The Intersection of Machine Learning and Type 1 Diabetes: A Critical Analysis of Machine Learning Driven Innovations

Student No. – 720017170

Project Type – Critical Review

CSC3029 - Declaration of Gen AI-tools

Tick all that apply:

AI-supported use is permitted in this assessment. I acknowledge the following uses of GenAI tools in this assessment:

☐ *I have used GenAI tools for brainstorming ideas.*

☐ *I have used GenAI tools to assist with research or gathering information.*

☐ *I have used GenAI tools to help me understand key theories and concepts.*

☐ *I have used GenAI tools to identify trends and themes as part of my data analysis.*

☒ *I have used GenAI tools to suggest a plan or structure for my assessment.*

☒ *I have used GenAI tools to give me feedback on a draft.*

☐ *I have used GenAI tool to generate image, figures or diagrams.*

☐ *I have used GenAI tools to proofread and correct grammar or spelling errors.*

☐ *I have used GenAI tools to generate citations or references.*

☐ *Other [please specificy]s*

☐ *I have not used any GenAI tools in preparing this assessment.*

☒ *I declare that I have referenced use of GenAI outputs within my assessment in line with the University referencing guidelines.*

List of Contents

List of Figures and Tables

Table of Abbreviations

Abbreviation	Explanation
T1DM	Type 1 Diabetes Mellitus
ML	Machine Learning
CVD	Cardiovascular Disease
CGM	Continuous Glucose Monitoring
EHR	Electronic Health Records
SVR	Support Vector Regression
NN	Neural Network
DL	Deep Learning
RF	Random Forest
XGBoost	Extreme Gradient Boosting
PH	Prediction Horizon
RMSE	Root Mean Squared Error
SHAP	Shapley Additive Explanations
FNN	Feed-forward Neural Network
EGA	(Clarke's) Error Grid
FCNN	Fast-adaptive and Confident Neural Network
RNN	Recurrent Neural Network
MAML	Model-Agnostic Meta-Learning
GRU	Gated Recurrent Unit
AUC	Area Under Curve
NHANES	National Health and Nutrition Examination Survey
HbA1c	Glycosylated Haemoglobin
EASD	European Association for the Study of Diabetes
ADA	American Diabetes Association
LIME	Local Interpretable Model-Agnostic Explanations
TIR	Time in Range
HCL	Hybrid Closed Loop
HIPAA	Health Insurance Portability and Accountability Act
GDPR	General Data Protection Regulation
MHRA	Medicines and Healthcare Products Regulatory Agency
SD	Standard Deviation

1. Lay Abstract

People with Type 1 Diabetes Mellitus (T1DM) are required to monitor and control their blood sugar levels to avoid immediate risks like low blood sugar levels and long-term complications, including heart disease. This review examines how machine learning (ML) models can support this by predicting blood sugar changes, identifying risks, and powering systems that automatically manage insulin delivery. Different models are suited to various scenarios: neural networks (NN) work best for short-term sugar level predictions, while models like support vector regression (SVR) are more stable over more extended periods. Deep learning methods can be very accurate but are often too complex for clinical use, which may delay adoption. Automated insulin systems using two hormones show the most promise for controlling blood sugar, but they are expensive and not easy to roll out in everyday care. Future work should aim to make ML tools easier to understand, reduce costs, and ensure they work reliably across different patient groups.

2. Scientific Abstract

Background: Machine learning (ML) has become more prevalent in managing Type 1 Diabetes Mellitus (T1DM), offering ways to predict glucose trends, detect risks, and improve insulin delivery.
Methods: This critical review explored studies from 2015 to 2025 across four key areas of T1DM management: glucose prediction, complication risk, closed-loop systems, and ethical considerations.
Results: Neural networks (NN) perform best for short-term glucose prediction, while support vector regression (SVR) offers more reliable results over longer time horizons. Models like extreme gradient boosting (XGBoost) have shown strong results in hypoglycaemia prediction (AUC = 0.99), and random forests (RF) have effectively predicted cardiovascular disease (CVD) and T1DM onset. Deep learning (DL) models outperform others in detecting diabetic retinopathy, but a lack of transparency limits their applications. Closed-loop systems that deliver insulin and glucagon provide better glucose control than single-hormone systems (TIR = 93.1%), though expensive and hard to scale. Key barriers include model complexity, limited training data, and regulatory constraints.
Conclusion: ML has the potential to support a range of T1DM applications, but challenges remain in making systems explainable, adaptable, and suitable for routine care. Improving trust, reducing computational demands, and validating models across diverse datasets will be essential for future progress.

3. Introduction

Type 1 Diabetes Mellitus (T1DM) is a chronic autoimmune disease resulting in the immune-mediated destruction of insulin-producing pancreatic β-cells (1). This loss of insulin function requires lifelong exogenous insulin replacement to regulate blood glucose (1). Around 8.4 million live with T1DM globally, with numbers expected to increase (2). Short-term complications include hypoglycaemia, hyperglycaemia and diabetic ketoacidosis — a build-up of ketones in the blood leading to a drop in blood pH (3, 4). Long-term risks such as cardiovascular disease (CVD), accounting for 35% of T1DM mortality (5), and retinopathy can significantly affect quality of life (6).

Management traditionally involves continuous glucose monitoring, insulin therapy, and behavioural adjustments. Technologies like insulin pumps and continuous glucose monitors (CGMs) have

improved accuracy and delivery, but consistent glucose control remains difficult for many patients (7). This has led to interest in tools supporting more adaptive and personalised care.

Although medical devices are widely discussed, ML systems have received far less attention in T1DM reviews. There is limited discussion around ML systems designed to predict glucose levels, assess risk, or automate insulin delivery. ML refers to how an artificial intelligence (AI) system acquires the ability to learn and improve from experience and observed data to predict new or unseen cases. As a branch of AI, it enables technologies to perform tasks typically associated with human reasoning or decision-making (8).

These methods are already being applied to data from CGMs, electronic health records (EHR), and wearable technology (9). Unlike traditional models, they can detect complex trends in behaviour and physiology that traditional systems may miss (10, 11), potentially leading to better care.

Recent studies highlight the potential of ML methods to predict glucose changes, detect risks, and manage insulin delivery through closed-loop systems (12, 13). Other models show promise for identifying complications like CVD and retinopathy earlier than standard methods (14, 15). However, clinical use remains limited by problems such as poor interpretability, inconsistent data, and lack of transparency (16, 17, 18).

As more people are diagnosed and healthcare systems face increasing demand, there is growing interest in using data-driven approaches to improve T1DM management. This review explores the use of ML in T1DM across three key areas: glucose prediction, risk modelling, and automated insulin delivery, with attention to the ethical and practical barriers to broader adoption. This provides a comprehensive insight into ML-driven strategies for future T1DM management.

4. Methods

4.1 Search Strategy

The search strategy applied within this critical review explores relevant literature within major academic databases, including PubMed, IEEE, and Google Scholar, using key terms such as “predictive analytics AND diabetes management” (81 results), “machine learning AND Type 1 diabetes” (45 results), “glucose AND prediction models” (979 results), “machine learning AND closed-loop insulin delivery” (7 results) and “machine learning AND automated treatment of diabetes” (166 results). The databases were selected for their comprehensive coverage of peer-reviewed biomedical and healthcare research. Boolean logic operators (e.g., “AND” and “OR”) were applied to refine the search, while filters restricted results to studies published from 2015 to 2025 to discover the most up-to-date research.

4. 2 Inclusion and Exclusion Criteria

Inclusion criteria focused on studies that:

- Were published within the last 10 years.
- Were in English
- Directly discussed ML or data science techniques in T1DM management.
- Applied robust methods, e.g. clinical trials, systematic reviews, and meta-analyses.

Exclusion criteria eliminated studies that:

- Were published before 2015
- Not accessible through the University of Exeter or open-source

4. 3 Quality Assessment

To determine the quality of the literature specific criteria were used, focusing on study design, sample size and strength of the analysis. Clinical trials and systematic reviews were given higher priority due to the robust nature of their methods and ability to generate high-level conclusions. Studies with larger sample sizes and diverse populations were favoured due to their increased statistical power and enhanced application of their findings. Additionally, studies containing rigorous statistical methods, such as cross-validation for ML models and multicentre clinical trials, provided more reliable and valid evidence. This method ensures that the studies included in the review contain the best available evidence on the application of ML in T1DM management.

4. 4 Thematic Organisation

The reviewed literature was categorised into four primary thematic areas: blood glucose concentration prediction, short-term and long-term risk prediction, automated treatment systems, and ethical and practical considerations. This thematic structure allows for an in-depth review of ML-driven innovations in T1DM management, focusing on critical areas where ML has made the most significant impact and allowing for a robust structure for the review.

5. Discussion

5.1 Blood Glucose Concentration Prediction

Blood glucose concentration prediction is a key area of research when considering the applications of ML in T1DM management. It allows for early interventions that aid patients in avoiding hypoglycaemic and hyperglycaemic events, reducing the risk of severe complications (19). ML models using CGM data have demonstrated promising results in short-term glucose prediction; however, differences in model architecture, feature selection, and dataset variability significantly impact predictive performance and can affect long-term performance (12). Accurate long-term prediction is more clinically challenging, as it must account for complex physiological dynamics such as insulin pharmacokinetics, dietary intake, physical activity, stress, and circadian rhythms, factors not easily inferred from CGM data alone (20). This section critically reviews the performance of ML models: SVR, NNs, and DL when used for blood glucose prediction, focusing on their strengths, limitations, and clinical application.

