Int. j. inf. tecnol.
https://doi.org/10.1007/s41870-025-02459-3

ORIGINAL RESEARCH

A novel deep learning model for early diabetes risk prediction using attention-enhanced deep belief networks with highly imbalanced data

Olusola Olabanjo[1,2,3] · Ashiribo Wusu[4] · Olufemi Olabanjo[5] · Mauton Asokere[3] · Oseni Afisi[6] · Boluwaji Akinnuwesi[7]

Received: 2 November 2024 / Accepted: 4 February 2025

Abstract Diabetes mellitus is a prevalent chronic illness with severe complications that demand timely diagnosis. This study introduces an attention-enhanced Deep Belief Network (DBN) for early diabetes risk prediction, designed to address challenges associated with highly imbalanced datasets. Using a dataset from Sylhet Diabetes Hospital, which includes symptom and demographic information from patients, we applied an ensemble feature selection approach to identify critical predictors. To address the class imbalance, Generative Adversarial Networks (GANs) were used to generate synthetic data, ensuring the model's robustness in identifying underrepresented cases. Additionally, a hybrid loss function combining cross-entropy and focal loss was implemented to improve classification, especially for hard-to-detect instances. Our results show that the attention-based DBN model, augmented with synthetic data from GANs and optimized with a hybrid loss function, achieves an AUC of 1.00, F1-score of 0.97, precision of 0.98, and recall of 0.95, outperforming several baseline models. This research offers a novel and effective approach for early diabetes detection, demonstrating potential for use as a clinical tool in preventive healthcare settings.

Keywords Deep belief networks (DBN) · Diabetes risk · Attention mechanism · Generative adversarial networks (GANs) · Imbalanced data · Hybrid loss function

1 Introduction

Diabetes mellitus (otherwise often referred to as diabetes) remains one of the popular life-threatening diseases which affects relatively 500 million people worldwide. It is a chronic disease associated with a high blood sugar level in a human's body [1, 2]. The pancreas is an organ in the human body that produces a special hormone known as insulin [3]. Insulin is released by the pancreas into the bloodstream, aiding in the transport of glucose into the cells [4]. Diabetes is a condition in which the pancreas is unable to make insulin or in which the body is unable to use insulin as it should. Due to its relatively long asymptomatic phase, its early detection has been receiving massive attention from both medical and non-medical scientists. Diabetes mellitus is known to manifest in two types: Type I and Type II [5, 6]. The former occurs when the pancreatic beta cells are mistakenly attacked by the immune system and the body produces too little – or none at all – insulin while in the case of the latter, the body does not produce enough or becomes actively resistant to insulin. The third, but not so common type of diabetes is

✉ Olusola Olabanjo
olola57@morgan.edu

1 Center for Equitable AI and Machine Learning Systems (CEAMLS), Morgan State University, Baltimore, MD 21251, USA

2 Department of Mathematics, Morgan State University, Baltimore, MD, USA

3 Department of Computer Science, Lagos State University, Lagos, Nigeria

4 Department of Mathematics, Lagos State University, Lagos, Nigeria

5 Department of Cybersecurity, University of Derby, Derby, UK

6 Department of Philosophy, Lagos State University, Lagos, Nigeria

7 Department of Computer Science, Faculty of Science and Engineering, University of Eswatini, Kwaluseni, Eswatini

Published online: 04 March 2025

gestational diabetes; the case of which a woman becomes diabetic during pregnancy due to hormonal changes. Diabetes mellitus is known to exhibit symptoms such as polyuria, polydipsia, polyphagia, sudden weight loss (usually Type I [7]), weakness, obesity (usually Type II [8]), delayed healing, visual blurring, itching, irritability, genital thrush, partial paresis, muscle stiffness, alopecia, etc [9].

The alarming fatality of this popular disease is evident from the facts that 85% of diabetic patients were from low- and middle-income countries and that its clinical detection takes so long that patients may gradually have started suffering from other diabetes-causing diseases such as heart attacks, stroke, hypertension, blurry vision, blindness, foot ulcer, amputation, kidney damage and other organ failures [10–12]. These symptoms set in due to the number of years (7–12) the disease has gone without notice or treatment. In fact, the degree of severity of its manifestation and associated complications correlates with its detection period. This makes early diagnosis, early commencement of treatment as well as early awareness of patient's risk factors to contribute to the reduction of its prevalence globally, thereby beneficial in terms of the patient's health and expenditure [13]. Identification of risk and protective factors is a key component in diseases which are incurable, confusable and takes a long time to manifest [14]. These factors promote awareness, prevents the disease, influence people's lifestyles towards avoiding the disease, fosters effective prevention and suggests routines that serve as positive countermeasures.

Several statistics- and machine-learning-based studies are being conducted daily to predict and diagnose diabetes [15–18]. The advent of technology has revolutionized many sectors including healthcare and medical technologies. It helps in the improvement of services offered to patients and serves as an efficient and effective measures of treating, diagnosing, service delivery, information handling, administration etc. [19, 20]. In the recent past, machine learning models have centered on the use of supervised deep learning and classical machine learning models for the prediction and determination of the Type-II diabetes risk factors. However, in this study, we propose an unsupervised approach to this study. A deep belief neural network is proposed to determine the risk factors of Type-II diabetes and the impact of ensemble feature selection was measured.

The application of machine learning models in non-evasive prediction of diabetes' risk factors have gained wide scholarly attention in the recent decades and this can be attributed to sophistication in compute devices and state-of-the-art machine learning algorithms. In this section, we zoom on various machine learning tools and algorithms that have been developed for the prediction of Type-II Diabetes mellitus and their accuracies are also discussed. Many notable risk assessment tools have been proposed, developed or/and deployed for a non-evasive determination of diabetes risk factors, some of which are Latin-America-FINDRISC (LAFINDRISC) [21], Risk Test by American Diabetes Association (RTADA) [22], Leicester Practice Risk Score [23], Test2Prevent [24]. These and many more have proven to be effective screener for assessing the risk of undiagnosed diabetes. The accuracies and applicable reliability of these tools are difficult to quantify because of the absence of Fasting Plasma Glucose (FPG) data or other related data. In terms of the core machine learning engine, many classical and state-of-the-art machine learning (ML) models have been proposed for a non-evasive early prediction of undiagnosed diabetes. They include, but are not limited to, individual models such as Artificial Neural Networks (ANNs) [25], k-Nearest Neighbors (kNN) [26, 27], Linear Regression [28], Logistic Regression [29], Naïve Bayes [30], Random Forests (RF) [31], Decision Trees (DT) [32], Support Vector Machine (SVM) [33], deep learning [34] among others. The maximum accuracy obtained for these classical models was 97.9%. Table 1 focuses on the accuracies and implementation details of some of the deep learning models which have been proposed for early risk factor detection of diabetes.

Structurally, our paper is organized into several key sections to present the research in a clear and logical manner. The Introduction section provides background information on diabetes mellitus and the importance of early detection. In the Related Works section, we review existing studies on diabetes prediction, focusing on machine learning and deep learning models, and identifying gaps that our work aims to address. The Methods section outlines the dataset, preprocessing steps, and the development of our Deep Belief Network (DBN) [46] model, including the novelty of the work. The Results and Discussion section presents the performance of our model compared to baseline models, discussing the significance of improvements like feature weighting and hybrid loss functions. Finally, the Conclusion summarizes our findings, emphasizing the contribution of our model to early diabetes risk prediction and its potential impact on medical practice.

