New AI for Diabetes Control and Management, Part I: A Survey and Perspective on LLM-Based Interpretation of Stress, Exercise, and Glucose Dynamics

Adina Chotbaeva, Mengmeng Zhang, Máté Siket, Jing Wang, Levente Kovács, György Eigner, Xuanting Ding, and Fei-Yue Wang

***Abstract*—Large language models (LLMs) offer new opportunities in diabetes research by integrating contextual behavioral data with physiological biosignals, supporting more personalized care. This review surveys recent studies on the application of LLMs to interpret interactions among stress, physical activity, and glucose dynamics across type 1, type 2, gestational, and monogenic diabetes (MODY). We examine methods for capturing contextual data beyond traditional patient diaries, including wearable sensors, lifestyle logs, and digital health tools, and discuss how these data are combined with continuous glucose monitoring. After retrieving 39 relevant studies and finally retaining 25 after screening, we summarize the current capabilities, limitations, and clinical implications of LLM-assisted multimodal approaches in diabetes management. The findings highlight both the promise and challenges of applying LLMs to synthesize heterogeneous data, providing insights for future research on enhancing individualized and evidence-based diabetes care.**

***Index Terms*—Large language models (LLMs), diabetes mellitus, multimodal data integration, digital health**

Manuscript received: 28 July 2025; revised: 24 August 2025; accepted: 9 September 2025. *(Corresponding author: Fei-Yue Wang.)*

Citation: A. Chotbaeva, M. Zhang, M. Siket, J. Wang, L. Kovács, G. Eigner, X. Ding, and F.-Y. Wang, New AI for diabetes control and management, part I: A survey and perspective on LLM-based interpretation of stress, exercise, and glucose dynamics, *Int. J. Intell. Control Syst.*, 2025, 30(3), 207–217.

Adina Chotbaeva is with Physiological Controls Research Center, Óbuda University, Budapest H-1034, Hungary, and also with Applied Informatics and Applied Mathematics Doctoral School, Óbuda University, Budapest H-1034, Hungary (e-mail: adina.chotbaeva@uni-obuda.hu).

Mengmeng Zhang is with State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhangmengmeng2022@ia.ac.cn).

Máté Siket and György Eigner are with Physiological Controls Research Center, Óbuda University, Budapest H-1034, Hungary, and also with John von Neumann Faculty of Informatics, Óbuda University, Budapest H-1034, Hungary (e-mail: siket.mate@uni-obuda.hu; eigner.gyorgy@uni-obuda.hu).

Jing Wang is with State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China (e-mail: wangjing2014@ia.ac.cn).

Levente Kovács is with Physiological Controls Research Center, Óbuda University, Budapest H-1034, Hungary, with Applied Informatics and Applied Mathematics Doctoral School, Óbuda University, Budapest H-1034, Hungary, and also with John von Neumann Faculty of Informatics, Óbuda University, Budapest H-1034, Hungary (e-mail: kovacs@uni-obuda.hu).

Xuanting Ding is with School of Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: dingxt24@mails.tsinghua.edu.cn).

Fei-Yue Wang is with School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China, and also with DeSci Center of Parallel Intelligence, Óbuda University, Budapest H-1034, Hungary (e-mail: feiyue.wang@ia.ac.cn).

Digital Object Identifier 10.62678/IJICS202509.10254

I. Introduction

DIABETES mellitus (DM) is a chronic metabolic disorder characterized by elevated blood glucose levels [1]. It affects millions of people worldwide and imposes a substantial burden on healthcare systems. The disease is associated with severe complications, including renal failure, lower limb amputations, and vision loss [2]. The prevalence of type 1 diabetes (T1D), type 2 diabetes (T2D), gestational diabetes mellitus (GDM), and monogenic forms continues to rise, underscoring the urgent need for effective prevention and management strategies [3]. Among these, type 2 diabetes accounts for approximately 90% of all cases [4]. Although monogenic diabetes is relatively rare (1%–2% of all cases), it is frequently underdiagnosed or misdiagnosed, emphasizing the importance of accurate identification for optimal treatment [5].

The escalating global prevalence of diabetes necessitates innovative solutions for its prevention and treatment. Digital health technologies, particularly large language models (LLMs), have emerged as promising tools in this regard. Sheng et al. [6] reviewed how LLMs, built on advanced natural language processing (NLP) techniques, assist both patients and healthcare professionals across diverse aspects of diabetes care, from risk prediction and personalized coaching to optimizing clinical workflows.

Despite advances in continuous glucose monitoring (CGM) and insulin delivery systems, major challenges persist in real-world contexts. One critical issue is data fragmentation: numerical biosignals from CGM and insulin pumps are often separated from qualitative, unstructured data such as patient diaries, wearable sensor outputs, and lifestyle logs (e.g., diet, exercise, stress, and sleep) [7]. Integrating these heterogeneous data streams is essential for accurately assessing a patient's metabolic state. Moreover, contextual factors such as stress, illness, or disrupted sleep strongly influence glycemic dynamics, yet remain difficult to quantify and incorporate into predictive models [8]. Current management strategies often lack the granularity required for personalized guidance, placing a high cognitive burden on patients (who must interpret complex data themselves) and on clinicians (who face fragmented records during brief consultations) [5].

Existing tools, such as CGM visualization platforms, rarely integrate contextual information from unstructured sources like patient diaries or social media. This omission is critical, as such data often contain valuable cues for personalizing therapy, including emotional states, stressors, subjective experiences of food, activity, and sleep. Patient-reported outcomes are especially important: perceived glucose variability and hyperglycemia are stronger predictors of daily diabetes-related distress than objective CGM metrics [9]. While initially recognized as a limitation in T1D management, similar issues of self-management, personalization, and monitoring limitations are common across all diabetes types.

These challenges highlight the need for a new paradigm in diabetes management. LLMs provide a promising solution by offering a unified framework for processing multimodal data. Their capacity to interpret unstructured text from patient diaries and link it with structured biosignals enables more holistic and personalized care. By bridging clinical metrics with the complex realities of daily life, large language models help to reduce cognitive burden and enhance decision support. Advanced models such as GPT-4 and Claude 3.5 Sonnet demonstrate sophisticated reasoning, contextual understanding, and question-answering abilities, making them promising candidates for patient counseling and clinical applications [10, 11].

As illustrated in Fig. 1, the gap between structured physiological signals and unstructured contextual narratives remains a major obstacle to personalized care. Large language models act as a transformative bridge by combining diary entries, social media reflections, and behavioral journals with CGM and electronie health record (EHR) data to generate holistic patient profiles.

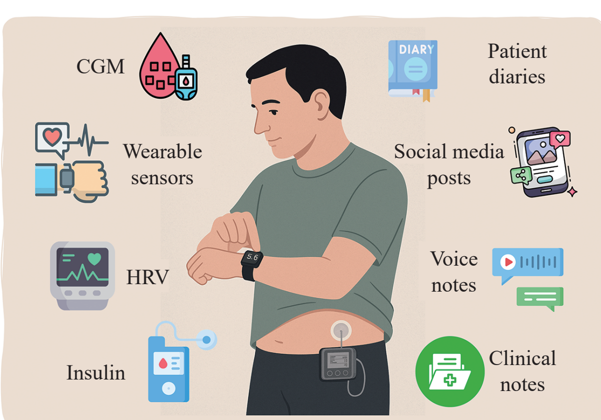


Fig. 1 Structures and unstructured data of diabates.

In parallel, specialized medical LLMs such as Meditron, BioMistral, and DeepDR-LLM are being developed by enriching general models with domain-specific medical knowledge [12–14]. These efforts expand the applicability of large language models beyond generic NLP tasks to high-stakes healthcare domains.

This review aims to systematically examine how LLMs, when integrated with advanced monitoring technologies, help decipher the interplay between stress, physical activity, and glucose dynamics across different forms of diabetes. By extending the scope beyond traditional AI-based glucose prediction or broad digital health studies, this work addresses multimodal AI applications in T1D, T2D, GDM, and monogenic diabetes, thereby contributing to the advancement of personalized and evidence-based care.

