Data-Driven Modeling and Prediction of Blood Glucose Dynamics: Machine Learning Applications in Type 1 Diabetes

Ashenafi Zebene Woldaregay [a], Eirik Årsand [b], Ståle Walderhaug [b, c], David Albers [d], Lena Mamykina [d], Taxiarchis Botsis [a], Gunnar Hartvigsen [a]

[a] *Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway*
[b] *Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*
[c] *SINTEF Digital, Software Engineering, Safety and Security, Tromsø, Norway*
[d] *Department of Biomedical Informatics, Columbia University, N.Y., USA*

Corresponding Author:
Ashenafi Zebene Woldaregay,
Medical Informatics and Telemedicine (MI&T) Group,
Department of Computer Science, University of Tromsø –The Arctic University of Norway
Email: ashenafi.z.woldaregay@uit.no

Abstract

Background

Diabetes mellitus (DM) is a metabolic disorder that causes abnormal blood glucose (BG) regulation that might result in short and long-term health complications and even death if not properly managed. Currently, there is no cure for diabetes. However, self-management of the disease, especially keeping BG in the recommended range, is central to the treatment. This includes actively tracking BG levels and managing physical activity, diet, and insulin intake. The recent advancements in diabetes technologies and self-management applications have made it easier for patients to have more access to relevant data. In this regard, the development of an artificial pancreas (a closed-loop system), personalized decision systems, and BG event alarms are becoming more apparent than ever. Techniques such as predicting BG (modeling of a personalized profile), and modeling BG dynamics are central to the development of these diabetes management technologies. The increased availability of sufficient patient historical data has paved the way for the introduction of machine learning and its application for intelligent and improved systems for diabetes management. The capability of machine learning to solve complex tasks with dynamic environment and knowledge has contributed to its success in diabetes research.

Motivation

Recently, machine learning and data mining have become popular, with their expanding application in diabetes research and within BG prediction services in particular. Despite the increasing and expanding popularity of machine learning applications in BG prediction services, updated reviews that map and materialize the current trends in modeling options and strategies are lacking within the context of BG prediction (modeling of personalized profile) in type 1 diabetes.

Objective

The objective of this review is to develop a compact guide regarding modeling options and strategies of machine learning and a hybrid system focusing on the prediction of BG dynamics in type 1 diabetes. The review covers machine learning approaches pertinent to the controller of an artificial pancreas (closed-loop systems), modeling of personalized profiles, personalized decision support systems, and BG alarm event applications. Generally, the reviewers will identify, assess, analyze, and discuss the current trends of machine learning applications within these contexts.

Method

A rigorous literature review was conducted between August 2017 and February 2018 through various online databases, including Google Scholar, PubMed, ScienceDirect, and others. Additionally, peer-reviewed journals and articles were considered. Relevant studies were first identified by reviewing the title, keywords, and abstracts as preliminary filters with our selection criteria, and then we reviewed the full texts of the articles that were found relevant. Information from the selected literature was extracted based on predefined categories, which were based on previous research and further elaborated through brainstorming among the authors.

Results

The initial search was done by analyzing the title, abstract, and keywords. A total of 624 papers were retrieved from *DBLP Computer Science* (25), *Diabetes Technology and Therapeutics* (31), Google Scholar (193), IEEE (267), *Journal of Diabetes Science and Technology* (31), PubMed/Medline (27), and ScienceDirect (50). After removing duplicates from the list, 417 records remained. Then, we independently assessed and screened the articles based on the inclusion and exclusion criteria, which eliminated another 204 papers, leaving 213 relevant papers. After a full-text assessment, 55 articles were left, which were critically analyzed. The inter-rater agreement was measured using a Cohen Kappa test, and disagreements were resolved through discussion.

Conclusion

Due to the complexity of BG dynamics, it remains difficult to achieve a universal model that produces an accurate prediction in every circumstance (i.e., hypo/eu/hyperglycemia events). Recently, machine learning techniques have received wider attention and increased popularity in diabetes research in general and BG prediction in particular, coupled with the ever-growing availability of a self-collected health data. The state-of-the-art demonstrates that various machine learning techniques have been tested to predict BG, such as recurrent neural networks, feed-forward neural networks, support vector machines, self-organizing maps, the Gaussian process, genetic algorithm and programs, deep neural networks, and others, using various group of input parameters and training algorithms. The main limitation of the current approaches is the lack of a well-defined approach to estimate carbohydrate intake, which is mainly done manually by individual users and is prone to an error that can severely affect the predictive performance. Moreover, a universal approach has not been established to estimate and quantify the approximate effect of physical activities, stress, and infections on the BG level. No researchers have assessed model predictive performance during stress and infection incidences in a free-living condition, which should be considered in future studies. Furthermore, a little has been done regarding model portability that can capture inter- and intra-variability among patients. It seems that the effect of time lags between the CGM readings and the actual BG levels is not well covered. However, in general, we foresee that these developments might foster the advancement of next-generation BG prediction algorithms, which will make a great contribution in the effort to develop the long–awaited, so-called artificial pancreas (a closed-loop system).

1. Introduction

Diabetes mellitus (DM) is a chronic metabolic disorder that results in abnormal blood glucose (BG) regulation, mostly either due to the failure of the body to secrete insulin (Type I) or the inability of the body to respond to insulin action (type II). People with diabetes are prone to an increased morbidity and mortality rate as compared to the normal population [1]. Regardless of its current prevalence and burden (415 million adults), the number of adults with diabetes is projected to reach 642 million by 2040 [2]. Apart from financial and other burdens faced by individual patients and families, countries and national health systems are substantially impacted, spending between 5% and 20% of their total health expenditure. For example, Solli et al. [3] estimated Norway's total expenditure, excluding secondary diagnoses, to be €293 million in 2005, representing about 1.4% of the total healthcare expenditures. According to Skrivarhaug et al. [4], the average incidence rate of type 1 diabetes per 100,000 person years in Norway was estimated to be 22.6 (95%, CI 21.4, 23.7) between 1989 and 1996; 28.4 (95%, CI 27.3, 29.6) between 1996 and 2004; and 32.7 (95%, CI 31.5, 34.0) between 2004 to 2012, justifying a slight increase every seven years, which shows the increasing coast of the national health expenditure associated with diabetes [4]. Moreover, the total cost of diagnosed diabetes in 2012 in the United States was estimated to be $245 billion, including $176 billion in direct medical costs and $69 billion in reduced productivity [5]. The complexity of diabetes prognosis and management has opened a way for artificial intelligence (AI) and machine learning techniques to become key technologies that provide solutions and

empower both diabetes patients and their healthcare givers in their everyday lives [6], which in turn have a great potential in minimizing the individual, social, and economic burden of the nation. Recently, in line with this trend, many publicly funded AI research projects have been carried out, including EMPOWER, MOBIGUIDE, COMMODITY12 EU, DIADVISOR, DIABEO, and PEPPER to help diabetic individuals [6].

Recently, the introduction of quantified-self, which aims to empower patients to make decisions about their own health condition through health data collection and documentation, has led to the rapid integration and use of wearable tools and sensors, point of care (POC) devices, and other body area networks for physiological monitoring and other health-related purposes [7, 8]. This has resulted in the accumulation of big personal health data that grow on a daily basis [9], which has created opportunities for further analysis of these data to capture relevant information for better self-monitoring, self-management, and treatments through different AI techniques [9, 10]. Recently, machine learning techniques—due to their adaptive nature in a world with dynamic environments and knowledge—have been successful at solving complex tasks that are difficult to model with other classical approaches. Machine learning and data mining strategies have become increasingly popular with their expanding application in general and within diabetes research in particular. Despite machine learning applications' increasing and expanding popularity in diabetes research in general and in BG prediction and dynamics modeling in particular, updated reviews that materialize the current trends in modeling options and strategies in the context of personalized BG prediction are lacking. However, several other reviews have been conducted on BG prediction and other techniques [11-13]. For example, Oviedo et al. [12] conducted a methodological review regarding the prediction models of BG levels, risks, and events. The reviewers assessed physiological models, data-driven models, and a hybrid approach, and their experimental setup and performance metrics were mainly focused on a closed-loop system (an artificial pancreas) [12]. Moreover, Zarkogianni et al. [13] carried out a critical literature review to pinpoint emerging technologies for diabetes management and advances, mainly focusing on sensors for physiological and lifestyle monitoring, models, and molecular biomarkers for predicting the onset and assessing the progress of DM and on modeling and control methods for regulating BG levels [13]. Furthermore, Kavakiotis et al. [11] performed a systematic review of machine learning and data mining techniques in diabetes research in the context of diabetes prediction and diagnosis, diabetes complications, genetic background and environment, and healthcare and management. Therefore, the objective of this review is to develop a compact guide regarding the modeling options and strategies of machine learning applications and a hybrid system focusing on BG prediction and modeling of personalized BG profiles in type 1 diabetes. The review covers the machine learning approaches that are pertinent to artificial pancreas controllers (closed-loop systems), models of personalized profile, personalized decision support systems, and BG alarm event applications. Generally, the reviewers will identify, assess, analyze, and discuss the current trends of machine learning applications for BG prediction in type 1 diabetes.

2. Machine Learning-based Data Mining Tasks for Type 1 Diabetes

The ubiquitous nature and widespread use of mobile health applications (mHealth apps), sensors, wearables, and POC devices for self-monitoring and management purposes have made possible the generation of automated and continuous personal data, which created the opportunity to use such collected personal data for the modeling of an artificial pancreas (a closed-loop system), a personalized BG profile, personalized decision support systems, and BG alarm event applications through data mining and machine learning techniques. Data mining approaches could be categorized as descriptive or unsupervised (i.e., clustering, association, and summarization) and predictive or supervised learning (i.e., classification and regression) [14]. In this regard, the most widely used machine learning based data mining tasks in the literatures are BG anomalies detection, BG prediction, and BG dynamics and decision making/education models, as shown in Figure 1. The scope of this review is mainly on BG prediction techniques focusing on different classes of machine learning algorithms, artificial neural networks, support vector machines, Bayesian neural networks, decision trees, and others. It should be noted that reinforcement learning is not under the scope of this review.

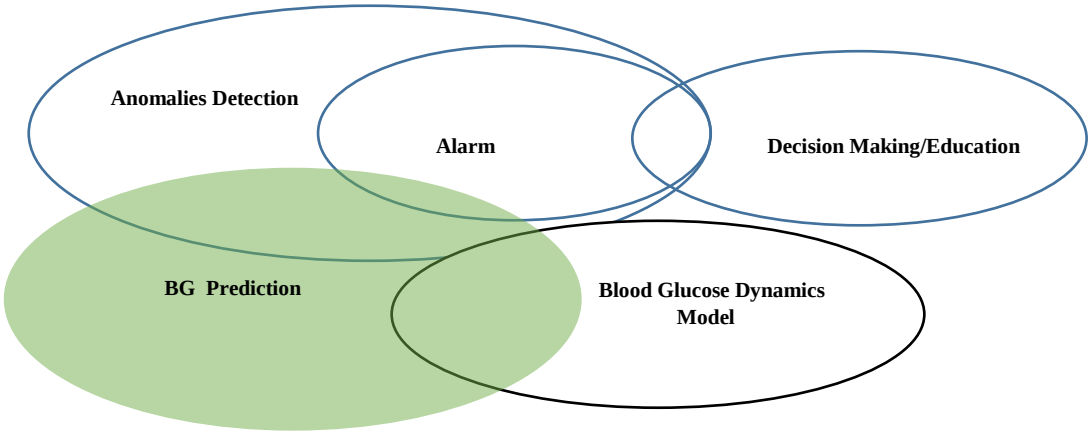

Figure 1: Most widely used machine learning-based data mining tasks, based on self-recorded data from people with type 1 diabetes (modified version of Figure 2 in [14]). The green ellipses indicate the scope of this review.

2.1. Blood Glucose Prediction

BG prediction involves forecasting an individual's BG levels based on past and current history (data) of the patient, mainly to provide the necessary alarm so as to avoid any further complications from hypoglycemia and hyperglycemia. Numerous factors can directly affect BG levels, among which the history of BG values, insulin, physical activity, and dietary intakes are the prominent ones. Moreover, they are also affected by other factors, such as an individual's body mass index, stress level, amount of sleeping time, presence of illness, medications, smoking habit, periods (menstruation), alcoholism, allergies, and altitude. An ideal BG predictor should incorporate as much information as possible to effectively track and predict BG levels, as shown in Figure 2 [15]. BG prediction approaches are broadly classified into three major categories: physiology based (knowledge based), data driven (empirical based), and hybrid (hybrid of the two), as shown in Figure 3 [12]. A physiology-based approach entirely relies on the individual's underlying physiological mechanisms and requires extensive knowledge of each underlying mechanism. It divides the individual BG metabolism into three different regulatory compartments: BG dynamics, insulin dynamics, and meal absorption dynamics [12] and uses various mathematical (differential) equations and probabilistic frameworks to model each compartment. The physiology-based approach is mainly categorized into two: the lumped (semi-empirical) model and the comprehensive model. The lumped model consists of fewer equations and parameters compared to the comprehensive model since most of the organs and tissues are lumped together. However, the comprehensive model is more complex because it considers various organs and tissues separately [16].

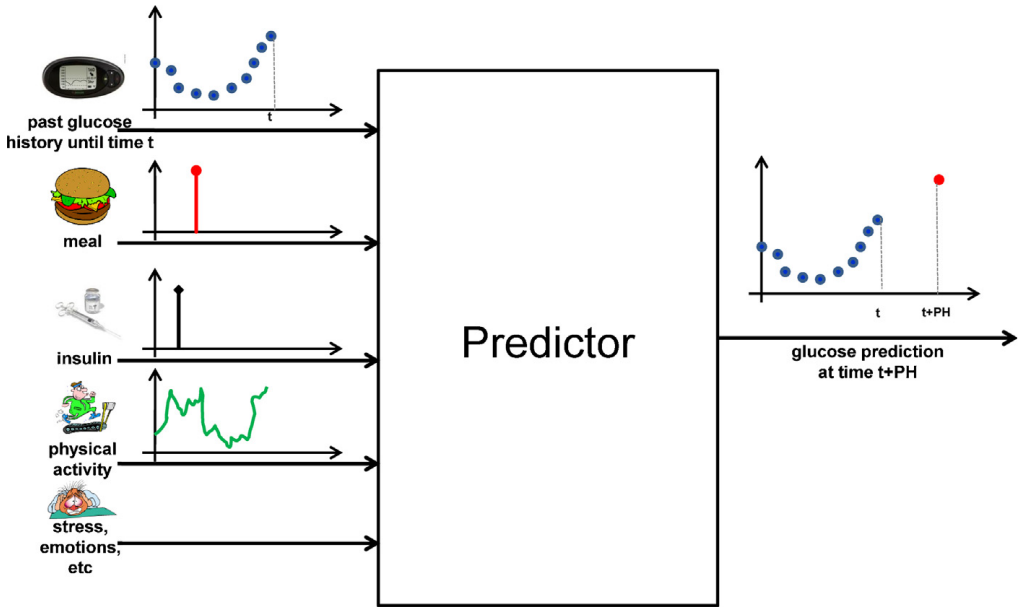

Figure 2: An ideal blood glucose predictor (Reprinted from [15]).

Unlike the physiology-based approach, the data-driven approach uses the individual's self-recorded historical data and requires little understanding of the underlying physiological mechanism; hence, it is commonly known as the black box approach [16]. Generally, it could be divided into three different models: a time series model, machine learning model, and hybrid model. Therefore, it is the purpose of this review to explore, assess, and analyze the state-of-the art machine learning techniques and the hybrid approach for BG prediction.

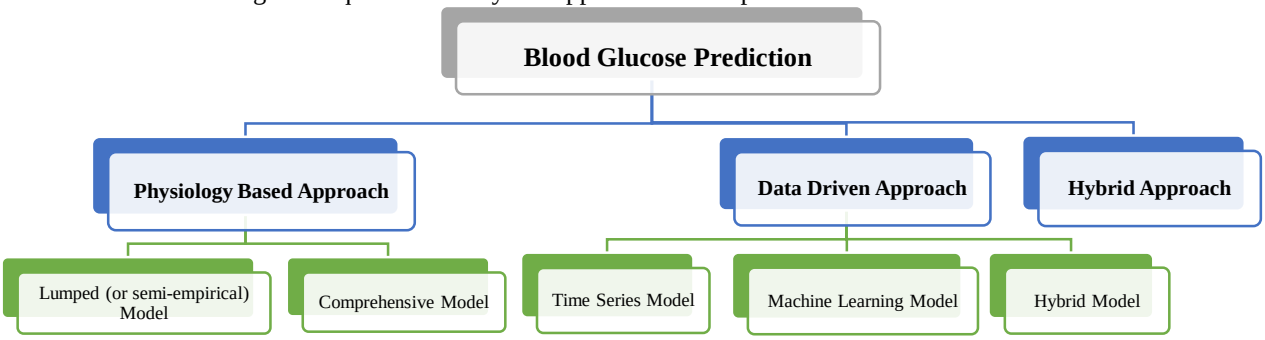

Figure 3: Taxonomy of blood glucose prediction approaches.

3. Method

The objective of this review is to develop a compact guide regarding modeling options and strategies of machine learning applications and a hybrid system focusing on BG prediction in type 1 diabetes. The review covers machine learning approaches pertinent to the controller of an artificial pancreas (closed-loop systems), modeling of a personalized profile, personalized decision-support systems, and BG alarm event applications. For the purpose of the study, a rigorous literature search was conducted between August 2017 and February 2018, through various online databases, including Google Scholar, IEEE Xplore, DBLP Computer Science Bibliography, ScienceDirect, PubMed/Medline, Journal of Diabetes Science and Technology, and *Diabetes Technology & Therapeutics*. Moreover, a reference list of the selected articles was used to extract additional articles to get a complete overview of the field. Peer-reviewed journals, articles, and conference proceedings published between 2000 and 2018 were considered. The inclusion and exclusion criteria were set up through rigorous discussions and brainstorming among the authors. Different combinations of the terms "prediction," "forecasting," "controller of artificial pancreas," "diabetes," "intelligent system," "hybrid system," "machine learning," "BG event alarm," "blood glucose control," "BG personalized decision system," "clinical," "closed-loop system," and "personalized profile" were used during the search. The terms were combined using "AND" and "OR" for a better searching strategy. The relevant articles were first identified by reviewing the title, keywords, and abstracts for a preliminary filter with our selection criteria, and then full text articles that seemed relevant were reviewed. Information from the selected literature was extracted based on some predefined categories, which were based on previous research and further refined through brainstorming sessions among the authors.

3.1. Inclusion and Exclusion Criteria

To be included in the study, the article should develop, test, and discuss machine learning and any of its hybrid algorithms in type 1 diabetes, focusing on the modeling of BG prediction. Therefore, studies that reside outside of these stated scopes are excluded from the review, including all articles written in languages other than English.

3.2. Data Categorization and Data Collection

Information was extracted from the selected articles based on the predefined parameters (variables) and categories. The categories were defined based on rigorous brainstorming and discussions among the authors. These categories were demarcated solely to collect the relevant data and to assess, analyze, and evaluate the model's characteristics and its experimental setup.

Age and Number of Subjects: This category was defined to assess, analyze, and evaluate the number of subjects involved in the algorithm development, thereby quantifying the level of algorithmic validity.