2 Review and meta-analyses of related works

Many studies have been published in the application of machine learning and deep learning models to the risk prediction of diabetes mellitus. Table 2 provides a comprehensive comparison of recent studies and highlights different aspects of each study, such as the year, method or

Table 1 Compressed summary of some deep learning models for a non-evasive risk prediction of diabetes

References	Techniques / ML Models	Methodology	Major outcomes	Data sources
[35]	Denosing AE	Normalization, training (704,587), validation (5000) and testing (76,214)	Performance was measured using AUC (0.907)	Mount Sinai Data Warehouse (ICD-9)
[36]	Modified Long Short-Term Memory (LSTM), Attention pooling layer	training, validation and testing: 2/3, 1/6 and 1/6 respectively from 53,208 admissions	Study produced maximum accuracy of 79%	EHR data from hospital patients
[37]	Restricted Boltzmann machine (RBM) and Recurrent Neural Network (RNN)	Feature selection, Min–Max normalization, train (80%), test (20%)	Sensitivity and precision: 90.66%, 75% respectively	PID Data from the UCI Repository
[38]	Modified 1-D CNN and FC layer	The data for training and testing: 15 samples, 10 samples; leave-one out cross-validation	AUC of Type I-Diabetes, Type II—Diabetes, healthy subjects: 0.9659, 0.9625, 0.9644	Breath samples collected by MOS sensors with 1000-s intervals
[39]	CNN, LSTM, and SVM	Heart rate variability (HRV) data from 71 ECG datasets. fivefold cross-validation was used	Validation accuracy of 95.7% was obtained	ECG data sampled at 500 Hz from 40 subjects
[40–44]	Deep Multi-Layer Perceptron (DMLP)	Train-test split, data transformation, k-fold cross validation, normalization, feature selection	Maximum accuracy 88.41%, maximum AUC 84.13%, Sensitivity 87.92%, f1 Score 0.808	PID, Practice Fusion Dataset and HER dataset of
[45]	Deep Belief Network	Min–max normalization; feature selection by PCA; pre-training for RBMs; supervised fine-tuning	Sensitivity: 100%, F1 score: 0.808	Practice Fusion dataset (9948 patients, ICD-9)

techniques used, best-performing model, dataset details, feature engineering approaches, loss functions, and performance metrics like accuracy, AUC (Area Under the Curve), precision, and recall. For example, studies using Random Forest (RF) models, such as those by Jiang Liangjun et al. [47] and Islam MM et al. [48] report high accuracy and recall, indicating the robustness of RF in diabetes prediction tasks. Additionally, some studies like that of Lugner Moa et al. [49] use advanced techniques like Shapley Additive Explanations for feature selection, showcasing efforts to improve model interpretability and accuracy.

Furthermore, the table shows how different datasets and feature engineering techniques impact model performance. Some studies, such as those by Li Lin et al. [50] use extensive datasets (e.g., millions of instances) to build models like XGBoost, achieving high AUC and accuracy, which demonstrates the potential of using large, diverse datasets for model robustness. On the other hand, smaller datasets, like those used by Sisodia Deepti and Sisodia Dilip Singh [51] achieve lower performance metrics, underscoring the challenge of data limitations.

Further advances include the use of more sophisticated models and hybrid techniques. For example, recent studies by Li et al. (2023) and Bhat et al. (2022) explored the benefits of incorporating deep learning methods and boosting algorithms such as LightGBM and XGBoost. These models leverage large datasets and complex interactions within the data, thus offering higher predictive accuracy and robustness. Yet, our study differentiates itself by combining an ensemble feature selection approach with synthetic data generated via Generative Adversarial Networks (GANs), alongside an attention-based deep learning model. This combination addresses both the accuracy and class imbalance issues highlighted in previous works, establishing a comprehensive approach to early diabetes detection that surpasses traditional methods.

Figure 1 illustrates the impact of feature selection, data augmentation, and model complexity on performance metrics (Accuracy, Precision, Recall, and F1-Score) for various models as reported in the literature. In the first chart, models without feature selection exhibit slightly higher average scores across all metrics than those with feature selection, indicating that including all features may sometimes yield better predictive results. The second chart shows that models utilizing data augmentation consistently outperform those without it, especially in terms of Accuracy and F1-Score, demonstrating the benefit of data augmentation for handling imbalanced datasets. The third chart compares simple and complex models, revealing that complex models achieve substantially better performance across both Accuracy and F1-Score, thus highlighting the

Table 2 Summary of related works on diabetes risk prediction using machine learning / deep learning models

Author(s)	Year	Subject addressed	Method/ Techniques	Best Model	Dataset description	Feature engineering	Loss function	Performance reported	Citation
Suryadevara Chaitanya Krishna [52]	2023	Predict risk of patient been diabetic or not	Decision tree model, KNN, RF, Logistic regression	DT	Hospital of Frankfurt, Germany located on Kaggle with 2000 samples with 9 features	NA	Sigmoid function	Acc = 98%	35
Xu Weifeng et al. [31]	2017	Risk prediction of type II diabetes	RF, ID3 model, Naive Bayes model and AdaBoost model	RF	School of medicine, University of Virginia which has 373 instances and 10 features after pre-processing stage	Randomization selection of features using Gini criterion for data splitting and tenfold classification	Discretization on continuous features	Acc = 85%	130
Jiang Liangjun et al. [47]	2023	Prediction of risk of diabetes using association between key life characteristic indicators with machine learning	RF	RF	Guangzhou Haizhu District Community Health Development Guidance Center with diabetic patients follow-up records from 2016 to 2023 with 252,176 instances with 9 features	f_classif evaluation function, Recursive Feature Elimination (RFE) algorithm and rank ordering of feature	NA	Acc = 91.24%, Recall = 94.22%, AUC = 91.15%	16
Lu Haohui et al. [53]	2022	Developed a predictive model for T2DM using socio-demographic and behavioral information and network attributes	LR,KNN, SVM,NB,DT,RF, XGBoost and ANN	RF	CBHS health funds in Australia, 124,000 de-identified patients gathered from 1995 to 2018 with 10 features using three(3) to build the model	Network features(central centrality and Eigenvector centrality) and patient features using	bootstrap aggregation for ensemble DT's, maximum depth = 10, use entropy criterion and the number of estimators = 200	AUC = 91%	108
Kopitar Leon et al. [54]	2020	Prediction of undiagnosed T2DM	LR, Glmnet, RF, XGBoost, LightGBM	Glmnet	Ten (10) healthcare centers in Slovenia with 3723 instances with 58 indicators and 1 output	Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB)	RMSE	AUC = 82%	282

Table 2 continued

Author(s)	Year	Subject addressed	Method/Techniques	Best Model	Dataset description	Feature engineering	Loss function	Performance reported	Citation
Lugner Moa et al. [49]	2024	Identification of predictive factors for the development of type 2 diabetes	XGboost	XGboost	Dataset obtained from UK Biobank with comprehensive biomedical database with 502,625 instances and 419 features	Hapley Additive explanation, grid search algorithm and hypercube sampling	NA	AUC = 90%, Acc = 0.92, Sensitivity = 0.62 and Specificity = 0.93	6
Li Lin et al. [50]	2023	Prediction of diabetes risk assessment model	CART, LightGBM, RF, XGBoost, LR	XGBoost	National physical examination (NPE) project in 2020 with 3,774,084 healthy individuals and 3,013,47 T2DM patients	Pearson's Correlation Coefficient, Univariate logistic regression and Multivariate logistic regression	NA	AUC = 91%,acc = 83%, precision = 28%,PPV = 0.98% and NPV = 91%	5
Birjais Roshan et al. [55]	2019	Whether a patient has diabetes or not with machine learning	GB, LR and NB	GB	Pima Indian diabetes dataset with 768 instances and 8 number of attributes	correlation-based approach	Empirical risk minimization principle	Acc = 86%	106
Bhat Salliah Shafi et al. [56]	2022	Development of methodology for diabetes disease risk prediction	RF, MLP, SVM, GB, DT, and LR	RF	Dataset has 403 instances each with 11 attributes collected from April 2021 to Feb 2022	Correlation matrix	NA	Acc = 98%	77
Lama Lara et al. [57]	2021	To determine if machine learning can predict Increase in type 2 diabetes or prediabetes risk without known abnormal glucose regulation	RF	RF	Stockholm Diabetes Preventive Program (SDPP) with 8000 instances with 9 features	SHAP-TreeExplainer	NA	AUC = 78%	20

