International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Reviewing Predictive Modeling for Type-3 Diabetes: Challenges and the Path Forward

Rakhi Kumari[1,*], Pratap Singh Patwal[2]

[1,2]Department of Computer Science and Engineering, Laxmi Devi Institute of Engineering and Technology

Alwar-Tijara-Delhi, Highway, Chikani, Rajasthan, India

[1,*]rkumari2070@gmail.com, [2]pratappatwal@gmail.com

---***---

Abstract -The growing number of people affected by diabetes and its complications calls for new ways to detect, manage, and treat the condition early. This paper presents a machine learning-based framework designed to predict and classify Type 1, Type 2, and Type 3 diabetes with improved accuracy while addressing key gaps in current research. The framework combines data types, such as clinical records, demographic details, and biosignal data from wearable devices, to ensure a comprehensive and reliable approach. It also uses advanced data processing methods like feature selection, dimensionality reduction, and handling imbalanced datasets to improve model performance.
The paper explores machine learning to highlight the importance of real-time monitoring, tracking data over time, and creating personalized risk assessments, especially for pregnant women. Additionally, it addresses practical challenges like making the models scalable, considering ethical issues, and integrating socioeconomic and biosignal data. Evaluation metrics like accuracy, precision, recall, and F1-score may ensure the model's reliability and effectiveness.
This paper shows how machine learning can revolutionize diabetes care by providing early warnings, reducing complications, and offering actionable insights for better management. By combining advanced algorithms with real-world usability, the proposed framework bridges the gap between research and practical healthcare, offering a scalable solution for improved diabetes management.

Keywords: Diabetes Mellitus, type 2 diabetes, Type-3 Diabetes, common metabolic disorder, diabetes treatment

I.INTRODUCTION

The prevalence of diabetes is rising, with 382 million cases globally in 2020 and projections reaching 592 million by 2035 [21]. Factors like urbanization, lifestyle changes, and aging contribute to this trend, creating challenges for management [22]. In 2016, WHO reported 8.5% of adults (17+) had diabetes, with 1.5 million direct and 2.3 million indirect fatalities in 2013 [21]. Despite advancements in care, delayed diagnosis often leads to severe complications [23].
Table 1 summarizes the characteristics, causes, and symptoms of Type 1, Type 2, and Type 3 (Gestational) Diabetes, with relevant references for clarity and further study.

Table 1 Overview of Diabetes Types and Common Symptoms

Type of Diabetes/ Common Symptoms	Description	Ref.
Type 1 Diabetes	Autoimmune destruction of pancreatic beta cells, requiring insulin therapy.	[33]
Type 2 Diabetes	Insulin resistance and insufficient production, often linked to lifestyle factors.	[33-35]
Type 3 Diabetes (Gestational Diabetes)	Temporary glucose intolerance during pregnancy, requiring glucose monitoring.	[33]
Symptoms	Frequent urination, chronic thirst, unexplained weight loss, fatigue, blurred vision, and delayed wound healing.	[33-35]

The three primary datasets (See Table 2) Diabetes_DataSet.csv, NCSU_Diabetes_Dataset, and Ohio_T3DM_DATASET [32] serve as the foundation for exploring diabetes patterns, building predictive models, and enhancing classification accuracy in diabetes research. Diabetes_DataSet.csv includes critical physiological and medical parameters such as glucose levels, BMI, and insulin measures, providing a robust basis for analyzing diabetes trends and identifying high-risk individuals. The NCSU_Diabetes_Dataset extends this research by incorporating cognitive assessments and patient history, including factors like MMSE scores and HbA1c levels, enabling the development of more nuanced predictive models. Lastly, the Ohio_T3DM_DATASET focuses on Type-3 Diabetes (Gestational Diabetes), with a unique emphasis on pregnancy-related glucose control and medical parameters, enhancing the accuracy of gestational

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

diabetes classification. Together, these datasets provide a comprehensive framework for advancing diabetes research and improving healthcare outcomes.

Table 2 Review and Explanation of Primary Datasets Used in Diabetes Research

Dataset Name	Purpose	Key Features
Diabetes_DataSet.csv [32]	Explore diabetes patterns and trends.	Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome.
NCSU_Diabetes_Dataset [37]	Build predictive models for diabetes classification.	Patient ID, Group (Pregnant/Non-Pregnant), Visit, Age, MMSE, HbA1c, Cognitive Function, Glycemic Control.
Ohio_T3DM_DATASET[36,37]	Enhance diabetes classification accuracy with emphasis on Type-3 Diabetes (Gestational).	MR Delay, M/F, Age, SES, PR Beta, HbA1c, Glycemic Control.

Diabetes affects 34.2 million people in the U.S., with healthcare costs reaching $3.8 trillion in 2020. Effective management involves early detection, education, and resource optimization, supported by ML tools for efficient diagnosis.

