ISSN: 2579-0730

International Digital Organization for Scientific Research IDOSR JBCP/26/102.1700
IDOSR JOURNAL OF BIOLOGY, CHEMISTRY AND PHARMACY 11(1):15-21, 2026.
https://doi.org/10.59298/IDOSR/JBCP/26/102.1521

Artificial Intelligence and Big Data in Predicting and Managing Obesity-Associated Diabetes

Mwende Muthoni D.

Faculty of Medicine Kampala International University Uganda

ABSTRACT

Artificial intelligence (AI) and big data are reshaping how we understand, predict, and manage obesity-associated type 2 diabetes (T2D). The diabesity phenotype arises from heterogeneous interactions among genes, behaviors, environments, and health-care systems. At scale, routinely collected data electronic health records (EHRs), pharmacy claims, continuous glucose monitoring (CGM), wearables, meal logs, imaging, multi-omics, and social determinants of health (SDOH) capture this complexity but are noisy, incomplete, and biased. Modern machine-learning (ML) methods can transform these substrates into actionable insights: predicting incident T2D and complications; detecting subclinical trajectories; stratifying patients into mechanistic endotypes; recommending individualized nutrition, activity, and pharmacotherapy; and monitoring for relapse or adverse events. Time-series deep learning, survival modeling, graph neural networks, and causal inference frameworks enable robust forecasting and counterfactual reasoning, while reinforcement learning (RL) personalizes dynamic regimens. However, translation hinges on trustworthy data engineering, external validation, calibration, explainability, privacy, and equity. Federated learning and differential privacy protect data; fairness auditing and participatory design mitigate bias; and MLOps governs monitoring, drift detection, and post-deployment updates. Integrating AI into clinical workflows requires human-in-the-loop decision support, interoperable standards, and pragmatic evaluation focused on outcomes that matter to patients and systems. This review synthesizes the data foundations, predictive analytics, digital phenotyping, and decision-support paradigms relevant to diabesity; outlines implementation, safety, and governance requirements; and maps a path toward multimodal, foundation-model–enabled "digital twins" that couple physiology with behavior to modify disease trajectories. Done well, AI augments not replaces clinicians and patients, enabling earlier intervention, precise therapy matching, and durable cardiometabolic risk reduction.

Keywords: machine learning; continuous glucose monitoring; digital phenotypes; federated learning; precision diabetes care

INTRODUCTION

Type 2 diabetes (T2D) linked to obesity often termed "diabesity" is common, costly, and heterogeneous. Despite effective lifestyle, pharmacologic, and surgical tools, many patients experience therapeutic inertia, fragmented care, and preventable complications. Three realities motivate AI-enabled transformation[1–4]. First, the biology is complex: insulin resistance (IR) and β-cell dysfunction emerge from interactions among adipose, hepatic, muscular, islet, vascular, immune, and neural networks shaped by diet, sleep/circadian patterns, physical activity, medications, psychosocial stress, and the built environment[5, 6]. Second, the data now exist to observe this complexity: EHRs record diagnoses, labs, vitals, procedures, medication exposures, and unstructured notes; CGM provides minute-level glucose dynamics; wearables capture steps, heart rate variability, sleep staging, and energy expenditure; nutrition apps reveal timing and composition; imaging quantifies visceral fat and hepatic steatosis; genomics/epigenomics/metabolomics detail molecular states; and SDOH datasets index neighborhood deprivation, food access, pollution, and transportation[7]. Third, traditional guidelines and single-risk-factor thresholds cannot fully exploit these data to tailor care in real time.

AI brings pattern recognition, forecasting, and decision optimization to this landscape. Supervised learning can predict incident T2D, hospitalization, hypoglycemia, or progression to insulin; survival models estimate time-to-event under competing risks; sequence models (temporal convolutional networks, transformers) learn from CGM and medication histories; and graph neural networks capture relationships among patients, clinicians, and codes[8]. Unsupervised and self-supervised learning extract latent structure for digital phenotyping endotypes that explain why one patient responds to a GLP-1 receptor agonist while another benefits more from an SGLT2 inhibitor or thiazolidinedione[8]. Causal inference methods targeted maximum

likelihood estimation, doubly robust learners, instrumental variables, and causal forests help move beyond correlation toward estimating individualized treatment effects. RL and contextual bandits personalize sequential decisions: meal timing, macronutrient distribution, activity prompts, and medication titration.

