Journal of Integrated Sciences
Special Issue, February 2026
ISSN: 2806-4801

ARTIFICIAL INTELLIGENCE IN PRE-DIABETES: CGM-DERIVED PHENOTYPES AND NON-INVASIVE BIOMARKERS FOR EARLY DETECTION AND RISK STRATIFICATION

Sajib Paul[1], Jafrin Zakaria[2], Antor Das[3], Jafidul Aziz[4], Asaduzzaman Anonno[5], Sayan Rahman Oni[6], Mehzabeen Ifty[1], Mamtaz Mariam Asha[7]

[1]School of Pharmacy, BRAC University, Dhaka 1212. Bangladesh
[2]Department of English and Humanities, BRAC University, Dhaka 1212, Bangladesh
[3]Department of Life Sciences, Independent University Bangladesh, Dhaka 1229, Bangladesh
[4]BRAC Business School, BRAC University, Dhaka 1212, Bangladesh
[5]Department of Computer Science and Engineering, BRAC University, Dhaka 1212, Bangladesh
[6]Department of Computer Science and Engineering, East West University, Aftabnagar Dhaka-1212, Bangladesh
[7]Department of Public Health, Daffodil International University, Savar, Dhaka 1216, Bangladesh

ABSTRACT

Prediabetes represents a critical window for preventing progression to type 2 diabetes mellitus (T2DM). Artificial intelligence (AI) offers novel opportunities for precision prevention through early detection, risk stratification, and individualized lifestyle interventions. We reviewed twenty-seven recent studies that explored applications of AI in prediabetes, including machine learning models leveraging electronic health records, deep learning *approaches using wearable sensors, and non-invasive digital biomarkers. Emerging evidence strengthens the case for AI-enabled prediabetes management, with area under the receiver operating characteristic curve (AUROC) values ranging from 0.79 to 0.91 and root mean squared error (RMSE) below 15 mg/dL, surpassing conventional methods. AI-driven clustering reveals distinct metabolic phenotypes, highlighting the potential for tailored preventive strategies. Despite promising performances, challenges such as external validation, model generalizability, fairness, and clinical implementation remain. Current literature is also limited by homogeneous study designs and poor external validation, restricting opportunities for clinical translation. Our findings underscore the transformative potential of AI in prediabetes management by uniquely integrating available evidence, while emphasizing the need for rigorous translational studies to achieve scalable, data-driven, and personalized prevention at the population level.*

Keywords: *Prediabetes, Artificial Intelligence, Precision Prevention, Digital Biomarkers, Continuous Glucose Monitoring, Personalized Intervention.*

Corresponding author: Sajib Paul can be contacted at paul.sajib4048@gmail.com

1. INTRODUCTION

Prediabetes refers to the condition of impaired fasting glucose (IFG) or impaired glucose tolerance (IGT), representing an intermediate metabolic state between normal blood glucose levels and type 2 diabetes mellitus (T2DM). Currently, 10–15% of adults worldwide are estimated to be prediabetic, and if their conditions are left unmanaged will progress to T2DM within five years (American Diabetes Association, 2024). Early diagnosis and early intervention play a central role in preventing this development, reducing cardiovascular risk, and mitigating long-term complications. However, the most common screening modalities, such as fasting plasma glucose, glycated hemoglobin (HbA1c), and oral glucose tolerance tests, still play the most important role in the process of prediabetes detection. Sadly, all of these techniques are constrained by their ability to capture both dynamic glycemic fluctuations and inter-individual variation (Zhang et al., 2023).

The complexity of disease prediction, diagnosis, and management can be addressed by Artificial Intelligence and, therefore, produce a transformative impact on the healthcare system. Regarding the creation of prediabetes detection, risk stratification, and individual prevention measures, AI creates opportunities that have never been observed previously. Its capacity to handle complex data sets can be employed to eliminate the shortcomings of conventional methods employed to identify and treat prediabetes by detecting patterns of glycemic variability and inter-individual variation. The significance of CGM, wearable devices, electronic healthcare records (EHR), and other digital biomarkers is to give signals about diabetes based on multifaceted, high-frequency data, which can be processed with various AI methods, such as machine learning (ML) and deep-learning frameworks, as opposed to the conventional methods (Cell Reports Medicine Editorial Team, 2023; Gao et al., 2023). Therefore, AI aids in predicting glycemic fluctuations, early identification of high-risk patients who are at risk of progressing to type 2 diabetes mellitus (T2DM), and the individualization of lifestyle interventions in real-time (Kim et al., 2025; Zhang et al., 2023). As a simple illustration, sophisticated deep-learning models such as long short-term memory (LSTM) networks and other recurrent neural networks can analyze the time-varying glucose dynamics variables (using CGM data), thereby enhancing the early detection and continuous risk assessment (Li et al., 2021; Wu et al., 2024). Moreover, noninvasive digital biomarkers can be inferred, e.g., glycemic data collected by wearables, and analyzed to provide scalable solutions to population-wide prediabetes screening without recourse to invasive procedures historically used (Farahmand et al., 2025; Wang et al., 2025).

