OPEN ACCESS Freely available online
Review Article
Vol. 1, Issue 2, July-December 2026 | ISSN: 3139-6267
Journal of Artificial Intelligence and Digital Health

A Systematic Review of Hybrid Intelligent Models for Early Detection of Diabetes Using Electronic Health Records

Umoh Augustine Uduak[1,4*], Joseph Daniel[2,4], Emmanuel Asuquo Ubong[3]

[1]Department of Information Systems, Faculty of Computing, University of Uyo, Uyo, Nigeria
[2]Department of Data Science, Faculty of Computing, University of Uyo, Nigeria
[3]School of Computing and Information Technology, Federal University of Technology, Ikot Abasi, Nigeria
[4]TET Fund Centre of Excellence in Computational Intelligence Research, University of Uyo, Uyo, Nigeria

***Correspondence to:** Umoh Augustine Uduak[1,4], [1]Department of Information Systems, Faculty of Computing, University of Uyo, Uyo, Nigeria; [4]TET Fund Centre of Excellence in Computational Intelligence Research, University of Uyo, Uyo, Nigeria, E-mail: uduakumoh@uniuyo.edu.ng

Received: May 17, 2026; **Manuscript No:** JAID-26-8772; **Editor Assigned:** May 20, 2026; **PreQc No:** JAID-26-8772 (PQ); **Reviewed:** May 28, 2026; **Revised:** July 01, 2026; **Manuscript No:** JAID-26-8772 (R); **Published:** July 20, 2026

Citation: Uduak UA, Daniel J, Ubong EA (2026). A Systematic Review of Hybrid Intelligent Models for Early Detection of Diabetes Using Electronic Health Records. J. Artif. Intell. Digit. Health. Vol.1 Iss.2, July (2026), pp:78-88.

ABSTRACT

Diabetes is a major global health concern due to its rising prevalence and the significant number of undiagnosed cases. Early detection using Electronic Health Records (EHRs) has become a key research focus, supported by advances in machine learning and artificial intelligence. This study presents a systematic review of hybrid intelligent models for early diabetes detection using EHR data from 2015 to 2026. The review analyzes various hybrid approaches, including ensemble learning, deep learning, Fuzzy logic-based, feature selection techniques, and optimization-based models. It also examines commonly used datasets, feature extraction methods, and evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. Findings indicate that hybrid intelligent models generally outperform traditional machine learning methods by improving predictive accuracy and capturing complex nonlinear relationships in clinical data. However, issues such as missing data, lack of standardization, interpretability, and limited external validation remain present. The study highlights the need for explainable AI, federated learning, and multimodal data integration to improve clinical applicability. Overall, this review provides insights into current methodologies and identifies future directions for developing more robust, scalable, and clinically applicable diabetes prediction systems.

Keywords: Diabetes; Prediction; Electronic Health Records (EHRs); Machine Learning; Fuzzy Logic; Early Diagnosis

INTRODUCTION

Diabetes places a heavy burden on patients and health systems because it is common, chronic, and linked to substantial morbidity, disability, and death worldwide [1,2]. Several recent studies describe diabetes as a major global public health challenge and note that its incidence and prevalence continue to rise, especially for type 2 diabetes. [3-7]. This burden is made worse by the fact that many people remain undiagnosed in the early stages. One review report that global diabetes rates have nearly doubled since 1980, reaching 9.3% among adults, and that about half of the 463 million people living with diabetes were unaware of their condition [8]. Other recent papers also stress that undiagnosed diabetes remains common, including in underserved groups with limited access to screening [9-10]. The reason early detection matters are clear: if diabetes is not detected and managed in time, prolonged hyperglycemia can damage multiple organs and lead to serious complications. The cited studies mention kidney failure, nephropathy, stroke, blindness, cardiac events, and other disabling outcomes as consequences of missed or delayed diagnosis [9,11,12]. In both type 2 diabetes and type 1 diabetes, authors argue that earlier identification can support timely intervention and improve outcomes. [4,13]. More broadly, timely detection is considered important for improving patient outcomes, reducing the healthcare burden, delaying disease progression, and enabling targeted prevention measures such as lifestyle changes and closer follow-up [14-17]. Taken together, this background shows why early diabetes detection is not just a technical prediction task but a clinical and public health priority. Because diabetes can remain silent until complications begin, there is a strong rationale for methods that can identify high-risk individuals or probable undiagnosed cases earlier, so that treatment, prevention, and monitoring can start before irreversible harm occurs [3,15,18]

To address these challenges, machine learning (ML) and artificial intelligence (AI) techniques have been widely applied in predictive healthcare analytics [19-20]. In particular, hybrid intelligent models combining ensemble learning, deep learning, feature selection, and optimization methods have demonstrated better performance than single classifiers by improving accuracy, robustness, and generalization in clinical decision-making [21]. Additionally, deep learning approaches are effective in capturing complex nonlinear relationships in clinical data [22]. Despite these advancements, there is still a limited comprehensive understanding of how different hybrid models perform in early diabetes detection using Electronic Health Records (EHRs). Existing studies differ in methodology, evaluation metrics, datasets, and validation strategies, making it difficult to identify the most effective approaches. Challenges such as model interpretability, external validation, data privacy, and clinical applicability also remain unresolved.

This study presents a systematic review of hybrid intelligent models for early diabetes detection using EHRs from 2015 to 2026. It identifies commonly used ML, Fuzzy logic-based and hybrid techniques, evaluates their performance using metrics such as accuracy, precision, recall, F1-score, and AUC, and examines datasets and validation methods. The review also highlights key limitations and research gaps, contributing to the development of more accurate, scalable, and clinically applicable intelligent healthcare systems.

LITERATURE REVIEW

Electronic Health Records are Central to Early Diabetes Detection

Electronic health records are especially important for early diabetes detection because they capture patients' health information across repeated clinical encounters, rather than at a single point in time. This longitudinal view can contain early indicators of disease, including changes in lab values, vital signs, diagnoses, medications, and other clinical signals that appear before diabetes is formally recognized in care. Several papers describe EHRs or EMRs as a core data source for detection and prognostic modeling, and note that these records are useful not only for diabetes prediction itself but also for identifying related clinical events and risks in routine practice. [23-26].

