Indonesian Journal of Innovation and Applied Sciences (IJIAS)
Journal Homepage: https://ojs.literacyinstitute.org/index.php/ijias
ISSN: 2775-4162 (Online)
Review Article

Volume 5	Issue 2	June (2025)	DOI: 10.47540/ijias.v5i2.1981	Page: 145 – 162

Managing Diabetes Using Machine Learning and Digital Twins

Sanele Hadebe[1], Belinda Ndlovu[1], Kudakwashe Maguraushe[2]
[1]Department of Informatics and Analytics, National University of Science and Technology, Zimbabwe
[2]School of Computing, College of Science, Engineering and Technology, University of South Africa
Corresponding Author: Belinda Ndlovu; Email: belinda.ndlovu@nust.ac.zw

ARTICLE INFO

Keywords: Artificial Intelligence, Diabetes Prediction, Digital Twins, Machine Learning.

Received : 22 April 2025
Revised : 11 May 2022
Accepted : 04 June 2022

ABSTRACT

Diabetes is a major public health problem worldwide, and early diagnosis will remain pivotal for intervention and management. This Systematic Literature Review (SLR), therefore, attempts to explore the prospects of integrating Machine Learning (ML) and Digital Twins (DT) to enable diabetes treatment through prediction and patient-specific modeling. This SLR contributes to the body of literature by examining how ML and DTs are being applied in diabetes treatment, identifying the opportunities and challenges that exist, and determining which algorithms are most commonly used. In contrast to SLRs that have been reviewed previously, this study considers Digital Twin-based technological perspectives, along with algorithmic evaluations of ML models, to provide an overall view of the potential for combined use in diabetes care. Following PRISMA guidelines, 11 relevant studies were selected from major academic databases. The study identified Random Forests, Gradient-Boosted Decision Trees, K-Nearest Neighbors, Time Series and Structured Analysis, Regression-based algorithms, and Artificial Neural Networks as machine learning algorithms commonly used to predict diabetes risk. The integration of ML and DT for diabetes management enables the personalization of patient management through virtual representations, real-time monitoring of an individual's glucose levels, simulation of disease progression, and prediction of subsequent treatment steps for proactive and immediate decision-making. Through this collaboration, simulations of various situations are performed, and the interventions are optimized to correspond with unique human physiological profiles for better patient outcomes. Based on the results, policymakers must balance data quality and patient privacy.

Introduction

Diabetes is a chronic condition that affects millions of people all over the world, which has a significant cost to health and management (Abdi et al., 2020; Li et al., 2020; Maguraushe & Ndayizigamiye, 2024; Murere et al., 2024). The condition is classified into two categories, namely Type 1, which results from insufficient insulin production by the pancreas, and Type 2 diabetes, which is common among adults (Mpofu et al., 2024; Mutunhu et al., 2022). Many additional health concerns arise from this condition, and these create a significant psychological and social burden on the patients as they are bound to face a variety of health issues throughout their lives (Chu et al., 2023; Ndhlovu et al., 2023; Mutunhu et al., 2023; Suryasa et al., 2021; Tomic et al., 2022). Unfortunately, traditional diabetes management methods, which primarily include standard clinical practices such as regular monitoring of blood glucose levels, dietary management, and medication adherence (Mtshali et al., 2024; Mutunhu et al., 2024b; Ndlovu et al., 2024), have generally been inadequate in providing holistic preventive and personalized care to their patients; thus, patient outcomes are far from optimal, and risks for developing complications, including cardiovascular disease, end-stage renal disease, retinopathy, and neuropathy, are prominent (Cappon & Facchinetti, 2024; Cellina et al., 2023). This pressing need calls for the establishment of innovative solutions for the enhancement of diabetes care and, thus, road maps for data-driven personalized medicine (Cellina et al., 2023; Murere et al., 2024).

Digital Twins (DT) are virtual representations of physical objects, capable of creating simulations and analyses of intricate systems (Agrawal et al., 2022; Chu et al., 2023; Singh et al., 2021). In diabetes management, DTs create an individual virtual profile of the person, such as age, gender, Body-Mass-Index (BMI), a history of high blood pressure, a history of heart attacks, a history of smoking, level of HbA1c, and blood glucose level to simulate and predict metabolic responses (Cellina et al., 2023). These profiles can be used to simulate disease progression. Additionally, they can be used in the modelling of behaviors and lifestyles that can provide a real-time representation of the patient's health (Meijer et al., 2023).

Medical Digital Twin (MeDigiT) consists of a virtual duplicate of each patient's physiological condition, originating from numerous data sources that provide individualized health care (Chu et al., 2023). Such a model utilizes data from electronic medical records, devices used for tracking physical exercise, and medical imaging, providing a full illustration of his healthy manner (Mosquera-Lopez & Jacobs, 2024). MeDigiT uses advanced modeling and simulation to simulate and even predict attendant outcomes, enabling providers to customize treatments and interventions (Cappon & Facchinetti, 2024). For instance, for diabetes management, MeDigiT can customize insulin doses, foresee implications, and design individual-specific treatment plans that enhance good patient outcomes and quality of life-healthy existence for others (Shamanna et al., 2024).

Machine learning (ML) is a growing field wherein algorithms analyze large datasets, capture non-linear correlations, and make reliable predictions (Frank et al., 2020; Moyo et al., 2022; Murere et al., 2024; Rane et al., 2024). In diabetes management, ML is employed to analyze complex patient data, such as clinical information and real-time health metrics, to create helpful clinical decision support (Mosquera-Lopez & Jacobs, 2024). It does so by predicting patient outcomes, improving treatment pathways, and personalizing interventions through pattern and relationship detection within the data. With this increased input from ML, MeDigiT is enabled to play an even stronger role in assessing and giving personalized healthcare solutions to diabetic patients (Chu et al., 2023).

However, previous Systematic Literature Reviews (SLRs) have not addressed the collective use of ML and DT for diabetes care in a truly holistic sense. Other SLRs have primarily focused on isolated ML approaches for glucose level prediction, utilizing only a few physiological parameters (Meijer et al., 2023; Mosquera-Lopez & Jacobs, 2024; Rane et al., 2024). Other SLRs mention the advantages of DTs but posit their lack of incorporation with broader health data and ML techniques (Chu et al., 2023) Furthermore, these reviews are inconsistent as far as evaluation metrics are concerned, thereby making it very difficult to even compare one with another (Cappon & Facchinetti, 2024). Thus, the use of ML and DT together to provide personalized and predictive care in the area of diabetes remains largely unexplored.

Therefore, this systematic literature review attempts to bridge this gap by reviewing the different avenues for diabetes management through the joint or separate application of ML and DT. The objectives are to explore the current application, opportunities, integration challenges, and highlight the particularities of the ML algorithms and DT technologies used. This review contributes to the body of literature by providing a holistic view on the potentialities of the ML-DT integration, along with insights and suggestions for further research as well as clinical practice in personalized diabetes care. Thus, the objective of this SLR is to analyze how ML and DT technologies are applied to the management of diabetes, identify the opportunities and challenges of ML and DT technologies offer to improve diabetes care, and determine the specific ML algorithms and DT technologies used for diabetes management.

METHODS

This review was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) methodology. The PRISMA process involves several stages, including outlining research questions, search terms, databases searched, criteria for eligibility, reports, findings, discussions, and gaps (Moher et al., 2009).

Database Name

To study the existing literature on diabetes prediction using ML along with DT, a systematic search of the following electronic databases was undertaken: PubMed, Springer, and IEEE Xplore.

Full Search Strategy

A search string was constructed and modified to suit each database syntax. The search string used a combination of keywords and search queries that are based on 'diabetes prediction', 'machine learning', and 'digital twins'. The variants of these terms were also used to get a wide coverage of publications in each of the databases. (“digital twin*” OR “digital shadow” OR “DT” OR “digital Avatar”) AND (“Machine Learning” OR “ML” OR “Artificial Intelligence” OR “AI” OR “Deep Learning” OR “DL”) AND (“Diabetes*” OR “Blood Glucose” OR “Blood Sugar” OR “T2DM”)

Inclusion and Exclusion

The following inclusion criteria were used:

1. Studies purely on ML and DT applied specifically for diabetes prediction.
2. Scholarly articles published within the years 2020 to 2025.
3. Articles written in the English language only due to language barriers.
4. Research that assesses the effectiveness and efficiency of ML integrated with DT for diabetes management.
5. Studies that apply ML and DT specifically for predicting diabetes.

