International Journal of Pharmaceutics and Drug Analysis

Content Available at www.ijpda.org ISSN: 2348:8948

GENERATIVE ARTIFICIAL INTELLIGENCE FOR DIABETES MELLITUS PREDICTION AND INTELLIGENT CLINICAL DECISION SUPPORT

SANGEETA LALWANI[1], SHASHI KUMAR SINGH[2]

[1]*Research Scholar, Future University, Bareilly.*
[2]*Dean (Engineering), Future University, Bareilly.*

Received: 14.04.2026 Revised: 03.05.2026 Accepted: 15.06.2026

ABSTRACT

Diabetes mellitus is a major chronic metabolic disorder characterized by persistent dysregulation of blood glucose and progressive cardiovascular, renal, neurological, and ophthalmic complications. The global prevalence of diabetes has increased substantially, reaching approximately 14% among adults in 2022, with nearly 830 million people living with the disease. The rapid expansion of electronic health records, continuous glucose monitoring, laboratory databases, medical imaging, wearable technologies, and patient-generated health data has created opportunities for artificial intelligence (AI)-based diabetes prediction and clinical decision support. Conventional machine learning models demonstrate potential for diabetes detection and complication prediction; however, challenges remain regarding heterogeneous datasets, external validation, calibration, dataset shift, interpretability, and clinical implementation. This study proposes a systematic framework integrating generative AI, particularly large language models (LLMs), with validated predictive models for diabetes risk assessment and intelligent clinical decision support. A secondary evidence-synthesis approach was used, integrating findings from systematic reviews, clinical trials, diabetes technology research, and recent LLM literature. Evidence indicates that only 36% of evaluated machine learning models for diabetes complications demonstrated clearly useful discrimination. Generative AI has shown promise in treatment support, although performance decreases with increasing clinical complexity. ChatGPT-4 achieved 69.2–84.6% agreement with clinician prescribing in selected monotherapy cases but only 47.1–49.8% for dual- or triple-therapy cases. The findings support using generative AI as an orchestration layer for prediction, evidence retrieval, contextualization, explanation, documentation, and workflow support rather than as a replacement for validated prediction or clinical judgment. Safe implementation requires calibrated models, authoritative retrieval, uncertainty representation, auditability, privacy, fairness assessment, human oversight, and prospective clinical validation.

Keywords: *Generative AI; Diabetes Mellitus; Large Language Models; Clinical Decision Support; Machine Learning; Retrieval-Augmented Generation.*

***CORRESPONDING AUTHOR**

Sangeeta Lalwani

DOI: https://doi.org/10.47957/ijpda.v14i3.751

PRODUCED AND PUBLISHED BY

South Asian Academic Publications

1. INTRODUCTION

1.1 Background

Diabetes mellitus is a group of disorders of glucose regulation, metabolic homeostasis showing a progressive impairment. The impact of the condition is not limited to glycemic abnormalities and has led to it being recognized as one of the major global health challenges, with the consequences reaching cardiovascular disease, chronic kidney disease, neuropathy, retinopathy, complications of lower-limb and premature mortality[1]. According to the World Health Organization, the prevalence of adults with diabetes has risen from around 200 million in 1990 to 830 million in 2022. During the same period the prevalence of adult diabetes rose from 7% to 14%.

The epidemiological burden demands more timely identification of people with increased risk and more rapid diagnostic evaluation and ongoing surveillance and treatment decisions based upon that assessment. In traditional health care systems, blood tests, physical checks, diagnosis and interpretation by clinicians are used. While these mechanisms are still critical, the quantity and diversity of current health information has made the situation ripe for making human decisions with the assistance of computational approaches.

Diabetics prediction has been increasingly studied using AI techniques. Non-linear relationships between demographic factors, anthropometric measures, laboratory measurements, comorbidities, medication history and longitudinal clinical events can be identified by machine-learning algorithms [2]. The capability can be extended to high-dimensional information such as retinal images, continuous glucose monitoring data and complex electronic health-record representations via deep-learning models.

In recent years, systematic evidence shows that diabetes detection and diabetes complications prediction and risk stratification have already been done on machine learning. Based on a systematic review of machine-learning and deep-learning approaches, it was possible to identify applications over several datasets and algorithmic families, such as convolutional neural networks, support-vector machines, logistic regression and gradient-boosting approaches. A second systematic review of prediction of diabetes complications that included 32 studies and 87 models found that 36% of models showed clearly useful discrimination, as defined by the reviewers.

The results highlight an important distinction between algorithmic development and clinical utility. Clinically effectiveness cannot be guaranteed by the predictive accuracy within a development data set. It can be highly discriminating, yet provide an under-specified estimate of the probability; it can be highly discriminating, yet it fails to work well across different populations; it can be highly discriminating, yet it does not yield actionable clinical information.

1.2 Emergence of Generative Artificial Intelligence

Generative AI adds features that are unique from traditional predictive machine learning. In general, the output of a predictive model will be a classification, probability, or numeric prediction of structured or multimodal input variables [3]. Generative models can convert clinical data into natural-language explanations, summaries, questions, and recommendations and workflow outputs.

One of the key generative-AI technologies is large language models. Their natural-language processing capabilities open the door for integration of structured prediction with clinical documentation and medical knowledge. A generative model may also be able to combine lab data with medication history, glucose monitoring trends, clinical notes, and guideline recommendations into an actionable output for diabetes treatment.

The possible use isn't limited to diagnosis. The areas of patient education, clinical documentation, treatment-plan analysis, risk communication, retrieval of guidelines, and summarization of clinical information over time that span a patient's care are where generative AI can help [4]. Recent insights from 2024 into the use of large language models in diabetes care highlighted relevant areas of diabetes management and the necessity for further clinical testing.

The 2026 Standards of Care in Diabetes highlight the need for technology and personalized diabetes care and the use of evidence-based clinical practice. These standards are revised every year by the American Diabetes Association and include: glucose monitoring, insulin delivery, automated insulin delivery, connected devices and diabetes self-management support software. This setting will create a vital context for the incorporation of AI into the current digital diabetes landscape.