5.1.1 Support Vector Regression

SVR is a method that constructs a hyperplane with the maximum margin between data points (21). It is commonly used for protein classification, gene expression analysis, and disease diagnosis (21). However, SVR applications within blood glucose prediction have been explored throughout the literature.

Hamdi et al. (20) investigated their application by building a model using SVR trained on CGM data. The model was evaluated across four prediction horizons (PH): 15, 30, 45 and 60 minutes, achieving root mean squared errors (RMSE) of 0.52, 0.60, 0.66, and 0.72 mmol/L, respectively (20). These

results indicate SVR's high accuracy on short-term predictions; however, performance declined when predicting longer-term blood glucose concentrations. Additionally, the study's small sample size (12 patients) limits statistical power and generalisability and could cause overfitting.

While SVR models are known for their interpretability and computational efficiency, their static nature can limit flexibility in modelling the dynamic patterns seen in blood glucose regulation (21). Compared to more adaptive models like NNs, SVR may underperform when used for longer-term prediction, where physiological changes such as insulin absorption or activity levels become harder to capture (20). These limitations are amplified when models are trained on smaller datasets, increasing the risk of overfitting, where the model becomes too tailored to the training data and struggles to generalise to unseen cases (22). Future research should consider optimising input features, expanding datasets, and applying regularisation methods to improve reliability. Incorporating interpretability tools such as Shapley additive predictions (SHAP) may also help close the gap between model outputs and clinical decision-making, giving practitioners greater insight into how predictions are formed and improving trust in the model's use (16).

5.1.2 Neural Networks

Although SVR shows some clinical applications, NNs have been widely researched in this area and could address the limitations of SVR. NNs are computational models containing interconnected processing units called neurons (23). Weights, a determinant of each connection's strength, connect neurons (23). Feedforward networks, characterised by their unidirectional data flow (23), have been extensively researched due to their ability to process sequential glucose trends and learn hidden correlations in CGM data (12).

A study by Allam et al. (24) developed a three-layer NN, illustrated by Figure 1, trained on CGM data and evaluated using three PHs: 15, 30 and 45 minutes (24). The model achieved RMSE of 0.15, 0.42, and 0.83 mmol/L across the PHs (24), outperforming SVR across PH = 15 and 30 but performing worse over PH = 45, indicating that SVRs are more applicable for long-term blood glucose prediction (PH > 30), contradicting the conclusions by Hamdi et al. (20, Table 1). However, the highly accurate NN performance may reflect limits in dataset composition and sample size, as another study using a feedforward neural network (FNN) concluded RMSE of 1.66 and 2.51 mmol/L when evaluated over 30 and 60-minute PHs and tested on a dataset of 451 patients rather than 9 (25). However, this study did not address clinical acceptability. To determine the clinical applicability of Allam et al.'s model (24), the predictions were evaluated using Clarke's Error Grid (EGA) (26), illustrated in Figure 2. This evaluation demonstrated that at PH = 15 and 30, all outputs were in zones A and B, the clinically accepted zones. However, at PH = 45, 2.7% of predictions fell into zone D, indicating a higher risk of incorrect treatments and detections than at PH = 15 and 30 (24). Since predictions with RMSE below 0.83 mmol/L aligned with zones A and B, this threshold may serve as an informal benchmark for clinical acceptability in similar models. However, this threshold is an estimation and not yet standardised in clinical ML guidelines.

Allam et al.'s (24) findings indicate that NNs lag behind SVR in long-term glucose prediction due to consistently higher RMSE at extended horizons. However, evaluating SVR using Clarke's EGA is an area for future research. Assessing the models across different PHs and methods is essential for determining their clinical prediction ability. While Clarke's EGA validates clinical acceptability, the

NN's lack of confidence intervals limits interpretability, a key barrier to real-world deployment. This challenge limits its application in clinical settings, as healthcare providers may hesitate to rely on predictions they cannot fully comprehend. Furthermore, despite their promising predictive abilities, NNs face challenges such as requiring large, diverse training datasets to generalise well (12, 20) and the risk of introducing biases from limited or skewed data sources (12), suggesting an area for future research where more diverse and larger data sets are implemented in clinical trials.

Future models should incorporate techniques such as SHAP, which can offer insights into how specific inputs influence the model's predictions and improve trust in its use in clinical settings (16). Additionally, addressing computational demands through optimisation techniques, such as pruning or quantisation, could facilitate integration into low-resource settings, e.g. closed-loop systems (27).

5.1.3 Deep Learning

Although NNs provide a solid foundation for blood glucose concentration prediction, their clinical application is often constrained by limited interpretability and small training datasets (12). DL, a subset of ML using multilayered NNs to simulate the decision-making of a human brain (28), is another branch of ML used in the literature for blood glucose concentration prediction. To address key limitations found in NNs and SVR (20, 24), Zhu et al. (12) created the fast-adaptive and confident neural network model (FCNN). This FCNN combines an attention-based recurrent neural network (RNN) with evidential regression and model-agnostic meta-learning (MAML). The architecture, shown in Figure 3, integrates bidirectional gated recurrent unit (GRU) layers with attention to extract features from CGM data and inputs, like meals and insulin, to make accurate predictions. The evidential regression layer generates confidence intervals for the predictions, and MAML allows the model to adapt quickly to new T1DM patients with minimal data.

The FCNN (12) was evaluated on three clinical datasets: an eight-week data set with 12 T1DM subjects, six-week clinical trial data from 12 subjects with insulin pumps of daily injections and a six-month data set from 35 subjects with extensive activity records (12). Accuracy was evaluated using RMSE over 30 and 60-minute PHs (12). The model performed best on the first dataset, achieving an RMSE of 1.03 mmol/L at PH = 30, indicating moderate short-term prediction accuracy, but at PH = 60, the RMSE increased to 1.39 mmol/L (12).

The FCNN produced higher mean RMSE than the NN and SVR (20, 24, Table 1), exceeding the 0.83 mmol/L clinical acceptability threshold inferred from Clarke's EGA evaluation of Allam et al.'s NN (24). This indicates lower predictive accuracy and a greater likelihood of clinically unacceptable glucose estimates, which could increase the risk of incorrect treatment decisions. Another study by Jaloli et al. (30) showed that DL methods perform better than SVR and NNs, as illustrated in Table 1. However, Jaloli's DL model has a very different architecture from Allam et al.'s, which could cause performance differences (24, 30). This is potentially attributed to incorporating interpretability tools in the FCNN, which can lead to decreased performance based on previous studies of interpretability-performance trade-offs (29).

Despite potentially causing the FCNN's lower performance, integrating confidence intervals improves its real-world applicability by providing clinicians with a sense of how certain or uncertain each prediction is. This added layer of transparency helps practitioners interpret the results more

confidently and reduces the risk of relying on potentially misleading outputs. Furthermore, only training on CGM data without contextual variables, such as physical activity, limits the model's predictive accuracy, suggesting that incorporating additional inputs could improve robustness and adaptability across diverse patient populations.

5.1.4 Comparisons and Future Directions

All three ML approaches demonstrate potential in T1DM management, but key differences affect their clinical utility. NNs, particularly Allam et al.'s model (24), show strong short-term prediction performance, reflected in the lowest mean RMSE and clinical acceptability (Table 1). However, accuracy drops at longer PHs, and a lack of interpretability limits real-world use. While slightly less accurate in the short term, SVR outperforms NNs in long-term predictions and offers greater explainability, which is important for clinical decision-making. DL models provide flexibility and built-in confidence estimation but generally show higher RMSE, possibly due to model complexity and variation in training data.

This review proposes a hybrid approach that combines SVR for long-term trend forecasting with an NN for short-term predictions. This approach draws on each method's complementary strengths to improve both predictive reliability and adaptability in clinical settings. Future studies should evaluate how such systems perform across varied patient groups and real-world scenarios.

To support deployment, explainability tools like SHAP and LIME should be tested alongside clinical intuition in prospective studies (16). Federated learning should be explored to allow model development across multiple sites while preserving data privacy (31). Finally, using standardised reporting frameworks, such as TRIPOD-AI, would improve transparency, reproducibility, and readiness for clinical implementation (32).

5.2 Short-Term and Long-Term Risk Prediction

Acute risks such as hyperglycaemia and hypoglycaemia can escalate quickly and lead to more complex health implications (3), such as diabetic ketoacidosis, CVD, and retinopathy (1, 2). This indicates that hypoglycaemia prediction, early diagnosis, and the prediction and monitoring of further complications are important study areas in T1DM management. ML has been extensively deployed in the literature to create tools that can process complex datasets and deliver accurate predictions, leading to earlier intervention. This section evaluates current ML methods in predicting hypoglycaemia, T1DM onset and its associated long-term complications.