The exclusion criteria were:

1. Non-peer-reviewed publications, editorials, commentaries, and opinion pieces.
2. Research that is too narrow and/or limited, such as case studies with a very limited focus.
3. Studies that focus exclusively on gestational diabetes.

Eligibility and Screening

The SLR process was initiated with a wide-ranging search for studies within the 3 databases. The first-round search yielded 560 publications in total, as follows: PubMed (158), IEEE Xplore (124), and Springer (278). The PRISMA guidelines were followed for the screening and eligibility assessment of studies included in the review to maintain the relevance and quality of selected studies. The first step involved the removal of duplicate studies. After duplicate study removal, 550 unique records remained.

The researchers then proceeded with title and abstract screening of the remaining 550 records, which aimed to discard all those studies that were irrelevant to the formulated study research questions. Resultantly, 539 records were excluded based on the following criteria:

1. Not relevant to the topic (410 records).
2. Not related to the objective of the study (9 records).
3. Not related to diabetes, ML and DT (120 records).

Studies Included

The remaining 11 studies were read in their full text, and all 11 studies were included. A detailed description of the process is depicted in the PRISMA flow diagram (see Fig. 1), where the number of records at each step in the screening and eligibility process is illustrated.

RESULTS AND DISCUSSION

The PRISMA guidelines were used, and Fig. 1 is the flow diagram for this study using PRISMA. Table 1 presents the papers that met the inclusion criteria.

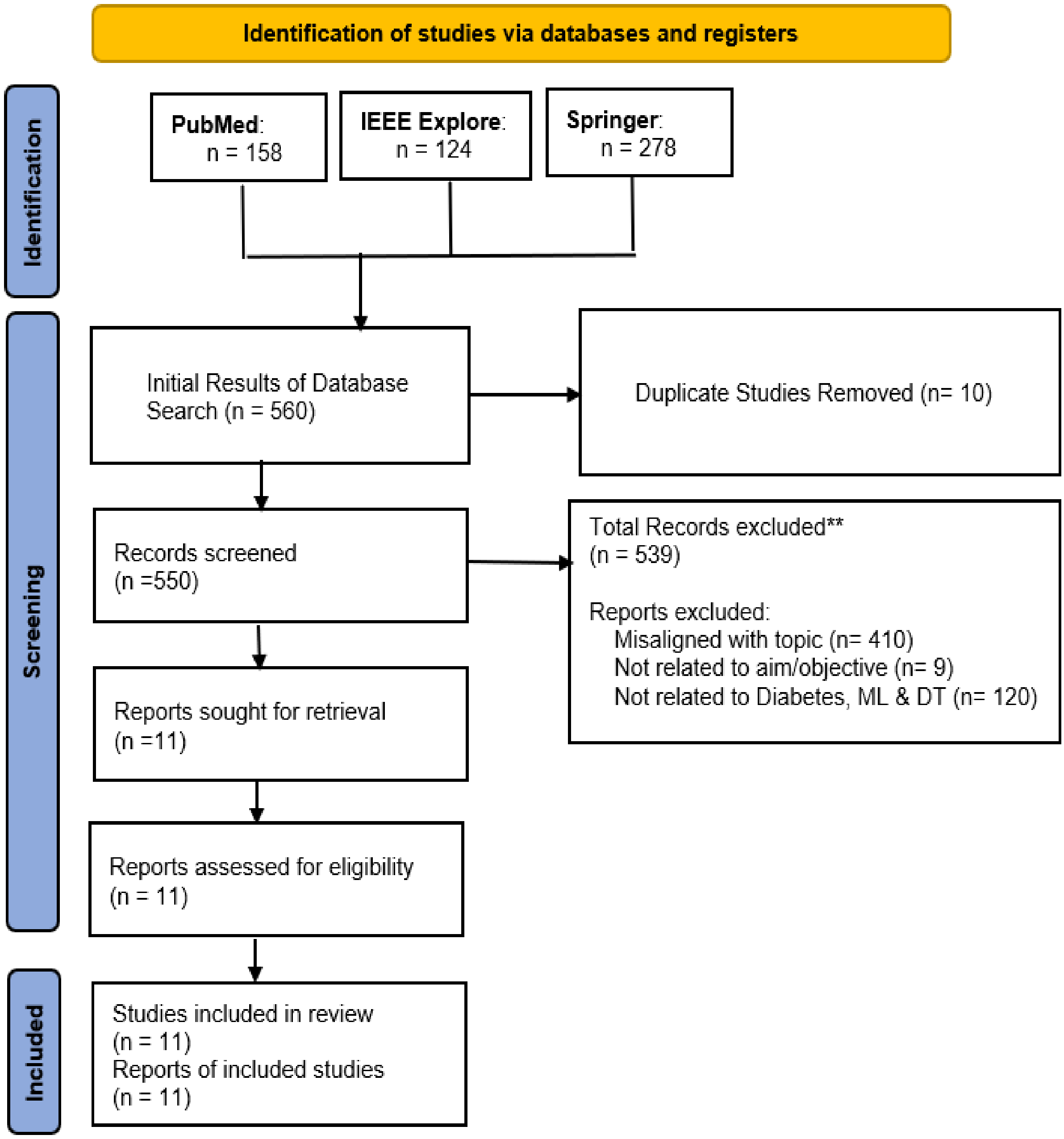
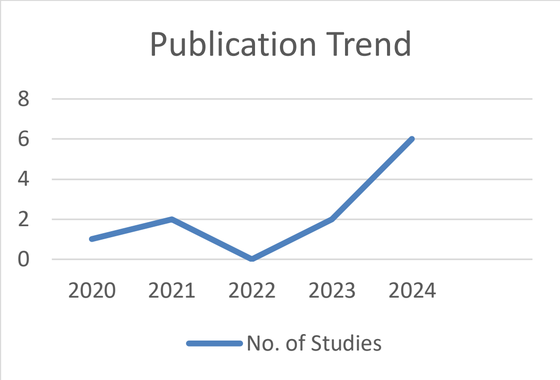


Figure 1. PRISMA Flow Diagram (Page et al. 2021)

Publication and Citation Trends

Fig. 2 shows that only one paper was published in 2020, and the paper has 79 citations. The year 2021 brought in more publications, and the figure reached 2, despite an impending fall in citations, which reduced to 55.