1.3 Research Problem

The research question is not just: Can artificial intelligence predict diabetes? Many studies have shown that machine learning algorithms can be useful in generating predictive classifications [5]. The larger question is how predictive AI technologies can be coupled with generative AI to provide clinically relevant decision support, but not be replaced by language generation alone.

The technological layers can be split into three layers. This is done in the first layer, which involves prediction, using statistical and machine-learning models to estimate the probability of developing diabetes or complications. The second is related to the retrieval of evidence, which involves the collection of clinical guidelines, validated knowledge sources and patient-specific evidence [6]. The third layer is generation, in which the LLM takes the retrieved information and the model's output and transforms it into a comprehensible clinical response.

A system without the first layer would generate good fluent clinical recommendations, but not very solid ones. If the second layer is skipped, a system can make recommendations that are not linked to the current standards of care. Such a system without the third layer can generate valid predictions that are hard to understand and/or to be incorporated into clinical routines.

1.4 Research Objectives

The primary objective of this paper is to develop an evidence-based framework for the application of generative artificial intelligence to diabetes mellitus prediction and intelligent clinical decision support.

The secondary objectives are to examine current AI approaches for diabetes prediction, assess evidence concerning generative AI in diabetes care, compare predictive and generative AI capabilities, identify measurable performance requirements, formulate an integrated architecture and establish methodological requirements for safe clinical implementation.

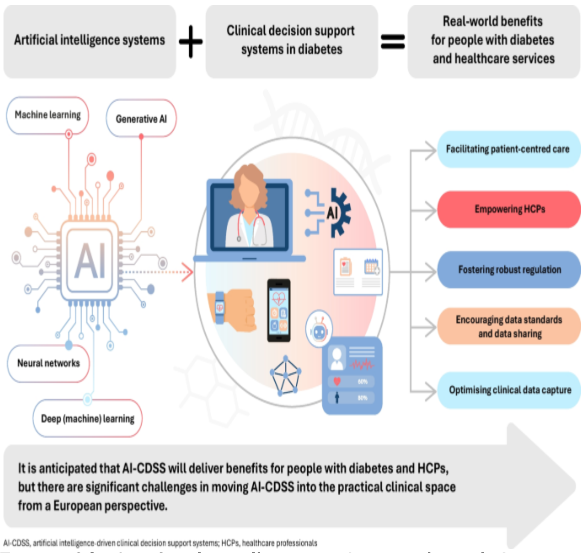

Figure 01: Artificial intelligence-driven clinical decision support systems to assist healthcare professionals and people with diabetes in Europe at the point of care

2. LITERATURE REVIEW

2.1 Artificial Intelligence in Diabetes Prediction

Due to recent developments in artificial intelligence, the diabetes prediction application has evolved from traditional statistical classification to ensemble learning, deep learning, and multimodal modelling. Structured variables like age, BMI, blood pressure, glucose levels, insulin levels, family history were early ones used. Later models include longitudinal aspects of clinical data, imaging and genetic information.

It is pointed out in the literature that the choice of algorithms depends greatly on the structure of the data sets [7]. Logistic regression is still employed due to its interpretability and transparency. Decision trees offer interpretable rule based structures and random forest are ensembles of decision trees, offering robustness. Gradient-boosting methods like Boost are able to model nonlinear relationships and interactions. Neural networks can encode complex relationships in a large data set, and convolutional networks are pertinent to the imaging of the retina and other spatial data sets.

According to a systematic review of diabetes complication prediction, the most commonly used diabetes models were neural network, of which 15 neural-network models were evaluated. Many common predictors were used, such as age and diabetes duration or body mass index. In that review, random forest performed the best overall in terms of the microvascular and macrovascular outcomes.

The very same evidence also shows a huge methodological flaw [8]. The majority of studies were rated as having high risk of bias and external validation was found to be limited. Therefore, prediction results that have been reported at the development stage of the model should not be interpreted as clinical effectiveness.

2.2 Deep Learning and Multimodal Diabetes Prediction

Deep learning affords the chance to learn information that may not be conveniently summarized in tabular predictors. Retinal images are a prominent example, since the microvascular damage caused by diabetes is associated with DR. AI-based imaging systems can detect visual patterns, which may aid screening and risk stratification, at a scale beyond what is currently possible with other technology.

Multimodal models are an additional step that takes into account clinical, laboratory, imaging and behavioural information [9]. The basic premise is that diabetes risk is not captured by a single factor, but by combination of several physiological and contextual factors.

The genetic and molecular information also has been examined. A recent review of genetic biomarkers and machine learning techniques highlighted that classical machine learning is still the current predominant approach, whereas deep learning and other more advanced techniques might open the door for the more complicated application of genetic data [10]. The review also noted that reporting on discrimination, calibration/classification measures is inconsistent, making it difficult to compare across studies.

This discovery is significant for generative AI as it can only give clinically accurate explanations if the predictive information is accurate. Poorly calibrated prediction is insufficient to make up for generative fluency.

2.3 Large Language Models in Diabetes Care

The large language models bring a new interaction paradigm. Clinicians can communicate with a language model using natural-language prompts, rather than having to access individual dashboards and structured data fields.

Summary, treatment plan interpretation, guideline retrieval, patient education, risk explanation, and medication plan suggestions are possible applications for diabetes [11]. A 2024 Study of the use of LLM's in Diabetes care noted that LLM's can support several aspects of diabetes care, including the lack of evidence in its use at this time.

Since then, there has been additional experimental evidence. The study used a retrospective study design, which included 85 adult patients with type 2 diabetes. However, the agreement was significantly different with different types of clinician prescriptions, depending on the complexity of treatment [12]. Among one-third of treatment-naive patients, agreement for monotherapy was 84.6% in one evaluation, and none of them agreed to triple therapy. For some treatment categories, few-shot prompting led to increased agreement.

The findings offer a glimpse into the capabilities and constraints of generative AI. For relatively standard decisions, the system can be used reasonably, while its reliability decreases as the number of medicines, the combination of medicines and the comorbidities and treatment history increase.