5.2.1 Short-Term Risk Prediction

Unlike general blood glucose prediction, which estimates continuous glucose concentrations over time, hypoglycaemia prediction focuses on identifying the probability that a patient's blood glucose concentration will fall into hypoglycaemia (33). Hypoglycaemia remains one of the most serious and immediate risks for individuals with T1DM, with patients averaging two symptomatic episodes per week and one severe episode, often with seizure or coma, per year (3). Reddy et al. (33) address this issue by developing and evaluating two algorithms, a decision tree and an RF, to predict hypoglycaemia at the start of exercise. Both models were trained using a dataset containing 154 observations of in-clinic aerobic exercise in 43 adults with T1DM from three studies (33). Evaluation,

using area under the curve (AUC), revealed an AUC of 0.79 for the decision tree, and the RF achieved a higher AUC of 0.84 (Table 2), showing considerable accuracy (34, Table 3). The increased performance of the RF could be attributed to its ability to handle complex, non-linear interactions between features and reduce overfitting by aggregating multiple decision trees (33), indicating that RFs may be more suitable for deployment within wearable technology for use during exercise. However, only focusing on aerobic exercise limits generalisability.

The limitations of the previously mentioned study are addressed in a paper by Duckworth et al. (35), who used extreme gradient boosting (XGBoost), a gradient-boosted decision tree utilising gradient descent (36), to predict hypoglycaemic events 60 minutes before their occurrence. This model achieved an AUC of 0.99 (35), indicating that XGBoost has excellent accuracy (34, Table 3) and outperforms decision trees and RFs when predicting hypoglycaemic events (Table 2), potentially due to increased algorithmic complexity. In contrast, a review by Mellor et al. (37) found that XGBoost displayed an AUC of 0.84, the same as the RF model, indicating the increased precision may be partially attributed to dataset-specific optimisations and simplicity of feature selection, raising concerns about overfitting when applied to larger, more diverse populations. Future research should incorporate variables such as the average daily risk range or % coefficient of variation to improve the feature selection used to represent glycaemic variability (35).

Duckworth et al. (35) address concerns with interpretability by integrating SHAP to identify which features have been most important in predicting risk, allowing clinical practitioners to develop early interventions and reduce their impact (16). Conversely, Reddy et al. (33) did not incorporate interpretability tools, limiting the clinical application of their findings, as models with strong interpretability are more likely to gain regulatory approval and clinician adoption.

Although the application of ML in short-term risk prediction is apparent within the literature, one consistent challenge is the occurrence of false positives, causing unnecessary alerts and potential mistreatment, which could cause more severe complications (38). Reddy et al. (33) reported 13% error rates, potentially attributed to individual variability or noise in the data. False negatives are very problematic, as missed hypoglycaemic events can lead to severe complications. Interpolation of CGM data may be a method to reduce false negatives. Russon et al. (39) found that CGM recording intervals of 15 minutes caused 10.6% of Level 2 hypoglycaemic episodes to be missed. Although interpolation reduced this to 3%, it also increased false positives, highlighting the delicate trade-off between model sensitivity and alert precision.

Approaches such as threshold tuning, reinforcement learning-based alert suppression, and hierarchical decision models should be considered to minimise alert fatigue while maintaining sensitivity (40). Another limitation that reduces the application of ML in short-term risk prediction is ML models' computational demands. Future models should focus on developing lightweight architectures or utilising edge AI to improve real-time performance while conserving device resources.

5.2.2 Long-Term Risk Prediction

In addition to their use in short-term risk prediction, ML approaches have also been extensively studied for long-term risk prediction tasks, including the early detection of T1DM onset. Precise

prediction of T1DM onset facilitates early intervention and identification of risk factors before irreversible destruction of β-islet cells, which can significantly improve T1DM management (41). A study by Dinh et al. (14) investigated the performance of XGBoost in predicting T1DM onset. Their model was built using the National Health and Nutrition Examination Survey (NHANES), which contains 123 variables, combined with lab data. The model achieved an error rate of 0.26 when tested on individuals without T1DM, indicating that it is moderately accurate when predicting T1DM outcomes (14). Although this model has a higher error rate when compared to other models (Table 2), Dinh et al. (14) improved the clinical application of their model by presenting feature importance data, allowing clinical practitioners to identify the most important risk factors and develop early intervention strategies. The model being trained on survey data combined with clinical data strengthens its clinical application, making it highly suitable for integration with EHR systems to support large-scale T1DM health screening. However, the NHANES dataset contains data from all diabetic patients, but the study demonstrates the ability of these models to predict diabetes onset, which could transfer over to T1DM.

Alazwari et al. (42) address the limitations of the previously mentioned study by developing two models, multiple linear regression (MLR) and an RF, to predict the age at onset of T1DM within a population of 359 individuals across three diabetes clinics, improving generalisability due to its multicentre nature. The MLR and RF achieved error rates of 0.19 and 0.11 (42), indicating that the MLR was moderately accurate, with the RF proving to be highly accurate when predicting the age of onset of T1DM and outperforming the XGBoost (14, Table 2). Contrastingly, a study concluded that XGBoost had a higher accuracy when predicting T1DM, concluding an error rate of 0.18 compared to 0.21 from the RF (43). This may be due to the lack of balance in the data, causing skewed results and highlighting the impact of dataset composition on model performance. Additionally, this paper was not peer-reviewed, which could limit the robustness of the conclusions. Research in this field is important to medical practitioners as it means individuals can be screened, allowing for early intervention programmes to be developed to avoid the long-term complications associated with T1DM. However, the research reported that additional variables, such as genetic risk factors, should be included in the analysis (42). Also, the multicentre nature of Alazwari et al.'s (42) study may impact data security, highlighting the potential need for federated learning as it can collaboratively train models while maintaining data privacy (31).

Although ML models have successful applications in predicting T1DM onset, ML methods are more commonly used when predicting T1DM progression and long-term complications such as CVD and retinopathy, allowing for early interventions that could reduce patient mortality. Dinh et al. (14) investigated the use of XGBoost, using a combination of survey data from the NHANES and lab data, to predict CVD in patients with T1DM and indicate the most influential predictors of CVD. After evaluation, this model achieved an AUC of 0.84 (14), which indicates considerable classification accuracy, with the most important predictors being age, LDL cholesterol and chest pain (14, 34, Table 3). This approach shows vast clinical applications as practitioners can use the important features to develop intervention plans, causing a reduction in CVD occurrences among patients with T1DM, leading to lower mortality rates. Conversely, Akella (44) found that support vector machines (SVM), k-nearest neighbours (KNN) and RFs achieved higher AUC than the XGBoost; 0.87, 0.85, and 0.88, illustrated in Table 2, highlighting the impact of model architecture on prediction accuracy. However, this was performed on a different dataset with no direct comparison to XGBoost,

highlighting an area for future research where multicentre clinical trials should be performed comparing the performance of SVM, KNN, RF and XGBoost on CVD prediction in T1DM patients.

Another complication researched in the literature is retinopathy and its progression. Rabhi et al. (45) developed a temporal DL framework, which analysed glycosylated haemoglobin (HbA1c) to predict retinopathy. This model achieved an AUC of 0.88 (45), demonstrating the capability of DL to identify complex, non-linear patterns within HbA1c data, achieving considerable predictive performance (34, Table 2). In contrast, another DL model displayed superior performance, achieving an AUC of 0.95 (15), as shown in Table 2. This may be attributed to the differences in methods, as this study used eye fundus images to achieve excellent prediction accuracy (15, 34, Table 3). However, the test dataset in this study was unbalanced and negatively skewed, raising potential concerns about overfitting (15). Therefore, cross-validation techniques or regularisation methods should be used to reduce overfitting. These studies display the potential of DL to identify retinopathy. However, future research could incorporate DL into smartphone fundoscopy, reducing the burden on healthcare institutions as the prevalence of T1DM increases and improving at-home diagnostic capabilities (46).

5.2.3 Future Directions

Across studies, model performance varied by task. XGBoost performed best in hypoglycaemia prediction (AUC = 0.99), while RFs excelled in T1DM onset tasks (error rate = 0.11) and CVD prediction (AUC = 0.88), indicating that future research should prioritise task-specific model selection. Future research should also focus on improving generalisability and reducing overfitting. For example, Duckworth et al.'s (35) XGBoost achieved excellent accuracy for hypoglycaemia prediction, but its dataset-specific optimisations raise concerns about overfitting. Similar issues were noted by Rabhi et al. (45) for retinopathy prediction.

Implementing k-fold cross-validation, L1 or L2 regularisation, and testing models on independent datasets could improve reliability. Instead of relying on single-institution datasets, researchers should conduct multicentre trials where ML models are tested on diverse patient populations across multiple hospitals, including federated learning, to maintain data security (31). However, communication latency and model drift must be addressed (31).

For clinical adoption, improving interpretability is essential. Tools like SHAP or Local Interpretable Model-Agnostic Explanations (LIME) could increase trust and support regulatory approval (47). Additionally, lightweight architectures involving fewer features could enhance real-time performance, allowing better incorporation into wearable technology, as Russon et al. (48) demonstrated that glycaemic events during exercise could be effectively predicted using only two features: start glucose and exercise duration.

5.3 Automated Treatment via Closed-Loop Systems

Hyperglycaemia and diabetic ketoacidosis pose challenges for T1DM management, as both can result in hospitalisation, cognitive impairment, or, in severe cases, death (1). Closed-loop systems aim to address this by combining data from CGMs with advanced ML algorithms to adjust insulin delivery automatically, leading to a decrease in hyperglycaemia due to more accurate control (49, 50). These approaches can potentially reduce patient intervention, which could be crucial for children or the elderly who cannot adhere to traditional T1DM management methods strictly. Also,

these systems can tailor insulin dosing to physiological needs and improve time in the normoglycemic range (TIR), a key metric for glycaemic control (51). This section critically reviews the literature evaluating the applications of closed-loop systems within T1DM management.