EHRs are also central because they are generated as part of normal healthcare delivery, which makes them practical for scalable early detection. Instead of relying only on dedicated screening studies, researchers can use routinely collected physiological and clinical data to identify high-risk patients and support prevention or earlier follow-up. This is why multiple studies frame EMR- or EHR-based diabetes prediction as a way to enable primary and secondary prevention, improve outcomes, and reduce burden on the healthcare system. [24,27,8]. Another reason EHRs matter is their size and richness. Modern health systems produce large amounts of structured and unstructured data, and this volume is difficult for clinicians to process manually. The cited literature argues that machine learning and deep learning are well suited to this setting because they can learn patterns from large-scale records and help turn raw healthcare data into clinically useful forecasts. In diabetes research specifically, reviews note that many automated detection studies rely mainly on EHR data, and recent papers describe deep analysis of EHRs as opening new opportunities for early detection and intervention [23,29,30,31]. This central role of EHRs also fits the broader move toward data-driven risk stratification in diabetes. Recent work states that AI can identify people at high risk by analyzing EHRs together with other health information, and that large-population data can reveal patterns that emerge before diabetes develops. For a review focused on hybrid intelligent models, this matter because EHRs provide the shared clinical backbone on which different modeling methods can be combined, compared, and translated into real clinical workflows [4,32,33].

Role of Artificial Intelligence, Machine Learning, and Hybrid Intelligent Models

Electronic health records (EHRs) are important for early diabetes detection because they contain patient information collected over multiple clinical visits rather than from a single encounter. This longitudinal data can reveal early signs of diabetes through changes in laboratory results, medications, diagnoses, and vital signs before the disease is formally diagnosed. Studies identify EHRs and EMRs as major sources for diabetes prediction and prognostic modeling, as they support the identification of disease patterns and related clinical risks in routine healthcare practice [23,24,25,26]. Because EHRs are generated during normal healthcare delivery, they also provide a practical and scalable approach for identifying high-risk patients and supporting preventive care without depending solely on dedicated screening programs [24,27,28].

EHRs are equally valuable because of the large amount of structured and unstructured data they contain. The growing volume of healthcare data is difficult to analyze manually, making machine learning and deep learning useful for discovering patterns and generating clinically relevant predictions. Existing studies show that many automated diabetes detection systems rely heavily on EHR data, while recent research highlights the potential of advanced EHR analysis for improving early diagnosis and intervention [23,29,30,31]. This supports the wider shift toward data-driven risk prediction in diabetes care, where AI methods use EHRs and related health information to identify individuals at risk before disease onset. For hybrid intelligent models, EHRs serve as the main clinical foundation that allows different analytical methods to be integrated and applied within real healthcare workflows [4,32,33].

Key Challenges and Research Gaps Motivating A Systematic Review

A major reason for this systematic review is that existing studies still show recurring weaknesses. Earlier research found that many diabetes prediction models were limited in scope, focused on single diabetes types, lacked interpretability, and struggled with complex or incomplete EHR data [34]. Data quality also remains a major challenge, as clinical datasets often contain missing values, and deleting incomplete records can reduce accuracy and introduce bias. Recent studies emphasize the need for better handling of missing data through imputation, NLP, and hybrid approaches that integrate structured and unstructured EHR information [35,36,37]. Another concern is the black-box nature of many machines learning and deep learning models, which limits trust and clinical adoption in early diabetes detection. Current research highlights the importance of explainability tools such as LIME and SHAP to improve transparency and usability [38]. In addition, the literature remains fragmented across datasets, methods, and definitions of "hybrid" models, making comparisons difficult and limiting understanding of which approaches are most effective for true early detection [39]. These issues justify a dedicated systematic review to identify methodological gaps, challenges in handling EHR data, and barriers to clinical implementation.

MATERIALS AND METHODS

This section presents an overview of the methodology used for the literature review, including the study objectives, researchquestions, inclusion and exclusion criteria, and the overall review framework. Subsection (A) describes the procedures followed during the literature review process. Subsection (B) explains the development of the research questions, while Subsection (C) outlines the article selection strategy. Subsection (D) presents the inclusion and exclusion criteria used in the study.

S/N	Author	Method	Result	Evaluation Metric	Type of Dataset	Problem Solved
1	Odigie (2025) [17]	Machine learning models (ensemble & deep learning)	Ensemble and deep learning models showed better predictive performance than single classifiers	Not explicitly stated (reported as high discrimination/ performance)	Electronic Health Records (EHR)	Early prediction of Type 2 Diabetes Mellitus using EHR data
2	Srinivasan (2023) [40]	Hybrid model (Decision Tree, Type-2 Fuzzy Expert System, Adaptive Neuro-Fuzzy Inference System)	Improved classification performance and higher prediction accuracy	Accuracy (model validation reported)	Pima Indians Diabetes Dataset	Early prediction and diagnosis of diabetes mellitus using AI-based classification techniques
3	Albahli (2020) [41]	Hybrid model (K-Means clustering + Random Forest + XGBoost)	Achieved higher predictive performance compared to conventional models	Accuracy = 97.53% (10-fold cross validation)	Not explicitly stated (diabetes clinical dataset)	Early prediction of Type 2 diabetes using hybrid machine learning approach
4	Ramesh and Lakshmanna (2024) [42]	Hybrid deep learning model (O-SBGC-LSTM enhanced with Eurygaster Optimization Algorithm; Graph Convolutional LSTM + fuzzy inference system)	Achieved >98% accuracy and outperformed classical machine learning models	Accuracy (>98%)	Not specified (diabetes-related clinical/health dataset)	Early detection and prediction of diabetes and its complications (especially Type 2 diabetes and related risks like heart disease)
5	Shaheen (2024) [43]	Ensemble deep learning models (Hi-Le: Highway + LeNet; HiTCLe: Highway + LeNet + TCN) with ProWSyn oversampling and SHAP explainability	HiTCLe outperformed individual models; improved diabetes detection and prediction performance	Accuracy (94%), F1-score (94–96%), Precision (94%), Recall (95%), K-Fold Cross Validation	Diabetes Prediction Dataset (imbalanced dataset)	Early detection and prediction of diabetes using ensemble deep learning with class imbalance handling

