

# Different Machine Learning Algorithms Involved in Glucose Monitoring to Prevent Diabetes Complications and Enhanced Diabetes Mellitus Management



Wai-kit Ming and Zonglin He

**Abstract** Diabetes mellitus (DM) is a group of metabolic disorders resulting from dysregulation of blood glucose (BG). Hence, it may lead to various vascular and neural complications involving multiple organ systems, either short- or long-term. During the past two decades, various computer-assisted systems based on machine learning algorithms have become available and have achieved satisfactory performance in glucose monitoring and predicting the prognosis of diabetic patients. The increased availability of multidimensional health data has shed light on machine learning for a novel BG prediction and diabetes management method. So far, various machine learning algorithms have been productive in predicting BG and diabetes progression and prognosis. Hence, machine learning algorithms have been regarded as accurate, with less operation cost and higher efficacy in predicting potential diabetes in the undiagnosed population, profiling personalized BG dynamics, establishing personalized decision support systems, and building BG alarm events in DM patients. However, real-world data concerning the efficacy of various machine learning algorithms in diabetes prediction and management is still limited, and internationally acceptable guidelines have not been established to estimate and quantify the potential lifestyle-relevant variables related to the BG level. This chapter has been written to address the current progress in the application of machine learning in glucose monitoring and DM management. Different machine learning algorithms have also been discussed on the validity and feasibility of the algorithms fit for purpose.

**Keywords** Machine learning · Surveillance · Prediction · Diabetes mellitus · Management

---

W. Ming (✉) · Z. He

Department of Preventive Medicine and Public Health, School of Medicine, Jinan University, Guangzhou, China

e-mail: [kming@connect.hku.hk](mailto:kming@connect.hku.hk)

Z. He

e-mail: [hezonglin0leon@stu2015.jnu.edu.cn](mailto:hezonglin0leon@stu2015.jnu.edu.cn)

## 1 Introduction

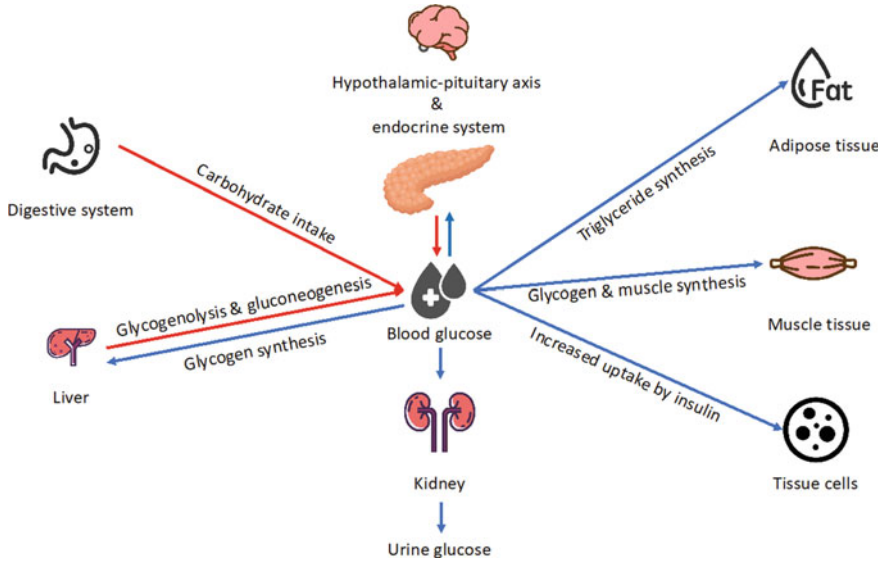
Diabetes mellitus (DM), which entails a global health crisis, is a group of metabolic disorders that results from dysregulation of blood glucose (BG), either due to the failure of the body to secrete insulin (type I diabetes mellitus, T1DM) or the inability of the body to respond to insulin action (type II diabetes mellitus, T2DM), as well as first recognition during pregnancy (Gestational Diabetes Mellitus, GDM) or other specific types [69]. The patients may present with chronic hyperglycemia, manifesting polydipsia, polyuria, and polyphagia.

Clinically, the current mainstream diagnostic investigation method of DM is venous plasma glucose measurement [28], and 2-h oral glucose tolerance test (OGTT) remains the internationally accepted gold standard for DM diagnosis, where the venous plasma glucose levels are obtained for fasting, as well as 1-h- and 2-h-post a certain amount of glucose intake (normally 50 g or 75 g).

The mainstream view of the pathophysiology of DM remained that genetic predisposition underlies DM development, where what control the biological steps of beta-cell action, insulin secretion, insulin interaction with tissue cells, insulin receptor production and insulin action inside the cells that were altered or mutated [22]. T2DM patients are getting increasingly insensitive to the physiological effects of insulin. Therefore, more insulin is needed to maintain the original effects of insulin to induce cells to uptake glucose [29]. Nevertheless, for T1DM patients, as their insulin production by beta cells is impaired, therefore, eventually, for both T1DM and T2DM patients, the pharmacological induction of insulin secretion or insulin absorption is no longer sufficient for maintaining the euglycemic state. External insulin supplementation is the *sine qua non* for diabetes management [39].