As they are entirely novel to the healthcare setting, their applications have to encounter problems related to clinical translations, including model calibration, external validation, equitable performance in other populations, and proven behavior-change impacts (Miller & Miller, 2024; Khalid et al., 2025). Unfortunately, in the absence of these building blocks, it remains true that the recommendations of the American Diabetes Association (ADA) are yet to internalize AI-based stratification (American Diabetes Association, 2024), thereby creating a critical disconnect between new discoveries and practice. This gap underscores the need to further investigate how these methods can be applied in prediabetes with particular attention to the above clinical translation issues in order to come up with minimum reporting requirements and inform future translational research.

The purpose of this review is to conduct a profound and committed examination of the use of different artificial intelligence methods in prediabetes. It entails mapping of evidence based on CGM-derived phenotypes to noninvasive digital biomarkers and their role in early detection, prediction of progression risk, phenotype discovery, and personalized preventive interventions. In this way, it will give power to the evidence-based incorporation of artificial intelligence into the management of prediabetes, will expand the range of overcoming clinical translation issues, and will eventually give directions to future research priorities.

2. RESEARCH METHODOLOGY

We conducted a structured scoping review to identify and analyze studies on the application of artificial intelligence in prediabetes, focusing on CGM-derived phenotypes, noninvasive digital biomarkers, and predictive or personalized interventions. As this study employed a scoping review approach, a formal PRISMA-based systematic screening with numerical reporting was not conducted. Instead, we adopted a thematic selection strategy, emphasizing studies directly relevant to AI applications in prediabetes and early detection. To enhance transparency, a simplified flow diagram (Figure 01) illustrates the screening and inclusion process.

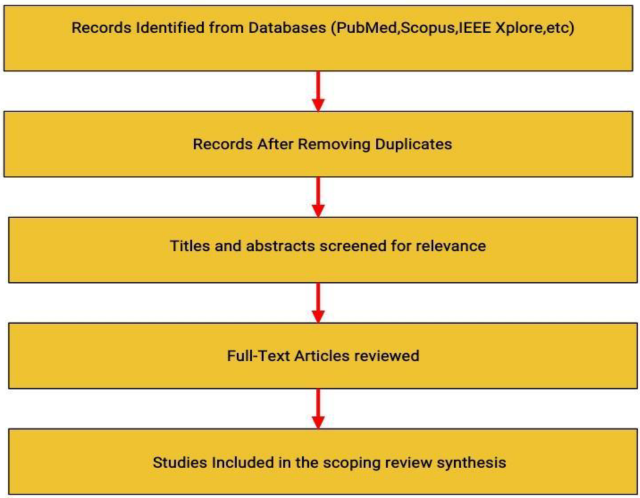

Figure 1. Study Section Flow Diagram (Scoping Review)

2.1 Search Strategy

We performed literature searches in three databases, including PubMed, Scopus, and IEEE Xplore, for publications from January 2010 to June 2025. In addition to database search, reference lists of relevant articles were also screened to identify additional studies. Keywords for searching were established through domain expertise and literature and were related to prediabetes, artificial intelligence, machine learning, deep learning, wearables, and continuous glucose monitoring.

2.2 Inclusion and Exclusion Criteria

To be eligible for inclusion in the review, studies have to meet the following criteria: peer-reviewed, English-language studies indexed in Scopus that applied artificial intelligence to prediabetes using CGM, wearable devices, electronic health records, or other noninvasive digital biomarkers. Exclusion criteria included studies focused exclusively on type 1 or type 2 diabetes, non-original research such as commentaries or abstracts, or non-peer-reviewed sources and non-English publications. Titles, abstracts, and full texts were independently screened by the reviewers.