Table 3 Classification and Diagnosis of Diabetes

Classification	Diagnosis Criteria
Normal	Fasting plasma glucose (FPG) < 100 mg/dL or 2-hour plasma glucose < 140 mg/dL
Prediabetes	FPG 100-125 mg/dL or 2-hour plasma glucose 140-199 mg/dL or HbA1c 5.7%-6.4%
Diabetes	FPG ≥ 126 mg/dL or 2-hour plasma glucose ≥ 200 mg/dL or HbA1c ≥ 6.5% or random glucose ≥ 200 mg/dL

Source: Adapted from American Diabetes Association. *Diabetes Care* [32]

Table 3 outlines the diagnostic criteria for Normal, Prediabetes, and Diabetes according to fasting plasma glucose (FPG), 2-hour plasma glucose, and HbA1c levels. Normal individuals exhibit a fasting plasma glucose (FPG) level of less than 100 mg/dL or a 2-hour plasma glucose level of less than 140 mg/dL. Prediabetes signifies an elevated risk, characterized by a fasting plasma glucose (FPG) of 100-125 mg/dL, a 2-hour plasma glucose level of 140-199 mg/dL, or a HbA1c between 5.7% and 6.4%. Diabetes is diagnosed when FPG is ≥ 126 mg/dL, 2-hour plasma glucose is ≥ 200 mg/dL, HbA1c is ≥ 6.5%, or random glucose is ≥ 200 mg/dL. These criteria, derived from ADA guidelines, assist healthcare practitioners in achieving precise classification and successful intervention.

Machine learning (ML) models have proven effective for disease prediction, offering cost-efficient solutions for diabetes diagnosis [24]. Diabetes mellitus (DM) affects energy conversion, with type 2 diabetes posing the greatest burden due to factors like lifestyle and genetics [25-30]. ML-driven systems aid in early detection, mitigating complications like cardiovascular issues, renal impairment, and neuropathy [31-33]. ML methodologies like random forests, logistic regression, and neural networks predict diabetes by analyzing diverse datasets. Predictive analytics helps identify high-risk individuals, providing personalized healthcare solutions.

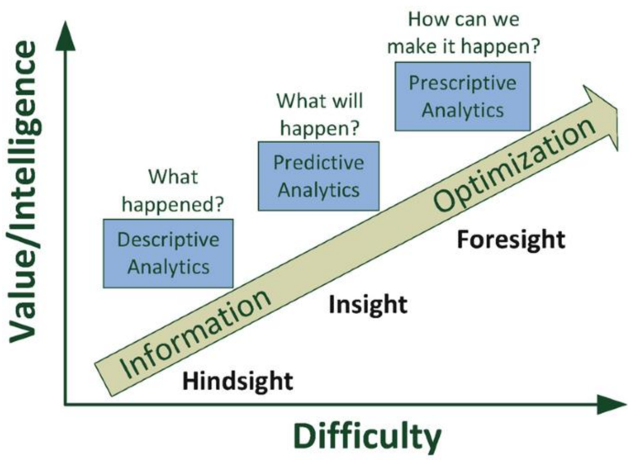

Fig. 1 Stages of Business Analytics and the Analytics Value-Difficulty Spectrum (Source: Gartner Inc.)

As illustrated in Figure 1, these categories escalate in both value and complexity. Organizations generally implement analytics in this order. Descriptive analytics centers on comprehending historical events to extract insights, whereas diagnostic analytics determines the causes of past incidents to anticipate or manage future results. Predictive analytics employs data-driven models to project future occurrences, and prescriptive analytics seeks to influence or optimize outcomes based on these predictions. Collectively, these four categories encapsulate the objectives of analytics.

This paper explores the use of machine learning (ML) algorithms to recognize and classify diabetes types, including Type-3 diabetes, and perform predictive analysis. By leveraging diverse datasets containing clinical, demographic, and biochemical features, the study aims to develop reliable predictive models for early detection, risk assessment, and personalized treatment. Advanced ML techniques, such as KNN, gradient boosting, logistic regression, and hybrid algorithms, are utilized to uncover factors influencing diabetes outcomes and provide insights into its pathophysiology.

Given the global rise of diabetes and its associated complications, this work emphasizes the need for effective prediction technologies. The proposed system uses ML to analyze a wide range of health-related factors, enabling timely interventions and tailored care. This innovative approach aims to improve healthcare outcomes, reduce complications, and enhance patients' quality of life by addressing diabetes with precision and efficiency.

The growing burden of diabetes underscores the potential of ML-based predictive models to identify individuals at risk of types 1, 2, and 3 diabetes. These models enable early intervention, personalized care, and better healthcare outcomes, reducing complications and financial strain [32]. This potential for significant improvement in patient outcomes is what motivates us to work in this area.

This research's significant contributions lie in the development of machine learning-based models for type 1, type 2, and type 3 diabetes, enabling early detection and personalized management. By integrating diverse datasets comprising demographic, clinical, and biochemical characteristics, the study enhances diabetes risk prediction accuracy. Advanced techniques like gradient boosting, KNN, and hybrid algorithms improve model reliability and interpretability. The research provides tools for early detection, targeted prevention, and optimized resource allocation, thereby reducing healthcare burdens. Moreover, it promotes patient-centric solutions through education and lifestyle guidance while contributing to healthcare technology with data-driven chronic disease management systems. A robust clinical implementation framework ensures the practical application of these models, thereby enhancing patient outcomes and evolving healthcare systems.