Yet predictive accuracy is insufficient. For clinical utility, models must be calibrated, externally validated, and explainable enough for safe action. They must handle missingness, label noise, and dataset shift; respect privacy and consent; and avoid amplifying structural bias. Federated learning allows cross-institutional training without centralizing raw data, while differential privacy and secure aggregation mitigate re-identification risk. Deployment requires interoperable standards (e.g., FHIR), clinician-friendly interfaces, and evaluation in pragmatic trials or stepped-wedge implementations that measure HbA1c, weight, cardiovascular/renal outcomes, utilization, and patient-reported outcomes[9, 10]. Post-deployment, MLOps monitors drift, fairness metrics, and safety signals, updating models with governance and auditability. This review proceeds in six parts. Section 1 outlines data foundations: multimodal sources, preprocessing, interoperability, and governance. Section 2 covers predictive analytics for risk, progression, and complications using time-series and graph methods. Section 3 details digital phenotyping and patient stratification to guide therapy. Section 4 focuses on decision support and personalization rules, recommendations, RL, and digital twins. Section 5 addresses implementation, safety, equity, privacy, and regulation, including federated learning and differential privacy. Section 6 looks ahead to multimodal foundation models, synthetic data, and clinician/copilot workflows. We close with a practical conclusion on building trustworthy, equitable AI that augments human expertise to prevent and manage diabesity.

2. Data Foundations for AI in Diabesity: Sources, Engineering, and Governance

Modern diabesity AI depends on multimodal data assembled over time:

EHR/claims: diagnoses (ICD), procedures (CPT), labs (HbA1c, lipids, ALT, eGFR), vitals, medication orders/fill patterns, and clinical notes. Free-text notes and scanned documents carry context (diet, adherence, housing) accessible via natural language processing (NLP)[11].

CGM/time series: minute-level glucose enabling features (time-in-range, glycemic variability, rate-of-change, dawn phenomenon indices) and raw sequences for deep models[12].

Wearables/phones: steps, heart rate, HRV, sleep duration/regularity/staging, circadian alignment, location-derived activity spaces, and app-based nutritional timing[12].

Imaging: DEXA, CT/MRI for visceral/subcutaneous fat and liver fat (MRI-PDFF), and ultrasound elastography for fibrosis[13].

Multi-omics: genomics (polygenic risk scores), epigenomics (methylation), transcriptomics/proteomics/metabolomics/lipidomics to capture pathways (inflammation, mitochondrial, adipokine signaling)[14].

SDOH/environment: neighborhood deprivation indices, food deserts, walkability, air quality, temperature, and transportation.

Engineering challenges. Data are messy: missing not at random, irregular sampling, coding drift, and site heterogeneity. Solutions include: (i) common data models (OMOP, FHIR) and concept mapping; (ii) robust imputation (Bayesian, multiple imputation with chains, deep generative models); (iii) label curation using weak supervision and clinician adjudication; (iv) temporal alignment ("index dates," sliding windows, landmarking) and feature stability selection to prevent leakage; (v) text processing with domain-adapted language models and de-identification; and (vi) edge ingestion for CGM/wearables with consent management[15].

Governance and consent. Transparent data provenance and dynamic consent allow patients to control secondary use. Data minimization and purpose limitation reduce risk. Audit trails document access and transformations. A data-trust or institutional review board oversees linkage of clinical and consumer data, with strong community representation[16].

Quality and drift. Pre-deployment data audits quantify completeness, missingness, and label reliability; post-deployment data drift monitors distribution changes (e.g., new assay platforms, medication guidelines) that can degrade model performance. Synthetic data can aid testing but must be rigorously privacy-screened[16].

Equity from the start. Representation of under-served groups, localization to diet/culture, and capture of language and health-literacy markers reduce downstream bias. Collecting SDOH enables deconfounding and fairness auditing.

3. Predictive Analytics for Risk, Progression, and Complications

Incident T2D prediction. Supervised models trained on EHR/claims (age, BMI, lipids, HbA1c, BP, meds, SDOH) forecast conversion from prediabetes to T2D. Gradient boosting, survival forests, and deep survival neural nets handle censoring and time-varying covariates; calibration (reliability curves, Brier scores) ensures risk probabilities are trustworthy. External validation across health systems and geographies tests transportability[17].