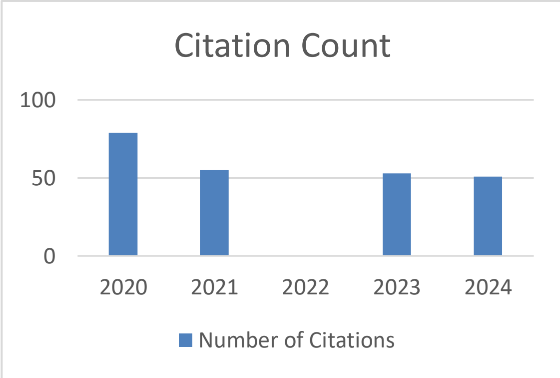


Figure 2. Publication and Citation Trends

Table 1. Papers included in the review

Author	Application	Opportunities	Challenges	Algorithms	Evaluation Metrics
(Zhang et al., 2024)	1. Patient-specific virtual profiles 2. Simulation of disease progression 3. Behavioral and lifestyle modelling	1. Personalized Disease Management 2. Real-Time Monitoring and Predictive Analytics 3. Integration of Multiomic Data 4. Improved Clinical Decision-Making 5. Identification of Novel Disease Targets	1. Data Sparsity 2. Data Quality 3. Complexity of Biological Systems 4. Data Integration 5. Generalizability 6. Interpretability of Predictions	1. Lasso Regression 2. Linear Regression 3. Logistic Regression 4. Random forests 5. Support Vector Machines (SVM)	R^2 (Coefficient of Determination). F1 Score.
(Shamanna et al., 2024)	1. Patient-specific virtual profiles 2. Simulation of disease progression 3. Behavioral and lifestyle modelling	1. Personalized Treatment Strategies 2. Predictive Analytics 3. Behavioral and Lifestyle Modelling 4. Real-Time Monitoring and Feedback 5. Enhanced Patient Engagement 6. Improved Clinical Outcomes	1. Data Quality 2. Data Integration 3. Personalization 4. Scalability 5. Generalizability 6. Technical Challenges 7. Ethical considerations 8. clinical acceptance	1. Gradient-Boosted Decision Trees 2. Artificial Neural Network (ANN) 3. Long Short-Term Memory (LSTM)	Health-Specific
(Joshi et al., 2023)	1. Patient-specific virtual profiles 2. Simulation of disease progression 3. Behavioral and lifestyle modelling	1. Personalized Nutrition and Lifestyle Recommendations 2. Prediction of Postprandial Glycemic Responses (PPGRs) 3. Real-Time Monitoring and Feedback 4. Simulation of Disease Progression 5. Improved Clinical Decision-making	1. Data Integration 2. Data Quality 3. Personalization 4. Technological Accessibility 5. Ethical considerations 6. Scalability		Health-Specific
(Thamotharan et al., 2023)	1. Patient-specific virtual profiles 2. Simulation of disease progression 3. Behavioral and lifestyle modelling	1. Personalized Treatment Plans 2. Real-Time Monitoring and Feedback 3. Predictive Analytics. 4. Enhanced Decision Support 5. Behavioral Insights and	1. Data Quality 2. Data Availability 3. Data Integration 4. Model complexity and interpretability 5. Personalization 6. Scalability 7. Continuous Learning	1. Long Short-Term Memory (LSTM) 2. Structured Time-Series Analysis	Accuracy Health-Specific

		Lifestyle Intervention 6. Research and Development	and Adaptation		
(Kulkarni et al., 2024)	1. Patient-specific virtual profiles 2. Simulation of disease progression 3. Behavioral and lifestyle modelling	1. Personalized Treatment Plans 2. Real-Time Monitoring and Feedback 3. Simulation of Disease Progression 4. Behavioral and Lifestyle Interventions 5. Improved decision making 6. Research and Development	1. Complex and Chaotic Medical Data 2. Model complexity and interpretability 3. Data Quality 4. Data Sparsity	1. Denoising Autoencoder (DAE) 2. Broad Learning System (BLS) 3. k-Nearest Neighbor (KNN)	Accuracy
(Hasib et al., 2024)	1. Patient-specific virtual profiles 2. Behavioral and lifestyle modelling.	1. Personalized Treatment Plan 2. Predictive Modelling 3. Real-Time Monitoring and Feedback. 4. Support for Self-Management	1. Limited Scope of Factors Considered. 2. Generalizability 3. Data Sparsity 4. Complex Patterns in Data	1. Artificial Neural Network (ANN) 2. KNN Regression 3. Random Forest 4. Decision Trees	Accuracy
(Sarani et al., 2024)	1. Patient-specific virtual profiles 2. Simulation of disease progression 3. Behavioral and lifestyle modelling	1. Personalized Treatment Plans 2. Enhanced Prediction and Prevention 3. Improved clinical outcomes 4. Personalized Lifestyle Recommendations 5. Data Visualization	1. Data Integration 2. Data Quality 3. Data and Algorithm Bias	1. Reinforcement Learning (RL) 2. Recurrent Neural Networks (RNNs)	Health-Specific
(Shamanna et al., 2024)	1. Patient-specific virtual profiles 2. Predictive modeling of glycemic responses	1. Personalized Nutrition 2. Predictive Glycemic Control 3. Real-Time Monitoring and Feedback. 4. Improved Metabolic Outcomes	1. Data Quality 2. Generalizability 3. Data Variability 4. Overfitting model	1. CatBoostRegressor 2. Random Forest 3. LSTM	1. Mean Squared Error (MSE) 2. Root Mean Squared Error (RMSE) 3. Mean Absolute Error (MAE) 4. R^2
(Paramesh et al., 2020)	1. Patient-specific virtual profiles 2. Simulation of disease progression 3. Behavioral and lifestyle model	1. Personalized Nutrition Guidance 2. Predictive Modelling 3. Real-Time Monitoring and Feedback 4. Long-Term Health Monitoring 5. Behavioral Insights	1. Data Quality -Data Integration 2. Ethical considerations 3. Technical challenges 4. Operational Challenges		
(Shamanna et al., 2021)	1. Patient-Specific Virtual Profiles 2. Simulation of Disease Progression 3. Behavioral and Lifestyle Modelling	1. Personalized Treatment Plans 2. Real-time Monitoring and Feedback 3. Predictive Analytics 4. Improved Decision Making 5. Improved clinical outcomes 6. Nutritional Guidance	1. Data Quality 2. Data Integration 3. Personalization 4. Scalability 5. Generalizability 6. Ethical consideration	1. Gradient-boosted decision trees 2. Artificial Neural Networks	Health-Specific

(Shamanna et al., 2021)	1. Patient-Specific Virtual Profiles 2. Simulation of Disease Progression 3. Behavioral and Lifestyle Modelling	1. Personalized Treatment Plans 2. Real-time Monitoring and Feedback 3. Predictive Analytics 4. Improved Clinical Outcomes 5. Enhanced Patient Engagement	1. Data Quality 2. Data Integration 3. Patient Heterogeneity 4. Scalability 5. Model complexity and interpretability	Statistical

There were no publications relating to the research topic in 2022. In 2023, two papers were published with 53 citations, which in essence was probably a tiny decrease in both number and impact from 2021. In 2024, 6 papers were published, but the citation count fell back to 51. The increase in publications in 2024, despite a slight drop in citations, indicates growing interest in the topic of research. This is a positive sign for the fields as it shows that more researchers are tackling the subject, which will eventually lead to further advancements and impacts.

Research Origins

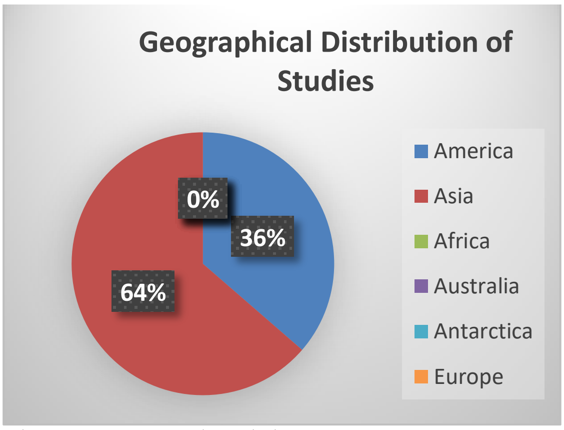


Figure 3. Research Origins

Figure 3 shows that almost two-thirds (approximately 64%) of the reviewed studies are associated with Asia's research ecosystem in diabetes management with ML and DT technologies. These studies have mainly focused on clinical applications, personalized treatment strategies, and decision support systems. There seems to be limited contributions from other continents, such as Africa, Antarctica, Australia, and Europe, revealing a promising gap for the pursuit of future inquiry.

DT Application in Studies

Four main application areas were identified, as depicted in Fig. 4.

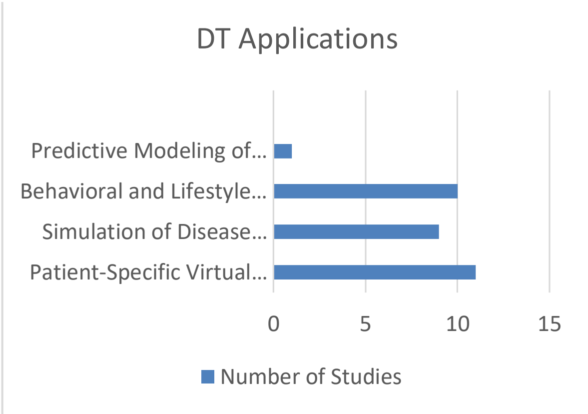


Figure 4. DT Application in Studies

All 11 studies involved people using patient-specific virtual profiles for individualized diabetes prediction based on their unique health data. Behavioral and lifestyle modeling were highlighted by approximately 27% (3) of the studies for the importance of personal habits and environmental factors in diabetes prediction. Only one study used the DT technology for the predictive modeling of glycemic responses in its integration with ML for diabetes management. On the other hand, approximately 18% (2) of the studies used a simulation of disease progression as to how predictive modeling would be important in showing how diabetes could develop over time.