II. Background

Effective diabetes management relies on accurate and timely monitoring of key physiological parameters (see Table 1 for an overview of non-invasive glucose and stress monitoring technologies). This section provides a comprehensive overview of the evolution of glucose and stress monitoring technologies, highlighting their underlying principles, advantages, and limitations.

A. Traditional Monitoring Methods

Traditional self-monitoring of blood glucose (SMBG) through finger pricks is invasive, often painful, and provides only sporadic snapshots of glucose levels. This approach frequently overlooks critical variability and trends in glycemia [25], which can lead to poor adherence and suboptimal disease management [26]. Although patient diaries contain rich contextual information, such as emotional states, stressors, and subjective experiences related to eating, they are unstructured and difficult to analyze computationally, limiting their utility in personalized therapy [11]. Moreover, subjective perceptions recorded in diaries, including perceived glucose variability, have been shown to be stronger predictors of daily diabetes-related distress than objective CGM metrics [9].

B. Continuous Glucose Monitoring

Continuous glucose monitoring systems have transformed diabetes management by providing real-time data, offering unprecedented insight into glycemic trends and variability [25]. CGM is increasingly considered the standard of care for patients requiring insulin therapy [27]. Nevertheless, current CGM visualization tools often can not fully integrate contextual information from unstructured sources such as diaries or social media, limiting their potential for truly personalized therapy [9]. Contextual factors, including emotional stress, lifestyle changes, or sleep patterns, profoundly influence glycemia [28], yet remain difficult to quantify. Using only objective CGM metrics makes it difficult to identify the underlying causes of glycemic fluctuations. As a result, it becomes harder to personalize management and evaluate how individuals respond to interventions. Additionally, the complexity of CGM data can overwhelm both patients and clinicians, creating challenges in

Table 1 Overview of non-invasive glucose and stress monitoring technologies.

Technology/method	Principle of operation	Key advantages	Key limitations/challenges	Illustrative studies/devices
Sweat biosensors	Measurement of glucose in sweat using electrochemical or optical sensors	Non-invasive, continuous monitoring, potential for wearable devices	Low glucose concentration, influence of sweat rate, contamination, need for calibration	Wearable microfluidic devices [15], graphene-based biosensors [16]
Tear biosensors	Measuring glucose in tears using contact lenses or paper sensors	Non-invasive, painless, continuous monitoring	Impact of tear collection methods, need to increase sensitivity, correlation with blood glucose levels	Smart contact lenses [17], paper colorimetric biosensors [18]
Saliva biosensors	Measurement of glucose in saliva using electrochemical or optical sensors	Non-invasive, easy sample collection	Low glucose concentration, influence of salivation stimulation, limited testing conditions	Non-enzymatic electrochemical sensors [19], smart toothbrushes/mouth guards [20]
Heart rate variability	Analysis of time intervals between successive heartbeats to assess autonomic nervous system activity	Non-invasive, reflects physiological stress and metabolic changes, potential for predicting hypoglycemia	Sensitivity to noise, requires complex analysis, not always directly correlated with glucose levels	Empatica E4 [21], Fitbit Sense [22], Apple Watch [23]
Accelerometer data	Monitoring physical activity and detecting abnormal movements (e.g., tremors during hypoglycemia)	Non-invasive, continuous collection of movement data, potential for insulin adjustment and stress detection	Need for data contextualization, differences between aerobic and anaerobic activity	Smartwatches [24], fitness trackers [24]

interpretation and treatment adjustment [29]. For example, mainstream CGM platforms such as the Dexcom G7 [30, 31] or Abbott FreeStyle Libre [32, 33] provide detailed glucose traces and trend alerts, yet still face limitations in contextual integration, underscoring the need for more comprehensive data fusion approaches.

C. *Wearable Physiological Sensors*

The shift from invasive, periodic measurements to continuous, and non-invasive monitoring through wearable devices marks a fundamental transition from reactive to proactive diabetes management. Devices such as smartwatches and fitness trackers (e.g., Fitbit Sense, Apple Watch, and Empatica E4) can continuously record vital signs and physiological parameters, including heart rate variability (HRV), physical activity, and sleep patterns. These wearables provide convenient, non-invasive, and cost-effective alternatives or supplements to traditional monitoring [25].

They enable real-time tracking of physical activity, nutrition, and medication adherence, supporting personalized recommendations [34]. AI-based systems can further integrate these data streams with glucose measurements to offer tailored guidance on diet, medication, and exercise [35, 36]. HRV analysis derived from wearables can be used to predict type 2 diabetes and detect hypoglycemia, as well as to assess diabetic autonomic neuropathy [35, 36]. Accelerometers embedded in wearables can detect involuntary tremors associated with hypoglycemia and continuously monitor activity patterns [25], informing dynamic adjustment of insulin delivery algorithms [37]. When combined with the analytical capabilities of LLMs, these physiological signals provide a more comprehensive view of patient health [37].

D. *Non-Invasive Blood Glucose Monitoring (NIBGM)*

Non-invasive blood glucose monitoring using biological fluids such as sweat, tears, and saliva is an active research area aiming at minimizing discomfort and reducing infection risk [34, 38]. Wearable microfluidic devices and biosensors are being developed to detect glucose in sweat, although low glucose concentrations and potential contamination remain significant challenges [25, 34].

Smart contact lenses and paper-based colorimetric biosensors for tear glucose monitoring offer non-invasive and painless diagnostics. Challenges include the influence of tear collection methods, the need for enhanced sensitivity, and the establishment of reliable correlations with blood glucose levels [25, 34, 39, 40]. Saliva-based glucose detection is also under investigation using non-enzymatic electrochemical sensors, optical sensors, and smart toothbrushes/mouthguards, yet these face similar issues with low concentrations and limited testing conditions [26].

Overall, non-invasive biosensors for biological fluids hold significant promise in terms of comfort and convenience, but they collectively face challenges related to sensitivity, interference from other substances, and contamination [25].

This section outlines the methodological framework of our literature review. Fig. 2 illustrates the overall workflow of data integration and personalized insight generation in LLM-based diabetes management, where structured physiological signals (e.g., CGM and HRV) and unstructured contextual data (e.g., diaries, social media posts, and clinical notes) are fused through LLM engines. This conceptual framework provides the foundation for the literature review methodology described in the following sections.

III. Methodology of Literature Review

A. *Search Strategy*

We conducted a structured literature search across Scopus, PubMed, and Web of Science to identify studies published up to August 2025. As illustrated in Fig. 3, search queries combined terms relate to large language models and diabetes (e.g., “LLM” or “large language model” and “diabetes”). This search yields 31 records. An additional 8 studies are identified through manual screening of references and domain-focused hand searching, resulting in a total of 39 candidate records for full-text assessment.

B. *Three-Layer Eligibility Standard*

All 39 full-text records are screened against a prespecified three-layer framework.

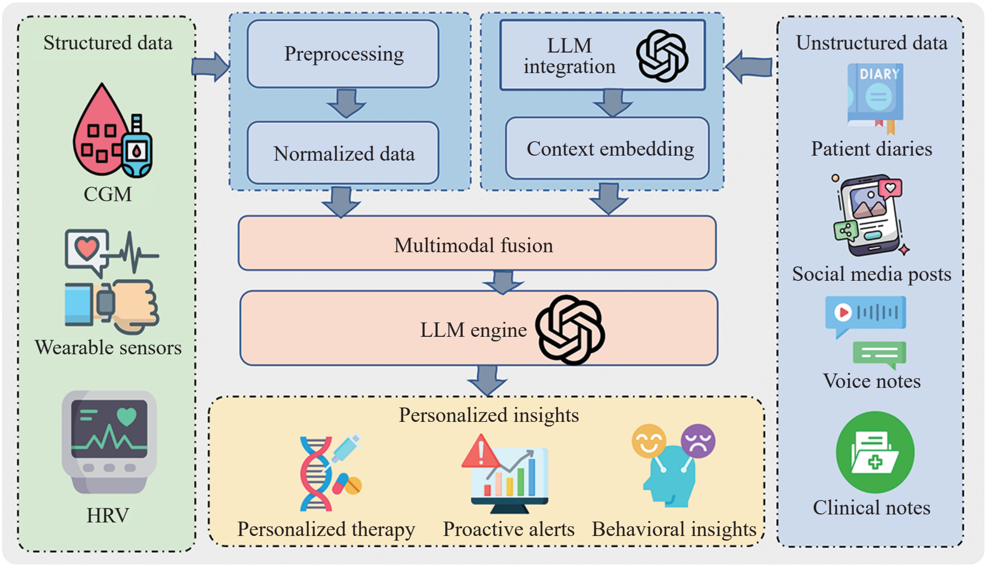


Fig. 2 Diabetes management data integration and personalized insights workflow.