6	Choubey and Paul (2016) [44]	Hybrid model (Genetic Algorithm for feature selection + Multilayer Perceptron Neural Network for classification)	Improved classification performance and reduced computation cost	Accuracy = 79.13%, ROC = 0.842	Pima Indians Diabetes Dataset	Diabetes diagnosis using feature selection and neural network classification
7	Rahimloo and Jafarian (2016) [45]	Hybrid model (Artificial Neural Network + Logistic Regression combination)	Hybrid model achieved lower error and better reliability than individual models	Error function: ANN = 0.1, Hybrid model = 0.0002	Medical diabetes dataset (not explicitly specified)	Diabetes prediction using hybrid statistical and neural network approach
8	Petrović (2024) [46]	Metaheuristic-optimized RNN (PSO, FA, GA, WOA, SCA) with attention mechanism for ECG anomaly detection	Improved performance; attention mechanism reduced error and enhanced model accuracy	Error reduced from 0.006837 to 0.002486	ECG time-series dataset	Early anomaly detection in cardiovascular signals using optimized deep learning models
9	Sharma et al. (2019) [47]	Hybrid Smart Data Mining + IoT framework (SMDIoT: data mining, IoT, chatbots, semantic analysis, granular computing, biosensors)	Proposed system improves real-time monitoring and decision support for patients; described as more effective and economical than traditional approaches	Not explicitly stated (conceptual framework; no fixed performance metric)	IoT biosensor + healthcare data (conceptual/real-time patient monitoring environment)	Early detection and monitoring of diabetes and cardiovascular diseases using integrated AI-IoT healthcare system
10	Olabanjo (2025) [48]	Attention-enhanced Deep Belief Network (DBN) with ensemble feature selection, GAN-based data augmentation, and hybrid loss (cross-entropy + focal loss)	Outperformed baseline models with very high predictive performance	AUC = 1.00, F1-score = 0.97, Precision = 0.98, Recall = 0.95	Sylhet Diabetes Hospital dataset (symptom + demographic data, imbalanced dataset)	Early prediction of diabetes risk using deep learning with imbalance handling and attention mechanism
11	Mohanty (2023) [49]	Deep learning models (VGG16 + XGBoost hybrid network, DenseNet121 CNN) with image preprocessing and class balancing	DenseNet121 outperformed other model and existing methods for diabetic retinopathy detection	Accuracy: VGG16+XGBoost = 79.50%, DenseNet121 = 97.30%	APTOS 2019 Blindness Detection Kaggle retinal image dataset (imbalanced)	Early detection and classification of diabetic retinopathy using deep learning on retinal images
12	Munadi (2022) [50]	Deep learning CNN models (ShuffleNet + MobileNetV2) with decision fusion for thermal image classification	Decision fusion improved performance and achieved perfect classification on dataset	Accuracy = 100%	Plantar thermogram (thermal imagingdataset for diabetic foot ulcers)	Early detection and classification of diabetic foot ulcers using thermal imaging and deep learning

13	Nahiduzzaman (2021) [51]	Hybrid CNN + SVD feature extraction with Extreme Learning Machine (ELM) classifier (with image preprocessing using CLAHE and Ben Graham method)	Outperformed existing methods for DR detection in both binary and multiclass classification	Accuracy: 99.73% (binary), 98.09% (APTOS-2019), 96.26% (Messidor-2); Recall up to 100%	APTOS-2019 and Messidor-2 retinal image datasets	Early detection and grading of diabetic retinopathy using hybrid deep learning and feature reduction approach
14	Dolo (2022) [52]	Federated Learning model (Federated Averaging with Differentially Private Stochastic Gradient Descent – DPSGDFedAvg)	Achieved acceptable prediction performance while preserving strong privacy protection	Accuracy = 60%–70%	Pima Indians Diabetes Dataset	Early detection of Type 2 diabetes with privacy-preserving machine learning approach
15	Bilal (2022) [53]	Two-stage deep learning model (U-Net for OD & BV segmentation + InceptionV3-based CNN-SVD hybrid for feature extraction and classification with data augmentation)	Achieved state-of-the-art DR detection performance across multiple datasets	Accuracy = 97.92% (EyePACS-1), 94.59% (Messidor-2), 93.52% (DIARETDB0)	Retinal image datasets (EyePACS-1, Messidor-2, DIARETDB0)	Automated early detection and classification of diabetic retinopathy using deep learning
16	Jayathilaka (2025) [54]	Hybrid deep learning CNN framework with attention mechanisms and feature fusion for multi-disease retinal classification (DR, AMD, Glaucoma, RVO)	High and consistent classification performance across multiple eye diseases; demonstrated clinical applicability	Accuracy: DR = 93%, AMD = 92%, Glaucoma = 94%, RVO = 94%	Retinal fundus image datasets (RFMiD, IDRiD, APTOS + clinical Sri Lankan dataset)	Early detection and classification of multiple retinal diseases using deep learning for blindness prevention
17	Moya-Albor (2024) [55]	Deep learning model using Knowledge Distillation (Inception-v3 teacher-student framework) with combined KL divergence + categorical cross-entropy loss	Proposed model outperformed baseline models in DR lesion detection and classification	Training accuracy = 99.01%, Validation accuracy = 97.30%	Fundus image datasets (imbalanced and balanced datasets)	Early detection and classification of diabetic retinopathy lesions using knowledge distillation and deep learning