Type of Input: This category was defined to assess, analyze, and evaluate the type of inputs used to develop the prediction algorithm. This includes the key diabetes parameters such as BG, insulin injection, physical activity, dietary information, and others.
Data Format or Type/Data Source/Data Size: This category is defined to assess, analyze, and evaluate the type of data format used as input to the prediction algorithm. This depends on the data sources (i.e., the type of diabetes technologies, mobile application, and others) used for data collection and algorithm development. The data can be from simulated in silico or in vivo (real) patients. It includes different data formats, such as continuous glucose monitoring devices (i.e. CGM), BG simulators (i.e., PADOVA), and m-Health applications (i.e., a diabetes diary) and others.
Input Preprocessing: This category defines the kind of preprocessing that the algorithm implements to avoid missing, sparse, and corrupted input data.
Class of Machine Learning: This category defines the class of machine learning algorithm used to train and test the prediction. It includes different classes of machine learning algorithms: artificial neural networks, support vector machines, Bayesian neural network, decision trees, and others.
Training/Learning Algorithm: This category defines the class of learning algorithms used to train the prediction algorithms. It includes different training algorithms, such as the back-propagation algorithms, kernels, optimization techniques, and others.
Validation techniques: This category defines the kind of validation approaches used to validate the model during training. This includes holdout, random subsampling, k-fold cross-validation, and others.
Prediction Horizon (PH): This category defines the extent of the lead time in which the prediction can be executed with the developed system. It shows how long the future BG can be predicted without losing considerable accuracy as compared with the individual's BG dynamics.
Performance Metrics/Evaluation Criteria: This category defines the type of evaluation metrics used to determine the accuracy of the implemented prediction algorithm. It includes different performance metrics such as mean square error (MSE), error grid analysis, and others.

3.3. Literature Evaluation

The evaluation and analysis of the literature were based on the categories and variables defined above to pinpoint the state-of-the art machine learning-based BG prediction techniques and their associated characteristics along with the experimental setups. The first analysis was conducted based on data characteristics and the type of input the prediction algorithms have used to reveal the state-of-the art inputs used in BG prediction. The second analysis was conducted based on the type of machine learning used in the algorithm development to pinpoint the most adopted class of machine learning in this specific task. The final analysis was conducted based on the performance metrics used to evaluate the performance of the developed system. This analysis will reveal important information regarding the available performance metrics to choose from, which is usually a confusing and difficult process, given the large number of performance metrics.

4. Results

4.1. Relevant Literature

The initial hit was vetted using the title, abstract, and keywords and retrieved a total of 624 papers from DBLP Computer Science (25), *Diabetes Technology and Therapeutics* (31), Google Scholar (193), IEEE (267*), Journal of Diabetes Science and Technology* (31), PubMed/Medline (27), and ScienceDirect (50). After removing duplicates from the list, 417 records remained. Then, we independently assessed the articles and screened them based on the inclusion and exclusion criteria, which eliminated another 204 papers, leaving 213 relevant papers. After a full-text assessment, 55 articles were left, which were critically analyzed, as shown in Figures 4 and 5 and Tables 1 and 2. The inter-rater agreement was measured using a Cohen Kappa test, and disagreements were resolved through discussions.

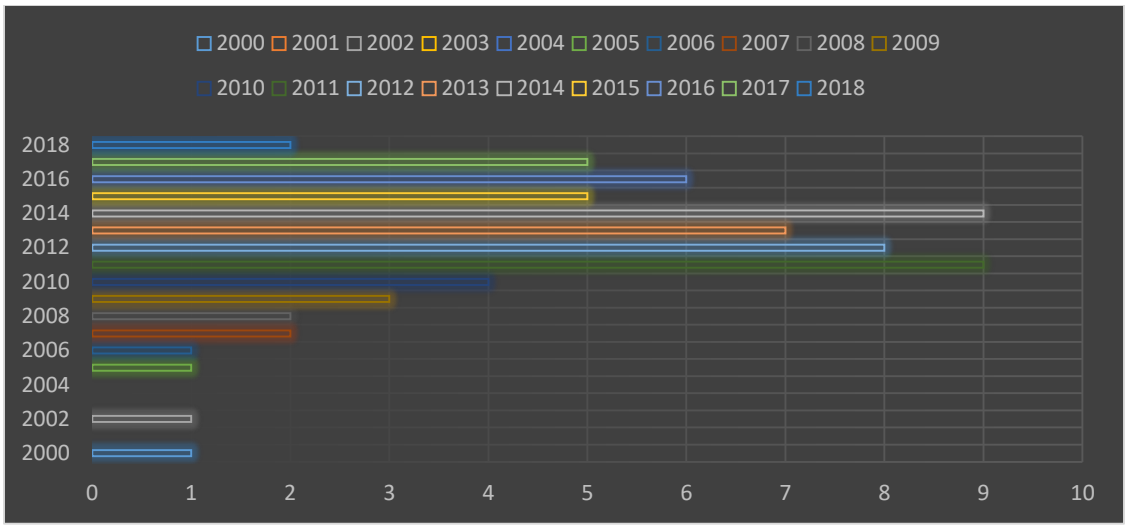

Figure 4: The number of articles included per year of publication.

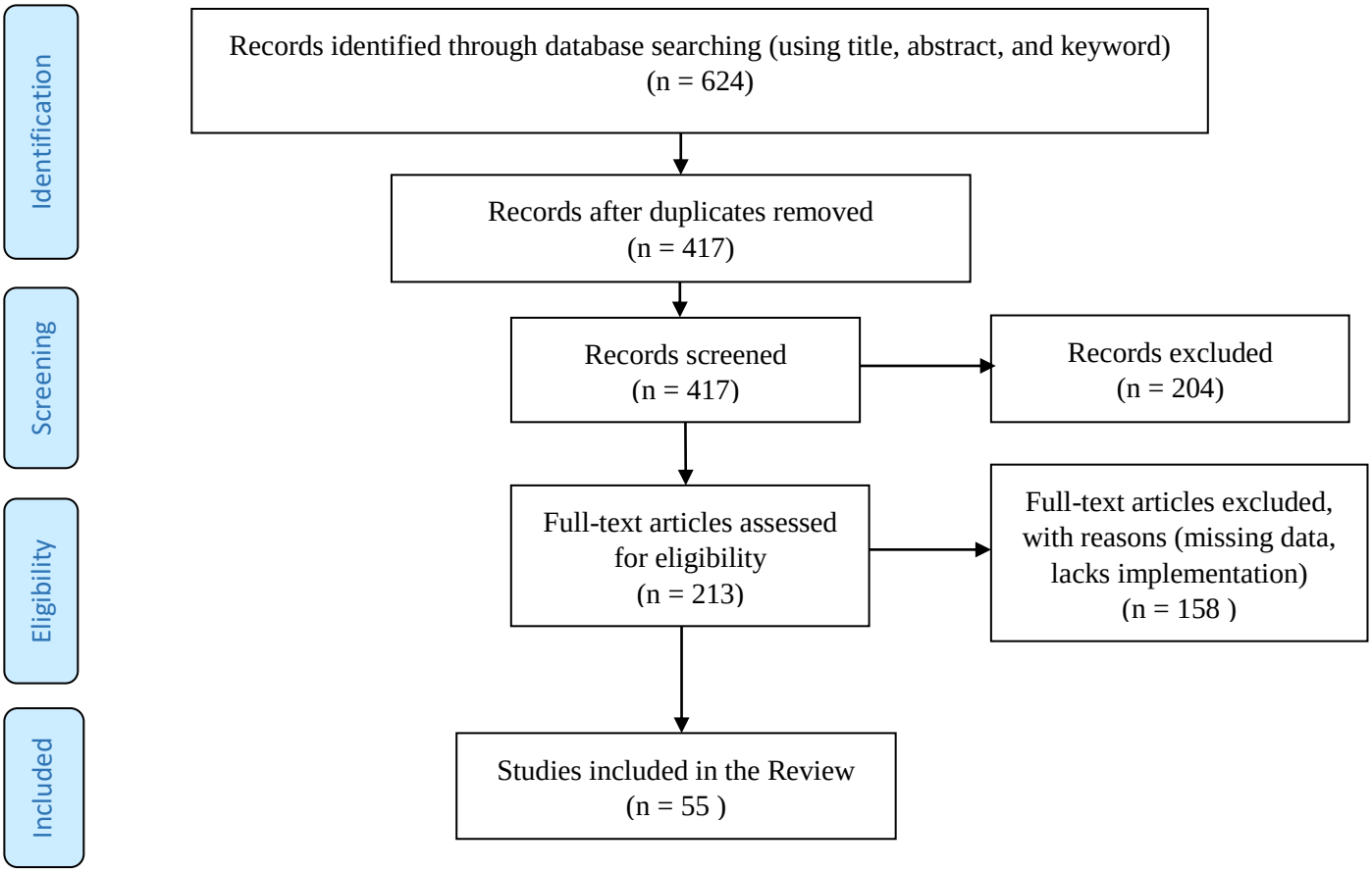

Figure 5: Diagram of the review process.

Table 1: Number of subjects used, data source, class of machine learning, prediction horizon, performance metrics and others information extracted from the literatures included in the study.

Ref.	Subject	Type of input	Data Format/Data source	Input Pre-processing	Class of Machine Learning	Prediction Horizon (PH)	Performance Metrics
[17]	9	Blood Glucose (BG)	Guardian Real Time CGM (Medtronic-Minimed)	Smoothing (LPF of order 11)	Recurrent neural network (RNN)	15, 30, 45, 60 min	Root Mean Squared Error (RMSE), FIT, Normalized Prediction Error (NPE), Clarke Error Grid approach (CEGA)
[18]	9	Blood Glucose (BG)	Guardian Real Time CGM (Medtronic Minimed)	Smoothing (LPF), noise and time lag reduction	Feed forward neural network model	15, 30, 45 min	RMSE, FIT, NPE, CEGA
[19]	1	BG, Insulin, diet, exercise, stress,	Data collected Manually	N/A	Radial-basis function (RBF) network	N/A	RMSE
[20]	5	CGM, insulins, diet	N/A	Filtering-Extended Kalman filter	Support Vector Regression- Gaussian kernel	30, 60 min	RMSE, expert references
[21]	N/A	N/A	N/A	N/A	Layer Recurrent Network (LRN), Elman net, and Nonlinear Autoregressive Neural Network with external input (NARX-net)	20, 40, 60, 80, 100, 120 min	N/A
[22]	30 in silico	BG, Insulin, diet	UVA/Padova simulator	Normalization	Online (real time) & adaptive Recurrent Neural network (RNN)	30 and 45 min	Continuous Glucose-Error Grid Analysis (CG-EGA), sensitivity and specificity
[23] & [24]	20 in silico, 15 real	BG, Insulin, diet	UVA/Padova, FreeStyle Navigator CGM	Scaling, adding a noise on the CGM using AR first order and Gaussian noise	Hybrid (Feed forward Neural network plus linear prediction algorithm) along with physiological model	30 min	RMSE, Temporal Gain (TG), Energy of the Second-Order Differences (ESOD), Index J
[25] & [15]	20 real	BG, derivative of BG, diet	Dexcom SEVEN PLUS CGM	Normalization, Scaling, Bayesian smoothing	Hybrid (Jump neural network along with physiological model)	30 min	RMSE, average TG, ESOD
[26]	10 real	BG, change in BG, physical activities	Guardian real-time CGM, SenseWear Armband	Normalization, quantizing input space	Feed forward neural network (FNN), a self-organizing map (SOM), a neuro-fuzzy network with wavelets as activation functions (WFNN), linear regression model (LRM),	30, 60 and 120 min	RMSE, Correlation Coefficient (CC), Mean Absolute Relative Difference (MARD), and CG-EGA
[27]	6 real	BG, change of BG, physical activity	Guardian real-time CGM SenseWear Armband	Normalization	Neuro-fuzzy model with wavelet activation functions (WFNN)- Gaussian functions as membership function	15, 30, 45, 60 min	RMSE, CC, MARD, and CG-EGA
[28]	12 real	BG, Insulin, diet	Insulin pumps and CGMS (First Department of Pediatrics, P&A Kyriakou Children's Hospital, Athens)	N/A	Hybrid-Compartmental model (MM)) and self-organizing map (SOM) - Vector Quantization Method	30 and 60 min	RMSE, CC, and CG-EGA
[29]	1 real	BG, Insulin, diet, exercise, stress,	Delft University of Technology	Feature extraction using principal component analysis	Wavelet Neural Network	Interval	RMSE
[30]	7,109 users	17 medical examination items	4 years Medical examination database	Feature extraction-random forest with importance score. Feature selection-Sequential Backward Selection (SBS) algorithm	Support vector machine (SVM) and random forest	N/A	ROC, Mean Absolute Error (MAE) and RMSE
[31]	10 real	BG	JDRF CGM Study Group	N/A	Ensemble approach Hybrid-fusion (AR, extreme learning machine, and support vector regression-kernel function (Gaussian))	15, 30 and 45 min	RMSE, relative error, CEGA, and J index.
[32]	10 real	BG, Insulin, exercise, diet, others	Automatic electronic recording device & paper records	N/A	Gaussian process regression	N/A	MAE
[33] & [34]	6 real	BG, Insulin, exercise, diet	Abbott Freestyle and the Dexcom Seven Plus	N/A	Ensemble approach (state-space-based model (SS), a recursive ARX model and a kernel-based predictor)	40 min	CEGA, RMSE
[35]	10 real	BG	Abbott Freestyle and the Dexcom Seven Plus	Data transformation	Kernel-function- Fully Adaptive Regularized Learning (FARL)	30, 60, 75 min	CEGA, Prediction Error Grid Analysis (PRED-EGA)
[36] & [37]	4 real	BG, Insulin	Self-Monitoring Blood Glucose (SMBG) - One Touch Ultra® glucometer, Medtronic CGMS	Smoothened using cubic splines Interpolation. De-noising with extended Kalman filter algorithm	Feature based Feed-forward Neural Network (FNN)	30,45 and 60 min	RMSE and Time delay
[38] & [39]	2 real, 1 in silico	Time, Diet, Insulin, BG, illness, stress, pregnancy	Diabetes simulator, Diabetic Outpatient Department of Glasgow's Royal Infirmary	N/A	Elman Recurrent Neural Network	1, 2, 4, 8 hours	Average error
[40]	1 in silico	BG, diet, insulin	Diabetes simulator, AIDA	N/A	Elman recurrent artificial neural networks (ANNs)	15, 30, 45, and 60 min, 2, 4, 6, 8, or 10 hours	RMSE
[41]	5 in silico	BG	Simulator program (www.2aida.net) and a CGM system - FreeStyle Libre	Cubic spline interpolation, scaling	Support Vector Regression (SVR)- Radial basis function (RBF) as kernel	30 and 60 min	Arithmetic mean of the relative error
[42]	10 real	BG, time, insulin, exercise, diet, Stress level	N/A	Normalization, renormalization	Feed-forward neural networks (MLP), Elman Recurrent Neural Network	N/A	N/A
[43]	5 real	BG, diet. Insulin	N/A	Extended Kalman filter, ARIMA	Hybrid (Generic physiological model & Support Vector Regression- Gaussian kernel)	30 and 60 min	RMSE, expert references
[44]	15 real	BG	Medtronic Guardian, Abbott Navigator	Noise pre-filtering- causal Kalman filtering method. cubic spline to recover missing samples	Feed-forward neural network model	15, 30 and 45 min	RMSE and prediction delay
[45]	27 real	BG, insulin, diet, lifestyle, emotional factors, hypo/hyperglycemic symptoms	CGMS Gold Medtronic	N/A	Feed-forward neural network model	75 min	CEGA, RMSE and Mean Absolute Difference percent (MAD%)
[46] & [47]	18 real	BG, insulin, diet, lifestyle, emotional factors, hypo/hyperglycemic symptoms	Medtronic CGM, electronic diary	N/A	Time-lagged feed-forward neural networks, Genetic algorithm- Determining Optimal step sizes and momentum values for the minimization of error	50, 75, 100, 120, 150 and 180 min	MAD%
[48]	5 real	Time, BG, and Electronic Medical Records data	Medtronic CGMS	Normalization	Feed-forward Neural Network	75 min	CEGA, Overall error (MAD%)
[49]	1 real	BG, insulin, diet, exercise	N/A	Fuzzy approximation of food, exercise	Hybrid- (Compartmental Model & Feed-forward neural network, fuzzy logic, and expert system)	N/A	Mean percent error (MPE) and mean absolute percent error (MAPE)