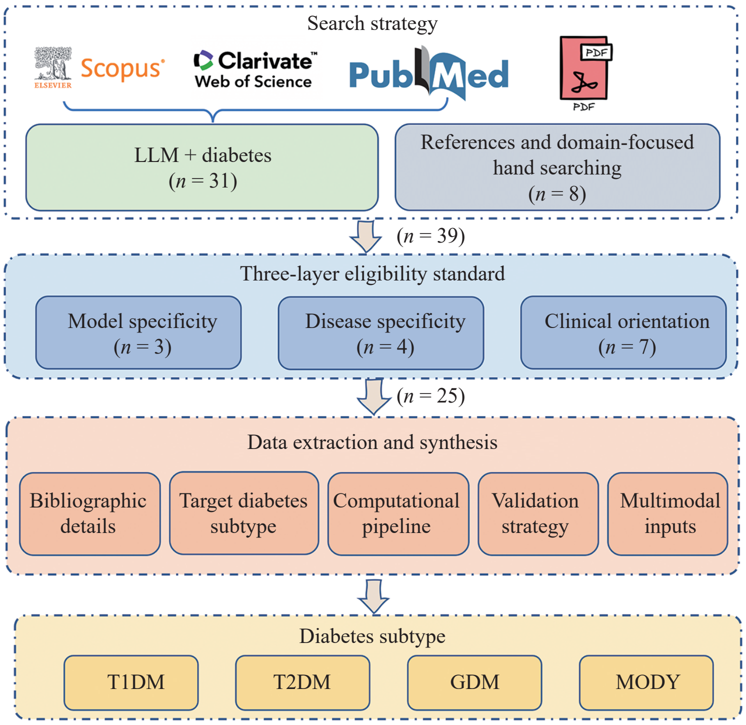


Fig. 3 Methods of literature review.

(1) Model specificity: Only studies that explicitly use large language models (instruction-following or generative NLP architectures) are included. Studies relying solely on traditional machine learning or vision-only pipelines without textual modeling are excluded.

(2) Disease specificity: Directly focus on diabetes, type 1 diabetes mellitus (T1DM), type 2 diabetes mellitus (T2DM), gestational diabetes mellitus, or MODY, excluding broader metabolic or endocrine contexts.

(3) Clinical orientation: Relate to a clinical or translational task (e.g., prognosis and risk stratification, therapeutic decision support, patient interaction, or complication monitoring) with empirical evaluation sufficient to appraise utility.

Records failing any single layer are excluded at that stage. The most common reason for exclusion is reliance on traditional machine learning (ML) pipelines without true LLM components (n = 3). Further 4 studies are excluded for focusing on broader metabolic disease without diabetes-specific stratification, and 7 records are excluded as purely conceptual or theoretical without empirical validation. After screening, 25 studies satisfy all three criteria and are retained as the final evidence pool.

C. Data Extraction and Synthesis

From each included study, we extracted bibliographic details, the target diabetes subtype, the design of the computational pipeline, the validation strategy, and whether multimodal inputs are incorporated. To enhance consistency and efficiency while maintaining rigor, a semi-automated tagging step is applied in which a large language model flags abstracts for subtype mentions and multimodal indicators; and all labels are subsequently verified and, where necessary, refined by manual review.

The final set of 25 studies is synthesized according to diabetes subtype (T1DM, T2DM, GDM, or MODY), reflecting differences in pathophysiology, care priorities, and sources of unstructured data. This subtype-driven organization supports a clinically oriented synthesis consistent with the structure adopted in the results section.

IV. Results of Literature Review

Diabetes mellitus is clinically classified into four major subtypes, T1DM, T2DM, GDM, and MODY, with distinct pathophysiology, disease trajectories, and therapeutic strategies [41, 42]. These differences shape both clinical priorities and the kinds of unstructured data generated: in T1DM, short-horizon glucose volatility under absolute insulin dependence yields high-frequency diaries and CGM annotations [43, 44]; in T2DM, long-term progression with heterogeneous comorbidities produces longitudinal EHR narratives and lifestyle reports [45, 46]; in GDM, the short yet high-risk pregnancy window generates antenatal notes and pregnancy-specific communications [47, 48]; and in MODY, family histories and genetic testing documents are central to diagnosis and management [49, 50]. Because LLMs are designed to parse and structure heterogeneous free-text and multimodal inputs, recent works have explored subtype-specific applications aligned to these data characteristics [51, 52]. In the following sections, we review how LLMs have been applied to address the unique data challenges and management needs of each diabetes subtype. See Table 2 for key LLM applications across different diabetes types, and see Table 3 for a comparative performance of LLMs in diabetes management tasks (illustrative).

A. T1DM: Diaries, Psychosocial Narratives, and Multimodal Short-Horizon Personalization

T1DM is uniquely defined by extreme short-horizon glycemic volatility and absolute dependence on exogenous insulin. Unlike type 2 diabetes, in which longitudinal monitoring dominates, T1DM management often hinges on anticipating excursions over the span of hours. Small contextual factors, such as skipping a meal, sudden physical exertion, and psychosocial stress, can trigger acute swings, underscoring the critical role of unstructured patient-generated data. Diaries, stress notes, exercise logs, and social media entries, when coupled with CGM, capture micro-contexts invisible to numerical biosignals alone [43]. Entries such as "exam anxiety", "skipped breakfast", and "post-exercise

Table 2 Key LLM applications across different diabetes types.

Diabetes type	Specific LLM applications	Illustrative examples/researches	Key outcomes/metrics
T1D	Stress identification, glucose prediction, personalized insulin adjustments, physical activity analysis	Linking "work anxiety" to glucose spikes (LLM+multimodal dataset), reducing glucose prediction MAE, dynamic insulin adjustment based on post-workout fatigue	F1-score for stress detection: 0.83–0.88, Glucose prediction MAE: 15.2–17.2 mg/dL
T2D	Clinical decision support, complication prediction, health literacy enhancement, personalized treatment recommendations	Insulin clinical decision support (iNCDSS) for insulin dosing, EHR data extraction for HbA1c/LDL prediction, RAG chatbots for T2D FAQs	Improved insulin management, enhanced predictive accuracy for metabolic markers, improved patient health literacy
GDM	Diabetes prevention, personalized nutrition counseling, reminder systems, emotional support	GPT-based chatbots for instant answers and advice, RCT evaluating GPT-counseling	Improved glycemic control and patient satisfaction (expected in RCT)
Monogenic diabetes	Risk assessment, genetic analysis, precision therapeutics, multiomics integration	AI/ML for early diagnosis based on genetics, multiomics for individualized dietary strategies	Improved diagnostic accuracy, enhanced personalized treatment plans, accelerated research
General applications	Multimodal data analysis (CGM, wearables, EHRs, logs), CDSS, patient engagement	GlucoLens for hyperglycemia prediction, LLMs for raw CGM data analysis, DeepDR-LLM for personalized recommendations	Hyperglycemia prediction accuracy 73.3%, 35% increase in patient engagement

Table 3 Comparative performance of LLMs in diabetes management tasks (illustrative).