DM without proper management may lead to a variety of vascular and neural complications involving multiple organ systems either in a short- or long-term manner, and it is the multiple complications secondary to DM that lead to the heavy burdens of the patients, causing increased medical cost and decreased quality of life [22]. In this sense, regular community-based screening and prompt diagnosis in undiagnosed patients, sufficient patient education and support, continuous medical care, and user-friendly continuous BG monitoring, as well as psychological dredge and social support, are required to prevent acute complications (e.g., ketoacidosis) and minimize the risk of long-term complications (e.g., nephropathy, retinopathy, diabetic foot, cardiovascular disease, or stroke) [14, 29].

Therefore, on the one hand, in the community, timely screening of diabetes in undiagnosed patients could help prevent further development of diabetic complications, hence reducing disease burden and improving quality of life. On the other hand, for DM patients, BG monitoring is of vital significance. It is acknowledged that optimizing glycemic control through lowering BG levels and minimizing glucose variability could prevent the development of microvascular complications and long-term macrovascular disease [47, 54]. BG serves as the most important risk factor and prognostic factor in DM patients owing to its predictive values in disease progression; it is difficult to manage because of its multifactorial nature, as well as inter-and

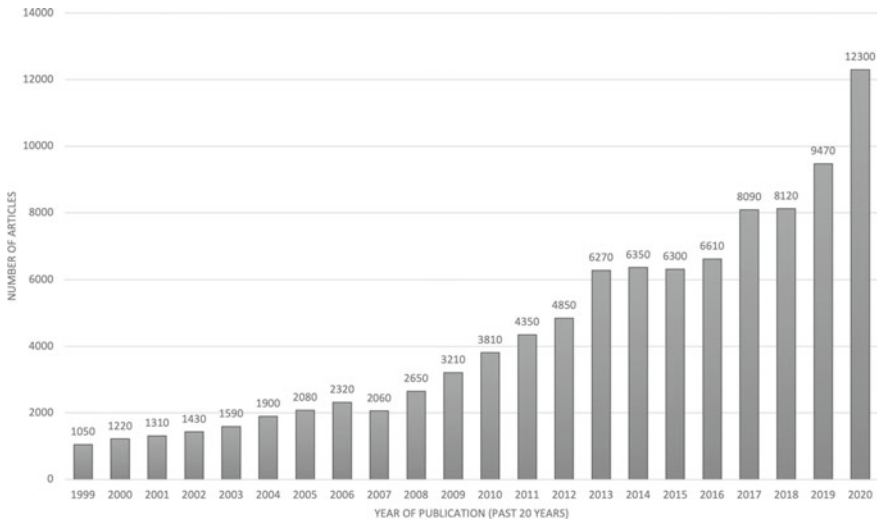


**Fig. 1** Systems and organs related to blood glucose level

intra-personal variability associated with nutritional, behavioral, and pharmaceutical management, as shown in Fig. 1 [54].

Specifically, timely acknowledgment of the fluctuation of blood glucose levels underlies the foundation of diabetes management. With proper and timely blood glucose monitoring, efficacious treatment, dysglycemia (especially undetected hypoglycemia) identification, and treatment plan modification (including medical nutrition therapy, exercise therapy, and pharmaceutical interventions) become possible. Normally, the blood glucose level is checked before the meal, 2 h post-meal, and before sleep [1].

The emergence of self-monitoring of blood glucose (SMBG) has inspired diabetes management in the previous decades, aspiring for euglycemia. Yet, its inconvenience in use may lead to incomplete BG data collection [1]. Moreover, portable blood glucose meters have allowed patients and healthcare workers to obtain dynamic blood glucose level data. With the development of technology, the advent of continuous glucose monitoring (CGM) has made surveillance of fluctuation pattern, frequency, level, and timing of BG level variation possible, and it is proven useful in alarming hypoglycemia. Nevertheless, the CGM devices could be expensive and require continued capillary glucose testing for calibration. Despite the gradual transition from SMBG to more advanced glucose monitoring devices, some reluctance to monitor the blood glucose has been noted given the costs, complexity in use, and low awareness of the necessity. Though SMBG has been available for many DM patients, the need for frequent testing and continuous replenishment of consumables has undermined the patients' compliance. Besides traditional serum glucose monitoring, novel materials have also inspired glucose monitoring. For instance, non-invasive and non-enzymatic



**Fig. 2** The number of published articles in Google Scholar includes “Blood glucose” and “Machine learning”

sensing using advanced nanomaterials gained popularity, despite lacking sufficient clinical evidence in the accuracy and stability of long-term glucose monitoring [15, 56, 59]. Hence, accurately monitoring the blood glucose while improving glycaemic control and the quality of life of these patients is now one of the biggest challenges in DM management. The recent boom in BG levels prediction arises with the explosion of interest in Artificial Pancreas Project, a closed-loop control system for BG control [60], and a gross estimate of the number of academic papers concerning “blood glucose” and “machine learning” in the Google Scholar database is shown in Fig. 2.

Artificial intelligence (AI) is progressively utilized in medicine to find patterns in complex sets of clinically collected data and self-monitored data to improve health outcomes [38]. Among many AI-based algorithms, machine learning (ML) can equip computers with the ability to learn without the need to be explicitly programmed in advance [49]. The ML algorithms provide the added value of the expertise of clinicians. It is better than using only one in disease treatment [11, 68], especially in better DM management and complications prevention (Gadekallu et al. 2020; [51]).

