Smart Health 32 (2024) 100457

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

Recent trends and techniques of blood glucose level prediction for diabetes control

Benzir Md. Ahmed [a,c,*], Mohammed Eunus Ali [a], Mohammad Mehedy Masud [b], Mahmuda Naznin [a]

[a] Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, 1000, Bangladesh
[b] College of IT, United Arab Emirates University, Al Ain, 15551, United Arab Emirates
[c] Department of Computer Science and Engineering, United International University, Dhaka, 1212, Bangladesh

ARTICLE INFO

Keywords:
Blood glucose prediction
Machine learning
Deep learning
Diabetes management

ABSTRACT

Diabetes, a metabolic disorder disease, can cause short-term acute or even long-term chronic complications in a patient's body. In 2021, 10.5% of the world's adult population had diabetes. These numbers are increasing day by day, which results in an associated increase of morbidity, mortality, and health care cost related to diabetes. Thus, a huge research effort has been carried out to manage diabetes. A precursor to diabetes management is to predict the future blood glucose levels based on a patient's past history. In this paper, we provide a comprehensive and systematic study of diabetes management, focusing on recent research towards blood glucose level prediction. In particular, we have categorized and presented existing recent research based on major clinical application domains, different input features, and major modeling techniques including physiological, data-driven, and hybrid models. We have summarized the performance analysis of different modeling techniques using different metrics, and critically analyzed these techniques from different perspectives. Finally, we have identified a number of research challenges and potential future works that range from data collection to model improvement for Type 2 Diabetes Mellitus. This review can be a good starting point for researchers and practitioners who are working in building data-driven computational models for diabetes management and blood glucose level prediction.

1. Introduction

Diabetes Mellitus (DM) is a group of chronic metabolic disorders caused due to a raised level of blood glucose (BG) in a human body (Association, 2014). Diabetic patients may develop acute complications such as *ketoacidosis* that occurs when the body produces a high level of ketones, and *hyperosmolar* that occurs when blood has an extremely high sugar level, or chronic complications from long-term diabetes. Long-term diabetes may alter tissues and organs, which may result in diseases such as cardiovascular disease (heart disease), diabetic retinopathy (eye disease), diabetic nephropathy (kidney disease), and diabetic neuropathy (nerve disease). In recent years, the number of people with diabetes has increased at a staggering speed which causes an increase in morbidity, mortality, and health care cost. According to International Diabetes Federation (IDF) Diabetes Atlas 2021 (Federation, 2021), only in 2021, there were about 536.6 million adults aged between 20–79 years (10.5% of all adults in this age group) had diabetes. Among them, 6.7 million adults were to die due to diabetes in 2021. This has triggered a considerable research effort to develop efficient approaches for diabetes management.

* Corresponding author at: Department of Computer Science and Engineering, United International University, Dhaka, 1212, Bangladesh.
E-mail address: benzir@cse.uiu.ac.bd (B.M. Ahmed).

https://doi.org/10.1016/j.smhl.2024.100457
Received 18 August 2023; Received in revised form 31 October 2023; Accepted 11 February 2024
Available online 15 February 2024

A large body of research focuses on developing various diabetes clinical application models. These clinical applications include diabetes diagnosis, adverse glycemic events detection, diagnosis of diabetes-related long-term complications, blood glucose level (BGL) prediction (BGLP), blood glucose control, daily decision support system, prediction of population risk stratification, diabetes related sensor technology, identification of biomarkers, etc. We now briefly discuss some of them. Some research works diagnose diabetes or pre-diabetes using various input features such as Electronic Health Records (EHRs) data fields (Miotto, Li, & Kidd, 2016; Pham, Tran, Phung, & Venkatesh, 2017), Breath Signal Acetone Levels (BSALs) (Lekha & Suchetha, 2017), and Heart Rate Variability (HRV). The works that predict adverse glycemic events i.e., hyperglycemia (BGL over the normal range) and hypoglycemia (BGL below the normal range) events include (Cappon, Facchinetti, Sparacino, Georgiou, & Herrero, 2019; Goutham, Ravi, & Kp, 2018; Oviedo et al., 2019a; Reddy et al., 2019; Vehi, Contreras, Oviedo, Biagi, & Bertachi, 2019). These works use a wide variety of input features including historical BGLs, insulin dosages, meal intakes, physical activities, and EHRs from hospitals or clinics.

Another group of research works in diabetes management focus on predicting long-term diabetic related complications including Diabetic Retinopathy (DR) (Arcadu et al., 2019; Raumviboonsuk et al., 2019), Diabetic Neuropathy (DuBrava et al., 2016; Williams et al., 2020), Diabetic Foot (DF) infection (Adam et al., 2018; Vega et al., 2020), Diabetic Nephropathy (DN) (Fiarni, Sipayung, & Maemunah, 2019; Huang, Huang, Lee, & Weng, 2015), and Cardiovascular Diseases (CVDs) (Jonnagaddala et al., 2015). A major group of research works in diabetes management focus on building models to predict BGL. These modeling techniques include physiological models (Bergman, Ider, Bowden, & Cobelli, 1979; Bock, Francois, & Gillet, 2015; Liu et al., 2019), data-driven models (Aliberti et al., 2019; Faruqui et al., 2019; Fox, Ang, Jaiswal, Pop-Busui, & Wiens, 2018; Sun, Jankovic, Bally, & Mougiakakou, 2018), and hybrid models (Gyuk, Vassanyi, & Kosa, 2019; Karim, Vassanyi, & Kosa, 2020; Zecchin, Facchinetti, Sparacino, & Cobelli, 2013). With the popularity and success of Deep Learning (DL) techniques, we have witnessed an explosive growth and usage of data-driven models in this domain recently. Since there have been a large number of research works in many sub-fields of diabetic management, it demands a systematic review of related works in this domain.

There have been a few literature reviews on diabetes related clinical applications. Oviedo, Vehi, Calm, and Armengol (2016) have conducted a review on modeling strategies only for *personalized* prediction of BGLs and adverse BG events. The reviewed modeling strategies include control-relevant models besides physiological models, data-driven models, and hybrid models, which are published between 2010 and 2016. Our study has included all BGL models including personalized and generalized models. Kavakiotis et al. (2017) have conducted a systematic review of applications of only data-driven techniques on diabetes prediction, diabetic complications, biomarkers, and health care management systems. In another review, Dankwa-Mullan, Rivo, Sepulveda, Park, Snowdon, and Rhee (2018) have covered application of data-driven models in the Decision Support System (DSS), diabetes related complications, and prediction of population risk stratification. Contreras and Vehi (2018) have reviewed research works that proposed Artificial Intelligence (AI) techniques specially data-driven techniques between 2010 and 2018 in seven diabetes management categories including BG control, BGL prediction, detection of adverse glycemic events, patient risk prediction, and DSS. They have a limited focus on BGL prediction. Woldaregay et al. (2019) have reviewed only the data-driven modeling in BGL Prediction related to Artificial Pancreas (AP) system, DSS, and adverse glycemic event detection. They have reviewed 55 articles between 2010 and 2018 covering four application areas. Zhu, Li, Herrero, and Georgiou (2020) have reviewed 40 articles investigating the applications of only the Deep Learning techniques on diagnosis of diabetes, glucose management, and diagnosis of diabetes-related complications. In the glucose management part, they have included the BGL prediction as a sub-part. Vettoretti, Cappon, Facchinetti, and Sparacino (2020) have provided a narrative review of the use of data-driven techniques for the development of DSS for BG control and BGL prediction. Wadghiri, Idri, Touria, and Hakkoum (2022) have presented a systematic review of ensemble BGL prediction models between 2000 and 2022. Felizardo, Garcia, Pombo, and Megdiche (2021) have presented a systematic review of data-based models on hypoglycemia prediction using real data between January 2014 to June 2020. A quick comparison among the existing review papers and the present study is also presented in Table 1.

Though the above reviews have covered many sub-topics from different dimensions, no existing review work focuses on a comprehensive approach of BGL prediction that include input features, modeling techniques, performance comparisons, etc. To fill up the above research gaps, we present a systematic review of recent methods in BGL prediction. In summary, our major contributions in this paper are listed as follows.

- We have presented and categorized major clinical application domains in diabetes management, focusing on BGL prediction based applications.
- We have organized related works of BGL prediction under two broad perspectives as follows: (1) input features, and (2) modeling techniques.
- As input features play a major role in the outcome of the BGL prediction, we have categorized existing research based on different types of input features for modeling.
- We have conducted an extensive review on recent findings on all major modeling techniques including physiological, data-driven, and hybrid models that have been used for the BGL prediction application.
- We have presented the performance analysis of major works in BGL prediction using different metrics.
- We have identified a number of research challenges, limitations, and the scope of future works including in the direction of BGL prediction of Type 2 Diabetes Mellitus (T2DM).

The rest of the paper is organized as follows. Section 2 provides some preliminary information on diabetes and highlights various clinical applications related to diabetes. In Section 3, the methodology of this survey is described. In Section 4, the role of different input features for BGL prediction is discussed. Section 5 discusses different modeling techniques for BGL prediction. Section 6 provides the comparative study of the performance measures, issues of different modeling techniques, and the scope of future works. Finally, Section 7 provides the conclusion.

Table 1
Existing review articles on BGL prediction vs. this article. In the "No. of Articles in Review" row, the value inside parentheses is the no. of articles on BGL prediction. Modeling Technique (MT), Machine Learning (ML), Deep Learning (DL).

Criteria	Oviedo et al. (2016)	Contreras and Vehi (2018)	Woldaregay et al. (2019)	Zhu et al. (2020)	Wadghiri et al. (2022)	This Study
Study year range	2010–2016	2010–2018	2000–2018	2016–2020	2000–2020	2015–2023
No. of Articles in Review	140	141(27)	55	40	32	58
BGL Prediction Focus	Sub	Limited	Main	Sub	Main	Main
Clinical Application Categorization	✗	✗	✗	✓	✗	✓
Input feature categorization	✗	✗	✗	✗	✗	✓
MT categorization	✓	✗	✓	✗	✓	✓
MT coverage	All	✗	ML	DL	Ensemble	All
Data Collection method	✗	✗	✓	✗	✗	✓
Meal data collection method	✗	✗	✓	✗	✗	✓
MT perform. analysis	✗	✗	✗	✗	✓	✓

Table 2
List of acronyms and their elaborations.

Full forms	Acronyms	Full forms	Acronyms
Auto-Regressive	AR	AR with Exogenous inputs	ARX
Carbohydrate	CHO	Cardiovascular Diseases	CVD
Decision Support System	DSS	Diabetic Foot	DF
Diabetic Nephropathy	DN	Diabetic Retinopathy	DR
Extreme Learning Machine	ELM	Galvanic Skin Response	GSR
Genetic Algorithm	GA	Generative Adversarial Network	GAN
Glycemic Index	GI	Grammatical Evolution	GE
Heart Rate	HR	Heart Rate Variability	HRV
Least Square Estimation	LSE	Particle Swarm Optimization	PSO
Prediction Horizon	PH	Recursive Least Squares	RLS
Recurrent Neural Network	RNN	Self-Organizing Map	SOM

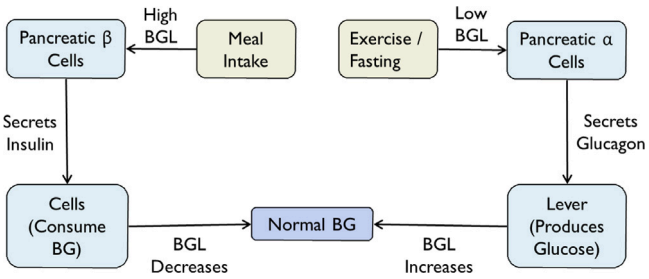


Fig. 1. An overview of Glucose–Insulin dynamics.

2. Diabetes: Background and clinical applications

In this section, we first briefly discuss the preliminaries about the physiological background of diabetes, and then discuss various clinical applications of diabetes.

2.1. Diabetes basics

The increased BGL that occurs in DM is due to (i) lack of secretion of insulin from pancreatic beta cells or (ii) insulin resistance of cells, where cells fail to use insulin to consume glucose from blood or both.

Glucose Insulin Dynamics: The glucose–insulin regulatory system in a healthy human body as seen in Fig. 1 provides the following impacts - (i) if there is high glucose in blood due to meal intake, the incretin hormones is secreted in the gut (a) to signal pancreas beta cell to release (bolus) insulin to match the glucose in the meal, (b) to signal pancreatic alpha cells to inhibit glucagon secretion to stop hepatic glucose production in the liver, (c) to slow gastric emptying to slow down postprandial glucose production. (ii) If there is high glucose in blood due to constant hepatic glucose production, pancreatic beta cells release constant drip of (basal) insulin to match that glucose. (iii) Insulin secreted from pancreatic beta cells allow the body cells to consume glucose from blood thus reducing glucose in the blood. (iv) If there is low glucose in blood due to physical exercise or fasting, pancreatic alpha cells release glucagon to stimulate hepatic glucose production, breaking the glycogen in the liver. Besides that, the kidney also contributes to the level of BG by determining the amount of glucose that can be reabsorbed by the blood and the amount that will be excreted through urine (see Table 2).

Types of Diabetes: The major two types of diabetes out of the various categories of diabetes according to the underlying cause of hyperglycemia listed by American Diabetes Association (ADA) (Association, 2019) are - *Type* 1 diabetes and *Type* 2 diabetes. *Type* 1 diabetes is caused by an autoimmune process of a person born with genetic predisposition to getting the disease initiated by some environmental trigger. This targets and starts destroying the pancreatic beta cells that produce insulin as foreign elements. This leads

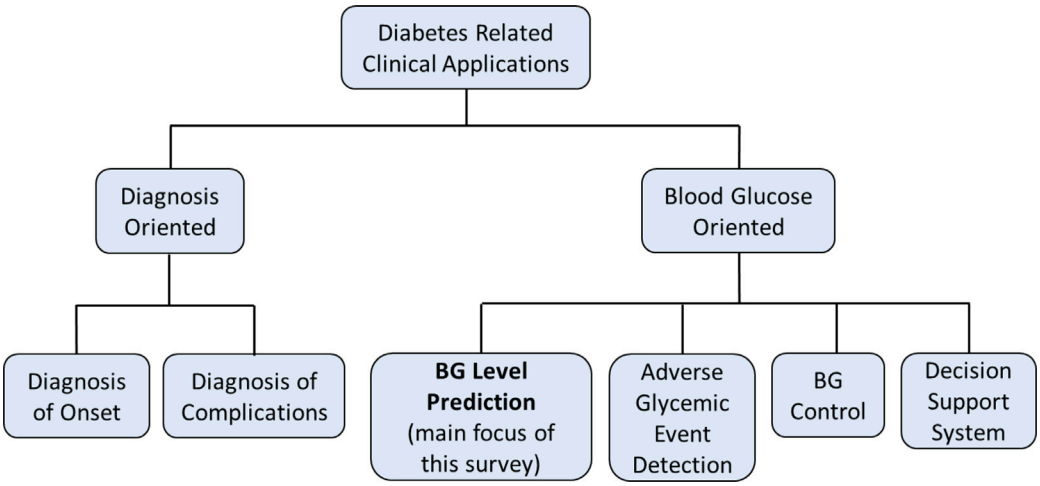


Fig. 2. A taxonomy of clinical applications related to diabetes. (This taxonomy narrows search to BGL prediction).

to a progressive decline of pancreatic insulin secretion and eventually to the zero production of insulin. This type of diabetes can be developed at any age, cannot be prevented or reversed, and needs insulin treatment for the rest of the person's life. Approximately, from 5 to 10% of all diabetic people have Type 1 Diabetes Mellitus (T1DM) (Association, 2019). *Type* 2 Diabetes Mellitus (T2DM) is caused by a pancreatic beta cell insufficiency that becomes unable to produce enough insulin or peripheral insulin resistance that reduces the ability of the cells to consume glucose from blood, or both. Obesity and the lack of exercise are two major factors to the development of Type 2 diabetes. Type 2 diabetes usually starts at late adulthood. However, it can be managed and controlled with a healthy diet, physical activity, and/or medication. Sometimes, it even can be remitted.

2.2. Diabetes related clinical applications

To prevent long-term diabetes related complications, one has to *proactively* manage diabetes. Among many things, diabetes management plan should include - (i) self-monitoring of BGL, (ii) a meal plan to manage weight and BGL, (iii) a physical exercise plan to lower BGL and increase sensitivity to medication, and (iv) diabetes medication plan in form of oral dose or insulin injection to lower BGL. Diabetes engages two things: (a) a *biological process* that includes hormone action and metabolism, and (b) individual *human behavior* that is driven by personal preferences. On one hand, every human being has a unique biological process. On the other hand, every human has his/her own unique personal preferences. Due to these two unique things, every individual has a unique case of diabetes. Hence, each individual needs a separate diabetes management plan. Diabetic people with uncontrolled glucose in blood are at a much higher risk of having life threatening complications in the long run than healthy people.

Diabetes management starts with the diagnosis of diabetes onset. After the diabetes onset is diagnosed, the next step is to control the BGL. If the BG is not managed well, various diabetes related complications may arise. Those complications need to be predicted before they happen so that necessary steps can be taken to avoid them or push them further into the future. Hence, the diabetes related clinical applications can be broadly divided into two domains which are - *diagnosis oriented*, and *blood glucose oriented*. The categorized clinical applications are presented in Fig. 2. Their associated research works are given in Table 3 and are described in short as follows.

2.2.1. Diagnosis oriented applications

The diagnosis oriented clinical applications can be broadly divided into two domains which are - *diagnosis of onset*, and *diagnosis of complications*. They are described briefly as follows.

Diagnosis of Diabetes Onset: Early diagnosis of diabetes or *pre-diabetes* is very important. If pre-diabetes can be diagnosed, appropriate actions e.g., change in diet, increase of physical exercises, etc. can be taken to avoid diabetes. If diabetes can be diagnosed at the early stage, change in lifestyle related to diet and physical activities can control the diabetes. It ultimately helps an individual to lead a normal life and to prevent diabetes related long-term complications. There is active research in this field. The work in Miotto et al. (2016) has used a 3-layer stack of denoising autoencoders on aggregated EHRs, Pham et al. (2017) have introduced a *DeepCare* prediction model that used Long Short Term Memory (LSTM) on discrete EHRs, Goutham et al. (2018) have presented a DL framework using LSTM and Convolutional Neural Network (CNN) on HRV signals, Spanig et al. (2019) have employed Deep NNs using only demographic and basic physiological (non-invasive) data, etc. It is observed that the research works have mostly used deep learning techniques to diagnose diabetes.