Progression and treatment response. Models forecast HbA1c trajectories, medication intensification, weight change, and likelihood of remission after lifestyle, pharmacotherapy (GLP-1/SGLT2/TZD), or bariatric surgery[18]. Counterfactual frameworks (causal forests, targeted learning) estimate individualized treatment effects, supporting therapy matching. Sequence models ingest longitudinal labs/meds to predict who benefits

from adding an SGLT2 inhibitor versus a GLP-1 RA given comorbid HF/CKD/ASCVD and prior responses[19].

Hypoglycemia and adverse events. CGM-based sequence models (temporal CNNs, LSTMs, transformers) predict near-term hypoglycemia and nocturnal events; joint models including insulin doses and meals offer proactive alerts. For SGLT2 inhibitors, risk screens for euglycemic ketoacidosis incorporate context (illness, fasting, low-carb diets). For TZDs, edema and fracture risk models guide selection[20].

Complications and organ risk. Multi-task learning predicts retinopathy, neuropathy, nephropathy progression, NAFLD/NASH transitions, and cardiovascular events. Imaging-augmented models incorporate CAC scores, liver fat/fibrosis measures. Graph neural networks capture relationships among diagnoses and outcomes; competing-risk survival models estimate probabilities across endpoints over time[21, 22].

Explainability and actionability. Beyond SHAP plots, counterfactual explanation ("what minimal change reduces risk?") enables coaching (e.g., earlier meal window, increased steps, medication add-on). Uncertainty quantification flags cases for human review[23].

Evaluation. Discrimination (AUROC, AUPRC), calibration (ECE), clinical utility (decision curves, NRI), and impact (change in HbA1c/weight/events in trials) form a complete scorecard. Subgroup performance and fairness metrics (equalized odds, demographic parity gaps) are mandatory[24].

4. Digital Phenotyping and Patient Stratification

Population-level averages obscure meaningful heterogeneity. Unsupervised learning (clustering, mixture models, topic models for codes, autoencoders, variational inference) on longitudinal EHR/CGM/wearable data discovers endotypes that share pathophysiology and treatment response[25]. Examples include: (i) predominant hepatic IR with fasting hyperglycemia and NAFLD; (ii) peripheral/muscle IR with postprandial spikes; (iii) β-cell-limited phenotype with low first-phase insulin; (iv) inflammasome-dominant phenotype with elevated CRP and sleep apnea; and (v) circadian misalignment with late eating and short sleep. Trajectory clustering groups patients by weight, HbA1c, or CGM time-in-range evolution[26].

Phenotype validation requires triangulation: biological plausibility (omics, imaging), differential outcomes in retrospective cohorts, and prospective enrichment in trials. Hybrid approaches combine clinician-defined rules with learned embeddings (semi-supervised learning). Phenotype portability is tested across institutions and cultures; label shift is handled by domain adaptation[27].

Therapy matching. Once stable, endotypes inform treatment policies: GLP-1 RA first for hyperphagic/obesity-dominant; SGLT2 inhibition for HF/CKD; TZD for severe IR/NAFLD with low edema risk; chrononutrition and sleep therapy for misalignment; bariatric surgery for high-BMI/poor medication response. RL can learn policy improvement from logged data while respecting safety constraints[28, 29].

N-of-1 learning. Within-person models built from CGM, meal logs, and wearables identify personal triggers (late rice/chapati meals, poor sleep, stress) and optimal countermeasures (fiber-first sequence, post-meal walk, earlier eating window). Bayesian hierarchical models borrow strength across similar patients while preserving individuality[30, 31].

Equity. Digital phenotypes must reflect diverse diets, languages, and schedules. Community co-design ensures clusters are interpretable and culturally relevant, preventing stigmatizing labels and ensuring access to matched interventions.

5. Decision Support, Personalization, and Digital Twins

Clinical decision support (CDS). Rules-plus-ML systems surface timely suggestions: add SGLT2 inhibitor in T2D with CKD; consider GLP-1 RA for ASCVD/obesity; evaluate OSA with STOP-Bang; propose eTRE for late eaters. Human-in-the-loop workflows let clinicians accept, modify, or reject suggestions with rationale capture for continual learning[32].

Recommender systems. Using CGM and meal/activity logs, recommend meals with predicted lower glucose spikes, suggest post-meal walks, or reorder meal sequence (vegetable/protein then starch). Contextual bandits balance exploration/exploitation to learn preferences and glycemic responses without large risks[33].

Reinforcement learning (RL). RL policies personalize insulin titration, metformin add-on timing, or GLP-1 RA dose advancement, subject to safety constraints and clinician oversight. For lifestyle, RL schedules activity prompts and meal-timing nudges based on chronotype and daily energy expenditure. **Safe RL** incorporates control barriers to avoid hypoglycemia or over-restriction[34].