Challenges Identified in Integrating ML and DT in Diabetes Management

Table 2 shows that data quality is the most common challenge when it comes to the integration of ML and DT for diabetes management, as seen by a total of 10 studies that support the notion. Data integration follows as the second most common challenge, with a record of 8 studies. Scalability is another issue, mentioned in 5 studies. Data sparsity is a concern in 3 studies. The generalizability of models is a major drawback, as noted by 5 studies. 3 different studies highlighted the complexity and interpretability of models as a contributing challenge. 2 studies stressed that ethical challenges arise when integrating ML and DT in diabetes management. One study emphasized clinical acceptance as a significant challenge in integrating ML and DT into diabetes management. Three studies reported personalization. Another study posed biological complexities as another challenge. Technical challenges associated with the integration of ML and DT in diabetes management were reported by two studies. A single study highlighted that operational challenges hinder integration, and another highlighted continuous learning and adaptation as a challenge in the integration process. Furthermore, another study suggested that a major challenge in the integration of ML and DT is overfitting models. Lastly, a single study mentioned a limited scope of factors as one of the challenges encountered in the integration of ML and DT for diabetes management.

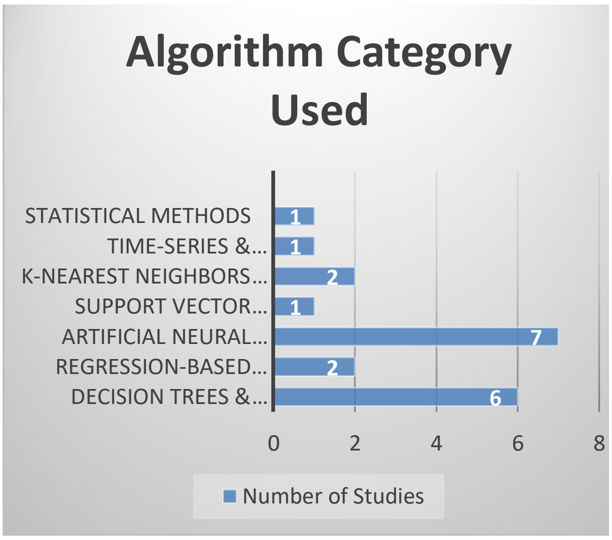


Figure 5. Algorithm Categories Used

Figure 5 shows the algorithm categories used with Artificial Neural Network methods (Custom ANNs, LSTM, RNN, Denoising Autoencoder [DAE], Broad Learning System [BLS], Reinforcement Learning [RL]) being used in 64% of the studies; followed by Decision Trees and Ensemble methods being used in 55% of the studies. Both KNN and Regression-based methods (Linear Regression, Logistic Regression, Lasso Regression [Optimized Linear/Logistic Regression]) were used in 18% of the studies, while statistical methods, Time-series & Structured Analysis, and support vector machines were used in one study.

Evaluation Metrics Used for the ML Algorithms

According to Table 3, the studies reviewed relied on various performance evaluation metrics for their models and interventions, and most of the health metrics reflect a considerable improvement. For example, one of the studies (Shamanna et al., 2021) established a 79.72% accuracy in predicting diabetes using the ANN model. One more study (Sarani et al., 2024) had the RMSE equal to 19.83 mg/dL in predicting glucose, thereby assuring the precision of the model in the blood glucose prediction process. Improvements in health indicators have included decreases by an average of 6.3% in BMI and from 9.0% down to 6.1% in HbA1c. Thus, it indicates the efficacy of the interventions impacting glycaemic control and improving all-around health outcomes in patients with type 2 diabetes. Time in Range (TIR) improved significantly, with patients achieving TIR percentages ranging from 86 to 97% (Shamanna et al., 2021). Other metric results such as F1 Score, were not explicitly mentioned.

Table 4. Opportunities of ML and DT in Diabetes Management

Opportunity	Author
Enhanced personalization and precision	(Alkhatib et al., 2017; Joshi et al., 2023; Sarani et al., 2024; Shamanna et al., 2021; Zhang et al., 2024)
Early detection and prevention	(Kulkarni et al., 2024; Shamanna et al., 2021; Thamotharan et al., 2023; Zhang et al., 2024)
Behavioral and lifestyle interventions	(Hasib et al., 2024; Joshi et al., 2023)
Scalability and Accessibility	(Hasib et al., 2024; Joshi et al., 2023)
Continuous improvement and learning	(Hasib et al., 2024; Joshi et al., 2023)

Applications of ML and DT in Diabetes Management

ML and DT are revolutionizing personalized healthcare in diabetes management by establishing virtual patient profiles, simulating disease progression, and modeling behavioral-lifestyle influences. This allows for the provision of more personalized care and effective interventions by the healthcare provider.

Patient-specific virtual profiles

Patient-specific virtual profiles refer to the DTs that are created for each patient (Sarani Rad et al., 2024). DTs provide highly detailed digital representations of individual patients, merging clinical and real-time data streams, including continuous glucose monitors, wearables, boost detection, and electronic health records (Kulkarni et al., 2024; Thamotharan et al., 2023). These virtual profiles offer a complete picture of a patient's health status and thus provide an opportunity for tailoring treatment to that patient. For instance, one study using DTs reported improvements of about 86 to 97% in Time in Range (TIR) among patients (Shamanna et al., 2021). DTs thus hold promise for better glycemic control and lower risk for complications related to diabetes.

Simulation of disease progression

Simulation of disease progression refers to the use of computational models, such as DTs, to predict how a disease, like diabetes, will evolve in an individual, based on their unique physiological data and response to interventions (Shamanna et al., 2024). ML algorithms, in particular those of predictive model, have been able to predict important health metrics such as HbA1c levels, glucose concentration, and all other clinically relevant variables with well-defined accuracy (Sarani Rad et al., 2024; Shamanna et al., 2021). These models utilize very large datasets to recognize patterns and relationships that can then be used to personalize treatment plans, enhancing their accuracy and effectiveness in diabetes management (Zhang et al., 2024). In diabetes reversal stage prediction, ANN model accuracy has been reported to be 79.72% in one study (Shamanna et al., 2021). Similarly, another study implied that the model is capable of accurately predicting blood glucose levels with an RMSE of 19.83 mg/dL in their glucose prediction (Sarani et al., 2024).

Behavioral and lifestyle modeling

Behavioral and lifestyle modeling involves the approach of addressing several elements of one's behavior and lifestyle: clinical care, nutritional education, physical activity, and psychological support to optimize diabetes prevention and management (Alkhatib et al., 2017; Maguraushe & Ndlovu, 2024). ML algorithms combined with DTs allow near real-time-to-real-time monitoring and adaptive changes to treatment plans (Hasib et al., 2024). This is very useful for managing type 2 diabetes, where individual responses to treatment are extremely variable, as treatment plans can also be automatically adapted (Joshi et al., 2023). Healthcare providers can make better decisions through the integration of ML algorithms and DTs, optimize treatment protocols, and thus improve patient outcomes. For example, a study using DTs was able to realize significant improvements in behavioral and lifestyle factors such as physical activity and dietary adherence (Joshi et al., 2023).

Opportunities of ML and DT in Diabetes Management

The application of ML and DT in Diabetes management opens doors to opportunities for new improvements toward an even more precise and personalized form of care that can yield better patient outcomes.