Diagnosis of Diabetes-related Long-term Complications: Prolonged untreated elevated BGL, i.e., hyperglycemia, causes long-term diabetes-related complications (Bailey et al., 2013). Prolonged exposure of tissues (cells) to high concentration of glucose can lead to a severe damage of the interior layer of blood vessels. This results in alteration of tissues and organs causing complications

Table 3
Categorized diabetes clinical applications. Diabetic Retinopathy (DR), Diabetic Neuropathy, Cardiovascular Disease (CVD), Diabetic Foot (DF) infection, Diabetic Nephropathy (DN).

Application		Ref.	Description
Diagnosis Oriented Applications	Diagnosis of Diabetes Onset	Goutham et al. (2018), Miotto et al. (2016), Pham et al. (2017), Spanig et al. (2019)	Diagnose diabetes, and prediabetes.
	Diagnosis of Diabetes-related Long-term Complications	DR - (Arcadu et al., 2019; Raumviboonsuk et al., 2019; Ting et al., 2019; Wan, Liang, & Zhang, 2018), Neuropathy - (DuBrava et al., 2016; Williams et al., 2020), CVD - (Jonnagaddala et al., 2015), DF - (Adam et al., 2018; Fraiwan et al., 2017; Goyal et al., 2017; Vega et al., 2020), DN - (Fiarni et al., 2019; Huang et al., 2015)	Diagnose diabetes-related long-term complications.
Blood Glucose Oriented Applications	BGL Prediction	Given in Table 7 and 8 in detail	Predicts BGL.
	Adverse Glycemic Event Detection	Cappon et al. (2019), Oviedo et al. (2019a), Reddy et al. (2019), Vehi et al. (2019)	Predicts hyperglycemia and hypoglycemia events.
	BG Control	Breton et al. (2012), Cappon, Vettoretti, Marturano, Facchinetti, and Sparacino (2018), Daskalaki, Diem, and Mougiakakou (2016), Dassau et al. (2008), Favero et al. (2015), Kircher, Mauseth, Bhatia, and Matheson (2008), Mauseth et al. (2015), Seo, Lee, Lee, Jin, and Park (2019), Yadav, Rani, and Singh (2016)	Find a solution e.g., proper insulin dosage amount to control normal BGL. Automatic Pancreatic (AP) system
	Daily DSS	Dagliati et al. (2018), Eghbali-Zarch, Tavakkoli-Moghaddam, Esfahanian, Sepehri, and Azaron (2018), Sun, Jankovic, Budzinski et al. (2018), Torrent Fontbona and Lapez (2018), Yom-Tov et al. (2017)	Provide daily recommendations to manage diabetes.

including DN, DR, neuropathy, or peripheral neuropathy. High glucose levels can also damage big blood vessels. This may result in cardiovascular diseases and stroke. Due to peripheral neuropathy, one loses the ability to sense traumatic pressures on foot. This wears away the epidermis and exposes the deeper layers creating the neuropathic ulcer, then foot infection, and eventually turning to amputation.

In the field of DR, researchers have used CNNs and transfer learning (Wan et al., 2018) to classify fundus images, have used DL models (Arcadu et al., 2019) to predict future DR progression using Color Fundus Photographs of DR patients, have used fundus photographs and a DL based system (Ting et al., 2019) as grading tool to determine the cardiovascular risk factors for DR, have used a DL algorithm (Raumviboonsuk et al., 2019) on gradable retina images to detect DR with specialist-level accuracy. In the field of diabetic neuropathy, researchers have developed a CNN model with data augmentation (Williams et al., 2020) for the diagnosis of diabetic neuropathy, have identified variables that are correlated with a diagnosis of Diabetic Peripheral Neuropathy (DuBrava et al., 2016) using RF modeling applied to EHR, have used a hybrid approach employing Machine Learning (ML) and rule-based text mining techniques (Jonnagaddala et al., 2015) to find CVD risk factors from clinical notes.

In the field of DF, researchers have used AI and DL techniques (Vega et al., 2020) for the classification of DF thermograms, and have presented a CNN architecture, DFUNet (Goyal et al., 2017) to diagnose DF Ulcers (DFU) using foot images. Fiarni et al. (2019) have constructed a prediction model for retinopathy, nephropathy, and neuropathy using k-means clustering and decision tree-based classification techniques. We observe that in the diagnosis of long-term complications articles ML and DL techniques are mostly used.

2.2.2. Blood glucose oriented applications

When diabetes is already diagnosed, lifestyle must be changed to reduce the BG. This ultimately helps a patient to avoid further long-term health hazards. Healthy lifestyle might even cure a patient from diabetes. Therefore, it is very important to explore the possible steps for the BG management domain which is full of diversified challenges and issues. This BG Management research

domain can be broadly divided into four sub-domains: BGL prediction, adverse glycemic event detection, BG control, and the daily decision support system. They are described in short as follows.

BGL Prediction: The basic goal of a glucose management plan is to keep the BGL within the normal range, which is between 90 mg/dl and 130 mg/dl before meals, and under 180 mg/dl two hours after the meals. BGL prediction has goals (i) to prevent *hypoglycemia*, (ii) to prevent *hyperglycemia*, or (iii) to prevent or delay long term diabetes related complications, such as *DR*, *neuropathy*, CVDs like *heart attack* or stroke, DN, etc. The previous researches on BGL prediction models have been extensively covered in Section 5.

Adverse Glycemic Event Detection: A healthy person's BG remains at the normal level. This is controlled by the insulin secreted from the pancreatic beta cell. For a diabetic person, due to one or multiple of the following problems: autoimmune system disorder, cell glucose resistance, and reduced insulin secretion capacity of pancreatic beta cell, normal BGL cannot be maintained. For many diabetic persons, the BGL goes over the normal level, which is *hyperglycemia*. Various complications occur due to hyperglycemia including CVDs, DR, DN, neuropathy, DF infection, etc. On the other hand, for many the BGL goes below the normal level, which is *hypoglycemia*. The complications that might occur due to prolonged hypoglycemia are seizure, unconsciousness, and in extreme cases even death. There are active researches in this field where Cappon et al. (2019) have employed an Extreme Gradient-Boosted (XGB) tree algorithm to classify the postprandial glycemic state, Oviedo et al. (2019a) have proposed bi-class Support Vector Classifier for predicting postprandial hypoglycemia, Reddy et al. (2019) have used Random Forest (RF) classifier to avoid exercise related hypoglycemia, Vehi et al. (2019) have proposed a Support Vector Machines (SVMs) with Radial Basis Function (RBF) kernel based model to predict hypoglycemic events during postprandial periods, and Artificial Neural Network (ANN) to predict hypoglycemic episodes overnight, etc.

BG Control: High level of sugar in blood causes long term complications. To prevent complications, a person needs to bring the high blood sugar level down within the normal range. To lower the BG, one can take diabetes medication in the form of oral dose or insulin, can do a rigorous exercise for a while, bring a change in his/her diet, or any combination of them. There are active researches in this field where Mauseth et al. (2015) have stress-tested the performance of a dose safety fuzzy logic controller (Dassau et al., 2008; Kircher et al., 2008) to highlight the deficits in the AP system, Daskalaki et al. (2016) have presented a Reinforcement Learning (RL)-based algorithm for personalized glucose regulation in T1DM, Favero et al. (2015) have used an improved version of Modular Model Predictive Control (Breton et al., 2012) to control a Continuous Subcutaneous Insulin Infusion (CSII) in a multi-center outpatient setting, Cappon et al. (2018) have developed an NN corrector consisting of a fully connected Feed-forward NN (FNN) with 3 hidden layers, Seo et al. (2019) have developed algorithms based on a RF, a SVM with a linear function or a RBF, a K-Nearest Neighbor (KNN), and a logistic regression to control postprandial hypoglycemia, etc.

Daily Decision Support System (DSS): Decision support systems are open loop systems that provide recommendations to a diabetic patient to manage diabetes. These systems provide the appropriate insulin dosage amount to ingest, sometimes along with the BGL prediction. Sometimes, it might recommend to take a 30 min brisk walk, or to talk to the doctor. There are active researches in this field where Yom-Tov et al. (2017) have proposed an RL algorithm to improve participants' compliance with the activity program, Sun, Jankovic, Budzinski et al. (2018) have proposed an adaptive basal-bolus algorithm based on RL to provide personalized glucose control, Torrent Fontbona and Lapez (2018) have proposed a Case-Based Reasoning (CBR) insulin recommender system to keep BGL in the target glycemic range longer, etc.

3. Overview of the review

3.1. Major focus

The field of diabetes has a vibrant research community working on various diabetes related clinical applications. Out of them, BGL prediction is the most popular field among the researchers. This is due to its use in other fields such as BG control, prediction of adverse glycemic events, daily support system, etc. A predicted BGL value can also empower a diabetic patient to self manage his/her BGL. The focus of this review is to present a comprehensive review of recent modeling techniques in BGL prediction.

3.2. Methodology

An extensive literature search has been conducted to find relevant articles on BGL prediction in diabetes research. In our search, we have used three online databases PubMed, Google Scholar, and ResearchGate. The search has been restricted to English language peer reviewed journals, articles, and conference proceedings published between 2015 and 2023. The search string was "blood glucose" AND (level OR dynamics OR concentration) AND (prediction OR forecasting). From our search, initially we have identified 409 articles. After removing the duplicates we have 316 unique articles. Then, we have manually inspected the articles' titles and abstracts to exclude

1. *the articles that are not specifically related to BGL prediction*,
2. *the articles that are actually book chapters*,
3. *the articles that use non-invasive procedures to collect BG readings*, and
4. *the articles that do not target T1DM or T2DM*.

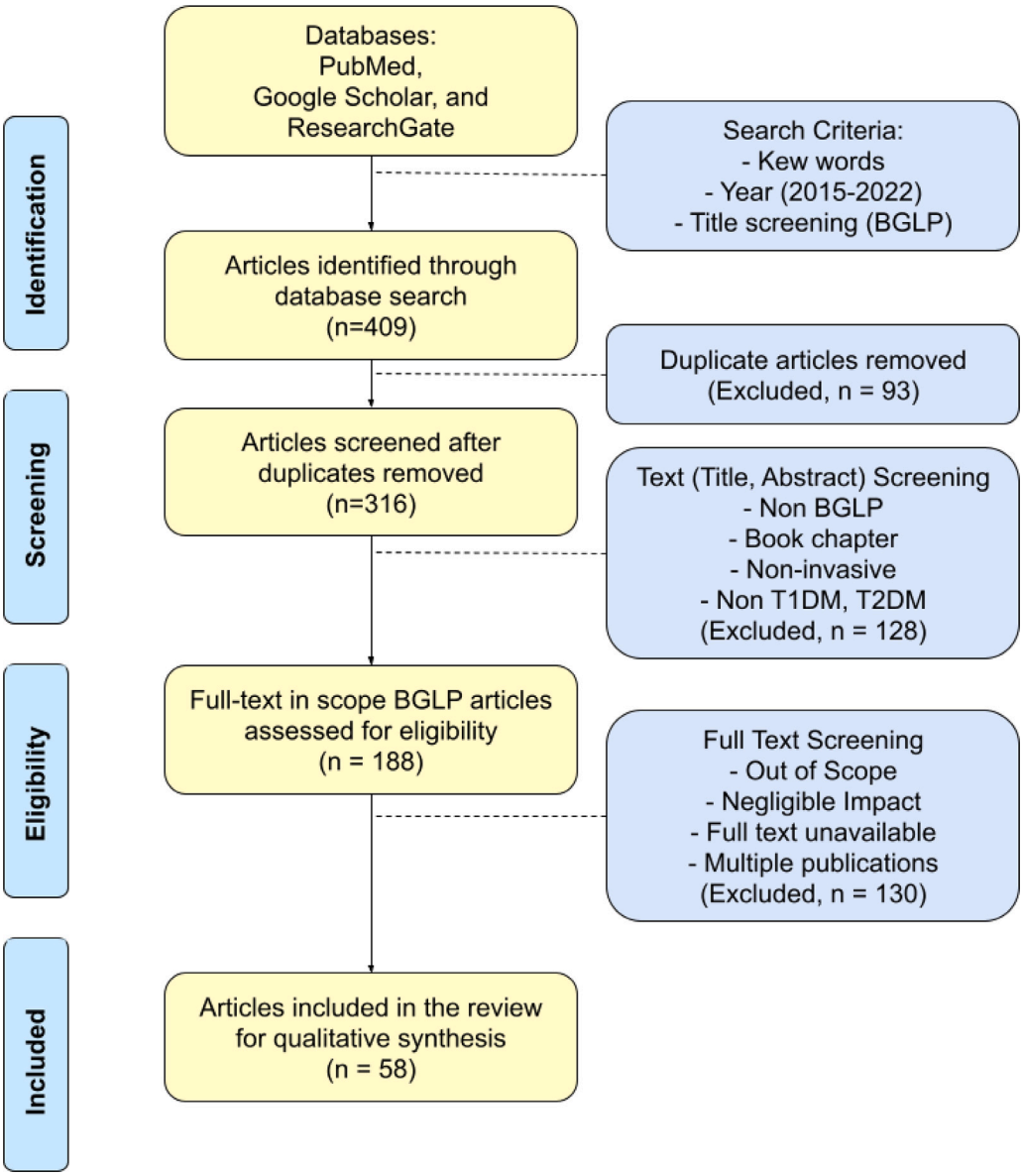
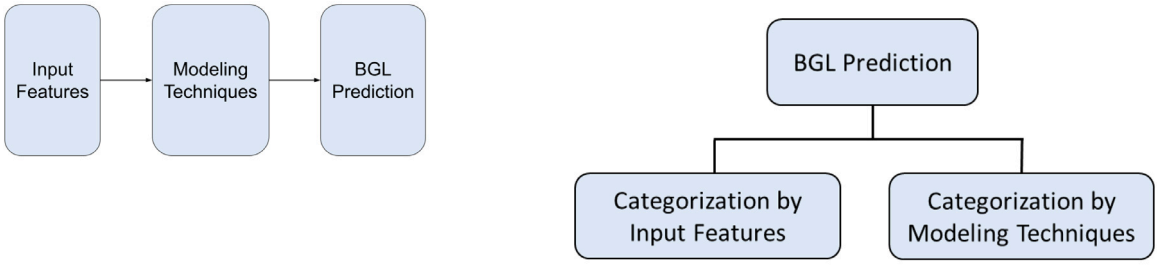


Fig. 3. Review Methodology.

The number of remaining BGLP papers is 188 after exclusion. According to the focus of this literature review, all types of modeling techniques for BGL prediction are included. The full text of the remaining 188 articles are then assessed for eligibility on the grounds of (i) in or out of scope for this review, (ii) level of impact, (iii) full text availability, and (iv) selection of a single article from the same study. After a thorough full text screening, 58 highly original papers with high impact are included in this review as shown in Table 7 and Table 8. The article selection process is briefly shown in Fig. 3.

3.3. Design of this study

A typical BGL prediction framework is given in Fig. 4(a). Both selection of contributing input features and design of efficient modeling techniques play a vital part in BGL prediction. We provided categorization of the reviewed research works based on the input features and the modeling techniques as shown in Fig. 4(b). In case of input features, the reviewed articles were categorized based on the most widely used input features: previous readings of BGL, insulin dosage amount, physical activity, and meal intake. A detailed study on this categorization is presented in Section 4. In case of modeling techniques, the reviewed articles were categorized based on all the currently used modeling techniques: physiological models, data driven models, and hybrid models. A detailed study on this categorization is presented in Section 5.



(a) A typical BGL prediction framework

(b) A taxonomy based on the way research works are categorized

Fig. 4. BGL prediction framework and primary taxonomy.

3.4. Performance metrics used

We used the two most widely used performance metrics in diabetes research: Root Mean Squared Error (RMSE) and Clarke Error Grid Analysis (CEG or EGA) to compare performances of different modeling techniques. RMSE quantifies the statistical accuracy of a modeling technique. It gives the prediction error of a computational method. RMSE is defined in Eq. (1). The lower the RMSE values are, the better the prediction performances get.

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n}(\hat{y}_i - y_i)^2} \tag{1}$$

where y_i are the actual BGL values, $\hat{y}_i$ are the predicted BGL values, n is the number of data samples.

Whereas, CEG quantifies the clinical accuracy of a modeling technique. It has five regions: A, B, C, D, and E. It provides the percentage (%) of predictions of the proposed models that fall in these regions. Higher CEG values for A and B are better. Using only a statistical accuracy metric is not enough for comparing BGL prediction performances. Even with a low RMSE, the errors can be in the critical hypo and hyperglycemia regions, which is not acceptable.

4. BGL prediction input data

In this section, we have discussed the datasets used in the literature, the preprocessing methods, and the input features selection process. We have also highlighted the input features used by different studies, and formatting of the input data before being fed to the models procedures.

4.1. Datasets used in literature

Various types of datasets including clinical, simulated, EHRs, and pathological test reports have been used in the literature. Among the reviewed articles, in 72.4% of the cases only clinical datasets are used. In 22.4% of the cases both clinical and simulated datasets are used. Due to the high rate of data generation by CGM devices per day, 96.5% of the clinical datasets have T1DM data. 53% of those datasets are proprietary. Rest of the datasets used by the literature are publicly available such as OhioT1DM, D1NAMO, UCI, DIAdvisor, DirecNet, and Tidepool. They are shown in Table 4. These publicly available datasets are the result of either individual research efforts of different research groups (UCI, OhioT1DM, D1NAMO), or European Union (EU) funded projects (DIAdvisor), or research collaboration among health professionals, health centers, and industries (JCHR (DirecNet, etc.), Tidepool). These datasets are mainly collected by following study protocols either in a hospital or clinic setup (Alfian et al., 2020; Aliberti et al., 2019; Bock et al., 2015; Kushner, Breton, & Sankaranarayanan, 2020; Li, Daniels, Liu, Herrero & and Georgiou, 2019; Li, Tuo, Wang, & Wang, 2020), or under the guidelines of a hospital yet in free living conditions (Frandes, Timar, Timar, & Lungeanu, 2017; Mosquera-Lopez, Dodier, Tyler, Resalat, & Jacobs, 2019), or in free living conditions (Hamdi et al., 2018; Martinsson, Schliep, Eliasson, & Mogren, 2020; Nemat, Khadem, Eissa, Elliott, & Benaissa, 2022; Saiti, Macas, Lhotska, Stechova, & Pithova, 2020; Yang, Li, Shi, & Xie, 2018; Zhu, Li, Herrero, Chen, & Georgiou, 2018). In the study protocols conducted in free living conditions yet under the guidelines of a hospital, only CGM data is collected in many cases (Armandpour, Kidd, Du, & Huang, 2021; Frandes et al., 2017). In quite a few of the cases (Martinsson et al., 2020; Mosquera-Lopez et al., 2019; Nemat et al., 2022; Xie & Wang, 2020), the patients were under insulin-pump therapy. In a few of the cases (Karim et al., 2020) T2DM patients' CGM data is used. However, there is no dataset (proprietary or public) created only from T2DM patients' data that includes discrete BG readings from glucometers. More efforts need to be put in this direction.

