabetes risk but are routinely used to assess the severity of Diabetic complications in prevalent cases as well.

The disease dynamics of T2D is very complicated and thus, a vast pool of research contributions involving ML literature are devoted for the early prediction, risk assessment, classification, prognosis and self-management of T2D. Many researchers have reviewed such extensive literature to reflect various state-of-the-art developments (Abhari et al. 2019; Contreras and Vehi 2018; Fregoso-Aparicio et al. 2021; Kavakiotis et al. 2017). Abhari et al. (2019) presented a systematic review of applications of AI methods in T2D care involving machine learning, fuzzy logic, expert systems and knowledge-based systems. The authors reviewed 31 articles selected after initial screening and final scrutiny on the basis of AI method, algorithm used, health application (screening/diagnosis, risk assessment, treatment, complication), clinical variables, best performing algorithm and best

accuracy obtained. The findings of the review revealed that Naïve Bayes and support vector machine (SVM) may generate better predictive accuracy due to their suitability to the variables used in T2D diagnosis. They also stressed that current research works have focused more on classification of patients based on diagnostic outcomes of T2D, however very little attention is given to estimating the probability of outcomes in real life. In another work by Fregoso-Aparicio et al. (2021), a systematic review of ML and deep learning models for T2D is presented involving 90 studies. The review results summarized the dataset characteristics, complementary techniques used for data imbalance issue and feature selection, data sampling method, ML model adopted as well as the performance/validation metrics considered by each study. The authors reported that decision tree models produced optimal performance however, selection of best feature subset and data balancing are necessary. Deep neural networks although highly accurate were more suitable for large datasets only. Further, they suggested incorporating at least three performance metrics along with Area Under the Receiver Operating Characteristic Curve (AUROC), a robust indicator to deal with heterogeneity in features. Contreras and Vehi (2018) carried out yet another review of AI applications in the field of diabetes management and clinical decision support. Based on their prime purpose, a functional taxonomy classifying different AI methods into learning, knowledge discovery or reasoning methods is also presented. The review identified following main application areas—blood glucose control, Blood glucose forecasting, adverse event detection, determination and recommendation of Insulin bolus, risk classification, diet, exercise and faults detection and finally lifestyle and day-to-day support in diabetic patients. Kavakiotis et al. (2017) systematically reviewed ML and data mining techniques for diabetes in the areas of prediction and diagnosis, related complications, genetic and environmental background, and clinical care and management. They found supervised learning methods dominating the diabetes research efforts (reported in 85% of the studies considered) than unsupervised learning (15%) with SVM as the most popular classifier used. De Silva et al. (2021) proposed a protocol for carrying out systematic review and meta-analyses of ML predictive modeling studies for T2D on the basis of their use and performance effectiveness at clinical as well as community levels. The authors state that most of the existing systematic reviews in the area

of T2D have considered traditional modeling studies only, with very less attention given to ML studies. They laid out analysis of candidate studies at two levels- at primary level, discriminative power, calibration and classification accuracy are used and at secondary level, predictors, algorithms, validation degree and intended use of the model are suggested for inclusion.

We present a comprehensive review of ML research efforts for T2D clinical decision support aimed at its prevention and management. The review included selected papers from the medical literature as well as recent papers from ML literature to identify and highlight both medical as well as technological gaps in ML predictive models for T2D. Three main application areas namely risk assessment, diagnosis and prognosis of T2D using ML have been identified and individual studies in each category are reviewed based on the dataset characteristics (size, features used), methodology (consisting of pre-processing, feature extraction and classification steps), techniques/algorithms used in each step and their performance. The review results are organized as follows- (1) risk assessment using statistical risk scores and ML based risk models (2) diagnosis using invasive and non-invasive features (3) prognosis of T2D incidence and prognosis of T2D to associated complications.

3 Materials and methods

A literature search using PubMed and Google Scholar databases was conducted to identify relevant research papers in the area of ML-based diabetes clinical decision support.

3.1 Search strategy

The search criterion comprised of four keywords—‘Diabetes’, ‘Prediction’, ‘Machine Learning’ and ‘Clinical decision support’, where AND operator was used to combine the former keyword with the latter three keywords individually. Since many of the existing T2D research works involving ML did not mention the diabetes type considered in either title or abstract, the usage of ‘Type 2 Diabetes’ as a search keyword was not found to be effective. The search was conducted from 2015 to present with English as the publication language. Figure 1 depicts the search strategy that involves initial screening based on titles, followed by abstracts and finally full text read as shown in Redundant articles were excluded at each stage of search process.

3.2 Inclusion and exclusion criteria

This review primarily focuses on T2D, which is acquired mainly through lifestyle habits and is preventable to a great

extent. Articles proposed for type 2 diabetes only were included for the study and articles focusing on type 1 diabetes or gestational diabetes were eliminated. We have considered research involving clinical decision support for T2D aimed at its prevention and management that includes diagnosis, risk assessment and prognosis. The articles proposed for treatment prediction, self-management and related complications and their severity assessment are excluded. Majority of the selected papers are standard journal publications with few notable conference papers. Diabetes health reports providing important facts and figures about the disease were also included. Articles explaining redundant methodologies are excluded.

4 Machine learning for clinical decision support

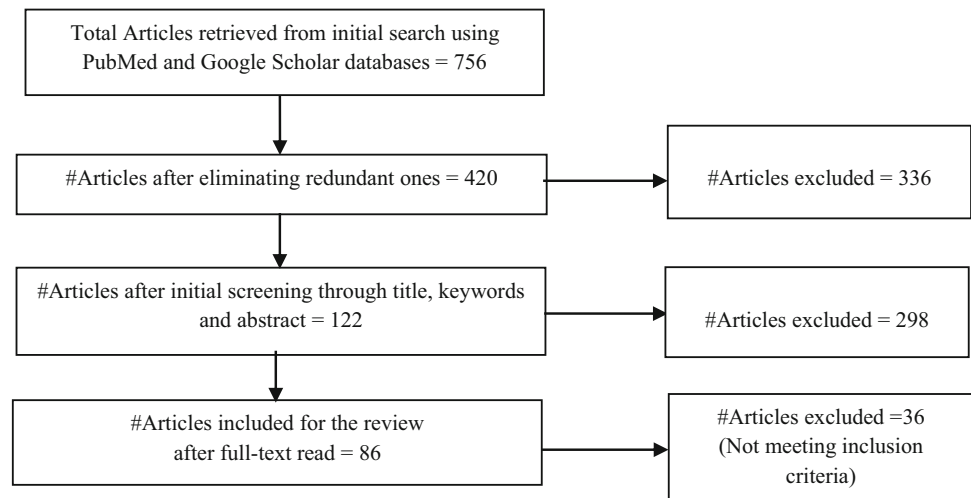
The tremendous growth of ML applications in healthcare and clinical decision-making is due to combined efforts of digitalization of healthcare sector, rise of big data technologies, ubiquitous computing and open access movements. Today, complex problems in healthcare domain are being approached with ML solutions due to the basic idea of learning from data and powerful learning algorithms that derive trends and generate predictions at big scales of data. This section provides an overview of machine learning and describes the general architecture of ML based clinical decision support system.

4.1 Overview

Machine learning relies on learning from example data (Dimopoulos et al. 2018), often from large datasets mimicking human learning capabilities. Traditional problem-solving using computer involved programming the logic required to solve specific tasks however, machine learning does not depend on explicit programming rather ML algorithms are built around generalizing the results from past, known examples by means of pattern recognition (Kaur and Kumari 2022). Medical predictive and diagnosis tasks require special expertise, case specific solutions and huge experience, which are fulfilled by machine learning algorithms (Dimopoulos et al. 2018). Learning is accomplished in two stages namely training and testing. While training involves learning the inherent patterns in the data through features, testing is used to test the algorithm performance with unseen data. Basically, ML algorithms are classified into three classes (Kaur et al. 2021) as follows:

1. *Supervised learning algorithms* make predictions by learning from labelled training data where each training example and its corresponding output (label) are fed

Fig. 1 Flow diagram of the process adopted for the search and scrutiny of research articles for the review



to the classification algorithm during the training phase and testing involves predicting the correct output label for new data.

2. *Unsupervised learning algorithms* involve learning from un-labelled training data usually by means of discovering hidden patterns in the data as well as clustering them based on their similarities.
3. *Reinforcement learning algorithms* involve learning in dynamic and more complex settings where the predictions are made by trial and error, and the learner is positively rewarded for correct predictions while being negatively rewarded (or penalized) for incorrect ones. Using this cue, the algorithm traverses in the direction of maximizing the overall positive rewards, thereby arriving at the correct output approximation.

Another hybrid mode of learning called semi-supervised learning involves a combination of both supervised and unsupervised learning over a dataset that consists of small portion of labelled training examples with remaining majority of the examples being un-labelled. It brings together the highly efficient and accurate generalization ability of supervised algorithms with realistic approximation provided by unsupervised ones. This approach is applicable in case of datasets containing large number of training examples with many features where generalization over such data is usually very difficult. In such context, supervised learning is applied over the small portion of the labelled training data and unsupervised learning is used to group the un-labelled training data into clusters of similar examples. Finally, the knowledge gained during supervised learning is utilized to generalize over clustered data. The recent trends of ever-increasing scale and complexity of data has given rise to another highly sophisticated learning paradigm called as deep learning (DL). A popular sub-domain of ML, DL is built over artificial neural networks with more than 3 hidden layers thus adding more

depth to the model as well as producing better optimizations of the predictions. The concept behind deep learning algorithms is to imitate the learning abilities of the human brain which involves several interconnections of neurons (a fundamental learning unit), of the range of millions or billions functioning all together in tandem. It is capable of representing complex and huge datasets and has gained a lot of momentum recently.