18	Zavala-Díaz (2026) [56]	Hybrid system (rule-based methods + information extraction + machine learning classifiers including Random Forest, SVM, Gradient Boosting)	Gradient Boosting and hybrid rule + ML approach achieved best performance in classifying clinical notes	Accuracy up to 0.999 (Gradient Boosting); average performance 0.966	Clinical and oncology notes dataset	Automated identification of diabetic patients using clinical note classification system
19	Devi (2026) [57]	Hybrid machine learning models (RF, CNN, GNN, LSTM, Logistic Regression, Naive Bayes; best: RF + GNN) using sensor-based physiological data	RF + GNN achieved best predictive performance among tested models	Accuracy = 93%, also evaluated using Precision, Recall, F1-score	Sensor-based physiological dataset (glucose, temperature, pulse rate from wearable devices)	Early prediction of diabetes risk using sensor-driven hybrid machine learning approach
20	Iftikhar (2026) [58]	Hybrid ensemble machine learning framework (Logistic Regression, Neural Networks, Decision Trees, SVM, Random Forest + SMOTE balancing)	Ensemble model outperformed individual classifiers in CKD prediction	Accuracy = 97.71%, Sensitivity = 99.84%, Specificity = 97.19%, F1-score = 98.19%, Brier score = 1.43%	Clinical CKD dataset (case-control study from District Buner, Pakistan)	Early prediction and diagnosis of Chronic Kidney Disease using ensemble machine learning system
21	Lad and Joshi (2026) [59]	Hybrid deep learning model (DCRNet: Dilated Convolution + LSTM architecture)	Outperformed state-of-the-art models in blood glucose prediction for both simulated and real patients	RMSE (simulated: 3.42–17.73 mg/dL), MAE (simulated: 2.11–11.78 mg/dL); RMSE (real: 12.57–34.41 mg/dL), MAE (real: 7.9–25.5 mg/dL)	UVA/Padova T1D simulator dataset and OhioT1DM real patient dataset	Short-term blood glucose level forecasting for Type 1 diabetes management
22	Mekale (2026) [60]	Hybrid framework (Graph Convolutional Network + Tabu Search optimization for feature/instance selection)	Outperformed existing studies in predictive performance for disease prognosis	Accuracy = 98.22%, Precision = 97.37%, Recall = 96.55%	Clinical dataset (patient-based graph structured data)	Prognosis and classification of hyperglucagonemia using graph-based deep learning and optimization techniques
23	Riaz (2026) [61]	Fuzzy logic-based computational model combining six readability metrics with fuzzy rules for unified readability scoring	Government websites were easier to read (6.46 ± 1.73), while non-government sites were more difficult (14.40 ± 2.33)	Readability score (mean ± SD)	Online diabetes-related web content (government and non-government websites)	Prediction of readability level of diabetes information to improve accessibility andpatient understanding

Table 1: Tabulated Literature on Early Detection of Diabetes

Research Questions Steps

Preparation Steps

The preparation stage involves identifying the major concepts of the study and defining the scope of the systematic review. This phase focuses on understanding how hybrid intelligent models are applied to Electronic Health Records (EHRs) for the early detection of diabetes and the challenges associated with their implementation.

RQ1: What types of hybrid intelligent models are used for early diabetes detection using Electronic Health Records?

RQ2: How do hybrid intelligent models improve diabetes prediction accuracy compared to traditional machine learning approaches?

RQ3: What types of EHR data and feature extraction techniques are commonly used in diabetes prediction studies?

RQ4: What challenges are associated with using EHR data for early diabetes detection, particularly regarding missing, incomplete, or unstructured records?

RQ5: Which evaluation metrics are most frequently used to assess the performance of hybrid intelligent models for diabetes prediction?

RQ6: What explainability or interpretability techniques are integrated into hybrid intelligent models to support clinical decision-making?

RQ7: What are the major research gaps and future directions identified in studies on hybrid intelligent models for early diabetes detection?

Article Selection Strategy

Execution Step

The execution stage involves implementing the planned research procedures, including data collection and application of the study design.

Search Strategy

Data collection was conducted using selective keyword-based searches across articles published between 2017 and 2025. Relevant studies were retrieved from research databases such as Google Scholar, IEEE Xplore, Elsevier, Springer, ScienceDirect, and PubMed. The searches focused mainly on article titles, abstracts, and keywords related to hybrid intelligent systems, diabetes prediction, machine learning, deep learning, and Electronic Health Records (EHRs).

The search process was guided by the research questions, while Boolean operators such as "AND" and "OR" were applied to refine and narrow the search results. Example search terms included: "hybrid intelligent models," "diabetes prediction," "early diabetes detection," "machine learning for diabetes," "deep learning," and "Electronic Health Records (EHRs)."

Digital Library Resources

Table 2 summarizes the digital library resources used for data collection, with emphasis on English-language publications obtained through keyword-based searches. The review covered studies published between January 2015 and December 2026 and included peer-reviewed journal articles, conference papers, academic books, review articles, and other scholarly materials. Major digital platforms consulted during this study included Google Scholar, Research Gate and Systematic Scholar supported by the Google search engine.

Language	Approach	Search Engine	Duration	Type	Digital library source	Number of Articles
English	Keywords, Titles	Google	2015- 2026	scholar articles, scientific journals, e-books, conferences, online workshops	Google Scholar	110
					Research Gate	150
					Systematic Scholar	120
					Total	380

Table 2: Resource Libraries

Search Terms

Search terms are combinations of keywords and phrases used in digital databases, search engines, and academic repositories to retrieve relevant literature related to a specific research area. These terms represent the core concepts of the study and guide the search process in identifying relevant studies on hybrid intelligent models, diabetes prediction, and Electronic Health Records (EHRs).