LLM model	Performance metric	Task/domain	Strengths	Weaknesses/challenges
ChatGPT-4o	F1-score: 0.85, MAE: 12.3 mg/dL (glucose prediction), patient engagement improvement: 35%	Stress detection in diaries, glucose prediction (HRV/diary integration), patient engagement	Advanced sentiment analysis, contextual understanding of emotional language, natural language explanations of glucose fluctuations	Difficulties with dialectal features, underestimation of delayed effects of hypoglycemia (MAE: 18.9)
Claude 3.5 Sonnet	F1-score: 0.88 (highest), MAE: no specific MAE value provided after preprocessing	Stress detection (implicit stressors), glucose prediction, physical activity impact analysis	Strong detection of implicit stressors (e.g., "can not sleep"), identification of behavioral barriers to physical activity	Requires pre-processed sensor inputs (limits real-time operation), verbose outputs, privacy concerns due to cloud processing
Gemini Advanced	F1-score: 0.82, latency reduction: 15%, hypoglycemia prediction error reduction: 22%	Stress detection (multi-lingual), glucose prediction (real-time integration), physical activity impact analysis, patient engagement	Multi-lingual stress marker detection (e.g., Spanish), real-time CGM data integration via Wear OS API, reduced post-workout hypoglycemia prediction errors (with LSTM), 20% hyperglycemia reduction in DiabetAI app	Overly literal interpretation of metaphors, limited customization due to proprietary model

fatigue" often precede hypoglycemia or hyperglycemia yet are omitted from structured fields, making them a primary target for LLM-enabled interpretation.

To operationalize this information, recent pipelines have adopted multi-stage architectures that begin with natural language preprocessing of diary narratives, followed by event detection through named entity recognition (NER) trained on health-related corpora [53]. Events are temporally tagged and aligned with CGM signals, producing structured predictors for forecasting models. These enriched datasets are then input into supervised learning frameworks, in which LLMs serve either as feature extractors or as forecasting engines. Validation studies demonstrate that diary-augmented models reduce mean absolute error in short-term glucose forecasting compared with the autoregressive baselines, particularly when stress-related and exercise-related narratives are included [54]. The inclusion of subjective textual cues allows models to anticipate glycemic excursions that numeric-only systems systematically miss. Importantly, these pipelines reduce the cognitive burden traditionally placed on patients, who otherwise must reconcile lifestyle notes and CGM outputs manually when adjusting insulin doses.

The integration of psychosocial data adds another dimension. Research has shown that perceived glucose fluctuations recorded in patient diaries are stronger predictors of diabetes distress than CGM traces alone [55]. Building on this insight, multimodal systems such as StressLLM fuse wearable-derived signals with textual descriptors of stress, anxiety, and fatigue [56]. These hybrid pipelines apply cross-attention mechanisms to jointly interpret physiological variability and psychosocial triggers, generating dual-layered outputs: one predicting near-term glycemic excursions, the other flagging elevated distress risk. This approach expands the therapeutic horizon of T1DM from purely metabolic safety toward holistic well-being, embedding mental health within glycemic control.

Scalability challenges have also prompted the exploration of population-level unstructured data, particularly from social media. GPT-4 zero-shot classifiers have outperformed support vector machine baselines in categorizing diabetes-related narratives across multiple tasks [57]. However, pipelines trained exclusively on synthetic labels generated by LLMs without human oversight produce unreliable results. This finding emphasizes the necessity of human-in-the-loop workflows, in which clinicians validate model outputs to ensure reliability and prevent unsafe recommendations. Such designs balance the scalability of LLM annotation with the fidelity required in clinical translation.

The predictive advantage of T1DM systems is further magnified when diaries and psychosocial narratives are fused with additional sensor data. Pipelines that integrate CGM, wearable accelerometry, and textual diaries employ three core transformations: normalization of biosignals, extraction of lifestyle and emotional context via NLP, and cross-modal merging into LLM-based predictors [58]. These multimodal architectures extend forecasting horizons beyond two to three hours while retaining interpretability, producing outputs that clinicians can audit in real time. Comparative analyses show that multimodal diary-wearable-CGM fusion consistently outperforms unimodal baselines in both accuracy and stability of predictions, thereby offering a robust framework for context-aware insulin dosing.

Taken together, the T1DM literature demonstrates a coherent translational arc. Diaries and stress logs, when processed through large language model based pipelines, improve forecasting precision; psychosocial narratives, when integrated with biosignals, expand prediction to encompass mental health; social media pipelines, when augmented with human oversight, extend these methods to broader populations; and multimodal fusion architectures elevate both accuracy and interpretability. The cumulative result is a short-horizon personalization paradigm in which large language models transform unstructured narratives into clinically actionable signals that sharpen insulin strategies, reduce cognitive burden, and integrate psychosocial care into everyday diabetes management.

B. T2DM: Longitudinal Risk Stratification, Therapeutic Navigation, and Translational Applications

In contrast to the short-horizon volatility of type 1 diabetes, T2DM demands sustained and longitudinal management in the presence of heterogeneous phenotypes and frequent multi-morbidity. The dominant unstructured data, EHR narratives, patient-provider communications, and lifestyle diaries, encode adherence patterns, comorbidity dynamics, and behavioral factors that rarely appear in structured fields, strongly influence outcomes [59]. The literature on large language model applications in type 2 diabetes mellitus can be traced along a continuum: beginning with prognosis modeling, progressing through therapeutic navigation and behavior-sensitive personalization, and extending into workflow integration and complication monitoring.

At the front end of this continuum lies prognosis modeling. Semi-supervised NER pipelines have been developed to parse sparse, multidimensional EHR records, and extracting phenotypic descriptors such as "family history of diabetes", "elevated HbA1c", and "irregular meal timing" [53]. Teacher-student frameworks reduce annotation burden while filtering noise, improving the fidelity of extracted features. These features form the input to predictive systems such as the Di-TaP model, which achieved an AUROC of 0.864 for five-year T2DM onset prediction, outperforming the ADA risk score at 0.771 [60]. Beyond accuracy, Di-TaP incorporates an explainable AI layer that highlights the relative contributions of individual drivers, HbA1c, BMI, and lifestyle irregularities, offering clinicians transparent evidence for proactive intervention. External validation across subgroups has shown improved calibration when multimodal signals such as wearable data are added, underscoring the importance of linking text-derived features to physiological context [54].

Building on the risk assessment, therapeutic navigation constitutes the next domain of application. Here, LLMs combined with knowledge graphs or retrieval-augmented architectures provide traceable and guideline-concordant recommendations for drug selection [52]. These hybrid systems mitigate hallucination by grounding responses in established protocols and offering justifications that can be audited. Comparative evaluations reveal meaningful contrasts: in simulations, GPT-4 frequently adopted more conservative prescribing strategies than physicians when faced with diagnostic uncertainty [61], serving as a safeguard against inappropriate polypharmacy. Against conventional rule-based clinical decision support systems, LLM hybrids demonstrate superior flexibility, updating recommendations dynamically as guidelines evolved rather than relying on static rules [61]. This comparative perspective suggests that LLMs are moving beyond passive summarizers toward "safety co-pilots" that not only synthesize evidence but also mitigate therapeutic risk.

Lifestyle and behavioral modeling extend this trajectory by linking patient-reported logs with metabolic outcomes. Pipelines process dietary diaries, activity records, and patient-reported outcomes, often in conjunction with CGM and wearable accelerometers. Through temporal alignment of narrative events with glycemic fluctuations, these systems have identified recurring behavior-to-risk pathways, such as late-night snacking, stress-induced eating, and disrupted sleep patterns that elevate next-day glucose levels [62]. Feature-engineered approaches emphasize interpretability, isolating behaviors that clinicians can target directly, while multimodal embedding models optimize predictive accuracy by jointly representing text and biosignals [63]. This duality reflects a central design tradeoff: balancing transparency with performance. In practice, behavior-aware pipelines transform lifestyle anecdotes into actionable pathways, enabling a shift from reactive glucose control to anticipatory and behavior-sensitive care.