[50]	4 real	BG, insulin, diet	Medtronic MiniMed	N/A	Hybrid – (Compartmental model & Feed Forward Neural Network, Recurrent Neural Network)	N/A	RMSE, CC
[51]	1 real	Time, BG, insulin injection, diet	Diary data	Normalization	Hybrid- (Compartmental Models (CMs) and Recurrent NN (RNN))	15 min	RMSE, CC, MAD%, Standard Deviation (SD)
[52]	25 real	BG, Rate of change in BG	DirectNet Central Laboratory	Normalization, smoothing (1-order Butterworth filter)	Deep Neural Network-classical radial basis function networks, judge predictor	30 min	PRED-EGA grid
[53]	2 real	BG, Insulin, diet, exercise	Medtronic Paradigm 522 insulin pump and CGMs, diary	Normalization	Adaptive network-based fuzzy inference system (ANFIS)- triangular & Linear membership function	120 min	Average error of prediction
[54]	70 real	BG, insulin, diet, exercise, sleep, hypoglycemic symptoms	Smartphone diary	pooled panel data (PPD) regression - Clustering	Support vector machine (SVM), Decision tree, random forest	N/A	MAE, RMSE, and coefficient of determination
[55]	4 in silico, 3 real	BG	AIDA simulator, Dexcom SEVEN	N/A	Hybrid- (Genetic Algorithms (GA) and Compartmental Model)	30 min	N/A
[56]	6 real	BG, insulin, diet, physical activity	Dexcom G4 Platinum CGM, Paradigm Veo - 754 insulin pumps, mylife Omnipod insulin pumps, AccuChek insulin pump, Fitbit Charge HR™ devices	N/A	Hybrid-Fused (ARX and Elman simple recurrent neural network) for prediction and Extreme Learning Machine for correction	15, 30 and 45 min	RMSE, CC and TL
[57]	15 real	BG, Insulin, diet, physical activity	Guardian Real-Time CGM, SenseWear Armband, paper diary	N/A	Single hidden layer feedforward neural networks (kernel RLS, Gaussian kernel)-Extreme Learning Machine	30 min	RMSE, TG, ESOD
[58]	15 real	BG, Insulin, diet, physical activity	Guardian Real-Time CGM, SenseWear Armband, paper diary	Ranking feature set- Random forests (RF) and RReliefF algorithms, bootstrap resampling	Support vector regression- radial basis function or Gaussian processes	30 and 60 min	Average RMSE
[59]	27	Time, BG, insulin, Physical activity	Guardian Real-Time CGM, SenseWear Armband, paper diary	N/A	Hybrid- (Random Forests regression technique & compartmental model)	15, 30, 60 and 120 min	RMSE, CEGA
[60] & [61] & [62] & [63]	2 real	BG, Insulin, diet, physical activity	Guardian Real-Time CGM, SenseWear Armband, paper diary	N/A	Hybrid- (Support Vector machines- linear kernel and Compartmental model)	15, 30, 60 and 120 min	RMSE
[64]	20 real	BG, Insulin, diet, physical activity	Medtronic Minimed Guardian Real-Time system	N/A	Feedforward neural networks-multilayer perceptron	N/A	RMSE, normalized RMSE, CC
[65]	25 in silico	BG	GlucoSim software	Digital Noise filtering techniques-Kalman filter	Time-lagged feed-forward NN	60 and 120 min	MAD%
[66]	18 real	BG, Insulin, diet, Heat flux, skin temperature, and METs (Metabolic Equivalent)	SenseWear Pro 2 armband, Guardian CGMS, food and insulin diary	N/A	Gaussian Processes (GPs)- Bayesian framework	25 min, 1 hour and 4 hours	N/A
[67] & [68]	23 real	BG and insulin	Medtronic insulin pumps, real-time CGM system	Smoothening and Filtering	Hybrid-Fused (autoregressive model with output correction – cARX, & recurrent neural network – RNN)-Data fusion (Genetic Algorithms (GA), & Genetic Programming (GP))	15, 30 and 45 min	RMSE, time lag (TL), and CC
[69]	10 real	BG, insulin, diet	Medtronic CGMs	Feature reduction through averaging	K-nearest neighbors regression, Random forest regression, Hybrid- (Symbolic regression by tree-based genetic programming & compartment models)	30, 60, 90 and 120 min	CEGA
[70] & [71]	100 virtual, 5 real	BG, insulin, diet	UVA/Padova T1D simulator	N/A	Hybrid (Genetic Programming-Grammatical Evolution & Physiological model)	N/A	CEGA, RMSE, Mean absolute deviation (MAD), & MARD
[72]	17 real	BG	Guardian CGM system	Stationarity & autocorrelation test- Augmented Dickey-Fuller (ADF)	Autoregressive neural network (AR-NN)	30, 60 and 90 min	Mean absolute error (MAE), and RMSE
[73]	1 real	BG, insulin, diet	Paradigm Real-time Insulin Pump, MiniMed CGM	N/A	Recurrent neural network - Neural Network Autoregressive external input, Recurrent Multilayer Perceptron (RMLP)	N/A	N/A
[74]	1 real	Times, insulin, diet, exercise, BG	N/A	N/A	Hybrid-(Recurrent neural network & Compartmental model), neuro-fuzzy time series models	N/A	The explained variance as a function of mean squared prediction error
[75]	2 real	BG, insulin, exercise, diet	Guardian Real-Time CGMS, SenseWear Body Monitoring System armband, diet manually collected by the patient	N/A	Hybrid-(Compartmental model & Support Vector Regression- linear kernel)	15, 30, 60 and 120 min	RMSE , CC, CEGA
[76, 77]	5 in silico, 8 in silico	BG, diet, insulin	AIDA simulator	N/A	Genetic programming-Grammatical Evolution (GE)	N/A	Mean percentage average error, CEGA
[78]	10 real	BG, Insulin, diet	N/A	Scaling and normalization, interpolation of missing values	Recurrent neural network (RNN)	30, and 60 min	RMSE
[79]	1 real	BG, insulin, diet	Paradigm Real-time Monitoring (CGM & Insulin Pump)	N/A	Neural Network Autoregressive external input (NNARX), Recurrent multilayer perceptron neural network	N/A	Mean Square Error (MSE) and absolute error
[28]	12 real	BG, insulin, diet	insulin pumps and CGMS (Diabetes Centre, First Department of Pediatrics, P&A Kyriakou Children's Hospital, Athens)	N/A	Hybrid- (Compartmental Models (CMs) and a Self-Organizing Map (SOM) - Vector Quantization Method)	30 and 60 min	RMSE, CC, CG-EGA
[80]	12 real	BG	Freestyle Navigator CGM System (Abbott Laboratories).	N/A	Support vector regression (SVR) based on differential evolution (DE) algorithms	15 min, 30 min, 45 min and 60 min	Root mean square error (RMSE), the mean absolute percentage error (MAPE) and the fitness degree (R2).
[81]	13 real	BG	Freestyle Navigator CGM System (Abbott Laboratories).	Spline interpolation method was used to overcome missing records	Feed forward Artificial Neural Networks (ANN)	15 min, 30 min, 45 min and 60 min	Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Fitness degree R2, the Relative Error Analysis (REA), Sum of Squares of Glucose Prediction Errors (SSGPE), CEGA.

Table 2: Subject age group, data size and validation approach used for development, and reported model performance extracted from the literatures.

Ref	Subject Age Group	Data size (Training, Validation & Testing)	Validation Approach	Mathematical Accuracy	Clinical Accuracy
[17]	N/A	The average duration of glucose measurements for each patient is 2 days, 288 samples for each day	Random subsampling: Data set consists of 4916 samples (4416 samples used to train, 500 samples used to test and validate)	15 min (FIT (%)-95.33, RMSE (mmol/L) - 0.14, NPE (%)-1.7), 30 min (FIT (%)-85.83, RMSE (mmol/L)- 0.42, NPE (%)-5.27), 45 min (FIT (%)-72.3, RMSE (mmol/L)- 0.84, NPE (%)-10.28), 60 min (FIT (%)-56.61, RMSE (mmol/L)- 1.32, NPE (%)-16.2)	Clarke's EGA:15 min (A- 100, B-0, C-0, D-0, E-0), 30 min (A- 98.6, B- 1.3, C-0, D-0, E-0),45 min (A- 91.5, B- 8.4, C-0, D-0, E-0),60 min (A-78.7, B- 19.3, C-0, D- 1.95, E-0),
[18]	N/A	The average duration of glucose measurements for each patient is 2 days, 288 samples for each day	Random subsampling: Data set consists of 4916 samples (4416 samples used to train, 500 samples used to test and validate)	15 min (FIT (%)-94.68, RMSE (mmol/L) - 0.15, NPE (%)-1.94), 30 min (FIT (%)-85.5, RMSE (mmol/L)- 0.42 , NPE (%)-5.37), 45 min (FIT (%)-72.1, RMSE (mmol/L)- 0.83, NPE (%)-10.2)	Clarke's EGA:15 min (A-100, B-0, C-0, D-0, E-0), 30 min (A-98.53, B-1.47, C-0, D-0, E-0),45 min (A-86.7, B-10.5, C-0, D-2.7, E-0)
[19]	N/A	A continuous period of 77 days from one patient	Repeated Hold-out Method: 20 holdout conditions, consist of a random distribution (train data set - 40%, validation data set - 30% and test data -30%)	Interval RMSE Validation data–Morning (0.0710), Afternoon (0.0491), Evening (0.0263), Night (0.0119)	N/A
[20]	N/A	Approximately 1,400 days' worth of clinical patient data	Random subsampling: A total of 200 timestamps in the dataset, 40 points per patient were manually selected to mimic a diverse set of circumstances.	N/A	Physician expert references
[21]	N/A	9 days dataset with 626 vectors	N/A	40 min (RMSE-0.313), 50 min (RMSE-0.338), 60 min (RMSE-0.346)	N/A
[22]	10 adults, 10 adolescents, and 10 children	Data of 8 days derived from a virtual population of 30 diabetes patients	Hold-out: The first 4 days per patient were used for training, whereas the remaining data were used for evaluation	Mean (SD) (30 min- (Adults (RMSE (mg/dL)- 2.8 (0.4)), Adolescents (RMSE-3.1 (0.8)), Children (RMSE-4.5 (2.2))), 45 min- (Adults (RMSE-4.0 (0.7)), Adolescents (RMSE-4.4 (0.8)), Children (RMSE -6.3 (3.0))	According to the CG-EGA: more than 89% of the predictions and 93–94% of the predictions for hypoglycemic range were clinically accurate for all the patients and PHs.
[23] & [24]	N/A	Simulated Data: 11 consecutive days of monitoring BG, insulin along with 3 meals per day	k-fold cross-validation	Mean ± SD:30 min (RMSE (mg/dl) -14±4.1)	N/A
[23] & [24]	N/A	Real Data: Fifteen type-1 diabetic patients were monitored for seven consecutive days (meals, carbohydrate intake for hypo-corrections, and insulin dosages)	k-fold cross-validation	Mean ± SD:30 min (RMSE (mg/dl) - 9.4±1.5)	Time lag (TG) in (min): 24.9 ± 4.4
[25] & [15]	N/A	20 T1D patients, monitored for 2 or 3 consecutive days in real-life conditions.	k-fold cross-validation (training set constituted by 70 % of the data, and the validation set constituted by the remaining 30 % of data), Bayesian regularization	Mean ± SD:30 min (RMSE (mg/dl)-16.6 ± 3.1)	Time lag (TG) in (min): 18.5 ± 3.4
[26] (FFNN)	7 males and 3 females (41.8 ± 14.39 of age)	Data corresponding to identical number of days (6) for each patient were used	10-fold cross-validation	Mean ± SD: 30min (RMSE - 13.31 ± 4.47), 60min (RMSE - 22.66 ± 6.86), 120min (RMSE - 37.62 ± 11.79)	CG-EGA-Accurate Readings (AR): 30 min (hypoglycemia (73.29%), euglycemia (88.46%), hyperglycemia (84%)), 60 min (hypoglycemia (54.25%), euglycemia (88.83%), hyperglycemia (83.95%)), 120 min (hypoglycemia (33.65%), euglycemia (88.18%), hyperglycemia (82.81%))
[26] (SOM)	7 males and 3 females (41.8 ± 14.39 of age)	Data corresponding to identical number of days (6) for each patient were used	10-fold cross-validation	Mean ± SD: 30min (RMSE - 11.42 ± 2.33), 60min (RMSE - 19.58 ± 3.80), 120min (RMSE - 31.00 ± 6.07)	CG-EGA-Accurate Readings (AR): 30 min (hypoglycemia (91.11%), euglycemia (91.86%), hyperglycemia (88.59%)), 60 min (hypoglycemia (78.47%), euglycemia (90.45%), hyperglycemia (86.96%)), 120 min (hypoglycemia (56.40%), euglycemia (88. 86%), hyperglycemia (84.73%))
[26] (WFNN)	7 males and 3 females (41.8 ± 14.39 of age)	Data corresponding to identical number of days (6) for each patient were used	10-fold cross-validation	Mean ± SD: 30min (RMSE - 15.22 ± 2.17), 60min (RMSE - 24.66 ± 3.39), 120min (RMSE - 39.59 ± 5.03)	CG-EGA-Accurate Readings (AR): 30 min (hypoglycemia (76.18%), euglycemia (89.48%), hyperglycemia (85.13%)), 60 min (hypoglycemia (64.74%), euglycemia (88.89%), hyperglycemia (84.57%)), 120 min (hypoglycemia (51.51%), euglycemia (87.34%), hyperglycemia (82.36%))
[27]	N/A	Data from the medical records of 6 T1DM patients for an observation period ranging from 7 to 15 days (mean ± standard deviation: 10.83 ± 3.86) were used.	10-fold cross-validation	N/A	CG-EGA- Zones A (15 min (94.35 ± 5.66), 30 min (86.70 ± 3.76), 45 min (78.08 ± 7.56), 60 min (71.89 ± 9.33))
[28]	7 male and 5 female (19.83 ± 12.28 of age)	Patients were monitored for a ten-day period.	Hold-out: 60% of the data for training purposes (model development), while the remaining 40% for testing (model evaluation)	30 min (RMSE (mean ± standard deviation (SD): 14.10 ± 4.57) and CC (mean ± SD: 0.94 ± 0.02)), 60 min (RMSE (mean ± SD: 23.19 ± 6.40) and the CC (mean ± SD: 0.84 ± 0.05))	CG-EGA-Accurate Readings (AR): 30 min (hypoglycaemia (81.06%), euglycemia (92.18%), hyperglycaemia (88.27%)), 60 min (hypoglycaemia (63.22%), euglycemia (92.18%), hyperglycaemia (87.19%))
[29]	N/A	Dataset from one patient covering a period of 77 days	Multi-fold cross validation- (10-fold cross validation)	RMSE (Morning (0.0450), Afternoon (0.0348), Evening (0.0330), Night (0.0170))	N/A
[30]	4,095 males, and 3,501 females (20-50 of age)	4 years medical examination data	Random subsampling: 2/3 of the data is training set and the rest testing set	FS-random Forest (AUC(74.92%), RMSE (0.5706), MAE(0.3200)), Random Forest (AUC(71.96%), RMSE (0.5996),MAE(0.3347)),FS-SVM (AUC(72.19%), RMSE (0.6672),MAE(0.3533)),SVM (AUC(71.67%), RMSE (0.6909),MAE(0.3685))	N/A
[31]	N/A	CGMS readings of each patient included 860 CGMS data points with 5-min sampling period (in total, 4,300 min)	Hold-out: 60% for training and the rest for testing and validation. These first 500 points (2,500 min) for each subject were used for training, and the other 360 points (1,800 min) are validation data.	RMSE (mg/dl): 30 min (19.0 ± 0.3)	CG-EGA- Zones A: (30 min (85.7±0.3)) J Index: (30 min (12.0 ± 2.1))
[32]	N/A	Each patient's medical history corresponds to a period from 116 (926 observations) to 149 (1327 observations) days of measurements.	Hold-out: 80% for model training and 20% for testing	Mean MAE: 21.5-23 mg/dl	N/A
[33]	N/A	20 datasets simulated, each 8 days long.	Hold-out: One of the 20 datasets was used for training and the others were considered test data.	RMSE [mg/dl]: 60 min (8.1)	N/A
[33] & [34]	3 Men /3 Women, 32–68 of age	Each trial ran over three days (Meal and insulin administration were noted in a logbook, glucose was monitored by Abbott Freestyle (DAQ) and Dexcom Seven Plus (DIAdvisor I) CGM systems	Hold-out: The model was trained on second trail data and validated on the third trail data	Median RMSE/ $RMSE_{best}$ [Min-Max]: 40 min (1.03 [0.75–1.04])	CG-EGA- Zones A: 40 min (95.5%)
[35]	Male and female between 18 -70 years old	The DAQ-trial clinical record of nearly 10 days of CGM data collected with Abbott's Freestyle Navigator (Δt = 10 (min)), and another record of CGM data were collected for three days with the use of DexCom SEVEN PLUS (Δt = 5 (min)).	Hold-out: One patient CGM-measurements collected during one day of the DAQ-trial with the use of an Abbott sensor as learning datasets and the rest for testing	N/A	CG-EGA- Zones A: 30 min (91.3%), 60 min (75.14%), 75 min (68.77%)
[36] & [37]	age10±4	One day (24 hr) data collected through self-monitoring blood glucose (SMBG), Medtronic CGMS and other sources.	Random subsampling: 50% of data is used for training, 25% for validation and 25% for testing	RMSE (mg/dl): 30 min (10),45 min (15), 60 min (20)	Time lag in minutes (Mean± SD): 30 min (3.2±2), 45 min (4.5±3), 60 min (7.6±4.1),
[38] & [39]	15 old girl and 32 old pregnant woman	Both patients regularly monitored and recorded, in a diary, their BG, insulin, diet and physical exercise for a 10 days period and 122 events in total.	Hold-out: Most of the data sets were used during training (97 events), with only a small number used to evaluate performance.	RMSE (mg/dl): One event-step prediction (27)	N/A
[40]	N/A	For a single patient Twenty-eight days of data were produced from AIDA	Random subsampling: Divide random training 60%, validation 20%, and testing 20%)	15, 30, 45, and 60 minutes ($RMSE_{5day}$ of 0.15 ± 0.04 SD mmol/L, and an $error_{max}$ of 0.27 mmol/L), 8 hr ($RMSE_{5day}$: 0.14 ± 0.16 SD mmol/L, $error_{max}$: 0.20 mmol/L), 10 hrs. ($RMSE_{5day}$: 0.14 ± 0.16 SD mmol/L, $error_{max}$: 0.36 mmol/L).	N/A
[41]	N/A	Simulated: For each patient, 25 days of BGLs were simulated	Five-fold cross validation	Arithmetic mean of the relative error over all samples of one day T: 30 minute (0.2-4 %),60 minute (0.3-7 %)	N/A
		1 patient CGM data amounted to 4635 readings over a period of roughly 35 days	Five-fold cross validation	Arithmetic mean of the relative error over all samples of one day T: 30 minute (19 %)	N/A
[42] Elman Recurrent	Woman, ages between 17 and 26 years	The recorded data that was used covers a continuous period of 75 days for some of patients and 135 days for	Hold-out: 75% of datasets used to training and the rest for testing	Mean of prediction Errors: (24.1449 (mg/dl))	N/A