Table 4
Publicly available datasets used by the reviewed articles. Glucose (G), Insulin (I), Meal (M), Exercise (E), Step Count (SC), Skin & Air Temp. (SAT), Jaeb Center for Health Research (JCHR).

Name	Data	Size	Ref.	Used by
JCHR (DirecNet, etc.)	G	Many children & other T1DM datasets	Online	Alfian et al. (2020), He and Wang (2020), Mhaskar, Pereverzyev, and Van der Walt (2017), Ozogur, Ozogur, and Orman (2020), Wang, Tong, and Yu (2020)
UCI	G, I, M, E, Hypoglycemic symptoms	70 T2DM, weeks to months	Online	Li and Fernando (2016)
OhioT1DM	G, I, M, SC, HR, GSR, SAT, sleep quality, work effort, time of exercise-sleep-work-stress-illness	12 T1DM, 8 weeks	Marling and Bunescu (2020)	Daniels, Herrero, and Georgiou (2021), DAntoni, Merone, Piemonte, Iannello, and Soda (2020), Deng et al. (2021), Dudukcu, Taskiran, and Yildirim (2021), Li, Liu, Zhu, Herrero and and Georgiou (2019), Martinsson et al. (2018), Mirshekarian, Shen, Bunescu, and Marling (2019), Nemat et al. (2022), Ozogur et al. (2020), Rabby et al. (2021), Rubin-Falcone, Fox, and Wiens (2020), Xie and Wang (2020), Yang, Yu, Ma, Wu, and Li (2022), Zhu, Li, Chen, Herrero, and Georgiou (2020), Zhu et al. (2018), Zhu, Li, Herrero, and Georgiou (2022), Zhu, Yao, Li, Herrero, and Georgiou (2020)
D1NAMO	G, M, electrocardiogram, breathing, accelerometer	9 T1DM, 20 healthy subjects, 4 days	Dubosson et al. (2018)	Saiti et al. (2020)
DIAdvisor	G, M	20 T1DM, 2 or 3 days	Online	Zecchin et al. (2013)
Tidepool	G, I, M, etc.	10000+ datasets of diff. sizes	Neinstein et al. (2016), Online	Mosquera-Lopez et al. (2019), Rubin-Falcone et al. (2020)

4.2. Preprocessing methods

Various studies have performed preprocessing methods to handle missing and noisy data. In the glucose data, missing data with large gaps are primarily handled by discarding data (Aliberti et al., 2019; Zhu et al., 2018), or by using transfer learning where data from most similar patients (Faruqui et al., 2019) are used. Missing data with short gaps is imputed using linear interpolation (Amar et al., 2020; Daniels et al., 2021; Nemat et al., 2022; Ozogur et al., 2020; Sun, Jankovic, Bally et al., 2018), spline interpolation (Alfian et al., 2020; Gyuk et al., 2019; Li, Liu et al., 2019), first order interpolation (Zhu et al., 2020, 2018), Kalman smoothing technique (Mirshekarian, Bunescu, Marling, & Schwartz, 2017; Mirshekarian et al., 2019; Rabby et al., 2021; Xie & Wang, 2020), last observation carried forward (Faruqui et al., 2019), modified Akima cubic Hermite interpolation (Liu et al., 2019), zero-phase fourth-order low-pass Butterworth filter (Mosquera-Lopez et al., 2019), and with zeros (Rubin-Falcone et al., 2020). In Sun, Jankovic, Bally et al. (2018), they have merged all the time-series data shorter than 1500 measurements to pre-train the models. To remove noise from data, some studies have used denoising filtering techniques such as Savitzky Golay technique (Alfian et al., 2020), median filter (Zhu et al., 2020, 2018), moving average filter (Li et al., 2020), Tikhonov regularization (Aliberti et al., 2019; Zhu et al., 2020), Gaussian filter (Li, Daniels et al., 2019), etc.

To reduce the adverse effect of imbalanced data, some studies (Deng et al., 2021) have used data augmentation on the minority class. Some studies have enriched their small input dataset with synthetic data using Markov chain sampling methods (Hidalgo et al., 2020). This has improved the performance of their prediction models. Many studies (Nemat et al., 2022; Rubin-Falcone et al., 2020) have re-sampled and aligned the discrete input features with the CGM time-series data. They also have re-framed the time-series data into supervised learning problems. To avoid biases from large input values they have normalized data using various normalization techniques such as min–max normalization (Alfian et al., 2020; Faruqui et al., 2019; Ozogur et al., 2020), scaling to be in the range [−1,1] (Mosquera-Lopez et al., 2019), etc.

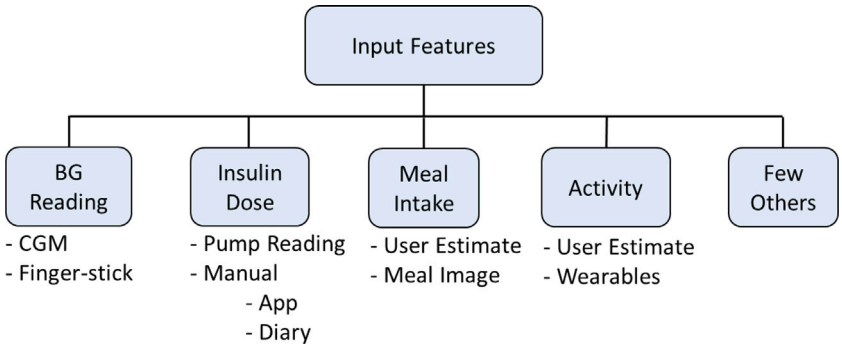


Fig. 5. A taxonomy based on the core input features.

4.3. Selection of input features

Majority of the articles (Armandpour et al., 2021; Daniels et al., 2021; Nemat et al., 2022) have focused on Type 1 Diabetes Mellitus (T1DM) due to the availability of a lot of data from CGM devices in a short time. CGM data is time-series data. There are a plethora of data driven techniques that can model BG time-series data effectively. Literature (see Tables 7 and 8) has proved that only glucose value as input feature is enough to predict future BGLs. Many studies (Alfian et al., 2020; Cichosz, Kronborg, Jensen, & Hejlesen, 2021; DAntoni et al., 2020; Fox et al., 2018; Hamdi et al., 2018; Nemat et al., 2022) have used only the history of BGLs as the input features to allow objective evaluation of the prediction models and to reduce the complexity and variability that come with multivariate variables. Collection and incorporation of the CGM glucose values as the only input is easy, makes the prediction system automatic without the need of human intervention, and is not influenced by human error. Some studies (Sun, Jankovic, Bally et al., 2018; Wang et al., 2020) have used only the history of BGLs as a first step of their research. They intend to include more input features in their future works.

Rabby et al. (2021) have investigated various combinations of only a few input features including CHO, insulin, exercise, and physical activity that are available in the OhioT1DM dataset on the ground that BG dynamics is influenced only by these factors. In their case, the optimal feature set is past history of glucose, bolus insulin, CHO, and step count. In the work by Gyuk et al. (2019) they have mentioned that they have not included physical activity and stress as input features on the ground that the existing literature do not have any reliable prediction methods that can model their effects on BGL. Another study (Li, Liu et al., 2019) has used glucose, meal, insulin and time stamps of those events to have a fair comparison with their baseline models. However, if any input feature influences BG, including it as input feature should help to build a more robust BGL prediction model.

4.4. Input features used in different prediction models

We have studied the articles that are selected for review thoroughly to find the data that are used as input features by the prediction systems. There are quite a few factors that influence the BGL. To better predict the BGL, enough data on these factors along with the previous readings of BGL are used as an input to the prediction system. The mostly used input feature by the reviewed articles is the pre-recorded *BGLs*. 96% of the reviewed research articles have mentioned this as an input feature. The second most used input feature is the history of *meal intake*. 52% of the reviewed research articles have used some forms of meal intake as input features. The third most used input feature is the history of *insulin dosage*. 50% of the reviewed articles have included this as an input feature. The fourth most used input feature is the history of *physical activities*. 14% of the reviewed articles have included this as input feature. Few other input features i.e., stress, heart rate, etc. are also used in the literature but less frequently. Fig. 5 presents the taxonomy of input features. A detailed study on these input features is presented below.

4.4.1. Blood glucose level

BGL is one of the most important factors that influence the prediction of BGL. An individual who has Type 1 Diabetes Mellitus (T1DM), must use a Continuous Glucose Monitoring (CGM) device to monitor his BGL. A CGM device used in Alfian et al. (2020), Bosoni et al. (2020), Fox et al. (2018), Karim et al. (2020), Li, Daniels et al. (2019), Liu et al. (2019), Oviedo et al. (2019b), Zecchin et al. (2013) and Mirshekarian et al. (2017) usually provides a BGL reading every five (5) minutes. That means, it can provide 288 readings per day, which is a lot of data points per day. For a T1DM patient, in most of the cases, 30 days of CGM data is enough to predict BGL.

On the other hand, an individual who has Type 2 Diabetes Mellitus (T2DM), does not need to measure continuous BG reading. Therefore, the patient may use a hand-held blood glucose meter used in Faruqui et al. (2019), Kushner et al. (2020) to measure own blood-glucose. Usually, the patient can measure before and after two hours of each meal, which amounts to 6 data points per day. That being the case, data collection for T2DM patients is more challenging than T1DM patients.

As CGM data from T1DM participants can be collected with less effort, the majority of the research in BGL prediction focuses on CGM based prediction. 83% of the reviewed articles have collected data from CGM devices. Whereas only 2% of the reviewed articles have used BGL readings from hand-held blood glucose meters from T2DM patients. 9% of the articles have used both. Few research works have not provided any information on the data collection process.

4.4.2. Insulin dosage

A meal ingestion at any time increases the BGL of a diabetic patient. Due to insulin secretion deficiency, a diabetic person needs to take insulin in the form of oral medication or exogenous insulin ingestion subcutaneously using subcutaneous insulin syringes or insulin pens (Georga, Protopappas, Polyzos, & Fotiadis, 2015; Gyuk et al., 2019; Kushner et al., 2020), or continuous subcutaneous insulin infusion using insulin pump (Aiello, Lisanti, Magni, Musci, & Toffanin, 2020; Kushner et al., 2020; Li, Daniels et al., 2019; Martinsson et al., 2020; Mirshekarian et al., 2019; Oviedo et al., 2019b; Sun, Jankovic, Bally et al., 2018; Xie & Wang, 2020). The insulin dosage allows body cells to consume glucose from blood. This reduces the glucose level in the blood (Lerner & Porte, 1971). However, in most of the clinical cases the dosage and time for each insulin ingestion are manually recorded by the diabetic patient using paper-diary or smart phone apps. In the case of an insulin pump, users need to enter the amount of insulin in the insulin pump. Some insulin pumps can store the dosage amount and time for a few months.

4.4.3. Physical activity

Physical activity or exercise is one of the most important factors that influences BGL. Physical activity triggers consumption of glucose from blood by the body's muscle cells, which results in reduction of BGL (Association, 2007; Balducci et al., 2014; Wasserman & BZ, 1994). Only 14% of the reviewed articles have used physical activity as input features. All the different ways of recording physical activity are discussed below.

1. *Recorded using Wearables*: Among the studies that have used exercise data, 75% have used (Bertachi, Biagi, Contreras, Luo, & Vehi, 2018; Daniels et al., 2021; Georga et al., 2015; Rabby et al., 2021; Xie & Wang, 2020; Zarkogianni et al., 2015) wearable devices. These studies (Bertachi et al., 2018; Daniels et al., 2021; Rabby et al., 2021; Xie & Wang, 2020) have used the publicly available OhioT1DM (Marling & Bunescu, 2020) dataset. Xie and Wang (2020) have used all the data that is available in the dataset. Others (Bertachi et al., 2018; Daniels et al., 2021; Rabby et al., 2021) have used only the step count and the duration of exercise. The studies in Georga et al. (2015), Zarkogianni et al. (2015) have recorded and used the calories burned during the activity. All these data are recorded as numerical values. They are then preprocessed before being fed to the prediction models.
2. *Manually Recorded by Patients*: In these studies (Faruqui et al., 2019; Li & Fernando, 2016), participants have manually recorded the physical activity data. In Faruqui et al. (2019), participants have recorded the duration of exercise. The duration of exercise is then converted to calories burned using a standard food nutrient chart. The recorded activity data in Li and Fernando (2016) is activity type which is a numerical value.

4.4.4. Meal intake

Meal intake is the primary source for the increase of glucose in the blood. A meal can have carbohydrate (CHO), protein, fat, fiber, and vitamins & minerals. Protein, fat, fiber, and vitamins & minerals do not contribute to the increase of glucose in blood. The CHO of meals is the sole source of increase of glucose level in blood (Association, 2001). To predict BGL accurately, the CHO content (in grams) of each meal must be computed. For an individual to self-manage his/her diabetes, he/she needs to estimate the CHO content (in grams) of each meal. In 75% of the research articles, the collected CHO amounts are directly estimated by the diabetic persons themselves either in the amount of grams or meal type and meal size. All the different ways of handling meal intakes used in the reviewed articles are discussed below. All the meal intake values are numerical values.

1. *Patient Estimated CHO Amount*: Patients have recorded the estimated CHO amount in grams using several ways including smart phone apps (Aiello et al., 2020; Faruqui et al., 2019; Kushner et al., 2020; Li, Daniels et al., 2019; Martinsson et al., 2020; Mirshekarian et al., 2019; Xie & Wang, 2020), or manually recorded in a diary (Zecchin et al., 2013), or special applications other than common smart device applications. In few studies (Faruqui et al., 2019), patients have received training sessions from dieticians on how to manually estimate the amount of CHO, fat, and calorie. In case of others, it is not mentioned how patients have estimated the CHO amount. There were few papers which did not mention the mode of recording the CHO amount (Li & Fernando, 2016; Oviedo et al., 2019b).
2. *Patient Estimated – Dietician Annotated*: Georga et al. (2015) have collected data from fifteen Type 1 diabetic patients in free living condition. The patients have recorded the meal intake, i.e., type of food, serving size, and time themselves. Later on, a dietician has analyzed the meal and has provided the estimated CHO amount. However, this data collection is a bit of a lengthy procedure, and it needs the approval of a Dietician.
3. *Patient Estimated – System Computed Macronutrient Amount*: Karim, Vassanyi, and Kosa (2019), Karim et al. (2020) have carried out the clinical studies at the Cardiac Rehabilitation Institute of the Military Hospital, Balatonfüred, Hungary. The patients have daily activities that are similar to free living conditions. In 2015 (Karim et al., 2019), they have collected a detailed dietary log of 21 days of 16 T2DM volunteer patients. In 2019 (Karim et al., 2020), they have collected a detailed dietary log of 84 days of five volunteer patients of both Type 1 and Type 2 diabetes. The patients have logged meal nutrient information using a smartphone application and a dietary expert database. Gyuk et al. (2019) are also a part of the same study that is carried out in 2015 at the same hospital above except that they have collected data of 26 T2DM patients.
4. *Blind Estimation of Meal Absorption*: Liu et al. (2019) have carried out their clinical trial at Imperial College Healthcare NHS Trust, London, UK, on ten T1DM adults, for two weeks. Due to the unavailability of any reliable information on meal absorption, they have assumed that breakfast and snacks are fast absorption meals, and lunch and dinner are medium absorption meals.

5. *CHO Estimation from Food Image*: 6.67% of the reviewed articles that have used meal as one of the input features have used food image as the initial source of meal consumed. The data used by Zecchin et al. (2013) was collected during the European Union Seventh Framework Program (FP7) DIAdvisor project. In that project, as mentioned by the authors, the CHO intake is extracted from the meal images taken by the patients with a smartphone camera. Though what has been done with the pictures have not been mentioned. Saiti et al. (2020) have used two sources of data - one in collaboration with FN Motol University Hospital for training purposes, and for the second they have used the open D1NAMO dataset. Both datasets have included food pictures. One has used the consumed CHO estimated by the physicians, and the other has used the estimated amount of calories, quality, and whether the food is balanced or not by dieticians.

4.4.5. Other features

Besides those four critical factors researchers have used other input features as well. Faruqui et al. (2019) besides BGL, exercise and meal intake, have used weight. They have observed that there is a lack of study regarding use of self-monitoring health behaviors and weight to predict BGL for T2DM patients. Gyuk et al. (2019) besides the critical four factors, have used medication logs to predict BGL. Xie and Wang (2020) besides using the main four factors have used all other data that are available in the OhioT1DM dataset that includes data from wearables (HR, GSR, skin temperature, and air temperature), sleep, work, stress, and illness. Kushner et al. (2020) besides BGL, insulin, meal, and step count, have used age, weight, duration, heart rate, and hemoglobin A1c (HbA1c). HbA1C is the average BG level of a person over the past two to three months. Along with BGL, insulin, and meal intake, weight of the patients are used by Li, Liu et al. (2019) to forecast accurate BGL.

4.5. Model ready input data format

In general, the input data to a prediction model is a multi-dimensional sequence. Each dimensional data comes from a different input feature. CGM time-series data is a temporal sequence with a fixed period usually of 5 min. On the other hand, insulin dosage amount, meal data such as CHO, fat, calorie, etc., physical exercise time duration and intensity, etc are discrete data. Hence, on the time domain CGM and other data are not aligned. Thus, before feeding these data to a prediction model, they need to be aligned. To align them, the discrete data needs to be re-sampled every fixed period of 5 min. Finally, the resulting uniform length input matrix is ready to be fed to the prediction models.