The paper is structured into five key sections. Section I provides an overview of fundamental concepts related to diabetes prediction and management, encompassing diabetes types, symptoms, healthcare strategies, machine learning methodologies, and research motivations. Section II discusses related work, followed by Section III, which identifies research gaps and outlines future directions. Section IV focuses on diabetes management through machine learning techniques, and Section V concludes the paper by addressing challenges and proposing future pathways.

II.RELATED WORK

This thesis builds on a comprehensive review of research in diabetes prediction and management. Khaled Alnowaiser et al. [1] proposed a Tri-Ensemble model with KNN imputation, emphasizing feature augmentation and ensemble learning to enhance prediction accuracy. Giovanni Annuzzi et al. [2] highlighted the role of dietary components in machine learning-based blood glucose prediction for Type 1 Diabetes, underscoring the importance of personalized healthcare solutions. Aditi Site et al. [3] explored multisensor data analysis for diabetes prediction, demonstrating the effectiveness of XGBoost in improving accuracy through feature engineering.

Pierluigi Francesco De Paola et al. [4] developed a model integrating interleukin-6 dynamics with glucose-insulin regulation, offering insights into exercise-induced diabetes management. Usama Ahmed et al. [5] showcased the potential of fused machine learning techniques to improve prediction accuracy, emphasizing their role in personalized healthcare and diagnostic tools. Hakim El Massari et al. [6] focused on ontology-based categorization methods and benchmarking studies, highlighting the need for accurate medical prediction models.

Julian Theis et al. [7] introduced a process mining and deep learning framework for predicting in-hospital mortality among diabetic ICU patients, leveraging EHR data to improve outcomes. John Daniels et al. [8] demonstrated the superiority of multitask learning over traditional methods for individualized blood glucose prediction, offering a promising avenue for personalized treatment strategies. Mohammad Zubair Khan et al. [9] investigated hybrid evolutionary approaches, combining convolutional neural networks with PSO-NNDP frameworks, achieving significant accuracy gains. Namho Kim et al. [10] developed a clustering-based model for personalized Hemoglobin A1c estimation, advancing AI-driven healthcare solutions for Type 2 Diabetes.

This literature review lays a robust foundation for the thesis, focusing on Hybrid Predictive Modeling for Type-3 Diabetes to enhance precision in diabetes classification and prediction.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Table 4 Summary of Research Studies on Diabetes Prediction and Management

Study Abstract	Contributions	Limitations	Ref.
Proposed a Tri-Ensemble Model with KNN-imputed features for improved diabetes prediction.	Improved diabetes prediction accuracy by combining imputation and ensemble learning techniques.	Model complexity may increase computational requirements.	[1].
Examined the role of nutritional factors in blood glucose prediction for Type 1 diabetes using machine learning.	Identified significant nutritional factors influencing blood glucose levels in Type 1 diabetes.	Limited to nutritional factors without considering other influencing variables.	[2].
Utilized multisensor data with machine learning to predict diabetes.	Introduced the use of multisensor data for enhancing diabetes prediction.	Dependent on multisensor data availability and quality.	[3].
Studied the long-term impact of physical activity on blood glucose regulation and diabetes progression.	Proposed a model to understand the impact of physical activity on diabetes progression.	Requires further validation for different demographics and activity levels.	[4].
Implemented a fused machine learning approach for accurate diabetes prediction.	Enhanced prediction accuracy through a novel fused machine learning method.	Limited generalizability for datasets with varying feature distributions.	[5].
Applied machine learning algorithms combined with ontology for diabetes prediction.	Integrated ontology with machine learning to refine diabetes prediction models.	Ontology integration may add complexity to implementation.	[6].
Used process mining and deep learning for in-hospital mortality prediction in diabetes ICU patients.	Improved prediction of in-hospital mortality for ICU diabetes patients.	Restricted to ICU settings; not generalizable to outpatient scenarios.	[7].
Developed a multitask learning approach for personalized blood glucose prediction.	Enhanced personalized blood glucose prediction with multitask learning.	Requires extensive data for personalized predictions.	[8].
Proposed a bio-inspired PSO-based approach to enhance neural network diabetes prediction systems.	Introduced PSO optimization for improved neural network-based predictions.	Optimization results may vary with different datasets.	[9].
Developed personalized Hemoglobin A1c estimation for Type 2 diabetes using sensor data.	Enhanced personalized diabetes care through HbA1c estimation.	Limited applicability for populations without adequate sensor data.	[10].
Proposed a deep learning model for predicting genome disorders, advancing diagnostic capabilities.	Enhanced genome disorder prediction accuracy using deep learning.	Model performance may vary with genetic data availability and quality.	[11].
Reviewed data mining techniques for diabetes detection and prediction comprehensively.	Highlighted strengths and weaknesses of various data mining techniques for diabetes.	Focuses mainly on data mining; lacks advanced machine learning comparisons.	[12].