The translational impact of these systems is reinforced by interventional evidence. In a randomized controlled trial, AI-assisted iNCDSS achieves non-inferior glycemic control compared with endocrinologist-directed titration in hospitalized T2DM patients while reducing clinician workload [64]. This study provides some of the clearest evidence that LLM-augmented systems can function safely in real-world clinical workflows. At the patient-facing interface, retrieval-augmented generation (RAG) chatbots have shown promise in delivering empathetic, contextually relevant, and clinically accurate responses to patient queries [65–67]. By grounding outputs in validated references, these systems improve health and literacy and foster adherence, and extend the reach of professional counseling beyond scheduled encounters.

Finally, the same principles are now being extended to complication monitoring. DeepDR-LLM, a multimodal framework, integrates retinal fundus images with textual EHR descriptors to support diabetic retinopathy screening [68]. Its pipeline encompasses preprocessing and normalization of both image and text inputs, embedding via vision transformers and LLM encoders, cross-modal fusion, and dual outputs: severity classification of retinopathy and personalized textual recommendations for follow-up. By bridging raw multimodal data with structured and auditable advice, DeepDR-LLM exemplifies how LLM-based systems expand beyond core metabolic management into the proactive detection and management of complications.

Taken together, the T2DM literature illustrates a coherent translational arc. Text-derived prognosis informs early risk stratification; guideline-grounded LLMs refine therapeutic navigation; behavior-aware models personalize management to daily realities; interventional trials validate safety and workflow feasibility; and multimodal frameworks extend decision support into complication monitoring. This integrated progression demonstrates how LLMs transform fragmented unstructured data into durable, interpretable, and clinically actionable pathways, repositioning T2DM care from reactive to anticipatory and from fragmented to integrative.

C. GDM: Antenatal Narratives, Digital Counseling, and Safety-Constrained Applications

In contrast with the sustained management required for T1DM and T2DM, GDM presents the unique challenge of a short yet high-risk pregnancy window [69, 70]. GDM generates unstructured data streams that are tightly coupled to pregnancy-specific care but often under-documented in structured records. These include antenatal visit notes describing dietary advice and exercise recommendations, patient diaries of meals and activity, and messages reporting stress, sleep disturbances, and anxiety. Because maternal glycemia is strongly influenced by such lifestyle and psychosocial contexts, these data hold clinical value yet are rarely integrated systematically. LLMs offer an avenue to structure and interpret these narratives, aligning them with guideline-based management protocols [47, 48].

The most prominent pipelines involve digital counseling systems powered by LLMs. These systems typically integrate four components: (1) retrieval of guideline-based responses, (2) generation of context-specific feedback, (3) scheduling of reminders for diet, exercise, and glucose testing, and (4) emotional support through empathetic dialogue. Randomized controlled trials are underway to evaluate their ability to improve glycemic control and patient satisfaction, with preliminary results suggesting potential to augment scarce clinical counseling resources [71]. Importantly, these pipelines are developed under heightened safety constraints, using retrieval-augmented architectures to anchor generated responses to official obstetric and diabetes guidelines, thereby mitigating the risk of misinformation in a high-stakes maternal-fetal context [72].

Despite this promise, the relative paucity of published GDM-focused LLM studies reflects three challenges. First, regulatory scrutiny surrounding pregnancy-related AI tools imposes stringent safety requirements. Second, clinical data remain fragmented between obstetric and endocrinology records, complicating end-to-end model training. Third, the domain demands guideline-anchored generation, as free-form responses can jeopardize maternal or fetal safety. Consequently, current works position LLMs not as autonomous advisors but as structured assistants that scale counseling time, improve adherence monitoring, and contextualize patient-reported narratives into guideline-concordant recommendations. The distinctive contribution in GDM, therefore, is the safe transformation of antenatal text into structured and auditable advice that supports adherence to diet, exercise, and stress management, with direct implications for maternal and fetal outcomes.

D. MODY: Narrative Suspicion, Genetic Reports, and Precision Therapeutics

Distinct from the common and well-characterized T1DM and T2DM, MODY is defined by its rarity and reliance on genetic and family-history narratives [73, 74]. MODY presents unique diagnostic and management challenges because of its rarity and reliance on text-intensive data sources. Family histories describing autosomal-dominant inheritance, narrative notes documenting atypical phenotypes such as young onset or lean body habitus, and long-form genetic testing reports are all crucial to distinguishing MODY from type 1 or type 2 diabetes. Yet, these materials are rarely operationalized into computable features, contributing to frequent mis-classification and under-diagnosis [49].

LLMs provide translational leverage in two primary ways. First, they transform narrative suspicion into structured prompts for genetic testing. For example, NLP pipelines can extract descriptors such as "strong paternal history" or "non-obese early-onset diabetes" from clinical notes and flag them against MODY suspicion criteria [45]. This enables clinicians to more systematically identify candidates for confirmatory genetic evaluation. Second, LLMs streamline the interpretation of genetic testing reports, which are often lengthy and filled with technical detail. By summarizing these documents into concise and guideline-concordant recommendations, large language models can clarify therapeutic implications, such as the sensitivity of HNF1A-MODY patients to sulfonylureas and the insulin requirements of GCK-MODY cases [50].

The scarcity of robust MODY-LLM studies are attributed to three structural barriers: the rarity of confirmed MODY cases, which limits the availability of paired text-genotype corpora; privacy restrictions surrounding genetic data; and the annotation burden posed by long genetic reports. Nevertheless, the potential impact is substantial. By accurately extracting and structuring genetic and narrative data, LLMs directly inform precision pharmacotherapy, reduce inappropriate insulin prescribing, and facilitate targeted screening within affected families. MODY thus represents a frontier application of LLMs: converting rare and complex textual evidence into actionable insights that reshape diagnostic accuracy and therapeutic strategy [49, 50].

V. Discussion

While large language models hold significant promise for diabetes management, their deployment raises important challenges and ethical considerations that shape future research directions.

A. Ethical Concerns and Data Quality

LLMs can inherit biases from unbalanced datasets, sampling methods, and inherent algorithmic prejudices, potentially resulting in unfair or discriminatory outcomes [11]. For example, dialectal variations [75], disparities related to insurance status, and underrepresented racial groups [76] may be poorly handled. Bias in training data can also lead to underestimation of rare complications.

Data quality is critical. Insufficient, biased, and noisy datasets, along with variability in medical images, can compromise generalizability and exacerbate algorithmic bias. Labeling errors and the inability to quantify complication severity further compound these issues. The quality of LLM outputs depends heavily on the reliability, specificity, and volume of input data [77].

Protecting patient privacy is equally important. Anonymizing sensitive data, such as diary entries, presents challenges, and cloud-based processing of some models (e.g., Claude 3.5 Sonnet) can limit their use in clinical applications [78]. Federated learning offers a promising solution by allowing decentralized model trained without sharing raw patient data [4, 28]. In this context, it is also important to clarify whether studies using LLMs with federated learning are included, to better assess their clinical applicability.

Beyond patient monitoring, LLMs are increasingly applied to fundamental research and drug discovery. For instance, the BertADP study [79] employs a fine-tuned protein language model to identify anti-diabetic peptides, outperforming conventional methods and accelerating therapeutic development. Looking forward, multi-omics integration represents a promising research direction. For example, combining genomic data with LLM analysis can aid in the diagnosis of MODY, providing more precise insights for personalized therapy.

B. Human-in-the-Loop Systems

Human-in-the-loop systems [75] emphasize that AI complements rather than replaces clinical expertise. Clinicians remain the final arbiters of AI-generated recommendations, ensuring patient-centered and safe care. Human-in-the-loop systems create a closed feedback loop where human evaluation continuously refines AI outputs, maintaining clinical accountability.

Effective training for clinicians to interpret AI recommendations enhances trust and usability [75]. Studies such as LLM-RawDMeth [80] underscore that while LLMs adhere to guidelines, they often lack individualized clinical evaluation. Human-in-the-loop approaches balance AI efficiency with indispensable human judgment, promoting safe, equitable, and patient-focused care.

C. Future Research Directions

Future research shall prioritize the following directions.

(1) Enhance the reliability of LLMs via RAG systems [13], which mitigate hallucinations and deliver timely and verifiable information.