In the present chapter, the following contents will be addressed: (1) the role of ML algorithms in DM management; (2) difference between various ML algorithms; (3) insights into future ML application.

## 2 The Role of ML Algorithms in DM Management

Specifically, the ML takes part in the DM managements mainly in three main aspects: (1) assisting precise BG level prediction; (2) detecting DM-associated complications and BG alarm event (BG anomalies); (3) establishing personalized decision support/education systems.

### 2.1 BG Levels Prediction

BG levels are variable and multifactorial, directly affected by insulin, physical activity, and dietary intakes, and influenced by numerous factors. Owing to the dynamic nature of BG levels, some scholars also conceptualized a physiological model that could consider the daily events that influence BG levels, including insulin uptakes, food intakes, exercise, sleep, and even seasonal variation [41].

A comprehensive understanding of the pathophysiology and biological mechanisms underlying DM development and progression is the foundation of incorporating physiological parameters in the ML algorithms. Generally, a physiology-based approach to ML strategies would fractionize the parameters related to BG regulation into three distinct categories, viz., BG dynamics, insulin dynamics, and meal absorption dynamics [40]. Two methods are generally used to incorporate the physiology-based data, namely the lumped (semi-empirical) model and the comprehensive model, where the former would only consist of a few equations and parameters, taking all the organs and tissues as a whole. At the same time, the latter manages data separately according to various organs and tissues [6].

Moreover, the increasing popularity of mobile health applications, biosensors, wearables, and many devices for self-monitoring and healthcare management has also made possible the generation and collection of automated and continuous health-related personal data to feed the ML algorithms [62], such as body mass index, stress level, amount of sleeping time, underlying diseases, medications use, smoking habit, menstruation, alcoholism, allergies, and geological factors [62].

Nevertheless, compared to the physiology-based approach, another approach coined the data-driven strategy also internalize self-collected data and other easily available parameters to predict BG. Regarded as the black box, although sometimes this approach can achieve a high accuracy rate, it is sometimes difficult to interpret the results since it lacks biological and physiological theoretical support underlying the mechanisms of the algorithms [62]. In sum, it could be divided into three different models, namely a time series model, machine learning model, and hybrid model.

Specifically, for DM patients, necessary alarms could be noted through BG prediction to avoid disease progression and over- or under-regulation of BG levels, causing hypoglycemia or hyperglycemia. Sudharsan et al. have shown that robust ML models for hypoglycemia prediction in T2DM patients could effectively identify vulnerable patients needing to manage hypoglycemia [55, 64]. Oviedo et al. conducted

a methodological review regarding the prediction models of BG levels, risks, and events. They found that the algorithms setup and performance metrics of the ML algorithms currently reported were mainly focused on a closed-loop system (an artificial pancreas) [40]. As reported by Woldaregay et al., in terms of BG level prediction, feedforward neural networks remain the most used algorithms (20%), followed by hybridization of the physiology-based model and machine learning techniques (19%), recurrent neural networks (18%), and support vector machines (SVMs) (11%) [62].

## 2.2 *Detection of DM-Associated Complications*

Continuously increased ML models attempting to manage DM-associated complications have been built and assessed. Studies have proven the efficacy of ML-assisted T2DM care programs in the community by identifying population-level effects and mostly benefited patient sub-groups [14, 66]. Makino et al. has demonstrated the use of machine learning (scikit-learn), building a prediction model from 24 factors of interest in predicting the progression of diabetic kidney disease, and an accuracy of 71% was achieved [34].

A systematic review by Kavakiotis et al. has summarized the efficacious role of ML and data mining techniques in diabetes screening and diagnosis and detection and management of complications [26]. Nevertheless, complications secondary to T1DM were scarcely investigated using ML prediction models [26]. T2DM prediction in the community is beneficial for the early detection of T2DM in populations with high-risk factors. It might robustly capture cases with early dysglycaemia but present with no obvious clinical symptoms [13].

On the other hand, pre-hospital screening is also an important application of ML algorithms. Haq et al. proposed a filter method based on a decision tree for incredibly important feature selection and incorporated two ensemble learning algorithms, Ada Boost and Random Forest, for feature selection; the proposed algorithm could reach a test accuracy of 99%, 99.8% with k-folds and 99.9% with LOSO validation, in identifying populations at risk of DM [23].

DM risk classification is vital and challenging, as the medical data is non-linear, non-normal, and complex [7, 35]. A variety of ML algorithms have been developed for the prediction and diagnosis of diabetes disease, viz., (1) supervised algorithms including decision tree (DT), random forest (RF), linear regression, logistic regression (LR), Gaussian process classification (GPC), naïve Bayes (NB), as well as neural networks like artificial neural network (ANN) and feedforward neural network (FFNN); and (2) unsupervised algorithms such as k-nearest neighborhood (KNN), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and support vector machine (SVM) [8, 35]. And the efficacy of such algorithms has been evaluated and reported by various researchers, with an accuracy of DM prediction ranging from 70 to 99% [4, 5, 9, 21, 25, 27, 33, 42, 43, 48, 58, 61, 65].