Neural Network		another. For each day we have recorded data just in the morning and afternoon and during this interval.			
[42] MLP Neural Network	Woman, ages between 17 and 26 years	The recorded data covers a continuous period of 75 days for some patients and 135 days for another. For each day data recorded just in the morning and afternoon and during this interval.	Hold-out: 75% of datasets used to training and the rest for testing	Mean of prediction Errors: (10.4023 (mg/dl))	N/A
[43]	N/A	A database of over 1,600 day worth of clinical data of patients	Random subsampling: an evaluation dataset of 200 timestamps, 40 points per patient and were manually selected to reflect the diverse set of situations the predictor would encounter in practice	RMSE(mg/dl)): 30 min (22.6), 60 min (35.8)	N/A
[44]	N/A	The Guardian dataset includes data from nine patients wore the CGM intermittently for 72 h/week over a 4-week period. The dataset includes 12 daily profiles for every patient: six full-day and six half-day recordings. The FreeStyle Navigator dataset includes data from six patients wore the CGM system for around 72 h. The dataset includes two complete daily profiles for every patient	Hold-out: For each dataset, three subjects (each with two daily profiles) were used for training the NNM. The validation set for the Guardian monitor included six patients and six profiles per patient. The validation set for the FreeStyle Navigator monitor included three patients and two profiles per patient.	Guardian CGM (RMSE (mean±SD mg/dL):15 min (9.74±2.71), 30 min (17.45±5.44), 45 min (25.08±8.73)), FreeStyle Navigator CGM (RMSE (mean±SD mg/dL): 15 min (10.38±3.15), 30 min (19.51±5.53), 45 min (29.07±6.77))	**Guardian CGM** (Delay (mean±SD min): Upward-15 min (3.92±1.21), 30 min (11.65±4.11), 45 min (16.46±4.79)), Downward-15 min (5.10±1.65), 30 min (15.92±3.64), 45 min (23.82±5.13)), **FreeStyle Navigator CGM** (Delay (mean±SD min): Upward-15 min (4.58±1.42), 30 min (7.26±3.34), 45 min (11.50±6.85)), Downward-15 min (5.37±2.16), 30 min (14.56±3.81), 45 min (28.67±4.62)),
[45]	N/A	Twenty-seven patients' data recorded using CGM and Pocket PC-based electronic diary facilitated documentation of insulin dosages, nutritional intake, hypoglycemic/ hyperglycemic symptoms, lifestyle/activities, and emotional factors.	Hold-out: Training set that included 23,432 vectors of CGM and electronic diary data collected in 17 patients. Evaluated using data derived from 10 patients not included in the model training set. This test dataset included 394.3 h of CGM and electronic diary data.	75 min (Overall MAD% (22.1), MAD% (non-hypoglycemic) (18.1), RMSE (mean±SD mg/dL) (43.9±6.5), RMSE (mean±SD mg/dL) (non-hypoglycemic) (43.0±6.4))	CG-EGA- 75 min: Zones A (62.3%), Zones (30.0%)
[46] & [47]	N/A	18 patients used CGM for a duration between 3 and 9 days	Hold-out: Training sets using 11–17 patients were used to generate neural network models. The performance of each neural network model was evaluated using the CGM and electronic diary data from a patient who was not included in the training data	Overall MAD(%) (50 min (6.7), 75 min (8.9), 100 min (11.7), 120 min (14.5), 150 min (16.6), 180 min (18.9))	N/A
[48] - A patient specific neural network model	38 years old	283.9 hr (3,407 data points) of CGM and concurrent medical records data from a 38 years old trauma patient (who had an intensive care stay of 16 days).	Hold-out: Trained using 243.6 hours (2,923 data points) of CGM and medical records data, 40.3 hours [484 data points] of CGM and medical records) was used to test the performance.	Overall error (MAD%): 75 min (7.9%)	75 min - CEGA revealed that 95.1% of the predictions fell within region A of the error grid and 4.9% fell within region B of the error grid.
[48] - feed forward neural network	N/A	556 hr (6,672 data points) of CGM and medical records data from 5 critical care patients	Hold-out: Trained using 515.7 hours (6,188 data points) of CGM and medical records data, 40.3 hours [484 data points] of CGM and medical records was used to test the performance.	Overall error (MAD%): 75 min (15.9%)	75 min - CEGA revealed that 69.8% of the predictions fell within region A of the error grid and 30.2% fell within region B of the error grid
[50] - Feed-Forward NN (FFNN)	Children (13-22), 2 female and 2 male	The patients were monitored for a period between 3 to 5 days.	Hold-out: The recorded days have been divided into two disjoint datasets: training, and testing sets. For each patient, the data of the first days (minimum 3 days) using CGMs have formed the training set, and the data of the last day using CGMs the testing set.	RMSE (mg/dl): 7.19	N/A
[50] - RNN using online Teacher-Forcing (TF) algorithm	Children (13-22), 2 female and 2 male	The patients were monitored for a period between 3 to 5 days.	Hold-out: The recorded days have been divided into two disjoint datasets: training, and testing sets. For each patient, the data of the first days (minimum 3 days) using CGMs have formed the training set, and the data of the last day using CGMs the testing set.	RMSE(mg/dl): 11.58	N/A
[51]	N/A	Data from a Type 1 diabetes patient covering a period of 69 days, have been used as input to the proposed system	Hold-out: From the available Type 1 diabetes patient data, the first 59 days, consisting of 275 glucose measurements, have been used in the training set, and the last 10 days, with 45 glucose measurements, in the testing set	On – Line RTRL – FR (RMSE(mg/dl): 41) On – Line RTRL - TF(RMSE(mg/dl): 45)	N/A
[52] Generic	Children less than 18 years old	Clinical data set provided by the DirectNet Central Laboratory, which lists BG levels of different patients taken at 5-min intervals with the CGM device	Hold-out: Consider the entire data for 30% of the patients as training data, and predict the BG level for the remaining 70% of patients in the data set.	N/A	Average PRED-EGA scores (in percent): Accuracy - 30 min (Hypoglycemia (86.41), Euglycemia (85.05), Hyperglycemia (62.24))
[53]	N/A	Data sets from two patients over eight weeks were used to prototype, train, and assess the mode. Data from the CGM and pump were used in conjunction with a diary of food consumed, carbohydrates, exercise type and duration, and meal times	Hold-out: For training, six sets of data containing insulin, carbohydrate, exercise level, and time of day were used. Each of the data sets contained 24 periods for every consequent 5-minute interval for a total duration of two hours (24 x 5 = 120 minutes).	Average error of prediction: 30 min (31 mg/dl), 60 min (57 mg/dl), 120 min (103 mg/dl)	N/A
[54]	N/A	Dataset includes 70 sets of data recorded from the UCI Machine Learning Repository	N/A	RMSE (SVM (68.76), DT (41.06), RF (39.73)), MAE (SVM (63.097), DT (36.423), RF (37.586))	N/A
[56]	Age 22-29 years	Sensor glucose, insulin pump data, food intake, and physical activity data from six individuals with T1D (HbA1c: 6.83 ± 0.75%; body mass index: 24.79 ± 4.71 kg/m2) under sensor-augmented pump (SAP) therapy	Hold-out: Data 48 h prior to the exercise intervention were used for training purposes, while data 35 h after the intervention were used for evaluation	RMSE(mg/dl)): Mean (SD) - 15 min (8.9 (1.70)), 30 min (18.9 (4.60)) and 45 min (21.6 (4.39))	TL (Min): Mean (SD)- 15 min (4.2 (2.04)), 30 min (9.2 (3.76)) and 45 min (10.0 (4.47)))
[57]	Age 40.3±13.5	Fifteen Type 1 diabetic patients, following multiple-dose insulin therapy and without significant micro- and macro-vascular complications, were monitored from 5 to 22 days (average 12.5±4.6) in free-living conditions.	10-fold cross- validation	RMSE (mg/dl): Mean ± SD - 30 min (6.1±1.6) $ESOD_{norm}$: Mean ± SD – 30 min (6.4±2.7)	TG (min): Mean ± SD – 30 min (7.7±3.7)
[58] Support vector regression	Age 40.3±13.6	Fifteen type 1 diabetic patients, following multiple dose insulin therapy and without significant micro- and macro-vascular complications, were monitored from 5 to 22 days (average 12.5 ± 4.6) in free-living conditions.	10-fold cross- validation	SVR—RF (RMSE (mg/dl): Mean ± standard – 30 min (5.7 ± 1.5), 60 min (6.4 ± 2.1)), SVR—RRF (RMSE (mg/dl): Mean ± standard – 30 min (5.9 ± 1.4), 60 min (6.8 ± 2.0))	N/A
[58] Gaussian processes	Age 40.3±13.6	Fifteen type 1 diabetic patients, following multiple dose insulin therapy and without significant micro- and macro-vascular complications, were monitored from 5 to 22 days (average 12.5 ± 4.6) in free-living conditions.	10-fold cross- validation	GP—RF (RMSE (mg/dl): Mean ± standard – 30 min (5.6 ± 1.7), 60 min (6.3 ± 2.6)), GP—RRF (RMSE (mg/dl): Mean ± standard – 30 min (5.9 ± 1.6), 60 min (6.8 ± 2.9))	N/A
[59]	Age 43.5±13.4, 12 female, 15 male	The dataset includes 27 type 1 diabetic patients following multiple-dose insulin therapy and was collected in the framework of the EU research project METABO from the participating clinical partners.	10-fold cross- validation	RMSE (mg/dl): 15 min (6.60), 30 min (8.15), 60 min (9.25) and 120 min (10.83)	CG-EGA: 15 min (Zone A (99.26), Zone B (0.62)),30 min (Zone A (98.23), Zone B (1.39)), 60 min (Zone A (97.59), Zone B (1.90)), and 120 min (Zone A (96.43), Zone B (2.75))
[60] & [61] & [62] & [63]	N/A	The data were collected from three type 1 diabetic patients over a period of 5, 11 and 13 days, respectively	V-fold cross validation	RMSE (mg/dl): 15 min (9.28), 30 min (15.59), 60 min (24.06) and 120 min (31.24)	CG-EGA: the vast majority of points belong to the zones A and B
	N/A	The data are collected from two type 1 diabetic patients over a period of 5 and 11 days, respectively.	Random subsampling: For the evaluation of the proposed method, the dataset of each patient is randomly split into training and test sets. The training set consists of the 66% of the data, while the remaining data are used for testing.	RMSE (mg/dl): 15 min (9.1), 30 min (14.8), 60 min (22.4) and 120 min (28.2)	N/A
	N/A	Seven patients with type 1 diabetes with an average observation period of 10 days (range from 5 – 14 days)	V-fold cross validation	RMSE (mg/dl): Mean ± SD - 15 min (9.51 ± 2.39)), 30 min (16.02 ± 3.55)), 60 min (24.81± 4.74)) and 120 min (36.15 ± 9.70))	CG-EGA: 15 min (Zone A (98.86 %), Zone B (1.08 %)),30 min (Zone A (92.54 %), Zone B (6.89 %)), 60 min (Zone A (80.02 %), Zone B (18.49 %)), and 120 min (Zone A (62.91 %), Zone B (33.78 %))
	12 female, 15 male, age 19-72 (43.5±13.4)	Data from 27 type 1 diabetics subject observation period of the study ranged from 5 to 22 days (average 13.42 ± 3.69).	Tenfold cross-validation	RMSE (mg/dl): 15 min (5.21), 30 min (6.03), 60 min (7.14) and 120 min (7.62) Mean Absolute Percentage Error: 15 min (2.06), 30 min (2.41), 60 min (2.79), 120 min (3.02)	CG-EGA: Accurate Reading - hypoglycaemia (15 min (96.76), 30 min (94.56), 60 min (92.16), 120 min (90.05)), euglycemia (15 min (96.56), 30 min (95.80), 60 min (94.93), 120 min (93.57)), hyperglycaemia (15 min(90.00), 30 min (89.28), 60 min (87.45), 120 min (84.85))
[64]	N/A	Clinical data from a group of 20 type 1 diabetes patients collected for a 30-day period	Hold-out: A subset of the first 3204 (11 days and 3 h) and subsequent 1200 (4 days and 4 h) samples were selected as training and validation data, respectively.	Overall Average RMS (mg/dl): One step prediction – 5.965	N/A
[65]	N/A	Data used here are 25sets of simulated blood glucose concentrations for 25 patients with various weights	Hold-out: 70% of data is used for training, 15% for validation and 15% for testing	NN with KF: MAD_{avg}% (60 min (29.10), 120 min (33.08))	N/A
[66]	9 males, 9 Females, Age 36.9±11.6 years	Patients monitored for approximately two weeks	Hold-out: The training data consisted of 6 days, while the validation set consisted of approximately 3 days.	N/A	N/A
[67] & [68]	17 to 70 years of age	Data collection time (days) –Training data (122 days), Evaluation set (111 days)	Hold-out: Half of the dataset for each patient was used in training and identification of the parameters of the prediction models, and the other half was used for evaluation of the model's performance.	RMSE (mg/dl): Median (5th–95th Percentiles) - 15 (11.9 (7.7–22.7)), 30 (18.9 (12.8–32.3)), and 45 min (26.1 (17.2–39.8)).	TL (Min): Median (5th–95th Percentiles) – 15 min (5.0 (0.5–10.0)), 30 min (10.0 (5.5–15.0)), 45 min (20.0 (10.0–25.0))

[69]	Female (80%), average age 42.3 (+/−11.07)	Ten complete days of data for each patient	Fold cross-validation	N/A	CG-EGA: – 30 min (Zone A (91.5%), + Zone B (7.0%)), 60 min (Zone A (75.5%), + Zone B (22.8%))
[70]	The population is characterized by 40% male and 60% female patients. The mean (SD) age is 42.9.	Real - Sets of daily vectors collected from 5 patients of the Hospital Clinic Universities of Barcelona for a period of 6 months.	Hold-out: 3-days series as the in-sample data for fitting the model. The models have been tested with the remainder data (out-of-sample data).	RMSE (mg/dl): 60 minutes (5.12)	CG-EGA: – 60 min (Zone A (61.98%), + Zone B (34.76%))
[71]	N/A	Virtual - Data obtained over 14 days for 100 virtual patients generated by the UVA/Padova T1D simulator	N/A	RMSE (mg/dl): 6-h prediction models (Nocturnal (11.80), Breakfast (22.09), Lunch (20.93), Dinner (29.00), Average-24 hr (20.96))	CG-EGA: – 6-h prediction models (Zone A +B (Nocturnal (99.37), Breakfast (98.68), Lunch (98.02), Dinner (97.16), Average-24 hr (98.31)),
[72]	8 women and 9 men (28 to 64 years (average, 41.2±13.36)	Seventeen type 1 diabetes patients were monitored for about 4 to 7 days (5.73±1.03 in average) in free-living conditions.	N/A	RMSE (mg/dl): Mean (SD)-30 min (2.37 (0.67)), 60 min (4.36 (3.86)), 90 min (22.23 (24.13))	N/A
[73]	23 years old	Collected data are selected for five normal days in the patient, under medical supervision, with standard ingestion (three meals per day and some snacks) and without exercise events.	N/A	N/A	N/A
[74]	Male	Data recorded for over a period of 63 days, which only had 463 blood glucose measurements is available in total which means that at 92% of the time steps, the blood glucose is unknown.	Hold-out: The first 42 days of the data set for training the models (containing 312 measurements of the blood glucose) and the following 21 days for testing (containing 151 blood glucose measurements)	One step ahead prediction: achieved the best prediction performance with an explained variance of 45.7%.	N/A
[75]	N/A	Data from two type 1 diabetic patients who were monitored over a period of 5 and 11 days, respectively	Leave-one-day out: the dataset of each patient is divided into two groups. The first group (i.e. test set) contains the data of i^{th} day, with i =1,…, k , where k is the total number of days and the second group contains the remaining data.	RMSE (mg/dl): Physical Activity Input – Sensor Data (15 min (11.13), 30 min (18.84), 60 min (28.79), 120 min (46.7)), Exercise Modelling (15 min (10.84), 30 min (17.92), 60 min (27.5), 120 min (43.34))	CG-EGA: - 60 min (Zones A (72.075%) and B (25.22%))
[76] & [77]	N/A	In silico patients 48-h dataset	Hold-out: The training dataset is formed by the 24-h records of five in silico patients. Testing dataset includes a different set of 24-h records for the same five in silico patients	Average Mean and standard deviation of percentage average error: (14.12 (2.11)).	CG-EGA: More than 95% of the data were found into Zone A
[78]	N/A	Dataset containing 400 timestamps collected from 10 T1D patients, 40 points from each.	Hold-out: Divide the datasets into 50/50%	RMSE (mg/dl): 30 min (21.4), 60 min (38.0)	N/A
[79]	Female patient, 23 years old	N/A	Hold-out	Mean value and standard deviation for the approximation error: 0.0039mg/dl (0.0209mg/dl)	N/A
[28]	7 male and 5 female (Age-19.83 ± 12.28)	Twelve Patients were monitored for a ten day period.	Hold-out: Data corresponding to the 60% of the monitored days were used for training purposes (model development), while the remaining 40% for testing (model evaluation)	RMSE (mg/dl): Mean ± standard deviation (SD) - 30 min (14.10 ± 4.57), 60 min (23.19 ± 6.40)	CG-EGA: Accurate readings – Hypoglycemia (30 min (81.06%), 60 min (63.22%)), Normal Glycaemia (30 min (92.18%), 60 min (91.71%)), Hyperglycemia (30 min (88.27%), 60 min (87.19%))
[80]	N/A	Twelve Patients were monitored for more than 14 days	Hold-out: 70% of his recorded blood glucose data for the training step and used the remaining 30% of glycemic data to test the predictions	Average RMSE (mg/dl): 15 min (9.44), 30 min (10.78), 45 min (11.82) and 60 min (12.95)	N/A
[81]	N/A	Twelve T1D patients (insulin-dependent diabetes) were using the Freestyle Libre System for 14 days. For each T1D patient, we have 1344 values of blood glucose. One more patient with three months long datasets.	Hold-out: The blood glucose measurements were divided into two parts for each patient: 2/3 of data (896 samples) for training step (70%) and 1/3 (448 samples) for testing step.	Average RMSE (mg/dl): 15 min (6.43), 30 min (7.45), 45 min (8.13) and 60 min (9.03)	CG-EGA: the majority of points are located in the zone A which verifies a clinically satisfactory result, while a minor quantity of points are in the other zones (B and D).

4.2. Evaluation and Analysis of the Literature

As defined earlier, the included literature studies were evaluated based on the data characteristics and types of inputs (diabetes parameters) used to train the algorithm, the class of machine learning techniques adopted for the task, and the performance metrics used to assess the prediction performance.

4.2.1. Data Characteristics and Input Parameters

4.2.1.1. Input Parameters

BG dynamics in people with diabetes is affected by various factors, such as the amount of insulin injection, the quantity of carbohydrate intake, the level and extent of physical activity, past readings of BG, stress (emotional feelings), any type of illness, alcohol consumption, smoking, menstruation, and others. An ideal BG predictor should incorporate all of these confounding variables for better estimation of the individual BG levels. According to the reviewed literature, BG, insulin, and diet comprise the most used group of parameters (29%), as shown in Figure 6. BG, insulin, diet, and physical activity make up the second most used group of parameters (25%). The use of only BG parameter ranked third (20%). The use of BG, insulin, diet, and physical activity, stress and, other groups of parameters ranked as the fourth most used (14%). The use of BG along with insulin and BG along with physical activity ranked equally as the fifth most used group of parameters (4%). The use of BG with diet and BG with insulin and physical activity ranked equally as the sixth most used parameters (2%).

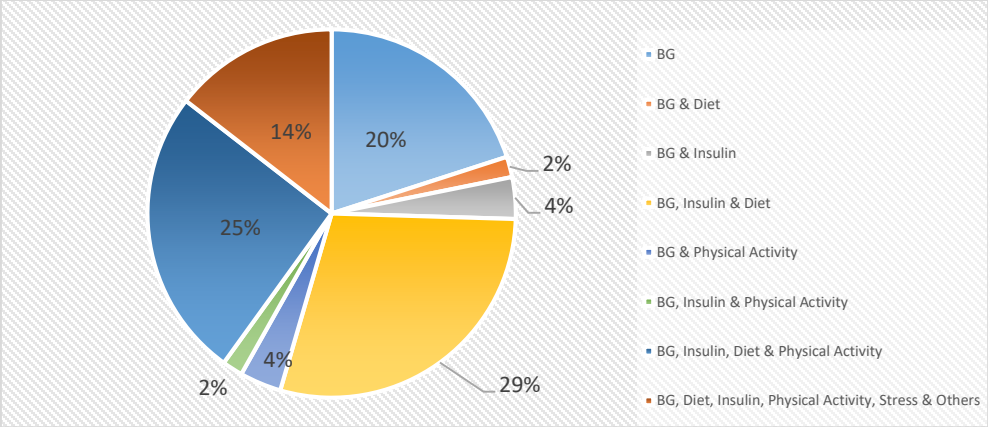

Figure 6: The type and number of input parameters used to train the models.

4.2.1.2. Data Characteristics

Data Sources

Different types of data sources have been exploited in the reviewed articles for recording various parameters for the prediction of BG, ranging from continuous glucose monitoring systems, insulin pumps, physical activity trackers, simulators for BG dynamics, and diaries to record daily events related to diet and insulin injections. The reviewed studies have used different brands of CGM systems to collect continuous subcutaneous concentrations of BG in the range of different minuets (i.e., 1–15 minutes). The most used CGM system is Guardian Real Time (MinMed, CGM), as shown in Table 3 below. The second most used CGM system is Dexcom SEVEN PLUS (CGM) (13.5%) and Abbott Freestyle (CGM) (13.5%). The third most used CGM system is FreeStyle Libre CGM and other unspecified brands (5.5%).

Table 3: Brands of CGM devices used in the studies.

CGM	Count	Percentage
Dexcom SEVEN PLUS (CGM)	5	13.5%
Guardian Real Time (MinMed, CGM)	23	62%
Abbott Freestyle (CGM)	5	13.5%
FreeStyle Libre CGM	2	5.5%
Other CGMs	2	5.5%

Regarding insulin pumps, different brands have been used. For example, the Paradigm Veo – 754 insulin pump is the most used brand (56%), as shown in Table 4 below. The second most used brands are Mylife Omnipod insulin pumps and AccuChek insulin pump (11%). Other non-specified brands of insulin pumps have also been used (22%).

