# Diabetes detection and prediction through a multimodal artificial intelligence framework

**Gururaj N. Kulkarni, Kelapati**
Department of Computer Science and Application, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, India

**Article Info**

***Article history:***

Received Jun 5, 2024
Revised Oct 22, 2024
Accepted Oct 30, 2024

***Keywords:***

Artificial intelligence
Diabetes detection
Healthcare
Machine learning
Multimodal framework
Prediction

**ABSTRACT**

Diabetes detection and prediction are crucial in modern healthcare, requiring advanced methodologies and comprehensive data analysis. This study aims to review the application of multi-parameters and artificial intelligence (AI) techniques in diabetes assessment, identify existing research limitations and gaps, and propose a novel multimodal framework for enhanced detection and prediction. The research objectives include evaluating current AI methodologies, analyzing multi-parameter integration, and addressing challenges in early detection and model evaluation. The study utilizes a systematic review approach, analyzing recent literature on AI-based diabetes detection and prediction, focusing on diverse data sources and machine learning (ML) techniques. Findings reveal a significant lack of integration of diverse data sources, limited focus on early detection strategies, and challenges in model evaluation. The study concludes with a proposed innovative framework for more accurate and personalized diabetes detection, contributing to the advancement of diabetes research and highlighting the potential of AI-driven healthcare interventions. This research underscores the importance of comprehensive data integration and robust evaluation methods in enhancing diabetes detection and prediction.



***Corresponding Author:***

Gururaj N. Kulkarni
Department of Computer Science and Application, Shri Jagdishprasad Jhabarmal Tibrewala University
Vidyanagri, Jhunjhunu Bisau Road, Chudela, District-Jhunjhunu, Rajasthan-333010, India
Email: gururajnkulkarni40@gmail.com

## 1. INTRODUCTION

Diabetes, a chronic metabolic disorder, affects millions of people worldwide, making it a significant public health concern [1]. It is characterized by elevated blood glucose levels resulting from either insufficient insulin production or the body's inability to use insulin effectively. There are primarily two types of diabetes: Type 1 and Type 2 as presented in Figure 1. Type 1 diabetes, often diagnosed in childhood or adolescence, occurs when the immune system mistakenly attacks and destroys insulin-producing cells in the pancreas, leading to insulin deficiency [2]. In contrast, Type 2 diabetes, more common in adults, develops when the body becomes resistant to insulin or doesn't produce enough insulin to maintain normal glucose levels [3]. The prevalence of diabetes has been steadily rising, attributed to factors such as sedentary lifestyles, unhealthy diets, obesity, and genetic predisposition [4]. According to global health statistics, diabetes affects approximately 10% of the adult population worldwide, with Type 2 diabetes accounting for the majority of cases [5]. The symptoms of diabetes can vary but often include increased thirst and hunger, frequent urination, unexplained weight loss, fatigue, blurred vision, and slow wound healing [6]. The causes of diabetes are multifactorial, involving a complex interplay of genetic, environmental, and lifestyle factors. Genetic predisposition, obesity, lack of physical activity, poor diet, and aging are among the key risk factors associated with the development of diabetes [7].

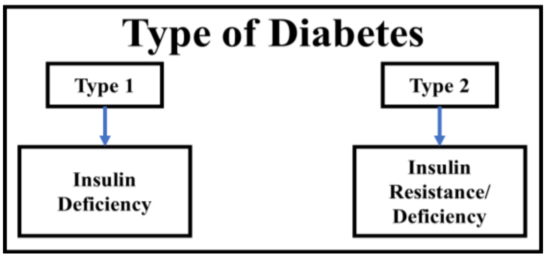


Figure 1. Types of diabetes

One of the most concerning aspects of diabetes is its potential complications, which can affect various organs and systems in the body. Long-term complications may include cardiovascular diseases, kidney damage, nerve damage (neuropathy), eye damage (retinopathy), foot ulcers, and increased risk of infections [8]. Diabetes also significantly contributes to morbidity and mortality rates globally, with a notable proportion of deaths attributed to diabetes-related complications [9]. Diagnosing diabetes involves various tests and assessments to measure blood glucose levels, insulin levels, and other relevant parameters [10]. Common diagnostic tests include fasting blood glucose (FBG) test [11], oral glucose tolerance test (OGTT) [12], glycated hemoglobin (HbA1c) test [13], and random blood glucose test [14]. These tests help healthcare professionals determine whether an individual has diabetes, prediabetes, or normal glucose metabolism. In recent years, machine learning (ML) [15], [16] and deep learning (DL) [17], [18] approaches have emerged as valuable tools for detecting and predicting diabetes. These computational techniques analyze large datasets containing demographic information (such as age, sex) [19], health parameters (glucose level, heart rate, and blood pressure) [20], biochemical parameters (glucose level, lipid profile) [21], physical activity parameters (exercise frequency, intensity) [22], ophthalmic parameters (retinal changes) [23], sleep patterns [24], and stress levels [25]. By integrating and analyzing these diverse data sources, ML and DL models can identify patterns, trends, and risk factors associated with diabetes onset, progression, and complications. However, most ML and DL research in diabetes detection has predominantly focused on demographic information, health parameters, and biochemical parameters called as longitudinal data [26], [27]. There is limited emphasis on incorporating sleep patterns, stress levels, and physical activity parameters into predictive models, despite their known influence on diabetes risk and management. Furthermore, the availability of multimodal datasets containing comprehensive information across these domains remains limited, posing challenges to developing robust and generalized ML/DL models for diabetes detection and prediction [28].