Digital twins. Multimodal patient-specific models simulate metabolic responses to diet patterns (low-carb vs Mediterranean), meal timing, pharmacotherapy combinations, or bariatric surgery. Twins integrate physiological models (glucose–insulin) with data-driven components to forecast HbA1c, weight, and organ outcomes; they support shared decision-making by visualizing trade-offs (e.g., LDL-C vs TG response)[35].

Closing the loop. For non-insulin-treated T2D, near-real-time "closed loop" coaching (CGM-triggered messages) reduces variability and time above range. For basal-bolus regimens, adaptive algorithms propose dose changes with clinician confirmation[36].

Pragmatic evaluation. A/B tests and stepped-wedge trials measure outcome changes and clinician/patient workload. User experience design minimizes alert fatigue and promotes adherence; behavioral economics (commitment devices, timely defaults) enhances engagement[37].

6. Implementation, Safety, Privacy, and Equity

Workflow integration. Embed CDS in EHR with single-click actions, order sets, and smart defaults; surface predictions at teachable moments (e.g., when reviewing CGM). Offer patient-facing summaries in plain language and local languages[38].

MLOps and safety. Monitor models for performance, calibration, data drift, and fairness. Establish incident response for model errors; log predictions, actions, and outcomes with auditability. Use champion–challenger deployments to update safely. Maintain model cards and data sheets documenting intent, training sets, and limitations[39].

Fairness and bias mitigation. Preempt bias via representative sampling, re-weighting, adversarial debiasing, and counterfactual fairness audits. Evaluate subgroup performance (sex, age, ethnicity, socioeconomic status) and mitigate disparities. Avoid using proxies for race as causal drivers; focus on SDOH and access variables to allocate resources equitably[40].

Privacy and security. Apply federated learning and secure aggregation for cross-site training; add differential privacy to updates; use homomorphic encryption for select computations. Enforce least-privilege access and continuous security testing. Provide transparent consent and opt-out pathways for secondary data use[41].

Regulatory and ethical considerations. Classify tools appropriately (risk-based) and align with clinical governance. Distinguish "locked" vs "adaptive" algorithms; define update cadence and re-approval processes. Ensure explainability at the right level: global feature importance for oversight; local rationale for actionability. Adopt participatory design with patients and clinicians; align incentives so AI reinforces—not replaces—relationships.

Equity and access. Design for low-resource settings: SMS-based coaching, offline-capable apps, edge AI on low-cost phones, integration with community health workers, and culturally adapted content. Evaluate affordability and digital literacy; provide loaner devices or clinic-based CGM trials where feasible. Measure patient-important outcomes (energy, sleep, function) alongside biomedical metrics[42, 43].

7. Horizons: Multimodal Foundation Models, Synthetic Data, and Copilot Workflows

Foundation models. Self-supervised pretraining on multimodal health data—notes (NLP), time series (CGM/wearables), images (liver MRI/ultrasound), and codes—produces generalizable representations that fine-tune for tasks (risk, response prediction, summarization). Contrastive learning aligns modalities (e.g., CGM ↔ meals, EHR ↔ imaging)[44].

Generative AI and synthetic data. Diffusion models and generative adversarial networks synthesize realistic CGM traces, EHR sequences, and liver fat images to augment training and stress-test pipelines. Privacy-preserving generation with membership-inference defenses is essential. Synthetic cohorts support power calculations and scenario planning (e.g., medication shortages)[45].

Causal and mechanistic hybrids. Blend structural causal models and physiological simulators with learned embeddings to improve counterfactual validity and extrapolation beyond observed data. Use in silico trials to propose adaptive trial designs focused on high-risk endotypes[46].

Copilot workflows. Large language model (LLM) copilots summarize longitudinal charts, draft patient-friendly plans, explain trade-offs, generate orders from CDS suggestions, and triage inbox messages always with clinician oversight. Patient-facing copilots translate CGM into daily actions, coordinate meal timing with chronotype, and support medication adherence[47].

Edge and ambient computing. On-device inference preserves privacy and reduces latency for CGM-triggered coaching; ambient sensors (home BP scales, smartwatches) feed continuous phenotyping. Interoperability via FHIR APIs enables plug-and-play modules[48].