4.6. Critical analysis

As a high volume of CGM data can be collected from T1DM patients, most of the existing research focuses on developing BGL prediction models from CGM data. However, the proposed models that have used T1DM patient data cannot be readily used to predict BGL for T2DM patients. More efforts should be given to collect data from T2DM patients. Insulin dosage recordings are mostly taken by the patients manually either by entering the amount into an insulin pump, or recording using a smartphone app, or a paper diary. The use of an insulin pump is probably overkill for the T2DM patients. A smart phone application based insulin recordings can be a feasible solution for collecting insulin intake of T2DM patients. To capture the physical activity related data of the patients, wearables such as *Fitbit* wrist bands are being used by the researchers for data input. More research needs to be aimed to identify the type of activities and intensity of the activities.

Meal consumption data are mostly recorded manually by the patients, where patients have estimated the CHO amount of their meals. Sometimes, dieticians annotate the CHO amount of the meals from meal pictures taken by the patients. However, these are either prone to be highly inaccurate or highly dependent on the knowledge of the experts. Predicting the CHO amount from the meal picture is an interesting and promising area of research. It removes the burden of manual intervention of estimating CHO amount from the patients. However, only a few papers have used meal images to record the meal intake of the diabetic patients. More research needs to be focused on capturing images of meals, predicting CHO amount, and finally predicting the BGL of the patients. The input features and the ways they have been recorded are given in Table 5.

5. Modeling techniques

In this section, we discuss the proposed modeling techniques for BGL prediction. Fig. 6 presents the taxonomy of modeling techniques. The comparative study of performance of those modeling techniques is analyzed in this Section and in Section 6.

BGL prediction is a challenging task as BGL depends upon many factors. Thus, BGL prediction models take different types of input data including daily amount of insulin ingestion, amount of meal intake, duration of physical exercise and the previous history of BGLs. Besides these four main BG movers, demographic data, HbA1c, sleep, stress, disease history, diabetes duration, smoking history, alcohol consumption history, etc. are also used. The volume of the input data also plays a key role in the BGL prediction, i.e., the more data is used, the more accurate the BGL prediction can be achieved. BGL prediction models also need to consider insulin absorption dynamics, insulin effects on glucose appearance rates and disposal rates, meal consumption dynamics, and variability in physiology of patients. Many research efforts have been focused on finding BGL prediction models. These BGL predictive models can be classified into the following three main categories depending on the computational methods they used: (i) *Physiological* models (PM), (ii) *Data Driven* models, and (iii) *Hybrid* models (HM). The categorized modeling techniques are given in Table 6 and described in detail as follows. All the articles that are studied are listed in Tables 7 and 8.

Table 5
Category of different input features.

Meal	Patient Estimated CHO Amount	Aiello et al. (2020), Faruqui et al. (2019), Kushner et al. (2020), Li, Daniels et al. (2019), Martinsson et al. (2020), Mirshekarian et al. (2019), Xie and Wang (2020), Zecchin et al. (2013)	Patients estimated the CHO amount in grams of their meals.
	Patient Estimated – Dietician Annotated	Georga et al. (2015)	Patient recorded meal type and size, dietician estimated CHO amount.
	Patient Estimated – System Computed	Gyuk et al. (2019), Karim et al. (2019, 2020)	Patient recorded meal type and size, system computed amount of CHO, protein, and lipids.
	Blind Estimation of Meal Absorption	Liu et al. (2019)	Assumed meal absorption rate: Fast, Medium, Slow.
	CHO Estimation from Food Image	Saiti et al. (2020), Zecchin et al. (2013)	CHO amount is computed from meal pictures taken by patients.
BG Reading	CGM	Alfian et al. (2020), Bosoni et al. (2020), Fox et al. (2018), Karim et al. (2020), Li, Daniels et al. (2019), Liu et al. (2019), Mirshekarian et al. (2017), Oviedo et al. (2019b), Zecchin et al. (2013)	Used CGM devices to record BGL.
	Fingertip BGL	Karim et al. (2019), Kushner et al. (2020), Martinsson et al. (2020), Xie and Wang (2020), Zhu et al. (2020)	Used ordinary fingertip blood glucose meters.
	FGM	Bosoni et al. (2020)	Used a flash glucose meter sensor under skin.
Insulin dosage	Insulin Pump Reading	Aiello et al. (2020), Kushner et al. (2020), Li, Daniels et al. (2019), Martinsson et al. (2020), Mirshekarian et al. (2019), Oviedo et al. (2019b), Sun, Jankovic, Bally et al. (2018), Xie and Wang (2020)	Used insulin pumps to apply insulin.
	Smartphone App	Aiello et al. (2020), Faruqui et al. (2019), Kushner et al. (2020), Li, Daniels et al. (2019), Martinsson et al. (2020), Mirshekarian et al. (2019), Xie and Wang (2020)	Recorded insulin dosages using a smartphone app.
	Diary Entry	Georga et al. (2015)	Record insulin dosages using a paper diary.
Physical Activity	Wearables	Bosoni et al. (2020), Georga et al. (2015), Mirshekarian et al. (2019), Xie and Wang (2020), Zhu et al. (2020)	Used fit band like wearables to record physical activity (step counts).
	Patient Estimated	Gyuk et al. (2019), Li, Daniels et al. (2019)	Patients manually estimated physical activity in minutes or calories burned.

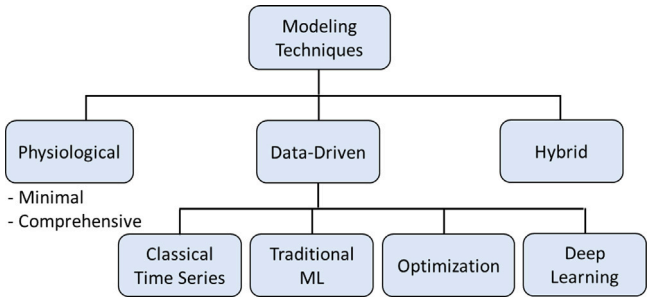


Fig. 6. An overview of different modeling techniques.

Table 6
Category of different modeling techniques.

Models	Category	Ref.	Description
Physiological	Minimal	Bergman et al. (1979), Bock et al. (2015), Liu et al. (2019)	Various organs and cells are lumped into one or more compartments.
	Comprehensive	Sorensen (1985)	All the organs that are involved in the glucose and insulin dynamics are modeled separately.
Data Driven	Traditional ML	Alfian et al. (2020), Cichosz et al. (2021), DAntoni et al. (2020), Frandes et al. (2017), Georga et al. (2015), Hamdi et al. (2018), Hidalgo et al. (2017), Marcus et al. (2020), Montaser et al. (2019), Ozogur et al. (2020), Velasco et al. (2017), Yang et al. (2018), Yu, Turksoy et al. (2018), Zarkogianni et al. (2015)	Used traditional machine learning techniques for prediction.
	Deep Learning	Aiello et al. (2020), Alfian et al. (2020), Aliberti et al. (2019), Armandpour et al. (2021), Dudukcu et al. (2021), Faruqui et al. (2019), Fox et al. (2018), Li, Daniels et al. (2019), Li, Liu et al. (2019), Martinsson et al. (2018), Mhaskar et al. (2017), Mirshekarian et al. (2017, 2019), Mosquera-Lopez et al. (2019), Nemat et al. (2022), Rabby et al. (2021), Wang et al. (2020), Xie and Wang (2020), Zhu et al. (2020, 2022)	Used DL techniques for prediction.
Hybrid	Physiological + ML	Bunescu, Struble, Marling, Shubrook, and Schwartz (2013), Karim et al. (2020), Saiti et al. (2020)	Minimal model and Traditional ML techniques.
	Physiological + Optimization	Contreras, Oviedo, Vettoretti, Visentin, and Vehi (2017), Gyuk et al. (2019)	Genetic algorithm, Grammatical evolution, etc. take physiological models as input.

5.1. Physiological models

Physiological models use the actual knowledge of physiology to capture the glucose–insulin dynamics that is described in Section 2.1. The major organs of the human body that are involved in the glucose–insulin dynamics include intestines, brain, lever, heart, kidney, pancreas, brain, lungs, vascular organs, peripheral tissues, etc. These models use multi-compartmental models including carbohydrate digestion model, glucose absorption model, insulin absorption model, insulin action model, etc. to mathematically represent some or all the above mentioned organs. The variations in these models depend on the underlying physiological aspects and the number of organs represented by the used compartments. Based on these two factors, the physiological models can further be categorized into (i) minimal models and (ii) comprehensive models (Balakrishnan, Rangaiah, & Samavedham, 2011).

5.1.1. Minimal models

In *minimal models*, various organs and cells are ignored due to their minimal contribution to the glucose-insulin dynamics. These models used a few compartments to represent the organs that contribute the most, nonetheless still achieving satisfactory results with computational efficiency. *Bergman's Minimal* model (Bergman et al., 1979) is one of the widely used minimal models. It uses the plasma glucose, the plasma insulin, the insulin action on glucose, and the meal absorption compartments.

Bock et al. (2015) have proposed a new linear dynamic Therapy Parameter-based compartmental Model (TPM) which is based on the standard therapy static prediction model and identifiable on BGL measurements only. The model and physician-set therapy parameters are intrinsically correlated. The model equations of the TPM are derived from the Bergman Minimal model (Bergman et al., 1979). They have proved that with only a few model parameters their simple model can perform slightly better than other models and may be safely used by physicians identifying therapy parameters. However, their validation data lacks variation. Liu et al. (2019) have used a compartmental composite model of glucose–insulin dynamics that used a deconvolution technique on CGM signals (Herrero et al., 2012). Additionally, they have used the CHO amount, meal absorption information, and insulin boluses as exogenous inputs. Their result shows that inclusion of meal absorption information improved the prediction accuracy from an RMSE of 32.69 mg/dL to 30.36 mg/dL for 60-min PH. Yet, their model may give significant prediction error in the presence of physical activity. Albers et al. (2017) have used (i) a data assimilation machine that uses a modified dual unscented Kalman filter and (ii) a Gaussian process regression model on two physiological models to produce accurate enough yet quick personalized models for T2DM individuals. They have applied their models on 23 days of carbohydrate (CHO) intake amount and discrete BG readings of 3 T2DM patients to predict two hours after meal BGL. This is the first attempt to predict two hours after meal BGL using discrete BG readings of T2DM patients.

5.1.2. Comprehensive models

In the comprehensive models, all the organs and the cells that are involved in the glucose and insulin dynamics are considered and modeled separately using separate compartments. Tiran, Galle, and Porte (1975) have firstly proposed a comprehensive physiological model. Guyton et al. (1978) and Cobelli, Federspil, Pacini, Salvan, and Scandellari (1982) have proposed other physiological comprehensive models. Sorensen (1985) have provided a detailed mathematical explanation of the work of Guyton et al. Due to their high complexity, comprehensive models are not being used by the larger research community.

5.2. Data driven models

In this section, different aspects of data driven models for prediction of BGL are discussed. The data driven models use a large volume of diabetic patients' related data to extract valuable patterns in the data. These models do not depend on underlying domain related expert knowledge i.e., human physiological information. There is no single data-driven model that has been used by the literature. Rather, we can broadly divide the proposed models into the following categories: *classical auto-regressive time series models*, *traditional machine learning models*, *optimization models*, and *Deep Learning models*. They are described as follows.

1. *Classical Auto-regressive (AR) Time Series Models*: Auto-regressive (AR) models represent random processes. In diabetes research, they are used to represent time-varying BG values. These models specify that the future predicted BG values depend linearly on the previous BG values. These models are used in Frandes et al. (2017), Montaser et al. (2019), Xie and Wang (2020), Yang et al. (2018). Frandes et al. (2017) have achieved a high prediction performance with an RMSE of 5.83 mg/dL applying the regime-switching AR model on 5.73 days BG values of 17 T1DM patients. They have shown that their model can provide fast prediction with high accuracy and less patient intervention. Montaser et al. (2019) have achieved a good performance when they have first clustered the glucose values, determined the local prediction model using Seasonal AutoRegressive Integrated Moving Average with eXogenous inputs (SARIMAX), and have finally obtained future BG value by integrating the local models using fuzzy membership. Majority of the articles that have used AR models usually use only BG values as input.
2. *Traditional Machine Learning (ML) Models*: Traditional Machine Learning (ML) models use a statistical model on a collected dataset to predict a value or class, or to cluster samples, etc. The studied works have used numerous ML techniques including Extreme Learning Machines (ELM), Linear Regression (LR), Recursive Least Squares (RLS), ridge regression and its variants, Support Vector Regression (SVR), fuzzy strategy, etc. Traditional ML models also include Artificial Neural Networks (ANNs) that mimic the way biological neurons signal one another. These ANN models comprise one input layer of nodes with the input data, usually one hidden layer of nodes with activation functions, and an output layer. These traditional ML models are used in DAntoni et al. (2020), Hamdi et al. (2018), Kushner et al. (2020), Li and Fernando (2016), Yu, Turksoy et al. (2018), Zarkogianni et al. (2015). Hamdi et al. (2018) have proposed an SVR model that is optimized by Differential Evolution (DE). They have achieved a good performance when applied on 14 days BG values of 12 T1DM patients with an RMSE of 10.78 mg/dL for 30-min PH. Zarkogianni et al. (2015) have achieved a similar performance applying Self-Organizing Maps (SOMs) on a slightly smaller cohort. On the other hand, Cichosz et al. (2021) have received just an above average performance of 21.63 mg/dL for 30-min PH applying ANN and a penalty score on a large dataset. Nevertheless, they have achieved a good performance of 99.72% correct prediction in the A+B regions in CEG for 30-min PH.
3. *Optimization Models*: Optimization methods are a global search strategy that searches for the best combination of the hyper-parameters in the input search space. These models are used in Faruqui et al. (2019), Hamdi et al. (2018), He and Wang (2020), Hidalgo et al. (2020, 2017), Ozogur et al. (2020), Velasco et al. (2017). In Hidalgo et al. (2017), Velasco et al. (2017), Genetic Programming (GP) and Grammatical Evolution (GE) are used as the prediction models. Velasco et al. (2017) have achieved an excellent 99.56% of correct prediction in the A+B regions in a CEG as GE is not susceptible to obtaining critical errors. Hidalgo et al. (2017) have presented a set of evolutionary methods based on GP and GE. They have trained their models on more than 10 days of history of BGL, insulin dosages, and meals of 10 T1DM patients as input. Their best model has achieved 91.21% correct predictions in the A + B regions of CEG for 2-hr PH and performs better than or equal to the traditional ML methods. In He and Wang (2020), Ozogur et al. (2020), PSO is used as an optimization algorithm to optimize model hyper-parameters. In He and Wang (2020), it is observed that PSO has improved the RMSE from 35.62 mg/dL to 15.29 mg/dL for 30-min PH. Faruqui et al. (2019) have used Bayesian optimization to optimize hyper-parameters.
4. *Deep Learning Models*: Deep Learning models are ANN models with a large number of hidden layers. The studied works have used numerous DL techniques including RNN, LSTM, Bidirectional LSTM (BiLSTM), CNN, Dilated CNN, attention mechanism, RNN Autoencoder, etc. These models are used in Faruqui et al. (2019), Martinsson et al. (2018), Mhaskar et al. (2017), Mirshekarian et al. (2019), Sun, Jankovic, Bally et al. (2018), Zhu et al. (2020) and Rabby et al. (2021). Wang et al. (2020) have achieved the best prediction performance with an RMSE of 3.03 mg/dL for 30-min PH. They have used Variational Modal Decomposition (VDM) to get a relatively stable BG time-series, LSTM to find the prediction model, and PSO to optimize the hyper-parameters. Alfian et al. (2020) with Multilayer Perceptron (MLP), Rabby et al. (2021) with stacked LSTM and kalman filter, Armandpour et al. (2021) with embedding vectors, Bidirectional RNN (BiRNN), and attention mechanisms have achieved similar prediction performances. Lately, there is a drastic increase in the use of DL methods specially the RNN based LSTM models in BGL prediction due to its success in capturing long-term temporal sequences. They are achieving good prediction performance as observed from above works. They can directly extract local and global features as required without the need of feature engineering.

Table 7
BGL prediction reviewed literature (Real data): Metric CEG has a % sign, Glucose (G), Insulin (I), Meal (M), Exercise (E), Auto Regressive Time Delayed (ARTiDe), Fuzzy C-Means (FCM), Physiological (Physio.), Ridge Regression (RR), Kernel RR (KRR), Relative Absolute Deviation (RAD), Sum of Squares Glucose Prediction Error (SSGPE), Temporal Gain (TG), Linear Regression (LR), Genetic Programming (GP), Markov Chain Monte Carlo (MCMC), Differential Evolution (DE), Performances (Perf.), Kernel Canonical Correlation Analysis based on PSO (PSO-KCCA), Echo State Network (ESN), Temporal Convolutional Network (TCN).