2.4 Generative AI and Clinical Decision Support

Creating plausible text isn't the only thing clinical decision support needs. A decision support system needs to bring in information that is accurate and clinically relevant, and that is properly contextualised and traceable to evidence [13].In a study that tested the performance of ChatGPT, Google and Llama 2 on 110 medical cases, GPT-4 performed better than GPT-3.5 in diagnosis and examination while Google outperformed in diagnosis; but these tasks were different for treatment. LLMs have also been found to have certain shortcomings in open-ended decision-making by wider clinical benchmarks. This is important because the clinical choices are often not definitive, but instead have multiple priorities and constraints which can vary from patient to patient. Current evidence indicates that generative AI should be considered more as assistant rather than a standalone clinician in the clinical decision support architecture.

2.5 Diabetes-Specific Generative Models

The generative systems for diabetes are important as they are not specifically designed to manage diabetes. A diabetes-related GPT developed using 65 information sources was evaluated with 420 diabetes-related questions (Ruma *et al*., 2026). The accuracy reported was 91.7%, with higher accuracy in general diabetes knowledge and nutrition than in questions related to insulin. The results show the significance of domain adaptation. With the development of knowledge retrieval, prompt engineering, controlled terminology and explicit source attribution, reliability can be enhanced for diabetes. But the accuracy in answering the questions does not demonstrate clinical effectiveness. Prospective evaluation of patient outcomes, clinician dependence, workflow and safety are needed for clinical deployment.

2.6 Generative AI and Personalized Treatment

Diabetes treatment personalisation is particularly challenging as it involves the glycaemic status, cardiovascular risks and renal function, weight considerations, medication effects, hypoglycaemia risk, cost, access and patient preferences [14].A 2025 paper looking at using GPT-4 to augment treatment plans designed by diabetologists noted that the model could successfully determine optimized treatment targets but performed less well at identifying the places for improvement. GPT-4's reasoning and explanations were found to be beneficial by an expert panel of academic diabetologists in about 40%. This evidence is to support the collaborative approach to the model, not the autonomous approach. The benefit of generative AI could be the most significant when it comes to collating evidence and highlighting review factors for the clinician.

3. RESEARCH GAP

Existing literature shows a divide between the predictive AI research and generative clinical AI research. Typically, predictive studies focus on discrimination, classification and model performance, while LLM studies focus on language generation, question answering, clinical reasoning and treatment recommendations. The gap in the literature is thus architectural/methodological. Minimal validation of diabetes prediction models with generative AI in one clinically constrained framework. The second gap has to do with evaluation. The accuracy, sensitivity, specificity, F1 score and AUC are some of the metrics we typically use to measure predictive models. In generative systems, there are extra criteria such as factuality, adherence to guidelines, correctness of citations, fullness and hallucination rate. A third gap relates to the evaluation along the way. Diabetes is a lifelong disease that needs to be monitored over and over. In a decision support system, the temporal context must be preserved, but cannot be passed on if it is outdated or inaccurate. A fourth gap is related to fairness. Diabetes prediction models may exhibit variability in performance among demographic, socioeconomic and health care access groups. Such a model can produce or accentuate differences in a generative system that is overlaid on it.

4. METHODOLOGY

4.1 Study Design

The current study is of structured evidence-synthesis and framework development design. It is not intended to depict a potential clinical trial nor be a new clinical prediction model. In contrast to this, the evidence published on AI-based diabetes prediction and the diabetes care facilitated by generative AI is summarised and an architecture built that has a clinical focus [15].The methodological approach is based on four evidence domains: diabetes epidemiology and clinical standards, traditional machine learning prediction process, generative-AI applications, and applications in clinical decision support. The 2026 American Diabetes Association Standards of Care were used as the primary current clinical reference document for diabetes diagnosis, management, and technology.

4.2 Evidence Identification

Searches were conducted on combinations of the terms: diabetes mellitus, type 1 diabetes, type 2 diabetes, artificial intelligence, machine learning, deep learning, generative artificial intelligence, large language models, ChatGPT, prediction and clinical decision support. Systematic reviews, peer reviewed clinical investigations, diabetes specific AI studies and new professional advice [16].The resulting evidence base was split into three groups: predictive AI, generative AI, diabetes technology and clinical implementation.

4.3 Analytical Framework

The proposed analytical framework has five functional stages: Data acquisition, predictive modelling, evidence retrieval, Generative reasoning and clinical oversight. The first data sources are Electronic Health Records, Laboratory System, Glucose Monitoring System, Imaging System, Wearable Devices and Patient

Reported Data. Data preprocessing includes dealing with missing, varying units, temporal alignment, duplication, and abnormal values. A validated predictive model, in turn, predicts diabetes risk or probability of complications. This prediction is used to query a retrieval layer which has access to existing clinical guidelines and validated institutional protocols [17].The structured information is then transformed into a clinical summary and decision-support output in the generative layer. The final product will include the predicted risk, evidence, uncertainty, and potential management considerations/information needing to be confirmed by the clinician. The final layer is human oversight. No treatment is automatically put in place without proper clinical review.

4.4 Performance Measures

Discrimination and calibration should be used to assess predictive performance. AUC reflects the discrimination; sensitivity and specificity reflect classification performance. The positive predictive value and negative predictive value provide other data about clinical screening.

For a binary classifier:

$$Sensitivity = \frac{TP}{TP + FN} Specificity$$
$$= \frac{TN}{TN + FP} Precision$$
$$= \frac{TP}{TP + FP} F1$$
$$= 2\left(\frac{Precision \times Recall}{Precision + Recall}\right)$$

Calibration should be evaluated in addition to discrimination for probabilistic models. High AUC, but poor calibration, can result in inappropriate risk estimates. Factual accuracy, adherence to guidelines, completeness, source-roundedness, and frequency and clinical acceptability of hallucinations should all be included as part of the performance in generative tasks [18].

4.5 Proposed Clinical Decision-Support Architecture

The proposed architecture is for prediction and generation to be separated.