2.3 Study Selection and Data Extraction

During the study selection process, reviewers focused on the AI model employed, data source, task focus (early detection, progression risk prediction, phenotype discovery, lifestyle feedback, or noninvasive glucose inference), performance metrics, validation strategy, and clinical relevance. Findings were synthesized based on five key domains: (1) early detection and screening, (2) progression risk prediction from prediabetes to type 2 diabetes, (3) personalized lifestyle feedback and just-in-time adaptive interventions, (4) phenotype discovery and metabolic subtyping, and (5) noninvasive glucose inference ("virtual CGM"). This approach allowed a comprehensive mapping of the evidence, identification of methodological gaps, and formulation of recommendations for minimum reporting standards in AI applications for prediabetes.

3. RESULTS

This section provides a comprehensive overview of the findings from our reviewed studies. Firstly, we have discussed the characteristics of our selected studies, and then various methods of artificial intelligence found in the studies were discussed.

Journal of Integrated Sciences
Special Issue, February 2026
ISSN: 2806-4801

Table 1. Summary of AI Applications in Prediabetes Across Domains

Task Domain	AI/ML Models	Data Sources	Performance	Representative Studies
Early Detection & Screening	Random Forest, XGBoost, CNN	EHR, wearable lifestyle data	AUROC 0.79–0.91	(Pukale et al., 2025; Shojaee et al., 2025; Zhang et al., 2025; Alonso et al., 2023)
Progression Risk (Prediabetes → T2DM)	Gradient Boosting, LightGBM, Logistic Regression	Longitudinal EHR, metabolic + dietary features	AUROC 0.81–0.84	(Chen et al., 2023; Diabetology & Metabolic Syndrome Authors, 2025; Shojaee et al., 2025)
Noninvasive Biomarker Inference ("Virtual CGM")	Deep Learning, CNNs, Multimodal Fusion	Wearables (PPG, HRV, sleep), life-long data	RMSE <15 mg/dL vs. CGM	(Alonso et al., 2023; Farahmand et al., 2025; Kim et al., 2025; Wang et al., 2025)
Phenotype Discovery & Metabolic Subtyping	Unsupervised Clustering, Curve-shape Modeling	CGM profiles, OGTT trajectories	Distinct phenotypes predictive of progression	(Shojaee et al., 2025; Zhou et al., 2024)
Just-in-Time Adaptive Interventions (JITAI)	Reinforcement Learning, Personalized feedback loops	Mobile health apps + CGM/wearables	Improved adherence, modest ↓ fasting glucose (3–6 mo)	(Alonso et al., 2023; Shojaee et al., 2025)

Source: The authors' own work.

3.1 Early Detection of Prediabetes

Machine learning approaches have demonstrated utility in detecting prediabetic states using routinely collected health data. Random forest and XGBoost models applied to electronic health records (EHR) achieved AUROC values ranging from 0.79 to 0.91, outperforming traditional logistic regression baselines (Pukale et al., 2025; Shojaee et al., 2025). Convolutional neural networks (CNNs) trained on wearable-derived lifestyle features also provided early warning of impaired glucose tolerance, highlighting the feasibility of AI-assisted community screening (Alonso et al., 2023; Kim et al., 2025). These findings suggest that AI can complement fasting glucose and HbA1c thresholds, particularly in populations where early dysglycemia is underdiagnosed (Zhang et al., 2025).

3.2 Progression Risk: From Prediabetes to Type 2 Diabetes

Several studies evaluated AI for predicting conversion from prediabetes to type 2 diabetes. Gradient boosting, LightGBM, and logistic regression models trained on longitudinal EHR and metabolic indicators consistently achieved AUROC values of 0.81–0.84 for 3–5-year progression prediction (Chen et al., 2023; Diabetology & Metabolic Syndrome Authors, 2025; Shojaee et al., 2025). Importantly, models integrating lifestyle and dietary features outperformed those based solely on biochemical markers (Chen et al., 2023). However, relatively few studies incorporated external validation cohorts, limiting generalizability (Gao et al., 2023).

3.3 Noninvasive Biomarker Inference (“Virtual CGM”)

Wearable-derived data streams such as photoplethysmography (PPG), heart rate variability, and sleep signals are increasingly used for glycemic inference. Deep learning models trained on multimodal wearable inputs approximated CGM profiles with root mean squared error (RMSE) <15 mg/dL in proof-of-concept studies (Alonso et al., 2023; Kim et al., 2025; Wang et al., 2025; Farahmand et al., 2025) While performance remains below invasive CGM, these approaches show promise for scalable, low-burden screening of prediabetes in community populations (Wang et al., 2025).