4.2 General architecture

Clinical decision support system refers to computerized healthcare systems that provide assistance/aid to clinicians, medical staff or patients themselves for the betterment of health and healthcare services, and implemented through modern technologies like AI, ML and data mining (Abhari et al. 2019; Kavakiotis et al. 2017; Sumathi and Meganathan 2019) by exploiting medical data (Sidey-Gibbons and Sidey-Gibbons 2019). When applied for disease prevention and management, clinical decision support using ML involves intelligent data analysis and inference to assist the physicians in clinical decision making (Barakat et al. 2010) by facilitating individualized risk profile study, tailor-made treatments, prognostic modeling of patient's health state etc. The general architecture of ML system for offering such clinical decision support with common constituent steps is depicted in Fig. 2 and explained as below-

1. *Data acquisition* Collection of healthcare data from available sources like hospital databases, online public repositories provided by research organizations, governments or scientists, healthcare surveys and audits data, data provided by insurance agencies, pharmaceutical companies and as such. The nature of the data depends on the problem and the objectives of the study. EHR, medical images from different modalities, prescriptions, data

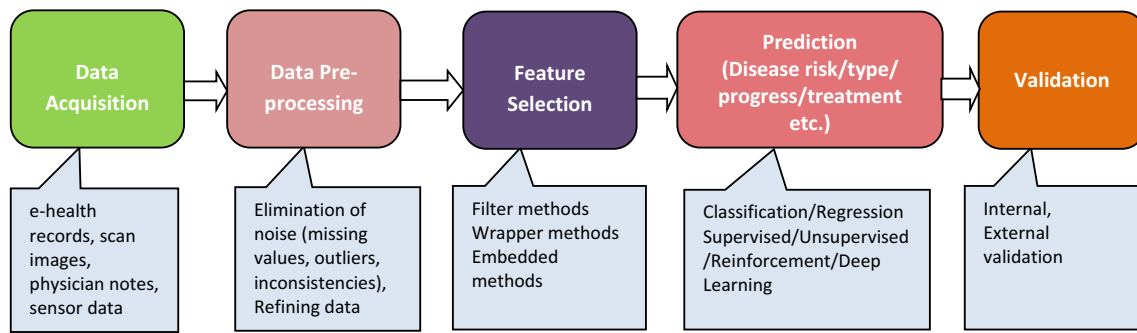


Fig. 2 General architecture of ML-based clinical decision support for disease prevention and management

collected by medical sensors are very popular and highly exploited sources of healthcare data in research studies nowadays.

2. *Data pre-processing* This refers to processing the raw input data and applying suitable transformations to refine it further for better or optimal results. Medical datasets are particularly characterized by lots of missing values, outliers as well as class imbalance. Elimination of such noise is accomplished by various pre-processing operations.
3. *Feature selection* Features are considered to be the building blocks of any ML algorithm, in general and are basically the intrinsic properties or characteristics of the dataset itself. Feature selection is a process of extracting relevant subset of features that best represent the various classes in the dataset as well as maximize the accuracy of the learning algorithm. There are three classes of feature selection techniques (Kaur et al. 2021)-
 - a. *Filter methods* Extract relevant features based on the inherent, statistical properties of the data to obtain a feature subset that is highly relevant with the target (output). Correlation, mutual information, variance are commonly used feature relevance measures in filter methods.
 - b. *Wrapper methods* Involve a search through the feature space and a classification algorithm to select a feature subset that results in maximum classification accuracy. Unlike filter methods, wrappers are considered to be highly dependent on the classification algorithm.
 - c. *Embedded methods* Comprise a combination of filter as well as wrapper methods where the search for optimal feature subset is an integral part of classifier training, rather than an independent step.
4. *Prediction* More commonly called as classification in ML terminology, it is comprised of two sub-steps namely—training and testing. Here, the original dataset is divided

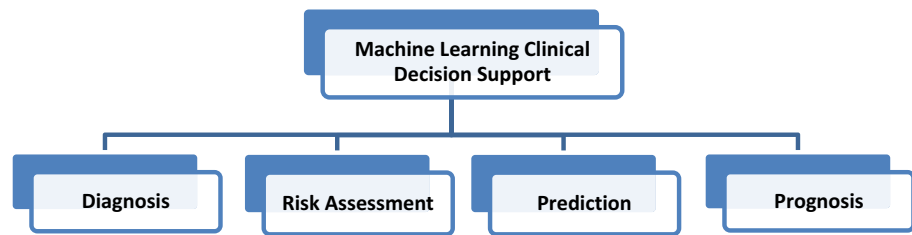
into training and testing datasets. During training, the training dataset is given as the input to the learning algorithm, which learns how to generalize or predict the outputs by means of pattern recognition. Testing involves determining the prediction performance of the learning algorithm over previously unseen examples from the testing dataset. The clinical decision support provided by ML predictive algorithms include prediction of disease risk, type, progress, treatment, complications and so on.

5. *Validation* Refers to the process of validating the classifier performance for accuracy and reliability of its predictions. Often, it is carried out on an internal test dataset, derived from the original dataset known as internal validation. When an external dataset, which is not a part of the input dataset is used to validate the results, it is called external validation. Validation is essential step, especially in diagnostic clinical decision support applications if it is to be adopted in clinical settings (Martinez-Millana et al. 2019; Noble et al. 2011).

ML-based clinical decision support applications can be categorized broadly into four classes—diagnosis, risk assessment, prediction and prognosis as shown in Fig. 3, that are not only limited to disease prevention and management but also aimed at healthcare service enhancement.

1. *Diagnosis* Refers to identification of a medical condition or disease based on the signs or symptoms. ML-based disease diagnosis can be used to develop mass screening tools, point-of-care portable diagnostic software packages for remote rural population, non-invasive or minimally invasive diagnosis systems for the elderly, specially-abled and unaffordable patients.
2. *Risk assessment* Objective assessment of risk factors driving the disease and their respective individual contribution to the disease progression. Prediction of disease incidence, progress, medication side effects, hospitalization, adverse events and so on include risk assessment as a crucial step.

Fig. 3 Classification of ML applications in clinical decision support



3. *Prediction* Forecasting of disease type, medical costs involved, potential treatment, medical facility requirement, cost v/s benefit estimates are some important problems in ML predictive models in healthcare.
4. *Prognosis* Refers to progression of a disease that marks the increasing severity and related health deterioration over time. Prognostic modeling of a disease, its related complications and future morbidities from a prior health state constitute ML prognostic clinical decision support.

5 Gaps in diabetes knowledge base, guidelines and medical practice

Both the incidence and prevalence rates of T2D have exponentially increased over the decades, and the very fact that diabetes has been considered a high priority chronic disease to be targeted by the world leaders itself indicates that it needs more investigation both in terms of medical advancements and technological interventions. The evolution of diabetes into newer subtypes, co-existing illnesses etc. have made risk-based patient classification particularly difficult. Many of these challenges stem from gaps in the medical knowledge surrounding T2D. This section highlights such gaps in diabetes knowledge base and resulting ambiguities in its guidelines and medical practice from related literature to identify the issues that can be addressed by exploiting ML predictive modeling.

In revised Diabetes classification by the World Health Organization (2019), it is reported that the conventional classification of diabetes into T1D and T2D based on age of its onset, amount of beta cell dysfunction, insulin resistance, diabetes related auto-antibodies and insulin dependency is no longer able to distinctly characterize each type as well as represent their phenotypes. It states that currently there are gaps in pathophysiology and etiology of diabetes. This is partly due to deviations in natural course of the disease itself, wherein recently T2D cases are observed in young adults and cases of T1D onset have surfaced in adulthood. Advances in molecular genetics has catalyzed the discovery of different variants of diabetes. In the wake of such changes in disease dynamics, patient specific risk assessment and targeted

intervention/treatment is absolute necessity to minimize the disease prevalence. In this regard, ML can be used to exploit the medical big data for determining relationships between different risk factors and their roles in diabetes onset and progression.

Yudkin (2016) highlights the issues around current pre-diabetes diagnostic criteria for prevention of diabetes. The author discusses the early onset micro and macro vascular complications of diabetes found in pre-diabetes stage and states that the majority of lifestyle intervention programs that enrolled prediabetic subjects at entry level may delay the incidence of diabetes but fail to reduce overall complications and mortality. Also, most of the subjects in such interventions have impaired glucose tolerance (IGT), with very few cases of impaired fasting glucose (IFG), but both being Pre-diabetes components. Further, HbA1C is used for pre-diabetes and diabetes diagnosis but it presents no evidence for management of pre-diabetes. The author thus suggests to revisit the current prediabetic diagnostic criteria since it could lead to overwhelming the medical system while no guarantee of long-term health benefits. In a counter point argument by Cefalu (2016), the author while acknowledging that existing pre-diabetes glycemic ranges do not lessen the complications related to diabetes apart from delaying its incidence states that is optimal for identifying intervention targets even at its lowest cut-off where risk is evident. The author suggests it can be considered for initial screening followed by evaluation with other risk factors like age, ethnicity etc. to detect likely subjects at lower thresholds of pre-diabetes to benefit from intervention. Previous two works thus indicate gaps in pre-diabetes diagnostic standards as well as risk assessment and reflect the scope for development of screening tools for pre-diabetes risks and its complications using ML to prevent T2D.

Tseng et al. (2019) investigated the awareness and implementation of Diabetes Prevention programs by Primary Care Physicians in US through pre-diabetes management. The findings of this study also report that there is a lack knowledge about pre-diabetes screening and clinical guidelines for its diagnosis and treatment. Reach et al. (2017) discussed a significant challenge in diabetes care called clinical inertia meaning delayed or lack of treatment intensification by clinicians at the right time. The authors described clinical

inertia at three levels namely- clinician level, patient level and system level. Among all the factors leading to clinician level clinical inertia, the over-estimation of medication side effects, difficulties in treatment progress tracking as well as uncertainty in spotting intervention candidates that can be attempted to be resolved using ML predictive modeling.