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Surveyed machine and deep learning models addressing childhood and adolescent obesity.	Identified gaps in obesity prediction models for younger populations.	Primarily surveys existing methods without proposing novel solutions.	[13].
Developed a method for type 2 diabetes prediction using an average weighted objective distance-based approach.	Introduced a novel prediction metric improving type 2 diabetes diagnosis.	Requires further testing on diverse populations and datasets.	[14].
Explored machine learning tools for predicting long-term risks of type 2 diabetes.	Provided tools to assess long-term risk factors effectively.	Limited generalizability for short-term diabetes risk prediction.	[15].
Studied physical activity and stress effects on glucose predictions for diabetes management.	Integrated physical and psychological factors into glucose concentration predictions.	Highly context-dependent on physical and psychological factors.	[16].
Proposed an early-stage risk prediction model for non-communicable diseases using machine learning.	Addressed non-communicable disease prevention with early prediction techniques.	Challenges in scaling the model for global application.	[17].
Introduced an ensemble machine learning approach for diabetes prediction.	Improved diabetes prediction accuracy with ensemble methods.	Model complexity may increase computational requirements.	[18].
Investigated advanced techniques for predicting the progression of type 2 diabetes.	Provided techniques to assess future risk progression in type 2 diabetes.	Focuses on progression; limited in addressing immediate diagnosis.	[19].
Developed predictive models linking metabolic syndrome to diabetes mellitus using machine learning.	Linked metabolic syndrome and diabetes risk effectively using predictive modeling.	Does not include non-metabolic factors affecting diabetes.	[20].

Atta-ur-Rahman et al. [11] introduced the Advance Genome Disorder Prediction Model (AGDPM), leveraging deep learning for improved accuracy in genetic disease predictions. Farrukh Aslam Khan et al. [12] reviewed data mining techniques for diabetes diagnosis, emphasizing algorithms like J48, KNN, and Random Forests, and recommending hybrid approaches for future work. Hera Siddiqui et al. [13] highlighted the role of machine learning in addressing childhood obesity, emphasizing demographic and psychological factors.

Pratya Nuankaew et al. [14] showcased the AWOD method's accuracy in Type 2 diabetes prediction by identifying key health factors. Nikos Fazakis et al. [15] evaluated long-term risk prediction systems, demonstrating the superiority of ensemble algorithms. Mert Sevil et al. [16] explored wearable device data for glucose prediction, advancing precision medicine for Type 1 Diabetes.

Rahatara Ferdousi et al. [17] emphasized the integration of AI in Health CPS, proposing a novel framework for disease prediction. Md. Kamrul Hasan et al. [18] presented an ensemble method with AUC-weighted soft voting, achieving improved diabetes forecasting accuracy. Md. Shafiqul Islam et al. [19] demonstrated effective T2DM prediction within 7–8 years using OGTT data and unique feature extraction techniques. Sajida Perveen et al. [20] investigated the link between metabolic syndrome and diabetes, highlighting HDL's unexpected correlation with diabetes onset and recommending advanced predictive methods. These studies collectively provide a robust foundation for advancing diabetes prediction and management techniques.

Pradhan et al. (2020) [21] demonstrated the effectiveness of Artificial Neural Networks (ANN) in predicting diabetes, showcasing their accuracy in managing complex health datasets. Liu et al. (2013) [22] extended the application of neural networks to forecasting cardiovascular autonomic dysfunction, while Temurtas et al. (2009) [23] highlighted the superiority of ANN over traditional diagnostic methods for diabetes.

Adeloye et al. (2017) [24] emphasized the healthcare challenges of Type 2 Diabetes in Nigeria, calling for advanced predictive methods. The International Diabetes Federation (IDF) [25, 26] provided a global perspective on diabetes trends, reinforcing the importance of AI-driven tools in addressing this growing burden.

Ismail et al. (2020) [27] analyzed lifestyle and genetic factors contributing to Type 2 Diabetes, stressing prevention strategies, while the American Diabetes Association (2021) [28] underlined the need for early intervention to mitigate complications. Licitra et al. (2017) [29] explored machine learning's role in diabetes management, highlighting ethical challenges, and Ismail et al. (2020) [30] discussed the need for secure health data systems to support AI integration in healthcare.

Table 4 provides an overview of various research efforts in the field. It details the researchers involved in each study and concisely summarizes their focus and objectives. Key findings, innovations, and advancements are highlighted in each research effort, emphasizing their contributions to diabetes prediction and management. Additionally, the table identifies challenges, constraints, or areas requiring further exploration, providing insights into the scope of future research.

III.RESEARCH GAP AND DIRECTION

A thorough analysis of the literature on machine learning for diabetes prediction and management reveals several research gaps. Firstly, while many studies employ personalized approaches, such as incorporating dietary and lifestyle factors, few explore dynamic models that adapt to real-time inputs like food intake, exercise, and stress levels, which hold significant potential for diabetes management. The integration of contextual awareness and biosignal data from wearable devices remains underexplored. Secondly, improving the interpretability of explainable AI techniques like SHAP and LIME is essential to enhance usability for healthcare professionals and patients. Thirdly, there is limited research on combining diverse data sources, such as clinical, wearable, and socioeconomic data, to develop robust and scalable models, with a heavy reliance on datasets like the Pima Indian Diabetes dataset, highlighting the need for diverse, real-world datasets.