(2) Integrate multimodal AI with established ML and deep learning (DL) techniques, such as LSTM networks for glucose forecasting or image-based deep learning for diabetic retinopathy detection [13].

(3) Merge multimodal AI with multi-omics datasets (e.g., genomics, proteomics, and metabolomics) to support precision nutrition and personalized therapies [81].

(4) Empower patients with artificial intelligence-driven

wearable self-monitoring devices to advance patient-centered diabetes care [4].

VI. Conclusion

The integration of large language models with advanced monitoring technologies, such as CGM, wearables, and non-invasive biosensors, is fundamentally transforming diabetes care. This convergence bridges the critical gap between quantitative physiological data and qualitative patient experiences, facilitating truly personalized, proactive, and patient-centered management.

LLMs demonstrate promise across all diabetes types: contextualizing stress and exercise in T1D, supporting prognosis, decision-making, and health literacy in T2D, assisting prevention and counseling in gestational diabetes, and enabling precision therapy in monogenic forms via multi-omics integration. Their capacity to analyze unstructured patient narratives alongside structured sensor data provides deep insights into complex glucose dynamics driven by real-world psychosocial and behavioral factors.

This review synthesizes the state-of-the-art applications of LLMs and advanced monitoring in diabetes care. Future research will focus on developing multimodal architectures, conducting rigorous clinical validation, including randomized controlled trials, and addressing ethical concerns such as bias and privacy through federated learning and explainable AI. The human-in-the-loop paradigm remains essential to ensure safe, effective, and trustworthy AI integration, highlighting collaboration between humans and machines rather than replacement. Advances in RAG systems and hybrid AI models will further enhance reliability, clinical utility, and adoption of large language model based interventions in diabetes management.

Acknowledgment

This work was supported by the Starting Research Grant of Óbuda University, the National Research, Development and Innovation Agency (No. 2024-1.1.1-KKV_FÓKUSZ-2024-00074), the National Research, Development and Innovation Fund of Hungary (No. TKP2021-NKTA-36), the National Natural Science Foundation of China (No. 82372051), the Beijing Natural Science Foundation (No. L222099), the Science and Technology Development Fund, Macao SAR (Nos. 0093/2023/RIA2, 0145/2023/RIA3, and 0157/2024/RIA2), the Capital's Funds for Health Improvement and Research (No. 2024-2-4015), and the Sichuan Key-Area Research and Development Program (No. 2024YFHZ0011).

References

[1] WHO, Diabetes [Online], https://www.who.int/news-room/fact-sheets/detail/diabetes, 2 August 2025.

[2] Y. Tsushima and N. Galloway, Glycemic targets and prevention of complications, *J. Clin. Endocrinol.*, 2025, 110(S2), S100–S111,

[3] Z. Guan, H. Li, R. Liu, C. Cai, Y. Liu, J. Li, X. Wang, S. Huang, L. Wu, D. Liu et al., Artificial intelligence in diabetes management: Advancements, opportunities, and challenges, *Cell Rep. Med.*, 2023, 4(10), 101213.

[4] M. Iftikhar, M. Saqib, S. N. Qayyum, R. Asmat, H. Mumtaz, M. Rehan, I. Ullah, I. Ud-Din, S. Noori, M. Khan et al., Artificial intelligence-driven transformations in diabetes care: A comprehensive literature review, *Ann. Med. Surg.*, 2024, 86(9), 5334–5342.

[5] S. Sakib, O. F. Akhand, and A. Abrar, From chat to checkup: Can large language models assist in diabetes prediction? arXiv preprint arXiv: 2506.14949, 2025.

[6] B. Sheng, Z. Guan, L. L. Lim, Z. Jiang, N. Mathioudakis, J. Li, R. Liu, Y. Bao, Y. M. Bee, Y.-X. Wang et al., Large language models for diabetes care: Potentials and prospects, *Sci. Bull.*, 2024, 69(5), 583–588.

[7] F. L. Schwartz, C. R. Marling, and R. C. Bunescu, The promise and perils of wearable physiological sensors for diabetes management, *J. Diabetes Sci. Technol.*, 2018, 12(3), 587–591.

[8] J. Saito and H. Kumano, Psychosocial factors influencing dietary management in patients with type 2 diabetes and healthy adults: An ecological momentary assessment approach, *Front. Psychol.*, 2025, 15, 1464542.

[9] D. Ehrmann, N. Hermanns, A. Schmitt, L. Klinker, T. Haak, and B. Kulzer, Perceived glucose levels matter more than CGM-based data in predicting diabetes distress in type 1 or type 2 diabetes: A precision mental health approach using *n*-of-1 analyses, *Diabetologia*, 2024, 67(11), 2433–2445.

[10] M. Sevil, M. Rashid, I. Hajizadeh, M. Park, L. Quinn, and A. Cinar, Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management, *IEEE Trans. Biomed. Eng.*, 2021, 68(7), 2251–2260.

[11] S. Aydin, M. Karabacak, V. Vlachos, and K. Margetis, Navigating the potential and pitfalls of large language models in patient-centered medication guidance and self-decision support, *Front. Med.*, 2025, 12, 1527864.

[12] J. Lee, H. Cha, Y. Hwangbo, and W. Cheon, Enhancing large language model reliability: Minimizing hallucinations with dual retrieval-augmented generation based on the latest diabetes guidelines, *J. Pers. Med.*, 2024, 14(12), 1131.

[13] A. Kelly, E. Noctor, L. Ryan, and P. van de Ven, The effectiveness of a custom AI chatbot for type 2 diabetes mellitus health literacy: Development and evaluation study, *J. Med. Internet Res.*, 2025, 27, e70131.

[14] F. Busch, L. Hoffmann, C. Rueger, E. H. van Dijk, R. Kader, E. Ortiz-Prado, M. R. Makowski, L. Saba, M. Hadamitzky, J. N. Kather et al., Current applications and challenges in large language models for patient care: A systematic review, *Commun. Med.*, 2025, 5(1), 26.

[15] W. Xu, L. Lu, Y. He, L. Cheng, and A. Liu, Long-term detection of glycemic glucose/hypoglycemia by microfluidic sweat monitoring patch, *Biosensors*, 2024, 14(6), 294.

[16] V. Lopes, T. Abreu, M. Abrantes, S. S. Nemala, F. De Boni, M. Prato, P. Alpuim, and A. Capasso, Graphene-based glucose sensors with an Attomolar limit of detection, *J. Am. Chem. Soc.*, 2025, 147(15), 13059–13070.

[17] J. Seo, J. Kang, J. Kim, H. Han, M. Park, M. Shin, and K. Lee, Smart contact lens for colorimetric visualization of glucose levels in the body fluid, *ACS Biomater. Sci. Eng.*, 2024, 10(6), 4035–4045.

[18] J. Xiao, Y. Liu, L. Su, D. Zhao, L. Zhao, and X. Zhang, Microfluidic chip-based wearable colorimetric sensor for simple and facile detection of sweat glucose, *Anal. Chem.*, 2019, 91(23), 14 803–14807.

[19] X. Liang, S. Zhang, S. Meng, R. Tan, K. Zhang, and J. Hu, A non-enzymatic flexible and wearable sensor based on thermal transfer printing technology for continuous glucose detection in sweat, *Microchem. J.*, 2024, 207, 111690.

[20] J. Kim, G. Valdés-Ramírez, A. J. Bandodkar, W. Jia, A. G. Martinez, J. Ramírez, P. Mercier, and J. Wang, Non-invasive mouthguard biosensor for continuous salivary monitoring of metabolites, *Analyst*, 2014, 139(7), 1632–1636.

[21] A. A. T. Schuurmans, P. De Looff, K. S. Nijhof, C. Rosada, R. H. J. Scholte, A. Popma, and R. Otten, Validity of the empatica E4 wristband to measure heart rate variability (HRV) parameters: A comparison to electrocardiography (ECG), *J. Med. Syst.*, 2020, 44(11), 190.

[22] H. Jin, Y. S. Abu-Raya, and H. Haick, Advanced materials for health monitoring with skin-based wearable devices, *Adv. Healthc. Mater.*, 2017, 6(11), 1700024.