6 Diabetes risk assessment

Risk assessment is an integral part of any T2D prevention strategy that gives the opportunity to proactively minimize the future health burden by lifestyle and behavior modifications if the risk factors identified early on in the patient. Many risk scoring systems as well as risk assessment tools have already been proposed in the literature.

6.1 Traditional diabetes risk scores

A large number of research studies were undertaken to design risk scores specifically for certain ethnic groups derived from respective study populations (Aekplakorn et al. 2006; Chen et al. 2010; Gray et al. 2010; Hippisley-Cox et al. 2009; Lindström and Tuomilehto 2003; Mohan et al. 2005; Schulze et al. 2007) and they are quite popular. Risk scores are tools that generate a numeric score characterizing the severity of disease risk for an individual based on easily obtainable information. They are more useful in initial population-level screening to be followed up with more objective tests for diagnosis (Pei et al. 2019). Since they are derived from study populations of different size and characteristics, their applicability is limited to similar populations.

Table 2 presents a comparative analysis of popular risk scores proposed in the literature, their study characteristics and performances. In a systematic review of different T2D risk scores and models (Noble et al. 2011), it is reported that most of them require data from certain lab tests that are not universally available and clarity of information regarding inclusion of target subjects and practical application of risk score/model is lacking. Further, follow up studies to assess the long-term benefits and external validation are not undertaken in many. Very few works reported robust discrimination and calibration properties. The authors conclude that usability and impact of risk scores and models are quintessential for their clinical use. Another work presented by Bethel et al. (2017) investigated whether incrementally updated clinical data can improve diabetes prediction as against historical data by subjecting a baseline model used in NAVIGATOR study to historical as well as updated risk factors from one year follow up data. Results indicate improved prediction accuracy with updated risk factor model. A novel T2D risk score for Mongolian population is proposed by Dugee et al. (2015) by adopting risk factors from well-established FRINDRISC and

Rotterdam risk scores that were originally developed for different ethnic groups. Further, novel risk factors are identified using logistic regression (LR) analysis and the resulting risk score is validated on Mongolian subjects, which produced better results than the risk scores considered.

6.2 Machine learning-based diabetes risk assessment

In recent times, ML methodologies are increasingly used for risk assessment of T2D. Xie et al. (2019) applied different ML models including artificial neural network, logistic regression, decision tree, support vector machine, random forest and Gaussian Naïve Bayes for prediction of diabetes. Risk analysis was performed by univariate and multivariate weighted logistic regression to determine the association of individual risk factors with T2D outcome. The authors identified sleep time and frequency of medical check-ups as novel T2D risk factors. Boutilier et al. (2021) presented ML-based risk stratification of diabetes and hypertension for early prediction of high-risk subjects at short term 2-month interval. They used decision trees (DT), regularized logistic regression, k-Nearest Neighbors (kNN), Random Forest (RF) and AdaBoost classifiers as well as compared their classification results with five different feature sets. The findings indicated Random Forest to be the optimal classifier and it produced better results with an improvement of 35.5% in AUC for diabetes when compared with risk scores from the US and UK. Luo (2016) presented a work on ML-based diabetes risk prediction for T2D and demonstrated an automatic explanation method for explaining type 2 diabetes risk. The concept was based on building separate risk prediction and explanation models to avoid compromising the predictive accuracy and proposed a novel associative rule classifier for explaining the risk. The method was demonstrated using champion predictive model of Practice Fusion dataset containing eight boosted regression trees and four random forests combined with cubic splines additive model. 415 rules were obtained after proposed rule pruning technique from a total of 178, 814 initial associative rules, which were validated by clinical experts. Syed and Khan (2020) employed Chi-squared test and Binary Logistic Regression for statistical analysis and prediction of risk factors. They also developed web-based API to predict patient specific T2D risk score in real time using decision forest. The feasibility and performance of risk scores for diabetes using routinely collected data like EHR has been investigated by Kaur and Kumari (2022) by testing six risk score models. The authors report that risk scores may provide reliable results against EHR only if missing data and inconsistencies are handled appropriately. They adopted an existing Bayesian network imputation model in R package to address missing values issue. A genetic risk score combined with existing ML predictive models is used to evaluate

Table 2 Popular diabetes risk scores proposed in the literature with their characteristics

Authors	Risk score	Study population	Risk factors	Performance
Mohan et al. (2005)	Indian Diabetes Risk Score	CURES population (size = 2350)	Age, abdominal obesity, family history of diabetes and physical activity	Sensitivity = 72.5% Specificity = 60.1%
Aekplakorn et al. (2006)	Australian Diabetes Risk Score	AusDiab dataset (size = 6060)	Age, sex, ethnicity, parental history of diabetes, history of high blood glucose, use of antihypertensive medications, smoking, physical inactivity and waist circumference	Sensitivity = 74% Specificity = 67.7%
Schulze et al. (2007)	Thai Risk Score	Thai cohort (size = 2677)	Age, sex, BMI, waist circumference, history of hypertension, history of diabetes in parents or siblings	Sensitivity = 77% Specificity = 60%
Hippisley-Cox et al. (2009)	QDScore (England and Wales)	QResearch Database Version 19 (size = 2,540,753)	Age, sex, ethnicity, body mass index, smoking status, family history of diabetes, Townsend deprivation score (socioeconomic indicator), treated hypertension, cardiovascular disease, and current use of corticosteroids	Variance = 51.53% in women, 48.16% in men Discrimination, <i>D</i> statistic = 2.11 in women, 1.97 in men
Chen et al. (2010)	Finnish Diabetes Risk Score	Drawn randomly from National Population Register (size = 4746)	Age, BMI, waist circumference, use of blood pressure medication, History of high blood glucose, physical activity, daily consumption of vegetables, fruits, or berries	Sensitivity = 78% Specificity = 77%
Gray et al. (2010)	German Diabetes Risk Score	[EPIC]-Potsdam study (size = 25,167)	Age, waist circumference, Height, history of Hypertension, Intake of red meat, whole-grain bread, coffee, alcohol consumption, physical activity, smoking	Sensitivity = 82.8% Specificity = 72.2%
Pei et al. (2019)	Leicester Diabetes Risk Score	ADDITION-Leicester study (size = 6186)	Age, sex, ethnicity, BMI, waist circumference, first degree family history of type 2 diabetes, antihypertensive medication or high blood pressure diagnosis	Sensitivity = 72.1% Specificity = 54.1%

the improvement in discriminative ability for T2D prediction by Wang et al. (2021) and is compared against traditional Cox proportional hazards model. In a similar work, Liu et al. (2019) proposed two non-invasive risk scores derived from REACTION study dataset and combined with ML ensemble models are T2D prediction. Lee and Kim (2016) proposed Naïve Bayes and Logistic Regression algorithms for T2D risk factor identification using phenotypes and anthropometric measures. The association of various phenotypes comprised of anthropometric measures along with triglyceride (TG) values was studied using binary logistic regression. The authors particularly investigated the role of hypertriglyceridemic waist (HW) phenotype defined as [waist circumference (WC) + TG levels] in increased prevalence of diabetes against individual phenotypes and triglycerides. Peddinti et al. (2017) assessed the role of various metabolites to determine future diabetes risk by employing mass spectrometry-based technique and logistic regression analysis. Classification was achieved using regularized least squares modelling.

7 Machine learning for diabetes diagnosis

Clinical diagnosis of T2D is carried out by standard diagnostic blood tests of FPG, OGTT or HbA1C. All these tests require pricking a patient to collect blood samples and check for normalcy of blood glucose levels. When a patient is diabetic, he/she requires regular monitoring of blood glucose to keep stable health devoid of complications. It means going through invasive blood tests at regular time periods advised by the physician. Although, many non-invasive or minimally invasive medical kits to blood glucose measurement by patients themselves are available and popular in the form of strips or devices, the availability of these kits to people of every economic stratum is far from confirmed and patients still need to visit physician for further treatment or consultation (Pei et al. 2019). In order to find an alternative, many researchers are focusing on development of non-invasive diabetes diagnosis methods.

7.1 Diabetes diagnosis using non-invasive features

ML based non-invasive diabetes diagnosis has been the research objective of many studies recently reported in the literature. While some researchers focused on building rapid but accurate diagnosis tools for mass screening purpose without any dependency on expensive lab procedures, others focused on particularly eliminating the plight of patients going through physical discomfort of invasive tests by finding alternative methods. The more common non-invasive features belonged to patient demographics, physical examinations, anamnesis/patient's medical history, data collected through questionnaire and so on. Recently, toenails (Carter

et al. 2019; Nirala et al. 2019), tongue features (Zhang et al. 2017) as well as iris images (Samant and Agarwal 2018) have been utilized for non-invasive T2D diagnosis. Carter et al. (2019) considered chemical composition analysis of different elements found in toenails of diabetic subjects for diagnosis using ML predictive modeling. A different approach proposed by Nirala et al. (2019) exploited the toenail photoplethysmogram (PPG) signal of subjects, which is an electro-optical modality able to detect diabetes and the related changes therein based on the blood volumetric changes in peripheral vascular bed. The authors analyzed the PPG waveform features and its derivatives that are indicative of arterial stiffness and characteristic curve properties found in Diabetic patients. Zhang et al. (2017) adopted a theoretical diagnostic concept followed in Traditional Chinese Medicine that diagnoses diseases based on tongue inspection. They extracted texture and colour features from tongue images of subjects to detect diabetes based on peculiar characteristics. A soft computing methodology based on iris images is proposed by Samant and Agarwal (2018) with statistical, textural and discrete wavelet transforms features extracted as per the iridology chart. Zhang et al. (2021) applied maximum likelihood ratio technique to exclude lab parameters and collinear variables while including only non-invasive features for diabetes prediction. For classification, a combination of bagging-boosting with stacking algorithm was employed. A summary of various works proposed for non-invasive diagnosis included in the study are summarized in Table 3 below.