Ref	Year	Input features	Dataset	Modeling technique	Perf. $\frac{RMSE}{CEG(\%)}$		
					PH=30	PH=60	PH=120
Physiological Models							
Bock et al. (2015)	2015	G	12 T1DM, 10 days	Physio.	98.00%	96.50%	87.00%
Albers et al. (2017)	2017	G	3 T2DM, 539 readings	Physio.	NA	NA	NA
Liu et al. (2019)	2019	G, I, M	10 T1DM, 2 weeks	Physio.	17.67	30.36	40.46
					99.05%	97.12%	95.27%
Traditional ML Models							
Georga et al. (2015)	2015	G, I, E, M	15 T1DM, 5 to 22 days	Extreme Learning Machine	6.10	–	–
Zarkogianni et al. (2015)	2015	G, ΔG, E	10 T1DM, 6 days	SOM	11.42	19.58	31.00
					93.84%	89.34%	82.37%
Li and Fernando (2016)	2016	G, I, E, M	70 subjects, months (Kahn, 1994)	Pre-clustered personalized LR model	–	–	36.70
Frandes et al. (2017)	2017	G	17 T1D, 5.73 days	Regime-switching AR	5.83	7.43	–
Velasco et al. (2017)	2017	G, I, M	1 T1DM, 46 days	GE, GP, MCMC	–	–	99.56%
Hidalgo et al. (2017)	2017	G, I, M	10 T1D, 10 days	GE, GP	98.18%	94.45%	91.21%
Hamdi et al. (2018)	2018	G	12 T1DM, 14 days	SVR, DE	10.78	12.95	–
Yu, Turksoy et al. (2018)	2018	G	3 T1DM, 3 days	RLS	11.71	–	–
					94.96%	–	–
Yu, Rashid et al. (2018)	2018	G	3 T1DM subjects, 60 h	Kernel sparse filtering algorithms	17.57	–	–
					93.83%	–	–
Yang et al. (2018)	2018	G	100 T1DM, T2DM, 4 Days	ARIMA	RAD:5.41	TG:0.48	–
					SSGPE:7.21	–	–
Montaser et al. (2019)	2019	G	10 T1DM, 1080 h	FCM, SARIMAX, Fuzzy strategy	7.75	12.17	–
Hidalgo et al. (2020)	2019	G,I,M	5 T1DM, 15 days	Markov, GE, RF, Ensemble	98.84%	97.77%	94.17%
Xie and Wang (2020)	2020	G, I, E, M	6 T1DM, 8 weeks	ARX, RR	19.48	–	–
Amar et al. (2020)	2020	G,I	141 T1DM, 64.4 days	ANN	24.27	42.87	–
					99.26%	95.82%	–
Marcus et al. (2020)	2020	G	11 T1DM, 7-50 days	4 KRR variants	20.48	–	–
					93.40%	–	–
DAntoni et al. (2020)	2020	G	6 T1DM, 8 weeks	ARTiDe Jump NN	18.80	–	–
Kushner et al. (2020)	2020	G, I, M	24 T1DM, 37.8 days	Physio. based FNN	–	28.00	38.00
					–	97.00%	94.00%
He and Wang (2020)	2020	G	10 T1D, 4 days	PSO-KCCA	16.40	–	–
Ozogur et al. (2020)	2020	G	6 T1DM, 8 weeks	Fuzzy PSO algorithm	27.46	–	–
					97.95%	–	–
Ozogur et al. (2020)	2020	G	89 T1DM, 26 h	Fuzzy PSO algorithm	38.93	–	–
					97.46%	–	–
Cichosz et al. (2021)	2021	G	225 T1DM, 6 months	ANN, penalty	21.63	37.68	–
					99.72%	97.96%	–
Deep Learning Models							
Mirshekarian et al. (2017)	2017	G, I, M	10 T1DM, 400 days	LSTM	21.4	38.0	–
Mhaskar et al. (2017)	2017	G	25 T1DM children, 4000 readings	Semi Supervised DL	93.23%	–	–
Martinsson et al. (2018)	2018	G	6 T1DM, 8 weeks	LSTM, Physio. loss function	20.10	33.20	–
Sun, Jankovic, Bally et al. (2018)	2018	G	20 T1DM, 1 week, 6975 readings	LSTM, BiLSTM	21.75	36.92	–
Zhu et al. (2018)	2018	G,I,M	6 T1DM, 8 weeks	Dilated CNN, WaveNet	21.72	–	–
Fox et al. (2018)	2018	G	40 T1DM, 3 years, 550k readings	RNN Autoencoder	MAPE:4.59		–
Aliberti et al. (2019)	2019	G	451 T1DM, 398578 samples	LSTM	5.93	13.21	–
					99.94%	99.46%	–
Li, Daniels et al. (2019)	2019	G, I, E, M	10 T1DM, 6 weeks	CNN	21.07	33.27	–
Li, Liu et al. (2019)	2019	G, I, M	10 T1DM, 180 days	Dilated CNN	19.19	31.78	–
Li, Liu et al. (2019)	2019	G, I, E, M	6 T1DM, 8 weeks	Dilated CNN	19.28	31.83	–

(*continued on next page*)

Table 7 (*continued*).

Ref	Year	Input features	Dataset	Modeling technique	Perf. $\frac{RMSE}{CEG(\%)}$		
					PH=30	PH=60	PH=120
Mirshekarian et al. (2019)	2019	G, I, M	6 T1DM, 8 weeks	LSTM	17.99	28.20	–
					–	97.31%	–
Faruqui et al. (2019)	2019	G, E, M	10 T2DM, 6 months	LSTM, Transfer learning	99.34%	–	–
Mosquera-Lopez et al. (2019)	2019	G, I	Train:124 T1DM,27466 days,Test:10 T2DM, 4 weeks	LSTM, Error correction	7.55	–	–
Martinsson et al. (2020)	2020	G	6 T1DM, 8 weeks	LSTM, Physio. loss function	18.87	31.40	–
Wang et al. (2020)	2020	G	56 T1DM, 125 h	VMD, LSTM, PSO	3.03	5.72	–
Alfian et al. (2020)	2020	G	12 T1DM children, 7 days	MLP	6.31	15.33	–
Rubin-Falcone et al. (2020)	2020	G	6 T1DM, 8 weeks	MLP	18.20	–	–
Zhu et al. (2020)	2020	G,I,E,M	6 T1DM, 8 Weeks	Dilated RNN	18.90	–	–
Zhu et al. (2020)	2020	G, I, M	12 T1DM, 8 Weeks	GAN	18.34	32.31	–
					98.82%	95.98%	–
Li et al. (2020)	2020	G	10 T1DM, 3 days	Improved ESN (an RNN variant)	34.17	–	–
Rabby et al. (2021)	2021	G,I,E,M	6 T1DM, 8 weeks	Stacked LSTM, Kalman filter	5.89	17.24	–
Dudukcu et al. (2021)	2021	G	12 T1DM, 8 weeks	LSTM, GRU, WaveNet	21.90	35.10	–
Deng et al. (2021)	2021	G	12 T1DM, 8 weeks	CNN, transfer learning, data augmentation	19.08	33.80	–
Daniels et al. (2021)	2021	G	12 T1DM, 8 weeks	DL, Transfer Learning	18.80	31.80	47.20
Armandpour et al. (2021)	2021	G	38 T1DM, 400K samples	Embedding vectors, BiRNN, Attention	8.91	14.81	–
Nemat et al. (2022)	2022	G	12 T1DM, 8 weeks	LR, LSTM, BiLSTM, Fusion	19.63	33.45	–
Yang et al. (2022)	2022	G,I,M,time	12 T1DM, 8 weeks	Channel DL	18.93	–	–
Zhu et al. (2022)	2022	G,I,M	12 T1DM, 8 weeks	Attention, BiRNN, Evidential Regression	18.64	31.07	–
Hybrid: Physiological + Traditional ML Models							
Zecchin et al. (2013)	2013	G, M	20 T1DM , 2 or 3 days	Linear, Meal Absorp., FNN, Skip conn.	16.60	–	–
Bertachi et al. (2018)	2018	G, I, E, M	6 T1DM, 8 weeks	Physio. features, ANN	19.33	31.72	–
Gyuk et al. (2019)	2019	I, M	26 T1DM and T2DM, 3 weeks	Physio., GA	–	–	
					–	–	
Karim et al. (2020)	2020	G, I, M	4-T2DM, 1-T1DM, 3 weeks	Physio., FNN	–	20.16	31.50
					–	96.46%	98.13%
Saiti et al. (2020)	2020	G, I, M	Train: 2 T1DM, 30 days; Test: 4 T1DM, 1 day	ARX, SVR; linear, bagging, boost metaregressor	17.66	35.65	–
					92.70%	71.79%	–

Researchers are recently combining different techniques to improve prediction performances, for example, Yu, Turksoy et al. (2018) have proposed a data-driven prediction strategy employing four adaptive filtering mechanisms: Recursive Least Squares (RLS), Extended RLS (EX-RLS), Kernel RLS (KRLS), extended KRLS (EX-KRLS) for training the candidate models using 3 days of CGM values of 10 adults as input. The outputs of the models with lowest prediction error and lowest learning error are then fused with a linear stack weighted method. Finally, an additional smoothing filter is applied to mitigate the noise and abrupt variability to predict the BGL. They have demonstrated that their fusion model is capable of accurate prediction. Hidalgo et al. (2020) have presented a three-step BGL prediction methodology that include (i) data augmentation using Markov chain, (ii) base BGL prediction model based on Grammatical Evolution (GE) and RF, and (iii) the final BGL prediction model using ensemble methods: bagging and Univariate Marginal Distribution Algorithm (UMDA). They have improved on their previous work (Hidalgo et al., 2017) and have achieved on the CEG plot for zone A+B, 98.84% correction prediction for 30-min PH, 97.11% correction prediction for 60-min PH, and 94.17% correction prediction for 120-min PH. In another work, Dudukcu et al. (2021) have proposed a prediction model that is a weighted fusion of three intermediate models using LSTM, Gated Recurrent Unit (GRU), and WaveNet, respectively. Their fusion model has provided promising results.

The proposed models in the reviewed articles need to be used in some sort of mobile platforms. Some of the research works have demonstrated that their models are computationally efficient to be used online in mobile platforms. In this work (Aliberti et al., 2019), the researchers have proposed a Nonlinear AutoRegressive (NAR) Neural Network and an LSTM network driven blood glucose prediction systems. These models are trained on glucose values of a large and heterogeneous cohort of 451 T1DM patients. The LSTM model outperforms the NAR model and has achieved an RMSE of 5.93 mg/dL for 30-min PH. However, the simple NAR model with just 8 regressors performs well for 30-min PH and looks promising for online BGL prediction. Another research group (Yu, Rashid et al., 2018) has developed adaptive kernel sparse filtering algorithms called *Approximate Linear Dependency KRLS* (ALD-KRLS), and *Surprise Criterion KRLS* (SC-KRLS) to provide real-time BGL prediction using 3 days of CGM values of 10 adults and has achieved a similar result with an RMSE of 17.57 mg/dL and 93.83% correct prediction in the A+B regions in CEG for 30-min

Table 8
BGL Prediction reviewed literature (Simulated data): Metric CEG has a % sign, Glucose (G), Insulin (I), Meal (M), Exercise (E), Least Squares (LS).

Ref	Year	Input features	Dataset	Modeling technique	Perf. $\frac{RMSE}{CEG(\%)}$		
					PH=30	PH=60	PH=120
Physiological Models							
Bock et al. (2015)	2015	G	10 T1DM, 4 days	Physio.	98.50%	97.25%	88.00%
Liu et al. (2019)	2019	G, I, M	10 T1DM, 2 weeks	Physio.	10.06	22.56	34.74
					99.62%	97.15%	94.16%
Traditional ML Techniques							
Zhao and Yu (2015)	2015	G, I	30 T1DM, 6 days	ARX based on LS	11.00	18.60	–
Yu, Turksoy et al. (2018)	2018	G	10 T1DM, 3 days	RLS and 3 variants	9.86	–	–
					99.65%	–	–
Yu, Rashid et al. (2018)	2018	G	10 T1DM, 6 days	Kernel sparse filtering algorithms	13.74	–	–
					91.96%	–	–
Deep Learning Models							
Mirshekarian et al. (2019)	2019	G, I, M	40 T1DM, 600 days	LSTM	1.23	2.27	–
Mirshekarian et al. (2019)	2019	G, I, M	10 T1DM, 90 days	LSTM	2.93	4.92	–
					–	99.47%	–
Li, Liu et al. (2019)	2019	G, I, M	10 T1DM, 180 days	Dilated CNN	8.80	19.90	–
Li, Daniels et al. (2019)	2019	G,I,E,M	10 T1DM, 360 days	CRNN	9.38	18.87	–
Zhu et al. (2020)	2020	G, I, M	10 T1DM, 360 days	Dilated RNN	7.8	–	–
Aiello et al. (2020)	2020	G, I, M	Train: 100 T1DM, 4 days, Test(real data): 1 T1DM, 30 days	Stacked LSTM	24.83	36.55	–
Hybrid Models							
Contreras et al. (2017)	2017	G, I, M	100 T1DM, 14 days	Physio., GE	20.96	–	–
					98.31%	–	–

PH. These models are insensitive to abnormal CGM readings and have improved computational efficiency. Frandes et al. (2017) have investigated the use of regime-switching Logistic Smooth Transition AR (LSTAR) predictive model on the multidimensional phase-space representation of 5.73 days of nonlinear chaotic CGM time series data of 17 T1DM patients. They have shown that their model can provide fast prediction and high accuracy with an RMSE of 5.83 mg/dL for 30-min PH.

It is also observed that some research works are using data preprocessing steps to improve prediction performance. Aliberti et al. (2019), have used a denoising preprocessing step using Tikhonov regularization (Ito & Jin, 2014) on their data. All their proposed Nonlinear AutoRegressive (NAR) Neural Network and LSTM network driven blood glucose prediction systems perform better due to the preprocessing step. Li, Liu et al. (2019) have introduced GluNet, a personalized deep NN framework consisting of data preprocessing, label transform and recover, dilated CNN, and post-processing to predict the probabilistic distribution of short-term BGLs based on historical data as from (Li, Daniels et al., 2019). More performance analysis on the data-driven models are discussed in Section 6.

5.3. Hybrid models

Hybrid models are usually a combination of two or more processes. In most of the cases, this model is a two step process. In the first phase or step, physiological models are designed to represent CHO digestion, glucose absorption, insulin absorption, exercise model, etc. In the second phase, the output from the physiological modes is provided as input to a data-driven model, which eventually predicts the BGL.

Zecchin et al. (2013) have used a linear prediction algorithm proposed in Sparacino et al. (2007) with past CGM values as input to produce the glucose prediction model. They have used the physiological meal simulation model proposed in Dalla Man, Camilleri, and Cobelli (2007) to get the glucose rate of appearance. Along with these two, other 6 inputs including insulin dosages are fed to a fully connected FNN that has one hidden layer with eight neurons with a tangent sigmoid activation function and an output layer. The final output comes by combining the linear glucose prediction model, the prediction error from the NN, and an additional skip connection (the jump part) from input of NN to the output.

Contreras et al. (2017) have applied a hybrid approach on simulated data. They have built physiological models (Insulin on board, CHO absorption) for insulin and meal. Then they take these models as input and have used symbolic regression through Grammatical Evolution (GE) (Hidalgo et al., 2014) to determine the underlying glucose dynamics. Grammatical evolution is a search algorithm that uses a grammar consisting of a set of rules to build expression for BG considering previous BGLs, CHO intake, and insulin dosage. These are used to generate predictive time series models according to a predefined objective function. They have used a penalizing fitness function as an objective function to stop deviations from the target BGL. Bertachi et al. (2018) have presented a model based on Artificial Neural Networks (ANN), where feature space has 5 features coming from CGM, Insulin Pump, Fit band readings, and physiological models. Gyuk et al. (2019) have proposed an algorithm that combined two things. One is a two compartmental algorithm (Arleth, Andreassen, Orsini-Federici, Timi, & Massi-Benedetti, 2000) for meal absorption. The other is an

advanced glucose control system that has used a small set of parameters yet maintains powerful descriptive capabilities based on Delay Differential Equations (DDE) (Palumbo, 2011). They have extended the combined model with personalized parameter training with Genetic Algorithm (GA) and Nelder–Mead method and used a diurnal parameter profile as a representation of the natural biorhythm. They have achieved an RMSE of 29.16 mg/dL and 92.5% clinically acceptable results for 60-min PH which proves the practical applicability of their model. However, they may have oversimplified their BGL control system with fewer parameters.

Karim et al. (2020) have used a two compartmental physiological model proposed in Arleth et al. (2000), which consists of stomach compartment, gastric emptying, intestinal compartment, breakdown of starch, and intestinal glucose absorption. This model takes the macro-nutrient content of every logged meal as input and provides the glucose absorption curve. This absorption curve, insulin dosage, and BGLs are then fed as input to a FNN with 20 hidden layers. Applying the NN model on 3 weeks' data of 5 patients they have achieved an RMSE of 20.16 mg/dL for 60-min PH and 39.17 mg/dL for 2-hr PH. Saiti et al. (2020) have used an open compartment model (Wilinska et al., 2005) to model the impact of bolus insulin absorption. The absorption curve, BG readings, and CHO amounts are fed to ARX and SVR models separately. They have used three ensemble methods: linear, bagging, and boost metaregressor for final prediction. They have demonstrated that both physiological models and ensemble methods have improved performance. Their best model has achieved an RMSE of 17.66 mg/dL for 30-min PH when trained on 30 days of BG, insulin, and meal data of 2 T1DM patients and tested on 1 day data of 4 T1DM patients.

5.4. Comparative analysis

Various modeling techniques have been used to predict BGL. These are being categorized in Table 6. Among them, DL techniques (46%) and traditional ML techniques (41%) have been used majority of the time followed by AR models (13%), optimization models (10%), hybrid models (6.5%) and physiological models (6.5%). Among the traditional ML techniques, ANN is the most popular followed by SVR models. Among the DL techniques, LSTM models are the most popular followed by RNN and CNN.

Physiological models have been used by the research community the least due to the need of expert knowledge, the lack of patient personalizing capabilities, and the inability to consider all the physiological mechanisms. Nonetheless, the research community is more willing to use them as a data pre-processing step of the data-driven models.

Following the current trend in other areas, the researchers are using data driven modeling techniques in the area of diabetes research. This is due to the recent (i) availability of plethora of data from Internet of Thing (IOT) devices like CGM devices, Insulin pumps, Wearable fit bands, EHRs, etc, and (ii) advent of high performing Graphical Processing Units (GPUs) to run the data driven techniques processing a lot of data in short period of time.

6. Discussion

6.1. Performance analysis

Blood glucose prediction is a challenging task due to quite a few influencing factors such as meal and insulin absorption rate, glucose consumption due to physical activity, etc. There are a lot of variations with respect to diabetes type, dataset type, dataset size, input features, prediction horizon, and performance metrics. These variations make it a challenging task to do a fair comparison among the predictive performances of different modeling techniques. A modeling technique with best performance with a proprietary dataset may have lower performance with other datasets than other modeling techniques. Yet, few intriguing insights can be taken away for future improvement.

1. *Influence of Input Features*: Input data used by reviewed articles vary. Mostly, they are various combinations of the four core input features: previous glucose readings, insulin dosages, physical activities, and meal intakes. Among the research articles that have used physiological models [Table 7], we can see that Bock et al. (2015) have used only previous glucose readings. Their model's correct prediction is 98% in the A+B regions in Clarke Error Grid (CEG) for 30-min PH. The model in Gyuk et al. (2019) has performed worse without using CGM readings as input. They have used only insulin and meal intake. The model proposed by Liu et al. (2019) has performed better when they have used more information in addition to previous CGM readings.
   From the performances of the proposed algorithms in Table 7, it is observed that all the high performing modeling techniques with respect to the statistical metric RMSE have used previous glucose readings. Majority of them have only used previous glucose readings. The best performing model by Wang et al. (2020) (RMSE: 3.03 mg/dL) has used VMD, LSTM, and PSO on 125 h of history of previous glucose readings of 56 T1DM patients. There are few second best performing models including (Frandes et al., 2017) (RMSE: 5.83 mg/dL) who has used Regime-switching AR on glucose readings only, Georga et al. (2015) (RMSE: 6.10 mg/dL) who has used ELM on BG, meal, insulin, and exercise data, Aliberti et al. (2019) (RMSE: 5.93 mg/dL) who has used LSTM on glucose readings only, Alfian et al. (2020) (6.31 mg/dL) who has used MLP on glucose readings only, Rabby et al. (2021) (RMSE: 5.89 mg/dL) who has used stacked LSTM, and kalman filter on BG, meal, insulin, and exercise data.
   Among the articles that have provided both statistical metric RMSE and clinical metric CEG, the best performance is observed by Aliberti et al. (2019). They have used an LSTM model only on previous history of BG values. They have achieved an RMSE of 5.93 mg/dL and correct prediction of 99.94% in the A+B regions in CEG for 30-min PH and an RMSE of 13.21 mg/dL and correct prediction of 99.46% in the A+B regions in CEG for 60-min PH. This is a very good prediction performance. They

have shown that LSTM can correctly predict future BG values with glucose values as the only input feature. Other closely performing models include (Liu et al., 2019) that has used a physiological model, Cichosz et al. (2021) that has used an ANN model and a penalty score for error, and Zhu et al. (2020) that has used GAN.