In addition, advanced feature engineering and dimensionality reduction methods, such as Random Projection, show promise but require further integration with deep learning frameworks. Longitudinal studies on diabetes progression and treatment efficacy are sparse, and hybrid and ensemble models, while effective, could benefit from optimization through metaheuristic approaches. Addressing class imbalances using techniques like SMOTE and ADASYN remains an area requiring frameworks for consistent performance across imbalanced datasets. Furthermore, ethical challenges, including data security, patient privacy, and algorithmic biases, persist as critical barriers to implementation.

Real-time predictive analytics for continuous diabetes monitoring and feedback systems are still in their infancy, particularly for clinical and non-clinical applications. Lastly, there is a need for comprehensive evaluation strategies that integrate clinical, operational, and economic outcomes to assess the practical feasibility of machine learning models in healthcare systems. These gaps underscore the necessity for further research to address the challenges and advance diabetes prediction and management.

Advancements in machine learning have opened new frontiers in diabetes prediction and management, yet significant research gaps remain that hinder the full realization of its potential. Addressing these gaps can pave the way for innovative solutions, improving both the accuracy and applicability of predictive models in real-world healthcare scenarios. By focusing on areas such as dynamic and real-time monitoring, a crucial aspect that keeps the audience engaged, enhancing interpretability, integrating diverse data sources, and addressing ethical concerns, researchers can contribute to a more holistic and effective approach to diabetes management. The following points outline key directions for future research, providing a roadmap for tackling existing challenges and advancing the field.

- **Dynamic and Real-Time Diabetes Management Models :** Develop adaptive machine learning models that respond to real-time inputs such as dietary intake, exercise, and stress levels. These models can leverage wearable devices and biosignal data to provide personalized diabetes care.
- **Improving Explainable AI for Healthcare :** Enhance the interpretability of tools like SHAP and LIME to make AI predictions more comprehensible for healthcare professionals and patients. This will increase trust and adoption of machine learning models in clinical settings.
- **Integration of Multi-Source Data :** Combine clinical data with information from wearable devices, socioeconomic profiles, and environmental factors to build comprehensive and scalable diabetes prediction models. Focus on diverse, real-world datasets to improve model generalizability.
- **Advanced Feature Engineering and Dimensionality Reduction:** Explore the synergistic application of dimensionality reduction techniques (e.g., Random Projection) with deep learning frameworks to improve predictive performance and computational efficiency.
- **Longitudinal Data Analysis:** Conduct studies that focus on tracking diabetes progression and treatment efficacy over time. This can lead to better

understanding of disease dynamics and long-term intervention strategies.

- **Optimization of Hybrid and Ensemble Models :** Investigate metaheuristic optimization techniques to enhance the performance of hybrid and ensemble machine learning models. This could improve their robustness and predictive accuracy in complex datasets.
- **Addressing Class Imbalances in Datasets:** Develop frameworks that ensure consistent model performance across imbalanced datasets. Explore advanced sampling techniques and model architectures that mitigate bias in predictions.
- **Ethics and Algorithmic Bias Mitigation:** Focus on data security, patient privacy, and reducing algorithmic biases to address ethical concerns. Develop transparent frameworks that comply with regulatory standards and ensure equitable healthcare solutions.
- **Real-Time Predictive Analytics and Feedback Systems:** Create systems that enable continuous monitoring and real-time predictions for diabetes care. These could integrate non-clinical and clinical data to provide timely feedback and interventions.
- **Holistic Evaluation Strategies:** Design evaluation frameworks that integrate clinical, operational, and economic outcomes to assess the practicality and impact of machine learning models on healthcare systems. This will help in their real-world implementation and scalability.

IV.DIABETES MANAGEMENT USING MACHINE LEARNING

The proposed approaches for enhancing diabetes subtype classification and prediction integrate cutting-edge data analysis tools with traditional machine learning techniques. Figure 2 illustrates a proposed framework for comprehensive diabetes prediction and risk assessment via machine learning. The proposed framework provides a systematic approach for diabetes prediction and management through the application of machine learning techniques. The approach begins with the acquisition of a comprehensive diabetes dataset that includes clinical, demographic, and biochemical data. This dataset receives comprehensive preprocessing, addressing missing values, normalizing data, and encoding categorical variables, so ensuring its analytical appropriateness and enhancing data dependability. Thereafter, exploratory data analysis (EDA) is conducted to identify patterns, correlations, and trends in the dataset, aiding in feature selection and improving data understanding.