Table 2 continued

Author(s)	Year	Subject addressed	Method/Techniques	Best Model	Dataset description	Feature engineering	Loss function	Performance reported	Citation
Bhat Salliah Shafi et al. [58]	2023	Risk assessment and prediction for diabetes mellitus	LR, GB and DT	DT	Pima Indian diabetes data set of 768 patients from the University of California Irvine with 568 diabetic and 200 are not	feature importance with highest f-score		Acc = 91%, Precision = 96%,Recall = 92%, and Fi score = 94%	15
Bhat Salliah Shafi et al. [59]	2022	Diabetes mellitus prediction with the diagnosis information of the patients	SVM, RF and KNN	RF	PIMA Indians Diabetes Dataset (PIDD) with 9 attributes and 2000 instances	correlation feature selection	NA	Acc = 97.75%	14
Swain Aparimita et al. [60]	2016	Prediction and classification of occurrence of diabetes mellitus	ANN and ANFIS	ANFIS	100 samples of patients with 6 features from Bhubaneswar, Odisha, India	Multilayer Perceptron (MLP)	RMSE, MSE	Acc = 90.32%	33
Prasetyo Simeon Yuda et al. [61]	2024	Prediction of diabetes risk with machine learning	Gaussian Naive Bayes, DT, and ANN	ANN	Behavioral Risk Factor Surveillance System (BRFSS) dataset with 253,680 individuals and 22 features	Standard Scaler	NA	Acc = 84.73%	0
Lai Hang et al. [62]	2019	Identification of Canadian patients at risk of having Diabetes Mellitus based on demographic information and laboratory results	LR, GB, DT and RF	GB	Canadian Primary Care Sentinel Surveillance Network (CPCSSN) with over 880,000 instances, 13,309 instances after preprocessing with 8 features	NA	Bernoulli loss function and tree-based learners	AUC = 84.7%, Sensitivity = 71.6%	264

Table 2 continued

Author(s)	Year	Subject addressed	Method/Techniques	Best Model	Dataset description	Feature engineering	Loss function	Performance reported	Citation
Samet Sarra et al. [63]	2021	Gives probability of developing diabetes mellitus using machine learning	RF,SVM, Bayesian Network (BN),KNN,DT, ANN, and LR, Hybrid model(Stacking)	Hybrid model	National Institute of Diabetes and Digestive and Kidney collected the dataset provided the Pima India Diabetes Dataset with 768 instances with 8 features	Correlation	NA	Precision = 91%,Recall = 91%,F1-score = 90%,Acc = 90.62%	11
Zarar Muhammad and Wang Yulin [64]	2023	Predict likelihood of diabetes occurrence	SVM, Decision Forest, LR, and ANN	ANN	PIMA Indians Diabetes Dataset (PIDD) with 9 attributes and 768 instances and Kaggle dataset (2000 instances with 9 features)	NA	NA	Acc = 98.8%	3
Khanam Jobeda Jamal and Foo Simon Y [65]	2021	Early prediction of diabetes mellitus	DT, KNN, RF, NB, AB, LR, SVM, and neural network (NN)	NN	Pima Indian Diabetes (PID) dataset with 768 patients and their 9 unique attributes	Pearson's correlation		Acc = 88.6%	341
Sisodia Deepti and Sisodia Dilip Singh [51]	2018	Predict the likelihood of occurrence of diabetes mellitus	DT, SVM and NB	NB	Pima Indians Diabetes Database (PIDD) which is sourced from UCI machine learning repository with 768 instances with 8 features with one target class	NA	NA	Acc = 76.30%,AUC = 82%,Recall = 76%,f1-score = 76%	922
Kaluarachchi KN, Premachandra KP and Dissanayake RBN [66]	2023	Low bias-variance trade-off for the diagnosis of type II diabetes with machine learning	LR, DT, RF, NB, KNN, SVM, GB, and ANN	RF	PIMA Indian Diabetes Dataset	NA	MSE	Acc = 83.12%	1

Table 2 continued

Author(s)	Year	Subject addressed	Method/Techniques	Best Model	Dataset description	Feature engineering	Loss function	Performance reported	Citation
Akther Mahnur et al. [67]	2023	Assessing one's diabetes risk prior to medical diagnosis with two datasets	SVM, RF, NB, DT, KNN, LR Adaboost, GB, XGBoost, and MLP	RF, GB	PIMA Indian with 768 instances with 8 features and a target class and Sylhet datasets. Dataset consist of 16 features with a target class with 720 instances	Correlation Matrix, Label Encoding	NA	Acc = 84.21%,Acc = 98.85%	2
Islam Md Rifatul et al. [48]	2023	Building and validation of ML model to detect diabetes in early stage	LR, KNN, Gaussian NB, RF,NB, SVM, DT,GB, Light GBM, XGBoost, AdaBoost, Categorical Boosting, MLP	GBM	Collected from Sylhet Diabetic Hospital situated in Sylhet, Bangladesh used by Early Stage Diabetes Risk Prediction with 520 instances with 17 features	Correlation Matrix	deviance	F1-scoce = 99.37% and AUC = 99.92%	7
Uddin Md Jamal et al. [68]	2023	Diabetes risk variables to examine the prevalence of diabetes in Bangladesh	DT, LR, SVM, GB, XGBoost, RF, and ensemble technique (ET	ET	Dataset from Bangladeshi population with 508 instances and 32 features	Recursive feature elimination (RFE)	lowest log-loss	Acc = 87.45%, Precision = 87.05%, Recall = 88%, ROC-AUC = 0.8745, F1-score = 87.52%, GM = 87.45%	18
Islam MM et al. [69]	2020	Predict the likelihood of having diabetes mellitus	NB, LR, and RF,DT	RF	Collected from Sylhet Diabetic Hospital situated in Sylhet, Bangladesh used for risk Prediction with 520 instances with 17 features	NA	NA	Acc = 99%	255
Islam Taminulet al. [70]	2023	Evaluate the effectiveness of machine learning model in predicting risk of diabetes mellitus	GNB, RF, SVM, LR, and the DT	RF	Diabetic patients in Bangladesh,3837 patient records in the dataset, 3057 of which correspond to affected cases and 396 were normal with 17 features	Bagging for DT	NA	Acc = 98%	1

Table 2 continued

Author(s)	Year	Subject addressed	Method/Techniques	Best Model	Dataset description	Feature engineering	Loss function	Performance reported	Citation
Xu Yundong and Nie Ying [71]	2024	Prediction of T2DM risk with SVM	SVM	SVM	1000 samples for the prediction	Data Correlation and Multicollinearity Analysis, Recursive Feature Elimination and Regularization	NA	F1-Score = 0.7968, Precision = 0.7786, Recall = 0.816	0
Nipa Nurjahan et al. [72]	2024	Machine learning based model that predict T2DM from early stage	LR, KNN, SVM, NB, DT,RF, stochastic gradient descent (SGD), Perceptron, AdaBoost, XGBoost, MLP, HGBC	ET, MLP, HGBC and LGBM	520 patients records from the University of California, Irvine (UCI) machine learning repository of Sylhet Diabetes Hospital, Sylhet and 558 patients records from Bangladesh	NA	NA	Acc = 97.11%,Acc = 96.42%, Acc = 94.90%	7

T2DM Type 2 diabetes mellitus (T2DM), *LR* Logistic Regression, *DT* Decision Tree, *KNN* K-Nearest Neighbours, *SVM* Support Vector Machine, *NB* Naive Bayes, *RF* Random Forest, *ANN* Artificial Neural Network, *Glmnet* Regularized Generalized Linear Model, *LightGBM* Light Gradient Boosting Machine, *XGBoost* eXtreme Gradient Boosting, RMSE—Root Mean Squared Error, CART classification and regression tree, *PPV* Positive Predictive Value, *NPV* Negative Predictive Value, *GB* Gradient Boosting, *Acc* Accuracy, *MLP* Multi-Layer Perceptron, *ANN* Artificial Neural Network, *ANFIS* Hybrid Adaptive Neuro-Fuzzy Inference, *RMSE* Root Mean Squared Error, *MSE* Mean Squared Error, *AUC* Area under the receiver operating characteristic curve, *ML*-Machine learning, *GM* Geometric mean, *GNB* Gaussian Naive Bayes, System, *HGBC* Hist gradiant boosting classifier, Light gradient boosting machine (LGBM), *NA* Not Applicable