The study by Zhu et al. (2018) has observed that insulin and CHO have significant impacts on BGL prediction. RMSE increases by 0.8 mg/dL when these two features are used with CGM data. In this study (Mosquera-Lopez et al., 2019), the authors have shown that the prediction accuracy is improved only by 1% after including insulin as an additional input to CGM readings. The authors in Zarkogianni et al. (2015) have observed that addition of physical activity data as input features leads to better BGL prediction performance specially in the hypoglycemic range. In another work (Zhu et al., 2020), the authors have experimented with different combinations of input features and have found that glucose, insulin, meal, and time combination performs the best. In the study by Karim et al. (2020), the authors have stipulated that the impact of some BG influencing input features such as emotions are hard to quantify. They also have observed that additional domain knowledge coming from an absorption model provides better predictability than the use of raw meal nutrient values. It is observed in the work by Mirshekarian et al. (2019) that the addition of meal and insulin data to the existing input features provided the biggest performance improvement. They have also noticed that the improvement from adding the time of the day is statistically significant, and addition of skin conductance and heart rate to the existing input features BG, insulin, meals, and the time of day improve prediction performance. Rubin-Falcone et al. (2020) have found only glucose (CGM and finger-stick), insulin, CHO, and time of the day to be helpful.

Xie and Wang (2020) have found that addition of heart rate as input feature only provides marginal improvement of prediction performance compared to insulin and CHO. They have suggested that this might be due to less frequent heart rate data compared to other input features data. In the study by Rabby et al. (2021), they have observed that among the data in the OhioT1DM dataset basal insulin, sleep, heart rate, GSR, and skin temperature either do not have any effect, or sometimes have adverse effects on the prediction performance. In their study, the most influencing feature combination is history of BG readings, CHO amount, bolus insulin, and step count. Some (Oviedo et al., 2016) say that using CHO and insulin together as input features is redundant and in addition to that multiple input features make the prediction task harder.

2. *Proprietary vs Publicly Available Datasets*: Majority of the reviewed articles have used proprietary datasets. The problem with the proprietary datasets is that it becomes difficult to compare the prediction models that use different datasets. There are a few publicly available datasets as listed in Table 4. Among those, the OhioT1DM dataset is the most popular (Bertachi et al., 2018; DAntoni et al., 2020; Li, Liu et al., 2019; Martinsson et al., 2018, 2020; Mirshekarian et al., 2019; Ozogur et al., 2020; Xie & Wang, 2020; Zhu et al., 2020) and is used 31.15% of the time. It would be fair to compare the performances of the research works that have used the OhioT1DM dataset. It is observed from the research works that have used the OhioT1DM dataset that the best performance is achieved in the research work done by Rabby et al. (2021) with an RMSE of 5.89 mg/dL. They have proposed a stacked LSTM based prediction model that they have trained on eight weeks of kalman filter smoothed CGM readings, insulin dosages, exercise, and CHO amount of six T1DM patients. They have demonstrated that kalman filter smoothing improved prediction performance. The second best performance is achieved in the research work of Mirshekarian et al. (2019) with an RMSE of 17.99 mg/dL for 30-min PH. They also have applied LSTM on previous glucose history, insulin dosage, physical activity, and meal. Performances of the rest of the works are close. Majority used DL methods e.g., LSTM, RNN, CNN, etc. Few of them used ARX, NN, and SVR.
3. *Influence of Size of Dataset*: Data-driven models require a lot of data. Usually, the more the data, the better the prediction performance gets. In the work of Ozogur et al. (2020), they have used two datasets: one has 8 weeks data of 6 T1DM patients, and another has 26 h data of 89 T1DM patients. Their model driven by Fuzzy particle swarm optimization algorithm performs better on the comparatively larger cohort of 8 weeks data of 6 T1DM patients and received an RMSE of 27.46 mg/dL. For the smaller dataset, with a lot more patients and less per patient data, this dataset suffers from inter patient variability. Whereas, DAntoni et al. (2020) have received similar performances when they used Auto Regressive Time Delayed (ARTiDe) Jump Neural Network (Jump NN) on a cohort of 33 patients for on average 15 days and on a different smaller cohort of 6 patients for 8 weeks. Mirshekarian et al. (2019) have achieved a performance improvement when they have used memory-augmented MemLSTM neural attention model over their previous LSTM model (Mirshekarian et al., 2017). This gives an observation that attention aware models generally improves BGL prediction performance.
4. *Simulated vs Real Data*: Collecting data that influences BG is difficult. In addition to that, the lack of publicly available data is a problem. Hence, few use simulated data. In the reviewed articles, 72.4% of the time real data is used, 5.2% of the time simulated data is used, and 22.4% of the time both types of data are used. Those simulated data are generated from the two simulators: UVA/Padova T1D and AIDA. It is observed that the performance of the modeling techniques when applied on simulated data [Table 8] is considerably better than when applied on real data [Table 7] across the board.
5. *Performance Variation by Prediction Horizon*: This study has recorded prediction performances for only three Prediction Horizons (PH): 30 min, 60 min, and 120 min. Among them PH 30 min is the most widely used. Whereas 120 min is the least popular. From Table 7, it is observed that in general, increasing the prediction horizon decreases performance. That means the prediction performance of BGL in two hours from current time is usually worse than in 30 min. This is due to the complexity of glucose–insulin dynamics. It involves quite a few factors. With a longer prediction horizon, it gets more challenging to predict the dynamics among these factors. Hence, prediction performance suffers. BGL two hours after meal is a standard metric used to check whether BGL is within the normal range or not. More research should include BGL prediction experiments for 120-min PH.

6. *Hyper-parameter Optimization by Optimization Algorithms*: It is observed that lately researchers (Hamdi et al., 2018; Ozogur et al., 2020) are using optimization algorithms such as GP, GE, DE, PSO, etc. to optimize the data-driven models. As can be observed in He and Wang (2020), Ozogur et al. (2020), Wang et al. (2020), these optimization algorithms have improved the prediction performances of the models.
7. *Computational Efficiency*: Some studies (Martinsson et al., 2018, 2020; Nemat et al., 2022) have investigated updating their training models with new data. In these cases, the retraining time is considerably less than the new data collection time. Hence, their system is feasible for even online re-training. The models proposed by Zhu et al. (2018) have high efficiency with small weight, less global training steps, and low time complexity of $O(n)$, where n is the number of layers. Li et al. (2020) have improved the time complexity of their proposed ESN model to $O(mn)$ from the complexity of the standard ESN $O(mn^2)$, where m is the number of training samples and n is the number of hidden nodes. The non-linear autoregressive (NAR) neural network model proposed by Aliberti et al. (2019) is a simple and computationally efficient model based on just 8 regressors. It provides satisfactory results for 30-min PH. The novel ANN model proposed by Amar et al. (2020) is fast and simple to deploy, converges fast, less prone to overfitting, and can be used on mobile devices. The model used in Li, Daniels et al. (2019) is light, computationally fast, and can be continuously updated as more data is available on the cloud. The models in these studies (Li, Daniels et al., 2019; Li, Liu et al., 2019; Martinsson et al., 2018, 2020; Yu, Rashid et al., 2018; Zecchin et al., 2013) are small enough to fit and used in mobile devices. The prediction models proposed by Wang et al. (2020) have high training time due to the use of PSO techniques. They intend to focus their future effort on the reduction of that training time.
8. *Generalized vs Personalized Models*: In the studied literature, the majority of the studies (Aliberti et al., 2019; Bock et al., 2015; Karim et al., 2020; Martinsson et al., 2020; Sun, Jankovic, Bally et al., 2018; Yang et al., 2018) have developed generalized models, quite a few studies (Faruqui et al., 2019; Kushner et al., 2020; Li, Liu et al., 2019; Zhu et al., 2018) have used personalized models, and some studies (Ozogur et al., 2020) have focused on both of the above aspects. To enhance the generalizability of their models the authors in Yang et al. (2018) have used an ARIMA model with adaptive orders. Martinsson et al. (2020) have observed that their models generalize well when they are trained on data from all patients. Zecchin et al. (2013) have presented that their jump NN model is able to generalize well when tested on a different dataset. They have claimed that the generalizability would improve if their model is trained on a larger dataset. Due to scarcity of per patient data they were not able to develop personalized models. The studies in Aliberti et al. (2019), Mosquera-Lopez et al. (2019) have used a publicly available large dataset to develop a generalizable glucose prediction model. However, Mosquera-Lopez et al. (2019) have further used personalized error correction and have improved the prediction performance by over 50%. Karim et al. (2020) have admitted that a prediction model trained with only 5 patients' data cannot generalize well.
   Few studies (Deng et al., 2021; Faruqui et al., 2019; Kushner et al., 2020; Zhu et al., 2020, 2018) have used transfer learning due to data scarcity to develop personalized prediction models. In the works of Zhu et al. (2020, 2018), 50% of the training data for each subject has been taken from other subjects. This expansion of training data has improved the generalization and prediction performance of their personalized models. In Ozogur et al. (2020), the authors have shown that personalized models perform better than generalized models. Armandpour et al. (2021) have developed personalized models without the need for a large dataset for individual patients using embedding vectors and shared parameters across the patients.
9. *Interpretation of the Prediction Models*: Majority of the studies have used machine learning and deep learning based prediction models. Reviewed studies have not used the interpretable machine learning techniques such as linear regression, decision tree, etc. As these techniques are not enough to develop prediction models with high accuracy. One study (Zarkogianni et al., 2015) has used a dimensionality reduction algorithm SOM which is interpretable. However, most of the models used in the literature are like black boxes. They are generally not interpretable. Interpretability methods can be utilized to interpret them. Nonetheless, they are not used by the current literature to provide interpretability of their prediction models.

6.2. Limitations

Diabetes research community is vibrant specially in the BGL prediction field. Still, there remain challenges. Due to lack of real data, some research works (Zhao & Yu, 2015) have used simulated data. However, simulated data cannot be readily used in clinical settings. Most if not all of the datasets used by the research community are in limited size. Due to the size of the datasets, validation data lacks variation. The proposed prediction models only perform well for their specific dataset. Due to inter patient variations of age, gender, history, glucose–insulin dynamics, etc there are substantial variations in the prediction performances. Hence, generalizability is not possible.

In addition to the small size of the datasets (Alfian et al., 2020), many datasets have missing data. Though there are effective preprocessing techniques to handle missing data their effectiveness have a limit. Due to missing data, prediction performance suffers. For CGM time-series data, for the sake of presenting good performing prediction models, many research works have assumed a stationary environment. However, this is not true over a long period of time. Hence, models proposed for a stationary environment do not work well over a long period of time.

Many research works (Alfian et al., 2020; Faruqui et al., 2019; Marcus et al., 2020; Nemat et al., 2022) have used secondary data from previous trials. Many of those trials are not designed to collect data for BG prediction purposes. Hence, many BG related data such as BG reading time, medication, etc are not well documented and available to use. Lately, data-driven models are extensively used by the research community. Domain knowledge can help the data-driven models to improve performances. Nevertheless, it is not an easy task to incorporate domain knowledge into data-driven methods.

In addition to the above mentioned limitations there are other limitations including limited forecasting window by many research works, limited applicability of the proposed linear models for the assumption of linearity of a nonlinear system, decreasing prediction accuracy with higher PH, potential incorrectness in data due to manual recording of data, etc.

6.3. Future works

A large variety of datasets, input feature sets, and modeling techniques have been used in BGL prediction. Researchers are also using different combinations of modeling techniques to improve prediction performance. We have identified a number of avenues of future works.

6.3.1. Related to dataset

Need More Research using Clinical Data: The results that are obtained from experiments using simulated data cannot be directly used in clinical settings. This is due to lack of inter patient variability in simulated data. Though we still need to use simulated data due to lack of real data more research works need to carry experiments using clinical data.

Prediction Models for T2DM Patients with Discrete Data: It is observed that majority of the prediction models proposed in the reviewed articles have used CGM time-series data. However, these models cannot be readily used for T2DM patients. More future research works need to be directed to use discrete T2DM patients' data.

Publicly Available Large Dataset: Due to the lack of publicly available large datasets with a lot of BG influencing input features many research works use small datasets or use datasets with only glucose readings from CGM devices. These models will fail to predict in many different new situations. Research community should collaborate to build large publicly available datasets with diverse feature sets from a large number of diverse patients using different CGM devices and glucometers so that models can be tested with randomly new and different types of large volumes of data for extensive validation. With the help of a large dataset developing a generalized model that can be used for a new patient with little data can be a potential future work.

6.3.2. Related to input feature set

More Features as Input: Majority of the research works have used CGM readings as the only input. Many of the works have stated that only historical glucose data is enough for BGL prediction. On the other hand, many factors including meal intake, insulin dosage, and physical activity influence BG. There are also some input features such as electrodermal activity, temperature acceleration, etc. that are collected by smart bands, some physical features such as age, gender, liver function, etc., some diabetes dynamics features such as plasma insulin estimation, glucose rate of appearance, etc. can be considered. Future works should put more effort to include more input features. The list of input features may vary depending on the specific target user group.

Appropriate Impact of Input Features: The different input features influence BGL differently, for example, meal increases BGL, insulin decreases BGL, physical exercise decreases BGL, etc. However, due to the simultaneous influences of multiple factors on BG, we do not get the expected impact. To see the expected impact, we need to separate the influence of different input features. This will be an interesting area of future research.

Intensity of Physical Activity: Now-a-days lots of data is generated from wearable devices. Therefore, more efforts are needed to be directed to identify the type and intensity of activities from wearable data.

Instantaneous Prediction of BGL from Meal Images: Self diabetes management is also becoming the new trend as it empowers a person to manage his/her diabetes. In this direction, instantaneous prediction of BGL from meal images is required, which can be a useful research direction.

Adaptive Feature Set: In all the currently proposed models, it is assumed that the feature set will remain the same throughout the use of the model by any patient. However, it will be more suitable if the feature set adapts and changes according to the new data being collected in real-time.

6.3.3. Related to preprocessing methods

Handle Data Scarcity with Data Augmentation: In BGL prediction research, we have scarcity of real data. However, we need a lot of real data for good performing prediction models. In this situation, data augmentation can be used as is used in the following works (Deng et al., 2021; Hidalgo et al., 2020; Velasco et al., 2017) to handle scarcity of data. More research works need to use data augmentation to improve prediction performance, and to find out how much data augmentation is useful.

CGM Device Errors Need to be Considered: Majority of the researchers use time series data from CGM devices. CGM time series data has noise, shifts, errors, etc. that can affect the accuracy of the prediction models. These errors are not considered by many researchers. However, to develop robust prediction models these errors need to be handled in the model development process. Whether the use of denoising, smoothing, or any other filters can handle these issues without increasing prediction errors in the critical hypo and hyperglycemic BGLs can be an interesting future work. Reduction of prediction errors in the hypo and hyperglycemic BGLs due to imbalanced data is by itself an area of future research.

6.3.4. Related to prediction methods

Suitable Modeling Techniques: Due to different individual lifestyles, inter patient variability, and different quality of the used datasets, one modeling technique is not expected to work well in all situations. Yet, it is possible to recommend different modeling techniques for different glucose regions, or in different situations. The model can be personalized for individual patients with patient specific number of epochs, number of hidden layers, number of neurons per layer, etc.

Hybridization of Different Modeling Techniques: There is no one or two modeling techniques that can be used in all the scenarios. Recently, a few researchers are combining or ensembling different modeling techniques in the search for better prediction performance. They are getting promising results. Due to the presence of inter person variability, combining physiological models with other data-driven models can be a potential option. Use of self-attention mechanisms, appropriate incorporation of domain knowledge to develop physiologically informed data-driven models, etc. can be other potential options. Therefore, more research effort needs to be focused in this direction.

Models for Non-stationary Environment: A model that is designed for a stationary environment will not work well for a long period of time. Hence, more works need to be dedicated to non-stationary circumstances.

Online Prediction Model Performances: Majority of the prediction models are tested against collected data using cross-validation or new unseen data. In diabetes research, these prediction models are supposed to run on devices with limited capability. How these models will perform on a limited capability device is unknown. More investigation needs to be carried out on the devices where these prediction models will be used. A potential solution will include multi-objective optimization approach. In this approach, one objective may be to develop computationally efficient model, and a second objective may be to develop a model with high prediction performance. A potential solution to develop computationally efficient model may include transfer learning. For some algorithms including Recursive Least-Squares (RLS) used in Yu, Rashid et al. (2018), Yu, Turksoy et al. (2018), computational complexity increases with dimensionality. How to address to select an appropriate training sample size can be a potential topic for future research.

6.3.5. Related to quality of BGL prediction

Estimation of the Probabilistic Distribution of BGL Predictions: Almost all of the reviewed articles have proposed only BGL prediction models. However, only the following works (Li, Liu et al., 2019; Martinsson et al., 2018; Mirshekarian et al., 2019; Zhu et al., 2022) have provided the probabilistic distribution of the predicted BGL values. This distribution provides the uncertainty in the prediction result. It tells us whether the predicted values can be trusted or not. More research works should provide the confidence level of its predictions to understand the usability of a model in clinical settings.

6.3.6. Others

When previous glucose history is used as the only input feature, there is scope for future research to find which modeling technique and how much previous history is enough to find the best performing prediction model. Other future works may include forecasting for longer PHs, including future events such as physical activity, fast acting CHO consumption to prevent onset of hypoglycemia, etc, including circadian variations in glucose absorption and insulin sensitivity as input features, balancing RMSE and CEG values for prediction performance, using more efficient optimization algorithms, using different data fusion techniques, fine-tuning hyper-parameters both for ensemble and non-ensemble methods, examining different models as base and meta-learners, etc.