5.3.1 Single-Hormone Closed-Loop Systems

Closed-loop systems have been extensively researched in the literature. Benhanmou et al. (52) conducted a 12-week, multicentre randomised crossover trial with 68 adults, comparing an ML-driven hybrid closed-loop (HCL) system to sensor-assisted pump therapy. The analysis revealed that the individuals using the HCL system saw a 68.5 (± 9.4) % TIR, with 29.5 (± 10.2) % spent in the hyperglycaemic range (>10 mmol/L), both showing significant improvements from sensor-assisted pump therapy (59.4 (± 10.2) % vs 68.5 (± 9.4) %, 36.3 (± 10.2) % vs 29.5 (± 10.2) %; $p < 0.0001$) (52). An HbA1c reduction of 0.3% was observed between the sensor-assisted pump therapy group and the HCL group; however, this was not statistically significant (-0.14 (± 0.6) % vs -0.29 (± 0.6) %; $p > 0.05$) but shows potential for long-term glycaemic control (52). Despite promising results, the study's short 12-week duration and limited sample size restrict its applicability to broader populations and long-term T1DM management. Another limitation stated within the paper was that the study lacked an appropriate control for a proper assessment of remote monitoring and that there was no evaluation of psychological and human factors (52).

McAuley et al. (13) expanded on and supported Benhamou et al.'s (52) research by conducting a larger 26-week study with 120 adults. As illustrated by Table 4, their results demonstrated an improvement in TIR to 69.9 (± 9.5) % and a reduction in hyperglycaemic time to 27.6 (± 9.5) %, both significantly different to the control (54.7 (± 12.7) % vs 69.9 (± 9.5) %, 40.3 (± 14.4) % vs 27.6 (± 9.5) %; $p < 0.0001$), further supporting the efficacy of HCL systems (13). A 0.4% HbA1c reduction was experienced between the groups. In contrast to the previously mentioned study (52), it was seen to be statistically significant (7.4 (± 0.8) % vs 7.0 (± 0.6) %; $p < 0.0001$), indicating an improvement in long-term glycaemic control due to it reflecting the average blood glucose concentration over the last 2 to 3 months (53). Furthermore, while the study population was more diverse, it only consisted of adults, meaning the efficacy may differ when used on other demographics. This indicates that further improvements in sample size and diversity are required to enhance generalisability and further reduce the hyperglycaemic time. Also, due to self-reported adherence, data bias could be introduced, and it may not reflect the accurate use of the HCL systems.

5.3.2 Dual-Hormone Closed-Loop Systems

While McAuley et al. (13) and Benhamou et al. (52) showed improvements in TIR and hyperglycaemic time, HCL systems have seen considerable advancements. Dual hormone systems integrating insulin delivery with glucagon administration potentially improve the TIR from single-hormone systems (54). These devices could fully automate the treatment of T1DM and remove the patient's burden by mimicking the body's natural counter-regulatory mechanisms.

Dual-hormone closed-loop systems have been deployed within the literature by Zhu et al. (54), who incorporated deep reinforcement learning to enhance insulin and glucagon delivery using the input of glucose and meal carbohydrates. The dual-hormone system outperformed the single-hormone HCLs, achieving a TIR of 93.1 (± 4.5) % among adults (54), as illustrated in Table 4. This represents a

23.1% improvement above the clinical benchmark set by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA), which recommend a TIR of ≥70% for individuals with T1DM to reduce the risk of complications (55). This indicates that the dual hormone system is more adept at regulating blood glucose concentrations, substantially reducing the risk of acute and chronic T1DM complications. The time spent in a hyperglycaemic range was reported to be 5.63 (± 3.29) % (54), almost five times lower than the 27.6 (± 9.5) % reported by the single-hormone system (13). The improvement in TIR and hyperglycaemic time is likely due to the advanced reinforcement learning algorithms employed in dual-hormone systems. These models dynamically adjust insulin and glucagon delivery in response to real-time glucose fluctuations, reducing both hyperglycaemic and hypoglycaemic episodes (54).

However, another study showed no significant differences between single and dual hormone systems in glycaemic control, and interestingly, around exercise and during the night, the single hormone system performed better (56). These differences in results may be due to limitations within the sample size of the study by Lindkvist (56), which consisted of only 11 adolescents. Furthermore, the discrepancy between Zhu et al. (54) and Lindkvist (56) highlights the limitations of in silico studies. While Zhu et al. (54) demonstrated exceptional TIR improvements, these results were based on computer simulations rather than clinical trials, which may not account for real-world physiological complexities such as variability in patient adherence, diet, and stress levels, identifying an area of future research where clinical trials can be conducted with these features included.

5.3.3 Future Directions

Future studies should investigate closed-loop systems in more diverse clinical trials, potentially involving children and the elderly, as this demographic could see the most benefits; however, this may introduce some ethical concerns. This could be implemented using multicentre studies to enhance generalisability across varied environments and behaviours (57). Trials across wider durations should be used to better evaluate these systems' long-term efficacy, allowing for more accurate conclusions on the application in T1DM management. Furthermore, cost-effective alternatives or non-invasive technologies should be explored by simplifying hardware, reducing the need for calibration and addressing the high cost (58).

Advancing ML algorithms to improve adaptability and interpretability will also be key. Explainable AI methods, such as SHAP or LIME, can build trust among healthcare providers, while adaptive learning models can enhance personalisation and system performance by minimising false alarms and optimising hormone delivery (47, 49).

5.4 Ethical and Practical Considerations

While ML's technical potential in T1DM management is well established, its ethical implementation in clinical contexts remains a critical area of concern. Concerns involve data privacy and security, algorithmic biases, model interpretability and transparency, and regulatory and ethical compliance.

5.4.1 Data Privacy and Security

Building ML models to assess areas of T1DM management requires vast amounts of sensitive patient data, which requires enhanced security measures to prevent data breaches, patient distrust, and privacy concerns (59).

Specific measures have been introduced to avoid these issues, such as the Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR); these guidelines outline standards that healthcare institutions must follow to manage data responsibly and transparently (59). Other methods have been developed, such as Qu et al. (60), who created a federated learning method that allows for decentralised data sharing without impacting privacy, making multicentre trials more accessible. However, some argue that federated learning may reduce model accuracy due to fragmented data training (61). These strategies indicate the importance of innovative data security and privacy approaches in developing ML within T1DM management.

5.4.2 Algorithmic Bias

Many studies within T1DM management using ML methods contain limitations in sample size, population diversity or dataset imbalance. These introduce algorithmic bias as issues can arise from datasets which inadequately represent diverse demographics, leading to skewed predictions and potential inequities in patient care (62). For example, models trained solely on data from homogeneous populations may fail to generalise to underrepresented groups, exaggerating health disparities and reducing the effectiveness of models (63). Also, bias within algorithms can be substantial, leading to increased false negatives, resulting in undiagnosed cases, leading to long-term complications, and delayed T1DM treatment (64).

These challenges require addressing for the seamless application of ML-driven innovations within T1DM management. For example, Friedler et al. (65) compared several fairness-enhancing algorithms, highlighting strategies that could be adapted to ensure unbiased healthcare outcomes in T1DM management. Furthermore, Agarwal et al. (66) highlight bias-mitigation techniques such as adversarial debiasing, which rebalances datasets by adjusting weight distributions in underrepresented subgroups, reducing false negative rates in minority patients with T1DM. However, debiasing techniques like adversarial debiasing can introduce instability and are still being evaluated in clinical datasets (67). These methods could ensure that ML technologies contribute to fair and effective T1DM management for all patients.

5.4.3 Model Interpretability and Transparency

ML models, for example, DL, operate as "black boxes" caused by their complex prediction algorithms (16). This leads to issues concerning interpretability and transparency, as clinicians may not understand the conclusions, leading to distrust and limiting their applications within T1DM management (68).

To address these challenges, models should incorporate SHAP or LIME, as these indicate the contribution of individual features to a model's predictions (47). Furthermore, some studies use attention mechanisms to improve their interpretability by allowing models to prioritise relevant data features, making decision-making processes more transparent (69). Finally, the model developed by

Zhu et al. (12) shows a good example of how DL can be used to obtain accurate predictions and output confidence intervals, which clinicians can interpret, giving them more trust in the projections.

5.4.4 Regulatory and Ethical Compliance

Before being applied to T1DM management, ML-driven innovations must meet stringent safety and performance requirements set by regulatory bodies such as the Medicines and Healthcare Products Regulatory Agency (MHRA), which oversees the approval of AI-based medical technologies in the UK (70). Technologies must follow ethical principles such as respecting patient autonomy, ensuring informed consent, and promoting transparency concerning the predictions (71). For example, there are many risks at play when using ML-driven innovations, such as false negatives, which can have immense impacts (72) and potentially inaccurate hormone administration; these issues must be discussed with the patient before they are used in practice. Furthermore, ensuring equal access to these technologies is vital for preventing disparities in T1DM outcomes between populations.

6. Conclusions

ML technologies are reshaping the landscape of T1DM management, offering tools for personalised, data-driven intervention across a range of clinical challenges. From glucose prediction and risk modelling to automated insulin delivery, the literature demonstrates a broad spectrum of promising applications.