### 3 Different Machine Learning Algorithms

ML algorithms were established to reproduce human neural networks in silico in the 1980s. ML is generally composed of three key components: learning algorithms, computational power, and data [18]. As a subset of AI, ML models can be regarded as algorithms that can either self-learn or learn from preset parameters. The main objective is to identify effective variables and the underlying correlation [36, 38]. ML models are normally developed through the following steps, namely, problem identification, goal setting, data collection and sorting, ML model building, validation, assessment of impact, deployment, and monitoring, as well as future modifying [12], as shown in Fig. 3.

The ultimate aim of establishing machine-learning algorithms is to provide optimal personalized decision support of DM management, specifically by developing better closed-loop insulin delivery systems taking into account glycemic variability in DM patients [62].

The value of health-related data to expedite precision medicine development has been well underlined [37, 46, 50]. Therefore, biomarkers and pharmacogenetics parameters may also be incorporated into the ML algorithms to predict management efficacy and responses in patients [19, 31], the onset the progression of the disease course, as well as BG levels [67].

Several factors may influence the eventual clinical implementation during the model design, such as data type and size, model interpretability, and the use of a balancing model. Nevertheless, every type of ML algorithm has its limitations, which

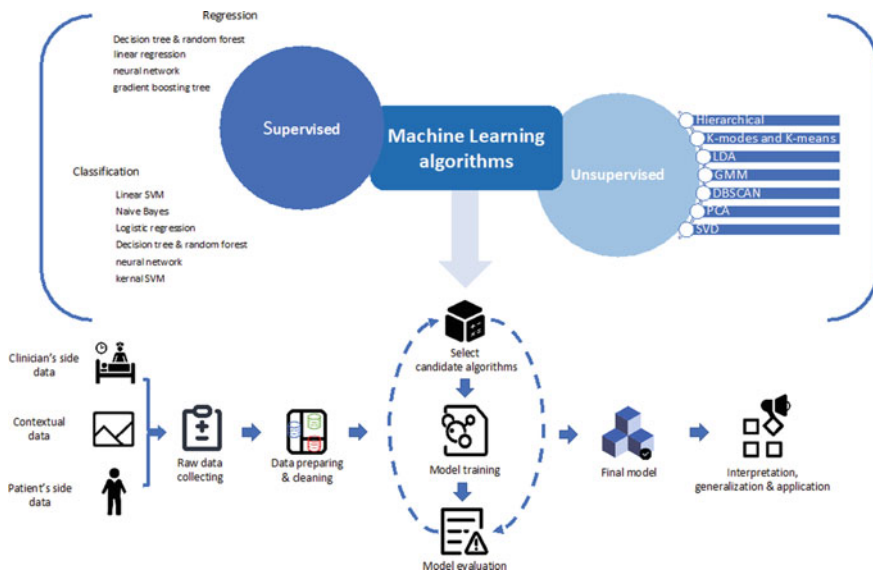


Fig. 3 A graphical summary of the machine learning algorithm process

may only work at full efficacy in specific circumstances. In a systematic review on ML models for community-based T2DM, ANN outperformed all the counterparts, closely followed by logistic regression, decision trees, and random forests [32]. There exists nothing like universally acceptable and ever-winning ML algorithms that fit in every situation. Therefore, to generate relevant and robust results, the currently available ML frameworks should be adjusted in a tailor-made manner to improve further productivity and efficiency [18].

### ***3.1 Artificial Neural Network (ANN)***

An artificial neural network (ANN) is a computational model inspired by biological nervous systems. It comprises various processing elements similar to neurons and axons-like connections called weights [44].

The topology of the ANN could be classified into two main types, namely the feedforward networks and recurrent/feedback networks. The feedforward network is the most used one, where feedback information could be sent back to the former level. In contrast, information could only be sent in one direction (forward) from the earlier stage to the next level in the forward network. Therefore, ANN has excellent efficacy and significant advantages and could adjust to the data flexibly to model and solve a real-world problem.

### ***3.2 Support Vector Machines (SVM) and Gaussian Process Regression***

Support vector machines (SVM), a supervised learning algorithm, have been largely utilized for various purposes, such as identification and recognition of patterns, categorization or classification, regression, and prediction [10]. The use of SVM could minimize the errors incurred by empirical classification.

Support vector regression (SVR) is the most widely used in BG level prediction and modeling among the many SVM algorithms. For instance, Reymann et al. has developed an SVT-based Mobile platform with a radial basis function as a kernel to predict BG levels [45].

Although Gaussian process regression is non-parametric, it could estimate uncertainty and capture noise and smoothness parameters from data input [62]. For instance, Tomczak et al. has reported the feasibility of Gaussian process regression in BG level prediction using categorical inputs such as the type of measurement (e.g., insulin dose, meal intake, physical exercise, pre-prandial BG measurement, and others) [53].



### ***3.3 Decision Tree and Random Forest***

A decision tree (DT) uses a structure built using input features to predict or classify the target outcomes using various input variables. The decision rules could be easily extracted, and hence it is generalized and extended for multiple kinds of application.

Random forests also called random decision forests, serve as an ensemble learning approach for classification and regression applications. It learns through a multitude of decision trees having been constructed, and it can thus directly start feature selection, generating the model of the class or the mean of prediction [24]. Two methods are generally utilized when measuring variable importance, namely the Gini importance index and permutation importance index [2].

For instance, Xiao et al. developed a kind of BG predictor using random forest and support vector regression to evaluate the improved performance gained using a mixed strategy to select an optimal feature pattern [63]. Moreover, Georga et al. predicted the BG levels using random forest regression in a multivariate and multidimensional dataset [17].