Performance of majority of the reviewed works was reported in terms of two common metrics—classification accuracy (ACC) and area under the receiver operating characteristic curve (AU-ROC/AUC). In binary classification, accuracy indicates the percentage of correct positive predictions against overall predictions. However, it may not be reliable in case of datasets with high imbalance. AUC is more precise indicator of classifier performance in terms of how well it can distinguish between different classes including positive as well as negative classes (discrimination). AUC also provides a means of optimal model selection when two or more models have similar accuracies. Few studies though have used metrics like Precision, Recall, Sensitivity, Specificity etc. We have considered accuracy and AUC for comparison of model performance unless in cases where not reported.

7.2 Diabetes diagnosis using invasive features

Considerable numbers of ML research efforts for diabetes diagnosis have relied on clinical features of patients that are collected by invasive lab procedures. The primary reason for inclusion of such features is that they provide solid evidence and medical context to characterize the diabetes that is universally established. Following are the brief account of such

Table 3 Summary of ML methodologies for diabetes diagnosis using non-invasive features

Authors	Non-invasive data source	Modality/device used	Dataset size	Features used	ML classifier	Performance
Pei et al. (2019)	Physical Examination Reports	–	4205	Age, gender, body mass index (BMI), hypertension, history of cardiovascular disease or stroke, family history of diabetes, physical activity, work stress, and salty food preference	J48 decision tree, AdaBoostM1, Sequential Minimal Optimization, Bayes Net, Naïve Bayes	Accuracy—0.9503 by J48 decision tree
Carter et al. (2019)	Toenails	Inductively coupled plasma mass spectrometry	40	Concentration of 22 elements found in toenails, age, gender, smoking history	Random Forests	AUC—0.90
Nirala et al. (2019)	Toenails	Photoplethysmography	141	Derivates of PPG signal waveforms	Support Vector Machine	Accuracy—97.87%
Zhang et al. (2017)	Tongue images	TDA-1 digital tongue instrument	296	Texture and color of tongue body and coating	Support Vector Machine	Accuracy—78.77%
Samant and Agarwal (2018)	Iris images	Through I-SCAN-2 Iris Scanner device	338	Statistical, textural and discrete wavelet transforms	Random Forests	Accuracy—89.63%
Zhang et al. (2021)	Retrospective Cohort data collected by survey	–	37,730	Demographic, medical and family history, anthropometric, lifestyle and diet predictors	Joint bagging boosting with stacking algorithm	AUC—0.885
Späinig et al. (2019)	Anamnesis and non-invasive physiological sensors	Speech synthesis and recognition system, ultrasonic sensor	4814	Age, gender, BMI, height, weight, waist circumference	Deep neural network and Support Vector Machine	AUC—0.84 and 0.703 by DNN (with and without HbA1C, respectively)
Farran et al. (2019)	Questionnaire data	–	1837	Age, sex, BMI, pre-existing hypertension, family history of hypertension	Logistic Regression, k-Nearest Neighbor (kNN), Support Vector Machine	AUC—0.83, 0.82 and 0.79 by SVM (for 3-, 5-, 7-year period respectively)

research works which describe a significantly novel or noteworthy methodology.

Zheng et al. (2017) proposed a semi-automated ML framework with feature engineering to detect borderline diabetics missed by conventional expert algorithms, by including self-reported diabetic symptoms and complications along with clinical features. A data-driven approach for prediction of cardiovascular disease, pre-diabetes and diabetes using exhaustive feature set comprised of demographic, dietary, physical examination, laboratory and questionnaire features is presented by Dinh et al. (2019). ML weighted ensemble model was employed for classification purpose which incorporated 131 variables for cardiovascular disease and 123 variables for pre-diabetes and diabetes, respectively. Bernardini et al. (2020b) sought to address the issues of high dimensionality, black box nature and overfitting, common to conventional ML models by proposing a novel sparse balanced support vector machine for diabetes detection. Least absolute shrinkage and selection operator (LASSO) regularization technique was employed to induce sparsity. Recursive rule extraction algorithm commonly applied for diabetes prediction yields a large rule set complicating the interpretability. Hayashi and Yukita (2016) proposed novel methodology combining Recursive rule extraction with J48graft classifier for diabetes prediction which generated a compact as well as interpretable rule set with 83.83% accuracy. Severeyn et al. (2020) considered k-means clustering applied over plasma glucose and insulin values of subjects to predict pre-diabetes and diabetes by analysis of the area under the glucose and insulin curves. Similarly, Abbas et al. (2019) proposed diabetes prediction using plasma glucose and insulin values obtained from OGTT test, with post load readings taken at three time points combined with patient demographics using SVM classifier. In a distinct work by Bernardini et al. (2020a, b), Triglyceride Glucose Index (TyG) a clinically significant but less exploited diabetes indicator is used to predict diabetes risk from historical EHR data using multiple-instance learning boosting algorithm. Stolfi et al. (2020) developed a computational model to simulate T2D patients to study the immunological and metabolic alterations expressed through clinical, physiological and behavioral features. They attempted to extend the EU-funded “Multi-scale Immune System Simulator for the Onset of Type 2 diabetes” (MISSION-T2D) model to build an economically-viable smartphone application for diabetes self-management. A summary of all the research works referred in this area are summarized in Table 4.

Casanova et al. (2016) presented a ML methodology for predicting diabetes in a high-dimensional data setup using Random Forest classifier and also performed variable ranking for risk estimation. Wearable medical sensors (WMS) are modern devices that contribute to medical big data nowadays have not been utilized much for integration into Clinical

Decision Support Systems. A novel work reported by Yin and Jha (2017) proposed such a health decision support system comprised of WMS and ML modules with ensembles of ML algorithms and demonstrated the feasibility for disease diagnosis using six categories including T2D. They also showed scalability and viable storage needs for implementation of such a system. A novel weighted average (soft voting approach) ensemble model is developed to improve diabetes prediction by addressing the low precision issue (Akula et al. 2019). Another study (Rahman et al. 2020) based on deep learning proposed a convolutional Long Short-Term Memory (ConvLSTM) for robust classification of Diabetic records and Boruta wrapper for feature selection. The proposed work reported 97.26% classification accuracy. Nguyen et al. (2019) presented a combined wide and deep learning model that can bring together the strengths of memorization and generalization power of wide and deep neural network, respectively to improve diabetes prediction. Deng et al. (2021) applied deep transfer learning to forecast blood glucose values in 5 min to 1 h interval along with data augmentation on Continuous Glucose Monitor readings from 30 min long time interval for T2D patients. Bala Manoj Kumar et al. (2020) proposed a deep neural network for diabetes prediction and feature importance concept using bagged Extremely Randomized Trees (ExtraTrees) and Random Forests for relevant feature selection. An ML-based diabetes self-management software is developed with blood glucose regulation based on diet management by Sowah et al. (2020). The core module of food classification and recommendation was implemented by Tensorflow network and kNN algorithm respectively and other modules comprised educational, medication reminder and physical activity tracker.

8 Prognostic modeling

Disease prognostic modeling involves modeling the progression of the disease into different stages, where successive stages correspond to more severe complications and health degradation. It aids in predicting future complication risk and treatment planning so as to reduce the severity. T2D prognostic modeling using ML can be classified under two categories—first, prognosis from prior health state (normoglycemia/related morbidity) to T2D incidence and second, prognosis of prevalent T2D leading to associated complications.

8.1 Normoglycemia/prior morbidity to type 2 diabetes incidence

Prognosis modeling to predict incident diabetes comprises of normoglycemia/prior morbidity as a precondition. Normoglycemia refers to normal blood glucose levels found in a

Table 4 Summary of ML methodologies for diabetes diagnosis using invasive, lab-based clinical features

Authors	Dataset size	Features	Feature selection technique	ML classifier	Performance
Zheng et al. (2017)	300	Demographic, self-reported symptoms and complications, diagnosis notes, Medication, Lab test parameters, disease diagnosis codes	By feature construction and correlation-based feature summarization	k-Nearest Neighbor, Naïve Bayes, J48, Random Forests, Support Vector Machine and Logistic Regression	Accuracy—0.99 by LR
Lee and Kim (2016)	11,937	Age, height, weight, BMI, FPG, SBP, DBP, TG, weight, patient phenotypes, anthropometric indices	—	Naïve Bayes and Logistic Regression	Accuracy—0.661, 0.735 by LR in men and women, respectively
Dinh et al. (2019)	21,091	Demographic, physical examination, dietary intake, lab test and questionnaire-based features	XGBoost Ensemble classifier	Random Forest, gradient boosted tree, logistic regression, support vector machine and their weighted ensemble model	Accuracy—95.7, 84.4% by XGBoost for diabetes and pre-diabetes, respectively
Bernardini et al. (2019b)	2433	Demographic (age, gender), monitoring (BP), clinical features (pathologies, exemptions, exam and drug prescriptions)	—	Sparse balanced support vector machine	AUC—81.43%
Hayashi and Yukita (2016)	768	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function	—	J48graft decision tree with recursive rule extraction algorithm	Accuracy—83.83%
Severeyn et al. (2020)	188	Insulin, blood glucose values obtained from 5-sample OGTT	—	K-means clustering	AUC—0.76
Abbas et al. (2019)	1492	Demographic features (age), insulin and plasma glucose before and after OGTT	Minimum redundancy maximum relevance	Support Vector Machine	Mean accuracy—96.80%
Bernardini et al. (2020a, b)	2433	Demographic (age, gender), monitoring (systolic and diastolic BP), laboratory features	—	Multiple Instance Boosting Algorithm	Accuracy—0.83, 0.70 with and without triglyceride-glucose index, respectively
Casanova et al. (2016)	3633	Demographic, anthropometric, blood markers, medical history, ECG data	—	Random Forests	AUC—0.82
Akula et al. (2019)	9948	Age, gender, systolic and diastolic blood pressure, height, weight, BMI	—	Weighted average ensemble model	Accuracy—85%
Rahman et al. (2020)	768	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function	Boruta wrapper	Convolutional long short-term memory	Accuracy—97.26%