The framework utilizes machine learning techniques to create predictive models, establishing separate classification models for Type 1 and Type 2 diabetes. These predictions pertain to Type 3 diabetes, a relatively under-researched kind, employing advanced classification techniques. Longitudinal data is employed to enhance the comprehension of the progression and risk factors associated with Type 3 diabetes. The methodology is explicitly formulated to forecast the risk of Type 3 diabetes in pregnant women, focusing on a critical and vulnerable population and underscoring the significance of the research for this demographic.

The subsequent improvement involves utilizing conditional rules or machine learning techniques to increase predictive accuracy and derive actionable insights. The final stage involves comprehensive evaluation and validation of the models using performance metrics such as accuracy, precision, recall, and AUC-ROC, ensuring reliability and generalizability. This system integrates advanced analytics, accurate forecasts, and risk assessment, precisely addressing shortcomings in Type 3 diabetes prediction and offering a comprehensive solution for diabetes management, thereby instilling confidence in the audience regarding the sophistication of the methodology.

Effective diabetes prediction and management require a comprehensive and systematic approach that leverages the power of machine learning and data analytics. The proposed framework aims to address this need by integrating diverse data sources, robust preprocessing techniques, and advanced machine learning algorithms to predict and assess the risks associated with Type 1, Type 2, and Type 3 diabetes. By incorporating components such as exploratory data analysis, longitudinal data tracking, and tailored predictive models, this framework not only enhances diagnostic accuracy but also provides actionable insights for personalized and proactive diabetes care. The following sections detail each component of the framework, emphasizing its role in improving prediction and risk assessment in clinical and non-clinical settings.

This framework outlines a comprehensive approach for diabetes prediction and risk assessment using machine learning techniques. Here's an explanation of each component (Depicted in Figure 2) :

Diabetes Dataset:

The process begins with the collection of diabetes-related datasets, which may include demographic data, clinical measurements, patient history, and biosignals from wearable devices.

Data Preprocessing

Data cleaning and transformation are performed to handle missing values, standardize variables, and prepare the dataset for analysis. This step ensures the quality and reliability of input data.

Exploratory Data Analysis (EDA)

Statistical and visual analysis of the dataset is conducted to uncover patterns, correlations, and trends. EDA helps identify relevant features for prediction and provides insights into data characteristics.

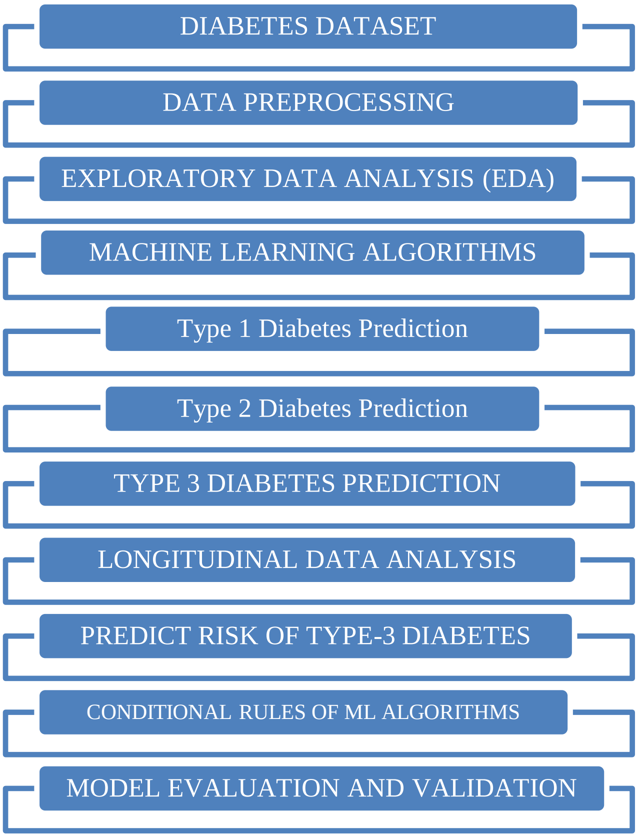

Fig. 2 Proposed Workflow for Diabetes Prediction and Risk Assessment Using Machine Learning

Machine Learning Algorithms

Various machine learning models are applied to predict diabetes outcomes. These algorithms are designed to analyze the processed data and classify diabetes into different types (Type 1, Type 2, and Type 3).

Type 1 and Type 2 Diabetes Prediction

Classification models are built specifically for predicting Type 1 and Type 2 diabetes. These models are trained on the dataset and optimized for accuracy.

Type 3 Diabetes Prediction

A dedicated component focuses on predicting Type 3 diabetes, which includes conditions like gestational diabetes or other diabetes-related cognitive impairments.

Longitudinal Data Analysis

This step involves analyzing time-series or longitudinal data to track diabetes progression and assess the risk over time, especially for chronic conditions.

Predict Risk of Type-3 Diabetes in Expectant Mothers

The framework evaluates the risk of Type 3 diabetes (such as gestational diabetes) in pregnant women using machine learning or conditional rules.

Conditional Rules or Machine Learning Algorithms

Either machine learning models or rule-based systems are applied to refine predictions and identify high-risk individuals with greater precision.

Model Evaluation and Validation

The final step involves assessing the model's performance using metrics such as accuracy, precision, recall, and F1-score. This ensures the robustness and reliability of the framework for real-world applications.