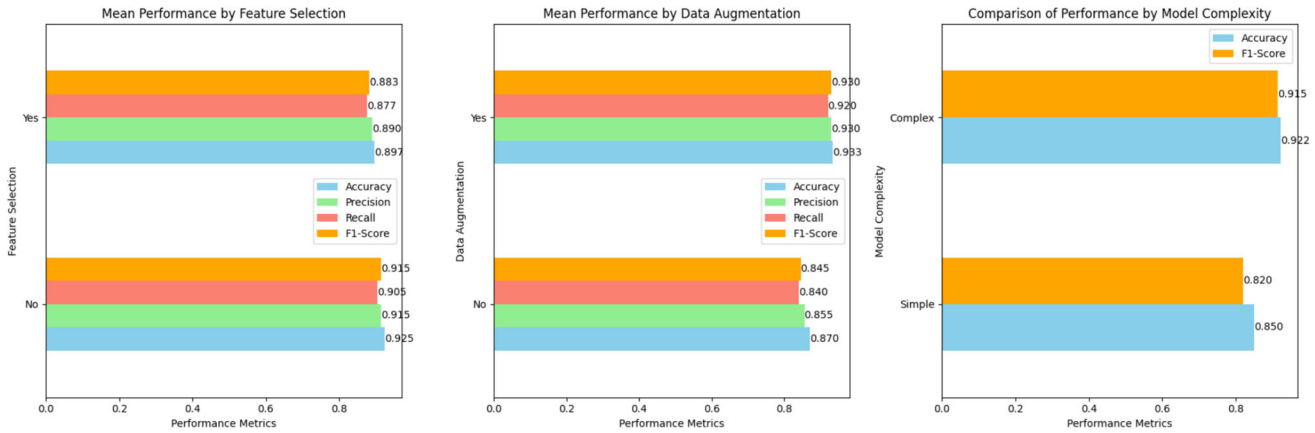

Fig. 1 Comparison of average reported model performance metrics based on **a** feature selection, **b** data augmentation and **c** model complexity

trade-off between model simplicity and predictive power. In this analysis, simple models refer to those with straightforward structures and fewer parameters, such as Logistic Regression, while complex models include advanced architectures with more layers or parameters, like Random Forest, XGBoost, Deep Belief Network (DBN), and LightGBM, which generally offer higher predictive performance.

The radar chart (Fig. 2) visualizes the performance of the top five models (Logistic Regression, Random Forest, XGBoost, Deep Belief Network, and LightGBM) across four metrics: Accuracy, Precision, Recall, and F1-Score. Each model's polygon shape indicates its strengths, with models like DBN and XGBoost extending further out, showing higher performance across most metrics. This allows a quick comparison to see which models are more balanced across all metrics or excel in specific areas.

The review of related works showed some areas of challenges which we seek to address in this paper. One of the key issues identified is the imbalance in the dataset, where the majority of instances belong to the non-diabetic class. This imbalance causes traditional machine learning

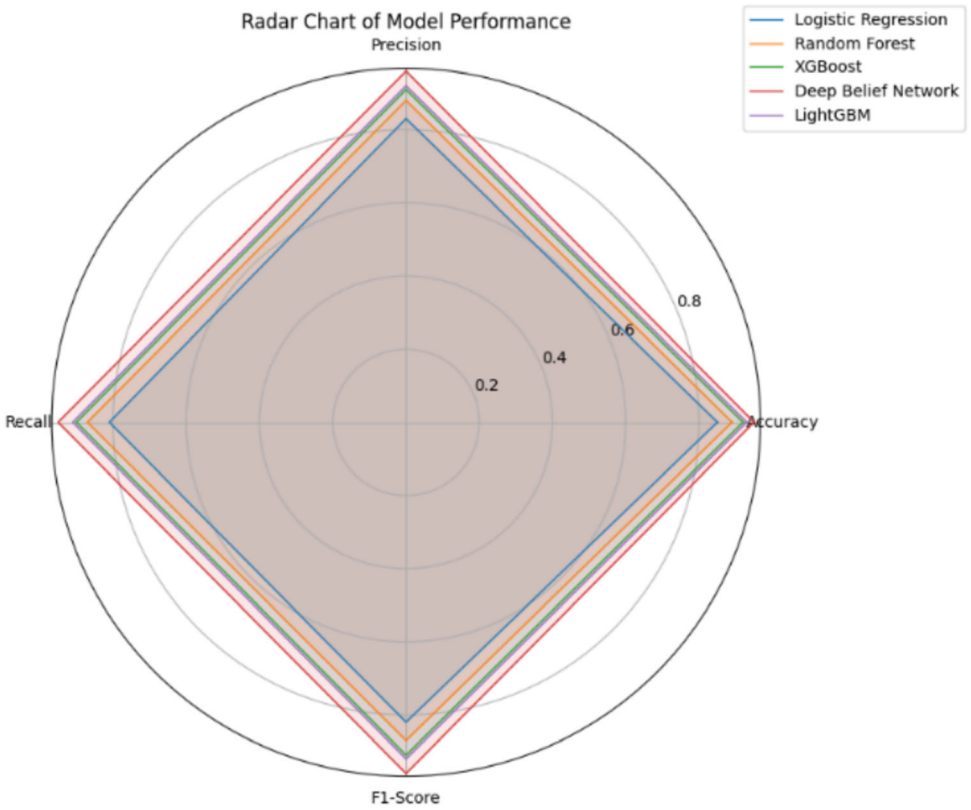

Fig. 2 Radar chat of model performance as reported in the literature

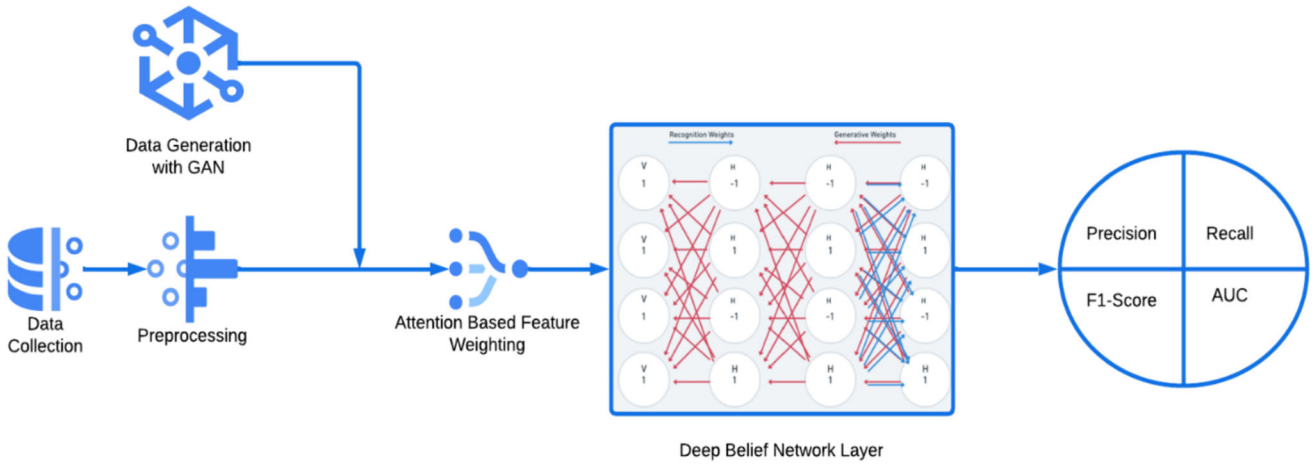

Fig. 3 Architecture of proposed model

models to perform poorly on minority classes [73, 74], leading to a high rate of false negatives [75, 76]. To address this, we integrated Generative Adversarial Networks (GANs) for data augmentation [77]. GANs generate synthetic data for the minority class helping to balance the dataset and ensure that the model is trained on a more representative sample [78]. This approach seeks to improve the model's ability to detect diabetes in underrepresented cases, ensuring a lower rate of false negatives and better overall classification performance.