Table 4 (continued)

Authors	Dataset size	Features	Feature selection technique	ML classifier	Performance
Nguyen et al. (2019)	9948	Demographic (age, sex), physical examination (body mass index (BMI) and blood pressure), physical, diagnosis codes, medication, lab variables	Manual feature engineering	Wide and deep learning ensemble model	Accuracy—84.28%, AUC—84.13%
Deng et al. (2021)	40	Demographics (age, gender), body compositions (body mass, height, body mass index), hormone levels, blood glucose data	—	Recurrent neural networks, gated convolutional neural networks, self-attention networks with transfer learning	Accuracy—89% AUC—0.86
Kumar et al. (2020)	768	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function	Extremely randomized trees	Deep neural network model	Accuracy—98.16%
Nilashi et al. (2017)	768	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function	—	Neural network	Accuracy—92.28%
Zhu et al. (2019)	768	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function	—	Principal component analysis + k-means + Logistic Regression	Accuracy—97.40%
Han et al. (2015)	7913	HbA1C, triglycerides, uric acid, high density lipoprotein, age, diastolic BP, cholesterol, waist, weight	Chi-square test, information gain-based filter technique, Random Forest	Support Vector Machine + Random Forests	Weighted mean precision and recall—94.2, 93.9%, respectively
Roopa and Asha (2019)	768	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function	PCA	Principal component analysis based Linear Regression Model	Accuracy—82.1%
Alharbi and Alghahhtani (2019)	200	Pregnancies, fasting plasma glucose, 1 h and 2 h post glucose load in OGTT, diastolic BP, HbA1C, BMI, age	Genetic algorithm	Extreme learning machine neural network	Accuracy—97.5%
Wang et al. (2019)	768	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function	—	Random Forests	Accuracy—87.1%
Mahboob Alam et al. (2019)	768	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function	PCA	K-means clustering, Random Forests, Artificial Neural Network for classification, Apriori algorithm for rule generation	Accuracy—75.7%

Table 5 Summary of ML methodologies for prognosis of normoglycemia/prior morbidity to incident type 2 diabetes

Authors	Prognosis modeling	Dataset size	Features	Feature selection technique	ML classifier	Performance
Yokota et al. (2017)	Pre-diabetes to diabetes	2105	Sex, age, BMI, smoking habit, physical activity, family history of diabetes up to third-degree relatives, systolic blood pressure, fasting and 1 and 2 h post-load plasma glucose during 75 g OGTT, HbA1c, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides and alanine aminotransferase	Stepwise backward elimination	Multivariate logistic regression	AUC—0.80
Perveen et al. (2019)	Metabolic syndrome to diabetes	4403	Sex, systolic blood pressure, diastolic blood pressure, high density lipoprotein, triglycerides, BMI, fasting blood sugar	–	Decision Tree and Naïve Bayes	AUC—79%
Anderson et al. (2016)	Normoglycemia to pre-diabetes and diabetes	24,331	Demographic, lab features, clinical observations, ICD-9 disease diagnosis codes, prescriptions	–	Ensemble model from reverse engineering and forward simulation platform	AUC—0.76 (normoglycemia to T2D and 0.70 (normoglycemia to pre-diabetes)
Cahn et al. (2020)	Pre-diabetes to diabetes	121,792	Demographics, diagnoses READ codes, medical interventions, hospitalizations, medications, laboratory test results	LASSO on Logistic Regression model	Gradient Boosted Trees	AUC—0.865 (internal validation); 0.907, 0.925 (in external validation datasets)

Table 5 (continued)

Prognosis from normoglycemia/prior morbidity to T2D incidence

Authors	Prognosis modeling	Dataset size	Features	Feature selection technique	ML classifier	Performance
Perveen et al. (2020)	Normoglycemia to T2D	1981	Age, gender, systolic and diastolic blood pressure, fasting blood glucose, triglycerides, BMI, high density lipoprotein, HbA1C	–	Gaussian hidden Markov model	AUC—86.9%
Choi et al. 2019	Normoglycemia (with cardiovascular risk) to T2D	8454	Sex, age, BMI, history of particular diseases, blood test results, pharmaceutical treatments for cardiovascular disease and T2DM	Information gain	Logistic Regression, linear and quadratic discriminant analysis, k-Nearest Neighbor	AUC—0.78 by Logistic Regression
Garcia-Carretero et al. 2020	Obese, hypertensive state to T2D	1647	Demographic, clinical and laboratory variables including those related to cardiovascular risk	–	k-Nearest Neighbor	Accuracy—0.977, AUC—0.89

healthy individual. Since the progression is gradual, prognosis of normoglycemia to future incident diabetes is significant to assess the inevitable risk posed by age as well as changing lifestyle habits. Certain prior health conditions or morbidities are closely related to diabetes onset like pre-diabetes, metabolic syndrome and hypertension. Further, these morbidities themselves are believed to add to diabetes risk. MeTS and cardiovascular disease have also been studied along with diabetes to investigate whether each one of them is responsible for the onset of the other (cause) or is the resultant condition arising from the disease (effect) (Perveen et al. 2019; Sumathi and Meganathan 2019; Choi et al. 2019). This section presents a discussion of ML-based diabetes prognosis works identified from the literature as follows—Yokota et al. (2017) developed risk scores for predicting Pre-diabetes to diabetes conversion using multivariate logistic regression analysis. They investigated the relationship of BMI difference (Final BMI – Initial BMI) during the observation period of 4.7 years with diabetes onset and concluded that patients with high BMI difference were more at risk and weight loss may substantially reduce the risk. In yet another work by Perveen et al. (2019), the relationship of different Metabolic Syndrome (MeTS) risk factors with incident diabetes is studied using logistic regression analysis. Naïve Bayes and Decision Tree algorithms were used for classification. The findings suggested that high levels of high-density lipoproteins (HDL) are strongly predictive of diabetes. Anderson et al. (2016) proposed a Reverse Engineering and Forward Simulation (REFS) analytics platform to predict pre-diabetes and T2D from a retrospective EHR dataset. In this approach, the REFS platform is used to build ensemble of predictive models based on Bayesian inference concepts. Cahn et al. (2020) applied gradient boosted trees model for modelling the progression of pre-diabetes to diabetes with Lasso on logistic regression used for feature extraction. Both internal and external validation of the results were performed. While an AUC of 0.865 was reported in internal dataset, external validation resulted in 0.907 and 0.925 AUC. Table 5 below summarizes existing research works adopting various ML methodologies for prognosis prediction of normoglycemic state or prior morbidity to incidence of T2D.

Prediction of incident diabetes over irregular, sparse longitudinal data is another challenge as many of the public medical datasets are longitudinally collected over different time periods where the recorded data may contain missing values (sparsity) and the time periods may be irregular. Perveen et al. (2020) attempted to solve this problem using Gaussian hidden Markov model (GaussianHMM) classifier. Novel polynomial approximation method based on Newton's Divide Difference method (NDDM) was employed to handle irregularity and sparsity.

8.2 Type 2 diabetes to related complications

The complications arising from diabetes are wide ranging affecting major organs and severe health degradation. Cardiovascular complications, Retinopathy, Nephropathy and Neuropathy are most common. Here, we describe the research efforts proposed for prediction of T2D complications identified from the literature review. Dalakleidi et al. (2017) proposed ensembles of artificial neural networks to estimate the 5-year risk of T2D and associated cardiovascular complications as a result. Dagliati et al. (2018) predicted diabetic retinopathy, neuropathy and nephropathy in T2D patients at different time intervals using Logistic Regression and stepwise feature selection. Clinically relevant Graphical nomograms were also presented. An evaluation of hybrid wavelet neural networks (HWNNs) and self-organizing maps (SOMs) to predict the onset of coronary heart disease and stroke as a long-term complication of diabetes was presented by Zarkogianni et al. (2018). Kowsher et al. (2019) proposed to prognosticate T2D and its treatment by applying various ML algorithms. Treatment prediction comprised of predicting the suitable medication based on the disease complications. Deep neural network produced the optimal performance at 95.14% accuracy among others. Aminian et al. (2020) used ML-based risk modeling of end-organ Diabetic complications in two groups of patients namely with and without undergoing metabolic surgery. Allen et al. (2022) predicted the risk of chronic kidney failure as a Diabetic complication using Gradient Boosted Trees and Random Forest for 5-year risk window. Sudharsan et al. (2015) developed ML probabilistic model using Random Forest, Support Vector Machine, k-Nearest Neighbor and Naive Bayes classifiers to predict hypoglycemic events from time-series blood glucose readings. A summary of existing ML prognostic methodologies proposed for predicting T2D complications is provided in Table 6 below.