V.CONCLUSION, CHALLENGES, AND PATH FORWARD

This study emphasizes the significant impact of machine learning in tackling the increasing difficulties associated with diabetes prediction and management, especially concerning Type-3 diabetes. The suggested framework combines various data sources, sophisticated preprocessing methods, and cutting-edge machine learning models to improve prediction accuracy and deliver actionable insights for tailored care. This study establishes a foundation for scalable, real-world applications in diabetes care by tackling gaps like real-time monitoring, contextual modeling, and the integration of biosignal data. Moreover, the emphasis on ethical considerations and the clarity of AI-driven solutions highlight the framework's practical value for healthcare professionals and patients.

Nonetheless, various obstacles persist. The dependence on varied and high-quality datasets poses a notable challenge, especially in achieving generalizability among diverse populations. To build trust in AI-driven healthcare solutions, it is essential to systematically address ethical issues like data privacy, security, and algorithmic biases. Moreover, attaining real-time predictive capabilities and incorporating longitudinal data to enhance the understanding of disease progression presents intricate challenges that necessitate additional investigation.

In the future, it is essential to create adaptive models to facilitate real-time monitoring and tailored interventions. Improving the clarity of machine learning predictions using tools such as SHAP and LIME is crucial for their successful incorporation into clinical workflows. Furthermore, collaboration across various fields, including data science, medicine, and policy-making, will be essential in transforming these technological innovations into practical healthcare solutions. By tackling these challenges, this study lays the groundwork for a strong, patient-focused strategy for diabetes prediction and management, ultimately enhancing outcomes and alleviating the worldwide healthcare burden of this chronic condition.

References

[1]. K. Alnowaiser, "Improving Healthcare Prediction of Diabetic Patients Using KNN Imputed Features and Tri-Ensemble Model," in IEEE Access, vol. 12, pp. 16783-16793, 2024, doi: 10.1109/ACCESS.2024.3359760.

[2]. 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.

[3]. A. Site, J. Nurmi and E. S. Lohan, "Machine-Learning-Based Diabetes Prediction Using Multisensor Data," in IEEE Sensors Journal, vol. 23, no. 22, pp. 28370-28377, 15 Nov.15, 2023, doi: 10.1109/JSEN.2023.3319360.

[4]. 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.

[5]. U. Ahmed et al., "Prediction of Diabetes Empowered With Fused Machine Learning," in IEEE Access, vol. 10, pp. 8529-8538, 2022, doi: 10.1109/ACCESS.2022.3142097.

[6]. El Massari, H., Sabouri, Z., Mhammedi, S. and Gherabi, N., 2022. Diabetes prediction using machine learning algorithms and ontology. Journal of ICT Standardization, 10(2), pp.319-337.

[7]. Theis, J., Galanter, W.L., Boyd, A.D. and Darabi, H., 2021. Improving the in-hospital mortality prediction of diabetes ICU patients using a process mining/deep learning architecture. IEEE Journal of Biomedical and Health Informatics, 26(1), pp.388-399.

[8]. Daniels, J., Herrero, P. and Georgiou, P., 2021. A multitask learning approach to personalized blood glucose prediction. IEEE Journal of Biomedical and Health Informatics, 26(1), pp.436-445.

[9]. Khan, M.Z., Mangayarkarasi, R., Vanmathi, C. and Angulakshmi, M., 2022. Bio-Inspired PSO for Improving Neural Based Diabetes Prediction System. Journal of ICT Standardization, 10(2), pp.179-199.

[10]. Kim, N., Seo, W., Kim, N.H. and Park, S.M., 2022. Toward Personalized Hemoglobin A1c Estimation for Type 2 Diabetes. IEEE Sensors Journal, 22(23), pp.23023-23032.

[11]. Nasir, M.U., Gollapalli, M., Zubair, M., Saleem, M.A., Mehmood, S., Khan, M.A. and Mosavi, A., 2022. Advance genome disorder prediction model empowered with deep learning. IEEE Access, 10, pp.70317-70328.

[12]. Khan, F.A., Zeb, K., Al-Rakhami, M., Derhab, A. and Bukhari, S.A.C., 2021. Detection and prediction of diabetes using data mining: a comprehensive review. IEEE Access, 9, pp.43711-43735.

[13]. Siddiqui, H., Rattani, A., Woods, N.K., Cure, L., Lewis, R.K., Twomey, J., Smith-Campbell, B. and Hill, T.J., 2021. A survey on machine and deep learning models for childhood and adolescent obesity. IEEE Access, 9, pp.157337-157360.

[14]. Nuankaew, P., Chaising, S. and Temdee, P., 2021. Average weighted objective distance-based method for type 2 diabetes prediction. IEEE Access, 9, pp.137015-137028.

[15]. Fazakis, N., Kocsis, O., Dritsas, E., Alexiou, S., Fakotakis, N. and Moustakas, K., 2021. Machine learning tools for long-term type 2 diabetes risk prediction. ieee Access, 9, pp.103737-103757.