Another challenge identified in the literature is the inability of traditional models to assign varying importance to different features during the learning process. In diabetes prediction, not all symptoms contribute equally to the outcome, and the model needs to prioritize more relevant features. To address this, we introduced an attention mechanism to dynamically weight features based on their importance to the classification task [79]. By calculating attention scores for each feature, the model focuses more on critical indicators and less on irrelevant or noisy features. This is intended to improve the interpretability of the model and ensures that critical symptoms for early diabetes detection are given priority in the decision-making process.

Lastly, we observed that existing models often struggle with optimizing performance when handling imbalanced data, leading to suboptimal results for minority classes. Traditional loss functions like cross-entropy tend to be biased towards the majority class [80]. To mitigate this, we implemented a hybrid loss function [81, 82], combining cross-entropy with focal loss. Focal loss adjusts the

Table 3 Description of Data Features

SN	Attributes	Datatype	Yes (as 1)	No (as 0)
	Age	$20 \leq$ Age ≤ 100		
	Sex	Male and Female	Male (328)	Female (192)
	Polyuria	Yes/No	258	262
	Polydipsia	Yes/No	233	287
	Sudden Weight Loss	Yes/No	217	303
	Weakness	Yes/No	305	215
	Polyphagia	Yes/No	237	283
	Genital Thrush	Yes/No	116	404
	Visual Blurring	Yes/No	233	287
	Itching	Yes/No	253	267
	Irritability	Yes/No	126	394
	Delayed Healing	Yes/No	239	281
	Partial Paresis	Yes/No	224	296
	Muscle Stiffness	Yes/No	195	325
	Alopecia	Yes/No	179	341
	Obesity	Yes/No	88	432
	Class	Positive/Negative	Positive (320)	Negative (200)

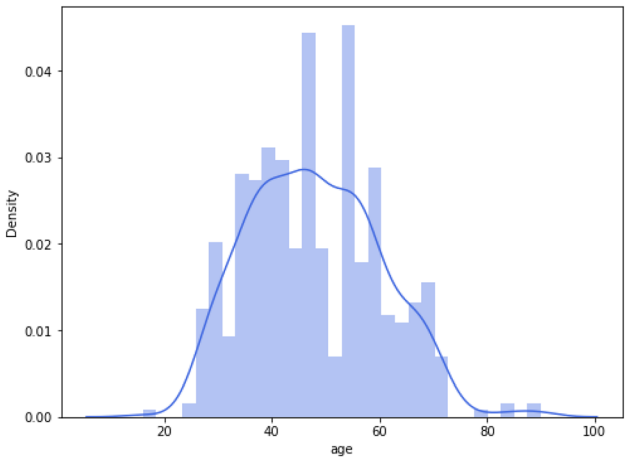

Fig. 4 Frequency Distribution of ages of subjects

model's focus by reducing the impact of easily classified samples, allowing the model to give more attention to harder-to-classify instances. This is intended to ensure a more balanced training process, improving both precision and recall, particularly for the minority class. Together, these improvements aim to enhance the model's robustness and accuracy in early diabetes detection.

3 Methods

3.1 Architecture of proposed model

The workflow of our proposed deep learning model for early diabetes risk prediction as seen in Fig. 3 comprises a sequence of processes designed to address the challenges of data imbalance, feature relevance, and model accuracy. Our model leverages several enhancements, including Generative Adversarial Networks (GANs) for data augmentation, an attention mechanism for prioritizing features, and a hybrid loss function that combines cross-entropy with focal loss. These components are integrated into the system workflow, which unfolds in several key stages.

The first stage, data preprocessing, ensures that the dataset is clean, standardized, and optimized for the training process. Categorical features, such as gender and

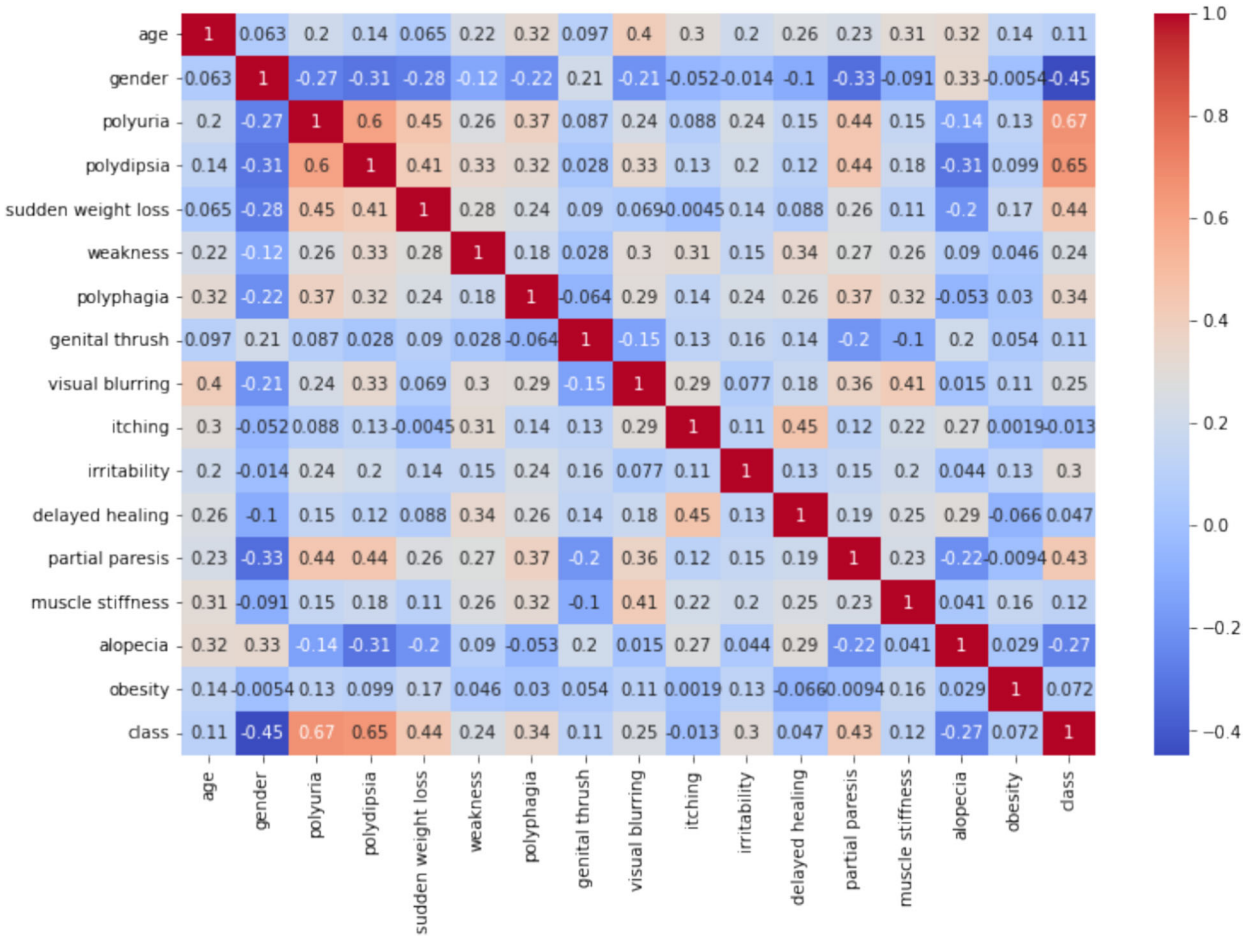

Fig. 5 Correlation heatmap of dataset features

symptom indicators, are encoded as binary variables, which simplifies model interpretation and processing. Additionally, continuous features like age are normalized to ensure consistent scaling across all input features, which allows the model to evaluate each feature's importance more effectively without any single feature dominating due to its scale. To address the dataset's inherent imbalance between diabetes-positive and diabetes-negative cases, synthetic data generation is performed using a Generative Adversarial Networks (GAN). This involves training a generator and a discriminator, where the generator produces synthetic samples that mimic the minority class data, and the discriminator attempts to distinguish between real and synthetic samples. By incorporating these synthetic data samples into the dataset, the system achieves a balanced representation, which aids the model in recognizing patterns within underrepresented classes and improves generalization in predictions. The system also incorporates an attention mechanism that dynamically assigns weights to each feature based on its relevance to diabetes prediction. This process calculates an attention score for each feature, effectively prioritizing indicators that are more closely correlated with diabetes outcomes. The attention scores are then used to create a context vector, a weighted combination of all features, which is passed into the Deep Belief Network. This approach allows the model to focus on critical features while reducing the impact of less relevant information, thereby improving the predictive accuracy.