### ***3.4 Logistic Regression***

Logistic regression (LR) is generally utilized for classification purposes, and the dependent variable ought to be categorical, owing to its significant role in classification compared to regression. With advantages in robustness and easy handling of non-linear data, the logistic regression could predict the probability of a binary variable (the dummy output variables) based on one or more predictor variables [57].

## **4 An Example of the Application of ML Algorithms Predicting BG Levels in Pregnant Women with GDM in Resource-Limited Regions**

Gestational diabetes mellitus (GDM) is glucose intolerance (hyperglycemia) with the first onset or discovered upon pregnancy. Unmanaged GDM could lead to severe adverse outcomes compromising both mothers and offspring. Nevertheless, pregnant women living in low- and middle-income areas or countries may fail to undergo routine antenatal examinations, leading to a missed diagnosis of GDM. The reluctance to experience the full course of oral glucose tolerance test (OGTT) or the unavailability of a sufficient testing kit may be blamed. To tackle the problems, an AI model that included 9 algorithms was trained using data collected from 12,304 pregnant women from November 2010 to October 2017 who underwent routine prenatal tests in the Obstetrics and Gynecology Department of the First Affiliated Hospital of Jinan University, Guangzhou, China.

The pregnant women's age and fasting blood glucose level were chosen as the critical parameters input for model building. For validation, fivefold cross-validation was conducted for the internal dataset. An external validation dataset constituted with 1655 cases collected from the electronic database of the Prince of Wales Hospital, Hong Kong SAR.

With 9 ML algorithms (SVM, RF, AdaBoost, kNN, NB, decision tree, LR, eXtreme gradient boosting, and gradient boosting decision tree) built, SVM reached the best performance, obtaining an accuracy of 88.7% in the external validation set.

Later, a mobile application was developed, and a prospective and multicenter study was conducted to test the clinical efficacy of the mobile application incorporated with the ML algorithms we developed in GDM screening for pregnant women in resource-limited areas, using only fasting blood glucose value and their age [52]. Although further experiments are needed, this study has provided direct evidence that ML algorithms could, on the one hand, provide a highly accurate diagnosis of undiagnosed patients with high efficacy, and on the other hand, render the cost at an extremely low level. Hence can become an appropriate tool used in the real world instead of merely an algorithm-chasing high performance in silico.

## 5 Outlook

Considering the ML algorithms involved in DM managements, several questions emerge (1) who is using the algorithms; (2) what kinds of data are input in the algorithms, and (3) how is the efficacy and interpretability of the models?

Owing to the "black box"-like low interpretability of ML algorithm, the promotion and further generalized application of ML is doubted, despite that the predictive performance is considerably convincing and promising [20, 30]. Nevertheless, machine learning is only effective when large samples are used due to the input data's multi-dimensionality [3], hence, the studies' small sample size models may be under-estimated. Moreover, ML models devoid of appropriate external validation suffer from limited applicability and extendibility and lacks clinical impact. Even if the ML model is suitable for clinical application, challenges exist in practice due to real-world scenarios' complexity and variability.

Moreover, the user of the ML algorithms matters. Although some ML models could achieve a prominent accuracy and clinical significance level, the data needed to feed the model may not be easily collected and utilized. Therefore, ML algorithm builders should consider the real-world situation and consider the future users of the algorithms instead of an utter inaccuracy chasing. Concerning the future users of the ML algorithms, the source of the parameters for the model building could be both from the patients' side (BG levels, insulin intake, calories intake each diet, exercise, and others) and from clinician's database (bio-physiological parameters, laboratory investigations, ancillary examinations, and others). Moreover, it is also necessary to consider any relevant contextual information, such as intra- and inter variability

among the patient's lifestyle changes, environmental factors, the time series (diurnal vs. nocturnal), and other relevant factors parameters [62].

Therefore, achieving a universal model that accurately predicts and easily collects data from the target population isn't easy. The high accuracy remains controversial if the algorithms were extrapolated to a larger population or a different population. Lacking specific clinical evidence so far, the ML algorithms could still not replace routine diabetes screening and diagnosis and provide clinical suggestions for management for potential DM-related complications predicted. In this sense, future studies should also value the interpretability and applicability of the ML algorithms developed. The assessment of the clinical efficacy and cost-effectiveness of the ML algorithms in the clinics remains urgently needed.

## 6 Conclusion

Machine learning algorithms have been regarded as accurate, with less operation cost and higher efficacy in predicting potential diabetes in undiagnosed populations, profiling personalized BG dynamics, establishing personalized decision support systems, and building BG alarm events in DM patients. However, real-world data concerning the efficacy and cost-effectiveness of various machine learning algorithms clinically is still limited, and internationally acceptable guidelines have not been established to estimate and quantify the potential lifestyle-relevant variables related to BG level.