Evaluation science. Move beyond AUROC to cost-effectiveness, clinician time saved, equity impact, and sustainability metrics. Open benchmarks with diverse populations and robust baselines accelerate progress while preventing hype cycles.

CONCLUSION

AI and big data can convert the complexity of obesity-associated T2D into timely, individualized action. The opportunity spans the care continuum: predict risk and trajectories before complications arise; stratify into mechanistic endotypes that guide therapy; personalize nutrition, activity, pharmacotherapy, and surgical referral; and monitor response to prevent relapse. Realizing this promise requires rigorous data engineering, external validation, calibration, and safety monitoring; governance that protects privacy; and design choices that promote fairness and clinician–patient partnership. Federated and privacy-preserving learning broaden access to diverse data, while MLOps ensures models remain reliable as practice and populations evolve. Looking forward, multimodal foundation models, causal-mechanistic hybrids, and copilot interfaces will enable digital twins that illuminate trade-offs and support shared decisions. The pragmatic mandate is to start simple—high-value CDS embedded in workflows then iterate with measurement and community co-design. Done thoughtfully, AI will help bend the curve on diabesity by delivering earlier, more precise, and more equitable metabolic care.

REFERENCES

1. AbdlWhab, H.M., Al-Saffar, A., Mahdi, O.A., Alameri, R.B.: The impact of insulin resistance and glycaemic control on insulin-like growth factor-1 in patients with type 2 diabetes: a cross-sectional study. Clin. Diabetes Endocrinol. 10, 36 (2024). https://doi.org/10.1186/s40842-024-00202-8
2. Alzaid, F., Fagherazzi, G., Riveline, J.-P., Bahman, F., Al-Rashed, F., Al-Mulla, F., Ahmad, R.: Immune cell–adipose tissue crosstalk in metabolic diseases with a focus on type 1 diabetes. Diabetologia. 68, 1616–1631 (2025). https://doi.org/10.1007/s00125-025-06437-z
3. Argyropoulou, D., Geladas, N.D., Nomikos, T., Paschalis, V.: Exercise and Nutrition Strategies for Combating Sarcopenia and Type 2 Diabetes Mellitus in Older Adults. J. Funct. Morphol. Kinesiol. 7, 48 (2022). https://doi.org/10.3390/jfmk7020048
4. Ashley, L., Robb, K.A., O'Connor, D.B., Platt, R., Price, M., Robinson, O., Travis, E., Lipscombe, L., Ajjan, R., Birch, R.: Increased Breast and Colorectal Cancer Risk in Type 2 Diabetes: Awareness Among Adults With and Without Diabetes and Information Provision on Diabetes Websites. Ann. Behav. Med. Publ. Soc. Behav. Med. 57, 386–398 (2023). https://doi.org/10.1093/abm/kaac068
5. Uranga, R.M., Keller, J.N.: The Complex Interactions Between Obesity, Metabolism and the Brain. Front. Neurosci. 13, (2019). https://doi.org/10.3389/fnins.2019.00513
6. Alum, E.U., Obasi, D.C., Abba, J.N., Aniokete, U.C., Okoroh, P.N., Akwari, A.Ak.: Evolving Paradigms in Nutrition Therapy for Diabetes: From Carbohydrate Counting to Precision Diets. Obes. Med. 100622 (2025). https://doi.org/10.1016/j.obmed.2025.100622
7. Ehrenstein, V., Kharrazi, H., Lehmann, H., Taylor, C.O.: Obtaining Data From Electronic Health Records. In: Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2 [Internet]. Agency for Healthcare Research and Quality (US) (2019)