[23] M. V. Perez, K. W. Mahaffey, H. Hedlin, J. S. Rumsfeld, A. Garcia, T. Ferris, V. Balasubramanian, A. M. Russo, A. Rajmane, L. Cheung et al.,

Large-scale assessment of a smartwatch to identify atrial fibrillation, *N. Engl. J. Med.*, 2019, 381(20), 1909–1917.

[24] L. Piwek, D. A. Ellis, S. Andrews, and A. Joinson, The rise of consumer health wearables: Promises and barriers, *PLoS Med.*, 2016, 13(2), e1001953.

[25] A. Y. Alhaddad, H. Aly, H. Gad, A. Al-Ali, K. K. Sadasivuni, J. J. Cabibihan, and R. A. Malik, Sense and learn: Recent advances in wearable sensing and machine learning for blood glucose monitoring and trend-detection, *Front. Bioeng. Biotechnol.*, 2022, 10, 876672.

[26] W. Zhang and M. Wang, Saliva glucose monitoring system, U.S. Patent 20140197042A1, 2014.

[27] C. Ji, T. Jiang, L. Liu, J. Zhang, and L. You, Continuous glucose monitoring combined with artificial intelligence: Redefining the pathway for prediabetes management, *Front. Endocrinol.*, 2025, 16, 1571362.

[28] D. C. Klonoff, R. M. Bergenstal, E. Cengiz, M. A. Clements, D. Espes, J. Espinoza, D. Kerr, B. Kovatchev, D. M. Maahs, J. K. Mader et al., CGM data analysis 2.0: Functional data pattern recognition and artificial intelligence applications, *J. Diabetes Sci. Technol.*, 2025, 14, 19322968251353228.

[29] E. Healey, A. Tan, K. Flint, J. Ruiz, and I. Kohane, Leveraging large language models to analyze continuous glucose monitoring data: A case study, *Sci. Rep.*, 2025, 15, 1143.

[30] S. K. Garg, M. Kipnes, K. Castorino, T. S. Bailey, H. K. Akturk, J. B. Welsh, M. P. Christiansen, A. K. Balo, S. A. Brown, J. L. Reid et al., Accuracy and safety of dexcom g7 continuous glucose monitoring in adults with diabetes, *Diabetes Technol. Ther.*, 2022, 24(6), 373–380.

[31] P. Avari, R. Pushparatnam, L. Leelarathna, T. Tan, A. H. Frankel, N. Oliver, and M. Reddy, Accuracy of the dexcom G7 continuous glucose monitoring sensors in people with diabetes undergoing hemodialysis (ALPHA-2 study), *Diabetes Technol. Ther.*, 2025, 27(5), 402–406.

[32] E. G. Wilmot, M. Evans, K. Barnard-Kelly, M. Burns, I. Cranston, R. A. Elliott, G. Gkountouras, N. Kanumilli, A. Krishan, C. Kotonya et al., Flash glucose monitoring with the FreeStyle libre 2 compared with self-monitoring of blood glucose in suboptimally controlled type 1 diabetes: The FLASH-UK randomised controlled trial protocol, *BMJ Open*, 2021, 11(7), e050713.

[33] A. A. Al Hayek, A. A. Robert, and M. A. Al Dawish, Evaluation of FreeStyle libre flash glucose monitoring system on glycemic control, health-related quality of life, and fear of hypoglycemia in patients with type 1 diabetes, *Clin. Med. Insights Endocrinol. Diabetes*, 2017, 10, 1179551417746957.

[34] Z. Du, F. Zhang, Y. Ge, Y. Liu, H. Yu, Y. Wang, R. Dalan, and X. Shen, Application of wearable devices in diabetes management, *Health Metab.*, 2025, 2, 7.

[35] V. S. Fengade, H. Swati, M. Chandak, R. Rattan, A. Singhal, P. Kamble, M. Phatak, and N. John, Development of enhanced machine learning models for predicting type 2 diabetes mellitus using heart rate variability: A retrospective study, *Cureus*, 2025, 17, e80933.

[36] A. Coppola, S. Conte, D. Pastore, F. Chiereghin, and G. Donadel, Multifractal heart rate value analysis: A novel approach for diabetic neuropathy diagnosis, *Healthcare*, 2024, 12(2), 234.

[37] B. Thapa, M. Rivas, H. Griffith, and H. Rathore, StressLLM: Large language models for stress prediction via wearable sensor data, in *Proc. 2025 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2025, 1–6.

[38] P. Z. Chan, E. Jin, M. Jansson, and H. S. J. Chew, AI-based noninvasive blood glucose monitoring: Scoping review, *J. Med. Internet Res.*, 2024, 26, e58892.

[39] W. Park, H. Seo, J. Kim, Y. M. Hong, H. Song, B. J. Joo, S. Kim, E. Kim, C. G. Yae, J. Kim, J. Jin, J. Kim et al., In-depth correlation analysis between tear glucose and blood glucose using a wireless smart contact lens, *Nat. Commun.*, 2024, 15(1), 2828.

[40] E. F. M. Gabriel, P. T. Garcia, F. M. Lopes, and W. K. T. Coltro, Paper-based colorimetric biosensor for tear glucose measurements, *Micromachines*, 2017, 8(4), 104.

[41] American Diabetes Association, Standards of medical care in diabetes—2018, *Diabetes Care*, 2018, 41(Suppl 1), S1–S159.

[42] M. J. Davies et al., Management of type 1 diabetes in adults: A consensus report, *Diabetologia*, 2020, 63, 2551–2562.

[43] J. Healey et al., Diary-integrated LLM pipelines for CGM interpretation, *Sci. Rep.*, 2025.

[44] L. Smith et al., Forecasting glucose excursions from diaries and CGM using LLMs, *JMIR Diabetes*, 2024.

[45] J. Zhou and L. Wang, Phenotyping algorithms from unstructured text for precision diabetes care, *J. Biomed. Inf.*, 2023, 136, 104251.

[46] W. Zhang et al., LLM-based prognosis prediction from free-text EHR notes, *JAMIA*, 2024.

[47] A. Smith and Y. Chen, Clinical notes and care challenges in gestational diabetes management, *J. Matern.-Fetal Med.*, 2021, 34(7), 1234–1242.

[48] F. Brown and X. Li, International guidelines on gestational diabetes care: Consensus and challenges, *Diabetes Care*, 2020, 43(6), 1073–1080.

[49] S. Fajans, G. Bell, and K. Polonsky, Monogenic diabetes: Diagnosis, clinical features, and management, *Lancet Diabetes Endocrinol.*, 2018, 6(8), 647–656.

[50] S. Ellard and K. Colclough, Precision medicine in monogenic diabetes: Translating genetic reports into therapy, *Diabetologia*, 2023, 66(4), 567–579.

[51] T. Dao, H. Lee, and Y. Zhang, LLM-powered multimodal AI for healthcare applications, *NPJ Digit. Med.*, 2024, 7(1), 123–135.

[52] J. Qin and P. Li, Knowledge graph-enhanced LLMs for safe therapeutic decision-making in diabetes, *Artif. Intell. Med.*, 2025, 150, 102576.

[53] L. Zuo and K. Zhang, Sner: Semi-supervised named entity recognition for diabetes-related clinical texts, *J. Biomed. Inf.*, 2025, 138, 104230.

[54] J. Lara-Abelenda et al., Personalized glucose forecasting with diary-augmented LLMs, *Diabetes Technol. Ther.*, 2025, 27, 210–221.

[55] H. Smith and R. Johnson, Perceived glucose fluctuations as predictors of diabetes distress, *Diabet. Med.*, 2023, 40, e15237.

[56] P. Thapa and Y. Wang, StressLLM: A multimodal large language model for stress-aware glucose prediction, in *Proc. ACM CHIL*, 2025, 315–327.

[57] C. Pan and D. Lee, Enhancing classification of health-related social media texts with GPT-4, *PLOS Digital Health*, 2024, 3, e001234.