Hence, this work aims to address several key issues in the field of diabetes detection and prediction as presented above. To achieve this goal, a comprehensive review is presented, focusing on the diverse parameters utilized in previous studies for diabetes detection and prediction over the years. By analyzing the methodologies and findings of these studies, valuable insights into the strengths and limitations of existing approaches are gained. Furthermore, this review identifies various datasets that have been used in diabetes research, along with their respective findings. Understanding the outcomes of these studies provides a basis for evaluating the effectiveness and applicability of different datasets in diabetes detection and prediction tasks. However, despite the progress made in diabetes research, there are notable limitations and challenges that need to be addressed. These limitations encompass issues such as the limited integration of multimodal data, insufficient emphasis on early detection strategies, and gaps in incorporating factors like sleep patterns, stress levels, and physical activity parameters into predictive models. Identifying these gaps, issues, and challenges is crucial for advancing the field of diabetes research and developing more accurate and robust predictive models. To bridge these gaps and overcome the identified challenges, this work proposes a multimodal framework for diabetes detection and prediction. This framework integrates diverse data sources, including demographic information, health parameters, biochemical markers, physical activity patterns, ophthalmic parameters, sleep patterns, and stress levels. By leveraging a multimodal approach, this framework aims to improve the accuracy, sensitivity, and specificity of diabetes detection and prediction models. Additionally, it seeks to enhance early identification strategies and provide a more comprehensive understanding of diabetes risk factors and progression pathways.

In section 2 of this manuscript, an extensive literature survey is conducted, focusing on the detection and prediction of diabetes using artificial intelligence (AI) techniques. Moving on to section 3, a novel multimodal framework for the detection and prediction of diabetes is presented. Finally, in section 4, the manuscript concludes by summarizing the overall work.

## 2. LITERATURE SURVEY

In this section an extensive literature survey is conducted, focusing on the detection and prediction of diabetes using AI techniques. The survey encompasses discussions on the utilization of various multi-parameters such as health parameters, sleep patterns, stress levels, physical activity parameters, biochemical markers, and ophthalmic parameters in diabetes research. Through this survey, the results and findings from previous studies are identified, providing insights into the effectiveness and applicability of AI-based approaches in diabetes detection and prediction tasks. Following the discussion of results and findings, the limitations identified from the literature survey are presented. These limitations highlight areas of improvement and challenges faced by existing research methodologies and approaches in diabetes detection and prediction using AI techniques. Understanding these limitations is crucial for developing more robust and accurate predictive models in future research works. Furthermore, in this section, the survey delves into the identification and discussion of gaps, issues, and challenges observed in the literature survey. These discussions shed light on the gaps in current research, issues faced in implementing AI techniques for diabetes detection, and challenges that need to be addressed for advancing the field of AI-based diabetes research.

- Detection and prediction of diabetics using AI techniques considering multi-parameters

This section delves into the detection and prediction of diabetes using AI by considering multi-parameters. These multi-parameters include demographic information (age, sex), health parameters (blood pressure, heart rate), sleep patterns (quality, duration), stress, (physiological response, cortisol level), physical activity parameter (exercise frequency, intensity), biochemical parameters (glucose level, lipid profiles) and ophthalmic parameters (retinal changes). Bodapati *et al.* [29], this work focussed on evaluating the ophthalmic parameter (i.e., diabetic retinopathy (DR)). In this work, they considered the Kaggle APTOS 2019 [30] dataset (image dataset) for evaluating the ophthalmic parameter. This work utilized various convolutional network (CovNet) approaches to extract the optimal characteristics from the dataset. These characteristics were used for training deep neural network (DNN) for identifying DR. This approach achieved accuracy of 0.9741 for identifying DR and accuracy of 0.817 for severity extent prediction. Cordeiro *et al.* [31], this work focused on electro cardi gram (ECG) signals (biomedical signals) which had various parameters other than ECG, i.e., heart-rate, blood glucose concentration level, weight, gender, height and age. This work focused on hyperglycaemia, a type of diabetes where sugar levels are very much high in comparison to a normal person. Further, they faced a challenge regarding the dataset, i.e., most of the normal ECG signals datasets such as Physionet Bank [32] do not include the glucose concentration levels. To address this issue, they generated a novel dataset which contains ECG signal having glucose concentration levels. They considered 1,119 individuals of both non-hyperglycaemias and hyperglycaemia individuals. They classified an individual as hyperglycaemias if the blood glucose concentration level was 100 mg/dL. Further, this work considered 10-fold cross validation (CV) DNN approach which achieved area-under the curve (AUC) of 0.9453.

Hervella *et al.* [33], considered ophthalmic parameter for detecting whether an individual is non-diabetic or diabetic. For this work, for training their approach, they utilized a multimodal dataset [34] which consisted of images related to retina. Further, testing was done on various DR image dataset, i.e., EyePACS-Kaggle [35], Messidor [36], Messidor-2 [37], and IDRiD [38]. These various datasets include image dataset which focus on the DR. This work main aim was to grade the DR images and detect a diabetic individual. Hence, they presented a model called as multi-modal image-encoding (MIE), which used the CovNet encoder for extracting features from the images. Further, the results for extraction of features using IDRiD showed that the MIE achieved accuracy of 0.6117 and DR AUC of 0.9190. Also, when extracting features from Messidor showed that MIE achieved accuracy of 0.6605 and DR AUC of 0.8439, respectively. When grading the IDRiD and Messidor, the MIE achieved accuracy of 0.6505 and 0.7255 respectively. Kulkarni *et al.* [39], focused on detecting Type-2 diabetics using ECG signals. In this work, they also considered other parameters other than ECG, i.e., blood pressure, body-mass-index, sex, age, heart, and other biochemical parameters. For this study they have generated their own dataset collected from Nagpur. The dataset included 1,262 subjects. Further, for detecting whether the subject is diabetic or non-diabetic, this work utilized various ML and DL approaches, i.e., long short-term memory (LSTM), convolutional neural network (CNN), and XGBoost (XGB). Also, this work proposed an algorithm called as DiaBeats which achieved accuracy of 0.968.