9 Discussion

Our review was focused on ML prevention and management efforts specifically for type 2 diabetes. We drew out some observations and findings from the review study as follows—majority of the included works focused on developing early, rapid or minimally invasive T2D prediction systems, with risk assessment considered to be implicitly carried out while prediction. A host of research works such as those presented in (Battineni et al. 2019; Farran et al. 2019; García-Ordás et al. 2021; Kopitar et al. 2020; Lai et al. 2019; Lee and Kim 2016; Maniruzzaman et al. 2018; Olivera et al. 2017; Pei et al. 2019; Zheng et al. 2017; Zou et al. 2018) have evaluated the predictive ability of various classifiers for T2D prediction with comparative analysis and some others have carried

Table 6 Summary of ML methodologies for prognosis of type 2 diabetes to related complications

Paper	Prognosis modeling	Dataset size	Features	Feature selection technique	ML classifier	Performance
Dalakeidi et al. (2017)	T2D to fatal/non-fatal cardiovascular disease incidence	768 (Pima dataset) 560 (Hippokrateion dataset)	Number of pregnancies, plasma glucose after 2-h OGTT, diastolic BP, triceps skin fold thickness, serum insulin, BMI, age, diabetes pedigree function (Pima dataset) Age, diabetes duration, BMI, systolic and diastolic blood pressure, HbA1C, blood glucose, total cholesterol, triglycerides, high-density and low-density lipoprotein (Hippokrateion dataset)	–	Ensembles of artificial neural networks	Accuracy—92.86%, AUC—0.739
Dagliati et al. (2018)	Diabetic retinopathy, neuropathy and nephropathy	943	Age, gender, time to diagnosis, BMI, HbA1C, lipid profile, smoking habit, antihypertensive therapy	Stepwise feature selection based on Akaike information criterion	Logistic Regression	Accuracy—0.77, 0.69, 0.74, 0.80, 0.79 (retinopathy, nephropathy, neuropathy, respectively) AUC—0.80, 0.734, 0.799 (retinopathy, nephropathy, neuropathy, respectively)
Zarkogianni et al. (2018)	T2D to fatal/non-fatal cardiovascular disease incidence	560	Age, BMI, HbA1C, diabetes duration, pulse pressure, fasting blood glucose, total cholesterol, high density lipoprotein cholesterol, triglycerides,	–	Ensemble models based on hybrid wavelet neural network (HWNN) and self organizing maps (SOM)	AUC—71.48 by hybrid ensemble

Table 6 (continued)

Paper	Prognosis modeling	Dataset size	Features	Feature selection technique	ML classifier	Performance
Kowsher et al. (2019)	T2D to complications and treatment prediction	9483	Fasting and 2 h post load blood glucose, BMI, duration of diabetes, age, sex, blood pressure, high cholesterol, heart diseases, kidney diseases, medications	–	Deep neural network (optimal classifier) along with a number of ML algorithms	Accuracy—95.14%
Aminian et al. (2020)	T2D related coronary artery events, heart failure, nephropathy and all-cause mortality with or without metabolic surgery	2287	Sex, age, BMI, BMI category, race, smoking status, medical history, HbA1C, systolic and diastolic BP, estimated glomerular filtration rate (eGFR), triglycerides, medication history	–	Multivariable time-to-event regression and Random Forest	AUC—0.81 (all-cause mortality), 0.67 (coronary artery events), 0.75 (heart failure), 0.76 (nephropathy) by Random Forest model
Allen et al. (2022)	T2D to 5-year risk of multi-stage chronic kidney disease (CKD)	62,994	Age, sex, BMI, systolic and diastolic blood pressure, blood urea nitrogen, creatinine, eGFR, high- and low-density lipoprotein, white cell count, medical history (acute kidney injury, chronic heart failure, reported smoking, alcoholism)	–	Gradient Boosted Trees and Random Forest	AUC—0.75 for any stage of CKD (internal validation), > 0.82 for advanced stages (external validation)
Sudharsan et al. (2015)	T2D to hypoglycemia	56,000 blood glucose data points	Self-monitored blood glucose readings, timestamp of reading, medication administration and dosage	–	Random Forest, support vector machine, k-Nearest Neighbor, Naïve Bayes	Sensitivity—92%, Specificity—90% with blood glucose readings and medication

out the effectiveness of classifier predictions with different combinations of pre-processing (Maniruzzaman et al. 2018; Wang et al. 2019) and feature selection techniques (De Silva et al. 2020; Roy et al. 2021; Rubaiat et al. (2018)). Logistic Regression, a traditional statistical technique for binary and multivariate analysis has been considered in many of the included works owing to its simplicity and ability to model the interrelationships between dependent and independent variables. Nusinovici et al. (2020) compared the efficiency of logistic regression with different ML models in predicting four different chronic diseases. They concluded that logistic regression is as efficient as ML models for disease risk prediction when the dataset considered has fewer incident cases and simple clinical predictors. On the other hand, a variety of ML algorithms have been applied to predict diabetes among which more common ones include Naïve Bayes, k-Nearest Neighbors, support vector machines, k-means clustering, decision trees, neural networks, ensemble models like random forests, gradient boosting etc.

The implicit risk assessment in many of the works if at all included was for the patient population in general and not at individual level, though some studies particularly investigated the role of certain specific risk factors in T2D development (Lee and Kim 2016; Peddinti et al. 2017; Perveen et al. 2019; Yokota et al. 2017). Many of the research works validated the results on test dataset derived from the original input dataset, with external validation on different dataset found in few of the works. A number of features were considered for predictive modeling of T2D including clinical (lab-based or physical examination parameters), lifestyle, demographic and some unconventional, non-invasive features (toenails, iris images), which makes comparison of results less straightforward owing to difference in accuracy with differing features. This was primarily due to use of different datasets that varied with respect to patient population and characteristics. It is thus imperative to validate the predictive ability and usefulness of ML predictive models on real life patient characteristics other than the input dataset considered.

In case of prognostic modeling, ML models have been proposed to predict prognosis of individuals with high risk to future incident T2D as well as prognosis of prevalent diabetes to potential complications. Both areas have huge clinical and economical value that target the core problem of identifying high risk candidates eligible for certain medical interventions/treatments so as to prevent future health degradation as well as healthcare costs.

An important and inevitable parameter of clinical decision support systems, i.e., interpretability is gaining lot of emphasis nowadays so as to develop explainable, white-box models. Kopitar et al. (2019) throws light on local v/s global interpretability with a case of diabetes prediction and stresses on the importance of local interpretation techniques for risk assessment at individual patient level. Table 7 enlists the

Table 7 Major gap areas and underlying challenges in existing ML clinical decision support for type 2 diabetes prevention and management

Gap areas	Underlying challenges
1. Medical gaps (diabetes knowledge base, guidelines and practice)	Etiological and pathophysiological causes of diabetes unknown Lack of knowledge about pre-diabetes risk factors Selection of intervention candidates Detection of early onset complications in pre-diabetes
2. Technological gaps (machine learning predictive modeling)	Data availability and pre-processing issues Patient-wise risk factor ranking Accuracy v/s interpretability tradeoff External validation of the model Practical usability and impact of the model

major gap areas and underlying challenges in ML-based diabetes clinical decision support.

10 Conclusion

In this paper, a review of ML research efforts specifically for type 2 diabetes clinical decision support is presented by firstly identifying and highlighting the medical gaps in diabetes knowledge base, guidelines and practice that can be addressed using ML, followed by review of recent papers in the areas of ML-based diabetes risk assessment, diagnosis and prognosis. The results of review have been summarized in three different application areas—(1) statistical and ML-based risk assessment, (2) non-invasive and invasive diagnosis methods, (3) prognosis modeling for predicting incidence of type 2 diabetes and that of related complications. Our paper provides the breadth of ML applications in clinical decision support aimed at prevention and management of type 2 diabetes and highlights the potential medical and technological gaps in existing research works that need to be resolved to build clinically usable and reliable ML clinical decision support models for type 2 diabetes care.

Funding The authors did not receive support from any organization for the submitted work.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest None.