The predictive engine is used for quantitative risk estimation. May be Logistic regression, Random forest, Gradient boosting or neural networks based on the data set. The evidence engine fetches clinical guidance that is relevant. The advantage of retrieval-augmented generation over unrestricted generation is that there is explicit evidence that comes from controlled sources that is fed into the model [19].The natural-language explanation is generated with the generative engine. It is guided to differentiate between accepted results, predicted models and recommendations based on guidelines, and uncertainty. The information is represented in the clinical interface for the clinician. The clinician is always the one who interprets and acts on the results.

5. RESULTS AND ANALYSIS

5.1 Evidence Synthesis

Current evidence shows that AI models for diabetes prediction are at a technically mature development phase, but are not fully validated for routine clinical use. The most compelling evidence is for traditional machine learning [20]. For diabetes complications, of the 32 studies identified in the systematic review, only 36% demonstrated clearly useful discrimination and most studies were at high risk of bias in the systematic review. Evidence created with generative AI is relatively new and comes in a variety of forms. The accuracy of the diabetes-specific GPT evaluation was 91.7% across 420 questions, while treatment-selection studies showed increasing level of agreement with increasing clinical complexity.

Table 01: Quantitative Evidence Base for AI in Diabetes

Evidence domain	Numerical evidence	Interpretation
Global people living with diabetes	830 million	Substantial population-level need
Adult diabetes prevalence, 1990	7%	Historical baseline
Adult diabetes prevalence, 2022	14%	Approximately doubled
ML complication-prediction studies	32	Broad research base
ML models evaluated	87	Multiple algorithmic approaches
Models with clearly useful discrimination	36%	Limited proportion
Neural-network models in reviewed complication studies	15	Frequently used approach
Patients in ChatGPT medication study	85	Early clinical evidence
ChatGPT monotherapy agreement, selected assessment	84.6%	Higher performance in simpler decisions
ChatGPT triple-therapy agreement	0%	Poor performance in complex regimen
Diabetes GPT evaluation questions	420	Domain-specific evaluation
Diabetes GPT reported accuracy	91.7%	Strong question-answering

		performance

Sources: WHO; systematic reviews and clinical studies cited above.

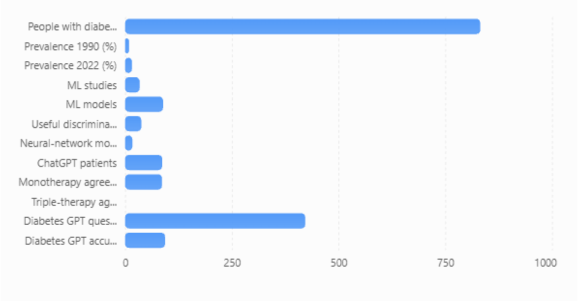

Figure 02: Quantitative Evidence Base for AI in Diabetes

5.2 Comparative Performance of AI Approaches

Table 02: Functional Comparison of AI Approaches

AI approach	Primary output	Principal strength	Principal limitation	Clinical role
Logistic regression	Risk probability	Interpretability	Limited nonlinear representation	Baseline prediction
Decision tree	Classification/rules	Explainability	Instability	Structured screening
Random forest	Risk/classification	Ensemble robustness	Lower transparency	Risk prediction
Gradient boosting	Risk score	Nonlinear interactions	Calibration requirements	High-performance prediction
Deep neural network	Probability/classification	Complex feature representation	Data and interpretability requirements	Multimodal prediction
LLM	Text/explanation	Natural-language synthesis	Hallucination and reasoning limitations	Clinical communication
Retrieval-augmented LLM	Evidence-grounded text	Source integration	Retrieval dependency	Decision support
Integrated predictive-generative AI	Risk + explanation	Combines quantitative and linguistic capabilities	System complexity	Intelligent CDS

The complexity of the features that need to be extracted and utilized in making the predictions. The complexity of the features that are required to be extracted and used to make predictions' Text/Explanation Natural language synthesis Limitation in hallucination and reasoning Clinical communication (Baraminic *et al.*, 2026)Leveraging past experiences and decisions to guide future choices. Using past experiences and decisions to inform future ones. Integrated predictive-generative AI Risk + explanation Combination of quantitative and linguistic skills Complexity of the system Intelligent CDS. The comparison shows that there is no one model class that meets all the requirements of diabetes decision support [21]. The value of predictive algorithms is in providing numerical risk estimates, with limited natural-language explanation. LLMs offer explanation and synthesis but in need external grounding for high-stakes decisions.

The integrated architecture proposed thus has the predictive model and the generative model as complementary rather than interchangeable elements.

5.3 Complexity-Dependent Generative AI Performance

The key take away from the reviewed evidence is the connection between the clinical complexity of the task and generative-AI performance. For example, a quantitative illustration is the 85-patient Study of Medication Selection using ChatGPT. For treatment-naive people, the level of agreement for selected monotherapy was 84.6%, for dual therapy there was significantly less agreement and triple therapy 0%.

Table 03: Relationship between Treatment Complexity and Generative-AI Concordance

Clinical decision category	Reported agreement pattern	Analytical interpretation
Treatment-naive monotherapy	69.2–84.6% across assessments	Relatively strong
Treatment-naive dual therapy	0–20.8% across assessments	Substantially weaker
Treatment-naive triple therapy	0%	Inadequate for autonomous support
Previously treated patients	Lower than treatment-naive cases	Historical context increases complexity
Few-shot	Improved some	Prompt/context

prompting	categories	engineering influences performance

Some categories benefited from few-shot prompting. Others benefited from the improved versions of few-shot prompting Prompt/context engineering had an impact on performance for some categories [22]

The pattern sets out that there is no single global measure of the accuracy of generative-AI performance. Clinical complexity needs to be a clear and explicit assessment criterion.

5.4 Domain-Specific Knowledge Performance

The GPT study is an additional significant result, specifically for diabetes research. The system had a higher accuracy rate with 420 diabetes-related questions (91.7%) with better accuracy for general knowledge and nutrition but low accuracy for insulin-related questions.

Table 04: Domain-Specific Generative-AI Performance Pattern

Domain	Reported performance pattern	Clinical significance
General diabetes knowledge	Stronger	Suitable for information synthesis
Nutrition	Stronger	Potential patient-education application
Diabetes technology	Evaluated	Requires current evidence retrieval
Insulin-related questions	Relatively weaker	Higher safety requirement
Complex treatment decisions	Variable	Clinician verification required

The results have shown that clinical risk should be the basis for the level of automation. A different operational threshold may be suitable for low-risk educational outputs vs insulin dosing and modification, and emergency decision support (Matahina *et al.*, 2025).