The next phase involves training the DBN model itself. The DBN undergoes layer-wise pretraining through Restricted Boltzmann Machines (RBMs), providing a strong foundation for processing high-dimensional data. Once pretraining is complete, the DBN is fine-tuned using backpropagation to minimize classification errors, refining the model's ability to accurately predict diabetes risk. To further enhance model performance, a hybrid loss function is applied, combining cross-entropy and focal loss. This hybrid approach allows the model to focus more on hard-to-classify instances by adjusting the weight given to samples based on their classification difficulty. This adjustment improves recall and reduces the rate of false negatives, making the model more reliable for detecting positive cases within imbalanced datasets.

The final stage involves evaluating the model's performance using key metrics such as precision, recall, F1-score, and AUC. These metrics offer a comprehensive view of the model's accuracy and robustness, particularly in early diabetes detection. Through this systematic workflow, our proposed system achieves high accuracy and interpretability, making it a valuable tool for predicting diabetes risk in complex medical datasets.

3.2 Dataset and preprocessing

The diabetes dataset used in this study was obtained from Kaggle (https://www.kaggle.com/datasets/ishandutta/early-stage-diabetes-risk-prediction-dataset) and it contains the response obtained from 520 subjects (who recently became diabetic or are currently showing symptoms of diabetes) using a direct questionnaire and approved by a medical doctor. This was released by Sylhet Diabetes Hospital of Sylhet, Bangladesh. It consists of the age, sex, Boolean response to each diabetes-related question and the class to which each person belongs after medical diagnosis (Positive or Negative). There are 16 attributes for each subject under consideration, the summary of which is presented in Table 3.

Figure 4 depicts the histogram showing the frequency distribution of the ages of subjects in the dataset, overlaid with a kernel density estimate (KDE) to provide a smooth estimate of the probability density function. The distribution is roughly centered around middle-aged individuals, with most subjects falling between the ages of 40 and 70. There are fewer subjects in the younger (under 20) and older (above 80) age ranges, as shown by the reduced frequencies at the tails of the distribution. This plot provides an overview of the age demographics in the dataset, which can be relevant for understanding age-related trends in diabetes prediction. The heatmap shown in Fig. 5 illustrates the correlation coefficients between the different features in the dataset. The correlation values range from -1 to 1, where values close to 1 indicate a strong positive correlation, and values close to -1 indicate a strong negative correlation. Features such as polyuria and polydipsia exhibit a strong positive correlation with the target class (diabetes diagnosis), indicating their significant relevance in predicting diabetes. In contrast, features like visual blurring and weakness show lower correlations. Negative correlations, such as between alopecia and the target class, suggest an inverse relationship. This heatmap helps in identifying multicollinearity between features, guiding the feature selection and model interpretation process.

The preprocessing stage of our diabetes prediction model begins with data cleaning and encoding categorical variables. The dataset contains both categorical variables (e.g., gender) and a continuous variable (age). We first ensure that all missing values are addressed, either by imputing them based on the most frequent values or by removing rows if necessary. Outliers in the continuous variable (age) are identified and handled using capping techniques. For categorical features such as gender, where values are "Male" and "Female", we apply binary encoding: Male is encoded as 1 and Female as 0. For the binary symptom variables, which consist of "Yes" and

"No", we encode "Yes" as 1 and "No" as 0. Formally, for any categorical feature X with categories [A, B] the encoding is defined as Eq. (1):

$$X\prime = \begin{cases} 1 \; if \; X = A \\ 0 \; if \; X = B \end{cases} \tag{1}$$

Next, we address feature scaling, especially for the continuous variable, age. Since age values vary significantly while the other features are binary, it is important to normalize age to ensure it does not dominate during training. We use Min–Max Normalization, which scales the non-categorical features to a range between 0 and 1, allowing it to be on the same scale as the binary features. The Min–Max normalization formula is given in Eq. (2):

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

where X is the original feature value, X_{min} and X_{max} are the minimum and maximum values of the features in the dataset respectively. This transformation ensures that all features contribute equally to the learning process.

3.3 Model development and prediction enhancements

In developing our early diabetes risk prediction model, we aimed to address key challenges associated with imbalanced data and feature importance in medical diagnosis. To achieve this, a novel Deep Belief Network (DBN) architecture enhanced by an attention mechanism was designed, which dynamically assigns weights to features based on their relevance to the prediction task. This approach allows the model to focus on the most important features, such as critical symptoms or demographic factors, while minimizing the influence of less relevant data. Additionally, the DBN was pre-trained using Restricted Boltzmann Machines (RBMs) and fine-tuned with backpropagation, providing a robust framework for high-dimensional data.

To further improve the model's handling of imbalanced data, we integrated Generative Adversarial Networks (GANs) for synthetic data augmentation, specifically generating additional samples for underrepresented classes, such as early-stage diabetes patients. This data augmentation step helps balance the dataset and improves the model's generalization capability. We also introduced a hybrid loss function that combines cross-entropy with focal loss, designed to mitigate the effects of class imbalance by focusing more on hard-to-classify examples. This combination of techniques ensures that our model is both accurate and resilient in detecting diabetes, even in challenging imbalanced datasets. Our improvements to the classical DBN model are discussed below:

3.3.1 Attention mechanism for feature weighting

The attention mechanism dynamically assigns weights to each feature based on its relevance to the prediction task. This process helps the DBN focus on the most critical features. Let $X = \{x_1, x_2, \ldots, x_n\}$ be the input feature vectors, where n is the number of selected features. For each feature x_i, we calculate an attention score, e_i, using a learnable weight matrix $\boldsymbol{W}_a$ and bias vector $\boldsymbol{b}_a$ such that Eq. (2) holds.

$$e_i = tanh(\boldsymbol{W}_a \cdot x_i + \boldsymbol{b}_a) \tag{3}$$

The attention weights, α_i, are then computed using the softmax function over the attention scores as given in Eq. (4)

$$\alpha_i = \frac{e^{e_i}}{\sum_j e^{e_j}} \tag{4}$$

The final context vector, c, which is a weighted combination of the input features is given in Eq. (5)

$$c = \sum_i \alpha_i x_i \tag{5}$$

which is then passed into our DBN for further processing.

3.3.2 DBN with attention mechanism

The DBN is a stack of Restricted Boltzmann Machines (RBMs)[83], where each RBM is trained as a layer-wise generative model. After the attention mechanism, the context vector c becomes the input to the DBN. Our RBM consists of visible units v and hidden units h. The energy function $E(v, h)$ is given in Eq. (6):

$$E(v, h) = -\sum_i v_i b_i - \sum_j h_j c_j - \sum_{i,j} v_i h_j w_{ij} \tag{6}$$

where v_i are the visible units, h_j are the hidden units, b_i and c_j are the biases for visible and hidden layers respectively, and w_{ij} are the weights between the visible unit v_i and the hidden units h_j. The conditional probabilities for the visible and hidden layers are respectively given in Eq. (7) and Eq. (8).

$$P(v_i = 1|h) = \sigma\left(b_i + \sum_j h_j w_{ij}\right) \tag{7}$$

$$P\left(h_j = 1|v\right) = \sigma\left(c_j + \sum_i v_i w_{ij}\right) \tag{8}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid activation function. Our proposed DBN has three RBMs with 250 units, 250 units and 500 units in the first, second and third hidden layers. The DBN performs layer-wise pretraining using contrastive divergence and is fine-tuned using backpropagation.