No single algorithm has demonstrated superiority across all applications. SVR consistently yields lower RMSE for long-term glucose prediction, making it ideal for extended predictions. NNs achieve high short-term accuracy with growing interpretability through techniques like confidence intervals. In risk modelling, XGBoost achieved excellent hypoglycaemia prediction. RFs have shown stronger results in predicting T1DM onset and CVD due to their robustness in handling complex variable interactions. Meanwhile, DL incorporated within dual-hormone closed-loop systems demonstrated the highest TIR performance compared to single-hormone systems, although high cost and algorithmic complexity remain key barriers to implementation.

A consistent limitation in existing models is the overreliance on CGM data, often to the exclusion of contextual lifestyle variables like physical activity and diet. These factors substantially influence glycaemic trends and should be incorporated to enhance model robustness and validity.

Translating ML into clinical practice requires models that are not only accurate but also interpretable, secure, and equitable. Adherence to privacy regulations such as the HIPAA and GDPR must be balanced with transparency and fairness. Federated learning offers a privacy-maintaining path toward diverse, multicentre collaboration, while fairness-aware frameworks and regular audits can help mitigate algorithmic bias.

As global T1DM prevalence continues to rise and healthcare systems face resource strain, ML presents a scalable opportunity to reduce clinical burden and deliver adaptive, individualised care. Future work should focus on computational efficiency, integration with telehealth and biosensors, and validation across diverse populations to bridge the gap between innovation and implementation.

7. References

1. Lucier J, Mathias PM. Type 1 diabetes. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2025
2. Gregory GA, Robinson TIG, Linklater SE, Wang F, Colagiuri S, de Beaufort C, et al. Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study. Lancet Diabetes Endocrinol. 2022;10(10):741–60. doi: 10.1016/S2213-8587(22)00218-2
3. Cryer PE. The barrier of hypoglycemia in diabetes. Diabetes. 2008;57(12):3169–76. doi: 10.2337/db08-1084
4. Kitabchi AE, Umpierrez G, Miles J, Fisher J. Hyperglycemic crises in adult patients with diabetes. Diabetes Care. 2009;32(12):e158–e158. doi: 10.2337/dc09-1563
5. Krolewski AS, Kosinski EJ, Warram JH, Leland OS, Busick EJ, Asmal AC, et al. Magnitude and determinants of coronary artery disease in juvenile-onset, insulin-dependent diabetes mellitus. Am J Cardiol. 1987;59(8):750–5. doi: 10.1016/0002-9149(87)91086-1
6. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019;157(107843):107843. doi: 10.1016/j.diabres.2019.107843
7. Heinemann L, Freckmann G, Ehrmann D, Faber-Heinemann G, Guerra S, Waldenmaier D. Real-time continuous glucose monitoring in the management of diabetes. Diabetes Res Clin Pract. 2018;139:78–82. doi: 10.1016/S0140-6736(18)30297-6
8. Oikonomou EK, Khera R. Machine learning in precision diabetes care and cardiovascular risk prediction. Cardiovasc Diabetol. 2023;22(1):259. doi: 10.1186/s12933-023-01985-3
9. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: Literature review. J Med Internet Res. 2020;22(5). doi: https://doi.org/10.2196/10775
10. Liu K, Li L, Ma Y, Jiang J, Liu Z, Ye Z, et al. Machine learning models for blood glucose level prediction in patients with diabetes mellitus: Systematic review and network meta-analysis. JMIR Med Inform. 2023;11:e47833. doi: 10.2196/47833
11. Navarro-Cerdán JR, Pons-Suñer P, Arnal L, Arlandis J, Llobet R, Perez-Cortes J-C, et al. A machine learning approach for type 2 diabetes diagnosis and prognosis using tailored heterogeneous feature subsets. Med Biol Eng Comput [Internet]. 2025; doi: 10.1007/s11517-025-03355-5
12. Zhu T, Li K, Herrero P, Georgiou P. Personalised blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning. IEEE Trans Biomed Eng. 2023;70(1):193–204. doi: 10.1109/TBME.2022.3187703
13. McAuley SA, Lee MH, Paldus B, Vogrin S, de Bock MI, Abraham MB, et al. Six months of hybrid closed-loop versus manual insulin delivery with fingerprick blood glucose monitoring in adults with type 1 diabetes: A randomized, controlled trial. Diabetes Care. 2020;43(12):3024–33. doi: 10.2337/dc20-1447
14. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak. 2019;19(1):211. doi: 10.1186/s12911-019-0918-5
15. Liu Y-P, Li Z, Xu C, Li J, Liang R. Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. Artif Intell Med. 2019;99(101694):101694. doi: 10.1016/j.artmed.2019.07.002

16. Ahmad MA, Teredesai A, Eckert C. Interpretable Machine Learning in Healthcare. In: 2018 IEEE International Conference on Healthcare Informatics (ICHI). IEEE; 2018.
17. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA 2017;318(6):517–8. doi: 10.1001/jama.2017.7797
18. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People-an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds Mach (Dordr). 2018;28(4):689–707. doi: 10.1007/s11023-018-9482-5
19. Diouri O, Cigler M, Vettoretti M, Mader JK, Choudhary P, Renard E, et al. Hypoglycaemia detection and prediction techniques: A systematic review on the latest developments. Diabetes Metab Res Rev 2021;37(7):e3449. doi: 10.1002/dmrr.3449
20. Hamdi T, Ben Ali J, Di Costanzo V, Fnaiech F, Moreau E, Ginoux J-M. Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. Biocybern Biomed Eng. 2018;38(2):362–72. doi: 10.1016/j.bbe.2018.02.005
21. What is support vector machine? [Internet]. Ibm.com. 2024 [cited 2025 Jan 28]. Available from: https://www.ibm.com/think/topics/support-vector-machine
22. What is overfitting? [Internet]. Ibm.com. 2024 [cited 2025 Feb 13]. Available from: https://www.ibm.com/think/topics/overfitting
23. Shanmuganathan S. Artificial neural network modelling: An introduction. In: Studies in Computational Intelligence. Cham: Springer International Publishing; 2016. p. 1–14.
24. Allam F, Nossair Z, Gomma H, Ibrahim I, Abd-el Salam M. Prediction of subcutaneous glucose concentration for type-1 diabetic patients using a feed forward neural network. In: The 2011 International Conference on Computer Engineering & Systems. IEEE; 2011. doi 10.1109/ICCES.2011.6141026
25. Ghimire S, Celik T, Gerdes M, Omlin CW. Deep learning for blood glucose level prediction: How well do models generalize across different data sets? PLoS One. 2024;19(9):e0310801. doi: 10.1371/journal.pone.0310801
26. Clarke WL. The original Clarke Error Grid Analysis (EGA). Diabetes Technol Ther. 2005;7(5):776–9. doi: 10.1089/dia.2005.7.776
27. Jouini O, Sethom K, Namoun A, Aljohani N, Alanazi MH, Alanazi MN. A survey of machine learning in edge computing: Techniques, frameworks, applications, issues, and research directions. Technologies (Basel). 2024;12(6):81. doi: 10.3390/technologies12060081
28. Holdsworth J. What is deep learning? [Internet]. Ibm.com. 2024 [cited 2025 Feb 6]. Available from: https://www.ibm.com/think/topics/deep-learning
29. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: Contextualizing explainable machine learning for clinical end use [Internet]. arXiv [cs.LG]. 2019. doi: 10.48550/arXiv.1905.05134
30. Jaloli M, Cescon M. Long-term prediction of blood glucose levels in type 1 diabetes using a CNN-LSTM-based deep neural network. J Diabetes Sci Technol. 2023;17(6):1590–601. doi: 10.1177/19322968221092785
31. Oyebode O, Fowles J, Steeves D, Orji R. Machine learning techniques in adaptive and personalized systems for health and wellness. Int J Hum Comput Interact. 2022;1–25. doi: 10.1080/10447318.2022.208 9085
32. Lu Y, Wan D. Enhancing methodological rigor in AI-driven systematic reviews: Addressing key limitations in predicting TACE outcomes. Dig Dis Sci. 2025; doi: 10.1007/s10620-025-08969-1