Yu *et al.* [40], considered various health parameters and demographic information for detecting Type-2 diabetes in patients. In this work they considered the research on early-life and aging-trends and effects (RELATE) dataset [41]. For detecting diabetes, they proposed an approach called diabetes-mellitus network (DMNet). The DMNet approach consisted of SMOTE-Tomek for feature extraction, Tandem-LSTM (T-LSTM) for capturing risk factors related to diabetes and finally MLP was used for classification. The results showed that DMNet achieved accuracy of 94%. Their main aim was to select optimal features for classification and handle class imbalance. Botella-Serrano *et al.* [42], this work studied how the sleep parameter can help to control Type-1 diabetes. This work continuously evaluated the glucose levels of the 25 patients for a duration of 14 days. Also, in this work, the biochemical parameter was also considered for the study which was continuously monitored using a fit band. The logistic regression (LR) was used for evaluating the data gathered from the patients. The results showed that patients who had good sleep improved their glycaemic control. Rodriguez-Leon *et al.* [43], presented a multimodal dataset which consisted of health parameters, ophthalmic parameters, biochemical parameters and demographic information for detecting Type-1 mellitus diabetes. Pai *et al.* [44], they presented a multimodal dataset which consisted of physical activity parameters and biochemical parameters for providing a better lifestyle for Type-2 diabetic patients.

Theis *et al.* [45], has considered using health parameters for evaluation of diabetics. This work considered medical-information-mart for intensive-care (MIMIC III) [46] dataset for evaluation of their work. This dataset consisted of various health related parameters which included clinical tests, clinical measurements, demographics, billing, pharmacotherapy, intervention methods, and medical information of an individual. This work utilized a DNN approach for prediction of death of patients suffering from diabetics. The results from this work showed that this approach achieved area-under the receiver-operating-characteristics (AUROC) of 0.873. Naseem *et al.* [47], this work considered using health parameters for predicting diabetes in individuals. This work considered a Kaggle dataset provided by National-Institute of Diabetes called as PIMA [48]. This dataset consisted of various parameters which include insulin level, blood-pressure, pregnancies, age, body-mass-indexes, glucose level, and skin thickness. For prediction, this work considered LR, support vector machine (SVM), LSTM, CNN and artificial neural network (ANN), and recurrent neural network (RNN). The findings showed that among all these approaches, the RNN achieved best accuracy, i.e., 0.81. Also, it was seen that ANN achieved better recall, i.e., 0.56.

Ahmed *et al.* [49], predicted diabetics using SVM and ANN. Further, they proposed an approach called as fused-method for diabetic prediction (FMDA). This work considered a dataset from University of California-Irvine (UCI) [50]. The dataset consisted of various health parameters, biochemical parameters, and demographic data. The results showed that SVM, ANN, and FMDA achieved accuracy of 89.10%, 92.31%, and 94.87% during testing. Jia *et al.* [51], considered to classify diabetes using PIMA dataset which consists of various health parameters. For classification, this work proposed an approach called as probabilistic-ensemble classification approach for diabetic individual (PE-DIM). Their main aim was to handle the class imbalance issue. This work was built by combining local median-based (LM) gaussian Naïve-Bayes (NB) (LMeGNB) and k-means synthetic-minority-over-sampling technique (SMOTE). The PE-DIM achieved accuracy of 0.9453 for the PIMA dataset during testing. Further, they evaluated their work on other datasets, i.e., on type-2 dataset called as RSMH [52] and Tabriz [53]. The RSMH dataset consisted of demographic, biochemical and health parameters, whereas the Tabriz dataset consisted of demographic and biochemical history and health parameters. The PE-DIM achieved better results for both the dataset, i.e., AUC of 0.9917 and 0.9982 for RSMH and Tabriz dataset, respectively.

Yadav *et al.* [54], focus was on detecting the hyper-parameters which can help prediction of diabetes by selecting optimal features. For evaluation of this work, they used they PIMA, and two Indian datasets which were focused on mellitus diabetes called as FHD and ADRC datasets. The FHD dataset consisted of 400 records having 7 attributes where 150 were diabetic patients and 250 were non-diabetic patients. Further, the ADRC dataset consisted of 583 records having 7 attributes where 167 were diabetic patients and 416 non-diabetic patients. The PIMA dataset in this work consisted of 768 records consisting of 9 attributes where 268 were diabetic and 500 were non-diabetic. All the datasets had health parameters and biochemical parameters. The FHD and ADRC dataset additionally consisted of demographic information. For balancing the features, the SMOTE was utilized. For selecting optimal features, they utilized wrapper and filtering method. Further for tuning the hyperparameter, Bayesian optimizing approach, grid and random search was utilized. Finally, they proposed an algorithm called as grey-wolf-optimization (GWO). Further, they used other ML approaches like random forest (RF), multi-layer-perceptron (MLP), decision-tree (DT), k-nearest-neighbour (KNN), NB, LR, ANN, SVM, and CNN. The results showed that the proposed approach using GWO achieved 0.98, 0.973, and 0.962 F-score for PIMA, ADRC and FHD datasets respectively. Himi *et al.* [55], considered demographic information, biochemical parameters, sleep patterns, and stress levels for predicting 12 kind of diseases which also included diabetes, called as medical AI (MedAi). This work constructed a dataset which consisted of various demographic information, sleeping patterns, biochemical parameters and stress level using data collected from a smart watch from 150 subjects. This work considered evaluating the dataset using various ML approaches which included gradient boosting (GB), support-vector-regression (SVR), SVM, KNN, RF, XGB, LR, and LSTM.