Enhanced Personalization and Precision

DTs provide true copies of individual patients. DTs handle the real-time integration of data from continuous glucose monitors, wearable devices, and electronic health records into fully-fledged virtual replicas of the patients (Alkhatib et al., 2017; Zhang et al., 2024). By pairing vast amounts of patient data with ML algorithms, DTs can discover unique patterns and links for formulating treatments highly personalized for each individual (Sarani et al., 2024; Shamanna et al., 2021). Such characteristics come in handy in the treatment of type 2 diabetes, a disease wherein treatments are highly individualized. For instance, according to research, these machines can predict the stages of diabetes reversal with an accuracy of 79.72%, while showing a precision of RMSE 19.83 mg/dL in predicting the glucose levels. Such predictions provide insight for healthcare providers adopting interventions more likely to yield the desired outcome (Fuyana et al., 2025), thereby improving diabetes management.

Early Detection and Prevention

According to research studies (Kulkarni et al., 2024; Thamotharan et al., 2023; Zhang et al., 2024), DTs can offer the opportunity for early detection and prevention of diabetic complications by simulating how the disease will progress and providing an overview of a patient's complete health status. In one study that used DTs, TIR percentages increased significantly, ranging from 86% to 97%

(Shamanna et al., 2021). This illustrates the possibility of DTs for benefitting glycemic control and may also serve to prevent diabetes-related complications, namely, cardiovascular diseases or chronic kidney disease.

Behavioral and Lifestyle Interventions

Continuous monitoring and adaptation of therapies can be used in conjunction with ML and DT technologies to meet the long and complex treatment courses of type 2 diabetes (Hasib et al., 2024; Joshi et al., 2023). By introducing these ML algorithms to DTs, one can leverage healthcare providers' information for evidence-based treatment optimization while improving the chances for successful treatment outcomes. For instance, DTs were reported to have significantly improved some behavior and lifestyle aspects, like increased physical activity and better dietary compliance (Joshi et al., 2023). It indicates that targeted interventions addressing some of the behavioral and lifestyle characteristics will be developed through DTs, ultimately aimed at combating the advancement of diabetes due to such factors.

Scalability and Accessibility

DTs can easily be scaled and made accessible to a much wider patient population, thereby broadening the reach and impact of personalized diabetes management (Hasib et al., 2024; Joshi et al., 2023). Such technologies can integrate within existing healthcare systems to enable healthcare providers to offer more effective and efficient care to patients with type 2 diabetes. For example, one study utilizing such tools reported very positive improvements in glycemic control and other health parameters in a wide variety of patient populations (Joshi et al., 2023). This shows that DTs could thus improve larger diabetes management outcomes.

Continuous Improvement and Learning

DT-based algorithms can continue to learn with the new data presented before them and, hence, constantly refine and improve models on diabetes management (Hasib et al., 2024; Joshi et al., 2023). This very capability ensures that the treatment plans get revised and updated along with changes in patient situations and healthcare practices. For instance, a study reported continuous improvement in prediction accuracy and treatment outcomes over time using ML-based DTs (Hasib et al., 2024). Therefore, it suggests that DTs can create dynamic and adaptable diabetes management systems along with patients and healthcare.

Challenges in ML and DT for Diabetes Prediction and Management

Though ML and DT have some potential opportunities, there are some obstacles to their introduction into diabetes prediction and management; hence, these should be addressed to realize the benefits of ML and DT in Diabetes management.

Data quality plays a vital role in determining the efficacy of ML models and decision trees. Imprecise or incomplete data are unable to yield good predictions or treatment plans (Mutunhu et al., 2024a; Sarani Rad et al., 2024; Shamanna et al., 2021). High-quality data would require rigorous processes of data collection, validation, and cleaning. For instance, a study mentioned the role of data quality in predicting reasonable diabetes reversal stages (Shamanna et al., 2021). The study highlighted that reliable predictive models require high-quality data, involving accurate measurements of HbA1c, glucose levels, and other clinical metrics, which are data of varying quality. Misclassification of diabetes stages as a result of erroneous data may direct treatment recommendations toward improper paths.

Data integration from sources such as electronic health records (EHRs), wearable devices, and continuous glucose monitors (CGMs) is necessary for creating a complete patient profile (Maguraushe & Ndlovu, 2024). However, due to different data formats, standards, and interoperability (Kulkarni et al., 2024; Thamotharan et al., 2023), data integration becomes a challenge. The study by Thamotharan et al. (2023) raises the issue of standardizing data integration for the enhancement of DTs' accuracy and reliability. It underscores the idea that integrating diverse sources of information contends with multiple technical barriers, which include making data consistent and compatible, thus compromising any overarching perspective on the patient's health situation envisioned by DTs.

Scalability is another very important challenge. ML models and DTs have shown promise in small-scale studies, whereas scaling these technologies to very large patient populations requires more computational resources and very efficient data management systems (Hasib et al., 2024; Joshi et al., 2023). A study by Faruqui et al. showed the immense need for scalable infrastructure to support the affluent uptake (Hasib et al., 2024). They further stated that scaling of DTs involves the increase of hardware but also the optimization of data storage and retrieval machinery. If ML and DTs have no scalable infrastructure, their benefits would remain confined to small, controlled environments with no reach to the larger patient population.

The ML model performance can be hindered in cases of data sparsity when some data types are simply missing or are not sufficiently represented in the existing data (Kulkarni et al., 2024; Sarani et al., 2024). Herein, this is most relevant for diabetes management, wherein data regarding certain physiological parameters may be limited. A study by Sarani Rad et al. (2024) pointed out data sparsity as being the formidable enemy of model accuracy. The study remarked that an absolute lack of good data on crucial parameters- insulin resistance or dietary intake would be calamitous to accurate predictive modeling. Thus, data sparsity poses tedious tasks of setting up new data collection methods and imputation techniques to improve the quality of the data.

Importance should be given to ensuring the generalizability of ML models and DTs across varying patient populations and settings. Models developed on a specific dataset may not perform effectively when used on a new population or in new settings (Joshi et al., 2023). Regarding diabetes management, a study planned on the generalizability of ML models showed the need for more diverse training datasets to confer some robustness to the model (Joshi et al., 2023).

ML models, and deep learning algorithms, in particular, are usually built on very complex architectures, rendering them more difficult to interpret. Such opaqueness can limit clinical acceptance because healthcare providers may be hesitant to put their trust in models whose actual decision mechanisms are poorly understood by them (Thamotharan et al., 2023). Correspondingly, DT ingests enormous amounts of data from a huge variety of sources, thus complicating any effort to assign causal importance to individual PDs of interest (Sarani Rad et al., 2024). Equally important issues pertain to ethics: the use of patient data raises serious concerns about privacy and protecting individuals' data, especially when these data are shared across diverse platforms and stakeholders (Shamanna et al., 2024). Patients' ethical standards and confidentiality should be enforced to allow a wide-ranging acceptance of ML and DT technologies.

Acceptance of ML and DT technologies among clinicians is further complicated by personalization issues (Joshi et al., 2023; Shamanna et al., 2021; Thamotharan et al., 2023). Genetic, environmental, and lifestyle factors are responsible for one's diabetes profile. Personalized predictions and interventions demand models that can account for that variability (Hasib et al., 2024). The complexity of biological systems complicates this, and diabetes is multidimensional, involving many intricate factors. Also, technical challenges in nature, like data quality and interoperability, make implementation difficult in these technologies (Shamanna et al., 2024). Operational challenges related to continuous learning and adapting add another layer of difficulty. Models need to include new data regularly to keep their accuracy and relevance, which entails strong infrastructure and resources (Paramesh et al., 2020). Overfitting- when models perform very well against training data but perform poorly with new data, another aspect where ML realizes that, in general, it seems to defeat the purpose concerning the reliability of diabetes predictions (Shamanna et al., 2021). Finally, the very limited set of factors used in the currently available models may very well ignore some of the very critical factors, leading to incomplete or incorrect predictions (Sarani et al., 2024). All these need to be solved to bring AI and DT closer to diabetes management.