5.5 Prediction-to-Decision Pipeline

The proposed framework is a structured decision-support process to predict diabetes.

Table 05: Proposed Intelligent Clinical Decision-Support Pipeline

Stage	Input	Computational process	Output
1	Patient information	Data normalization	Standardized record
2	Clinical variables	Predictive model	Diabetes risk probability
3	Risk estimate	Stratification	Low/intermediate/high risk
4	Patient context	Evidence retrieval	Relevant guidelines
5	Retrieved evidence + patient data	LLM synthesis	Clinical explanation
6	Model uncertainty	Uncertainty generation	Confidence statement
7	Clinical review	Human verification	Final decision
8	Outcome data	Feedback evaluation	Model monitoring

This pipeline helps stop the generative model from being a standalone medical expert. The model can not only take in structured predictions, but also retrieve evidence.

6. DISCUSSION

6.1 Integration of Predictive and Generative AI

The key takeaway from this analysis is that generative AI is more likely to be effective when adopted on top of existing predictive and knowledge retrieval capabilities, and not as a standalone diagnostic tool (Kucera *et al.*, 2023).Traditional prediction models attempt to answer questions like, "Is there a high risk for diabetes in any given individual?". How to communicate the result and what contextual factors should be taken into consideration can be answered by generative AI. This difference results in a two-level intelligence model. Quantitative intelligence assesses risk and generative intelligence is used to interpret and communicate that risk [23].The separation further enhances system governance. The output of the prediction model can be audited independently from the generative model, if the prediction model is independently validated. Errors can consequently be categorized as data errors, prediction errors, retrieval errors or generation errors.

6.2 Clinical Decision Support Rather Than Autonomous Decision-Making

There is no supporting evidence from the reviewed literature for autonomous generative-AI treatment decisions. The decrease in concordance with increasing complexity of treatment is of special note. It is important to note that a clinical decision support system should make recommendations in a structured way, not as a prescription. The output should include the clinical information used, evidence source, risk, uncertainty and factors that need to be assessed by the clinician based on the model [24].The 2026 ADA Standards emphasize the critical role of person-cantered and evidence-based diabetes care and the use of diabetes technologies. Therefore, AI should be integrated to bolster existing clinical governance not replace it.

6.3 Retrieval-Augmented Generation

One of the key technical solutions to minimize unsupported responses is retrieval-augmented generation. Rather than relying on the parameters that the LLM has been learned from, the system will look into the latest evidence in the accepted diabetes management guidelines and institutional protocols to suggest a diabetes management question [25].The evidence retrieved may include diagnostic criteria, treatment recommendations, medication safety information, monitoring requirements and technology guidance. The generation layer will then respond to the retrieved evidence with an answer. This architecture is especially significant due to the fact that diabetes guidelines are modified. The ADA Standards are revised on a yearly basis or as soon as new evidence or regulations require such changes. The biggest risk with the static model is that it can become stale since it is not updated.

6.4 Explainability

In an integrated AI system, there are two facets of explainability. The first relates to explainability with prediction. The system should be able to recognize factors that help determine the risk score. Model-relevant variables, for example, age, glycaemic measurements, body mass index and diabetes duration are presented where appropriate [26].The second relates to generative explainability. The LLM needs to provide the rationale for the selection of the clinical consideration and the evidence source for the clinical consideration in the output. It's important to note that explainability does not mean free-flowing chain-of-thought disclosure. The clinical systems need short, portable, and verifiable justifications not to the inescapable traces of internal reasoning. A suitable explanation should include the evidence, relevant clinical factors and uncertainty.

6.5 Calibration and Clinical Utility

It has been shown in the literature that AUC alone is not enough. In the case of genetic biomarkers, the systematic review focused specifically on discrimination, calibration and classification measures, and in the case of machine learning, it highlighted the need for such measures [27]. A clinically useful diabetes prediction system should therefore be accompanied by the report of calibration plots, calibration slope, calibration intercept and an appropriate decision-curve analysis in addition to discrimination measures. An overconfident prediction can be worse than an uncertain prediction, when deciding on a high-risk clinical situation. Therefore, the probabilities provided to the generative system should be calibrated and not raw classifier scores.

6.6 Fairness and Generalizability

One of the key challenges to making clinical AI a reality is its generalizability. The impact of a model in one health system may differ from another due to differences in population characteristics, access to health care services, measurement of services, and disease prevalence. Fairness should thus be considered in the development of models. The sensitivity, specificity and calibration and false negative rate should be compared among relevant demographic and clinical groups and differences should be explored [28].It becomes more relevant with the inclusion of generative AI as the language layer can magnify system errors. A model that generates less intense clinical recommendations for a specific population will then generate less intense risk predictions for that population.

6.7 Data Privacy and Security

The diabetes AI systems need to be provided with sensitive clinical data. Any data that includes electronic health records, continuous glucose-monitoring data, lab data, and medication information must be handled properly with security measures [29].

Key considerations include data minimization, encryption, access control, audit logging, and secure model deployment. When external generative models are implemented, there needs to be a clear policy on data transmission, retention and secondary use in healthcare organisations.

6.8 Human Factors

The interaction design is as important as the model's performance for clinical usefulness. If clinicians receive too many alerts from a highly accurate model, or the information they receive from the model is hard to interpret, the model can fail. The interface should thus emphasize brief summaries of risk, references to evidence, and markers of uncertainty along with explanations that are clinically relevant [30].Another issue to consider is automation bias. Clinicians might take the AI recommendations at face value, especially if they are grammatically correct and seem authoritative. The clinician might take AI recommendations for granted, especially if they are grammatically correct and appear authoritative. The interface should clearly demarcate information generated by the model from information verified by the clinician.