Annuzzi *et al.* [56], considered evaluating their model by considering health parameters and biochemical parameters. This work considered two datasets, i.e., DirectNet [57], and Ai4pg [58] dataset. In this work, they proposed an ML approach, i.e., feed-forward-neural-network (FFNN). The evaluation was done on inter-subjective analysis and intra-subjective analysis. The root mean squared error (RMSE) metric was considered for evaluation. Also, the tests were conducted for 15, 30, 45, and 60 mins and the RMSE was evaluated. The FFNN approach achieved RMSE of 4.14, 8.30, 13.72 and 16.69 for 15, 30, 45 and 60 mins respectively for DirectNet dataset and for Ai4pg, the results were evaluated by considering various features and its RMSE score. Their main aim was to select optimal features for predicting the glucose level of a patient. De Paola *et al.* [59], focused on evaluating the role of physical activity for prediction of glucose levels in Type-2 diabetic patients. This work was a simulation approach where they focused on how the interleukin-6 when injected can help to increase the insulin level in an individual body during a physical activity. The findings show that this approach established a base for using physical activity parameters for detection and prediction of diabetes.

## 3. FINDINGS

The results and findings from the above literature survey are given in Table 1 in Appendix. In Table 1, the datasets, methods, parameters used, and the findings are discussed in brief.

### 3.1. Limitations

The literature survey reveals several limitations in the existing research related to the detection and early identification of diabetes using multimodal datasets and a comprehensive set of parameters. Firstly, there is a notable scarcity of studies that have extensively utilized multimodal datasets encompassing various data types such as health parameters, sleep patterns, stress levels, physical activity, biochemical markers, and ophthalmic parameters. The lack of integration and analysis of multiple data sources affects the development of holistic and accurate diagnostic models for diabetes. This limitation suggests a gap in research that could potentially yield more robust and comprehensive approaches for diabetes detection. Secondly, within the limited studies that do consider multimodal datasets, there is a gap of work specifically focused on the early identification of diabetes. Early detection is crucial for timely intervention and management, yet the current literature lacks sufficient emphasis on leveraging diverse data modalities for early diagnosis. This gap is significant as early identification can lead to better outcomes for individuals at risk of developing diabetes or those in the early stages of the disease.

Moreover, the literature also highlights a lack of comprehensive studies that simultaneously consider demographics, health parameters, biochemical parameters, physical activity, and ophthalmic parameters for identifying diabetes. Diabetes is a complex metabolic disorder influenced by various factors, including age, gender, lifestyle, and physiological markers. However, the existing research often overlooks the synergistic effects of these diverse parameters, leading to incomplete assessments of diabetes risk and progression. The limitations identified in the literature survey underscore the need for more extensive and integrated research efforts that leverage multimodal datasets, prioritize early identification strategies, and consider a comprehensive range of demographic and physiological parameters. Addressing these gaps can significantly enhance the accuracy, timeliness, and effectiveness of diabetes detection and management approaches.

### 3.2. Gaps, issues, and challenges

The gaps, issues and challenges identified from the above section is as;

a. Limited use of multimodal datasets
   - Insufficient exploration and utilization of multimodal datasets combining diverse data types such as health parameters, sleep patterns, stress levels, physical activity, biochemical markers, and ophthalmic parameters.
   - Lack of integrated analysis across multiple data sources, affecting the development of comprehensive diagnostic models for diabetes.

b. Negligence of early identification
   - Scarcity of research focused specifically on early identification of diabetes, which is crucial for timely intervention and improved patient outcomes.
   - Missed opportunities to leverage diverse data modalities for early detection and intervention strategies.

c. Inadequate consideration of demographics and other parameters
   - Limited studies that simultaneously consider demographics (age, gender), health parameters, biochemical parameters, physical activity, and ophthalmic parameters for diabetes identification.
   - Failure to address the synergistic effects of these diverse parameters, leading to incomplete assessments of diabetes risk and progression.

d. Need for advanced analytical techniques
   - Lack of adoption of advanced analytical techniques such as ML algorithms, DL models, and AI for processing and interpreting multimodal data in diabetes research.
   - Challenges in developing accurate and scalable predictive models due to data complexity, heterogeneity, and volume.

Addressing these gaps, issues, and challenges is essential for advancing the field of diabetes detection, promoting early identification strategies, and improving patient outcomes through personalized and data-driven healthcare interventions. Hence, in the next section, a possible solution to address these gaps, issues and challenges is presented.

## 4. MULTIMODAL FRAMEWORK

In this section, a novel multimodal framework is introduced for the detection and prediction of diabetes, as presented in Figure 2. The framework begins by incorporating a comprehensive multimodal dataset that encompasses a wide range of parameters crucial for diabetes assessment. These parameters include demographic information such as age, gender, and ethnicity, health metrics like glucose levels, blood pressure, and heart rate, biochemical markers such as lipid profiles and insulin levels, physical activity data including exercise frequency and intensity, ophthalmic parameters like retinal changes, and data on sleep patterns and stress levels. Once the multimodal dataset is structured, a ML approach will be employed for the task of detecting and predicting diabetics. ML algorithms are particularly well-suited for handling complex and multidimensional data, making them an ideal choice for analyzing the diverse parameters present in the multimodal dataset. Techniques such as supervised learning, ensemble methods, and DL may be utilized within the ML framework to extract patterns, correlations, and predictive insights from the data. Following the application of ML algorithms, the framework proceeds to evaluate the performance of the diabetes detection and prediction models. This evaluation is conducted using a range of performance metrics, including accuracy, precision, recall, and F-score, among others. These metrics provide quantitative measures of the model's effectiveness in correctly identifying diabetic individuals, capturing true positive and true negative rates, and minimizing false positives and false negatives. Additionally, the evaluation process can consider other relevant parameters and aspects, such as model robustness, scalability, interpretability, and computational efficiency. By examining these metrics comprehensively, the framework aims to assess the overall effectiveness and reliability of the ML-based approach in diabetes detection and prediction tasks. The presented multimodal framework integrates diverse data sources, leverages advanced ML techniques, and employs rigorous evaluation measures to enhance the accuracy, reliability, and clinical utility of diabetes detection and prediction models. This holistic approach not only addresses the complexity of diabetes as a multifactorial condition but also provides a systematic and data-driven methodology for improving healthcare outcomes in diabetic care and management.