8. Yang, H., Li, J., Liu, S., Yang, X., Liu, J.: Predicting Risk of Hypoglycemia in Patients With Type 2 Diabetes by Electronic Health Record–Based Machine Learning: Development and Validation. JMIR Med. Inform. 10, e36958 (2022). https://doi.org/10.2196/36958
9. Van Calster, B., Steyerberg, E.W., Wynants, L., van Smeden, M.: There is no such thing as a validated prediction model. BMC Med. 21, 70 (2023). https://doi.org/10.1186/s12916-023-02779-w
10. Chen, R.J., Wang, J.J., Williamson, D.F.K., Chen, T.Y., Lipkova, J., Lu, M.Y., Sahai, S., Mahmood, F.: Algorithm fairness in artificial intelligence for medicine and healthcare. Nat. Biomed. Eng. 7, 719–742 (2023). https://doi.org/10.1038/s41551-023-01056-8
11. Laws, M.B., Michaud, J., Shield, R., McQuade, W., Wilson, I.B.: Comparison of Electronic Health Record–Based and Claims-Based Diabetes Care Quality Measures: Causes of Discrepancies. Health Serv. Res. 53, 2988–3006 (2018). https://doi.org/10.1111/1475-6773.12819
12. Yoo, J.H., Kim, J.H.: Time in Range from Continuous Glucose Monitoring: A Novel Metric for Glycemic Control. Diabetes Metab. J. 44, 828–839 (2020). https://doi.org/10.4093/dmj.2020.0257
13. Grąt, K., Grąt, M., Rowiński, O.: Usefulness of Different Imaging Modalities in Evaluation of Patients with Non-Alcoholic Fatty Liver Disease. Biomedicines. 8, 298 (2020). https://doi.org/10.3390/biomedicines8090298
14. O'Sullivan, S., Qi, L., Zalloua, P.: From omics to AI—mapping the pathogenic pathways in type 2 diabetes. FEBS Lett. n/a,. https://doi.org/10.1002/1873-3468.70115
15. Olaniran, O.R., Alzahrani, A.R.R.: Bayesian Random Forest with Multiple Imputation by Chain Equations for High-Dimensional Missing Data: A Simulation Study. Mathematics. 13, 956 (2025). https://doi.org/10.3390/math13060956
16. Bernardo, B.M.V., Mamede, H.S., Barroso, J.M.P., dos Santos, V.M.P.D.: Data governance & quality management—Innovation and breakthroughs across different fields. J. Innov. Knowl. 9, 100598 (2024). https://doi.org/10.1016/j.jik.2024.100598
17. Hageh, C.A., Henschel, A., Zhou, H., Zubelli, J., Nader, M., Chacar, S., Iakovidou, N., Hatzikirou, H., Abchee, A., O'Sullivan, S., Zalloua, P.A.: Improving T2D machine learning-based prediction accuracy with SNPs and younger age. Comput. Struct. Biotechnol. J. 27, 2772–2781 (2025). https://doi.org/10.1016/j.csbj.2025.06.038
18. Gazda, C.L., Clark, J.D., Lingvay, I., Almandoz, J.P.: Pharmacotherapies for Post-Bariatric Weight Regain: Real-World Comparative Outcomes. Obes. Silver Spring Md. 29, 829–836 (2021). https://doi.org/10.1002/oby.23146
19. Gulati, K., Wijndaele, K., Webb, J., von Arx, L., Seif, M., Jennison, T., Geneidat, A., Wild, R., Wood, R., Khunti, K.: Achievement of HbA1c and weight targets in adults with type 2 diabetes on once weekly injectable glucagon-like peptide-1 receptor agonist therapy in UK primary care: A retrospective, real-world study. Diabetes Obes. Metab. 27, 2086–2095 (2025). https://doi.org/10.1111/dom.16201
20. Fleischer, J., Hansen, T.K., Cichosz, S.L.: Hypoglycemia event prediction from CGM using ensemble learning. Front. Clin. Diabetes Healthc. 3, 1066744 (2022). https://doi.org/10.3389/fcdhc.2022.1066744
21. Ghandian, S., Thapa, R., Garikipati, A., Barnes, G., Green-Saxena, A., Calvert, J., Mao, Q., Das, R.: Machine learning to predict progression of non-alcoholic fatty liver to non-alcoholic steatohepatitis or