- (Hybrid Intelligent Models OR Hybrid AI Systems) AND (Diabetes Prediction OR Early Diabetes Detection) AND (Electronic Health Records OR EHR Data) AND (Machine Learning OR Deep Learning)
- (Machine Learning OR Deep Learning OR Ensemble Learning) AND (Diabetes Risk Prediction OR Diabetes Diagnosis) AND (Hybrid Models OR Integrated AI Systems) AND (Electronic Health Records OR Clinical Data)

- (Artificial Intelligence in Healthcare OR Hybrid Learning Systems) AND (Diabetes Detection OR Glucose Risk Prediction) AND (EHR Data OR Medical Records) AND (Data Imputation OR Missing Data Handling)
- (Explainable AI OR Interpretable Machine Learning) AND (Diabetes Prediction OR Clinical Decision Support Systems) AND (Hybrid Intelligent Models OR Ensemble Methods) AND (Healthcare Data Analytics OR Electronic Health Records)

Study Selection Based on Inclusion and Exclusion Criteria

The PRISMA flow diagram illustrates the systematic selection of studies used in this review. A total of 380 records were initially identified from databases, with no additional records obtained from registers, websites, organisations, or citation searching as in Figure 1. During the identification stage, 50 duplicate records were removed, leaving 330 records for screening. In the screening phase, all 330 records were assessed based on their titles and abstracts, and 180 records were excluded due to irrelevance to the study focus on hybrid intelligent models and EHR-based diabetes prediction. A total of 150 reports were then sought for retrieval, while 10 reports could not be accessed. The remaining 140 full-text reports were assessed for eligibility, of which 60 were excluded for not meeting the inclusion criteria. Finally, 80 studies satisfied all eligibility requirements and were included in the systematic review for qualitative synthesis and analysis.

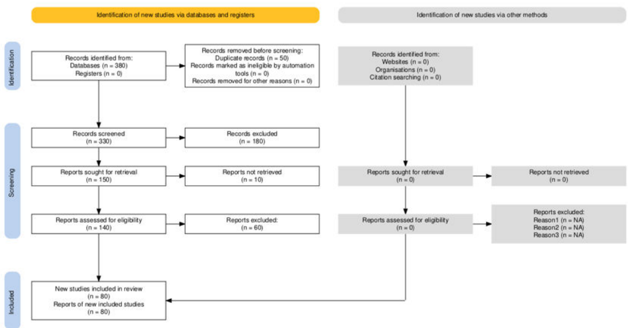

Figure 1: Prisma flow diagram of study selection

RESULTS AND DISCUSSION

RQ1

What types of hybrid intelligent models are used for early diabetes detection using Electronic Health Records?

This systematic review highlights several important findings regarding hybrid intelligent models for early diabetes detection using Electronic Health Records (EHRs). The study shows that commonly applied hybrid intelligent models include ensemble learning approaches such as Random Forest, Gradient Boosting, and XGBoost combinations, deep learning-based hybrids such as CNN–LSTM and DBN with attention mechanisms, feature selection integrated models such as GA–MLP and LASSO-based classifiers, optimization-driven hybrids involving techniques such as Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and Tabu Search combined with classifiers, as well as privacy-preserving approaches such as Federated Learning integrated with Differential Privacy. Among these approaches, ensemble learning and deep learning hybrid models are most frequently reported due to their strong predictive performance and ability to model complex clinical data relationships.

RQ2

How do hybrid intelligent models improve diabetes prediction accuracy compared to traditional machine learning approaches?

Hybrid intelligent models enhance diabetes prediction accuracy compared to traditional machine learning methods by combining the strengths of multiple algorithms, which improves generalization and reduces overfitting and bias. These models also strengthen feature representation and selection, enabling more effective learning of complex nonlinear relationships within clinical datasets. As a result, hybrid approaches consistently outperform single classifiers such as Support Vector Machines, K-Nearest Neighbors, and Logistic Regression. Reported improvements include high accuracy levels (typically between 90% and 99%), improved F1-score and recall values, and strong performance on imbalanced datasets.

RQ3

What types of EHR data and feature extraction techniques are commonly used in diabetes prediction studies?

The most commonly used EHR data types include demographic information such as age, gender, and BMI, clinical laboratory results such as glucose levels, HbA1c, and lipid profiles, patient medical history, lifestyle-related factors, and in some studies, sensor-based physiological data as illustrated in Figure 2. Frequently applied feature extraction and transformation techniques include feature selection methods such as Genetic Algorithms (GA), LASSO regression, Random Forest importance ranking, and mutual information, as well as dimensionality reduction techniques such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD). In addition, deep learning techniques such as Convolutional Neural Networks (CNNs) and Deep Belief Networks (DBNs) support automated feature learning, while data balancing techniques such as SMOTE and GAN-based augmentation improve dataset representation.

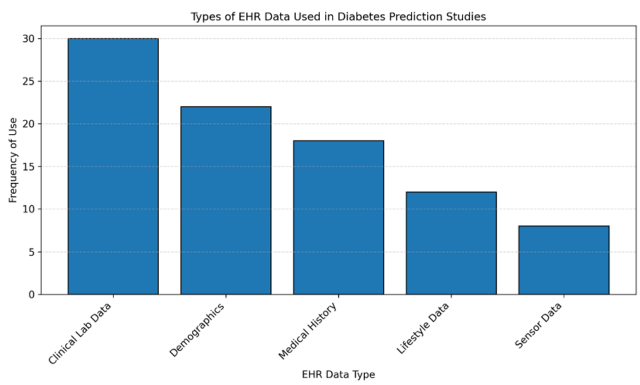

Figure 2: Types of HER Data

RQ4

What challenges are associated with using EHR data for early diabetes detection, particularly regarding missing, incomplete, or unstructured records?

Several challenges are associated with the use of EHR data in diabetes prediction. These include missing or incomplete records, inconsistent and noisy data entries, high-dimensional feature spaces, and class imbalance between diabetic and non-diabetic cases. Additional considerations include data privacy and security requirements, lack of standardized data formats across healthcare institutions, and limited availability of external validation datasets, all of which influence model generalization and real-world deployment.

RQ5

Which evaluation metrics are most frequently used to assess the performance of hybrid intelligent models for diabetes prediction?