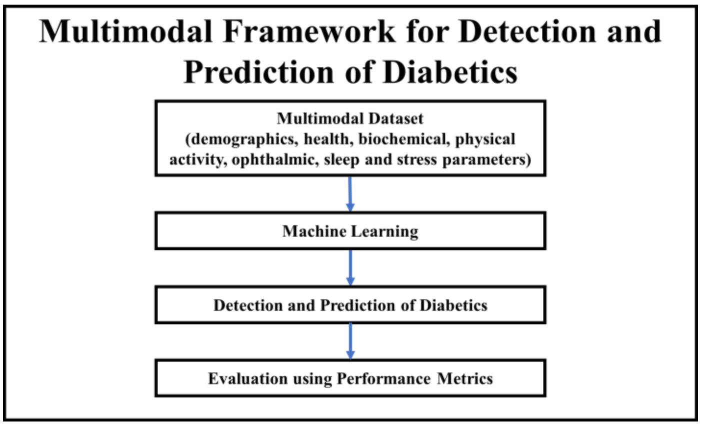


Figure 2. Multimodal framework for detection and prediction of diabetics

## 5. CONCLUSION

This work has delved into the intricate landscape of diabetes detection and prediction, particularly focusing on the utilization of AI techniques and multimodal datasets. Diabetes is a critical health issue with rising prevalence worldwide, making the development of accurate detection and prediction methodologies essential for improving patient outcomes. The study's comprehensive review of existing literature provided valuable insights into the diverse parameters and methodologies employed in previous diabetes assessment studies. The importance of considering multi-parameters, such as demographic, health, biochemical, physical activity, ophthalmic, sleep, and stress factors, was highlighted, along with the significance of advanced AI techniques like ML and DL in analyzing complex datasets and improving predictive accuracy. The limitations and challenges identified in existing research, such as the need for better integration of multimodal data, early detection strategies, and comprehensive model evaluation, underscored critical areas for improvement. To address these gaps, the study proposed a novel multimodal framework for diabetes detection and prediction, emphasizing the integration of diverse data types, application of ML approaches, and thorough performance evaluation using key metrics like accuracy, precision, recall, and F-score. However, the proposed framework offers a more holistic and personalized approach, which is crucial for enhancing the accuracy and efficacy of diabetes management. Future research should explore enhancing the multimodal framework by incorporating additional data sources, such as genetic markers, environmental factors, and patient lifestyle information. Integrating real-time data streams from wearable devices, mobile apps, and electronic health records (EHRs) could enable continuous monitoring and proactive intervention strategies. Furthermore, exploring advanced ML and DL techniques, including ensemble methods, transfer learning, and reinforcement learning, could further improve the predictive power and generalizability of diabetes detection models. Continued research and collaboration in this field can pave the way for improved patient outcomes, enhanced clinical decision-making, and better management of diabetes and related metabolic disorders.

## APPENDIX

Table 1. Results and findings

| Reference | Datasets | Methods | Parameters | Findings |
|---|---|---|---|---|
| [19] | Kaggle APTOS 2019 | CovNet approaches for feature extraction | Ophthalmic parameter | Achieved accuracy of 0.9741 for identifying DR and accuracy of 0.817 for severity extent prediction. |
| [31] | Novel dataset with ECG signals and glucose concentration levels | 10-fold cross-validation DNN approach | Biochemical parameter, demographics, health parameter | Achieved AUC of 0.9453 for classifying hyperglycaemia individuals. |
| [33] | Multimodal dataset for training | Multi-modal image-encoding | Ophthalmic parameter | MIE achieved accuracy of 0.6117 and DR AUC of 0.9190 when extracting features from IDRiD. Achieved accuracy of 0.6605 and DR AUC of 0.8439 when extracting features from Messidor. |
| [39] | Dataset collected from Nagpur | LSTM, CNN, XGBoost | Biochemical parameter, demographics, health parameter | Proposed DiaBeats algorithm achieved accuracy of 0.968 for detecting Type-2 diabetics. |
| [40] | RELATE dataset | DMNet approach | Health parameters, demographics | DMNet achieved 94% accuracy in detecting Type-2 diabetes. |
| [42] | Glucose and sleep monitoring dataset | LR | Sleep parameters, biochemical parameters | Patients with good sleep showed improved glycemic control. |
| [43] | Multimodal dataset with health, ophthalmic, biochemical information | - | Health parameters, ophthalmic parameters, biochemical parameters | Aimed at detecting Type-1 mellitus diabetes. |
| [44] | Multimodal dataset with physical activity and biochemical information | - | Physical activity parameters, biochemical parameters | Aimed at providing a better lifestyle for Type-2 diabetic patients. |
| [45] | MIMIC III dataset | DNN approach | Biochemical parameters, demographics. | Achieved AUROC of 0.873 for predicting death in diabetic patients. |
| [47] | PIMA dataset | LR, SVM, LSTM, CNN, ANN, RNN | Health parameters, biochemical parameters | RNN achieved the best accuracy of 0.81 among all approaches. |
| [49] | UCI dataset | SVM, ANN, fused-method for diabetic prediction (FMDA) | Health parameters, biochemical parameters, demographics | FMDA achieved accuracy of 94.87%. |
| [51] | PIMA dataset | PE-DIM approach | Health parameters | PE-DIM achieved accuracy of 0.9453 for the PIMA dataset and AUC of 0.9917 and 0.9982 for other datasets. |

Table 1. Results and findings (*Continued*)

| Reference | Datasets | Methods | Parameters | Findings |
|---|---|---|---|---|
| [54] | PIMA, FHD, ADRC datasets | GWO algorithm | Health parameters, biochemical parameters, demographics | GWO achieved 0.98, 0.973, and 0.962 f-score for PIMA, ADRC, and FHD datasets respectively. |
| [55] | Dataset collected from smartwatches | Various ML approaches | Demographic info, sleep patterns, biochemical parameters, stress level | Evaluated using GB, SVR, SVM, KNN, RF, XGB, LR, LSTM. |
| [56] | DirectNet, Ai4pg datasets | FFNN | Health parameters, biochemical parameters | FFNN achieved RMSE scores for glucose level prediction. |
| [59] | Simulation dataset | - | Physical activity parameters | Established the role of physical activity in predicting glucose levels in Type-2 diabetic patients. |

## REFERENCES


[1] X. Lin *et al.*, "Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025," *Scientific Reports*, vol. 10, no. 1, p. 14790, Sep. 2020, doi: 10.1038/s41598-020-71908-9.