Table 4: Brands of insulin pumps used in the studies.

Insulin Pump	Count	Percentage
Paradigm Veo - 754 insulin pumps	5	56%
Mylife Omnipod insulin pumps	1	11%
AccuChek insulin pump	1	11%
Other Insulin Pumps	2	22%

Physical activity is one of the key factors having a significant effect on BG dynamics. To this end, some studies have tried to measure the amount of physical activity using wearable fitness trackers, such as SenseWear Armband (89%), as shown in Table 5 below. It collects data using five sensors: heat flux, skin temperature, near body temperature, galvanic skin response, and a two-axis accelerometer. Some other studies also have used Fitbit Charge HR devices (11%) to record the amount and duration of physical activities.

Table 5: Brands of physical activity trackers used in the studies.

Physical Activity	Count	Percentage

Fitbit Charge HR™ devices	1	11%
SenseWear Armband	8	89%

Some studies have used either electronic or paper diary to record daily events such as meal and insulin injection. Most of the studies used a paper diary (50%), where the patient is expected to record his/her daily carbohydrate intake and insulin injection, as shown in Table 6 below. Some other studies have used a smartphone diary (22%) and other unspecified forms of diaries (28%).

Table 6: Types of diary used in the studies.

Diary	Count	Percentage
Paper Diary	9	50%
Smartphone Diary	4	22%
Others	5	28%

Some other studies have used a virtual patient's (in silico patient) data simulated from different tpes of BG simulator software. Among them, AIDA is the most used (50%), followed by UVA/Padova (30%), as shown in Table 7 below. The third most used software types are Glucosim and other non-specified simulators (10%).

Table 7: Types of blood glucose simulator used in the studies

Simulator	Count	Percentage
UVA/Padova	3	30%
AIDA	5	50%
Glucosim	1	10%
Other simulators	1	10%

Data Preprocessing

It is obvious that all machine learning algorithms tend to learn much quicker and effectively when the inputs and targets are preprocessed before training the network. For example, normalizing the input between 0 and 1 helps to keep the network weights from becoming too large [82]. In this regard, various data preprocessing techniques have been used in the reviewed articles, such as a smoothing using low pass filter [18], normalization [22], scaling [23], the interpolation of missing values [44, 78], and the differentiation of BG values. Others, featuring extraction and reduction [29, 69], include using the fuzzy approximation of food and exercise data [49], ranking candidate feature sets using random forest and RReliefF algorithms [58], and noise filtering using a Kalman filter [44, 65].

4.2.2. Class of machine learning

Various classes of machine learning techniques have been used in general dynamic system modeling, regression, and prediction services. However, for BG prediction, feed-forward neural networks are the most used techniques (20%), as shown in Figure 7. The hybridization of the physiology-based model and machine learning techniques is the second most used approach (19%). Recurrent neural networks in various forms ranked as the third most used technique (18%). Support vector machines (SVMs) ranked as the fourth most used technique (11%). Genetic programming techniques, most notably grammatical evolution, ranked as the fifth most used technique (6%). Autoregressive neural networks and neuro-fuzzy networks are the sixth most used techniques (5%). Self-organizing maps (SOMs) ranked seventh (4%). Extreme learning machines, kernel functions, Gaussian processes, genetic algorithms, and random forests ranked eighth (2%). Jump neural networks and deep neural networks ranked ninth (1%).

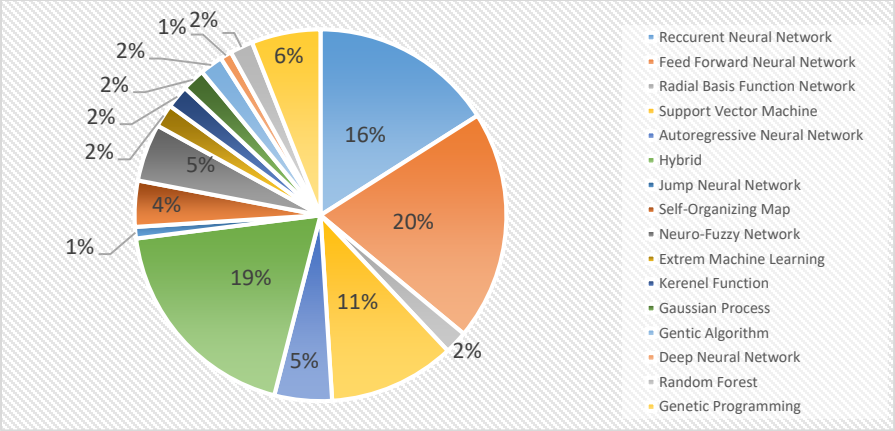

Figure 7: Classes of machine learning techniques used in the modeling of blood glucose prediction.

4.2.3. Performance Metrics

Performance metrics are necessary steps that should be carefully chosen based on the developed model under consideration. Various performance metrics are used to assess the predictive power of the developed model. However, choosing the appropriate metrics depends on the type of application the model is intended to use. Based on the reviewed articles, the performance metrics used to assess the predictive performance of the final model can be categorized into two groups: mathematical evaluation criteria (empirical accuracy) and clinical evaluation criteria (clinical accuracy). The mathematical evaluation criteria (empirical accuracy) are simply used to evaluate the numerical accuracy without giving due consideration to the clinical significance. This group of metrics includes root mean square error, correlation coefficient, FIT, normalized prediction error (NPE), and geometric mean. The clinical evaluation criteria (clinical accuracy), however, give due consideration to their significance in terms of clinical usability and include error grid analysis, average time gain, mean absolute relative difference, expert comparison, and J index. Generally, the most popular performance metric is root mean square error (36%), followed by Clarke error grid analysis (19%), as shown in Figure 8. The third most popular metric is correlation coefficient (12%), followed by temporal gain (8%). The fifth most popular metric is mean absolute error and mean absolute difference percent (5%). The sixth most used metrics are mean absolute relative difference, energy of the second order difference, and mean squared error (3%). The seventh most used metrics are normalized prediction error, expert reference, and J index (2%).

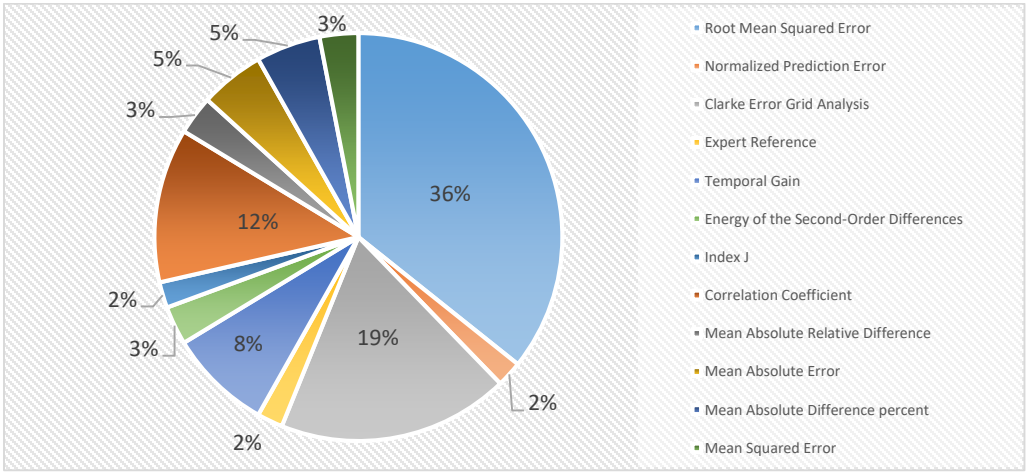

Figure 8: The performance metrics used to assess the predictive power of the developed models.

4.2.4. Prediction horizon (PH)

Prediction horizon (PH) is the lead time in which the model is able to forecast BG levels in the future. A wide range of prediction horizons have been reported from 15 minutes up to 2 hours and more. A satisfactory accuracy level is achieved for 15- and 30-minute PH with subsequent degradation afterwards. It is obvious that due to the limited number of available confounding factors in the data used to train the model, it is natural to expect degradation in prediction power as the prediction horizon increases. However, an increase in temporal gain (TG) enhances the clinical usability of the prediction services by prolonging the time necessary to take the required action during a critical situation but at the expense of clinical accuracy. Therefore, it is necessary to compromise on these complementary issues to achieve a reasonable predictive power and prediction horizon.

4.3. Strategies to Integrate the Effect of Diet, Physical Activity, Stress, and Infections

BG is affected by many factors, and its dynamics resemble a complex nonlinear process. However, the prominent parameters include the insulin medication, physical activity (exercise), the type of diet, the illness, and stress. A successful BG predictor is at least expected to integrate the continuous effect of these factors on BG dynamics. In this regard, integrating these effects has been attempted, however, with certain limitations. The proposed approaches to integrating the effects of diet, physical activity (exercise), illness, and stress into the BG predictor are described in the following section.

4.3.1. Proposed Approach to integrating the effect of diet (meals)

Diet has a tremendous effect on the level of BG. A certain diet could consist of various nutritional contents, such as carbohydrates, proteins, fiber, fat, and others. For people with diabetes, carbohydrates could be regarded as the most widely adopted means of meals planning. However, many factors affect the impact of carbohydrates on BG levels, (e.g., the type of carbohydrate and glycemic index). The carbohydrate glycemic index determines the rate at which it is digested, absorbed, and metabolized, thereby defining the rate at which it affects BG levels. According to the reviewed literature. as shown in Table 8 below, diet information, such as the number and time of meals, is mostly estimated in carbohydrate (grams) and manually recorded by patients using either a paper diary or an electronic diary. However, several researchers, including Zecchin et al. [25], have attempted to automate the procedure of recording diet information. Zecchin and colleagues [25] exploited information regarding the amount of ingested carbohydrates using a meal picture taken by a standard mobile phone. Generally, the integrations to the machine learning models are performed using directly the carbohydrates (grams), calculating the calorific values of meals, or relying on a compartmental model to estimate glucose absorption into the blood from the gut.

Table 8: Proposed approach to integrating estimation of food intake (Real subjects)

References	Proposed Approach
[19]	Carbohydrate consumption estimated in grams (g).
[20]	Carbohydrate consumption estimated in grams (g).
[25] & [15]	Patients used a mobile phone to take a meal picture, which was used to extract information about ingested carbohydrate (g). A physiological model was exploited to suitably preprocess information regarding the timing and amount of ingested carbohydrate within the meal.
[28]	Carbohydrate consumption estimated in grams (g). Compartmental models (CMs) was used to estimate glucose absorption into the blood from the gut.
[29]	Patients self-recorded information regarding the amount of ingested carbohydrates (in grams).
[33] & [34]	Patients received standardized meals (breakfast, lunch and dinner), where the amount of carbohydrates included in each meal was about 40 (45 in DAQ), 70 and 70 g, respectively.
[38] & [39]	A diet vector was formed containing total carbohydrates (g) and time of meal. A chart devised by the University of Glasgow Dietetics' Department was used to calculate and subdivide the calorific values of meals into energy, total carbohydrate, sugar carbohydrate, protein, fat, fiber and sodium.
[42]	Carbohydrate consumption estimated in grams (g).
[43]	Carbohydrate consumption measured in grams (g).
[45]	Pocket PC-based electronic diary was used to record nutritional content (patient self-recorded information).
[46] & [47]	An electronic diary documenting carbohydrate intake (patient self-recorded information) was used to record meal information.
[49]	Four types of food that correspond to the description of food units (compendium of food measures with associated carbohydrate composition) provided by the Canadian Diabetes Association were used. A fuzzy approximation was used to estimate food content (simple sugars (in grams), fruits (in grams), milk products (in grams), starch (in grams)).
[50]	Amount of carbohydrate ingested was estimated and a compartmental model was used to capture its effect on BG dynamics.
([51]	The effect of carbohydrate intake on BG absorption from the gut was estimated using compartmental model.
[53]	Patient self-recorded information - Patient dairy was used to record carbohydrates (in grams).
[56]	Food intake considered in terms of carbohydrates ingested.
[57]	Patient self-recorded information – a paper diary was used to record food intake information (type of food, serving sizes and time). The amount of carbohydrate in each meal was post-analyzed by a dietician. A compartmental model was used to estimate rate of appearance (R_a) of meal-derived glucose into plasma within a time interval.
[58]	Patient self-recorded information – a paper diary was used to record information on food intake (type of food, serving sizes and time). The amount of carbohydrate in each meal was post-analyzed by a dietician. A compartmental model was used to estimate rate of appearance (R_a) of meal-derived glucose into plasma within a time interval.
[59]	Patient self-recorded information – a paper diary was used to record daily food intake information (type, amount and time). The specific amount of carbohydrate in each meal was post-analyzed by a dietician. *Lehmann's* [83] compartmental model was used to estimate rate of appearance of exogenous (meal-derived) glucose in plasma (R_a), in which the rate of gastric emptying is a function of meal carbohydrate content.
[60] & [61] & [62] & [63]	Patient self-recorded information – a paper diary was used to record information on food intake (type of food, serving sizes and time). A dietician analyzed the food composition (i.e. calories, carbohydrates, fat etc.) and a compartmental model was used to estimate ingestion and absorption of carbohydrates (Meal Model).
[64]	The software described in Percival [84] was used to calculate the carbohydrate (CHO) intake of each patient (g CHO).
[66]	Patient self-recorded information – Participants used a diary to record a detailed food intake in terms of carbohydrate ingested. *Lehmann's* compartmental model [83] was used to estimate the amount of glucose entering the bloodstream via the guts.
[69]	The amount of carbohydrate was estimated by the patients. A Gaussian process enhanced with compartmental model was used to create smoothed time series of the original carbohydrate and insulin records.
[70]	The carbohydrate intake was estimated by patients and measured in grams (g).
[73]	Carbohydrate consumption estimated in grams (g). *Lehmann's* compartmental model [83] was employed to approximate continuous glucose absorption by gut, after carbohydrate oral consumption.
[74]	The amounts and times of food intake (fast, intermediate, and slow-carbohydrates) and compartmental model was used to estimate its effects on BG dynamics.
[75]	Patients manually notified food ingested (serving sizes, and time of each meal or snack), and insulin injections (type, dose and time). A dietician analyzed the food composition (calories, carbohydrates, fat etc.). *Lehmann's* [83] compartmental model was used to describe time course of glucose appearance in plasma after meal intake. The amount of glucose in the gut, q_{gut}, after ingestion of a meal containing D grams of glucose-equivalent carbohydrates was defined.
[76]	The amount of carbohydrates units ingested.
[78]	The amount of carbohydrates units ingested. A physiological model from *duke's PhD thesis* [85] was used to describe the BG dynamics upon food intake
[79]	The patients provide amount of ingested carbohydrates (in grams) and course of time. *Lehmann's* compartmental model [83] was employed to approximate glucose absorption by intestine after oral ingestion.
[28]	Meal information (amount of carbohydrates ingested and time of meal). A compartmental model was used to estimate glucose absorption from the gut. The model of glucose intestinal absorption was represented by a three-compartment nonlinear model - the stomach (solid and liquid phases) and intestine. The model assumed a constant rate of intestinal absorption but described gastric emptying rate to be dependent on total amount of nutrient in the stomach.

As a substitute for actual meal data, simulated meal information has been generated and used to develop and test the predictor algorithm, as shown in Table 9 below. This simulation considers a meal scenario representing the major meals of breakfast, lunch, dinner, and snacks, taking into account the different amounts of carbohydrate and meal time. The simulation is carried out by either varying the amount of carbohydrates for each meal or by varying the time

of meals and sometimes considering both. The variations are generated based on a preset range of values to randomly select from or using a random variable drawn from a uniform distribution.

Table 9: Proposed approach to integrating estimation of food intake (Virtual subjects)

References	Proposed Approach
[22]	A scenario representing the major meals were simulated: breakfast, lunch, dinner, and snack. The daily variations in quantity and timing regarding each meal was handled through a randomized approach: each kind of meal was assigned a range of possible quantities of grams of CHO and timings. For each day and meal, the approximate number of grams of CHO and the timing was chosen randomly from the respective ranges.
[23] & [24]	A simulation scenario consisted of the three major meals per day in a range of time intervals, Breakfast (45 + *u* g), Lunch (75 + *u* g) and dinner (85 + *u* g), where *u* was defined as is a random variable drawn from a uniform distribution.
[33]	A simulation scenario consisted of the three major meals per day with a range of amount of carbohydrates (g) and fixed timing. A random amount of carbohydrate (g) was selected from the specified range: Breakfast (08:00, 45±5 g), Lunch (12:30, 70±10 g), Dinner (19:00, 80±10 g).
[40]	A simulation scenario consisted of 2 meals (carbohydrate (in grams)) per day and random timings 13:30 (Meal 1) and 19:00 (Meal 2) ±30 minutes.
[41]	A simulation scenario consisted of a meal measured in carbohydrate intakes (in grams).
[71]	A simulation scenario consisted of the major three meals (carbohydrate (in grams)) per day with average intakes: breakfast (50 g), lunch (60 g), dinner (63.5 g), and a coefficient of variation (CV) of 20%, sampled using a Gaussian distribution. The information of carbohydrate intake also includes 20-g hypo treatments that were generated every 20 min when the glucose concentration fell below 60 mg/dl, as indicated by the SMBG measurements. Transformation of each meal events into a continuous signal that describes the absorption of the carbohydrate was carried out using a physiological model.
[76]	The simulation scenario was carried out by varying the parameters of the simulator to account for realistic situations, varying carbohydrates and/or insulin units to represent realistic situations like bigger or smaller meals and little changes in the insulin doses.

4.3.2. Proposed approach to integrating the effect of physical activity (exercise)

Physical activity (exercise) has a significant effect on BG dynamics [86]. However, the effect of physical activity (exercise) on BG varies considerably based on many factors, such as the type of activity, amount and intensity of activity, and duration. According to the reviewed literature, as shown in Table 10 below, physical activity information has been recorded either manually using paper and electronic diaries or automatically using wearables, such as SenseWear Armband and Fitbit Charge HR devices. A wide variety of physical activity (exercise) information has been considered, such as the degree of activity, the sum of energy expenditure during a range of time intervals**,** standard tables for caloric use during exercise, the metabolic equivalent of task (MET), the heat flux (h_f), and skin temperature (S_t) variables. The integration to the machine learning model is carried out by relying on four main approaches: using the scale and time of activity, directly feeding the recorded time and duration of the activity, directly feeding the data from the wearables, and using an exercise compartmental model to estimate the effect on BG levels.