7. Conclusion

In this paper, we have conducted a comprehensive literature review on the latest efforts on diabetes management as a part of clinical applications. We have analyzed the existing recent works from different perspectives that include clinical applications, input features, modeling techniques, and the performances. At first, we have presented the detailed study, and designed the taxonomy from the perspective of different clinical applications. Among the clinical applications, BGL prediction is found to be the most popular among the researchers. Then, we have discussed the details of different input features such as meal intake, historical glucose level, etc. used in our focused BGL prediction application. We have observed a lack of research on T2DM though over 90% of diabetic patients have T2DM. Due to high data point rate from CGM devices, majority of the research works used T1DM data. There are few publicly available T1DM datasets including OhioT1DM. Moreover to the best of our knowledge, there is no publicly available T2DM dataset. T2DM data collection is challenging as it takes a much longer data collection period. There is a dire need of publicly available benchmark T2DM dataset. More research efforts are needed to be directed to collect data from T2DM patients. We have discussed and compared the performances of different types of modeling techniques from different perspectives. We have identified a number of limitations in the existing prediction models. Finally, we have presented a few avenues of potential future research. It is expected that this study will help the researchers to keep working in search of an effective for all generalized BGL prediction model to help the diabetic patients to manage their diabetes well.

Funding information

This literature review is part of a research project funded by the Institute for Advanced Research (IAR) at United International University, Grant/Award Number: UIU/IAR/01/2021/SE/11.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Adam, M., Hagiwara, Y., Tan, J. H., Ng, E., Acharya, U. R., Oh, S. L., et al. (2018). Automated characterization of diabetic foot using nonlinear features extracted from thermograms. *Infrared Physics & Technology*, *89*, http://dx.doi.org/10.1016/j.infrared.2018.01.022.

Aiello, E., Lisanti, G., Magni, L., Musci, M., & Toffanin, C. (2020). Engineering applications of artificial intelligence therapy-driven deep glucose forecasting. *Engineering Applications of Artificial Intelligence*, *87*, Article 103255. http://dx.doi.org/10.1016/j.engappai.2019.103255.

Albers, D. J., Levine, M., Gluckman, B., Ginsberg, H., Hripcsak, G., & Mamykina, L. (2017). Personalized glucose forecasting for type 2 diabetes using data assimilation. *PLoS Computational Biology*, *13*(4), Article e1005232.

Alfian, G., Syafrudin, M., Anshari, M., Benes, F., Dwi Atmaji, F., Fahrurrozi, I., et al. (2020). Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features. *Biocybernetics and Biomedical Engineering*, *40*, 1586–1599. http://dx.doi.org/10.1016/j.bbe.2020.10.004.

Aliberti, A., Pupillo, I., Terna, S., Macii, E., Di Cataldo, S., Patti, E., et al. (2019). A multi-patient data-driven approach to blood glucose prediction. *IEEE Access*, *7*, 69311–69325.

Amar, Y., Shilo, S., Oron, T., Amar, E., Phillip, M., & Segal, E. (2020). Clinically accurate prediction of glucose levels in patients with type 1 diabetes. *Diabetes Technology & Therapeutics*, *22*(8), 562–569.

Arcadu, F., Benmansour, F., Maunz, A., Willis, J., Haskova, Z., & Prunotto, M. (2019). Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *npj Digital Medicine*, *2*, http://dx.doi.org/10.1038/s41746-019-0172-3.

Arleth, T., Andreassen, S., Orsini-Federici, M., Timi, A., & Massi-Benedetti, M. (2000). A model of glucose absorption from mixed meals. In *Proc. 4th IFAC modelling and control in biomedical systems symposium*: In *Proc. 4th IFAC modelling and control in biomedical systems symposium*: *vol. 33*, 307–312.http://dx.doi.org/10.1016/S1474-6670(17)35533-7,

Armandpour, M., Kidd, B., Du, Y., & Huang, J. Z. (2021). Deep personalized glucose level forecasting using attention-based recurrent neural networks. In *2021 international joint conference on neural networks* (pp. 1–8). IEEE.

Association, A. (2001). Postprandial blood glucose. *Clinical Diabetes*, *19*, 127–130. http://dx.doi.org/10.2337/diaclin.19.3.127.

Association, A. (2007). Standards of medical care for diabetes. *Diabetes Care*, *17*, 1514–1522. http://dx.doi.org/10.2337/diacare.27.2007.s15.

Association, A. D. (2014). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, *37*(Supplement 1), S81–S90. http://dx.doi.org/10.2337/dc14-S081.

Association, A. D. (2019). Classification and diagnosis of diabetes: Standards of medical care in diabetes—2019. *Diabetes Care*, *42*(Supplement 1), S13–S28. http://dx.doi.org/10.2337/dc19-S002.

Bailey, C., Aschner, P., Del Prato, S., LaSalle, J., Ji, L., & Matthaei, S. (2013). Individualized glycaemic targets and pharmacotherapy in type 2 diabetes. *Diabetes and Vascular Disease Research*, *10*, 397–409. http://dx.doi.org/10.1177/1479164113490765.

Balakrishnan, N., Rangaiah, G., & Samavedham, L. (2011). Review and analysis of blood glucose (BG) models for type 1 diabetic patients. *Industrial and Engineering Chemistry Research*, *50*, http://dx.doi.org/10.1021/ie2004779.

Balducci, S., Sacchetti, M., Haxhi, J., Orlando, G., d'Errico, V., Fallucca, S., et al. (2014). Physical exercise as therapy for type 2 diabetes mellitus. *Diabetes/metabolism Research and Reviews*, *30*, http://dx.doi.org/10.1002/dmrr.2514.

Bergman, R., Ider, Y., Bowden, C., & Cobelli, C. (1979). Quantitative estimation of insulin sensitivity. *The American Journal of Physiology*, *236*, E667–77. http://dx.doi.org/10.1152/ajpendo.1979.236.6.E667.

Bertachi, A., Biagi, L., Contreras, I., Luo, N., & Vehi, J. (2018). Prediction of blood glucose levels and nocturnal hypoglycemia using physiological models and artificial neural networks. In *KHD@iJCAI* (pp. 85–90).

Bock, A., Francois, G., & Gillet, D. (2015). A therapy parameter-based model for predicting blood glucose concentrations in patients with type 1 diabetes. *Computer Methods and Programs in Biomedicine*, *118*, 107–123. http://dx.doi.org/10.1016/j.cmpb.2014.12.002.

Bosoni, P., Meccariello, M., Calcaterra, V., Larizza, C., Sacchi, L., & Bellazzi, R. (2020). Deep learning applied to blood glucose prediction from flash glucose monitoring and fitbit data. In *Artificial intelligence in medicine: 18th international conference on artificial intelligence in medicine, AIME 2020, minneapolis, MN, USA, August 25–28* (pp. 59–63). http://dx.doi.org/10.1007/978-3-030-59137-3_6.

Breton, M., Farret, A., Bruttomesso, D., Anderson, S., Magni, L., Patek, S., et al. (2012). Fully integrated artificial pancreas in type 1 diabetes: Modular closed-loop glucose control maintains near normoglycemia. *Diabetes*, *61*, 2230–2237. http://dx.doi.org/10.2337/db11-1445.

Bunescu, R., Struble, N., Marling, C., Shubrook, J., & Schwartz, F. (2013). Blood glucose level prediction using physiological models and support vector regression. In *2013 12th international conference on machine learning and applications*: *vol. 1*, (pp. 135–140). http://dx.doi.org/10.1109/ICMLA.2013.30.

Cappon, G., Facchinetti, A., Sparacino, G., Georgiou, P., & Herrero, P. (2019). Classification of postprandial glycemic status with application to insulin dosing in type 1 diabetes- An in silico proof-of-concept. *Sensors*, *19*, 3168. http://dx.doi.org/10.3390/s19143168.

Cappon, G., Vettoretti, M., Marturano, F., Facchinetti, A., & Sparacino, G. (2018). A neural-network-based approach to personalize insulin bolus calculation using continuous glucose monitoring. *Journal of Diabetes Science and Technology*, *12*, 265–272. http://dx.doi.org/10.1177/1932296818759558.

Cichosz, S. L., Kronborg, T., Jensen, M. H., & Hejlesen, O. (2021). Penalty weighted glucose prediction models could lead to better clinically usage. *Computers in Biology and Medicine*, *138*, Article 104865.

Cobelli, C., Federspil, G., Pacini, G., Salvan, A., & Scandellari, C. (1982). An integrated mathematical model of the dynamics of blood glucose and its hormonal control. *Mathematical Biosciences*, *58*(1), 27–60. http://dx.doi.org/10.1016/0025-5564(82)90050-5.

Contreras, I., Oviedo, S., Vettoretti, M., Visentin, R., & Vehi, J. (2017). Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models. *PLoS One*, *12*(11), Article e0187754. http://dx.doi.org/10.1371/journal.pone.0187754.

Contreras, I., & Vehi, J. (2018). Artificial intelligence for diabetes management and decision support: Literature review. *Journal of Medical Internet Research*, *20*, http://dx.doi.org/10.2196/10775.

Dagliati, A., Sacchi, L., Tibollo, V., Cogni, G., Teliti, M., Martinez-Millana, A., et al. (2018). A dashboard-based system for supporting diabetes care. *Journal of the American Medical Informatics Association: JAMIA*, *25*, http://dx.doi.org/10.1093/jamia/ocx159.

Dalla Man, C., Camilleri, M., & Cobelli, C. (2007). A system model of oral glucose absorption: Validation on gold standard data. *IEEE Transactions on Bio-Medical Engineering*, *53*, 2472–2478. http://dx.doi.org/10.1109/TBME.2006.883792.

Daniels, J., Herrero, P., & Georgiou, P. (2021). A multitask learning approach to personalized blood glucose prediction. *IEEE Journal of Biomedical and Health Informatics*, *26*(1), 436–445.

Dankwa-Mullan, I., Rivo, M., Sepulveda, M., Park, Y., Snowdon, J., & Rhee, K. (2018). Transforming diabetes care through artificial intelligence: The future is here. *Population Health Management*, *22*, http://dx.doi.org/10.1089/pop.2018.0129.

DAntoni, F., Merone, M., Piemonte, V., Iannello, G., & Soda, P. (2020). Auto-regressive time delayed jump neural network for blood glucose levels forecasting. *Knowledge-Based Systems*, *203*, Article 106134. http://dx.doi.org/10.1016/j.knosys.2020.106134.

Daskalaki, E., Diem, P., & Mougiakakou, S. (2016). Model-free machine learning in biomedicine: Feasibility study in type 1 diabetes. *PLoS One*, *11*, Article e0158722. http://dx.doi.org/10.1371/journal.pone.0158722.

Dassau, E., Zisser, H., Palerm, C., Buckingham, B., Jovanovic, L., & Doyle, F. (2008). Modular artificial beta-cell system: A prototype for clinical research. *Journal of Diabetes Science and Technology*, *2*, 863–872. http://dx.doi.org/10.1177/193229680800200518.

Del Favero, S., Place, J., Kropff, J., Messori, M., Keith-Hynes, P., Visentin, R., et al. (2015). Multicenter outpatient dinner/overnight reduction of hypoglycemia and increased time of glucose in target with a wearable artificial pancreas using modular model predictive control in adults with type 1 diabetes. *Diabetes, Obesity & Metabolism*, *17*, http://dx.doi.org/10.1111/dom.12440.

Deng, Y., Lu, L., Aponte, L., Angelidi, A. M., Novak, V., Karniadakis, G. E., et al. (2021). Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. *NPJ Digital Medicine*, *4*(1), 109.

Dubosson, F., Ranvier, J.-E., Bromuri, S., Calbimonte, J.-P., Ruiz, J., & Schumacher, M. (2018). The open D1NAMO dataset: A multi-modal dataset for research on non-invasive type 1 diabetes management. *Informatics in Medicine Unlocked*, *13*, 92–100.

DuBrava, S., Mardekian, J., Sadosky, A., Bienen, E., Parsons, B., Hopps, M., et al. (2016). Using random forest models to identify correlates of a diabetic peripheral neuropathy diagnosis from electronic health record data. *Pain Medicine*, *18*, 107–115. http://dx.doi.org/10.1093/pm/pnw096.

Dudukcu, H. V., Taskiran, M., & Yildirim, T. (2021). Blood glucose prediction with deep neural networks using weighted decision level fusion. *Biocybernetics and Biomedical Engineering*, *41*(3), 1208–1223.

Eghbali-Zarch, M., Tavakkoli-Moghaddam, R., Esfahanian, F., Sepehri, M., & Azaron, A. (2018). Pharmacological therapy selection of type 2 diabetes based on the swara and modified multimoora methods under a fuzzy environment. *Artificial Intelligence in Medicine*, *87*, http://dx.doi.org/10.1016/j.artmed.2018.03.003.

Faruqui, S. H. A., Meka, R., Alaeddini, A., Li, C., Shirinkam, S., & Wang, J. (2019). Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: Secondary analysis of a randomized controlled trial. *JMIR Mhealth and Uhealth*, *7*, http://dx.doi.org/10.2196/14452.

Federation, I. D. (2021). *IDF Diabetes Atlas* (10th edn.). International Diabetes Federation, URL https://www.diabetesatlas.org.

Felizardo, V., Garcia, N. M., Pombo, N., & Megdiche, I. (2021). Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction–A systematic literature review. *Artificial Intelligence in Medicine*, *118*, Article 102120.

Fiarni, C., Sipayung, E. M., & Maemunah, S. (2019). Analysis and prediction of diabetes complication disease using data mining algorithm. *Procedia Computer Science*, *161*, 449–457. http://dx.doi.org/10.1016/j.procs.2019.11.144.

Fox, I., Ang, L., Jaiswal, M., Pop-Busui, R., & Wiens, J. (2018). Deep multi-output forecasting: Learning to accurately predict blood glucose trajectories. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1387–1395). http://dx.doi.org/10.1145/3219819.3220102.

Fraiwan, L., Alkhodari, M., Ninan, J., Mustafa, B., Saleh, A., & Ghazal, M. (2017). Diabetic foot ulcer mobile detection system using smart phone thermal camera: A feasibility study. *Biomedical Engineering Online*, *16*, 117. http://dx.doi.org/10.1186/s12938-017-0408-x.

Frandes, M., Timar, B., Timar, R., & Lungeanu, D. (2017). Chaotic time series prediction for glucose dynamics in type 1 diabetes mellitus using regime-switching models. *Scientific Reports*, *7*(1), 1–10.

Georga, E., Protopappas, V., Polyzos, D., & Fotiadis, D. (2015). Online prediction of glucose concentration in type 1 diabetes using extreme learning machines. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society* (pp. 3262–3265). http://dx.doi.org/10.1109/EMBC.2015.7319088.

Goutham, S., Ravi, V., & Kp, S. (2018). Diabetes detection using deep learning algorithms. *ICT Express*, *4*, http://dx.doi.org/10.1016/j.icte.2018.10.005.

Goyal, M., Reeves, N., Davison, A., Rajbhandari, S., Spragg, J., & Yap, M. H. (2017). DFUNet: Convolutional neural networks for diabetic foot ulcer classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *4*, http://dx.doi.org/10.1109/TETCI.2018.2866254.

Guyton, J., Foster, R., Soeldner, J., Tan, M., Kahn, C., Koncz, L., et al. (1978). A model of glucose-insulin homeostasis in man that incorporates the heterogeneous fast pool theory of pancreatic insulin release. *Diabetes*, *27*, 1027–1042. http://dx.doi.org/10.2337/diabetes.27.10.1027.

Gyuk, P., Vassanyi, I., & Kosa, I. (2019). Blood glucose level prediction for diabetics based on nutrition and insulin administration logs using personalized mathematical models. *Journal of Healthcare Engineering*, *2019*, 1–12. http://dx.doi.org/10.1155/2019/8605206.

Hamdi, T., Ben Ali, J., Costanzo, V., Fnaiech, F., Moreau, E., & Ginoux, J.-M. (2018). Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. *Biocybernetics and Biomedical Engineering*, *38*, http://dx.doi.org/10.1016/j.bbe.2018.02.005.

He, J., & Wang, Y. (2020). Blood glucose concentration prediction based on kernel canonical correlation analysis with particle swarm optimization and error compensation. *Computer Methods and Programs in Biomedicine*, *196*, Article 105574.

Herrero, P., Calm, R., Vehi, J., Armengol, J., Georgiou, P., Oliver, N., et al. (2012). Robust fault detection system for insulin pump therapy using continuous glucose monitoring. *Journal of Diabetes Science and Technology*, *6*, 1131–1141. http://dx.doi.org/10.1177/193229681200600518.

Hidalgo, J. I., Botella, M., Velasco, J. M., Garnica, O., Cervigon, C., Martınez, R., et al. (2020). Glucose forecasting combining markov chain based enrichment of data, random grammatical evolution and bagging. *Applied Soft Computing*, *88*, Article 105923.

Hidalgo, J. I., Colmenar, J. M., Kronberger, G., Winkler, S. M., Garnica, O., & Lanchares, J. (2017). Data based prediction of blood glucose concentrations using evolutionary methods. *Journal of Medical Systems*, *41*(9), 1–20.

Hidalgo, I., Colmenar, J., Risco-Martin, J., Cuesta-Infante, A., Maqueda, E., Botella, M., et al. (2014). Modeling glycemia in humans by means of grammatical evolution. *Applied Soft Computing*, *20*, 40–53. http://dx.doi.org/10.1016/j.asoc.2013.11.006.

Huang, G.-M., Huang, K.-Y., Lee, T.-Y., & Weng, J. (2015). An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients. *BMC Bioinformatics*, *16*, 1–10. http://dx.doi.org/10.1186/1471-2105-16-S1-S5.

Ito, K., & Jin, B. (2014). *Inverse problems: tikhonov theory and algorithms: vol. 22*, World Scientific.

Jonnagaddala, J., Liaw, S.-T., Ray, P., Kumar, M., Dai, H., & Hsu, C.-Y. (2015). Identification and progression of heart disease risk factors in diabetic patients from longitudinal electronic health records. *BioMed Research International*, *2015*, 1–10. http://dx.doi.org/10.1155/2015/636371.

Kahn, M. (1994). *UCI machine learning repository: Diabetes*.