[16]. Sevil, M., Rashid, M., Hajizadeh, I., Park, M., Quinn, L. and Cinar, A., 2021. Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management. IEEE Transactions on Biomedical Engineering, 68(7), pp.2251-2260.

[17]. Ferdousi, R., Hossain, M.A. and El Saddik, A., 2021. Early-stage risk prediction of non-communicable disease using machine learning in health CPS. IEEE Access, 9, pp.96823-96837.

[18]. Hasan, M.K., Alam, M.A., Das, D., Hossain, E. and Hasan, M., 2020. Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access, 8, pp.76516-76531.

[19]. Islam, M.S., Qaraqe, M.K., Belhaouari, S.B. and Abdul-Ghani, M.A., 2020. Advanced techniques for predicting the future progression of type 2 diabetes. IEEE access, 8, pp.120537-120547.

[20]. Perveen, S., Shahbaz, M., Keshavjee, K. and Guergachi, A., 2018. Metabolic syndrome and development of diabetes mellitus: predictive modeling based on machine learning techniques. IEEE Access, 7, pp.1365-1375.

[21]. Pradhan, N., Rani, G., Dhaka, V. S., &Poonia, R. C. (2020). Diabetes prediction using artificial neural network. Deep Learning Techniques for Biomedical and Health Informatics, 121, 327–339. https://doi.org/10.1016/B978-0-12-819061-6.00014-8

[22]. Liu, J., Tang, Z. H., Zeng, F., Li, Z., & Zhou, L. (2013). Artificial neural network models for prediction of cardiovascular autonomic dysfunction

in general Chinese population. BMC Medical Informatics and Decision Making, 13(1). https://doi.org/10.1186/1472-6947-13-80

[23]. Temurtas, H., Yumusak, N., &Temurtas, F. (2009). A comparative study on diabetes disease diagnosis using neural networks. Expert Systems with Applications, 36(4), 8610–8615. https://doi.org/https://doi.org/10.1016/j.eswa.2008.10.032

[24]. Adeloye, D., Ige, J. O., Aderemi, A. V, Adeleye, N., Amoo, E. O., Auta, A., & Oni, G. (2017). Estimating the prevalence , hospitalisation and mortality from type 2 diabetes mellitus in Nigeria : a systematic review and meta-analysis. 1–16. https://doi.org/10.1136/bmjopen-2016-015424

[25]. "Types of diabetes," International Diabetes Federation. Available: https://www.idf.org/aboutdiabetes/what-is-diabetes.html.

[26]. "International Diabetes Federation—facts and figures," International Diabetes Federation. Available: https://www.idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html. Accessed: Mar. 23, 2021.

[27]. L. Ismail, H. Materwala, and J. Al Kaabi, "Association of risk factors with type 2 diabetes: a systematic review," Computational and Structural Biotechnology Journal, vol. 19, pp. 1753–1766, 2020, doi: 10.1016/j.csbj.2021.03.003.

[28]. "Complications of type 2 diabetes," American Diabetes Association, 2021. Available: https://www.diabetes.org/diabetes/complications.

[29]. L. Licitra, A. Trama, and H. Hosni, "Benefits and risks of machine learning decision support systems," JAMA: The Journal of the American Medical Association, vol. 318, no. 23, p. 2354, 2017, doi: 10.1001/jama.2017.16627.

[30]. L. Ismail, H. Materwala, A. Karduck, and A. Adem, "Requirements of health data management systems for biomedical care and research: scoping review," Journal of Medical Internet Research, vol. 22, no. 7, 2020, doi: 10.2196/17508.

[31]. A. Mujumdar and V. Vaidehi, "Diabetes Prediction using Machine Learning Algorithms," in Proceedings of the International Conference on Recent Trends in Advanced Computing, vol. 165, pp. 292–299, 2019.

[32]. American Diabetes Association; Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 1 January 2006; 29 (suppl_1): s43–s48. https://doi.org/10.2337/diacare.29.s1.06.s43

[33]. Jamison, D.T.; World Bank; Disease Control Priorities Project. (Eds.) Disease Control Priorities in Developing Countries, 2nd ed.;OCLC: Ocm64588990; Oxford University Press: Oxford, UK; World Bank: New York, NY, USA; Washington, DC, USA, 2006.

[34]. World Health Organization. Noncommunicable Diseases Country Profiles 2018; World Health Organization: Geneva, Switzerland, 2018.

[35]. Ellulu, M.S., Samouda, H. Clinical and biological risk factors associated with inflammation in patients with type 2 diabetes mellitus. BMC EndocrDisord 22, 16 (2022). https://doi.org/10.1186/s12902-021-00925-0

[36]. IDF Diabetes Atlas. Available online: https://diabetesatlas.org/ (accessed on Nov 20 2024).

[37]. Guariguata, L.; Whiting, D.R.; Hambleton, I.; Beagley, J.; Linnenkamp, U.; Shaw, J.E. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res. Clin. Pract. 2014, 103, 137–149. DOI: 10.1016/j.diabres.2013.11.002