## References

1. Aikens, J.E., Zivin, K., Trivedi, R., Piette, J.D.: Diabetes self-management support using mHealth and enhanced informal caregiving. *J. Diabetes Complications* **28**(2), 171–176 (2014)
2. Al Mehedi Hasan, M., Nasser, M., Pal, B.: On the KDD'99 dataset: support vector machine based intrusion detection system (ids) with different kernels. *Int. J. Electron. Commun. Comput. Eng.* **4**(4), 1164–1170 (2013)
3. Al-Jarrah, O.Y., Yoo, P.D., Muhaidat, S., Karagiannidis, G.K., Taha, K.: Efficient machine learning for big data: a review. *Big Data Res.* **2**(3), 87–93 (2015)
4. Alić, B., Gurbeta, L., Badnjević, A.: Machine learning techniques for classification of diabetes and cardiovascular diseases. In: 2017 6th Mediterranean Conference on Embedded Computing (MECO). IEEE (2017)
5. Aslam, M.W., Zhu, Z., Nandi, A.K.: Feature generation using genetic programming with comparative partner selection for diabetes classification. *Expert Syst. Appl.* **40**(13), 5402–5412 (2013)
6. Balakrishnan, N.P., Rangaiah, G.P., Samavedham, L.: Review and analysis of blood glucose (BG) models for type 1 diabetic patients. *Ind. Eng. Chem. Res.* **50**(21), 12041–12066 (2011)
7. Barakat, N., Bradley, A.P., Barakat, M.N.H.: Intelligent support vector machines for diagnosis of diabetes mellitus. *IEEE Trans. Inf Technol. Biomed.* **14**(4), 1114–1120 (2010)

8. Bashir, S., Qamar, U., Khan, F.H.: IntelliHealth: a medical decision support application using a novel weighted multi-layer classifier ensemble framework. *J. Biomed. Inform.* **59**, 185–200 (2016)
9. Beloufa, F., Chikh, M.A.: Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm. *Comput. Methods Programs Biomed.* **112**(1), 92–103 (2013)
10. Campbell, C., Ying, Y.: Learning with support vector machines. *Synthesis Lectures Artif. Intell. Mach. Learn.* **5**(1), 1–95 (2011)
11. Chen, J.H., Asch, S.M.: Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N. Engl. J. Med.* **376**(26), 2507 (2017)
12. Chen, P.-H.C., Liu, Y., Peng, L.: How to develop machine learning models for healthcare. *Nat. Mater.* **18**(5), 410–414 (2019)
13. Choi, S.H., Kim, T.H., Lim, S., Park, K.S., Jang, H.C., Cho, N.H.: Hemoglobin A1c as a diagnostic tool for diabetes screening and new-onset diabetes prediction: a 6-year community-based prospective study. *Diabetes Care* **34**(4), 944–949 (2011)
14. Contreras, I., Vehi, J.: Artificial intelligence for diabetes management and decision support: literature review. *J. Med. Internet Res.* **20**(5), e10775 (2018)
15. Dayakar, T., Rao, K.V., Park, J., Sadasivuni, K.K., Rao, K.R.: Non-enzymatic biosensing of glucose based on silver nanoparticles synthesized from *Ocimum tenuiflorum* leaf extract and silver nitrate. *Mater. Chem. Phys.* **216**, 502–507 (2018)
16. Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P.K.R., Srivastava, G.: Deep neural networks to predict diabetic retinopathy. *J. Ambient Intell. Humanized Comput.* 1–14 (2020)
17. Georga, E. I., Protopappas, V.C., Polyzos, D., Fotiadis, D.I.: A predictive model of subcutaneous glucose concentration in type 1 diabetes based on random forests. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE (2012)
18. Gruson, D., Helleputte, T., Rousseau, P., Gruson, D.: Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. *Clin. Biochem.* **69**, 1–7 (2019)
19. Guebila, M.B., Thiele, I.: Model-based dietary optimization for late-stage, levodopa-treated, Parkinson’s disease patients. *NPJ Syst. Biol. Appl.* **2**(1), 1–8 (2016)
20. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. *ACM Comput. Surv. (CSUR)* **51**(5), 1–42 (2018)
21. Guo, Y., Bai, G., Hu, Y.: Using bayes network for prediction of type-2 diabetes. In: 2012 International Conference for Internet Technology and Secured Transactions, IEEE (2012)
22. Guthrie, R.A., Guthrie, D.W.: Pathophysiology of diabetes mellitus. *Crit Care Nurs Q* **27**(2), 113–125 (2004)
23. Haq, A.U., Li, J.P., Khan, J., Memon, M.H., Nazir, S., Ahmad, S., Khan, G.A., Ali, A.: Intelligent machine learning approach for effective recognition of diabetes in e-healthcare using clinical data. *Sensors* **20**(9), 2649 (2020)
24. Hasan, M.A.M., Nasser, M., Ahmad, S., Molla, K.I.: Feature selection for intrusion detection using random forest. *J. Inf. Secur.* **7**(3), 129–140 (2016)
25. Kahramanli, H., Allahverdi, N.: Design of a hybrid system for the diabetes and heart diseases. *Expert Syst. Appl.* **35**(1–2), 82–89 (2008)
26. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I.: Machine learning and data mining methods in diabetes research. *Comput. Struct. Biotechnol. J.* **15**, 104–116 (2017)
27. Kayaer, K., Yildirim, T.: Medical diagnosis on Pima Indian diabetes using general regression neural networks. In: Proceedings of the international conference on artificial neural networks and neural information processing (ICANN/ICONIP) (2003)
28. Kerner, W., Brückel, J.: Definition, classification and diagnosis of diabetes mellitus. *Exp. Clin. Endocrinol. Diabetes* **122**(7), 384–386 (2014)
29. Khazai, N.B., Hamdy, O.: Inpatient diabetes management in the twenty-first century. *Endocrinol. Metab. Clin. North Am.* **45**(4), 875–894 (2016)