33. Reddy R, Resalat N, Wilson LM, Castle JR, El Youssef J, Jacobs PG. Prediction of hypoglycemia during aerobic exercise in adults with type 1 diabetes. J Diabetes Sci Technol. 2019;13(5):919–27. doi: 10.1177/1932296818823792
34. Çorbacıoğlu ŞK, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value: A guide to interpreting the area under the curve value. Turk J Emerg Med. 2023;23(4):195–8. doi: 10.4103/tjem.tjem_182_23
35. Duckworth C, Guy MJ, Kumaran A, O'Kane AA, Ayobi A, Chapman A, et al. Explainable machine learning for real-time hypoglycemia and hyperglycemia prediction and personalized control recommendations. J Diabetes Sci Technol. 2024;18(1):113–23. doi: 10.1177/19322968221103561
36. Kavlakoglu E, Russi E. What is XGBoost? [Internet]. Ibm.com. 2025 [cited 2025 Apr 22]. Available from: https://www.ibm.com/think/topics/xgboost
37. Mellor J, Kuznetsov D, Heller S, Gall M-A, Rosilio M, Amiel SA, et al. Risk factors and prediction of hypoglycaemia using the Hypo-RESOLVE cohort: a secondary analysis of pooled data from insulin clinical trials. Diabetologia. 2024;67(8):1588–601. doi: 10.1007/s00125-024-06177-6
38. Bertachi A, Viñals C, Biagi L, Contreras I, Vehí J, Conget I, et al. Prediction of nocturnal hypoglycemia in adults with type 1 diabetes under multiple daily injections using continuous glucose monitoring and physical activity monitor. Sensors (Basel). 2020;20(6):1705. doi: 10.3390/s20061705
39. Russon CL, Vaughan N, Carr ALJ, Pulsford RM, Allen M, Andrews RC. An increase in recording interval in continuous glucose monitors results in the identification of fewer hypoglycaemic episodes but interpolation can help to identify some of these missed episodes. Diabetes UK Professional Conference 2022. 2022; doi: 10.1111/dme.14810
40. Johnson AEW, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. Proc IEEE Inst Electr Electron Eng. 2016;104(2):444–66. doi: 10.1109/JPROC.2015.2501978
41. Qu H-Q, Qu J, Glessner J, Liu Y, Mentch F, Chang X, et al. Improved genetic risk scoring algorithm for type 1 diabetes prediction. Pediatr Diabetes. 2022;23(3):320–3. doi: 10.1111/pedi.13310
42. Alazwari A, Abdollahian M, Tafakori L, Johnstone A, Alshumrani RA, Alhelal MT, et al. Predicting age at onset of type 1 diabetes in children using regression, artificial neural network and Random Forest: A case study in Saudi Arabia. PLoS One. 2022;17(2):e0264118. doi: 10.1371/journal.pone.0264118
43. Liu S. Diabetes prediction by KNN, SVM, random forest and XGBoost. Highlights in Science, Engineering and Technology. 2023;72:1113–20. doi: 10.54097/8h8dff76
44. Akella A, Akella S. Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution. Future Sci OA. 2021;7(6):FSO698. doi: 10.2144/fsoa-2020-0206
45. Rabhi S, Blanchard F, Diallo AM, Zeghlache D, Lukas C, Berot A, et al. Temporal deep learning framework for retinopathy prediction in patients with type 1 diabetes. Artif Intell Med. 2022;133(102408):102408. doi: 10.1016/j.artmed.2022.102408
46. Vaughan N. Review of smartphone fundoscopy for diabetic retinopathy screening. Surv Ophthalmol. 2024;69(2):279–86. doi: 10.1016/j.survophthal.2023.10.006
47. Ahmed S, Shamim Kaiser M, Hossain MS, Andersson K. A comparative analysis of LIME and SHAP interpreters with explainable ML-based diabetes predictions. IEEE Access. 2024;1–1. doi: 10.1109/access.2024.3422319

48. C.L. Russon, R.C. Andrews, R.M. Pulsford, N. Vaughan, M. Allen. Glycaemic events during exercise can be effectively predicted with machine learning using only start glucose and duration. Diabetologia. 2022;65((Suppl 1):S324). doi: 10.1007/s00125-022-05755-w
49. Haidar A. The Artificial Pancreas: How Closed-Loop Control Is Revolutionizing Diabetes. Vol. 36, IEEE control systems. New York: IEEE; 2016. 28–47 p.
50. Marks BE, Williams KM, Sherwood JS, Putman MS. Practical aspects of diabetes technology use: Continuous glucose monitors, insulin pumps, and automated insulin delivery systems. J Clin Transl Endocrinol. 2022;27(100282):100282. doi: 10.1016/j.jcte.2021.100282
51. Beck RW, Riddlesworth T, Ruedy K, Ahmann A, Bergenstal R, Haller S, et al. Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections: The DIAMOND randomized clinical trial: The DIAMOND randomized clinical trial. JAMA. 2017;317(4):371–8. doi: 10.1001/jama.2016.19975
52. Benhamou P-Y, Franc S, Reznik Y, Thivolet C, Schaepelynck P, Renard E, et al. Closed-loop insulin delivery in adults with type 1 diabetes in real-life conditions: a 12-week multicentre, open-label randomised controlled crossover trial. Lancet Digit Health. 2019;1(1):e17–25. doi: 10.1016/S2589-7500(19)30003-2
53. What is HbA1c? [Internet]. Diabetes UK. 2017 [cited 2025 Feb 5]. Available from: https://www.diabetes.org.uk/about-diabetes/looking-after-diabetes/hba1c
54. Zhu T, Li K, Georgiou P. A dual-hormone closed-loop delivery system for Type 1 Diabetes using deep reinforcement learning [Internet]. arXiv [q-bio.QM]. 2019. doi: 10.48550/arXiv.1910.04059
55. Holt RIG, DeVries JH, Hess-Fischl A, Hirsch IB, Kirkman MS, Klupa T, et al. The management of type 1 diabetes in adults. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia. 2021;64(12):2609–52. doi: 10.1007/s00125-021-05568-3
56. Lindkvist EB, Laugesen C, Reenberg AT, Ritschel TKS, Svensson J, Jørgensen JB, et al. Performance of a dual-hormone closed-loop system versus insulin-only closed-loop system in adolescents with type 1 diabetes. A single-blind, randomized, controlled, crossover trial. Front Endocrinol (Lausanne). 2023;14:1073388. doi: 10.3389/fendo.2023.1073388
57. Mughal IS, Patanè L, Caponetto R. A comprehensive review of models and nonlinear control strategies for blood glucose regulation in artificial pancreas. Annu Rev Control [Internet]. 2024;57(100937):100937. doi: 10.1016/j.arcontrol.2024.100937
58. le Roux CW, Steen O, Lucas KJ, Startseva E, Unseld A, Hussain SA, et al. Subgroup analysis by sex and baseline BMI in people with a BMI ≥27 kg/m2 in the phase 2 trial of survodutide, a glucagon/GLP-1 receptor dual agonist. Diabetes Obes Metab. 2025; doi: 10.1111/dom.16167
59. Mehmood A, Natgunanathan I, Xiang Y, Hua G, Guo S. Protection of big data privacy. IEEE Access. 2016;4:1821–34. doi: 10.1109/access.2016.2558446
60. Qu Y, Gao L, Luan TH, Xiang Y, Yu S, Li B, et al. Decentralized privacy using blockchain-enabled federated learning in fog computing. IEEE Internet Things J . 2020;7(6):5171–83. doi: 10.1109/jiot.2020.2977383
61. Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, et al. Advances and open problems in federated learning [Internet]. arXiv [cs.LG]. 2019. doi: 10.48550/arXiv.1912.04977
62. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866–72. doi: 10.7326/M18-1990

63. Wang A, Ramaswamy VV, Russakovsky O. Towards intersectionality in machine learning: Including more identities, handling underrepresentation, and performing evaluation. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. New York, NY, USA: ACM; 2022
64. Uddin MA, Islam MM, Talukder MA, Hossain MAA, Akhter A, Aryal S, et al. Machine learning based diabetes detection model for false negative reduction. Biomed Mater Devices. 2024;2(1):427–43. doi: 10.1007/s44174-023-00104-w
65. Friedler SA, Scheidegger C, Venkatasubramanian S, Choudhary S, Hamilton EP, Roth D. A comparative study of fairness-enhancing interventions in machine learning. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. New York, NY, USA: ACM; 2019.
66. Agarwal R, Bjarnadottir M, Rhue L, Dugas M, Crowley K, Clark J, et al. Addressing algorithmic bias and the perpetuation of health inequities: An AI bias aware framework. Health Policy Technol. 2023;12(1):100702. doi: 10.1016/j.hlpt.2022.100702
67. Zhang BH, Lemoine B, Mitchell M. Mitigating unwanted biases with adversarial learning [Internet]. arXiv [cs.LG]. 2018. doi: 10.48550/arXiv.1801.07593
68. Valente F, Paredes S, Henriques J, Rocha T, de Carvalho P, Morais J. Interpretability, personalization and reliability of a machine learning based clinical decision support system. Data Min Knowl Discov. 2022;36(3):1140–73. doi: 10.1007/s10618-022-00821-8
69. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med. 2018;15(11):e1002689. doi: 10.1371/journal.pmed.1002689
70. Hassan N, Slight R, Morgan G, Bates DW, Gallier S, Sapey E, et al. Road map for clinicians to develop and evaluate AI predictive models to inform clinical decision-making. BMJ Health Care Inform. 2023;30(1):e100784. doi: 10.1136/bmjhci-2023-100784
71. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Soc Sci Med. 2020;260(113172):113172. doi: 10.1016/j.socscimed.2020.113172
72. Wang F, Casalino LP, Khullar D. Deep learning in medicine-promise, progress, and challenges. JAMA Intern Med. 2019;179(3):293–4. doi: 10.1001/jamainternmed.2018.7117