References

- Abbas HT, Alic L, Erraguntla M, Ji JX, Abdul-Ghani M, Abbasi QH, Qaraqe MK (2019) Predicting long-term type 2 diabetes with support vector machine using oral glucose tolerance test. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0219636>
- Abhari S, Kalhori SRN, Ebrahimi M, Hasannejadasi H, Garavand A (2019) Artificial intelligence applications in type 2 diabetes mellitus care: focus on machine learning methods. *Healthc Inform Res*. <https://doi.org/10.4258/hir.2019.25.4.248>
- Aekplakorn W, Bunnag P, Woodward M et al (2006) A risk score for predicting incident diabetes in the Thai population. *Diabetes Care*. <https://doi.org/10.2337/dc05-2141>
- Akula R, Nguyen N, Garibay I (2019) Supervised machine learning based ensemble model for accurate prediction of type 2 diabetes. In: Conference proceedings—IEEE SOUTHEASTCON (Vol. 2019-April). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/SoutheastCon42311.2019.9020358>
- Alharbi A, Alghahtani M (2019) Using genetic algorithm and ELM neural networks for feature extraction and classification of type 2-diabetes mellitus. *Appl Artif Intell* 33(4):311–328. <https://doi.org/10.1080/08839514.2018.1560545>
- Allen A, Iqbal Z, Green-Saxena A et al (2022) Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus. *BMJ Open Diabetes Res Care* 10:e002560. <https://doi.org/10.1136/bmjdr-2021-002560>
- Aminian A, Zajichek A, Arterburn DE et al (2020) Predicting 10-year risk of end-organ complications of type 2 diabetes with and without metabolic surgery: a machine learning approach. *Diabetes Care* 43(4):852–859. <https://doi.org/10.2337/dc19-2057s>
- Anderson JP, Parikh JR, Shenfeld DK et al (2016) Reverse engineering and evaluation of prediction models for progression to type 2 diabetes: an application of machine learning using electronic health records. *J Diabetes Sci Technol* 10(1):6–18. <https://doi.org/10.1177/1932296815620200>
- Bala Manoj Kumar P, Srinivasa Perumal R, Nadesh RK, Arivuselvan K (2020) Type 2: diabetes mellitus prediction using deep neural networks classifier. *Int J Cognit Comput Eng* 1:55–61. <https://doi.org/10.1016/j.ijcce.2020.10.002>
- Barakat N, Bradley AP, Barakat MNH (2010) Intelligent support vector machines for diagnosis of diabetes mellitus. *IEEE Trans Inf Technol Biomed* 14(4):1114–1120. <https://doi.org/10.1109/TITB.2009.2039485>
- Battineni G, Sagaro GG, Nalini C, Amenta F, Tayebati SK (2019) Comparative machine-learning approach: a follow-up study on type 2 diabetes predictions by cross-validation methods. *Machines*. <https://doi.org/10.3390/machines7040074>
- Bernardini M, Morettini M, Romeo L, Frontoni E, Burattini L (2020a) Early temporal prediction of type 2 diabetes risk condition from a general practitioner electronic health record: a multiple instance boosting approach. *Artif Intell Med*. <https://doi.org/10.1016/j.artmed.2020.101847>
- Bernardini M, Romeo L, Misericordia P, Frontoni E (2020b) Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine. *IEEE J Biomed Health Inform* 24(1):235–246. <https://doi.org/10.1109/JBHI.2019.2899218>
- Bethel MA, Hyland KA, Chacra AR et al (2017) Updated risk factors should be used to predict development of diabetes. *J Diabetes Complicat* 31(5):859–863. <https://doi.org/10.1016/j.jdiacomp.2017.02.012>
- Boutillier JJ, Chan TCY, Ranjan M, Deo S (2021) Risk Stratification for early detection of diabetes and hypertension in resource-limited settings: machine learning analysis. *J Medi Int Res*. <https://doi.org/10.2196/20123>
- Bowen ME, Schmittiel JA, Kullgren JT, Ackermann RT, O'Brien MJ (2018) Building toward a population-based approach to diabetes screening and prevention for US adults. *Curr Diabetes Rep*. <https://doi.org/10.1007/s11892-018-1090-5>
- Cahn A, Shoshan A, Sagiv T, Yesharim R, Goshen R, Shalev V, Raz I (2020) Prediction of progression from pre-diabetes to diabetes: development and validation of a machine learning model. *Diabetes Metab Res Rev*. <https://doi.org/10.1002/dmrr.3252>
- Carter JA, Long CS, Smith BP, Smith TL, Donati GL (2019) Combining elemental analysis of toenails and machine learning techniques as a non-invasive diagnostic tool for the robust classification of type-2 diabetes. *Expert Syst Appl* 115:245–255. <https://doi.org/10.1016/j.eswa.2018.08.002>
- Casanova R, Saldana S, Simpson SL et al (2016) Prediction of incident diabetes in the Jackson heart study using high-dimensional machine learning. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0163942>
- Cefalu WT (2016) “Prediabetes”: are there problems with this label? No, we need heightened awareness of this condition! *Diabetes Care* 39(8):1472–1477. <https://doi.org/10.2337/dc16-1143>
- Chen L, Magliano DJ, Balkau B et al (2010) AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures. *Med J Aust* 192(4):197–202. <https://doi.org/10.5694/j.1326-5377.2010.tb03478.x>
- Choi BG, Rha SW, Kim SW, Kang JH, Park JY, Noh YK (2019) Machine learning for the prediction of new-onset diabetes mellitus during 5-year follow-up in non-diabetic patients with cardiovascular risks. *Yonsei Med J* 60(2):191–199. <https://doi.org/10.3349/ymj.2019.60.2.191>
- Contreras I, Vehi J (2018) Artificial intelligence for diabetes management and decision support: literature review. *J Med Int Res*. <https://doi.org/10.2196/10775>
- D'Angelo G, Palmieri F (2020a) Knowledge elicitation based on genetic programming for non destructive testing of critical aerospace systems. *Futur Gener Comput Syst* 102:633–642. <https://doi.org/10.1016/j.future.2019.09.007>
- D'Angelo G, Palmieri F (2020b) Discovering genomic patterns in SARS-CoV-2 variants. *Int J Intell Syst* 35(11):1680–1698. <https://doi.org/10.1002/int.22268>
- Dagliati A, Marini S, Sacchi L et al (2018) Machine learning methods to predict diabetes complications. *J Diabetes Sci Technol* 12(2):295–302. <https://doi.org/10.1177/1932296817706375>
- Dalakleidi K, Zarkogianni K, Thanopoulou A, Nikita K (2017) Comparative assessment of statistical and machine learning techniques towards estimating the risk of developing type 2 diabetes and cardiovascular complications. *Expert Syst*. <https://doi.org/10.1111/exsy.12214>
- De Silva K, Jönsson D, Demmer RT (2020) A combined strategy of feature selection and machine learning to identify predictors of prediabetes. *J Am Med Inform Assoc* 27(3):396–406. <https://doi.org/10.1093/jamia/ocz204>
- De Silva K, Enticott J, Barton C, Forbes A, Saha S, Nikam R (2021) Use and performance of machine learning models for type 2 diabetes prediction in clinical and community care settings: protocol for a systematic review and meta-analysis of predictive modeling studies. *Digit Health*. <https://doi.org/10.1177/20552076211047390>

- Deng Y, Lu L, Aponte L, Angelidi AM, Novak V, Kamiadakis GE, Mantzoros CS (2021) Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. *NPJ Digit Med*. <https://doi.org/10.1038/s41746-021-00480-x>
- Dimopoulos AC, Nikolaidou M, Caballero FF et al (2018) Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. *BMC Med Res Methodol*. <https://doi.org/10.1186/s12874-018-0644-1>
- Dinh A, Miertschin S, Young A, Mohanty SD (2019) A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Med Inform Decis Mak*. <https://doi.org/10.1186/s12911-019-0918-5>
- Dugee O, Janchiv O, Jousilahti P, Sakhiya A, Palam E, Nuorti JP, Peltonen M (2015) Adapting existing diabetes risk scores for an Asian population: a risk score for detecting undiagnosed diabetes in the Mongolian population. *BMC Public Health*. <https://doi.org/10.1186/s12889-015-2298-9>
- Farran B, AlWotayan R, Alkandari H, Al-Abdulrazzaq D, Channanath A, Thanaraj TA (2019) Use of non-invasive parameters and machine-learning algorithms for predicting future risk of type 2 diabetes: a retrospective cohort study of health data from Kuwait. *Front Endocrinol*. <https://doi.org/10.3389/fendo.2019.00624>
- Fregoso-Aparicio L, Noguez J, Montesinos L, García-García JA (2021) Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetol Metab Syndr* 13(1):148. <https://doi.org/10.1186/s13098-021-00767-9>. PMID:34930452;PMCID:PMC8686642
- García-Carretero R, Vigil-Medina L, Mora-Jimenez I, Soguero-Ruiz C, Barquero-Perez O, Ramos-Lopez J (2020) Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population. *Med Biol Eng Compu* 58(5):991–1002. <https://doi.org/10.1007/s11517-020-02132-w>
- García-Ordás MT, Benavides C, Benítez-Andrades JA, Alaiz-Moretón H, García-Rodríguez I (2021) Diabetes detection using deep learning techniques with oversampling and feature augmentation. *Comput Methods Progr Biomed*. <https://doi.org/10.1016/j.cmpb.2021.105968>
- Gray LJ, Taub NA, Khunti K et al (2010) The Leicester Risk Assessment score for detecting undiagnosed type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. *Diabet Med* 27(8):887–895. <https://doi.org/10.1111/j.1464-5491.2010.03037.x>
- Han L, Luo S, Yu J, Pan L, Chen S (2015) Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes. *IEEE J Biomed Health Inform* 19(2):728–734. <https://doi.org/10.1109/JBHI.2014.2325615>
- Hayashi Y, Yukita S (2016) Rule extraction using Recursive-Rule extraction algorithm with J48graft combined with sampling selection techniques for the diagnosis of type 2 diabetes mellitus in the Pima Indian dataset. *Inform Med Unlocked* 2:92–104. <https://doi.org/10.1016/j.imu.2016.02.001>
- Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P (2009) Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ* (online) 338(7698):811–816. <https://doi.org/10.1136/bmj.b880>
- Kaur H, Kumari V (2022) Predictive modelling and analytics for diabetes using a machine learning approach. *Appl Comput Inform* 18(1–2):90–100. <https://doi.org/10.1016/j.aci.2018.12.004>
- Kaur A, Guleria K, Kumar Trivedi N (2021). Feature selection in machine learning: methods and comparison. In: 2021 international conference on advance computing and innovative technologies in engineering, ICACITE 2021 (pp. 789–795). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICACITE51222.2021.9404623>
- Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I (2017) Machine learning and data mining methods in diabetes research. *Comput Struct Biotechnol J*. <https://doi.org/10.1016/j.csbj.2016.12.005>
- Kopitar L, Cilar L, Kocbek P, Stiglic G (2019) Local vs. global interpretability of machine learning models in type 2 diabetes mellitus screening. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics*, vol 11979 LNAI. Springer, pp 108–119. https://doi.org/10.1007/978-3-030-37446-4_9
- Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G (2020) Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci Rep*. <https://doi.org/10.1038/s41598-020-68771-z>
- Kowsher M, Turaba MY, Sajed T, Mahabubur Rahman MM (2019) Prognosis and treatment prediction of type-2 diabetes using deep neural network and machine learning classifiers. In: 2019 22nd international conference on computer and information technology, ICCIT 2019. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICCIT48885.2019.9038574>
- Lai H, Huang H, Keshavjee K, Guergachi A, Gao X (2019) Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocr Disord*. <https://doi.org/10.1186/s12902-019-0436-6>
- Lee BJ, Kim JY (2016) Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on Machine Learning. *IEEE J Biomed Health Inform* 20(1):39–46. <https://doi.org/10.1109/JBHI.2015.2396520>
- Lindström J, Tuomilehto J (2003) The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 26(3):725–731. <https://doi.org/10.2337/diacare.26.3.725>
- Liu Y, Ye S, Xiao X, Sun C, Wang G, Wang G, Zhang B (2019) Machine learning for tuning, selection, and ensemble of multiple risk scores for predicting type 2 diabetes. *Risk Manag Healthc Policy* 12:189–198. <https://doi.org/10.2147/RMHP.S225762>
- Luo G (2016) Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. *Health Inform Sci Syst*. <https://doi.org/10.1186/s13755-016-0015-4>
- Mahboob Alam T, Iqbal MA, Ali Y et al (2019) A model for early prediction of diabetes. *Inform Med Unlocked*. <https://doi.org/10.1016/j.imu.2019.100204>
- Maniruzzaman M, Rahman MJ, Al-MehediHasan M, Suri HS, Abedin MM, El-Baz A, Suri JS (2018) Accurate diabetes risk stratification using machine learning: role of missing value and outliers. *J Med Syst*. <https://doi.org/10.1007/s10916-018-0940-7>
- Martinez-Millana A, Argente-Pla M, Martinez BV, Salcedo VT, Merino-Torres JF (2019) Driving type 2 diabetes risk scores into clinical practice: performance analysis in hospital settings. *J Clin Med*. <https://doi.org/10.3390/jcm8010107>
- Mohan V, Deepa R, Deepa M, Somannavar S, Datta M (2005) A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects. *J Assoc Phys India* 53:759–763
- Nguyen BP, Pham HN, Tran H et al (2019) Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. *Comput Methods Progr Biomed*. <https://doi.org/10.1016/j.cmpb.2019.105055>
- Nilashi M, Ibrahim O, Dalvi M, Ahmadi H, Shahmoradi L (2017) Accuracy improvement for diabetes disease classification: a case on a public medical dataset. *Fuzzy Inf Eng* 9(3):345–357. <https://doi.org/10.1016/j.fiae.2017.09.006>
- Nirala N, Periyasamy R, Singh BK, Kumar A (2019) Detection of type-2 diabetes using characteristics of toe photoplethysmogram by applying support vector machine. *Biocybern Biomed Eng* 39(1):38–51. <https://doi.org/10.1016/j.bbe.2018.09.007>
- Noble D, Mathur R, Dent T, Meads C, Greenhalgh T (2011) Risk models and scores for type 2 diabetes: systematic review. *BMJ* (online). <https://doi.org/10.1136/bmj.d7163>