The most frequently used evaluation metrics in hybrid diabetes prediction studies include accuracy, which is the most widely reported metric, followed by precision, recall (sensitivity), F1-score, AUC-ROC, and specificity. In time-series glucose prediction studies, additional performance measures such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are also applied to evaluate forecasting accuracy

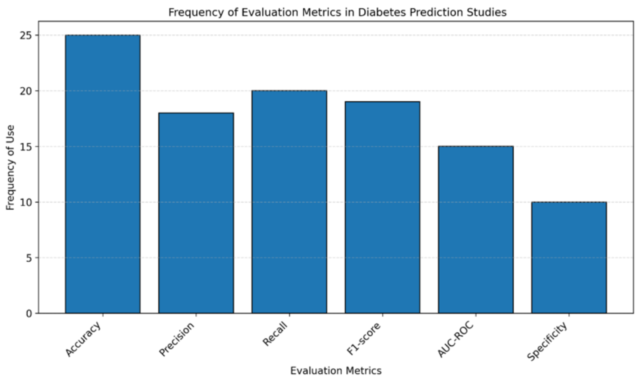

Figure 3: Evaluation Metrics used in Diabetes prediction

RQ6

What explainability or interpretability techniques are integrated into hybrid intelligent models to support clinical decision-making?

Explainability and interpretability techniques commonly integrated into hybrid models include SHAP (Shapley Additive Explanations), feature importance analysis from tree-based models such as Random Forest and XGBoost, rule-based decision systems, and attention mechanisms in deep learning architectures. These techniques improve transparency and support clinical understanding of model predictions, thereby enhancing trust and usability in healthcare environments.

RQ7

What are the major research gaps and future directions identified in studies on hybrid intelligent models for early diabetes detection?

Key research gaps identified in the literature include limited external validation across diverse populations, reduced interpretability in some deep learning-based models, limited real-world clinical deployment, inconsistent evaluation methodologies, and concerns related to data privacy in centralized learning systems. Future research directions emphasize federated learning for privacy-preserving prediction, explainable AI (XAI)-driven hybrid models, integration of multimodal data sources such as EHR, sensors, and imaging, development of real-time clinical decision support systems, and the establishment of standardized benchmarking frameworks to improve consistency and reproducibility across studies.

LIMITATIONS REPORTED IN THE STUDIES

Across the reviewed literature, several recurring limitations were identified in hybrid intelligent models for early diabetes detection using Electronic Health Records (EHRs). A major limitation is the presence of missing, incomplete, and noisy clinical data, which affects model reliability and reduces predictive consistency. Many studies also report issues related to high-dimensional datasets, which increase computational complexity and may lead to overfitting if not properly addressed. Another key limitation is the lack of standardized datasets and evaluation protocols, as most studies use different datasets such as Pima Indians, OhioT1DM, or proprietary EHRs, making direct comparison difficult. In addition, many models suffer from limited external validation, as they are often tested on single datasets without cross-population generalization. Interpretability remains a major concern, especially in deep learning and ensemble-based hybrid systems, where model decisions are often not transparent to clinicians. Furthermore, privacy and security constraints in EHR data limit data sharing and centralized model training, affecting scalability. Finally, several studies highlight limited real-world clinical deployment, indicating a gap between experimental performance and practical healthcare implementation.

CONCLUSIONS

This systematic review highlights the growing importance of hybrid intelligent models in improving early diabetes detection using Electronic Health Records (EHRs). The findings show that hybrid approaches particularly those combining ensemble learning, deep learning, optimization techniques, and feature selection methods consistently outperform traditional machine learning models in terms of accuracy, robustness, and generalization. EHR data plays a central role in enabling early detection due to its longitudinal and multi-dimensional nature, which captures early indicators of diabetes progression. The review also confirms that commonly used evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC are essential for performance assessment across studies. However, challenges such as data quality issues, lack of interpretability, inconsistent evaluation standards, and limited clinical deployment remain significant barriers. Addressing these challenges is essential for translating hybrid intelligent models into reliable clinical decision-support systems.

ACKNOWLEDGEMENTS

The authors sincerely appreciate TETFund for providing a supportive research environment through the TETFund Centre of Excellence in Computational Intelligence Research at the University of Uyo. They also gratefully acknowledge the support of the University of Uyo management.