8. Figures

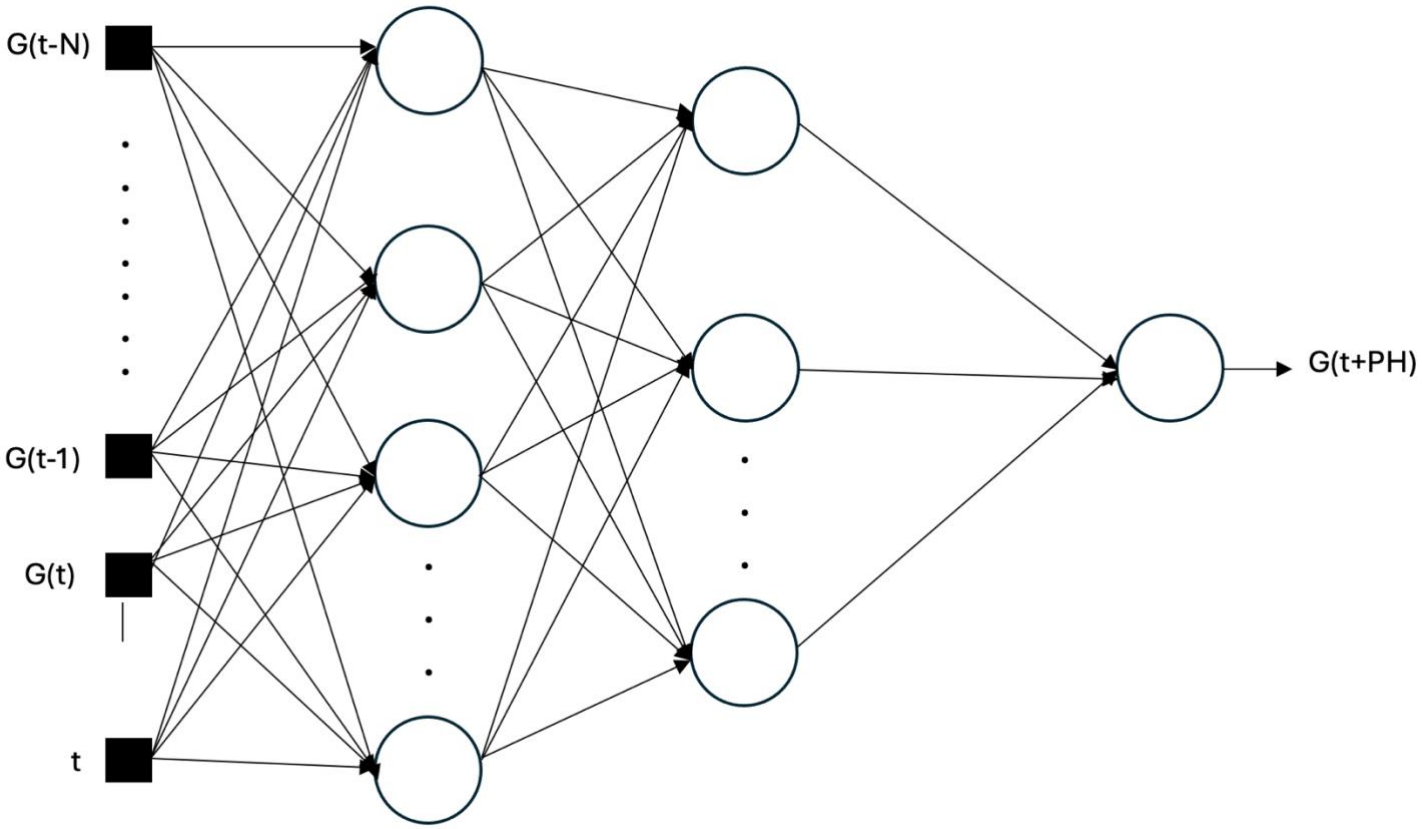

Figure 1. Architecture of a Neural Network, developed by Allam et al (Adapted from 24). The model uses previous blood glucose measurements; G(t-N) represents blood glucose concentration at time–N minutes ago. G(t-1) represents blood glucose concentration at time-1 minute ago. G(t) represents blood glucose concentration at the current time, and t represents current time; used for temporal learning, G(t+PH) represents the predicted blood glucose concentration at time+PH. Each circle represents a neuron within the hidden layers, with weighted connections capturing patterns in the temporal glucose data. Abbreviations: PH, prediction horizon; G, glucose; t, time.

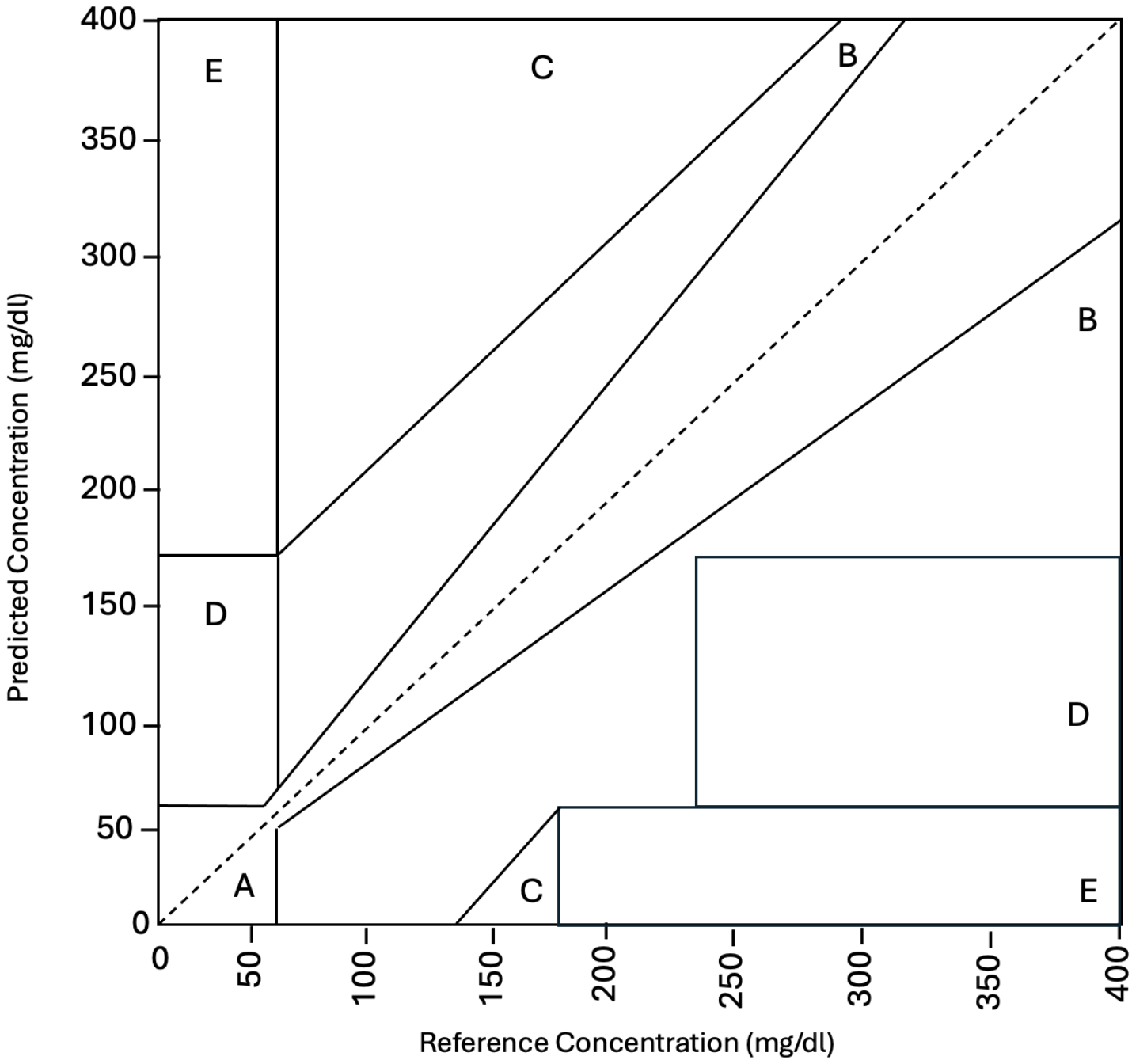

Figure 2. Clarkes Error Grid (Adapted from 26). Evaluates the clinical accuracy of glucose prediction models by categorising predicted blood glucose concentrations against reference measurements into zones (A-E). Zone A represents accurate predictions with no clinical risk, Zone B indicates benign errors with minimal clinical impact, Zones C and D represent errors that could cause to incorrect treatments and detections, and Zone E indicates predictions with severe consequences.