Table 10: Proposed approach to integrating physical activity/exercise (Real subjects)

References	Proposed Approach
[19]	Patient self-recorded exercise information using a scale from 1 to 5 was used to express the degree of physical exercise, where 1 means at rest and 5 means doing heavy physical exercise.
[26]	Energy expenditure as a result of daily physical activity or exercise events was recorded with a resolution time of 1 min using SenseWear Armband. The integration was carried out considering the sum of energy expenditure during a time period [t—150 min, t—120 min], which takes into account physical activity during the latest 30 min with a lag time equal to 120 min.
[27]	Energy expenditure as a result of daily physical activity or exercise events was recorded with a resolution time of 1 min using SenseWear Armband.
[29]	Patient self-recorded exercise information, where the integration was carried out using a scale of 1 to 5 to describe the intensity and duration. In this regard, a value of 1 indicates no exercise at all, and a value of 5 means heavy physical exercise based on interval of time.
[38] & [39]	Patient self-recorded physical exercise based on duration (mins), mobility (0-3), strength (0-3), endurance (0-3) and time of exercise (absolute), where mobility, strength, and endurance are expressed using a scale from 0 to 3. An exercise vector containing all these parameters was used to create a serious of events, which was fed to the model.
[42]	Patient self-recorded exercise information using a scale from 1 to 4 was used to express the magnitude and duration, where 1 means doing nothing and 4 signifies heavy exercise.
[45]	Patient self-recorded exercise information using a Pocket PC-based electronic diary to record lifestyle/activities.
[46] & [47]	Patient self-recorded exercise information using an electronic diary to record lifestyle (activities and events)
[49]	The intensity of physical exercise was considered based on three possible levels (moderate (baseball, walking), hard (hockey, basketball), strenuous (jogging, rowing)) and the patient was excepted to perform a fuzzy approximation based on these three levels.
[53]	Standard tables for caloric use was used to properly use the physical exercise data from the patients.
[56]	Physical activity data was recorded using Fitbit Charge HR devices
[57]	SenseWear Armband was used to compute and record energy expenditure. The integration was carried out considering the energy expenditure calculated cumulatively every 10 min over the last three hours based on the instantaneous (i.e. per minute) energy expenditure estimated by the SenseWear Armband.
[58]	SenseWear Armband was used to compute and record energy expenditure every 1 min. The integration was carried out considering the energy expenditure calculated cumulatively every 10 min over the last three hours based on the instantaneous (i.e. per minute) energy expenditure estimated by the SenseWear Armband.
[59]	SenseWear Armband was used to compute and record energy expenditure every 1 min. The integration was carried out based on cumulative energy expenditure (SEE) to consider the short-term effects of physical activities and exercise on glucose variability by expressing the energy expenditure over the last 3 hrs.in the form of a vector calculated cumulatively every 10 min.
[60] & [61] & [62] & [63]	SenseWear Armband was used to compute and record energy expenditure every 1 min using the five sensors: heat flux, skin temperature, near body temperature, galvanic skin response and a two-axis accelerometer. The integration was carried out in two different approaches: using the SensWear armband data directly or using the compartmental models to describe the effect of physical activities on glucose-insulin metabolism (Exercise Model): ❖ The first approach utilizes, the Metabolic Equivalent of Task (MET), heat flux (h_f) and skin temperature (S_t) variables, as recorded by the SenseWear armband, as inputs to the model. ❖ The second approach utilizes the output (G_{exer}, I_e)from the exercise compartmental models.
[64]	The amount and degree of physical exercise was computed based on an assumption of three level code: low (50–150 cal/30 min), moderate (150–200 cal/30 min) and strong (>200 cal/30 min). The integration was carried out using a mathematical interpolation function that considers the variation in physical exercise on a range of 0–2 (Exc. level: 0 (low), 1 (moderate), 2 (strong)), and the rates of carbohydrate and calorie consumption as described in literatures [86, 87].
[66]	SenseWear armband was used to compute and record physical activity using the 5 different sensors; transversal acceleration (measure of movement), longitudinal acceleration (measure of movement), heat flux (average heat dissipated or absorbed by the arm), galvanic skin response (electrical conductivity between two points on the arm), and skin and near-body temperature. A proprietary algorithm was employed used to estimate the physical activity energy expenditure by combining these different physiological signals.
[88]	The times and durations of physical exercise (regular, or intense)
[75]	SenseWear armband was used to compute and record physical activity using the 5 different sensors; heat flux, skin temperature, near body temperature, galvanic skin response and a two-axis accelerometer. The integration was carried out using two different approaches to investigate the dynamic effect of exercise on glucose variation: ❖ In the first approach, the Metabolic Equivalent of Task (MET), the heat flux (h_f) and the skin temperature (S_t) variables, as recorded by the SenseWear armband, are used directly as inputs to the model. ❖ However, in the second approach the output from the exercise compartmental models was used. The study reveals an important observation: using directly the real sensor data or the output from the exercise compartmental models, produce a relatively small difference in performance (RMSE).

4.3.3. Proposed Approach to integrating the effect of stress and Infections

Emotional stress or physiological stress caused by illness and infections could affect BG homeostasis due to hormonal changes in the body. In this regard, recent findings have indicated that emotional stress and infection have a strong impact on glycemic control, resulting in elevated BG (hyperglycemia) [89-92]. Only a few researchers have tried to integrate the effect of stress and illness, as shown in Table 11 below, using a scale to indicate the level of the stress

and binary values to indicate the absence and presence of illness. However, these studies have limitations in terms of quantifying the approximate effect of stress and illness in response to BG levels.

Table 11: Proposal to integrating presence of stress and illness (Real subjects)

References	Proposed Approach	
	Stress	**Illness and pregnancy**
[19]	A scale of one up to five was used to define the presence and magnitude of stress: one means relaxing and five means heavy stress.	N/A
[29]	Patient self-recorded stress information on a scale of one to five based on interval of time: value of one indicates very relaxing, and the value of five means heavy stress	N/A
[38] & [39]	A scale of zero to one was used to indicate the absence and presence of a stress: zero indicates absence and one indicates presence of stress.	A scale of zero to one is used to indicate the absence and presence of illness, surgery, and pregnancy: zero indicates absence and one indicates presence of stress.

4.4. A Comparative Assessment of Prediction Performances

The predictive performance of a BG predictor is affected by many technical factors and design choices, apart from the factors that affect BG dynamics, such as the type of machine learning, data size, prediction horizon (PH) and validation approaches considered, and others. We have used the most popular performance metrics, RMSE, to compare the reported performance levels by different researchers. According to the reviewed literature, different classes of machine learning and types of validation approaches have been tested and confirmed as being up to the task, as shown in Table 12 below. Please note that these performances are achieved based on real subject data.

Table 12: Reported predictive performance (RMSE) considering the class of machine learning employed, validation strategies and prediction horizon considered (Real Subjects).

Reference	Machine Learning	Validation	RMSE (mg/dl)(Mean/ Mean ± SD)				
			15 min	30 min	45 min	60 min	120 min
[17]	Recurrent Neural Network (RNN)	Random subsampling	**2.52**	**7.56**	**15.12**	**23.76**	
[18]	Feed Forward Neural Network (FFNN)	Random subsampling	**2.70**	**7.56**	**14.94**		
[23] & [24]	Hybrid (Feed Forward Neural Network Plus Linear Prediction Algorithm) along with Physiological Model	k-fold cross-validation		**9.4±1.5**			
[25] & [15]	Hybrid (Jump Neural Network along with Physiological Model)	k-fold cross-validation		**16.6 ± 3.1**			
[26]	Feed Forward Neural Network (FFNN)	10-fold cross-validation		**13.31 ± 4.47**		**22.66 ± 6.86**	**37.62 ± 11.79**
	Self-organizing map (SOM)			**11.42 ± 2.33**		**19.58 ± 3.80**	**31.00 ± 6.07**
	A Neuro-Fuzzy Network with Wavelets as Activation Functions (WFNN)			**15.22 ± 2.17**		**24.66 ± 3.39**	**39.59 ± 5.03**
[28]	Hybrid (Compartmental model (MM)) and Self-Organizing Map (SOM) - Vector Quantization Method)	Hold-out		**14.10 ± 4.57**		**23.19 ± 6.40**	
[31]	Ensemble Approach Hybrid-Fusion (AR, Extreme Learning Machine, and Support Vector Regression-Kernel Function (Gaussian))	Hold-out		**19.0 ± 0.3**			
[36] & [37]	Feature based Feed-Forward Neural Network (FFNN)	Random subsampling		**10.00**	**15.00**	**20.00**	
[43]	Hybrid (Generic Physiological Model & Support Vector Regression- Gaussian kernel)	Random subsampling		**22.6**		**35.8**	
[44]	Feed-Forward Neural Network model using Guardian CGM	Hold-out	**9.74±2.71**	**17.45±5.44**	**25.08±8.73**		
	Feed-Forward Neural Network Model using FreeStyle Navigator CGM		**10.38±3.15**	**19.51±5.53**	**29.07±6.77**		
[56]	Hybrid-Fused (ARX and Elman simple Recurrent Neural Network) for prediction and Extreme Learning Machine for correction	Hold-out	**8.9 ±1.70**	**18.9± 4.60**	**21.6 ± 4.39**		
[57]	Single hidden layer Feedforward Neural Networks - (kernel RLS, Gaussian kernel) - Extreme Learning Machine	10-fold cross- validation		**6.1±1.6**			
[58]	Support Vector Regression (SVR—RF)	10-fold cross- validation		**5.7 ± 1.5**		**6.4 ± 2.1**	
	Support Vector Regression (SVR—RRF)			**5.9 ± 1.4**		**6.8 ± 2.0**	
[58]	Gaussian Processes (GP—RF)	10-fold cross- validation		**5.6 ± 1.7**		**6.3 ± 2.6**	
	Gaussian Processes (GP—RRF)			**5.9 ± 1.6**		**6.8 ± 2.9**	
[59]	Hybrid- (Random Forests Regression technique & Compartmental Model)	10-fold cross- validation	**6.60**	**8.15**		**9.25**	**10.83**
[60] & [61] & [62] & [63]	Hybrid- (Support Vector Machines- Linear kernel and Compartmental Model)	v-fold cross validation	**9.28**	**15.59**		**24.06**	**31.24**
	Hybrid- (Support Vector Machines- Linear kernel and Compartmental Model)	Random subsampling	**9.1**	**14.8**		**22.4**	**28.2**

	Hybrid- (Support Vector Machines- Linear kernel and Compartmental Model)	v-fold cross validation	**9.51 ± 2.39**	**16.02 ± 3.55**		**24.81± 4.74**	**36.15 ± 9.70**
	Hybrid- (Support Vector Machines- Linear kernel and Compartmental Model)	10-fold cross-validation	**5.21**	**6.03**		**7.14**	**7.62**
[67] & [68]	Hybrid-Fused (Autoregressive Model with output correction – cARX, & Recurrent Neural Network – RNN)-Data fusion (Genetic Algorithms (GA), & Genetic Programming (GP))	Hold-out	**11.9**	**18.9**	**26.1**		
[70] & [71]	Hybrid (Genetic Programming - Grammatical Evolution & Physiological model)	Hold-out				**5.12**	
[75]	Hybrid-(Compartmental model & Support Vector Regression- linear kernel) using Physical Activity – Sensor Data as Input	Leave-one-day out	**11.13**	**18.84**		**28.79**	**46.7**
	Hybrid-(Compartmental model & Support Vector Regression- linear kernel) using Exercise Modelling		**10.84**	**17.92**		**27.5**	**43.34**
[78]	Recurrent Neural Network (RNN)	Hold-out		**21.4**		**38.0**	
[28]	Hybrid- (Compartmental Models (CMs) and a Self-Organizing Map (SOM) - Vector Quantization Method)	Hold-out		**14.10 ± 4.57**		**23.19 ± 6.40**	
[80]	Support Vector Regression (SVR) based on Differential Evolution (DE) Algorithms	Hold-out	**9.44**	**10.78**	**11.82**	**12.95**	
[81]	Feed Forward Artificial Neural Networks (FFNN)	Hold-out	**6.43**	**7.45**	**8.13**	**9.03**	

4.4.1. Validation Strategies

The purpose of training any machine is to be able to get better predictions of the testing values by enabling the machine to generalize from the training examples of all possible inputs. However, the complementary issues of overfitting and under training are challenging, given the large degree of variability in most machine learning algorithms. The training algorithm should be able to avoid overfitting with enough generalization power by relying on a third set of datasets known as validation sets [82]. The main purpose of cross-validation is to evaluate the generalization power of the algorithm and to compare and find the best algorithm among a set of multiple algorithms [93]. There should be enough data size in the three datasets (training, testing, and validation sets) so that the machine does not overfit or undertrain. The proportion of splitting these datasets into these three groups is up to the expert and is dependent on the model selection [82]. Various strategies have been proposed for cross-validation, including re-substitution validation, hold-out validation, k-fold cross-validation, leave one out cross-validation, and repeated k-fold cross-validation [93]. In this regard, the most popular strategies used in the reviewed articles are various forms of the k-fold cross-validation [25, 26], and hold-out [18, 19]. K-fold cross-validation strategy involves splitting the datasets into randomly partitioned k equal subsets and using one set as a validation sets and the rest for training, repeating the same process for all the different subsets [82]. However, hold-out involves portioning the datasets into non-overlapping subsets, where the first subset is entirely used for training and the rest for testing [93]. In this regard, among the most widely used validation approaches (k-fold cross-validation, random sampling, and hold-out), k-fold cross-validation achieves the best performance for all prediction horizons except for a prediction horizon of 15 min, where random subsampling achieves better performance levels, as shown in Figure 9 below.

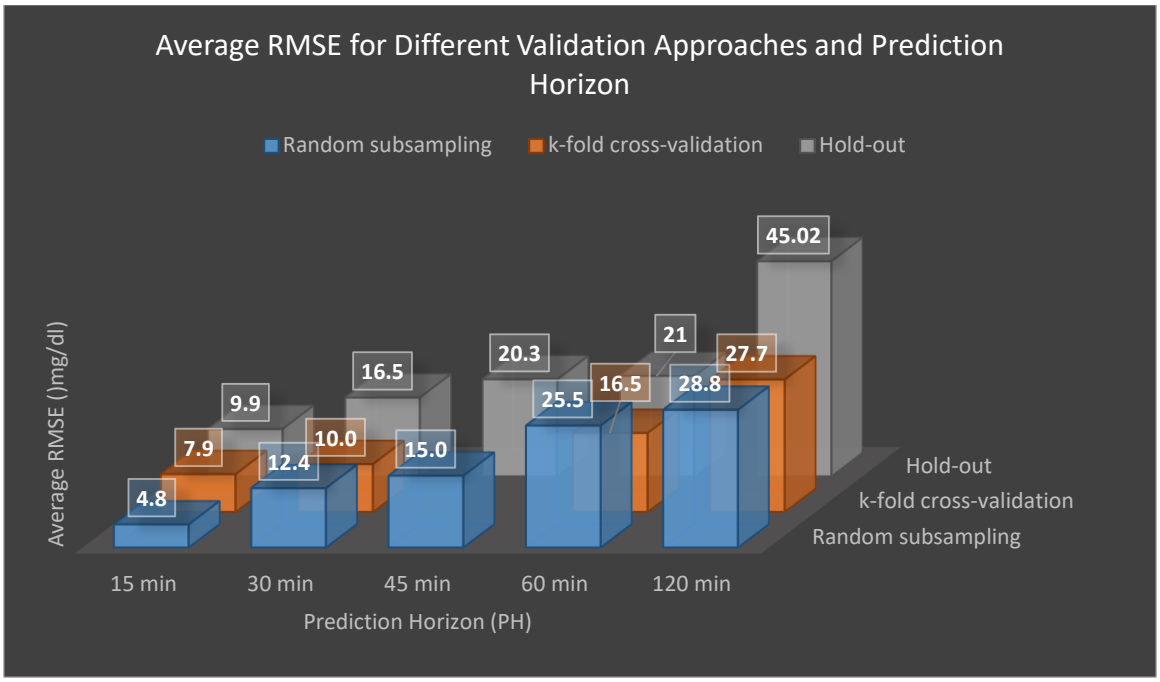

Figure 9: A comparative assessment of the predictive performance based on the adopted validation strategies and prediction horizon (PH) using data of real subjects. The average RMSE for each prediction horizon is computed by averaging the performance levels reported by different authors, taking into account the proposed validation approaches.

5. Discussion

5.1. Principal findings

Recently, machine learning has received wider attentions for modeling and the prediction of BG dynamics. Regardless of its popularity, no recent reviewers have analyzed and reflected on the current development. As far as our knowledge is concerned, no review has focused mainly on modeling and the prediction of BG dynamics using machine learning techniques. Therefore, the purpose of this review is to assess and analyse the state-of-the-art machine learning applications in BG prediction pertinent to the controller of an artificial pancreas (a closed-loop system), modeling of a personalized profile, personalized decision support systems, and BG alarm event applications. It serves as a compact guide regarding modeling options and strategies of machine learning applications and the hybrid system in type 1 diabetes. Some recent reviews have been conducted on diabetes in general and on BG prediction in particular [11-13]. For instance, Oviedo et al. [12] recently conducted a methodological review on BG prediction, focusing mainly on a closed-loop system (an artificial pancreas), which assessed a variety of approaches, including physiological models, data-driven models, and hybrid approaches. Kavakiotis et al. [11] also performed a systematic review of the applications of machine learning and data mining techniques in diabetes research in the context of diabetes prediction and diagnosis, diabetes complications, genetic background and environment, and healthcare and management.

BG dynamics are affected by numerous factors, such as the history of BG values, the insulin medication, physical activity, and dietary intake. Moreover, they are also affected by other factors, such as an individual's body mass index, stress level, amount of sleeping time, presence of illness, some medications, smoking habit, periods (menstruation), alcoholism, allergies, and altitude. In principle, a successful BG predictor is expected to incorporate as much information as possible to effectively track and predict BG levels. However, due to the complexity of BG dynamics, it remains difficult to achieve an accurate prediction in every circumstance. Most of the available BG prediction algorithms have their own limitations, working better in some specific circumstances. The reported BG prediction algorithms have explored various classes of machine learning, input parameters, and training algorithms. Most of these studies have neglected the effect of physical activity on BG dynamics, and only a few studies have considered the effect of uncontrollable patient parameters, such as stress, illness, and others. Generally, the reported BG prediction

algorithms can be categorized under different scenarios, such as real time (online) versus offline, the age group (children, adult, and old), BG regions (hypo/eu/hyperglycemia events), the time of day (diurnal vs. nocturnal), generalizability (generic vs. specific), free-living versus non-free-living conditions, and the evaluation approach (in vivo vs in silico). Accordingly, most researchers have considered separate age groups, which are typically related with the dynamics and active lifestyles adopted by each group. Few attempts have been made to develop real-time algorithms that perform under free-living conditions. Moreover, most of the reported algorithms perform better in either of these BG regions (hypo/eu/hyperglycemia events). Furthermore, most of the algorithms rely on in silico evaluations, which further put the clinical significance in question. Also, the lack of a well-defined approach to estimate carbohydrate intake is an issue; it is mainly done manually by the individual users and is prone to an error that can severely affect the predictive performance. The lack of a universal approach to estimate and quantify the approximate effect of physical activities, stress, and infection incidence on the BG level is another challenge. For instance, regarding physical activity integration, a wide variety of approaches have been proposed, such as using a scale and level code to quantify the degree and duration of physical activity, the sum of energy expenditures during an interval of time, physical activity, caloric use based on standard table, MET, exercise compartmental models, a proprietary algorithm to estimate the physical activity energy expenditure by combining different physiological signals such as transversal acceleration (a measure of movement), heat flux (the average heat dissipated or absorbed by the arm), longitudinal acceleration (measure of movement), skin and near-body temperature, and galvanic skin response (electrical conductivity between two points on the arm). It seems that almost all the studies have followed quite different approaches, and this poses a challenge in regarding one's approach as universal. In addition, also few studies have been done regarding model portability that can capture the inter- and intra-variation among the patients. It also seems that the effect of time lags between the CGM reading and the actual BG levels is not well covered. Generally, the review indicates the lack of a one-fits-all algorithm that performs better under totally free-living conditions.