ML Algorithms Applied in ML and DT for Diabetes Management

Lasso Regression is a linear model that carries out L1 regularization, being capable of shrinking some coefficients down to zero and performing feature selection in the process (Yang et al., 2025; Zhang et al., 2025). It is mostly useful for datasets with many features in the search for the most relevant ones to predict the outcome (Xia et al., 2023). In diabetes, Lasso Regression can help in determining some important factors that affect blood glucose levels and other metabolic markers (Zhang et al., 2024). The limitations of the Lasso Regression factor are the sensitivity in the selection of the regularization parameter and the diminishing degree of interpretability due to feature elimination (Zhang et al., 2024). It is being used in predicting responses to glycemia and in the optimization of treatment protocols, considering the most influential predictors (Thamotharan et al., 2023).

Linear Regression is an essential algorithm that can predict a specific dependent variable with the help of one or more independent variables (Lederer, 2022; Ogundokun et al., 2020). Such techniques are straightforward, easy to interpret, and therefore are mainly used for preliminary analysis in diabetes management (Zhang et al., 2024). Linear Regression can be used for the prediction of a continuous outcome, like blood glucose, based on previous records and other covariates. However, it is a very simple technique that may be useful for the interpretation and conclusions, but it assumes that all variables have a linear relationship with each other, which may not always be the case; it's also overfitting (Y. Zhang et al., 2024). Nevertheless, it provides a common model for preliminary studies of comparisons with more complicated algorithms (Hasib et al., 2024).

Logistic regression is used to classify binary events, such as the occurrence of diabetes or the possibility of remission (Zhang et al., 2024). It can take the form of a function that gives the likelihood of any event concerning the input variables (Chiramba et al., 2024). Due to its robustness and interpretability, Logistic Regression is very much applicable to diabetes care in clinical research (Austin et al., 2022). It assumes, however, a linear relation between the log of the odds of the outcome concerning the predictors, which may often not prove to be true. Such studies are already showing the odds ratio of diabetes-related complications and investigating whether interventions are working or not for patients (Sarani et al., 2024).

In addition to regression-based algorithms, Random Forests, which are ensemble methods of sorts that join multiple decision trees, are also used for their robustness in dealing with non-linear data and reducing overfitting (Shamanna et al., 2021; Mukura & Ndlovu, 2023; Zhang et al., 2024). Applications in personalized treatment planning and predictive modeling, including glycemic responses, have been increasingly experimented with (Shamanna et al., 2021; Shamanna et al., 2024), but the interpretability may become more limited in highly complex modeling.

Support Vector Machines (SVMs) are classification tools useful due to their capability of finding optimal hyperplanes in high-dimensional space (Zhang et al., 2024). They have been considered for both regression and classification tasks aimed at forecasting the clinical trajectories of key variables such as glycated hemoglobin (HbA1c), glucose, insulin, HOMA-IR, and estimated glomerular filtration rate (eGFR) (Zhang et al., 2024); however, they need careful parameter tuning and can be computationally expensive.

The k-Nearest Neighbors (KNN) algorithm has been utilized in some regression tasks, for example, in predicting glycemic levels (Hasib et al., 2024; Kulkarni et al., 2024). Due to the ability to work with large amounts of data, KNN is always bad on scalability. This means that computing power is largely tortured by calculation when a lot of datasets are involved. Gradient-boosted Decision Trees for predictive analytics and personalized treatment plans performed fairly well but were also subject to careful calibration to avoid overfitting (Shamanna et al., 2021; Sibindi et al., 2024; Shamanna et al., 2024). The relative strengths and weaknesses determine algorithm suitability for application in diabetes management; hence, choosing algorithms based on the characteristics and requirements of the problem is crucial.

Further applied in this study are Denoising Autoencoders (DAE), Broad Learning Systems (BLS), and Reinforcement Learning (RL) programs specifically developed for addressing different challenges in diabetes management, such as the treatment of noisy data and treatment strategy optimization within time (Kulkarni et al., 2024). CatBoostRegressor is yet another important algorithm for the extension to gradient boosting that can properly consider categorical features (Shamanna et al., 2024). Combined with the DT technology, these algorithms enable the creation of patient-specific virtual profiles and simulation of disease progression, personalized treatment plans, and change-behavior interventions (Paramesh et al., 2020; Shamanna et al., 2024; Zhang et al., 2024). The choice of an algorithm for a specified problem is dependent very much on the application of the algorithm, characteristics of the data, and evaluation metrics; hence the importance of choosing the right method for the specified problem.

Implications of research findings

Theoretical Implications

The most outstanding classes in algorithms are ANNs, Decision Trees & Ensemble Methods, with the number of studies being 7 and 6, respectively. Therefore, this strong dependence on these techniques for predictive analytics and personalized treatment planning suggests that these methods may be effective; however, others may be underexploited.

With ANNs and ensemble methods being by far the most commonly used, an opportunity for further diversity in model platforms has been dismissed. Time-Series & Structured Analysis and Support Vector Machines (SVMs) among other algorithms, are featured in only one study each, thus signaling a potential gap in further exploration of the applicability of these methodologies. Statistical methods could serve, for example, as robust frameworks for hypothesis testing and causal inference, which are important for gaining an understanding of the mechanisms governing the progression of diabetes. Likewise, time-series analysis could allow for a more thorough investigation of the temporal behavior of glycemic response and lifestyle factors, in turn helping screen for predictive model accuracy.

In addressing such gaps, it is recommended that future research include a wider range of algorithms and analytics techniques. It can be expected that a broader application of statistical methods and time-series analysis can provide models that predict outcomes and also explain the relationships among the various factors affecting diabetes management. The strengthening of model performance and robustness, especially concerning the handling of complex, high-dimensional data, may also benefit from the combination of SVMs and other less-studied methods.

Moreover, it is very important to have a systematic comparison of the performance of different algorithms. To find models that are most useful for a given task, researchers evaluate and benchmark them using metrics such as accuracy, F1 score, R2, and others. The comparative analysis becomes necessary so that for such particular concerns as diabetes management, an algorithm of high performance is not selected at the expense of letting an accurate and reliable model pass out. This approach to the theoretical implications, algorithm performance comparison, and improving data collection processes could increase the effectiveness of ML and DT technologies toward treating diabetes more holistically in the future.

Practical Implications

This research project highlights the real-world applications of ML and DT technologies for diabetes management in countries around the world. Research conducted in America and Asia, as well as some other countries, has demonstrated that there is interest in using these approaches to combat the increasing burden of diabetes. In America, this has been focused on therapeutic decisions and simulation models of disease progression tailored to virtual patient profiles. This should increase health outcomes by considering individual patient characteristics and needs to improve treatments.

The studies discussed have contributed substantially to an understanding of the complex interplay between genetics, environment, lifestyle, and the progression of diabetes. The virtual avatars of patients and simulation models give clinicians a solid framework with which to visualize and understand potential outcomes for different treatment strategies, thereby improving the overall decision-making process in the clinic. Digital infrastructures deliver care in real time and enable feedback for early intervention and chronic disease management. Patients receive health diagnoses and therapy advice that is tailored to suit their needs. While this may not address psychological or social aspects of the condition, these interventions improve the utility of models greatly in diabetes management, especially in behavioral and lifestyle management.

However, several hurdles have to be overcome before realizing the full potential of these technologies. The major urgent challenges include data quality and integration, generalizability itself, and specific issues arising from the complexity and heterogeneity of the data sets involved. Predictions will need a lot of reliability and interpretability to win the trust of health professionals and patients. Lastly, ethical issues, data privacy, and compliance considerations will need the utmost care to ensure the responsible use of such technologies.