30. Kuhn, M.: Building predictive models in R using the caret package. *J. Stat. Softw.* **28**(5), 1–26 (2008)
31. Lee, K., Kim, B., Choi, Y., Kim, S., Shin, W., Lee, S., Park, S., Kim, S., Tan, A.C., Kang, J.: Deep learning of mutation-gene-drug relations from the literature. *BMC Bioinform.* **19**(1), 1–13 (2018)
32. Lee, W. K., Forbes, A., Demmer, R.T., Barton, C., Enticott, J., De Silva, K.: Use and performance of machine learning models for type 2 diabetes prediction in community settings: a systematic review and meta-analysis. *Int. J. Med. Inform.* 104268 (2020)
33. Li, W., Li, Y., Hu, C., Chen, X., Dai, H.: Point process analysis in brain networks of patients with diabetes. *Neurocomputing* **145**, 182–189 (2014)
34. Makino, M., Yoshimoto, R., Ono, M., Itoko, T., Katsuki, T., Koseki, A., Kudo, M., Haida, K., Kuroda, J., Yanagiya, R.: Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Sci. Rep.* **9**(1), 1–9 (2019)
35. Maniruzzaman, M., Kumar, N., Abedin, M.M., Islam, M.S., Suri, H.S., El-Baz, A.S., Suri, J.S.: Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. *Comput. Methods Programs Biomed.* **152**, 23–34 (2017)
36. Mansour, R., Eghbal, K., Amirhossein, H.: Comparison of artificial neural network, logistic regression and discriminant analysis efficiency in determining risk factors of type 2 diabetes (2013)
37. Marjanovic, S., Ghiga, I., Yang, M., Knack, A.: Understanding value in health data ecosystems: a review of current evidence and ways forward. *Rand. Health Quart.* **7**(2) (2018)
38. Mitchell, T.M.: Machine learning and data mining. *Commun. ACM* **42**(11), 30–36 (1999)
39. Nam, S., Chesla, C., Stotts, N.A., Kroon, L., Janson, S.L.: Barriers to diabetes management: patient and provider factors. *Diabetes Res. Clin. Pract.* **93**(1), 1–9 (2011)
40. Oviedo, S., Vehí, J., Calm, R., Armengol, J.: A review of personalized blood glucose prediction strategies for T1DM patients. *Int. J. Numer. Methods Biomed. Eng.* **33**(6), e2833 (2017)
41. Plis, K., Bunescu, R., Marling, C., Shubrook, J., Schwartz, F.: A machine learning approach to predicting blood glucose levels for diabetes management. In: Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence (2014)
42. Polat, K., Güneş, S.: An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Process.* **17**(4), 702–710 (2007)
43. Polat, K., Güneş, S., Arslan, A.: A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square support vector machine. *Expert Syst. Appl.* **34**(1), 482–487 (2008)
44. Reinhardt, A., Hubbard, T.: Using neural networks for prediction of the subcellular location of proteins. *Nucleic Acids Res.* **26**(9), 2230–2236 (1998)
45. Reymann, M.P., Dorschky, E., Groh, B.H., Martindale, C., Blank, P., Eskofier, B.M.: Blood glucose level prediction based on support vector regression using mobile platforms. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE (2016)
46. Russell, S., Norvig, P.: Artificial intelligence: a modern approach (2002)
47. Sagar, R.C., Abbas, A., Ajjan, R.: Glucose monitoring in diabetes: from clinical studies to real-world practice. *Practical Diabetes* **36**(2), 57–62 (2019)
48. Sagir, A.M., Sathasivam, S.: Design of a modified adaptive neuro fuzzy inference system classifier for medical diagnosis of Pima Indians Diabetes. In: AIP Conference Proceedings, AIP Publishing LLC (2017)
49. Sammut, C., Webb, G.I.: Encyclopedia of Machine Learning. Springer Science & Business Media (2011)
50. Sanchez-Pinto, L.N., Luo, Y., Churpek, M.M.: Big data and data science in critical care. *Chest* **154**(5), 1239–1248 (2018)
51. Shah, N., Srivastava, G., Savage, D.W., Mago, V.: Assessing Canadians health activity and nutritional habits through social media. *Front. Public Health* **7**, 400 (2020)