3.4 Phenotype Discovery and Metabolic Subtyping

AI has enabled unsupervised clustering and curve-shape modeling of glucose trajectories to reveal heterogeneity within prediabetes. Building on these findings, CGM trajectories reveal distinct metabolic phenotypes, including differential postprandial responses and insulin resistance profiles predictive of progression to T2DM (Shojaee et al., 2025; Zhou et al., 2024). Non-invasive digital biomarkers derived from wearable devices, such as photoplethysmography, heart rate variability, and sleep patterns, closely approximated CGM profiles with root mean squared error below 15 mg/dL (Kim et al., 2025; Miller & Miller, 2024). Integration of multiple sensor-derived signals further improved predictive accuracy, supporting the potential of multimodal AI models for early detection and risk stratification (Alonso et al., 2023; Kim et al., 2025). These findings emphasize that CGM-derived phenotypes and non-invasive biomarkers can enhance early detection and inform personalized preventive strategies in prediabetes.

3.5 Just-in-Time Adaptive Interventions (JITAI)

Integration of AI with mobile health platforms has supported the delivery of JITAIs for lifestyle modification. Reinforcement learning–driven feedback loops improved adherence to diet and physical activity recommendations, with pilot studies reporting modest reductions in fasting glucose over 3–6 months (Alonso et al., 2023; Shojaee et al., 2025). While these interventions demonstrate translational potential, evidence remains preliminary and requires validation in larger randomized controlled trials (Shojaee et al., 2025).

3.6 Evidence Synthesis

Across predictive modeling studies, AI achieved AUROC values of 0.79–0.91 for early detection and progression risk prediction. Noninvasive biomarker inference studies reported CGM approximation with RMSE <15 mg/dL, while unsupervised clustering uncovered reproducible prediabetes subtypes. However, fewer than half of the studies performed external validation, and very few reported calibration, fairness, or real-world behavioral endpoints (Wu et al., 2024; Gao et al., 2023). Future research must emphasize transparent reporting, reproducibility, and clinically relevant outcomes such as sustained normoglycemia and long-term lifestyle adherence, consistent with ADA recommendations (American Diabetes Association, 2024).

4. DISCUSSION

The integration of artificial intelligence (AI) into prediabetes management represents a transformative advancement in personalized healthcare. Machine learning (ML) and deep learning (DL) models have demonstrated high efficacy in early detection, risk stratification, and personalized intervention for individuals at risk of progressing to type 2 diabetes mellitus (T2DM) (Pukale et al., 2025; Chen et al., 2023; Diabetology &

Metabolic Syndrome Authors, 2025; Shojaee et al., 2025; Alonso et al., 2023). By leveraging high-frequency data from continuous glucose monitoring (CGM), wearable sensors, and electronic health records (EHR), AI uncovers patterns beyond the reach of conventional approaches (Alonso et al., 2023; Kim et al., 2025; Miller & Miller, 2024; Wang et al., 2025).

AI-based early detection methods, including random forest and XGBoost models applied to EHR data, achieved AUROC values of 0.79–0.91, consistently outperforming traditional logistic regression approaches (Pukale et al., 2025; Shojaee et al., 2025). Convolutional neural networks (CNNs) trained on lifestyle data from wearables provided timely warnings of impaired glucose tolerance, demonstrating the feasibility of AI-assisted community screening (Kim et al., 2025; Alonso et al., 2023). Similarly, gradient boosting and LightGBM models trained on longitudinal EHR and metabolic data accurately predicted 3–5-year progression to T2DM, with models integrating lifestyle and dietary features outperforming those using only biochemical markers (Chen et al., 2023; Diabetology & Metabolic Syndrome Authors, 2025; Shojaee et al., 2025).

Noninvasive digital biomarkers, such as virtual CGM derived from wearable devices, offer scalable solutions for population-level screening without invasive procedures (Alonso et al., 2023; Farahmand et al., 2025; Kim et al., 2025; Wang et al., 2025). Deep learning models approximating CGM profiles achieved root mean squared errors below 15 mg/dL in proof-of-concept studies (Kim et al., 2025; Miller & Miller, 2024). Unsupervised clustering and trajectory analysis of glucose patterns revealed heterogeneous prediabetes phenotypes, which may inform more tailored preventive strategies (Shojaee et al., 2025; Zhou et al., 2024).