6.9 Role in Primary Care

Primary care is a potentially important deployment environment, as primary care physicians are often faced with patients who have multiple risk factors and limited information.AI can provide a summary of a history of records and flag missing investigations. Using predictive AI, risk for diabetes can be identified. The patient context can be linked to existing guidelines by retrieval mechanisms [31]This can lead to a reduced burden of information processing without shifting responsibility for clinical accountability to the model.

6.10 Role in Precision Diabetes Care

Adopting a precision medicine approach to interpretation of clinical and biological information is individualized. Diabetes is an ideal disease for such strategies because the metabolic profile of patient phenotypes differs significantly, as do the risk and progression of diabetes, and the response to therapy and complications [32].A multimodal AI system might integrate clinical history, glucose profiles, lab data, imaging, medication and behavioural data.

The multimodal representation can then be turned into a clinically interpretable patient profile by generative AI. This leads to a new framework that shifts from population level risk prediction to personalized decision support.

7. Proposed Generative-AI Framework for Diabetes Prediction

7.1 Data Layer

The data layer integrates electronic health records, laboratory systems, continuous glucose monitoring, wearable devices, retinal imaging, medication histories and patient-generated information. Data preprocessing should normalize the units, handle missing data, detect duplicate data and ensure the temporal ordering.

7.2 Predictive Layer

The predictive layer is a prediction of risk for diabetes or risk of complications.

The type of algorithm should be determined by the type of prediction problem. Logistic regression gives an easily understood baseline [33]. Random forest and gradient boosting are able to capture the relationship between two variables when it is nonlinear. High-dimensional multimodal data could be suitable for deep-learning methods. The prediction engine should be validated internally and externally before clinically implementing.

7.3 Knowledge Layer

Current diabetes guidelines, institutional protocols and validated evidence are stored in the knowledge layer. An authoritative clinical resource is needed for the knowledge layer, and the 2026 ADA Standards reflect a type of authoritative clinical resource.

7.4 Generative Layer

The generative model takes four main inputs: patient context, predictive output, retrieved evidence and clinical task. It then generates a structured response with the clinical summary, interpretation of the risk, evidence-based considerations and uncertainty [34].The model should never be allowed to create patient data that isn't available.

7.5 Safety Layer

The safety layer is the layer that performs the validation prior to the response that reaches the clinician. It verifies that the proposed recommendation is supported by evidence retrieved, that contraindications are identified and that uncertainty is expressed appropriately. High risk recommendations should be the mandatory clinician review.

7.6 Feedback Layer

The last layer is for clinician corrections and clinical outcomes. Feedback can be used to monitor the model and for performance reassessment. Should not automatically update a clinical model with no controlled validation.

8. EVALUATION FRAMEWORK

Table 06: Proposed Evaluation Metrics

Evaluation dimension	Metric	Target purpose
Discrimination	AUC	Risk-ranking ability
Sensitivity	%	Case identification
Specificity	%	False-positive control
Calibration	Calibration slope/intercept	Probability reliability
Clinical utility	Decision-curve analysis	Net clinical benefit
Generative factuality	% supported statements	Hallucination control
Guideline adherence	% compliant outputs	Evidence conformity
Citation correctness	% verified sources	Traceability
Completeness	Expert score	Clinical information coverage
Safety	Unsafe-output rate	Risk monitoring
Fairness	Group-specific metrics	Equity assessment
Workflow	Time per case	Operational efficiency
Human factors	Clinician acceptance	Usability
Longitudinal stability	Performance drift	Deployment monitoring

Decision-curve analysis and Net clinical benefit represent clinical utility measures. Clinical There is a broad range of variability in the level of performance drift between different products. The performance drift varies significantly between different products. This assessment framework goes beyond traditional machine learning assessment frameworks because intelligent clinical decision making is a socio-technical system.

9. CLINICAL IMPLEMENTATION CONSIDERATIONS

The implementation should be done in a phased manner and not directly into clinical practice.

The first step should be to look at the past and evaluate it. The second phase should be a silent prospective deployment, in which the AI system provides outputs without impacting the decision making process [35]. The third stage should evaluate clinician interaction and workflow. The fourth stage can evaluate clinical outcomes in the context of controlled implementation. The staged model enables technical errors to be caught before any consequences for the patient. Generative AI should also be linked with the latest evidence. The development of diabetes technology provides a good example of why it is not enough to rely on static knowledge systems. Insulin pumps, insulin pens, continuous glucose monitoring, automated insulin delivery, and connected insulin pens and digital diabetes support systems are part of the technology standards that the ADA will support and recommend for 2026. An intelligent system needs to be maintained throughout by the addition of knowledge.

10. LIMITATIONS

The major drawback of this paper is that it is a proposed structured synthesis and a framework for the future rather than a future clinical trial (Ghobrial *et al*., 2025). The numbers cited in the results section should only be interpreted as performance based on published studies and are not to be interpreted as performance that is obtained with a newly developed model. The second constraint is the fact that there is heterogeneity of the existing studies. There are significant variations in the data used, outcome definition, population, data preprocessing, and evaluation metrics used in diabetes prediction studies. Direct comparison is not, therefore, possible. The third constraint on the use of generative AI is the dynamic nature of the field. Model versions, prompting strategies, retrieval mechanisms, and clinical interfaces can have a significant impact on performance. The fourth limitation relates to the difference between accuracy in the answering of questions and the clinical use [36]. A model can have a high accuracy answering a question but be less successful in making more complex decisions for a patient. A fifth limitation is that much of the evidence base is relatively early stage, and is not as extensive as the evidence base for traditional clinical decision-support systems. Requirements for external validation and prospective clinical evaluation are still critical. External validation was clearly stated as a requirement for clinical implementation in the diabetes systematic review of diabetes complication prediction.

11. FUTURE RESEARCH DIRECTIONS

Future research needs to focus on multimodal diabetes FDMs, which can incorporate structured clinical information, glucose monitoring data, imaging and longitudinal information. Prospective evaluation should assess model accuracy as well as diagnostic delay, treatment appropriateness, clinician workload, patient outcomes and utilization of healthcare. Standardized measurements for generative AI in diabetes should also be established for research [37]. These should encompass regular cases, rare cases, multimorbidity, drug interactions, renal impairment, pregnancy, insulin use and emergency situations. Uncertainty-aware generation is another topic of priority. This should differentiate between evidence that is very reliable, model-based predictions and situations that need further information. Another research trajectory is federated learning, which can support the development of the model without having to centralize the sensitive patient data. The emergence of explainable multimodal architectures is also needed. A clinician should be able to determine the source of an output, either from an actual laboratory value, glucose patterns, imaging data, medication history, or retrieved clinical guidance.