[58] A. Mamun et al., Glucolens: An explainable multimodal system for hyperglycemia prediction, *IEEE J. Biomed. Health Inform.*, 2025.

[59] L. Garcia and M. Patel, Phenotyping type 2 diabetes from unstructured EHR narratives, *J. Am. Med. Inform. Assoc.*, 2023, 30, 1579–1590.

[60] X. Yan et al., Di-tap: Deep learning model for five-year T2DM risk prediction, *Nat. Med.*, 2024, 30, 245–256.

[61] S. Kelly and T. Nguyen, Effectiveness of custom AI systems for clinical drug selection in T2DM, *Lancet Digit. Health*, 2025.

[62] X. Li and H. Chen, Behavior-to-risk pathways in diabetes: LLM-based lifestyle modeling, *Diabetes Care*, 2025, 48, 991–1003.

[63] Y. Soumma et al., Counterfactual modeling of lifestyle triggers with LLMs, *NPJ Digit. Med.*, 2025.

[64] Z. Ying et al., Real-time AI-assisted insulin titration in hospitalized T2DM: A randomized trial, *JAMA Network Open*, 2025, 8, e252341.

[65] L. Ji and Y. Zhao, Retrieval-augmented generation for patient-facing diabetes chatbots, *NPJ Digit. Med.*, 2025, 8, 59.

[66] S. Kelly, A. Wong, and P. Lee, Evaluating patient satisfaction with llmbased chatbots for diabetes management, *J. Med. Internet Res.*, 2025, 27(3), e55678.

[67] A. Hussain et al., Retrieval-augmented generation for diabetes patient education, *BMC Med. Inf.*, 2025.

[68] S. Choi and H. Park, DeepDR-LLM: Multimodal integration for diabetic retinopathy detection, *IEEE Trans. Med. Imaging*, 2025, 44, 2023–2036.

[69] W. Zhao, A. Gupta, and J. Torres, Artificial intelligence in gestational diabetes: A systematic review, *BMJ Open Diabetes Res. Care*, 2025, 13(1), e003456.

[70] H. Al-Saad, M. Chen, and E. Rossi, Digital interventions and AI models for gestational diabetes care: Current evidence and gaps, *Diabetes Res. Clin. Pract.*, 2025, 210, 110123.

[71] M. Garcia and R. Patel, Evaluating GPT-based digital counseling for gestational diabetes: Study protocol for a randomized trial, in *Proc. International Conference on Digital Health*, Nanjing, China, 2024.

[72] L. Thompson and W. Zhao, Ethical and regulatory frameworks for AI use in pregnancy: Implications for gestational diabetes, *BMJ Health Care Inf.*, 2022, 29(1), e100512.

[73] P. Sharma, R. Singh, and E. Andersson, Advances in genetic testing and AI applications for monogenic diabetes, *Front. Endocrinol.*, 2025, 16, 1256.

[74] J. Murphy, R. Patel, and L. Chen, Clinical practice guidelines for monogenic diabetes: Diagnosis and management, *Endocr. Rev.*, 2023, 44(5), 789–812.

[75] J. M. Pavon, D. Schlientz, M. L. Maciejewski, N. Economou-Zavlanos, and R. H. Lee, Large language models in diabetes management: The need for human and artificial intelligence collaboration, *Diabetes Care*, 2025, 48(2), 182–184.

[76] W. Hussain and J. Grundy, Advice for diabetes self-management by ChatGPT models: Challenges and recommendations. arXiv: 2501.07931, 2025.

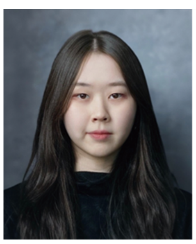
[77] F. F. Falah and N. Arbangi, The role of AI in chronic disease management: A literature review, *Int. J. Sci. Adv.*, 2025, 6(2), 293–296.

[78] O. Agnes Regina, Artificial intelligence in diabetes care: Transforming diagnosis, management, and research—a mini review, *Idosr J. Comput. Appl. Sci.*, 2024, 9(1), 11–14.

[79] X. Xie, C. Wu, Y. Qi, S. Liu, J. Huang, H. Lyu, F. Dao, and H. Lin, BertADP: A fine-tuned protein language model for anti-diabetic peptide prediction, *BMC Biol.*, 2025, 23(1), 210.

[80] J. F. Gaitan-Guerrero, C. Martínez-Cruz, M. Espinilla, D. Díaz-Jiménez, and J. L. López, A novel fine-tuning and evaluation methodology for large language models on IoT raw data summaries (LLM-RawDMeth): A joint perspective in diabetes care, *Comput. Methods Programs Biomed.*, 2025, 269, 108878.

[81] S. Mahdavi, Advancing nutritional genomics in the era of big data, multiomics and artificial intelligence, *BMJ Nutr. Prev. Health*, 2025.



Adina Chotbaeva is currently pursuing the PhD degree in applied informatics and applied mathematics at Óbuda University, Budapest, Hungary. Her research interests include how physical activity and other factors affect blood glucose levels.

Mengmeng Zhang received the BS degree in computer science from Zhuke Zhen College, Zhejiang University, Hangzhou, China, in 2022. She is currently pursuing the PhD degree in pattern recognition and intelligent system at Institute of Automation, Chinese Academy of Sciences, China, and at School of Artificial Intelligence, University of Chinese Academy of Sciences, China. Her research interests include the intersection of computer vision and medical imaging applications, universal medical image segmentation models, synthetic data-assisted segmentation, and visual language models for diagnosis and reasoning based on medical images.

Máté Siket is currently a postdoctoral researcher at Physiological Controls Research Center and John von Neumann Faculty of Informatics, Óbuda University, Budapest, Hungary. His research interests include the intersection of modern therapies for diabetes and engineering methods. He develops control algorithms for automated insulin delivery systems, estimation algorithms of blood glucose dynamics, and mathematical models. He actively develops an open-source metabolic simulator in which physical activities, meal intakes, and different insulin therapies can be developed and trialed.

Jing Wang is currently pursuing the PhD degree in intelligent science and systems study at Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Macao, China, and is an engineer at State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include parallel intelligence, smart medical, and large language model.

Levente Kovács received the DSc degree from Hungarian Academy of Sciences, Hungary, in 2024 and the PhD degree from Budapest University of Technology and Economics, Hungary, in 2008. He is currently a full professor at Óbuda University, Hungary, where he has served as rector since 2019 and founded the Physiological Controls Research Center. His research interests include modern control theory and physiological controls, particularly personalized healthcare and tumor control. He is a Stanford top 2% cited researcher and was the recipient of a prestigious ERC Starting Grant in 2015. He is an IEEE senior member and has held leadership roles, including chair of the IEEE Hungary Section. He is a recipient of the Dennis Gábor Award and holds multiple Doctor Honoris Causa titles.

György Eigner received the PhD degree from John von Neumann Faculty of Computer Science, Doctoral School of Applied Informatics and Applied Mathematics, Óbuda University, Hungary. His research interests include the application of advanced control methods in physiological relationships. He is a member of the Board of Governors of the IEEE System, Human and Cybernetics Society as a representative of the at-large members and chair of the Young Professionals Subcommittee of the SMCS. He is also an associate editor of the international journal *ACTA Polytechnica Hungarica* and a member of the Robotics College, Óbuda University, Hungary.

Xuanting Ding is currently pursuing the BS degree at Department of Automation, School of Information Science and Technology, Tsinghua University, Beijing, China. Her research interests include digital biology and intelligent health.

Fei-Yue Wang received the PhD degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990. He is currently a professor at State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, at School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, and also at Macau Institute of Systems Engineering, Macau University of Science and Technology, Macao, China. His research interests include methods and applications for parallel systems, social computing, and knowledge automation. He received the IEEE ITS Outstanding Application and Research Awards in 2009, 2011, and 2015, and the IEEE SMC Norbert Wiener Award in 2014. In 2021, he became the IFAC Pavel J. Nowacki Distinguished Lecturer. He is a fellow of the International Council on Systems Engineering, International Federation of Automatic Control, American Society of Mechanical Engineers, and American Association for the Advancement of Science.