[2] B. O. Roep, S. Thomaidou, R. V. Tienhoven, and A. Zaldumbide, "Type 1 diabetes mellitus as a disease of the β-cell (do not blame the immune system?)," *Nature Reviews Endocrinology*, vol. 17, no. 3, pp. 150–161, Mar. 2021, doi: 10.1038/s41574-020-00443-4.

[3] The Lancet, "Diabetes: a defining disease of the 21$^{st}$ century," *The Lancet*, vol. 401, no. 10394, p. 2087, Jun. 2023, doi: 10.1016/S0140-6736(23)01296-5.

[4] R. Pradeepa and V. Mohan, "Epidemiology of type 2 diabetes in India," *Indian journal of ophthalmology*, vol. 69, no. 11, pp. 2932–2938, Nov. 2021, doi: 10.4103/ijo.IJO_1627_21.

[5] J. Xie *et al.*, "Global burden of type 2 diabetes in adolescents and young adults, 1990-2019: systematic analysis of the global burden of disease study 2019," *Bmj*, p. e072385, Dec. 2022, doi: 10.1136/bmj-2022-072385.

[6] K. B. L. Devi, K. T. Meitei, and S. J. Singh, "Prevalence of type 2 diabetes and its signs and symptoms among the meiteis of Manipur, India," *Journal of the Anthropological Survey of India*, vol. 72, no. 1, pp. 59–70, Jun. 2023, doi: 10.1177/2277436x221136955.

[7] I. Kyrou *et al.*, "Sociodemographic and lifestyle-related risk factors for identifying vulnerable groups for type 2 diabetes: a narrative review with emphasis on data from Europe," *BMC Endocrine Disorders*, vol. 20, no. S1, p. 134, Mar. 2020, doi: 10.1186/s12902-019-0463-3.

[8] Y. Li *et al.*, "Diabetic vascular diseases: molecular mechanisms and therapeutic strategies," *Signal Transduction and Targeted Therapy*, vol. 8, no. 1, p. 152, Apr. 2023, doi: 10.1038/s41392-023-01400-z.

[9] K. L. Ong *et al.*, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the global burden of disease study 2021," *The Lancet*, vol. 402, no. 10397, pp. 203–234, 2023, doi: 10.1016/S0140-6736(23)01301-6.

[10] American Diabetes Association, "Classification and diagnosis of diabetes: standards of medical care in diabetes—2022," *Diabetes Care*, vol. 45, no. Supplement_1, pp. S17–S38, Jan. 2022, doi: 10.2337/dc22-S002.

[11] X. Tao *et al.*, "Predicting three-month fasting blood glucose and glycated hemoglobin changes in patients with type 2 diabetes mellitus based on multiple machine learning algorithms," *Scientific Reports*, vol. 13, no. 1, p. 16437, Sep. 2023, doi: 10.1038/s41598-023-43240-5.

[12] P. L. Lee, K. W. Wang, and C. Y. Hsiao, "A noninvasive blood glucose estimation system using dual-channel PPGs and pulse-arrival velocity," *IEEE Sensors Journal*, vol. 23, no. 19, pp. 23570–23582, Oct. 2023, doi: 10.1109/JSEN.2023.3306343.

[13] Z. Sekyonda, R. An, A. Avanaki, A. Fraiwan, and U. A. Gurkan, "A novel approach for glycosylated hemoglobin testing using microchip affinity electrophoresis," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 5, pp. 1473–1480, May 2023, doi: 10.1109/TBME.2022.3218501.

[14] S. J. J. Kirubakaran, M. A. Bennet, and N. R. Shanker, "Antiallergic abdominal belt for human glucose level measurement using microwave active sensor antenna," *IEEE Sensors Journal*, vol. 23, no. 14, pp. 16227–16237, Jul. 2023, doi: 10.1109/JSEN.2023.3280031.

[15] L. P. Nguyen *et al.*, "The utilization of machine learning algorithms for assisting physicians in the diagnosis of diabetes," *Diagnostics*, vol. 13, no. 12, p. 2087, Jun. 2023, doi: 10.3390/diagnostics13122087.

[16] H. El-Sofany, S. A. El-Seoud, O. H. Karam, Y. M. Abd El-Latif, and I. A. T. F. Taj-Eddin, "A proposed technique using machine learning for the prediction of diabetes disease through a mobile app," *International Journal of Intelligent Systems*, vol. 2024, pp. 1–13, Jan. 2024, doi: 10.1155/2024/6688934.

[17] A. E. S. El-Bashbishy and H. M. El-Bakry, "Pediatric diabetes prediction using deep learning," *Scientific Reports*, vol. 14, no. 1, p. 4206, Feb. 2024, doi: 10.1038/s41598-024-51438-4.

[18] O. R. Shahin, H. H. Alshammari, A. A. Alzahrani, H. Alkhiri, and A. I. Taloba, "A robust deep neural network framework for the detection of diabetes," *Alexandria Engineering Journal*, vol. 74, pp. 715–724, Jul. 2023, doi: 10.1016/j.aej.2023.05.072.