fibrosis. JGH Open Open Access J. Gastroenterol. Hepatol. 6, 196–204 (2022). https://doi.org/10.1002/jgh3.12716
22. Veldhuizen, G.P., Lenz, T., Cifci, D., van Treeck, M., Clusmann, J., Chen, Y., Schneider, C.V., Luedde, T., de Leeuw, P.W., El-Armouche, A., Truhn, D., Kather, J.N.: Deep learning can predict cardiovascular events from liver imaging. JHEP Rep. 7, 101427 (2025). https://doi.org/10.1016/j.jhepr.2025.101427
23. Ponce-Bobadilla, A.V., Schmitt, V., Maier, C.S., Mensing, S., Stodtmann, S.: Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. Clin. Transl. Sci. 17, e70056 (2024). https://doi.org/10.1111/cts.70056
24. Alemu, Y.M., Alemu, S.M., Bagheri, N., Wangdi, K., Chateau, D.: Discrimination and calibration performances of non-laboratory-based and laboratory-based cardiovascular risk predictions: a systematic review. Open Heart. 12, e003147 (2025). https://doi.org/10.1136/openhrt-2024-003147
25. Qiu, J., Hu, Y., Li, L., Erzurumluoglu, A.M., Braenne, I., Whitehurst, C., Schmitz, J., Arora, J., Bartholdy, B.A., Gandhi, S., Khoueiry, P., Mueller, S., Noyvert, B., Ding, Z., Jensen, J.N., de Jong, J.: Deep representation learning for clustering longitudinal survival data from electronic health records. Nat. Commun. 16, 2534 (2025). https://doi.org/10.1038/s41467-025-56625-z
26. Jung, I., Koo, D.-J., Lee, W.-Y.: Insulin Resistance, Non-Alcoholic Fatty Liver Disease and Type 2 Diabetes Mellitus: Clinical and Experimental Perspective. Diabetes Metab. J. 48, 327–339 (2024). https://doi.org/10.4093/dmj.2023.0350
27. van Hilten, A., van Rooij, J., BIOS consortium, Ikram, M.A., Niessen, W.J., van Meurs, J.B.J., Roshchupkin, G.V.: Phenotype prediction using biologically interpretable neural networks on multi-cohort multi-omics data. NPJ Syst. Biol. Appl. 10, 81 (2024). https://doi.org/10.1038/s41540-024-00405-w
28. Patel, R., Wadid, M., Makwana, B., Kumar, A., Khadke, S., Bhatti, A., Banker, A., Husami, Z., Labib, S., Venesy, D., Fonarow, G., Kosiborod, M., Al-Kindi, S., Bhatt, D.L., Dani, S., Nohria, A., Butler, J., Ganatra, S.: GLP-1 Receptor Agonists Among Patients With Overweight or Obesity, Diabetes, and HFpEF on SGLT2 Inhibitors. JACC Heart Fail. 12, 1814–1826 (2024). https://doi.org/10.1016/j.jchf.2024.07.006
29. Patoulias, D., Michailidis, T.: SGLT-2 Inhibitor and GLP-1 Receptor Agonist Treatment for Patients with Nonalcoholic Fatty Liver Disease and Type 2 Diabetes Mellitus: Is Their Combination the Optimal Treatment Option? J. Clin. Transl. Hepatol. 10, 574–576 (2022). https://doi.org/10.14218/JCTH.2022.00278
30. Askari, M.R., Rashid, M., Sun, X., Sevil, M., Shahidehpour, A., Kawaji, K., Cinar, A.: Detection of Meals and Physical Activity Events From Free-Living Data of People With Diabetes. J. Diabetes Sci. Technol. 17, 1482–1492 (2022). https://doi.org/10.1177/19322968221102183
31. Askari, M.R., Rashid, M., Sun, X., Sevil, M., Shahidehpour, A., Kawaji, K., Cinar, A.: Meal and Physical Activity Detection from Free-Living Data for Discovering Disturbance Patterns of Glucose Levels in People with Diabetes. BioMedInformatics. 2, 297–317 (2022). https://doi.org/10.3390/biomedinformatics2020019
32. Li, J., Albajrami, O., Zhuo, M., Hawley, C.E., Paik, J.M.: Decision Algorithm for Prescribing SGLT2 Inhibitors and GLP-1 Receptor Agonists for Diabetic Kidney Disease. Clin. J. Am. Soc. Nephrol. CJASN. 15, 1678–1688 (2020). https://doi.org/10.2215/CJN.02690320
33. Zheng, M., Ni, B., Kleinberg, S.: Automated meal detection from continuous glucose monitor data through simulation and explanation. J. Am. Med. Inform. Assoc. JAMIA. 26, 1592–1599 (2019). https://doi.org/10.1093/jamia/ocz159
34. Jafar, A., Kobayati, A., Tsoukas, M.A., Haidar, A.: Personalized insulin dosing using reinforcement learning for high-fat meals and aerobic exercises in type 1 diabetes: a proof-of-concept trial. Nat. Commun. 15, 6585 (2024). https://doi.org/10.1038/s41467-024-50764-5
35. Silfvergren, O., Simonsson, C., Ekstedt, M., Lundberg, P., Gennemark, P., Cedersund, G.: Digital twin predicting diet response before and after long-term fasting. PLoS Comput. Biol. 18, e1010469 (2022). https://doi.org/10.1371/journal.pcbi.1010469
36. Dermawan, D., Kenichi Purbayanto, M.A.: An overview of advancements in closed-loop artificial pancreas system. Heliyon. 8, e11648 (2022). https://doi.org/10.1016/j.heliyon.2022.e11648
37. Herrin, J., Finney Rutten, L.J., Ruddy, K.J., Kroenke, K., Cheville, A.L.: Pragmatic cluster randomized trial to evaluate effectiveness and implementation of EHR-facilitated collaborative symptom control in cancer (E2C2): addendum. Trials. 24, 21 (2023). https://doi.org/10.1186/s13063-022-06983-6
38. Solomon, J., Dauber-Decker, K., Richardson, S., Levy, S., Khan, S., Coleman, B., Persaud, R., Chelico, J., King, D., Spyropoulos, A., McGinn, T.: Integrating Clinical Decision Support Into Electronic Health Record Systems Using a Novel Platform (EvidencePoint): Developmental Study. JMIR Form. Res. 7, e44065 (2023). https://doi.org/10.2196/44065
39. Davis, S.E., Greevy, R.A., Lasko, T.A., Walsh, C.G., Matheny, M.E.: Detection of Calibration Drift in Clinical Prediction Models to Inform Model Updating. J. Biomed. Inform. 112, 103611 (2020). https://doi.org/10.1016/j.jbi.2020.103611
40. Cary, M.P., Zink, A., Wei, S., Olson, A., Yan, M., Senior, R., Bessias, S., Gadhoumi, K., Jean-Pierre, G., Wang, D., Ledbetter, L.S., Economou-Zavlanos, N.J., Obermeyer, Z., Pencina, M.J.: Mitigating Racial and