Karim, R., Vassanyi, I., & Kosa, I. (2019). Long-acting insulin management for blood glucose prediction models.. *Biomedical Research*, *30*, http://dx.doi.org/10.35841/biomedicalresearch.30-18-1190.

Karim, R., Vassanyi, I., & Kosa, I. (2020). After-meal blood glucose level prediction using an absorption model for neural network training. *Computers in Biology and Medicine*, *125*, Article 103956. http://dx.doi.org/10.1016/j.compbiomed.2020.103956.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, *15*, http://dx.doi.org/10.1016/j.csbj.2016.12.005.

Kircher, R., Mauseth, R., Bhatia, S., & Matheson, D. (2008). Fuzzy logic controller for insulin dosing. In *Diabetes technology meeting*.

Kushner, T., Breton, M., & Sankaranarayanan, S. (2020). Multi-hour blood glucose prediction in T1D: A patient-specific approach using shallow neural network models. *Diabetes Technology & Therapeutics*, *22*, http://dx.doi.org/10.1089/dia.2020.0061.

Lekha, S., & Suchetha, M. (2017). Real-time non-invasive detection and classification of diabetes using modified convolution neural network. *IEEE Journal of Biomedical and Health Informatics*, *22*(5), 1630–1636.

Lerner, R., & Porte, D. (1971). Relationships between intravenous glucose loads, insulin responses and glucose disappearance rate. *Journal of Clinical Endocrinology & Metabolism*, *33*, 409–417. http://dx.doi.org/10.1210/jcem-33-3-409.

Li, K., Daniels, J., Liu, C., Herrero, P., & Georgiou, P. (2019). Convolutional recurrent neural networks for glucose prediction. *IEEE Journal of Biomedical and Health Informatics*, *24*, 603–613. http://dx.doi.org/10.1109/JBHI.2019.2908488.

Li, J., & Fernando, C. (2016). Smartphone-based personalized blood glucose prediction. *ICT Express*, *2*, http://dx.doi.org/10.1016/j.icte.2016.10.001.

Li, K., Liu, C., Zhu, T., Herrero, P., & Georgiou, P. (2019). GluNet: A deep learning framework for accurate glucose forecasting. *IEEE Journal of Biomedical and Health Informatics*, *24*(2), 414–423. http://dx.doi.org/10.1109/JBHI.2019.2931842.

Li, N., Tuo, J., Wang, Y., & Wang, M. (2020). Prediction of blood glucose concentration for type 1 diabetes based on echo state networks embedded with incremental learning. *Neurocomputing*, *378*, 248–259.

Liu, C., Vehi, J., Avari, P., Reddy, M., Oliver, N., Georgiou, P., et al. (2019). Long-term glucose forecasting using a physiological model and deconvolution of the continuous glucose monitoring signal. *Sensors*, *19*, 4338. http://dx.doi.org/10.3390/s19194338.

Marcus, Y., Eldor, R., Yaron, M., Shaklai, S., Ish-Shalom, M., Shefer, G., et al. (2020). Improving blood glucose level predictability using machine learning. *Diabetes/Metabolism Research and Reviews*, *36*, http://dx.doi.org/10.1002/dmrr.3348.

Marling, C., & Bunescu, R. (2020). The OhioT1DM dataset for blood glucose level prediction: Update 2020. In *CEUR workshop proceedings*: *vol. 2675*, (p. 71). NIH Public Access.

Martinsson, J., Alexander, S., Bjorn, E., Christian, M., Simon, P., & Olof, M. (2018). Automatic blood glucose prediction with confidence using recurrent neural networks. In *3rd international workshop on knowledge discovery in healthcare data, kDH@ IJCAI-ECAI 2018, 13 July 2018* (pp. 64–68).

Martinsson, J., Schliep, A., Eliasson, B., & Mogren, O. (2020). Blood glucose prediction with variance estimation using recurrent neural networks. *Journal of Healthcare Informatics Research*, *4*, http://dx.doi.org/10.1007/s41666-019-00059-y.

Mauseth, R., Lord, S., Hirsch, I., Kircher, R., Matheson, D., & Greenbaum, C. (2015). Stress testing of an artificial pancreas system with pizza and exercise leads to improvements in the system's fuzzy logic controller. *Journal of Diabetes Science and Technology*, *9*, http://dx.doi.org/10.1177/1932296815602098.

Mhaskar, H., Pereverzyev, S., & Van der Walt, M. (2017). A deep learning approach to diabetic blood glucose prediction. *Frontiers in Applied Mathematics and Statistics*, *3*, http://dx.doi.org/10.3389/fams.2017.00014.

Miotto, R., Li, L., & Kidd, B. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, *6*, 26094. http://dx.doi.org/10.1038/srep26094.

Mirshekarian, S., Bunescu, R., Marling, C., & Schwartz, F. (2017). Using LSTMs to learn physiological models of blood glucose behavior. In *2017 39th annual international conference of the IEEE engineering in medicine and biology society* (pp. 2887–2891). http://dx.doi.org/10.1109/EMBC.2017.8037460.

Mirshekarian, S., Shen, H., Bunescu, R., & Marling, C. (2019). LSTMs and neural attention models for blood glucose prediction: Comparative experiments on real and synthetic data. In *2019 41st annual international conference of the IEEE engineering in medicine and biology society* (pp. 706–712). http://dx.doi.org/10.1109/EMBC.2019.8856940.

Montaser, E., Diez, J.-L., Rossetti, P., Rashid, M., Cinar, A., & Bondia, J. (2019). Seasonal local models for glucose prediction in type 1 diabetes. *IEEE Journal of Biomedical and Health Informatics*, *24*(7), 2064–2072.

Mosquera-Lopez, C., Dodier, R., Tyler, N., Resalat, N., & Jacobs, P. (2019). Leveraging a big dataset to develop a recurrent neural network to predict adverse glycemic events in type 1 diabetes. *IEEE Journal of Biomedical and Health Informatics*.

Neinstein, A., Wong, J., Look, H., Arbiter, B., Quirk, K., McCanne, S., et al. (2016). A case study in open source innovation: developing the tidepool platform for interoperability in type 1 diabetes management. *Journal of the American Medical Informatics Association*, *23*(2), 324–332.

Nemat, H., Khadem, H., Eissa, M. R., Elliott, J., & Benaissa, M. (2022). Blood glucose level prediction: advanced deep-ensemble learning approach. *IEEE Journal of Biomedical and Health Informatics*, *26*(6), 2758–2769.

Oviedo, S., Contreras, I., Quiros, C., Gimenez, M., Conget, I., & Vehi, J. (2019a). Risk-based postprandial hypoglycemia forecasting using supervised learning. *International Journal of Medical Informatics*, *126*, http://dx.doi.org/10.1016/j.ijmedinf.2019.03.008.

Oviedo, S., Contreras, I., Quiros, C., Gimenez, M., Conget, I., & Vehi, J. (2019b). Risk-based postprandial hypoglycemia forecasting using supervised learning. *International Journal of Medical Informatics*, *126*, http://dx.doi.org/10.1016/j.ijmedinf.2019.03.008.

Oviedo, S., Vehi, J., Calm, R., & Armengol, J. (2016). A review of personalized blood glucose prediction strategies for T1DM patients. *International Journal for Numerical Methods in Biomedical Engineering*, *33*, http://dx.doi.org/10.1002/cnm.2833.

Ozogur, H., Ozogur, G., & Orman, Z. (2020). Blood glucose level prediction for diabetes based on modified fuzzy time series and particle swarm optimization. *Computational Intelligence*, *37*, http://dx.doi.org/10.1111/coin.12396.

Palumbo, P. (2011). Glucose control by subcutaneous insulin administration: A DDE modelling approach. (pp. 1471–1476). ISBN: 9783902661937, http://dx.doi.org/10.3182/20110828-6-IT-1002.01374.

Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2017). Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of Biomedical Informatics*, *69*, http://dx.doi.org/10.1016/j.jbi.2017.04.001.

Rabby, M. F., Tu, Y., Hossen, M. I., Lee, I., Maida, A. S., & Hei, X. (2021). Stacked LSTM based deep recurrent neural network with kalman smoothing for blood glucose prediction. *BMC Medical Informatics and Decision Making*, *21*, 1–15.

Raumviboonsuk, P., Krause, J., Chotcomwongse, P., Sayres, R., Raman, R., Widner, K., et al. (2019). Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *npj Digital Medicine*, *2*, http://dx.doi.org/10.1038/s41746-019-0099-8.

Reddy, R., Resalat, N., Wilson, L., Castle, J., El Youssef, J., & Jacobs, P. (2019). Prediction of hypoglycemia during aerobic exercise in adults with type 1 diabetes. *Journal of Diabetes Science and Technology*, Article 193229681882379. http://dx.doi.org/10.1177/1932296818823792.

Rubin-Falcone, H., Fox, I., & Wiens, J. (2020). Deep residual time-series forecasting: Application to blood glucose prediction.. In *KDH@ ECAI* (pp. 105–109).

Saiti, K., Macas, M., Lhotska, L., Stechova, K., & Pithova, P. (2020). Ensemble methods in combination with compartment models for blood glucose level prediction in type 1 diabetes mellitus. *Computer Methods and Programs in Biomedicine*, *196*, Article 105628. http://dx.doi.org/10.1016/j.cmpb.2020.105628.

Seo, W., Lee, Y.-B., Lee, S., Jin, S.-M., & Park, S.-M. (2019). A machine-learning approach to predict postprandial hypoglycemia. *BMC Medical Informatics and Decision Making*, *19*, http://dx.doi.org/10.1186/s12911-019-0943-4.

Sorensen, J. T. (1985). *A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes*. Massachusetts Institute of Technology.

Spanig, S., Emberger-Klein, A., Sowa, J.-P., 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. *Artificial Intelligence in Medicine*, *100*, Article 101706. http://dx.doi.org/10.1016/j.artmed.2019.101706.

Sparacino, G., Zanderigo, F., Corazza, S., Maran, A., Facchinetti, A., & Cobelli, C. (2007). Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Transactions on Bio-Medical Engineering*, *54*, 931–937. http://dx.doi.org/10.1109/TBME.2006.889774.

Sun, Q., Jankovic, M., Bally, L., & Mougiakakou, S. (2018). Predicting blood glucose with an LSTM and bi-LSTM based deep neural network. In *2018 14th symposium on neural networks and applications* (pp. 1–5). http://dx.doi.org/10.1109/NEUREL.2018.8586990.

Sun, Q., Jankovic, M., Budzinski, J., Moore, B., Diem, P., Stettler, C., et al. (2018). A dual mode adaptive basal-bolus advisor based on reinforcement learning. *IEEE Journal of Biomedical and Health Informatics*, *PP*, 1. http://dx.doi.org/10.1109/JBHI.2018.2887067.

Ting, D., Cheung, C., Nguyen, D. Q., Sabanayagam, C., Lim, G., Lim, Z., et al. (2019). Deep learning in estimating prevalence and systemic risk factors for diabetic retinopathy: A multi-ethnic study. *npj Digital Medicine*, *2*, 24. http://dx.doi.org/10.1038/s41746-019-0097-x.

Tiran, J., Galle, K., & Porte, D. (1975). A simulation model of extracellular glucose distribution in the human body. *Annals of Biomedical Engineering*, *3*, 34–46. http://dx.doi.org/10.1007/BF02584487.

Torrent Fontbona, F., & Lapez, B. (2018). Personalised adaptive CBR bolus recommender system for type 1 diabetes. *IEEE Journal of Biomedical and Health Informatics*, *23*, 387–394. http://dx.doi.org/10.1109/JBHI.2018.2813424.

Vega, I., Hernandez-Contreras, D., Peregrina-Barreto, H., De, J., Rangel-Magdaleno, J., & Ramirez-Cortes, J. (2020). Deep learning classification for diabetic foot thermograms. *Sensors*, *20*, 1762. http://dx.doi.org/10.3390/s20061762.

Vehi, J., Contreras, I., Oviedo, S., Biagi, L., & Bertachi, A. (2019). Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning. *Health Informatics Journal*, *26*, http://dx.doi.org/10.1177/1460458219850682.

Velasco, J. M., Garnica, O., Contador, S., Lanchares, J., Maqueda, E., Botella, M., et al. (2017). Data augmentation and evolutionary algorithms to improve the prediction of blood glucose levels in scarcity of training data. In *2017 IEEE congress on evolutionary computation* (pp. 2193–2200). IEEE.

Vettoretti, M., Cappon, G., Facchinetti, A., & Sparacino, G. (2020). Advanced diabetes management using artificial intelligence and continuous glucose monitoring sensors. *Sensors*, *20*, 3870. http://dx.doi.org/10.3390/s20143870.

Wadghiri, M., Idri, A., Touria, E. I., & Hakkoum, H. (2022). Ensemble blood glucose prediction in diabetes mellitus: A review. *Computers in Biology and Medicine*, *147*, Article 105674. http://dx.doi.org/10.1016/j.compbiomed.2022.105674, URL https://www.sciencedirect.com/science/article/pii/S0010482522004620.

Wan, S., Liang, Y., & Zhang, Y. (2018). Deep convolutional neural networks for diabetic retinopathy detection by image classification. *Computers & Electrical Engineering*, *72*, 274–282. http://dx.doi.org/10.1016/j.compeleceng.2018.07.042.

Wang, W., Tong, M., & Yu, M. (2020). Blood glucose prediction with VMD and LSTM optimized by improved particle swarm optimization. *IEEE Access*, *8*, 217908–217916.

Wasserman, D., & BZ, Z. (1994). Exercise in individuals with IDDM. *Diabetes Care*, *17*, 924–937. http://dx.doi.org/10.2337/diacare.17.8.924.

Wilinska, M., Chassin, L., Schaller, H., Schaupp, L., Pieber, T., & Hovorka, R. (2005). Insulin kinetics in type-1 diabetes: Continuous and bolus delivery of rapid acting insulin. *IEEE Transactions on Bio-Medical Engineering*, *52*, 3–12. http://dx.doi.org/10.1109/TBME.2004.839639.

Williams, B., Borroni, D., Liu, R., Zhao, Y., Zhang, J., Lim, J. Z. M., et al. (2020). An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: A development and validation study. *Diabetologia*, *63*, http://dx.doi.org/10.1007/s00125-019-05023-4.

Woldaregay, A., Arsand, E., Walderhaug, S., Albers, D., Mamykina, L., Botsis, T., et al. (2019). Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artificial Intelligence in Medicine*, *98*, http://dx.doi.org/10.1016/j.artmed.2019.07.007.

Xie, J., & Wang, Q. (2020). Benchmarking machine learning algorithms on blood glucose prediction for type I diabetes in comparison with classical time-series models. *IEEE Transactions on Biomedical Engineering*, *67*(11), http://dx.doi.org/10.1109/TBME.2020.2975959.

Yadav, J., Rani, A., & Singh, V. (2016). Performance analysis of fuzzy-PID controller for blood glucose regulation in type-1 diabetic patients. *Journal of Medical Systems*, *40*, http://dx.doi.org/10.1007/s10916-016-0602-6.

Yang, J., Li, L., Shi, Y., & Xie, X. (2018). An ARIMA model with adaptive orders for predicting blood glucose concentrations and hypoglycemia. *IEEE Journal of Biomedical and Health Informatics*, *23*(3), 1251–1260. http://dx.doi.org/10.1109/JBHI.2018.2840690.

Yang, T., Yu, X., Ma, N., Wu, R., & Li, H. (2022). An autonomous channel deep learning framework for blood glucose prediction. *Applied Soft Computing*, *120*, Article 108636.

Yom-Tov, E., Feraru, G., Kozdoba, M., Mannor, S., Tennenholtz, M., & Hochberg, I. (2017). Encouraging physical activity in patients with diabetes: Intervention using a reinforcement learning system. *Journal of Medical Internet Research*, *19*, Article e338. http://dx.doi.org/10.2196/jmir.7994.

Yu, X., Rashid, M., Feng, J., Hobbs, N., Hajizadeh, I., Samadi, S., et al. (2018). Online glucose prediction using computationally efficient sparse kernel filtering algorithms in type-1 diabetes. *IEEE Transactions on Control Systems Technology*, *28*(1), 3–15. http://dx.doi.org/10.1109/TCST.2018.2843785.

Yu, X., Turksoy, K., Rashid, M., Feng, J., Hobbs, N., Hajizadeh, I., et al. (2018). Model-fusion-based online glucose concentration predictions in people with type 1 diabetes. *Control Engineering Practice*, *71*, 129–141. http://dx.doi.org/10.1016/j.conengprac.2017.10.013.

Zarkogianni, K., Mitsis, K., Litsa, E., Arredondo, M., Fico, G., Fioravanti, A., et al. (2015). Comparative assessment of glucose prediction models for patients with type I diabetes mellitus applying sensors for glucose and physical activity monitoring. *Medical & Biological Engineering & Computing*, *53*, http://dx.doi.org/10.1007/s11517-015-1320-9.

Zecchin, C., Facchinetti, A., Sparacino, G., & Cobelli, C. (2013). Jump neural network for online short-time prediction of blood glucose from continuous monitoring sensors and meal information. *Computer Methods and Programs in Biomedicine*, *113*, http://dx.doi.org/10.1016/j.cmpb.2013.09.016.

Zhao, C., & Yu, C. (2015). Rapid model identification for online subcutaneous glucose concentration prediction for new subjects with type I diabetes. *IEEE Transactions on Biomedical Engineering*, *62*(5), 1333–1344.

Zhu, T., Li, K., Chen, J., Herrero, P., & Georgiou, P. (2020). Dilated recurrent neural networks for glucose forecasting in type 1 diabetes. *Journal of Healthcare Informatics Research*, *4*, http://dx.doi.org/10.1007/s41666-020-00068-2.

Zhu, T., Li, K., Herrero, P., Chen, J., & Georgiou, P. (2018). A deep learning algorithm for personalized blood glucose prediction. In *KHD@ IJCAI* (pp. 64–78).

Zhu, T., Li, K., Herrero, P., & Georgiou, P. (2020). Deep learning for diabetes: A systematic review. *IEEE Journal of Biomedical and Health Informatics*, *PP*, 1. http://dx.doi.org/10.1109/JBHI.2020.3040225.

Zhu, T., Li, K., Herrero, P., & Georgiou, P. (2022). Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning. *IEEE Transactions on Biomedical Engineering*, *70*(1), 193–204.

Zhu, T., Yao, X., Li, K., Herrero, P., & Georgiou, P. (2020). Blood glucose prediction for type 1 diabetes using generative adversarial networks. In *CEUR workshop proceedings*: *vol. 2675*, (pp. 90–94).