[19] T. A. Ojurongbe *et al.*, "Predictive model for early detection of type 2 diabetes using patients' clinical symptoms, demographic features, and knowledge of diabetes," *Health Science Reports*, vol. 7, no. 1, Jan. 2024, doi: 10.1002/hsr2.1834.

[20] S. P. Chatrati *et al.*, "Smart home health monitoring system for predicting type 2 diabetes and hypertension," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 3, pp. 862–870, Mar. 2022, doi: 10.1016/j.jksuci.2020.01.010.

[21] O. H. Jasim, M. M. Mahmood, and A. H. Ad'hiah, "Significance of lipid profile parameters in predicting pre- diabetes," *Archives of Razi Institute*, vol. 77, no. 1, pp. 267–274, 2022, doi: 10.22092/ARI.2021.356465.1846.
[22] L. Dénes-Fazakas, M. Siket, L. Szilágyi, L. Kovács, and G. Eigner, "Detection of physical activity using machine learning methods based on continuous blood glucose monitoring and heart rate signals," *Sensors*, vol. 22, no. 21, p. 8568, Nov. 2022, doi: 10.3390/s22218568.
[23] A. I. Khan *et al.*, "Computational approach for detection of diabetes from ocular scans," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–8, May 2022, doi: 10.1155/2022/5066147.
[24] C. Antza, G. Kostopoulos, S. Mostafa, K. Nirantharakumar, and A. Tahrani, "The links between sleep duration, obesity and type 2 diabetes mellitus," *Journal of Endocrinology*, vol. 252, no. 2, pp. 125–141, Feb. 2022, doi: 10.1530/JOE-21-0155.
[25] M. Sevil, M. Rashid, I. Hajizadeh, M. Park, L. Quinn, and A. Cinar, "Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 7, pp. 2251–2260, Jul. 2021, doi: 10.1109/TBME.2020.3049109.
[26] F. Mohsen, H. R. H. Al-Absi, N. A. Yousri, N. El Hajj, and Z. Shah, "A scoping review of artificial intelligence-based methods for diabetes risk prediction," *npj Digital Medicine*, vol. 6, no. 1, p. 197, Oct. 2023, doi: 10.1038/s41746-023-00933-5.
[27] J. Chaki, S. Thillai Ganesh, S. K. Cidham, and S. Ananda Theertan, "Machine learning and artificial intelligence based diabetes mellitus detection and self-management: a systematic review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3204–3225, Jun. 2022, doi: 10.1016/j.jksuci.2020.06.013.
[28] T. Shaik, X. Tao, L. Li, H. Xie, and J. D. Velásquez, "A survey of multimodal information fusion for smart healthcare: mapping the journey from data to wisdom," *Information Fusion*, vol. 102, p. 102040, Feb. 2024, doi: 10.1016/j.inffus.2023.102040.
[29] J. D. Bodapati *et al.*, "Blended multi-modal deep convnet features for diabetic retinopathy severity prediction," *Electronics (Switzerland)*, vol. 9, no. 6, p. 914, May 2020, doi: 10.3390/electronics9060914.
[30] Mariaherrerot, "APTOS-2019 dataset," *Kaggle*, 2022. https://www.kaggle.com/datasets/mariaherrerot/aptos2019.
[31] R. Cordeiro, N. Karimian, and Y. Park, "Hyperglycemia identification using ecg in deep learning era," *Sensors*, vol. 21, no. 18, p. 6263, Sep. 2021, doi: 10.3390/s21186263.
[32] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.cir.101.23.e215.
[33] Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, "Multimodal image encoding pre-training for diabetic retinopathy grading," *Computers in Biology and Medicine*, vol. 143, p. 105302, Apr. 2022, doi: 10.1016/j.compbiomed.2022.105302.
[34] S. Hajeb Mohammad Alipour, H. Rabbani, and M. R. Akhlaghi, "Diabetic retinopathy grading by digital curvelet transform," *Computational and Mathematical Methods in Medicine*, vol. 2012, pp. 1–11, 2012, doi: 10.1155/2012/761901.
[35] J. Cuadros and G. Bresnick, "EyePACS: An adaptable telemedicine system for diabetic retinopathy screening," *Journal of Diabetes Science and Technology*, vol. 3, no. 3, pp. 509–516, May 2009, doi: 10.1177/193229680900300315.
[36] E. Decencière *et al.*, "Feedback on a publicly distributed image database: the messidor database," *Image Analysis and Stereology*, vol. 33, no. 3, pp. 231–234, Aug. 2014, doi: 10.5566/ias.1155.
[37] M. D. Abràmoff *et al.*, "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA Ophthalmology*, vol. 131, no. 3, pp. 351–357, Mar. 2013, doi: 10.1001/jamaophthalmol.2013.1743.
[38] P. Porwal *et al.*, "IDRiD: diabetic retinopathy – segmentation and grading challenge," *Medical Image Analysis*, vol. 59, p. 101561, Jan. 2020, doi: 10.1016/j.media.2019.101561.
[39] A. R. Kulkarni *et al.*, "Machine-learning algorithm to non-invasively detect diabetes and pre-diabetes from electrocardiogram," *BMJ Innovations*, vol. 9, no. 1, pp. 32–42, Jan. 2022, doi: 10.1136/bmjinnov-2021-000759.
[40] Z. Yu, W. Luo, R. Tse, and G. Pau, "DMNet: a personalized risk assessment framework for elderly people with Type 2 diabetes," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, pp. 1558–1568, Mar. 2023, doi: 10.1109/JBHI.2022.3233622.
[41] M. Mceniry, "Research on early life and aging trends and effects (RELATE): a cross-national study harmonized cross-national relate data," 2024. https://hrsdata.isr.umich.edu/data-products/research-early-life-and-aging-trends-and-effects-relate-files (accessed Apr. 22, 2024).
[42] M. Botella-Serrano, J. M. Velasco, A. Sánchez-Sánchez, O. Garnica, and J. I. Hidalgo, "Evaluating the influence of sleep quality and quantity on glycemic control in adults with type 1 diabetes," *Frontiers in Endocrinology*, vol. 14, Feb. 2023, doi: 10.3389/fendo.2023.998881.
[43] C. Rodriguez-Leon *et al.*, "T1DiabetesGranada: a longitudinal multi-modal dataset of type 1 diabetes mellitus," *Scientific Data*, vol. 10, no. 1, p. 916, Dec. 2023, doi: 10.1038/s41597-023-02737-4.
[44] A. Pai *et al.*, "Multimodal digital phenotyping of diet, physical activity, and glycemia in Hispanic/Latino adults with or at risk of type 2 diabetes," *npj Digital Medicine*, vol. 7, no. 1, p. 7, Jan. 2024, doi: 10.1038/s41746-023-00985-7.
[45] J. Theis, W. L. Galanter, A. D. Boyd, and H. Darabi, "Improving the in-hospital mortality prediction of diabetes ICU patients using a process mining/deep learning architecture," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 388–399, Jan. 2022, doi: 10.1109/JBHI.2021.3092969.
[46] A. Johnson, T. Pollard, and R. Mark, "MIMIC-III clinical database," *Physionet.org*, 2023. http://dx.doi.org/10.13026/CD7Z-WG25 (accessed Sep. 04, 2016).
[47] A. Naseem, R. Habib, T. Naz, M. Atif, M. Arif, and S. A. Chelloug, "Novel internet of things based approach toward diabetes prediction using deep learning models," *Frontiers in Public Health*, vol. 10, Aug. 2022, doi: 10.3389/fpubh.2022.914106.
[48] P. D. Kumar, "Pima indians diabetes database," *Kaggle*, 2020. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.
[49] U. Ahmed *et al.*, "Prediction of diabetes empowered with fused machine learning," *IEEE Access*, vol. 10, pp. 8529–8538, 2022, doi: 10.1109/ACCESS.2022.3142097.
[50] "UCI Machine Learning Repository," *Archive*. https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+dataset .
[51] Z. W. L. Jia, S. Lv, and Z. Xu, "PE_DIM: an efficient probabilistic ensemble classification algorithm for diabetes handling class imbalance missing values," *IEEE Access*, vol. 10, pp. 107459–107476, 2022, doi: 10.1109/ACCESS.2022.3212067.
[52] B. A. Tama, R. F. S., and H. Hermansyah, "An early detection method of type-2 diabetes mellitus in public hospital," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 9, no. 2, p. 287, 2011, doi: 10.12928/telkomnika.v9i2.699.
[53] M. Heydari, M. Teimouri, Z. Heshmati, and S. M. Alavinia, "Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran," *International Journal of Diabetes in Developing Countries*, vol. 36, no. 2, pp. 167–173, 2015, doi: 10.1007/s13410-015-0374-4.