Any successful BG prediction algorithm should at least consider both the patient's controllable (BG, insulin, diet, physical activity, and others) and uncontrollable parameters (stress, infections, medications, and others). Moreover, it is also necessary to consider any relevant contextual information, such as intra- and inter-variability among the patient's changes in lifestyles, the time of day (diurnal vs nocturnal) and others. Future scholars need to reflect on a longer prediction horizon (giving more response time) with reasonable clinical accuracy, approaches to improve time lags from the CGM, real-time capability under free living conditions, and a thorough validation using real patients (clinical trials) with ample subjects over a longer period of time. Moreover, the predictor should give proper weights and penalties for errors in hypoglycemia, euglycemia, and hyperglycemia regions. It is also necessary to consider a proper way to estimate the amount and effect of dietary consumption and physical activity during integration with the machine learning model. Stress and infections have a prominent effect on BG dynamics, which could in turn affect the predictor performances. To this end, it is necessary to test and assess the effects of a change in either lifestyle or physiology (infections) on the predictor performance using subjects who are monitored within different periods. Furthermore, the effect of different CGM devices on the quantitative performance of the prediction algorithms should be explored along with the associated time lag.

5.2. Summary of Existing Efforts (Machine Learning Techniques)

5.2.1. Artificial Neural Network (ANN)

An artificial neural network (ANN) is a computational model consisting of various processing elements known as neurons and a scaled connection between them called weights [94]. Various forms of artificial neural networks are used, but the network topology could be generally categorized as feed-forward networks (SLP, MLP, and radial basis function) and recurrent/feedback networks (Elman net, Kohonen's SOM, and Hopfield Networks). The feed- forward network is the most common topology, where it consists of a connection between different neurons that are directed only in one direction (forward) from the earlier stage to the next level. Recurrent or feedback network topology involves at least one feedback loop in the architecture [94]. Both of these network topologies have been successfully employed in modeling and for the prediction of BG levels in type 1 diabetes patients. Regarding the feed-forward network, for example, Allam et al. and others [18, 44, 45] have developed a feed-forward neural network from CGM data using the back propagation Levenberg-Marquardt optimization training algorithm. Pappada et al. and others [46, 47, 65] have also proposed time-lagged feed-forward neural networks trained through a back-propagation gradient descent algorithm, which is capable of storing previous values of data within the network. Zainuddin et al. [29] have proposed a wavelet neural network, integrating different wavelet families as an activation function for modeling BG dynamics trained through pseudo-inverse with fixed parameter initialization. Zarkogianni et al. [27] developed a

seven-layer neuro-fuzzy network using wavelets as an activation function and Gaussian function as a membership function trained through a gradient-based algorithm with an adaptive learning rate. Compared to these shallow networks, Mhaskar et al. [52] proposed a semi-supervised deep learning neural network with a judge predictor based on the function approximation on data-defined manifolds, using diffusion polynomials. Baghdadi et al. [19] implemented a radial basis function network using Gaussian function in the hidden layer neuron. Georga et al. [57] investigated the applicability of an extreme learning machine (ELM), specifically an online sequential ELM (OS-ELM) and online sequential ELM kernels (KOS-ELM) for training single hidden-layer feed-forward neural networks. In a surgical care setting, Pappada et al. [48] trained a feed-forward network from CGM data for bedside monitoring using a back-propagation training algorithm. Apart from the feed-forward network, recurrent or feedback networks have been utilized in BG prediction; that is, recurrent neural networks, autoregressive neural networks and self-organizing maps. For example, Daskalaki et al. [22] developed an online adaptive ANN-based model using a fully connected, multilayered ANN with two feedback loops trained through a teacher-forced, real-time, recurrent algorithm. Sandham et al. [38, 42] and Robertson et al. [40] have used an Elman recurrent network trained through a backpropagation gradient descent algorithm with momentum and an adaptive learning rate and Levenberg-Marquardt algorithm, respectively. Alanis et al. [73, 79] developed an autoregressive version of a neural network called neural network autoregressive external input (NNARX), which is trained through an extended Kalman filter (EKF) algorithm. Chernetsov et al. [21] performed a comparative analysis of three recurrent or feedback networks: the layer recurrent network (LRN), Elman net, and nonlinear autoregressive network (NARX-net). They investigated the effect of different learning algorithms, network architectures, prediction horizons, data sample sizes, and tapped delay line lengths on the performance of the network. Moreover, Zarkogianni et al. [26] conducted a comparative analysis of four machine learning techniques in the modeling of BG dynamics: a feed-forward neural network (FNN) trained through a backpropagation algorithm, a self-organizing map (SOM) achieved by applying a vector quantization method, a neuro-fuzzy network using wavelets as activation functions (WFNN), and a linear regression model (LRM). They used CGM data and also explored the effect of physical activity data collected from a SenseWear Armband. The study demonstrated the superiority of SOM and its ability to capture both the complexity of the dynamics and also the inter- and intra-variations among the patients [26].

5.2.2. Support Vector Machines (SVM), Kernel Function (KF), and Gaussian Process Regression

Support vector machines have been widely exploited across a wide range of applications, such as pattern identification and recognition, categorization or classification, regression, and prediction [95]. Support vector regression (SVR) is the most widely used class of SVMs in BG prediction and modeling. In this regard, for example, Reymann et al. [41] investigated the applicability of BG prediction using a mobile platform based on SVR, with radial basis function (RBF) as a kernel. Moreover, Li et al. [54] tried to use pooled patient data to capture patient similarities, which led to the development of a personalized BG prediction model using smartphone-collected data based on SVR. Georga et al. [58] investigated the potential performance enhancement from using a feature ranking algorithm, random forests (RF), and RReliefF algorithms, where the predictor is based on an SVR-exploiting Gaussian radial basis function (RBF) as a kernel.

As a solution to the artificial neural network requirement of larger training data and much more information to learn, Naumova et al. developed a novel fully adaptive regularized learning (FARL) approach using meta-learning to choose the kernels and regularization parameters in kernel-based regularization learning algorithms [35].

Gaussian process regression is a useful nonparametric regression tool that has been widely adopted in various applications, such as a vital-sign "early warning system," patient physiological monitoring, disease prediction, and the discovery of biomarkers in microarray gene expression data. Tomczak et al. [32] investigated the applicability of Gaussian process regression in BG prediction coping with categorical inputs. The input consisted of the data, time, code (categorical), and BG level (numeric). The categorical code was used to describe the type of measurement (e.g., insulin dose, meal intake, physical exercise, pre-prandial BG measurement, and others). The covariance function was proposed to deal with the categorical inputs [32].

5.2.3. Genetic Programming and Genetic Algorithms

An evolutionary algorithm (EA) is a biologically inspired approach to problem solving [96]. The two most used variants of EA in BG prediction and modeling approaches are genetic programming (GP) and genetic algorithms (GA). Hidalgo et al. [76, 77] used a genetic programming-based symbolic regression known as grammatical evolution to develop an individualized model of BG dynamics. Moreover, Contreras et al. [71] used the grammatical evolution approach to develop a standalone BG prediction model. Furthermore, Hidalgo et al. [69] assessed the performance of

different predictors, genetic programming, random forests, k-nearest neighbors, and grammatical evolution along with a new enhanced modeling algorithm, a variant of grammatical evolution that uses optimized grammar, and a variant of tree-based genetic programming that uses a three-compartment model for carbohydrate and insulin dynamics.

5.2.4. Random Forest (RF)

Random forests or random decision forests are an ensemble approach of learning for classification and regression applications, which learns by constructing a multitude of decision tress generating the mode of the class or the mean of prediction. In this regard, for example, Xao et al. [30] developed a random forest regression and support vector regression-based BG predictors and assessed the performance improvement gained through the selection of an optimal feature representative using a combined approach of feature importance scores of ensemble learning and a sequential backward selection (SBS) algorithm. Furthermore, Georga et al. [59] used a random forest regression to predict BG levels with a multivariate dataset containing a subcutaneous glucose profile, plasma insulin concentration, intestinal absorption of meal-derived glucose, and daily energy expenditure.

5.2.5. Hybrid Approach

Hybridization involves combining two or more different approaches, either at the preprocessing, feature extraction, or learning stage when looking for improved performance. The majority of the BG prediction models involve the hybridization of physiological (compartmental) models along with different machine learning techniques. Regarding support vector regression, for example, Plis et al. [43] combined support vector regression along with a physiological model, where the latter generates informative input features to be used to train the SVR model. Furthermore, Georga et al. [60-63] combined support vector regression with compartmental models, which are used to quantify the absorption of subcutaneously administered insulin, glucose from the gut following a meal, and the effects of exercise on plasma glucose and insulin dynamics. Regarding the hybridization of an artificial neural network with other approaches, some researchers have reported success in this direction. For example, Mougiakakou et al. [50] combined an artificial neural network with a compartmental model, where the latter is used to estimate the effect of food on BG levels and the influence of injected insulin on plasma insulin concentration; this output along with the previous BG measurements were used to train the ANN model. Mougiakakou et al. [51] further investigated the combination of a recurrent neural network along with three compartmental models, which estimated the effect of short-acting (SA) insulin intake on blood insulin concentration, intermediate-acting (IA) insulin intake on blood insulin concentration, and, carbohydrate intake on BG absorption from the gut. Zecchin et al. [23, 24] combined an artificial neural network and a physiological model to exploit meal information to be used along with the CGM data. Moreover, Zecchin et al. [15, 25] further explored the applicability of a jump neural network, which is feed by a meal physiological model and CGM data, and compared their result with a previously proposed artificial neural network [23, 24]. Briegel et al. [74] explored a nonlinear state space model for modeling an individual BG dynamic using a compartmental model and an artificial neural network. Furthermore, Otto et al. [49] developed a hybrid model combining an artificial neural network and fuzzy logic, where the fuzzy logic was used to approximate food, insulin, and the level of exercise. Several researchers have attempted to hybridize genetic programming along with physiological models. For example, Contreras et al. [70] developed a hybrid model using a genetic programming-based algorithm known as grammatical evolution and a physiological model. Self-organizing maps (SOMs) have been used to develop a hybrid model along with a physiological model. For example, Zarkogianni et al. [28] used the physiological model to simulate the subcutaneous insulin kinetics and glucose absorption from the gut into the blood, which are in turn fed into the SOM. Jankovic et al. [56] developed a two-layer (prediction and correction layer) online adaptive personalized BG prediction model. The prediction layer consisted of an autoregressive model with external input (ARX) and an artificial neural network, which made the first estimates and then the output was further optimized in the second (correction) layer through an extreme learning machine (ELM).

5.2.6. Ensemble Approach – Merging different Predictors for Performance Improvement

Due to the complexity of BG dynamics, it remains difficult to achieve an accurate prediction in every circumstance (i.e., hypo/eu/hyperglycemia events). One prediction model can have a better prediction power in either of these circumstances, and the other model can achieve better predictive power where the first model fails to accurately predict. Therefore, it is natural to look for opportunities to exploit the strengths from these different predictors to achieve better predictive power in most of these circumstances, which has led to ensemble approaches. An ensemble approach is generally favored when one is interested in merging two or more different predictors for improved performance. Various approaches have been taken to ensemble predictors (e.g., heuristic algorithms; bagging,

boosting, and weighted majorities; the Bayesian model averaging approach, and online versions of these) [33]. The main differences between these approaches are how the weights are determined to achieve the best possible predictive power. In this regard, for example, Wang et al. [31] proposed a novel approach that is able to combine several prediction algorithms, where the adaptive weight of each algorithm is determined through an inversely proportional relationship to its sum of the squared prediction errors. The proposed approach was tested using an autoregressive (AR) model, an extreme learning machine, and support vector regression and achieved a satisfactory result [31]. Moreover, Stahl et al. [34] proposed a novel Bayesian approach to merge multiple predictors by using recursive weighting for a single prediction through a regularized optimization technique. Stahl et al. [33] further investigated a novel merging approach that combines elements from switching and averaging techniques to form a soft switcher in a Bayesian framework. Botwey et al. [66] investigated three different data fusion techniques to merge two predictors, an autoregressive model with output correction, cARX, and a recurrent neural network, RNN, based on the Dempster-Shafer evidential theory (DST), genetic algorithms (GA), and genetic programming (GP). Moreover, Daskalaki et al. [68] merged an autoregressive approach with an output correction module (cARX) model, and recurrent neural network (RNN) models, where the fusion is implemented using a linear combination of the two models' output and the balancing factor (weight) is determined through a customized cost function.

Conclusion

The purpose of this review was to assess and analyze the state-of-the-art machine learning applications in BG predictions pertinent to the controller of an artificial pancreas (a closed loop system), the modeling of a personalized profile, personalized decision support systems, and BG alarm event applications. Previously, a number of different approaches have been tested to develop a BG prediction model. However, due to the complexity of BG dynamics, it remains difficult to achieve a universal model that produces an accurate prediction under free living conditions and in every circumstance (i.e., hypo/eu/hyperglycemia events). Recently, machine learning has been used to address these tasks and has demonstrated to hold a promising future given the ever-growing availability of self-collected health data. The ubiquitous nature and widespread use of mobile health applications (mHealth apps), sensors, wearables, and other POC devices for self-monitoring and management purposes have made possible the generation of automated and continuous personal data, which have created an opportunity to use much more detailed data to better train the machine learning model. Various machine learning techniques have been tested to predict BG, such as, recurrent neural networks, feed-forward neural networks, support vector machines, self-organizing maps, Gaussian processes, genetic algorithms and programming, and deep neural networks. These techniques use various groups of input parameters and training algorithms. The main limitation of the current approaches is the lack of a well-defined approach to estimate carbohydrate intake, which is mainly done manually by the individual users and is prone to an error that can severely affect the predictive performance. Moreover, there is the lack of a universal approach to estimate and quantify the approximate effect of physical activities, stress, and infection incidence on the BG level. Almost all the studies have quite different approaches, and this poses a challenge in terms of regarding one approach as universal. None of the researchers have assessed model predictive performance during stress and infection incidences in a free-living condition, which should be taken into account in future studies. Furthermore, little has been done regarding model portability that can capture the inter- and intra-variations among patients. It seems that the effect of time lags between the CGM reading and the actual BG levels is also not well covered. However, in general, we foresee that these developments might foster the next generation of BG prediction, which should result in a great contribution in the effort to develop the long-awaited so-called artificial pancreas (a closed-loop system).

Acronyms

Artificial Neural Network (ANN)
Decision Tree (DT)
Support vector machine (SVM)
Recurrent neural network (RNN)
Extreme learning machine (ELM)
Gaussian Process (GAP)
Genetic Programming (GP)
Genetic algorithm (GA)

Blood Glucose (BG)
Continuous Glucose Monitoring (CGM)
Self-Monitoring Blood Glucose (SMBG)
Self-organizing map (SOM)
Random Forest (RF)

References

1. Ogurtsova, K., J.D. da Rocha Fernandes, Y. Huang, U. Linnenkamp, L. Guariguata, N.H. Cho, D. Cavan, J.E. Shaw, and L.E. Makaroff, *IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040.* Diabetes Res Clin Pract, 2017. **128**: p. 40-50.
2. (IDF), I.D.F. *IDF diabetes atlas - 7th edition*. 2015; Available from: http://www.diabetesatlas.org/.
3. Solli, O., T. Jenssen, and I.S. Kristiansen, *Diabetes: cost of illness in Norway.* BMC Endocr Disord, 2010. **10**: p. 15.
4. Skrivarhaug, T., L.C. Stene, A.K. Drivvoll, H. Strom, G. Joner, and G. Norwegian Childhood Diabetes Study, *Incidence of type 1 diabetes in Norway among children aged 0-14 years between 1989 and 2012: has the incidence stopped rising? Results from the Norwegian Childhood Diabetes Registry.* Diabetologia, 2014. **57**(1): p. 57-62.
5. American Diabetes, A., *Economic costs of diabetes in the U.S. in 2012.* Diabetes Care, 2013. **36**(4): p. 1033-46.
6. Diabetes, P.o.t.s.W.o.A.I.f., *1st Workshop on Artificial intelligence for Diabetes (ECAI 2016)*. 2016: Zenodo.
7. Béranger, J., *1 - The Shift towards a Connected, Assessed and Personalized Medicine Centered Upon Medical Datasphere Processing*, in *Big Data and Ethics*. 2016, Elsevier. p. 1-95.
8. Swan, M., *The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery.* Big Data, 2013. **1**(2): p. 85-99.
9. Béranger, J., *Introduction*, in *Big Data and Ethics*. 2016, Elsevier. p. xi-xxxvi.
10. Mohammadi, D., *Turning big data into personalised diabetes care.* The Lancet Diabetes & Endocrinology, 2015. **3**(12): p. 935-936.
11. Kavakiotis, I., O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, *Machine Learning and Data Mining Methods in Diabetes Research.* Comput Struct Biotechnol J, 2017. **15**: p. 104-116.
12. Oviedo, S., J. Vehi, R. Calm, and J. Armengol, *A review of personalized blood glucose prediction strategies for T1DM patients.* Int J Numer Method Biomed Eng, 2016.
13. Zarkogianni, K., E. Litsa, K. Mitsis, P.Y. Wu, C.D. Kaddi, C.W. Cheng, M.D. Wang, and K.S. Nikita, *A Review of Emerging Technologies for the Management of Diabetes Mellitus.* IEEE Transactions on Biomedical Engineering, 2015. **62**(12): p. 2735-2749.
14. Banaee, H., M.U. Ahmed, and A. Loutfi, *Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges.* Sensors (Basel), 2013. **13**(12): p. 17472-500.
15. Zecchin, C., A. Facchinetti, G. Sparacino, and C. Cobelli, *Jump neural network for online short-time prediction of blood glucose from continuous monitoring sensors and meal information.* Computer Methods and Programs in Biomedicine, 2014. **113**(1): p. 144-152.
16. Balakrishnan, N.P., G.P. Rangaiah, and L. Samavedham, *Review and Analysis of Blood Glucose (BG) Models for Type 1 Diabetic Patients.* Industrial & Engineering Chemistry Research, 2011. **50**(21): p. 12041-12066.
17. Allam, F., Z. Nossai, H. Gomma, I. Ibrahim, and M. Abdelsalam, *A Recurrent Neural Network Approach for Predicting Glucose Concentration in Type-1 Diabetic Patients*, in *Engineering Applications of Neural Networks: 12th INNS EANN-SIG International Conference, EANN 2011 and 7th IFIP WG 12.5 International Conference, AIAI 2011, Corfu, Greece, September 15-18, 2011, Proceedings Part I*, L. Iliadis and C. Jayne, Editors. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 254-259.
18. Allam, F., Z. Nossair, H. Gomma, I. Ibrahim, and M. Abd-el Salam. *Prediction of subcutaneous glucose concentration for type-1 diabetic patients using a feed forward neural network*. in *Computer Engineering & Systems (ICCES), 2011 International Conference on*. 2011. IEEE.
19. Baghdadi, G. and A.M. Nasrabadi. *Controlling Blood Glucose Levels in Diabetics By Neural Network Predictor*. in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2007. Lyon, France.
20. Bunescu, R., N. Struble, C. Marling, J. Shubrook, and F. Schwartz, *Blood Glucose Level Prediction Using Physiological Models and Support Vector Regression.* 2013: p. 135-140.