Ethnic Bias and Advancing Health Equity in the Development, Evaluation, and Deployment of Clinical Algorithms in Healthcare: A Scoping Review and Implications for Policy. Health Aff. Proj. Hope. 42, 1359–1368 (2023). https://doi.org/10.1377/hlthaff.2023.00553

41. Shen, J., Zhao, Y., Huang, S., Ren, Y.: Secure and Flexible Privacy-Preserving Federated Learning Based on Multi-Key Fully Homomorphic Encryption. Electronics. 13, 4478 (2024). https://doi.org/10.3390/electronics13224478
42. Mahmood, H., Mckinstry, B., Luz, S., Fairhurst, K., Nasim, S., Hazir, T.: Community health worker-based mobile health (mHealth) approaches for improving management and caregiver knowledge of common childhood infections: A systematic review. J. Glob. Health. 10, 020438 (2020). https://doi.org/10.7189/jogh.10.020438
43. Alum, E.U.: Optimizing patient education for sustainable self-management in type 2 diabetes. Discov. Public Health. 22, 44 (2025). https://doi.org/10.1186/s12982-025-00445-5
44. Liu, Z., Alavi, A., Li, M., Zhang, X.: Self-Supervised Contrastive Learning for Medical Time Series: A Systematic Review. Sensors. 23, 4221 (2023). https://doi.org/10.3390/s23094221
45. Park, H.Y., Bae, H.-J., Hong, G.-S., Kim, M., Yun, J., Park, S., Chung, W.J., Kim, N.: Realistic High-Resolution Body Computed Tomography Image Synthesis by Using Progressive Growing Generative Adversarial Network: Visual Turing Test. JMIR Med. Inform. 9, e23328 (2021). https://doi.org/10.2196/23328
46. Petersen, M.L., van der Laan, M.J.: Causal Models and Learning from Data. Epidemiol. Camb. Mass. 25, 418–426 (2014). https://doi.org/10.1097/EDE.0000000000000078
47. Gaber, F., Shaik, M., Allega, F., Bilecz, A.J., Busch, F., Goon, K., Franke, V., Akalin, A.: Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis. NPJ Digit. Med. 8, 263 (2025). https://doi.org/10.1038/s41746-025-01684-1
48. Espinoza, J., Xu, N.Y., Nguyen, K.T., Klonoff, D.C.: The Need for Data Standards and Implementation Policies to Integrate CGM Data into the Electronic Health Record. J. Diabetes Sci. Technol. 17, 495–502 (2021). https://doi.org/10.1177/19322968211058148

CITE AS: Mwende Muthoni D. (2026). Artificial Intelligence and Big Data in Predicting and Managing Obesity Associated Diabetes. IDOSR JOURNAL OF BIOLOGY, CHEMISTRY AND PHARMACY 11(1):15-21. https://doi.org/10.59298/IDOSR/JBCP/26/102.1521