Integration of AI with mobile health platforms enabled just-in-time adaptive interventions (JITAI) for lifestyle modification. Reinforcement learning–driven feedback improved adherence to diet and physical activity recommendations, producing modest reductions in fasting glucose over 3–6 months (Alonso et al., 2023; Shojaee et al., 2025). Despite this potential, these interventions require validation in larger randomized trials (Shojaee et al., 2025).

Challenges remain in external validation, model calibration, fairness across populations, and demonstration of long-term clinical outcomes (Gao et al., 2023; Wu et al., 2024). The lack of AI-based risk-stratification tools in the 2024 American Diabetes Association guidelines highlights the continued lack of connection between new evidence and clinical practice (American Diabetes Association, 2024). These deficiencies will need stringent validation, standardized reporting structures, and systematic implementation. Effective AI integration can reduce the global burden of type 2 diabetes mellitus (T2DM) by combining information-driven insights with evidence-based clinical interventions, thus enhancing individual patient care.

5. CONCLUSION

The current review explains why artificial intelligence (AI) has significant potential to transform the management of prediabetes. The AI-based approaches allow early identification, accurate risk stratification, identification of specific prediabetes phenotypes, and the provision of contextsensitive and personalized interventions. Moreover, virtual continuous glucose monitoring and just-in-time adaptive interventions increase patient engagement and offer scalable solutions to preventive care.

Although the current evidence is promising, several concerns exist, such as validation in heterogeneous populations, the evaluation of long-term clinical outcomes, and ethical and equity concerns. Future, real-world research that includes multimodal data will be instrumental in achieving the potential of AI in precision prevention. Combining real-world clinical interventions with evidence-based information, AI provides an opportunity to reduce the worldwide incidence of type 2 diabetes and improve personalized care. This review identifies a route to scalable, data-driven prediabetes management by combining AI-driven insights with personalized interventions. These strategies may allow the prevention of type 2 diabetes to be detected earlier, preventive measures to be tailored, and eventually decrease the global burden of the disease.

REFERENCES

Alonso, A., Chen, Y., Zhang, L., et al. (2023). *Reinforcement learning–driven feedback for lifestyle modification in prediabetes*. NPJ Digital Medicine, 6, 93. https://doi.org/10.1038/s41746-023-00956-y

American Diabetes Association. (2024). *Standards of care in diabetes—2024*. Diabetes Care, 47(Suppl. 1), S1–S168. https://doi.org/10.2337/dc24-S002

Basiri, R., & Cheskin, L. J. (2024). *Enhancing individualized nutrition therapy with real-time CGM feedback in prediabetes*. Nutrients, 16(23), 4005. https://doi.org/10.3390/nu16234005

Cell Reports Medicine Editorial Team. (2023). *CGMap: Large-scale characterization of glucose dynamics in non-diabetic adults using continuous monitoring*. Cell Reports Medicine, 4(6), 101266. https://doi.org/10.1016/j.xcrm.2023.101266

Chen, Y., Wang, F., Xu, J., et al. (2023). *Machine learning models integrating dietary indicators improve prediction from prediabetes to T2DM*. Nutrients, 17(6), 947. https://doi.org/10.3390/nu17060947

Diabetology & Metabolic Syndrome Authors. (2025). *ML-based stratification of prediabetes and T2DM progression*. Diabetology & Metabolic Syndrome, 17, 1786. https://doi.org/10.1186/s13098-025-01786-6

Farahmand, E., Rahimi Azghan, R., Chatrudi, N. T., et al. (2025). *AttenGluco: Multimodal transformer-based glucose forecasting on AI-READI dataset*. https://arxiv.org/abs/2502.09919

Gao, M., Zhang, R., & Chen, Z. (2023). *AI-enabled classification of elevated glucose using non-invasive signals*. JMIR AI, 2, e48340. https://doi.org/10.2196/48340

Ji, C., Jiang, T., Liu, L., Zhang, J., & You, L. (2025). *Continuous glucose monitoring combined with artificial intelligence: Redefining the pathway for prediabetes management*. Frontiers in Endocrinology, 16, Article 1571362. https://doi.org/10.3389/fendo.2025.1571362

Kim, J., Lee, H., & Park, S. (2025). *Virtual CGM from life-log data using deep learning*. Scientific Reports, 15, 1367. https://doi.org/10.1038/s41598-025-01367-7