REFERENCES

1. Emerole CG, Osim EE, Emerole CO. The Burden and Management of Chronic Kidney Disease in Nigeria: The Role of the Physiologist in a Growing Public Health Challenge. Int. J. of Multidisciplinary and Current Research. 2026;14(2).
2. Noori M, Sattarpour R, Moradi-Lakeh M, Baradaran HR, Tehrani-Banihashemi A, Yarahmadi R, et al. The burden of type 2 diabetes attributable to air pollution in Iran at national and provincial level: a review of GBD 2021 findings. Journal of Diabetes & Metabolic Disorders. 2026;25(1):80.
3. Olwendo A. Comparison of machine learning methods for the prediction of type 2 diabetes in primary care setting using EHR data. Journal of Agriculture, Science and Technology. 2024;23(1):24-36.
4. Mohsen F, Al-Absi HR, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. npj Digital Medicine. 2023;6(1):197.
5. Ye B, Shen Z, Ji A, Jiao Y, Niu Y, Zuo K, Hu P, Li W. Advances in Electronic Health Record Analysis for Diabetes Risk Prediction: A Machine Learning-Enhanced Framework. In2024 5th International Conference on Artificial Intelligence and Computer Engineering (ICAICE) 2024 (pp. 830-839). IEEE.
6. Voskergian D, Bakir-Gungor B, Yousef M. Engineering novel features for diabetes complication prediction using synthetic electronic health records. Frontiers in Genetics. 2025;16:1451290.
7. Farnoosh R, Abnoosian K, Isewid RA. Two machine-learning hybrid models for predicting type 2 diabetes mellitus. Journal of Medical Signals & Sensors. 2025;15(4):11.
8. Pratiwi NK. In-Depth Exploration and Comparison of Machine Learning Performances for Early-Stage Diabetes Risk Prediction. Jurnal Infotel. 2024;16(2):353-68.
9. Ansari GA, Shafi S, Ansari MD, Shadab A. Advanced supervised machine learning methods for precise diabetes mellitus prediction using feature selection. Frontiers in Medicine. 2025;12:1620268.
10. Sulfikar A, Rachmawaty R, Sjattar EL. Emerging digital health approaches for the detection of undiagnosed type 2 diabetes mellitus in underserved populations: A scoping review. Nursing Practice Today. 2026.
11. Hassan F, Shaheen ME. Predicting diabetes from health-based streaming data using social media, machine learning and stream processing technologies. International Journal of Engineering Research and Technology. 2020;13(8):1957-67.
12. Sabitha E, Durgadevi M. Improving the diabetes diagnosis prediction rate using data preprocessing, data augmentation and recursive feature elimination method. International Journal of Advanced Computer Science and Applications. 2022;13(9).
13. Mittal R, Weiss MB, Rendon A, Shafazand S, Lemos JR, Hirani K. Harnessing machine learning, a subset of artificial intelligence, for early detection and diagnosis of type 1 diabetes: a systematic review. International journal of molecular sciences. 2025 ;26(9):3935.
14. Pinjari SS & Vani N (2023). An In-depth Comparison of Diabetes Detection Using Machine Learning. International Journal for Research in Applied Science and Engineering Technology.
15. Gupta P, Sindhu R. Diabetes prediction using machine learning. Journal of Electrical Systems. 2024;20(7s):2244-57.
16. Fan Y. Diabetes diagnosis using a hybrid CNN LSTM MLP ensemble. Scientific Reports. 2025;15(1):26765.
17. Odigie OC, Ayankoya FY, Kuyoro SO, Abiodun AG. A Systematic Review of Machine Learning Models for Predicting Type 2 Diabetes Mellitus Using Electronic Health Records. Asian Journal of Electrical Sciences. 2025;14(2):28-34.
18. Giri S. AI-Driven predictive models for early detection of diabetes: A review study. Int. J. Comput. Sci. Mob. Comput.. 2024;13(9):24-33.
19. Owolabi A, Osman AB. Artificial Intelligence and Big Data Analytics for Early Disease Prediction in Healthcare Systems.
20. Goraya A. The Role of Artificial Intelligence and Machine Learning in Modern Healthcare Systems. Smart Chips for Smart Devices: VLSI Design for Next-Generation IoT Solutions. 2026 :103-21.
21. Elabd E, Hamouda HM, Ali MM, Hamid AS, Fouad Y. Air quality index AQI classification based on hybrid particle swarm and grey wolf optimization with ensemble machine learning model. Scientific Reports. 2026.
22. Udoinyang U, Eyoh I, Umoh U, Ubong E, Ene E, Chibuikem O, Akponome B. OPTIMAL DEEP NEURAL NETWORK PARAMETERS FOR POWER LOSS MINIMIZATION ANALYTICS. Nigerian Journal of Technology. 2025;44(4):647-59.
23. Swinckels L, Bennis FC, Ziesemer KA, Scheerman JF, Bijwaard H, de Keijzer A, et al. The use of deep learning and machine learning on longitudinal electronic health records for the early detection and prevention of diseases: scoping review. Journal of medical Internet research. 2024;26:e48320.
24. Perveen S, Shahbaz M, Ansari MS, Keshavjee K, Guergachi A. A hybrid approach for modeling type 2 diabetes mellitus progression. Frontiers in Genetics. 2020;10:1076..
25. Gulhane RA, Gupta SR. An Early Disease Prediction and Risk Analysis of Diabetic Mellitus using Electronic Medical Records. InIOP Conference Series: Materials Science and Engineering 2021 (Vol. 1085, No. 1, p. 012023). IOP Publishing.
26. Jin Y, Li F, Vimalananda VG, Yu H. Automatic detection of hypoglycemic events from the electronic health record notes of diabetes patients: empirical study. JMIR medical informatics. 2019;7(4):e14340.
27. Alhassan Z, Watson M, Budgen D, Alshammari R, Alessa A, Al Moubayed N. Improving current glycated hemoglobin prediction in adults: Use of machine learning algorithms with electronic health records. JMIR Medical Informatics. 2021;9(5):e25237.
28. Zafari H, Li J, Zulkernine FH, Kosowan L, Singer A. Predictive Modeling of Diabetes using EMR Data. InHEALTHINF 2022 (pp. 211-218).
29. Zanelli S, Ammi M, Hallab M, El Yacoubi MA. Diabetes detection and management through photoplethysmographic and electrocardiographic signals analysis: A systematic review. Sensors. 2022;22(13):4890.
30. Chaki J, Ganesh ST, Cidham SK, Theertan SA. Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review. Journal of king saud university-computer and information sciences. 2022;34(6):3204-25.
31. Pang H, Zhou L, Dong Y, Chen P, Gu D, Lyu T, Zhang H. Electronic health records-based data-driven diabetes knowledge unveiling and risk prognosis. arXiv preprint arXiv:2412.03961. 2024.
32. Ghosh K, Chandra S, Ghosh S, Ghosh US. Artificial intelligence in personalized medicine for diabetes mellitus: a narrative review. Cureus. 2025;17(9).
33. Fadhlullah AF, Widiyaningtyas T. Comparative Analysis of Decision Tree and Random Forest Algorithms for Diabetes Prediction. JTAM (Jurnal Teori dan Aplikasi Matematika). 2024;8(4):1121-32.
34. Ali R, Hussain J, Siddiqi MH, Hussain M, Lee S. H2RM: a hybrid rough set reasoning model for prediction and management of diabetes mellitus. Sensors. 2015;15(7):15921-51.