Table 4 Results of voting ensemble feature selection method on the Dataset (✘ means disqualified, while ✔ means qualified, ✔✔ means strongly qualified)

SN	Attributes	Chi square	Mutual information gain	variance threshold	Voting
	Age	✘	✔	✘	✘
	Sex	✔	✔	✔	✔✔
	Polyuria	✔	✔	✔	✔✔
	Polydipsia	✔	✔	✔	✔✔
	Sudden Weight Loss	✔	✔	✔	✔✔
	Weakness	✔	✔	✔	✔✔
	Polyphagia	✔	✔	✔	✔✔
	Genital Thrush	✘	✔	✔	✔
	Visual Blurring	✔	✘	✔	✔
	Itching	✔	✘	✘	✘
	Irritability	x	✔	✔	✔
	Delayed Healing	✔	✘	✔	✔
	Partial Paresis	✔	✔	✘	✔
	Muscle Stiffness	✔	✔	✔	✔✔
	Alopecia	✔	✔	✔	✔✔
	Obesity	✘	✘	✘	✘

Table 5 Performance analysis of selected classical ML models

Models	F-Measure			Recall			Precision		
	All	13 Best	8 Best	All	13 Best	8 Best	All	13 Best	8 Best
Decision Tree	0.72	0.86	0.86	0.62	0.69	0.84	0.72	0.61	0.88
Random Forest	0.79	0.77	0.69	0.76	0.72	0.77	0.65	0.7	0.87
Logistic Regression	0.86	0.77	0.77	0.59	0.72	0.69	0.67	0.78	0.91
Support Vector Machine	0.66	0.89	0.77	0.86	0.69	0.88	0.58	0.68	0.83
k-Nearest Neighbors	0.72	0.89	0.86	0.74	0.66	0.88	0.72	0.8	0.91

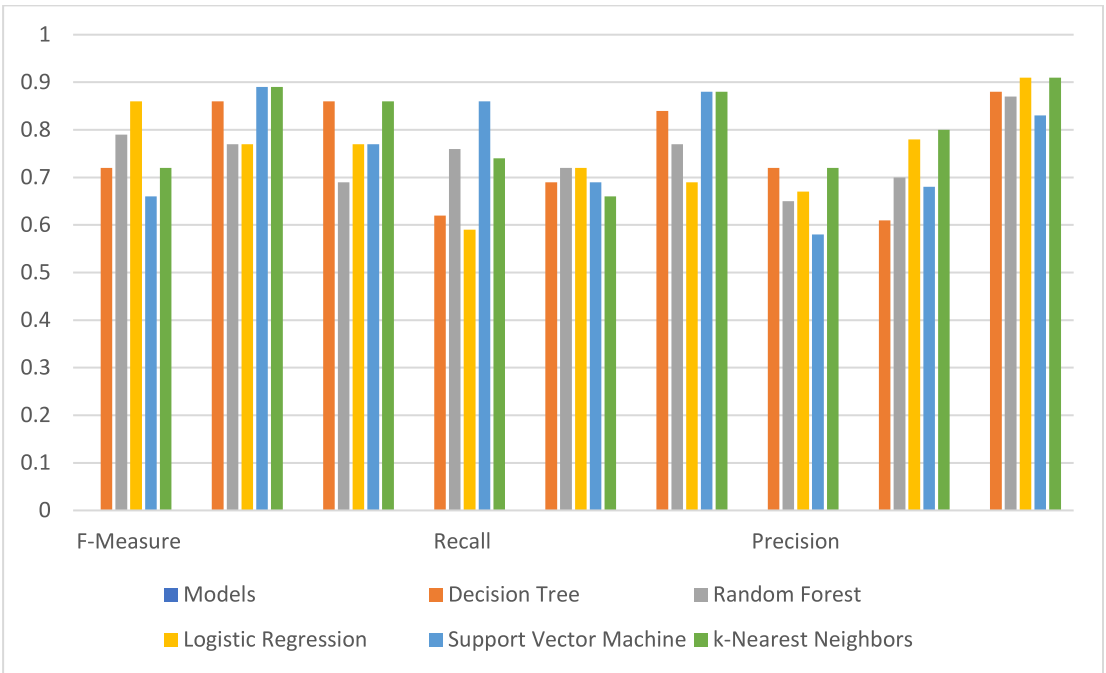

Fig. 6 Performance result of the classical ML models for diabetes prediction

Table 6 Performance of our incremental experiments with DBN

Experimental setups	Performance metrics			
	Precision	Recall	F1-Score	AUC
Exp. 1 → Classical DBN	0.85	0.82	0.83	0.87
Exp. 2 → Exp 1 + GAN	0.89	0.88	0.88	0.91
Exp. 3 → Exp 2 + Attention Mechanism	0.91	0.89	0.91	0.93
Exp. 4 → Exp 3 + Hybrid Loss Function	0.98	0.95	0.97	1.00

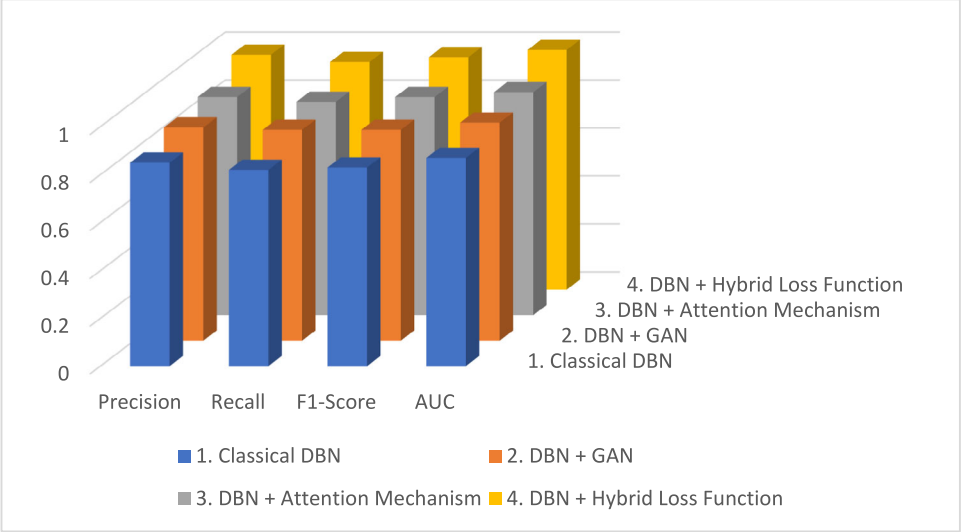

Fig. 7 Diagrammatic representation of the performance metrics

3.3.3 *Synthetic data augmentation with generative adversarial network (GAN)*

The dataset used in this study is highly imbalance and to address the class imbalance, we used a Generative Adversarial Network (GAN)[84] to generate synthetic data to augment our dataset. The generator, G, takes in a random noise vector $z \sim P_z$ and generates $G(z)$ synthetic samples. $D(x) = P(real|x)$ is the discriminator which receives both real data x and generated data $G(z)$ and tries to classify whether the data is real or synthetic. The generator and discriminator play a minimax game such that Eq. (9) holds:

$$\min_G \max_D \underset{x \sim P_{data}}{\vee E} \left[log\, D(x) + \underset{Z \sim P_Z}{\vee E} [(\log(1 - D(z)))] \right] \quad (9)$$

3.3.4 *Hybrid loss function*

To further handle the imbalanced nature of the diabetes dataset, a hybrid loss function is used, combining cross-entropy loss and focal loss. This loss function ensures the model pays more attention to hard-to-classify cases. The cross-entropy loss is given as Eq. (10)

$$L_{CE} = -\sum_i y_i \log p_i \quad (10)$$

where y_i is the true label for sample i, p_i is the predicted probability for the correct class. The focal loss adjusts the contribution of each sample based on the confidence of the prediction such that Eq. (11) holds:

$$FL(p_t) = -(1 - p_t)^{\gamma} \log p_t \quad (11)$$

where p_t is the predicted probability for the correct class and γ is the focusing parameter which helps to reduce the weight of well-classified examples and emphasizes hard-to-classify ones. The hybrid (final) loss L is a combination of cross-entropy and focal loss given in Eq. (12):

$$L = \alpha L_{CE} + (1 + \alpha) FL \quad (12)$$

where α is a hyperparameter which balances the two loss functions and can be tuned based on validation performances to optimize the trade-off between the two loss components.