21. Chernetsov, S., A. Emelyanov, and A. Karpenko. *Research of neural network-based blood glucose level forecasting systems for insulin-dependant diabetes patients*. in *Proceedings of the 6th International Workshop on Wearable, Micro, and Nano Technologies for Personalized Health*. 2009. Oslo, Norway.
22. Daskalaki, E., A. Prountzou, P. Diem, and S.G. Mougiakakou, *Real-time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients.* Diabetes Technol Ther, 2012. **14**(2): p. 168-74.
23. Zecchin, C., A. Facchinetti, G. Sparacino, G.D. Nicolao, and C. Cobelli, *Neural Network Incorporating Meal Information Improves Accuracy of Short-Time Prediction of Glucose Concentration.* IEEE Transactions on Biomedical Engineering, 2012. **59**(6): p. 1550-1560.
24. Zecchin, C., A. Facchinetti, G. Sparacino, G.D. Nicolao, and C. Cobelli. *A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information*. in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011. Boston, MA, USA.
25. Zecchin, C., A. Facchinetti, G. Sparacino, and C. Cobelli, *Jump Neural Network for Real-Time Prediction of Glucose Concentration*, in *Artificial Neural Networks*, H. Cartwright, Editor. 2015, Springer New York: New York, NY. p. 245-259.
26. Zarkogianni, K., K. Mitsis, E. Litsa, M.T. Arredondo, G. Ficomicron, A. Fioravanti, and K.S. Nikita, *Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring.* Med Biol Eng Comput, 2015. **53**(12): p. 1333-43.
27. Zarkogianni, K., K. Mitsis, M.T. Arredondo, G. Fico, A. Fioravanti, and K.S. Nikita. *Neuro-fuzzy based glucose prediction model for patients with Type 1 diabetes mellitus*. in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. 2014. Valencia, Spain
28. Zarkogianni, K., E. Litsa, A. Vazeou, and K.S. Nikita. *Personalized glucose-insulin metabolism model based on self-organizing maps for patients with Type 1 Diabetes Mellitus*. in *13th IEEE International Conference on BioInformatics and BioEngineering*. 2013. Chania, Greece
29. Zainuddin, Z., O. Pauline, and C. Ardil, *A Neural Network Approach in Predicting the Blood Glucose Level for Diabetic Patients.* World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2009. **3**(2): p. 436-443.
30. Xao, W., F. Shao, J. Ji, R. Sun, and C. Xing. *Fasting Blood Glucose Change Prediction Model Based on Medical Examination Data and Data Mining Techniques*. in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*. 2015. Chengdu, China.
31. Wang, Y., X. Wu, and X. Mo, *A novel adaptive-weighted-average framework for blood glucose prediction.* Diabetes Technol Ther, 2013. **15**(10): p. 792-801.
32. Tomczak, J.M., *Gaussian Process Regression with Categorical Inputs for Predicting the Blood Glucose Level*, in *Advances in Systems Science: Proceedings of the International Conference on Systems Science 2016 (ICSS 2016)*, J. Świątek and J.M. Tomczak, Editors. 2017, Springer International Publishing: Cham. p. 98-108.
33. Ståhl, F., R. Johansson, and E. Renard, *Ensemble Glucose Prediction in Insulin-Dependent Diabetes*, in *Data-driven Modeling for Diabetes: Diagnosis and Treatment*, V. Marmarelis and G. Mitsis, Editors. 2014, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 37-71.
34. Stahl, F., R. Johansson, and E. Renard, *Bayesian combination of multiple plasma glucose predictors.* Conf Proc IEEE Eng Med Biol Soc, 2012. **2012**: p. 2839-44.
35. Naumova, V., S.V. Pereverzyev, and S. Sivananthan, *A meta-learning approach to the regularized learning-case study: blood glucose prediction.* Neural Netw, 2012. **33**: p. 181-93.
36. Shanthi, S. and D. Kumar, *Prediction of blood glucose concentration ahead of time with feature based neural network.* Malaysian Journal of Computer Science, 2013. **25**(3).
37. Shanthi, S., P. Balamurugan, and D. Kumar. *Performance comparison of featured neural network with gradient descent and levenberg-marquart algorithm trained neural networks for prediction of blood glucose values with continuous glucose monitoring sensor data*. in *2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET)*. 2012. Tiruchirappalli, India
38. Sandham, W.A., E.D. Lehmann, D.J. Hamilton, and M.L. Sandilands. *Simulating and predicting blood glucose levels for improved diabetes healthcare*. in *4th IET International Conference on Advances in Medical, Signal and Information Processing - MEDSIP 2008*. 2008. Santa Margherita Ligure, Italy
39. Sandham, W.A., D.J. Hamilton, A. Japp, and K. Patterson. *Neural network and neuro-fuzzy systems for improving diabetes therapy*. in *Proceedings of the 20th Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286). 1998. Hong Kong, China, China.
40. Robertson, G., E.D. Lehmann, W. Sandham, and D. Hamilton, *Blood Glucose Prediction Using Artificial Neural Networks Trained with the AIDA Diabetes Simulator: A Proof-of-Concept Pilot Study.* Journal of Electrical and Computer Engineering, 2011. **2011**: p. 11.
41. Reymann, M.P., E. Dorschky, B.H. Groh, C. Martindale, P. Blank, and B.M. Eskofier, *Blood glucose level prediction based on support vector regression using mobile platforms.* Conf Proc IEEE Eng Med Biol Soc, 2016. **2016**: p. 2990-2993.
42. Quchani, S.A. and E. Tahami, *Comparison of MLP and Elman Neural Network for Blood Glucose Level Prediction in Type 1 Diabetics*, in *3rd Kuala Lumpur International Conference on Biomedical Engineering 2006: Biomed 2006, 11 – 14 December 2006 Kuala Lumpur, Malaysia*, F. Ibrahim, et al., Editors. 2007, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 54-58.
43. Plis, K., R.C. Bunescu, C. Marling, J. Shubrook, and F. Schwartz. *A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management*. in *AAAI Workshop: Modern Artificial Intelligence for Health Analytics.* 2014.
44. Perez-Gandia, C., A. Facchinetti, G. Sparacino, C. Cobelli, E.J. Gomez, M. Rigla, A. de Leiva, and M.E. Hernando, *Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring.* Diabetes Technol Ther, 2010. **12**(1): p. 81-8.
45. Pappada, S.M., B.D. Cameron, P.M. Rosman, R.E. Bourey, T.J. Papadimos, W. Olorunto, and M.J. Borst, *Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes.* Diabetes Technol Ther, 2011. **13**(2): p. 135-41.
46. Pappada, S.M., B.D. Cameron, and P.M. Rosman, *Development of a Neural Network for Prediction of Glucose Concentration in Type 1 Diabetes Patients.* Journal of diabetes science and technology (Online), 2008. **2**(5): p. 792-801.
47. Pappada, S.M. and B.D. Cameron, *Neural Network Modeling Approaches for Patient Specific Glycemic Forecasting*, in *Patient-Specific Modeling in Tomorrow's Medicine*, A. Gefen, Editor. 2012, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 505-529.
48. Pappada, S.M., M.J. Borst, B.D. Cameron, R.E. Bourey, J.D. Lather, D. Shipp, A. Chiricolo, and T.J. Papadimos, *Development of a neural network model for predicting glucose levels in a surgical critical care setting.* Patient Saf Surg, 2010. **4**(1): p. 15.
49. Otto, E., C. Semotok, J. Andrysek, and O. Basir, *An intelligent diabetes software prototype: predicting blood glucose levels and recommending regimen changes.* Diabetes Technol Ther, 2000. **2**(4): p. 569-76.
50. Mougiakakou, S.G., A. Prountzou, D. Iliopoulou, K.S. Nikita, A. Vazeou, and C.S. Bartsocas, *Neural network based glucose - insulin metabolism models for children with Type 1 diabetes.* Conf Proc IEEE Eng Med Biol Soc, 2006. **1**: p. 3545-8.
51. Mougiakakou, S., K. Prountzou, and K. Nikita, *A real time simulation model of glucose-insulin metabolism for type 1 diabetes patients.* Conf Proc IEEE Eng Med Biol Soc, 2005. **1**: p. 298-301.
52. Mhaskar, H.N., S.V. Pereverzyev, and M.D. van der Walt, *A Deep Learning Approach to Diabetic Blood Glucose Prediction.* Frontiers in Applied Mathematics and Statistics, 2017. **3**.
53. Mathiyazhagan, N. and H.B. Schechter. *Soft computing approach for predictive blood glucose management using a fuzzy neural network*. in *2014 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*. 2014. Boston, MA, USA
54. Li, J. and C. Fernando, *Smartphone-based personalized blood glucose prediction.* ICT Express, 2016. **2**(4): p. 150-154.
55. Khaled, E., R. Tim, and B. Osama, *Online Prediction of Blood Glucose Levels using Genetic Algorithm*, in *Biologically-Inspired Techniques for Knowledge Discovery and Data Mining*, A. Shafiq, et al., Editors. 2014, IGI Global: Hershey, PA, USA. p. 299-310.
56. Jankovic, M.V., S. Mosimann, L. Bally, C. Stettler, and S. Mougiakakou. *Deep prediction model: The case of online adaptive prediction of subcutaneous glucose*. in *2016 13th Symposium on Neural Networks and Applications (NEUREL)*. 2016. Belgrade, Serbia.
57. Georga, E.I., V.C. Protopappas, D. Polyzos, and D.I. Fotiadis. *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 (EMBC)*. 2015. Milan, Italy
58. Georga, E.I., V.C. Protopappas, D. Polyzos, and D.I. Fotiadis, *Evaluation of short-term predictors of glucose concentration in type 1 diabetes combining feature ranking with regression models.* Medical & Biological Engineering & Computing, 2015. **53**(12): p. 1305-1318.

59. Georga, E.I., V.C. Protopappas, D. Polyzos, and D.I. Fotiadis, *A predictive model of subcutaneous glucose concentration in type 1 diabetes based on Random Forests.* Conf Proc IEEE Eng Med Biol Soc, 2012. **2012**: p. 2889-92.
60. Georga, E.I., V.C. Protopappas, D. Polyzos, and D.I. Fotiadis. *Predictive modeling of glucose metabolism using free-living data of type 1 diabetic patients*. in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 2010. Buenos Aires, Argentina
61. Georga, E.I., V.C. Protopappas, and D. Polyzos. *Prediction of glucose concentration in type 1 diabetic patients using support vector regression*. in *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*. 2010. Corfu, Greece
62. Georga, E.I., V.C. Protopappas, and D.I. Fotiadis, *Glucose prediction in type 1 and type 2 diabetic patients using data driven techniques*, in *Knowledge-oriented applications in data mining*, K. Funatsu, Editor. 2011, InTech.
63. Georga, E.I., V.C. Protopappas, D. Ardigò, M. Marina, I. Zavaroni, D. Polyzos, and D.I. Fotiadis, *Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression.* IEEE Journal of Biomedical and Health Informatics, 2013. **17**(1): p. 71-81.
64. Fernandez de Canete, J., S. Gonzalez-Perez, and J.C. Ramos-Diaz, *Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes.* Comput Methods Programs Biomed, 2012. **106**(1): p. 55-66.
65. Ahmed, H.B. and A. Serener, *Effects of External Factors in CGM Sensor Glucose Concentration Prediction.* Procedia Computer Science, 2016. **102**(Supplement C): p. 623-629.
66. Valletta, J.J., A.J. Chipperfield, and C.D. Byrne, *Gaussian Process modelling of blood glucose response to free-living physical activity data in people with type 1 diabetes.* Conf Proc IEEE Eng Med Biol Soc, 2009. **2009**: p. 4913-6.
67. Botwey, R.H., E. Daskalaki, P. Diem, and S.G. Mougiakakou. *Multi-model data fusion to improve an early warning system for hypo-/hyperglycemic events*. in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2014. Chicago, IL, USA
68. Daskalaki, E., K. Nørgaard, T. Züger, A. Prountzou, P. Diem, and S. Mougiakakou, *An Early Warning System for Hypoglycemic/Hyperglycemic Events Based on Fusion of Adaptive Prediction Models.* Journal of Diabetes Science and Technology, 2013. **7**(3): p. 689-698.
69. Hidalgo, J.I., J.M. Colmenar, G. Kronberger, S.M. Winkler, O. Garnica, and J. Lanchares, *Data Based Prediction of Blood Glucose Concentrations Using Evolutionary Methods.* Journal of Medical Systems, 2017. **41**(9): p. 142.
70. Contreras, I., S. Oviedo, M. Vettoretti, R. Visentin, and J. Vehi, *Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models.* PLoS One, 2017. **12**(11): p. e0187754.
71. Contreras, I. and J. Vehi, *Mid-Term Prediction of Blood Glucose from Continuous Glucose Sensors, Meal Information and Administered Insulin*, in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016: MEDICON 2016, March 31st-April 2nd 2016, Paphos, Cyprus*, E. Kyriacou, S. Christofides, and C.S. Pattichis, Editors. 2016, Springer International Publishing: Cham. p. 1137-1143.
72. Frandes, M., B. Timar, and D. Lungeanu, *A Risk Based Neural Network Approach for Predictive Modeling of Blood Glucose Dynamics*. Exploring Complexity in Health: An Interdisciplinary Systems Approach, ed. Alexander Hoerbst, et al. Vol. 228. 2016: Studies in Health Technology and Informatics.
73. Alanis, A.Y., E.N. Sanchez, E. Ruiz-Velazquez, and B.S. Leon. *Neural model of blood glucose level for Type 1 Diabetes Mellitus Patients*. in *The 2011 International Joint Conference on Neural Networks*. 2011. San Jose, CA, USA.
74. Briegel, T. and V. Tresp, *A Nonlinear State Space Model for the Blood Glucose Metabolism of a Diabetic (Ein nichtlineares Zustandsraummodell für den Blutglukosemetabolismus eines Diabetikers).* at - Automatisierungstechnik, 2002. **50**(5/2002).
75. Georga, E.I., V.C. Protopappas, and D.I. Fotiadis, *Predictive Metabolic Modeling for Type 1 Diabetes Using Free-Living Data on Mobile Devices*, in *Wireless Mobile Communication and Healthcare: Second International ICST Conference, MobiHealth 2010, Ayia Napa, Cyprus, October 18-20, 2010. Revised Selected Papers*, J.C. Lin and K.S. Nikita, Editors. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 187-193.
76. Hidalgo, J.I., J.M. Colmenar, J.L. Risco-Martin, A. Cuesta-Infante, E. Maqueda, M. Botella, and J.A. Rubio, *Modeling glycemia in humans by means of Grammatical Evolution.* Applied Soft Computing, 2014. **20**: p. 40-53.

77. Hidalgo, J.I., J.M. Colmenar, J.L. Risco-Martín, E. Maqueda, M. Botella, J.A. Rubio, A. Cuesta-Infante, O. Garnica, and J. Lanchares, *Clarke and parkes error grid analysis of diabetic glucose models obtained with evolutionary computation.* 2014: p. 1305-1312.
78. Mirshekarian, S., R. Bunescu, C. Marling, and F. Schwartz, *Using LSTMs to learn physiological models of blood glucose behavior.* Conf Proc IEEE Eng Med Biol Soc, 2017. **2017**: p. 2887-2891.
79. Ruiz-Velázquez, E., A.Y. Alanis, R. Femat, and G. Quiroz. *Neural modeling of the blood glucose level for Type 1 Diabetes Mellitus patients*. in *2011 IEEE International Conference on Automation Science and Engineering*. 2011. Trieste, Italy.
80. Hamdi, T., J. Ben Ali, V. Di Costanzo, F. Fnaiech, E. Moreau, and J.-M. Ginoux, *Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm.* Biocybernetics and Biomedical Engineering, 2018. **38**(2): p. 362-372.
81. Ben Ali, J., T. Hamdi, N. Fnaiech, V. Di Costanzo, F. Fnaiech, and J.-M. Ginoux, *Continuous blood glucose level prediction of Type 1 Diabetes based on Artificial Neural Network.* Biocybernetics and Biomedical Engineering, 2018.
82. Marsland, S., *Machine learning: an algorithmic perspective*. Chapman & Hall/CRC Machine Learning & Pattern Recognition, ed. R. Herbrich and T. Graepel. 2014, New York, USA: Chapman and Hall/CRC 205.
83. Lehmann, E.D. and T. Deutsch, *A physiological model of glucose-insulin interaction in type 1 diabetes mellitus.* Journal of Biomedical Engineering, 1992. **14**(3): p. 235-242.
84. Percival, M.W., W.C. Bevier, Y. Wang, E. Dassau, H.C. Zisser, L. Jovanovic, and F.J. Doyle, 3rd, *Modeling the effects of subcutaneous insulin administration and carbohydrate consumption on blood glucose.* J Diabetes Sci Technol, 2010. **4**(5): p. 1214-28.
85. Duke, D.L., *Intelligent Diabetes Assistant: A Telemedicine System for Modeling and Managing Blood Glucose*. 2010, Carnegie Mellon University: Pittsburgh, Pennsylvania.
86. Roy, A. and R.S. Parker, *Dynamic modeling of exercise effects on plasma glucose and insulin levels.* J Diabetes Sci Technol, 2007. **1**(3): p. 338-47.
87. Riddell, M. and B.A. Perkins, *Exercise and glucose metabolism in persons with diabetes mellitus: perspectives on the role for continuous glucose monitoring.* J Diabetes Sci Technol, 2009. **3**(4): p. 914-23.
88. Briegel, T. and V. Tresp, *A Nonlinear State Space Model for the Blood Glucose Metabolism of a Diabetic (Ein nichtlineares Zustandsraummodell für den Blutglukosemetabolismus eines Diabetikers).* at-Automatisierungstechnik Methoden und Anwendungen der Steuerungs-, Regelungs-und Informationstechnik, 2002. **50**(5/2002): p. 228.
89. Botsis, T., A.M. Lai, W. Palmas, J.B. Starren, G. Hartvigsen, and G. Hripcsak, *Proof of concept for the role of glycemic control in the early detection of infections in diabetics.* Health Informatics Journal, 2012. **18**(1): p. 26-35.
90. Woldaregay, A.Z., E. Årsand, A. Giordanengo, D. Albers, L. Mamykina, T. Botsis, and G. Hartvigsen. *EDMON-A Wireless Communication Platform for a Real-Time Infectious Disease Outbreak De-tection System Using Self-Recorded Data from People with Type 1 Diabetes*. in *Proceedings from The 15th Scandinavian Conference on Health Informatics 2017 Kristiansand, Norway, August 29–30, 2017*. 2018. Linköping University Electronic Press.
91. Woldaregay, A., E. Årsand, T. Botsis, and G. Hartvigsen, *An Early Infectious Disease Outbreak Detection Mechanism Based on Self-Recorded Data from People with Diabetes.* Studies in health technology and informatics, 2017. **245**: p. 619-623.
92. Marcovecchio, M.L. and F. Chiarelli, *The effects of acute and chronic stress on diabetes control.* Sci Signal, 2012. **5**(247): p. pt10.
93. Yadav, S. and S. Shukla, *Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification.* 2016: p. 78-83.
94. Shanmuganathan, S. and S. Samarasinghe, *Artificial Neural Network Modelling*. Studies in Computational Intelligence Vol. 628. 2016.
95. Campbell, C. and Y. Ying, *Learning with Support Vector Machines*. 2011: Morgan \\& Claypool Publishers. 100.
96. Wong, M.L. and K.S. Leung, *Data Mining using Grammar based Genetic Programming and Applications.* Genetic Programming Series, ed. K. John. 2002, New York / Boston / Dordrecht / London / Moscow: Kluwer Academic Publishers.