52. Shen, J., Chen, J., Zheng, Z., Zheng, J., Liu, Z., Song, J., Wong, S.Y., Wang, X., Huang, M., Fang, P.H., Jiang, B., Tsang, W., He, Z., Liu, T., Akinwunmi, B., Wang, C.C., Zhang, C.J.P., Huang, J., Ming, W.K.: An innovative artificial intelligence-based app for the diagnosis of gestational diabetes mellitus (GDM-AI): development study. *J. Med. Internet Res.* **22**(9), e21573 (2020)
53. Ståhl, F., Johansson, R., Renard, E.: Ensemble glucose prediction in insulin-dependent diabetes. In: *Data-Driven Modeling for Diabetes*, pp. 37–71. Springer.
54. Stratton, I.M., Adler, A.I., Neil, H.A.W., Matthews, D.R., Manley, S.E., Cull, C.A., Hadden, D., Turner, R.C., Holman, R.R.: Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *BMJ* **321**(7258), 405–412 (2000)
55. Sudharsan, B., Peebles, M., Shomali, M.: Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. *J. Diabetes Sci. Technol.* **9**(1), 86–90 (2014)
56. T, D., Rao, K.V., Bikshalu, K., Malapati, V., Sadasivuni, K.K.: Non-enzymatic sensing of glucose using screen-printed electrode modified with novel synthesized CeO(2)@CuO core shell nanostructure. *Biosens. Bioelectron.* **111**, 166–173 (2018)
57. Tabaei, B.P., Herman, W.H.: A multivariate logistic regression equation to screen for diabetes: development and validation. *Diabetes Care* **25**(11), 1999–2003 (2002)
58. Temurtas, H., Yumusak, N., Temurtas, F.: A comparative study on diabetes disease diagnosis using neural networks. *Expert Syst. Appl.* **36**(4), 8610–8615 (2009)
59. Thatikayala, D., Ponnamma, D., Sadasivuni, K.K., Cabibihan, J.-J., Al-Ali, A.K., Malik, R.A., Min, B.: Progress of advanced nanomaterials in the non-enzymatic electrochemical sensing of glucose and H<sub>2</sub>O<sub>2</sub>. *Biosensors* **10**(11), 151 (2020)
60. Trevitt, S., Simpson, S., Wood, A.: Artificial pancreas device systems for the closed-loop control of type 1 diabetes: what systems are in development? *J. Diabetes Sci. Technol.* **10**(3), 714–723 (2016)
61. Wettayaprasit, W., Sangket, U.: Linguistic knowledge extraction from neural networks using maximum weight and frequency data representation. In: *2006 IEEE Conference on Cybernetics and Intelligent Systems*. IEEE (2006)
62. Woldaregay, A.Z., Årsand, E., Walderhaug, S., Albers, D., Mamykina, L., Botsis, T., Hartvigsen, G.: Data-driven modeling and prediction of blood glucose dynamics: machine learning applications in type 1 diabetes. *Artif. Intell. Med.* **98**, 109–134 (2019)
63. Xiao, W., Shao, F., Ji, J., Sun, R., Xing, C.: 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)*. IEEE (2015)
64. Xu, Q., Wang, L., Sansgiry, S.S.: A systematic literature review of predicting diabetic retinopathy, nephropathy and neuropathy in patients with type 1 diabetes using machine learning. *J. Med. Artif. Intell.* **3**(MI), 6 (2020)
65. Yilmaz, N., Inan, O., Uzer, M.S.: A new data preparation method based on clustering algorithms for diagnosis systems of heart and diabetes diseases. *J. Med. Syst.* **38**, 48–59 (2014)
66. You, Y., Doubova, S.V., Pinto-Masis, D., Pérez-Cuevas, R., Borja-Aburto, V.H., Hubbard, A.: Application of machine learning methodology to assess the performance of DIABETIMSS program for patients with type 2 diabetes in family medicine clinics in Mexico. *BMC Med. Inform. Decis. Mak.* **19**(1), 1–15 (2019)
67. Zarkogianni, K., Litsa, E., Mitsis, K., Wu, P.-Y., Kaddi, C.D., Cheng, C.-W., Wang, M.D., Nikita, K.S.: A review of emerging technologies for the management of diabetes mellitus. *IEEE Trans. Biomed. Eng.* **62**(12), 2735–2749 (2015)
68. Zhang, G., Yuan, J., Yu, M., Wu, T., Luo, X., Chen, F.: A machine learning method for acute hypotensive episodes prediction using only non-invasive parameters. *Computer Methods Progr. Biomed.* **200**, 105845 (2021)
69. Zhou, B., Lu, Y., Hajifathalian, K., Bentham, J., Di Cesare, M., Danaei, G., Bixby, H., Cowan, M.J., Ali, M.K., Taddei, C.: Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* **387**(10027), 1513–1530 (2016)

**Dr. Wai-kit Ming** is an Assistant Professor in Epidemiology and Public Health at the City University of Hong Kong. Before joining the City University of Hong Kong, he was the Head of Community Medicine at the School of Medicine and the Assistant Dean of the International School at Jinan University, Guangzhou, China. He was a visiting research fellow at the University of Oxford and the National Perinatal Epidemiology Unit in the United Kingdom. He served as the CEO of a medical device company and has thirteen national patents for medical devices. His research team has extensive interdisciplinary collaboration experience, including artificial intelligence medicine. He has had a long history of cooperation with the United States, the United Kingdom, Australia, and local universities on national-based clinical research projects. Dr. Ming's work revolves around AI medicine research that has major population health impacts locally and internationally. He has had two scholarly papers published as covers of a journal and over fifty other scholarly papers published, mainly in JCR Q1 and Q2 journals.

**Dr. Zonglin He** having received MBBS from the International School of Jinan University in 2021, Dr. He is now pursuing a research postgraduate degree in the Hong Kong University of Science and Technology, focusing on the research of stem cell biology. He is an experienced researcher with demonstrated history of conducting research in the field of public health, health economics, as well as clinical research and clinical translation of basic medical research. So far, he has published 15 SCI/SSCI articles, with 235 citation (H-index 6).