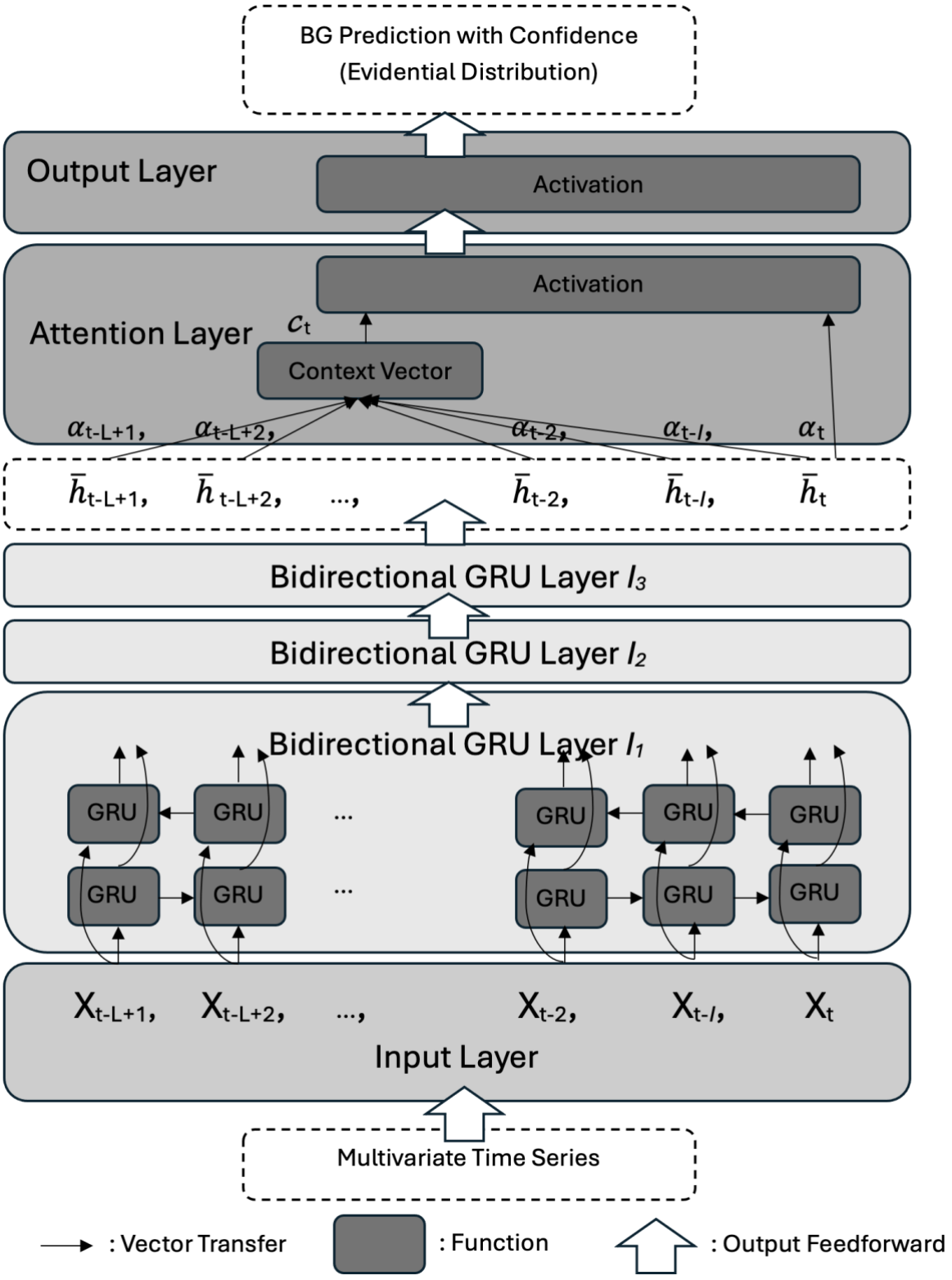

Figure 3. Diagram of a Deep Learning Model, developed by Zhu et al. (Adapted from 12). This model uses multivariate time series inputs, which are processed through three bidirectional GRU layers to capture patterns in the data over time. An attention layer identifies and emphasises the most relevant hidden states, creating a context vector that is passed to the output layer. The model outputs blood glucose predictions, along with confidence intervals based on an evidential distribution. Arrows represent data flow between layers, while shaded boxes represent key model components. Abbreviations: GRU, gated recurrent unit; BG, blood glucose.

Table 1. Comparison of Machine Learning Models for Blood Glucose Prediction.

		RMSE values in mmol/L ± (SD) across different PH					
Model	**N**	**PH = 15**	**PH = 30**	**PH = 45**	**PH = 60**	**PH = 90**	**Mean RMSE**
SVR (20)	12	0.52 † ^	0.60 † ^	**0.66 † ⁺ ^**	**0.72 † ⁺ ^**	*N/A*	0.63 (± 0.07) ^
NN (24)	9	**0.15 † ⁺ ^**	**0.42 † ⁺ ^**	0.83 †	*N/A*	*N/A*	**0.47 (± 0.28) ⁺ ^**
NN (25)	451	*N/A*	1.23 (± 0.01)	*N/A*	1.95 (± 0.02)	*N/A*	1.59 (± 0.36)
FCNN (12)	12	*N/A*	1.03 (± 0.14)	*N/A*	1.72 (± 0.20)	*N/A*	1.21 (± 0.35)
DL (30)	168	*N/A*	0.54 (± 0.07) ^	*N/A*	0.92 (± 0.14)	**1.30 (± 0.18)** ⁺	0.92 (± 0.31)
SVR (30)	168	*N/A*	0.96 (± 0.10)	*N/A*	1.41 (± 0.23)	1.61 (± 0.27)	1.33 (± 0.27)
NN (30)	168	*N/A*	0.59 (± 0.07) ^	*N/A*	0.99 (± 0.12)	1.38 (± 0.18)	0.99 (± 0.32)

RMSE in mmol/L ± (standard deviation) presented for each model across multiple PHs. Lower RMSE indicates higher predictive accuracy. N used to train each model is included. ⁺ shows the best-performing model for each time horizon and overall mean. *N/A* indicates that data for that prediction horizon was not available or not reported**.** † Standard deviation not reported for this model in the source study. ^ indicates a prediction below the proposed clinical acceptability threshold of 0.83 mmol/L. Abbreviations: SVR, support vector regression; NN, neural network; FCNN, fast-adaptive and confident neural network; DL, deep learning; RMSE, root mean squared error; PH, prediction horizons; SD, standard deviation; N, number of participants.

Table 2. Performance of Machine Learning Models in Short and Long-Term Complication Prediction.

Model	Application	Evaluation
Decision Tree (33)	Hypoglycaemia Prediction	AUC = 0.79
RF (33)	Hypoglycaemia Prediction	AUC = 0.84
XGBoost (35)	Hypoglycaemia Prediction	**AUC = 0.99**[+]
XGBoost (37)	Hypoglycaemia Prediction	AUC = 0.84
XGBoost (14)	T1DM Onset Prediction	Error Rate = 0.26
MLR (42)	T1DM Onset Prediction	Error Rate = 0.19
RF (42)	T1DM Onset Prediction	**Error Rate = 0.11**[+]
XGBoost (43)	T1DM Onset Prediction	Error Rate = 0.18
RF (43)	T1DM Onset Prediction	Error Rate = 0.21
DL (45)	Retinopathy Prediction – HbA1c Data	AUC = 0.88
DL (15)	Retinopathy Prediction – Eye Fundus Data	**AUC = 0.95**[+]
XGBoost (14)	CVD Prediction	AUC = 0.84
SVM (44)	CVD Prediction	AUC = 0.87
KNN (44)	CVD Prediction	AUC = 0.85
RF (44)	CVD Prediction	**AUC = 0.88**[+]

Summary of ML models evaluated for hypoglycaemia prediction, T1DM onset detection, retinopathy identification, and CVD risk prediction. Performance is reported using AUC or prediction error rate, depending on the study. AUC closer to 1.00 and error rate closer to 0.00 indicate better predictive performance. [+] indicates the best-performing model within each complication category. Abbreviations: RF, random forest; XGBoost, extreme gradient boosting; MLR, multiple linear regression; DL, deep learning; SVM, support vector machine; KNN, K-nearest neighbours; CVD, cardiovascular disease; AUC, area under curve; HbA1c, glycosylated haemoglobin; T1DM, type 1 diabetes mellitus.

Table 3. Interpretation of AUC Scores for Model Performance.

AUC	**Interpretation Suggestion**
0.9 ≤ AUC	Excellent
0.8 ≤ AUC < 0.9	Considerable
0.7 ≤ AUC < 0.8	Fair
0.6 ≤ AUC < 0.7	Poor
0.5 ≤ AUC < 0.6	Fail

AUC is a commonly used metric for evaluating classification model performance. Higher AUC indicate better prediction ability. The interpretation thresholds provided are from (34). Abbreviations: AUC, area under the curve.

Table 4. Performance of Closed-Loop Systems

Study	Type of Closed-Loop System	TIR (%)	HbA1c Reduction (%)	Time in Hyperglycaemic Range (>10 mmol/L) (%)
Benhamou et al. (52)	Single Hormone HCL	68.5 (± 9.4)	0.3	29.5 (± 10.2)
McAuley et al. (13)	Single Hormone HCL	69.9 (± 9.5)	**0.4**[+]	27.6 (± 9.5)
Zhu et al. (54)	Dual Hormone Closed-Loop System	**93.1 (± 4.5)** [+^]	*N/A*	**5.63 (± 3.29)** [+]
Lindkvist (56)	Single Hormone Closed-Loop System	79.6 (± 4.4) ^	*N/A*	33.2 (± 17.1)
Lindkvist (56)	Dual Hormone Closed-Loop System	66.8 (± 9.9)	*N/A*	11.5 (± 7.67)

Compares the performance of single- and dual-hormone closed-loop systems in T1DM management, using metrics (± SD) from Benhamou et al. (52), McAuley et al. (13), Zhu et al. (54), and Lindkvist (56). Metrics include TIR (%), HbA1c reduction (%), and Time in Hyperglycaemic Range (>10 mmol/L, %). 'N/A' indicates unavailable or unreported data. + denotes the best-performing result in each metric. ^ indicates TIR above the 70% threshold set by the EASD and ADA. Abbreviations: TIR, time in range; HbA1c, glycosylated haemoglobin; HCL, hybrid closed loop; EASD, European association for the study of diabetes; ADA, American diabetes association; SD, standard deviation; T1DM, type 1 diabetes mellitus.