- Nusinovici S, Tham YC, Chak Yan MY et al (2020) Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol* 122:56–69. <https://doi.org/10.1016/j.jclinepi.2020.03.002>
- Olivera AR, Roesler V, Iochpe C, Schmidt MI, Vigo Á, Barreto SM, Duncan BB (2017) Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes—ELSA-Brasil: accuracy study. *Sao Paulo Med J* 135(3):234–246. <https://doi.org/10.1590/1516-3180.2016.0309010217>
- Peddinti G, Cobb J, Yengo L et al (2017) Early metabolic markers identify potential targets for the prevention of type 2 diabetes. *Diabetologia* 60(9):1740–1750. <https://doi.org/10.1007/s00125-017-4325-0>
- Pei D, Gong Y, Kang H, Zhang C, Guo Q (2019) Accurate and rapid screening model for potential diabetes mellitus. *BMC Med Inform Decis Mak*. <https://doi.org/10.1186/s12911-019-0790-3>
- Perveen S, Shahbaz M, Keshavjee K, Guergachi A (2019) Metabolic syndrome and development of diabetes mellitus: predictive modeling based on machine learning techniques. *IEEE Access* 7:1365–1375. <https://doi.org/10.1109/ACCESS.2018.2884249>
- Perveen S, Shahbaz M, Saba T, Keshavjee K, Rehman A, Guergachi A (2020) Handling irregularly sampled longitudinal data and prognostic modeling of diabetes using machine learning technique. *IEEE Access* 8:21875–21885. <https://doi.org/10.1109/ACCESS.2020.2968608>
- Rahman M, Islam D, Mukti RJ, Saha I (2020) A deep learning approach based on convolutional LSTM for detecting diabetes. *Comput Biol Chem*. <https://doi.org/10.1016/j.compbiolchem.2020.107329>
- Rampone S, Valente A (2012) Neural network aided evaluation of landslide susceptibility in southern Italy. *Int J Mod Phys C*. <https://doi.org/10.1142/S0129183112500027>
- Reach G, Pechtner V, Gentilella R, Corcos A, Ceriello A (2017) Clinical inertia and its impact on treatment intensification in people with type 2 diabetes mellitus. *Diabetes Metab*. <https://doi.org/10.1016/j.diabet.2017.06.003>
- Roopa H, Asha T (2019) A linear model based on principal component analysis for disease prediction. *IEEE Access* 7:105314–105318. <https://doi.org/10.1109/ACCESS.2019.2931956>
- Roy K, Ahmad M, Waqar K et al (2021) An enhanced machine learning framework for type 2 diabetes classification using imbalanced data with missing values. *Complexity*. <https://doi.org/10.1155/2021/9953314>
- Rubaiat SY, Rahman MM, Hasan MK (2018) Important feature selection accuracy comparisons of different machine learning models for early diabetes detection. In: 2018 international conference on innovation in engineering and technology, ICIET 2018. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/CIET.2018.8660831>
- Samant P, Agarwal R (2018) Machine learning techniques for medical diagnosis of diabetes using iris images. *Comput Methods Progr Biomed* 157:121–128. <https://doi.org/10.1016/j.cmpb.2018.01.004>
- Schulze MB, Hoffmann K, Boeing H et al (2007) An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 30(3):510–515. <https://doi.org/10.2337/dc06-2089>
- Severejn E, Wong S, Velásquez J, Perpiñán G, Herrera H, Altuve M, Díaz J (2020) Diagnosis of type 2 diabetes and pre-diabetes using machine learning. In: IFMBE proceedings, vol 75. Springer, pp 792–802. https://doi.org/10.1007/978-3-030-30648-9_105
- Sidey-Gibbons JAM, Sidey-Gibbons CJ (2019) Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. <https://doi.org/10.1186/s12874-019-0681-4>
- Sowah RA, Bampoe-Addo AA, Armoo SK, Saalia FK, Gatsi F, Sarkodie-Mensah B (2020) Design and development of diabetes management system using machine learning. *Int J Telemed Appl*. <https://doi.org/10.1155/2020/8870141>
- Spänig S, Emberger-Klein A, Sowa JP, Canbay A, Menrad K, Heider D (2019) The virtual doctor: an interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. *Artif Intell Med*. <https://doi.org/10.1016/j.artmed.2019.101706>
- Stolfi P, Valentini I, Palumbo MC, Tieri P, Grignolio A, Castiglione F (2020) Potential predictors of type-2 diabetes risk: machine learning, synthetic data and wearable health devices. *BMC Bioinform*. <https://doi.org/10.1186/s12859-020-03763-4>
- Sudharsan B, Peebles M, Shomali M (2015) Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. *J Diabetes Sci Technol* 9(1):86–90. <https://doi.org/10.1177/1932296814554260>
- Sumathi A, Meganathan S (2019) Semi supervised data mining model for the prognosis of pre-diabetic conditions in type 2 diabetes mellitus. *Bioinformatics* 15(12):875–881. <https://doi.org/10.6026/97320630015875>
- Syed AH, Khan T (2020) Machine learning-based application for predicting risk of type 2 diabetes mellitus (t2dm) in Saudi Arabia: a retrospective cross-sectional study. *IEEE Access* 8:199539–199561. <https://doi.org/10.1109/ACCESS.2020.3035026>
- Tseng E, Greer RC, O'Rourke P et al (2019) National survey of primary care physicians' knowledge, practices, and perceptions of prediabetes. *J Gen Intern Med* 34(11):2475–2481. <https://doi.org/10.1007/s11606-019-05245-7>
- Wang Q, Cao W, Guo J, Ren J, Cheng Y, Davis DN (2019) DMP_MI: an effective diabetes mellitus classification algorithm on imbalanced data with missing values. *IEEE Access* 7:102232–102238. <https://doi.org/10.1109/ACCESS.2019.2929866>
- Wang Y, Zhang L, Niu M et al (2021) Genetic risk score increased discriminant efficiency of predictive models for type 2 diabetes mellitus using machine learning: cohort study. *Front Public Health*. <https://doi.org/10.3389/fpubh.2021.606711>
- World Health Organization (2016) Global report on diabetes
- World Health Organization (2019) Classification of diabetes mellitus
- Xie Z, Nikolayeva O, Luo J, Li D (2019) Building risk prediction models for type 2 diabetes using machine learning techniques. *Prev Chronic Dis*. <https://doi.org/10.5888/pcd16.190109>
- Yin H, Jha NK (2017) A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles. *IEEE Trans Multi-Scale Comput Syst* 3(4):228–241. <https://doi.org/10.1109/TMCS.2017.2710194>
- Yokota N, Miyakoshi T, Sato Y, Nakasone Y, Yamashita K, Imai T, Hirabayashi K, Koike H, Yamauchi K, Aizawa T (2017) Predictive models for conversion of pre-diabetes to diabetes. *J Diabetes Complicat* 31(8):1266–1271. <https://doi.org/10.1016/j.jdiacomp.2017.01.005>
- Yudkin JS (2016) “Prediabetes”: are there problems with this label? Yes, the label creates further problems! *Diabetes Care* 39(8):1468–1471. <https://doi.org/10.2337/dc15-2113>
- Zarkogianni K, Athanasiou M, Thanopoulou AC (2018) Comparison of machine learning approaches toward assessing the risk of developing cardiovascular disease as a long-term diabetes complication. *IEEE J Biomed Health Inform* 22(5):1637–1647. <https://doi.org/10.1109/JBHI.2017.2765639>
- Zhang J, Xu J, Hu X, Chen Q, Tu L, Huang J, Cui J (2017) Diagnostic method of diabetes based on support vector machine and tongue images. *Biomed Res Int*. <https://doi.org/10.1155/2017/7961494>
- Zhang L, Wang Y, Niu M, Wang C, Wang Z (2021) Nonlaboratory-based risk assessment model for type 2 diabetes mellitus screening in Chinese rural population: a joint Bagging-Boosting Model. *IEEE J Biomed Health Inform* 25(10):4005–4016. <https://doi.org/10.1109/JBHI.2021.3077114>

- Zheng T, Xie W, Xu L et al (2017) A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int J Med Inform* 97:120–127. <https://doi.org/10.1016/j.ijmedinf.2016.09.014>
- Zhu C, Idemudia CU, Feng W (2019) Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Inform Med Unlocked*. <https://doi.org/10.1016/j.imu.2019.100179>
- Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H (2018) Predicting diabetes mellitus with machine learning techniques. *Front Genet*. <https://doi.org/10.3389/fgene.2018.00515>