Table 2. Challenges Identified

Challenge	Number of Studies	References
Data Quality	10	(Zhang et al., 2024), (Sarani Rad et al., 2024), (Shamanna et al., 2024), (Joshi et al., 2023), (Thamotharan et al., 2023), (Kulkarni et al., 2024), (Shamanna, et al., 2024), (Paramesh et al., 2020), (Shamanna et al., 2021), (Shamanna, Joshi, et al., 2021)
Data Integration	8	(Zhang et al., 2024), (Sarani Rad et al., 2024), (Shamanna et al., 2024), (Joshi et al., 2023), (Thamotharan et al., 2023), (Paramesh et al., 2020), (Shamanna et al., 2021), (Shamanna, Joshi, et al., 2021)
Scalability	5	(Shamanna et al., 2024), (Joshi et al., 2023), (Thamotharan et al., 2023), (Shamanna et al., 2021), (Shamanna, Joshi, et al., 2021)
Data Sparsity	3	(Zhang et al., 2024), (Kulkarni et al., 2024), (Hasib et al., 2024)
Generalizability	5	(Zhang et al., 2024), (Shamanna et al., 2024), (Hasib et al., 2024), (Shamanna, Joshi, Thajudeen, et al., 2024), (Shamanna et al., 2021)
Model Complexity and Interpretability	3	(Thamotharan et al., 2023), (Kulkarni et al., 2024), (Shamanna et al., 2021)
Ethical considerations	2	(Paramesh et al., 2020), (Shamanna et al., 2021)
Clinical Acceptance	1	(Shamanna et al., 2024)
Personalization Challenges	4	(Joshi et al., 2023), (Thamotharan et al., 2023), (Shamanna et al., 2021)
Complexity of Biological Systems	1	(Zhang et al., 2024)
Technical Challenges	2	(Shamanna et al., 2024), (Paramesh et al., 2020)
Operational challenges	1	(Paramesh et al., 2020)
Continuous Learning and Adaptation	1	(Thamotharan et al., 2023)
Overfitting model	1	(Shamanna et al., 2024)
Limited Scope of Factors Considered	1	(Hasib et al., 2024)

Table 3. Evaluation Metrics

Metric	Number of Studies
F1 Score	1
R^2 (Coefficient of Determination)	2
Accuracy	3
Mean Squared Error (MSE)	1
Root Mean Squared Error (RMSE)	1
Mean Absolute Error (MAE)	1
Specific Health Metrics (e.g. HbA1c, BMI, TIR)	5
Statistical	1

CONCLUSION

This SLR examined the management of diabetes using ML and DT technologies. The combination of ML and DT provides a lot of opportunities which include personalized treatment plans, improved clinical outcomes, and improved decision-making. ANNs have been increasingly used in the training of ML models that serve as diabetes risk predictors.

Different evaluation metrics were used, and in some studies, no ML-specific metrics were mentioned, but only health-specific metrics like HbA1c, BMI, TIR, etc. were mentioned. This further makes it hard to evaluate, compare, and determine the efficient algorithm among others as other studies used standard ML evaluation metrics like F1 Score, MSE, RMS, and Accuracy. Hence, a comparative analysis should be conducted, using the standard ML evaluation metrics to determine the best-performing model among the ML models. Furthermore, the prediction results should be presented as percentage outputs, that is, the likelihood of getting diabetes based on the patient data input rather than just class variables as it will help in determining intervention measures to be taken. Lastly, the collaboration between Data Scientists and other stakeholders should be encouraged to enforce proper data collection procedures for high-quality data acquisition.

This study has inherent limitations. First, the analysis was restricted to three databases, which may not have included other relevant research on the prediction of diabetes. Second, the emphasis given to English-language literature would hold a language bias, excluding important findings published in other languages. Finally, the majority of evidence gathered from published studies may contain publication bias, as grey literature and studies that did not get published were excluded from this study. While the present study was therefore limited in some respects, it does offer insight into the application of ML and DT in diabetes management and serves as a platform for subsequent research and development in this essential area of preventive health care.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

1. Abdi, A., Jalilian, M., Sarbarzeh, P. A., & Vlaisavljevic, Z. (2020). Diabetes and COVID-19: A systematic review on the current evidences. *Diabetes Research and Clinical Practice*, *166*, 108347.
2. Agrawal, A., Fischer, M., & Singh, V. (2022). Digital Twin: From Concept to Practice. *Journal of Management in Engineering*, *38*(3).
3. Alkhatib, A., Tsang, C., Tiss, A., Bahorun, T., Arefanian, H., Barake, R., Khadir, A., & Tuomilehto, J. (2017). Functional foods and lifestyle approaches for diabetes prevention and management. In *Nutrients* (Vol. 9, Issue 12). MDPI AG.
4. Austin, A. M., Ramkumar, N., Gladders, B., Barnes, J. A., Eid, M. A., Moore, K. O., Feinberg, M. W., Creager, M. A., Bonaca, M., & Goodney, P. P. (2022). Using a cohort study of diabetes and peripheral artery disease to compare logistic regression and machine learning via random forest modeling. *BMC Medical Research Methodology*, *22*(1), 300.
5. Cappon, G., & Facchinetti, A. (2024). Digital Twins in Type 1 Diabetes: A Systematic Review. In *Journal of Diabetes Science and Technology*. SAGE Publications Inc.
6. Cellina, M., Cè, M., Alì, M., Irmici, G., Ibba, S., Caloro, E., Fazzini, D., Oliva, G., & Papa, S. (2023). Digital Twins: The New Frontier for Personalized Medicine? *Applied Sciences*, *13*(13), 7940.
7. Chu, Y., Li, S., Tang, J., & Wu, H. (2023). The potential of the Medical Digital Twin in diabetes management: a review. In *Frontiers in Medicine* (Vol. 10). Frontiers Media SA.
8. Frank, M., Drikakis, D., & Charissis, V. (2020). Machine-learning methods for computational science and engineering. In *Computation* (Vol. 8, Issue 1). MDPI Multidisciplinary Digital Publishing Institute.
9. Fuyana, C., Ndlovu, B., Dube, S., Maguraushe, K., & Malungana, L. (2025). *Optimizing HIV Care Through Machine Learning-Assisted Prediction and Personalized Treatment BT - Evolution in Computational Intelligence*. Springer Nature Singapore.
10. Hasib, S., Faruqui, A., Alaeddini, A., Du, Y., Li, S., Sharma, K., & Wang, J. (2024). *Nurse-in-the-Loop Artificial Intelligence for Precision*

Management of Type 2 Diabetes in a Clinical Trial Utilizing Transfer-Learned Predictive Digital Twin.