Khalid, M., Rahman, M., & Akter, S. (2025). *Interpretable machine learning model for prediabetes risk prediction in population health surveys*. BMC Public Health, 25, 22419. https://doi.org/10.1186/s12889-025-22419-7

Li, X., Ding, F., Zhang, L., Zhao, S., Hu, Z., Ma, Z., ... Zhao, Y. (2025). *Interpretable machine learning method to predict the risk of pre-diabetes using nationwide cross-sectional data: Evidence from CHNS*. BMC Public Health, 25, 1145. https://doi.org/10.1186/s12889-025-22419-7

Li, X., Hu, J., & Li, T. (2021). *Machine learning for short-term glucose prediction using accelerometer and CGM data*. PLOS ONE, 16(6), e0253125. https://doi.org/10.1371/journal.pone.0253125

Liarakos, A. L. L., Lim, J. Z. M., Leelarathna, L., & Wilmot, E. G. (2024). *Technology use in type 2 diabetes and prediabetes: A narrative review*. Diabetologia, 67, 2059–2074. https://doi.org/10.1007/s00125-024-06203-7

Liu, X., & Zhang, J. (2024). *Continuous glucose monitoring: A transformative approach to prediabetes detection*. Journal of Multidisciplinary Healthcare, 17, 5513–5519. https://doi.org/10.2147/JMDH.S493128

Luo, J., Kumbara, A., Shomali, M., Han, R., Iyer, A., Agarwal, R., & Gao, G. (2025). *A large sensor foundation model pretrained on continuous glucose monitor data for diabetes management* (arXiv:2412.09727). arXiv. https://doi.org/10.48550/arXiv.2412.09727

Miller, E., & Miller, K. (2024). *Detection and intervention: Use of continuous glucose monitoring in the early stages of type 2 diabetes*. Clinical Diabetes, 42(3), 398–407. https://doi.org/10.2337/cd23-0077

Pukale, S., Jagdale, S., Bhandari, A., & Shinde, S. (2025). *AI-powered early detection of diabetes using machine learning on electronic health records*. International Journal of Scientific Research in Science, Engineering and Technology, 12(2), 578–584. https://doi.org/10.32628/IJSRSET25122171

Shojaee, T., Patel, K., Rahman, S., et al. (2025). *Machine learning models for predicting 5-year progression from prediabetes to diabetes: External validation and web tool*. Journal of Medical Internet Research, 27, e73190. https://doi.org/10.2196/73190

Tatli, D., Papapanagiotou, V., Liakos, A., Tsapas, A., & Delopoulos, A. (2024). *Prediabetes detection in unconstrained conditions using wearable sensors* (arXiv:2410.02692). arXiv. https://doi.org/10.48550/arXiv.2410.02692

Wang, S., Liu, Q., & Zhao, Y. (2025). *Predicting interstitial glucose from wearables: Toward digital biomarkers for prediabetes*. Scientific Reports, 15, 14172. https://doi.org/10.1038/s41598-025-14172-z

Wu, Y., Chen, H., & Li, M. (2024). *LSTM networks for hypoglycemia prediction generalize across diabetes subtypes*. JMIR Medical Informatics, 12, e56909. https://doi.org/10.2196/56909

Yuan, L., Wang, Y., Xing, M., Liu, T., & Xiang, D. (2025). *Global research trends in AI-assisted blood glucose management: A bibliometric study*. Frontiers in Endocrinology, 16, 1579640. https://doi.org/10.3389/fendo.2025.1579640

Zhang, J., Zhang, Z., Zhang, G., Ge, X., Sun, R., & Zhai, X. (2023). *Early detection of type 2 diabetes risk: Limitations of current diagnostic criteria*. Frontiers in Endocrinology, 14, 1260623. https://doi.org/10.3389/fendo.2023.1260623

Zhang, K., Qi, Y., Wang, W., Tian, X., Wang, J., Xu, L., & Zhai, X. (2025). *Future horizons in diabetes: Integrating AI and personalized care*. Frontiers in Endocrinology, 16, Article 1583227. https://doi.org/10.3389/fendo.2025.1583227

Zhou, H., Li, J., Wang, X., & Liu, Y. (2024). *Glucose-curve-shape phenotyping predicts metabolic subtypes across glycemic states*. Nature Biomedical Engineering. Advance online publication. https://doi.org/10.1038/s41551-024-01311-6