12. CONCLUSION

While generative artificial intelligence offers a powerful new computational tool for predicting diabetes mellitus and for intelligent clinical decision support, its use is fundamentally distinct from a traditional predictive model. Predictive machine learning provides quantitative risk estimation, while generative AI provides contextual synthesis, explanation, evidence integration and natural-language communication. Current evidence shows although it is possible to develop machine learning models that deliver useful prediction of diabetes and diabetes complications, there are limitations such as external validation, calibration and bias. Of the 32 studies and 87 models analysed, 36% of the models showed clear usefulness in discrimination, with random forests showing the best overall performance for the outcomes analysed. There are limited examples of generative-AI evidence that shows promise for diabetes knowledge and some clinical tasks, along with clear limitations. 1) The accuracy of diabetes-specific GPT evaluation in 420 questions was reported as 91.7%, and for each of the treatment-selection research, the accuracy rate decreased significantly with the increase in treatment complexity. The results suggest that overall language proficiency should not be used as a measure of clinical reliability. The proposed architecture thus places generative AI on top of the validated prediction models and authoritative knowledge retrieval. The predictive layer is responsible for risk estimation, the knowledge layer is responsible for the current clinical evidence, the generative layer is responsible for combining the information, and the clinician gives the final interpretation and decision authority. It is therefore not an autonomous generative physician that is the most suitable model for clinical deployment, but a predictive-generative clinical decision-support system. Such a system can be integrated with quantitative risk estimation, evidence-based clinical synthesis, while keeping the human touch. There is a need for further external validation, calibration, multimodal integration, grounding of guidelines, communication of uncertainty, and prospective clinical evaluation in the future development of AI for diabetes care, as well as for fairness and privacy. This kind of development is clinically relevant and can have a huge impact on the future of diabetes technology, which is rapidly evolving as well as the prevalence of diabetes in the world. WHO statistics show that by 2022 there were about 830 million individuals with diabetes. The prevalence of diabetes – with over 830 million people affected in 2022 – calls for better detection and management. Generative AI can play a role in achieving this goal as a managed component in an evidence-based clinical architecture. It can provide most benefits when it is used to complement validated predication, to structure complex clinical information and enhance communication, not in lieu of diagnostic testing or professional clinical judgment.

13. FUNDING

Nil

14. CONFLICT OF INTEREST

Not Declared

15. AUTHOR CONTRIBUTION

Both are contributed equally

16. ACKNOWLEDGEMENT

Not Declared

17. INFORM CONSENT

Not Applicable

18. ETHICAL STATEMENT

Not Applicable.

19. REFERENCES

1. Deng, M., Yang, R., Zheng, X., Deng, Y. and Jiang, J., 2025. Artificial intelligence in diabetes care: from predictive analytics to generative AI and implementation challenges. *Frontiers in Endocrinology*, *16*, p.1620132.
2. Nimri, R. and Phillip, M., 2025. Enhancing care in type 1 diabetes with artificial intelligence driven clinical decision support systems. *Hormone Research in Paediatrics*, *98*(4), pp.384-395.
3. Osenga, A., Olawale, D.B., Gore, M., Rotifer, O.B., Egbon, E., Chinwag, V. and Bossio, S., 2025. Generative artificial intelligence in predictive analysis of diabetes and its complications: a narrative review. *Annals of translational medicine*, *13*(5), p.59.
4. Yang, H., 2026, June. Application of Artificial Intelligence and Large Language Models in Clinical Decision-Making for Diabetes. In *Exploring Science Academic Conference Series* (Vol. 14, pp. 211-219).
5. Yang, H., 2026, June. Application of Artificial Intelligence and Large Language Models in Clinical Decision-Making for Diabetes. In *Exploring Science Academic Conference Series* (Vol. 14, pp. 211-219).
6. Göktaş, M. and Bilgehan, T., 2025. A new era in diabetes management: generative artificial intelligence. *Artificial Intelligence Theory and Applications*, *5*(1), pp.63-81.
7. Henley, S., 2023. Generative Artificial Intelligence-based clinical decision support in screening, prevention, and treatment choices in medical care. *Contemporary Readings in Law and Social Justice*, *15*(1), pp.27-44.
8. Anand, A., Kumar, R., Singh, N., Dutta, B. and Rajput, O., 2025. Exploring generative AI and IoMT in diabetes prediction. *Generative Artificial Intelligence in Healthcare*, pp.132-150.
9. Rajashekar, N., 2025. *Generative Artificial Intelligence in Clinical Decision Support-Quantitative and Qualitative Analyses* (Doctoral dissertation, Yale University).
10. Dey, A.K., 2023. ChatGPT in diabetes care: an overview of the evolution and potential of generative artificial intelligence model like ChatGPT in augmenting clinical and patient outcomes in the management of diabetes. *International journal of diabetes and technology*, *2*(2), pp.66-72.
11. Ghorbian, M. and Ghorbian, S., 2025. Empowering Clinical Decision-Making with Generative AI in Intelligent Decision Support Systems. In *Generative Artificial Intelligence (AI) Approaches for Industrial Applications* (pp. 241-259). Cham: Springer Nature Switzerland.
12. Vehi, J., Mujahid, O., Beneyto, A. and Contreras, I., 2025. Generative artificial intelligence in diabetes healthcare. *Iscience*, *28*(8).
13. Hu, J., Ren, L., Wang, T. and Yao, P., 2025. Artificial intelligence-assisted clinical decision-making: a perspective on advancing personalized precision medicine for elderly diabetes patients. *Journal of Multidisciplinary Healthcare*, pp.4643-4651.
14. Ruma, U. and Masood, W., 2026. AI-driven treatment response prediction for chronic disease management using generative artificial intelligence and large language models. *International Journal of Recent Trends in Multidisciplinary Research*, *6*(2), pp.58-65.
15. Kucera, J., Sedlackova, A.N. and Negoianu, A.E., 2023. Generative Artificial Intelligence-based Clinical Decision Support in Patient Data Collection and Analysis, in Physiological Parameter Monitoring, and in Image-based Disease Diagnosis. *Contemporary Readings in Law and Social Justice*, *15*(1), pp.116-133.
16. Grupac, M., Zauskova, A. and Nica, E., 2023. Generative artificial intelligence-based treatment planning in clinical decision-making, in precision medicine, and in personalized healthcare. *Contemporary Readings in Law and Social Justice*, *15*(1), pp.45-62.
17. Misir, A., 2025. Artificial intelligence and machine learning in diabetic foot ulcer care: advances in diagnosis, treatment, prognosis, and novel therapeutic strategies. *Journal of Diabetes Science and Technology*, p.19322968251363632.