[54] P. Yadav, S. C. Sharma, R. Mahadeva, and S. P. Patole, "Exploring hyper-parameters and feature selection for predicting non-communicable chronic disease using stacking classifier," *IEEE Access*, vol. 11, pp. 80030–80055, 2023, doi: 10.1109/access.2023.3299332.

[55] S. T. Himi, N. T. Monalisa, M. Whaiduzzaman, A. Barros, and M. S. Uddin, "MedAi: a smartwatch-based application framework for the prediction of common diseases using machine learning," *IEEE Access*, vol. 11, pp. 12342–12359, 2023, doi: 10.1109/ACCESS.2023.3236002.

[56] G. Annuzzi *et al.*, "Impact of nutritional factors in blood glucose prediction in type 1 diabetes through machine learning," in *IEEE Access*, vol. 11, pp. 17104-17115, 2023, doi: 10.1109/ACCESS.2023.3244712.

[57] "Diabetes research in children network - Public Site," public.jaeb.org. https://public.jaeb.org/direcnet/stdy/167 (accessed Apr. 22, 2024).

[58] L. Angrisani *et al.*, "Neural network-based prediction and monitoring of blood glucose response to nutritional factors in type-1 diabetes," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Ottawa, ON, Canada, 2022, pp. 1-6, doi: 10.1109/I2MTC48687.2022.9806611.

[59] P. F. De Paola, A. Paglialonga, P. Palumbo, K. Keshavjee, F. Dabbene and A. Borri, "The long-term effects of physical activity on blood glucose regulation: a model to unravel diabetes progression," in *IEEE Control Systems Letters*, vol. 7, pp. 2916-2921, 2023, doi: 10.1109/LCSYS.2023.3290774.


## BIOGRAPHIES OF AUTHORS

**Gururaj N. Kulkarni** presently working as assistant professor in BLDEA Smt Bangaramma Sajjan Arts, Commerce and Science College for Women, Vijayapur BCA Department and currently pursuing Ph.D. in Computer Science and Application in Artificial Intelligence and Machine Learning from Shri Jagdishprasad Jhabarmal Tibrewala University (JJTU) in Jhunjhunu, Rajasthan. With 9-year teaching experience, he has published international, national papers and attended conferences. And also, he has two patents. He can be contacted at email: gururajnkulkarni40@gmail.com.

**Dr. Kelapati** assistant professor and head of the department, Department of Computer Science and Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, Chudela, Vidyanagri, Jhunjhunu, Rajasthan 333010 (India). She has been published number of research papers in national and international journals. She has also many national and international conferences successfully completed. She has successfully completed FDP which is organized by NPTEL–AICTE, Indian Institute of Technology. She has also attended FIP one-month online Faculty Induction Programme Organized by Malaviya Mission Teacher Training Center (formerly UGC- Human Resourse Development Centre), Jamia Millia Islamia, New Delhi. She can be contacted at email: kelapatipoonia@gmail.com.