11. Joshi, S., Shamanna, P., Dharmalingam, M., Vadavi, A., Keshavamurthy, A., Shah, L., & Mechanick, J. I. (2023). Digital Twin-Enabled Personalized Nutrition Improves Metabolic Dysfunction-Associated Fatty Liver Disease in Type 2 Diabetes: Results of a 1-Year Randomized Controlled Study. *Endocrine Practice*, *29*(12), 960–970.
12. Kulkarni, C., Quraishi, A., Raparthi, M., Shabaz, M., Khan, M. A., Varma, R. A., Keshta, I., Soni, M., & Byeon, H. (2024). Hybrid disease prediction approach leveraging digital twin and metaverse technologies for health consumer. *BMC Medical Informatics and Decision Making*, *24*(1).
13. Lederer, J. (2022). *Linear Regression* (pp. 37–79).
14. Li, H., Tian, S., Chen, T., Cui, Z., Shi, N., Zhong, X., Qiu, K., Zhang, J., Zeng, T., Chen, L., & Zheng, J. (2020). Newly diagnosed diabetes is associated with a higher risk of mortality than known diabetes in hospitalized patients with COVID‐19. *Diabetes, Obesity and Metabolism*, *22*(10), 1897–1906.
15. Maguraushe, K., & Ndlovu, B. M. (2024). The use of smart technologies for enhancing palliative care: A systematic review. Digital Health, 10, 20552076241271836.
16. Maguraushe, K., Ndayizigamiye, P. (2024). Towards a Smart Healthcare System for Non-Communicable Diseases (NCDs) Management: A Bibliometric Analysis. In: Masinde, M., Möbs, S., Bagula, A. (eds) *Emerging Technologies for Developing Countries. AFRICATEK 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol 520. Springer, Cham.
17. Meijer, C., Uh, H. W., & el Bouhaddani, S. (2023). Digital Twins in Healthcare: Methodological Challenges and Opportunities. In *Journal of Personalized Medicine* (Vol. 13, Issue 10). Multidisciplinary Digital Publishing Institute (MDPI).
18. Mosquera-Lopez, C., & Jacobs, P. G. (2024). Digital twins and artificial intelligence in metabolic disease research. In *Trends in Endocrinology and Metabolism* (Vol. 35, Issue 6, pp. 549–557). Elsevier Inc.
19. Moyo, N., Moyo, S., & Mutunhu, B. (2022). Mask-Up: A Face Mask Alert App Using Machine Learning. *2022 IST-Africa Conference, IST-Africa 2022*, 1–8.
20. Mpofu, L., Ndlovu, B., Dube, S., Muduva, M., Jacqueline, F., & Maguraushe, K. (2024, April 23). Predictive Model for Hospital Readmission of Diabetic Patients. *Proceedings of the International Conference on Industrial Engineering and Operations Management*.
21. Mtshali, N.C.W., Ndayizigamiye, P., Govender, I., Maguraushe, K. (2024). Fostering Youth Wellbeing Through mHealth Apps: Embracing Physical Activity for a Healthier Lifestyle. In: Sharma, S.K., Dwivedi, Y.K., Metri, B., Lal, B., Elbanna, A. (eds) *Transfer, Diffusion and Adoption of Next-Generation Digital Technologies. TDIT 2023. IFIP Advances in Information and Communication Technology,* vol 698. Springer, Cham.
22. Murere, I., Ndlovu, B., Dube, S., Muduva, M., & Jacqueline Kiwa, F. (2024, July 16). Comparative Analysis of Machine Learning Techniques for Predicting Diabetes. *Proceedings of the International Conference on Industrial Engineering and Operations Management*.
23. Mutunhu, B., Chipangura, B., & Singh, S. (2024a). An Exploration of Opportunities for Quantified-Self Technology in Diabetes Self-Care: A Systematic Literature Review. *J Health Inform Afr*, *11*(2), 17–30.
24. Mutunhu, B., Chipangura, B., & Singh, S. (2024b). Towards a quantified-self technology conceptual framework for monitoring diabetes. *Suid-Afrikaanse Tydskrif Vir Natuurwetenskap En Tegnologie*, *43*(1), 69–84.
25. Mutunhu, B., Chipangura, B., & Twinomurinzi, H. (2022). Internet of Things in the Monitoring of Diabetes. *International Journal of Health Systems and Translational Medicine*, *2*(1), 1–20.
26. Mutunhu, B., Chipangura, B., & Twinomurinzi, H. (2023). *A Systematized Literature Review: Internet of Things (IoT) in the Remote Monitoring of Diabetes* (pp. 649–660).
27. Ndlovu, B. M., Chipangura, B., & Singh, S. (2024). Factors Influencing Quantified

SelfTechnology Adoption in Monitoring Diabetes. In X.-S. Yang, S. Sherratt, N. Dey, & A. Joshi (Eds.), *Proceedings of Ninth International Congress on Information and Communication Technology* (pp. 469–479). Springer Nature Singapore.

28. Ndhlovu, P., Maguraushe, K., Ndayizigamiye, P. and Idemudia, E. C. (2023). The effect of smartwatch features on patient-centred healthcare. (2023). *ACIS 2023 Proceedings*. 19.
29. Ogundokun, R. O., Lukman, A. F., Kibria, G. B. M., Awotunde, J. B., & Aladeitan, B. B. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infectious Disease Modelling*, *5*, 543–548.
30. Paramesh Shamanna, Banshi Saboo, Suresh Damodharan, Jahangir Mohammed, Maluk Mohamed, Terrence Poon, Nathan Kleinman, & Mohamed Thajudeen. (2020). Reducing HbA1c in Type 2 Diabetes Using Digital Twin Technology-Enabled Precision Nutrition: A Retrospective Analysis. In *Brill's Studies in Intellectual History: Vol. 270/21* (pp. 1–467). Brill Academic Publishers.
31. Rane, N. L., Paramesha, M., Choudhary, S. P., & Rane, J. (2024). *Partners Universal International Innovation Journal (PUIIJ) Machine Learning and Deep Learning for Big Data Analytics: A Review of Methods and Applications*.
32. Sarani Rad, F., Hendawi, R., Yang, X., & Li, J. (2024). Personalized Diabetes Management with Digital Twins: A Patient-Centric Knowledge Graph Approach. *Journal of Personalized Medicine*, *14*(4).
33. Shamanna, P., Dharmalingam, M., Sahay, R., Mohammed, J., Mohamed, M., Poon, T., Kleinman, N., & Thajudeen, M. (2021). Retrospective study of glycemic variability, BMI, and blood pressure in diabetes patients in the Digital Twin Precision Treatment Program. *Scientific Reports*, *11*(1).
34. Shamanna, P., Erukulapati, R. S., Shukla, A., Shah, L., Willis, B., Thajudeen, M., Kovil, R., Baxi, R., Wali, M., Damodharan, S., & Joshi, S. (2024). One-year outcomes of a digital twin intervention for type 2 diabetes: a retrospective real-world study. *Scientific Reports*, *14*(1).
35. Shamanna, P., Joshi, S., Dharmalingam, M., Vadavi, A., Keshavamurthy, A., Shah, L., Samajdar, S. S., & Mechanick, J. I. (2024). Digital Twin in Managing Hypertension Among People With Type 2 Diabetes: 1-Year Randomized Controlled Trial. *JACC: Advances*, *3*(9).
36. Shamanna, P., Joshi, S., Shah, L., Dharmalingam, M., Saboo, B., Mohammed, J., Mohamed, M., Poon, T., Kleinman, N., Thajudeen, M., & Keshavamurthy, A. (2021). Type 2 diabetes reversal with digital twin technology-enabled precision nutrition and staging of reversal: a retrospective cohort study. *Clinical Diabetes and Endocrinology*, *7*(1).
37. Shamanna, P., Joshi, S., Thajudeen, M., Shah, L., Poon, T., Mohamed, M., & Mohammed, J. (2024). Personalized nutrition in type 2 diabetes remission: application of digital twin technology for predictive glycemic control. *Frontiers in Endocrinology*, *15*, 1485464.
38. Singh, M., Fuenmayor, E., Hinchy, E. P., Qiao, Y., Murray, N., & Devine, D. (2021). Digital twin: Origin to future. In *Applied System Innovation* (Vol. 4, Issue 2). MDPI AG.
39. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Health and Treatment of Diabetes Mellitus. *International Journal of Health Sciences*, *5*(1), I–V.
40. Thamotharan, P., Srinivasan, S., Kesavadev, J., Krishnan, G., Mohan, V., Seshadhri, S., Bekiroglu, K., & Toffanin, C. (2023). Human Digital Twin for Personalized Elderly Type 2 Diabetes Management. *Journal of Clinical Medicine*, *12*(6).
41. Tomic, D., Shaw, J. E., & Magliano, D. J. (2022). The burden and risks of emerging complications of diabetes mellitus. In *Nature Reviews Endocrinology* (Vol. 18, Issue 9, pp. 525–539). Nature Research.
42. W.C Mukura, N., & Ndlovu, B. (2023, September 12). Performance evaluation of artificial intelligence in decision support system for heart disease risk prediction. *Proceedings of the International Conference on Industrial Engineering and Operations Management*.
43. Xia, L., Nan, B., & Li, Y. (2023). Debiased lasso for generalized linear models with a diverging number of covariates. *Biometrics*, *79*(1), 344–357.
44. Yang, X., Lan, W., Lin, C., Zhu, C., Ye, Z., Chen, Z., & Zheng, G. (2025). Atrial

fibrillation risk model based on LASSO and SVM algorithms and immune infiltration of key mitochondrial energy metabolism genes. *Scientific Reports*, *15*(1), 6681.

45. Zhang, E., Goto, R., Sagan, N., Mutter, J., Phillips, N., Alizadeh, A., Lee, K., Blanchet, J., Pilanci, M., & Tibshirani, R. (2025). *LLM-Lasso: A Robust Framework for Domain-Informed Feature Selection and Regularization*.
46. Zhang, Y., Qin, G., Aguilar, B., Rappaport, N., Yurkovich, J. T., Pflieger, L., Huang, S., Hood, L., & Shmulevich, I. (2024). A framework towards digital twins for type 2 diabetes. *Frontiers in Digital Health*, *6*.