18. Rabbani, S.A., El-Tanani, M., Sharma, S., Rabbani, S.S., El-Tanani, Y., Kumar, R. and Saini, M., 2025. Generative artificial intelligence in healthcare: applications, implementation challenges, and future directions. *BioMedInformatics*, *5*(3), p.37.
19. Fadul, N., Alaskar, M.F., Jillahi, K.B. and El-Khaled, D.B., 2025. Generative AI in Healthcare: An Analytical Review of Models, Clinical Applications, and Decision-Support Implications. *Journal of Future Artificial Intelligence and Technologies*, *2*(4), pp.587-615.
20. Dubský, M., Liegertová, M., Bém, R., Sojáková, D., Fejfarová, V. and Jude, E.B., 2026. Artificial intelligence in diabetes care. *Diabetes, Obesity and Metabolism*, *28*(6), pp.4587-4601.
21. de Melo, P. and Rose, M.S., 2025. Accurate classification of diabetes via PM Generative AI. *Advances in Bioscience and Biotechnology*, *16*(9), pp.379-409.
22. Bajramagic, M., Battelino, T., Cos, X., Cote, M., Cui, N., Forbes, A., Galderisi, A., Heinemann, L., Hussain, S., Imbert, J. and Holm Jönsson, C., 2026. Artificial intelligence-driven clinical decision support systems to assist healthcare professionals and people with diabetes in Europe at the point of care: a Delphi-based consensus roadmap. *Diabetologia*, *69*(2), pp.259-273.
23. Kovacova, M., Kevicky, F. and Popescu, G.H., 2023. Generative artificial intelligence-driven healthcare systems in patient record analysis, in disease diagnosis and monitoring, and in customized treatment plans. *Contemporary Readings in Law and Social Justice*, *15*(1), pp.152-170.
24. Ahmed, Y.K. and Naji, A.N., 2025. Smart feature extraction using deep learning for early diagnosis of chronic diseases in next-generation medical decision support systems. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *14*(1), p.140.
25. Mashatian, S., Armstrong, D.G., Ritter, A., Robbins, J., Aziz, S., Alenabi, I., Huo, M., Anand, T. and Tavakolian, K., 2025. Building trustworthy generative artificial intelligence for diabetes care and limb preservation: a medical knowledge extraction case. *Journal of diabetes science and technology*, *19*(5), pp.1264-1270.
26. Deng, M., Yang, R., Zheng, X., Deng, Y. and Jiang, J., 2025. Artificial intelligence in diabetes care: from predictive analytics to generative AI and implementation challenges. *Frontiers in Endocrinology*, *16*, p.1620132.
27. Nimri, R. and Phillip, M., 2025. Enhancing care in type 1 diabetes with artificial intelligence driven clinical decision support systems. *Hormone Research in Paediatrics*, *98*(4), pp.384-395.
28. Osonuga, A., Olawade, D.B., Gore, M., Rotifa, O.B., Egbon, E., Chinwah, V. and Boussios, S., 2025. Generative artificial intelligence in predictive analysis of diabetes and its complications: a narrative review. *Annals of translational medicine*, *13*(5), p.59.
29. Yang, H., 2026, June. Application of Artificial Intelligence and Large Language Models in Clinical Decision-Making for Diabetes. In *Exploring Science Academic Conference Series* (Vol. 14, pp. 211-219).
30. Henley, S., 2023. Generative Artificial Intelligence-based clinical decision support in screening, prevention, and treatment choices in medical care. *Contemporary Readings in Law and Social Justice*, *15*(1), pp.27-44.
31. Anand, A., Kumar, R., Singh, N., Dutta, B. and Rajput, O., 2025. Exploring generative AI and IoMT in diabetes prediction. *Generative Artificial Intelligence in Healthcare*, pp.132-150.
32. Rajashekar, N., 2025. *Generative Artificial Intelligence in Clinical Decision Support-Quantitative and Qualitative Analyses* (Doctoral dissertation, Yale University).
33. Dey, A.K., 2023. ChatGPT in diabetes care: an overview of the evolution and potential of generative artificial intelligence model like ChatGPT in augmenting clinical and patient outcomes in the management of diabetes. *International journal of diabetes and technology*, *2*(2), pp.66-72.
34. Ghorbian, M. and Ghorbian, S., 2025. Empowering Clinical Decision-Making with Generative AI in Intelligent Decision Support Systems. In *Generative Artificial Intelligence (AI) Approaches for Industrial Applications* (pp. 241-259). Cham: Springer Nature Switzerland.
35. Vehi, J., Mujahid, O., Beneyto, A. and Contreras, I., 2025. Generative artificial intelligence in diabetes healthcare. *Iscience*, *28*(8).
36. Hu, J., Ren, L., Wang, T. and Yao, P., 2025. Artificial intelligence-assisted clinical decision-making: a perspective on advancing personalized precision medicine for elderly diabetes patients. *Journal of Multidisciplinary Healthcare*, pp.4643-4651.
37. Ruma, U. and Masood, W., 2026. AI-driven treatment response prediction for chronic disease management using generative artificial intelligence and large language models. *International Journal of Recent Trends in Multidisciplinary Research*, *6*(2), pp.58-65.