35. Mallika C, Selvamuthukumaran S. Hybrid Online Model for Predicting Diabetes Mellitus. Intelligent Automation & Soft Computing. 2022;31(3).
36. Masconi KL, Matsha TE, Erasmus RT, Kengne AP. Effects of different missing data imputation techniques on the performance of undiagnosed diabetes risk prediction models in a mixed-ancestry population of South Africa. PloS one. 2015;10(9):e0139210.
37. Hegde H, Shimpi N, Glurich I, Acharya A. Tobacco use status from clinical notes using Natural Language Processing and rule based algorithm. Technology and Health Care. 2018;26(3):445-56.
38. Zaferani N, Afrash MR, Moulaei K. Predicting and classifying type 2 diabetes using a transparent ensemble model combining random forest, k-nearest neighbor, and neural networks. Scientific Reports. 2025.
39. Woldaregay AZ, Årsand E, Botsis T, Albers D, Mamykina L, Hartvigsen G. Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. Journal of medical Internet research. 2019 ;21(5):e11030.
40. Shrinivasan L, Verma R, Nandeesh MD. Early prediction of diabetes diagnosis using hybrid classification techniques. IAES International Journal of Artificial Intelligence. 2023;12(3): 1139-48.
41. Albahli S. Type 2 machine learning: an effective hybrid prediction model for early type 2 diabetes detection. Journal of Medical Imaging and Health Informatics. 2020 ;10(5):1069-75.
42. Ramesh B, Lakshmanna K. A novel early detection and prevention of coronary heart disease framework using hybrid deep learning model and neural fuzzy inference system. IEEe Access. 2024;12:26683-95.
43. Shaheen I, Javaid N, Alrajeh N, Asim Y, Aslam S. Hi-le and hitcle: Ensemble learning approaches for early diabetes detection using deep learning and explainable artificial intelligence. IEEE access. 2024;12:66516-38.
44. Choubey DK, Paul S. GA_MLP NN: a hybrid intelligent system for diabetes disease diagnosis. International Journal of Intelligent Systems and Applications. 2016 ;8(1):49-59.
45. Rahimloo P, Jafarian A. Prediction of diabetes by using artificial neural network, logistic regression statistical model and combination of them. Bulletin de la Société Royale des Sciences de Liège. 2016;85:1148-64.
46. Petrovic A, Jovanovic L, Venkatachalam K, Zivkovic M, Bacanin N, Budimirovic N. Anomaly detection in electrocardiogram signals using metaheuristic optimized time-series classification with attention incorporated models. International Journal of Hybrid Intelligent Systems. 2024;20(2):159-83.
47. Sarwar A, Ali M, Manhas J, Sharma V. Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. International Journal of Information Technology. 2020;12(2):419-28.
48. Olabanjo O, Wusu A, Olabanjo O, Asokere M, Afisi O, Akinnuwesi B. A novel deep learning model for early diabetes risk prediction using attention-enhanced deep belief networks with highly imbalanced data. International Journal of Information Technology. 2025;17(4):1933-55.
49. Mohanty C, Mahapatra S, Acharya B, Kokkoras F, Gerogiannis VC, Karamitsos I, Kanavos A. Using deep learning architectures for detection and classification of diabetic retinopathy. Sensors. 2023;23(12):5726.
50. Munadi K, Saddami K, Oktiana M, Roslidar R, Muchtar K, Melinda M, et al. A deep learning method for early detection of diabetic foot using decision fusion and thermal images. Applied Sciences. 2022;12(15):7524.
51. Nahiduzzaman M, Islam MR, Islam SR, Goni MO, Anower MS, Kwak KS. Hybrid CNN-SVD based prominent feature extraction and selection for grading diabetic retinopathy using extreme learning machine algorithm. IEEE Access. 2021 ;9:152261-74.
52. Dolo B, Loukil F, Boukadi K. Early detection of diabetes mellitus using differentially private SGD in federated learning. In2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA) 2022 (pp. 1-8). IEEE.
53. Bilal A, Zhu L, Deng A, Lu H, Wu N. AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning. Symmetry. 2022;14(7):1427.
54. Jayathilaka S, Balaruban D, Kumanayake I, Elladeniya A, Wijendra D, Krishara J, De Silva M. An Integrated Deep Learning Framework for Early Detection of Vision Disorders. In2025 International Research Conference on Smart Computing and Systems Engineering (SCSE) 2025 (pp. 1-6). IEEE.
55. Moya-Albor E, Lopez-Figueroa A, Jacome-Herrera S, Renza D, Brieva J. Computer-aided diagnosis of diabetic retinopathy lesions based on knowledge distillation in fundus images. Mathematics. 2024;12(16):2543.
56. Zavala-Díaz J, Olivares-Rojas JC, Gutiérrez-Gnecchi JA, Téllez-Anguiano AC, Alcaraz-Chávez JE, Reyes-Archundia E. Clinical notes classification system for automated identification of diabetic patients: hybrid approach integrating rules, information extraction and machine learning. Journal of Intelligent & Fuzzy Systems. 2026;50(2):573-83.
57. Devi VK, Umamaheswari E, Utkarsh M, Gao XZ, Bhat M. Sensor-driven artificial intelligence for early diabetes prediction using machine learning models. InAdaptive AI in Sensor Informatics 2026 (pp. 181-202). Elsevier.
58. Iftikhar H, Hashem AF, Mohamud LA, Al-Moisheer AS, Medina RI, López-Gonzales JL. An intelligent ensemble machine learning model for early detection of chronic kidney disease in aging populations. Scientific Reports. 2026.
59. Lad K, Joshi M. DCRNet: Hybrid Deep Learning Architecture for Forecasting of Blood Glucose. Journal of Electronics, Electromedical Engineering, and Medical Informatics. 2026;8(1): 53-68.
60. Mekale SS, Chakraborty M, Mukherjee C. TS-GCN: a hybrid intelligent framework combining Tabu search clustering and graph convolutional network for prognosis of hyperglucagonemia. International Journal of Information Technology. 2026;18(1): 263-9.
61. Riaz S, Naveed MS, Sanjrani AA, Shoaib A. A novel model to predict the readability of online diabetes information based on fuzzy logic. PeerJ Computer Science. 2026 ;12:e3535.