By integrating the attention mechanism, synthetic data augmentation using GANs, and a hybrid loss function, our improved model becomes more robust, and better suited for imbalanced datasets.

Algorithm 1 Improved DBN with Attention Mechanism and GAN-based Data Augmentation for Diabetes Prediction

Improved DBN with Attention Mechanism and GAN-based Data Augmentation for Diabetes Prediction

1. **Preprocessing**
   Handle missing values (remove or impute).
   Encode categorical features (Yes/No to 1/0, Male/Female to 1/0).
   Normalize continuous features (e.g., Age) using Min-Max scaling.

2. **Initialize Attention Mechanism**
   Input: Feature matrix X for each patient.
   For each feature x_i:
   Compute attention score: $e_i = tanh(W_a * x_i + b_a)$
   Normalize attention scores: $\alpha_i = softmax(e_i)$
   Compute context vector: $c = sum(\alpha_i * x_i)$

3. **DBN Pretraining**
   Pretrain each layer in the Deep Belief Network (DBN) using Restricted Boltzmann Machines (RBMs).
   Use contrastive divergence to update the weights of each RBM.

4. **Fine-tuning the DBN**
   After pretraining, fine-tune the DBN using backpropagation to minimize classification error.

5. **GAN-based Data Augmentation**
   Initialize a generator $G(z)$ and discriminator $D(x)$.
   Generate synthetic samples using the generator: $G(z)$ = synthetic sample.
   Train the discriminator $D(x)$ to classify real and synthetic samples.
   Augment the dataset with synthetic samples from the GAN.

6. **Hybrid Loss Function**
   Compute Cross-Entropy Loss:
   $L_{CE} = -sum(y_i * log(p_i))$
   Compute Focal Loss:
   $FL(p_t) = -(1 - p_t)^{\gamma} * log(p_t)$
   Combine into hybrid loss: $L = \alpha * L_{CE} + (1 - \alpha) * FL$.

7. **Model Training**
   Train the DBN using the hybrid loss function and augmented dataset.

8. **Model Evaluation**
   Evaluate model performance on the test set using Precision, Recall, and F1-score.

9. **Output**
   Output the predicted class probabilities (diabetes-positive/negative).

4 Results and discussion

4.1 Results of baseline models

In this study, we developed a voting ensemble feature selection method which consisted of Chi-Square (CS), Mutual Information Gain (MIG) and Variance Threshold (VT) methods. Top ten methods were selected and prepared to run in the DBN pretraining for the prediction of diabetes mellitus. The parameters were tuned to achieve the most optimal accuracy obtainable. The top ten feature sets were then passed through five benchmark models (KNN, Linear SVM, Logistic Regression, Decision Trees and Random Forests) for performance comparison. We also performed correlation analysis by plotting the correlation matrix in order to determine prior to modeling if there is any overfitting. The categorical nature of the diabetes dataset required that Spearman correlation be used and not Pearson [85]. The result of the voting ensemble feature selection process is given in Table 4. The ensemble voting screened out age, itching and obesity as possible early predictors of diabetes mellitus with obesity having the lowest rank by our three feature rankers. The feature sets which were voted by our stack selectors are sex, polyuria, polydipsia, sudden weight loss, weakness, polyphagia, genital thrush, visual blurring, irritability, delayed healing, partial paresis, muscle stiffness and

Table 7 Comparison of Machine Learning Approaches for Diabetes Prediction and the Proposed Deep Belief Network (DBN) Model

Authors	Problem Addressed	Strengths	Comparison with proposed DBN
Palabaş Tugba [89]	Early detection of diabetes mellitus modeled with efficient machine learning techniques	Use ensemble techniques to improve the performance of the model	Improved DBN was built with ensemble feature techniques which gave an increase in model performance
Yadu Shweta, Chandra Rashmi, Sinha Vivek Kumar [90]	Finding the most effective method for risk prediction of diabetes mellitus	Applied graph analysis technique on dataset with random forest model for improved performance	Improve accuracy with a voting ensemble feature selection method for DBN
Taminul Islam et al. [70]	Early risk prediction of diabetes mellitus with evaluation of machine learning models	Ensemble model (Random Forest) gave the best performance	Feature weighting, Attention mechanism and hybrid loss function gave improved DBN
Darmawan Irfan, Gunawan Ricky Indra, Rahmatulloh Alam [91]	Effectiveness of several machine learning models in performing diabetes detection at an early stage	Compared 14 classification models with consciousness of the time of execution	DBN model gave a better accuracy
Islam MM et al. [69]	Predict the risk of having diabetes mellitus	Proposed best algorithm for the prediction of newly created datasets made for diabetic risk prediction	Improved DBN was built with Generative Adversarial Networks (GANs) which gave a better and more accurate result
Tahsin Mohammad et al. [92]	Predict the likelihood of having diabetes mellitus	Application of stacking to classifier model for better performance	Higher and better model was built with best feature selection, hybrid loss function and appropriate treatment for missing and imbalance dataset
Akter Laboni [93]	Prediction of diabetes at early stage relying on patients' different clinical symptoms with machine learning	Selection of significant features with Extreme Gradient Boosting (XGBoost) with specificity of 94.74%	DBN performed better with recall of 95%
Taser Pelin Yildirim [94]	Diabetes mellitus risk prediction	Utilized bagging and boosting approaches on different Six (6) decision tree models	Better and enhanced DBN model built utilizing an hybrid loss function
Nipa Nurjahan et al. [72]	Early prediction of diabetes mellitus risk using multiple datasets	Usage of shapley additive explanation to show the important of each feature added	Most relevant features were added in the feature selection process with the novel voting ensemble feature selection method used for DBN
Ma Juncheng [95]	Need for early prediction of diabetes mellitus with machine learning methods	Machine learning methods for prediction of DM with NN with 96%	Enhanced DBN built with a better performance with f1-score of 1

alopecia. These were then prepared for the pretraining of our DBN model.

In our experimental setup, the performance of the classical models was measured in three ways: all features in the original dataset, all qualified (including strongly qualified) features, and the strongly qualified features only. Table 5 and Fig. 6 show the results of various experiments with some classical non-deep classification models.

4.2 Performance of DBN model

In this study, we conducted four experiments to progressively improve the performance of a Deep Belief Network (DBN) model for diabetes prediction. The baseline model, referred to as the Classical DBN, served as a foundation for comparison and was implemented in *Experiment 1*. In *Experiment 2*, we introduced a Generative Adversarial Network (GAN) to augment the dataset by generating synthetic samples for the minority class, addressing class imbalance and improving model robustness. Building on this, *Experiment 3* incorporated an attention mechanism into the DBN, allowing the model to dynamically assign higher importance to more relevant features, enhancing its ability to identify critical patterns in the data. Finally, in *Experiment 4*, we applied a hybrid loss function, combining cross-entropy with focal loss, to focus the model on hard-to-classify examples and further mitigate the effects of class imbalance. This systematic approach allowed us to progressively refine the DBN model's performance across a series of experiments. In the tuning of our DBN model,

experimentally selected the values of our parameters and the best for our model and data were identified. Tuning is a crucial stage to avoid fitting problem. For instance, the choice of the number of hidden layers was carefully selected before too small results in underfitting while too large results in overfitting. In this study, we used Rectified Linear Unit (ReLU) as our hidden activation functions, with three hidden layers with 250, 250, 500 as the total number of hidden units in the neural networks, Sigmoid as our input activation function with 20 RBM epochs, 100 batch size and a global learning rate of 0.06.

Table 6 presents the performance metrics of our incremental experiments with Deep Belief Networks (DBN) for diabetes prediction. The experiments are organized in a progressive manner, with each new setup adding an enhancement to the base DBN model. The Classical DBN, which serves as the baseline, achieves a precision of 0.85, recall of 0.82, F1-score of 0.83, and an AUC of 0.87. In the second experiment, by integrating a Generative Adversarial Network (GAN) to address class imbalance, we observe an improvement in all metrics. Precision increases to 0.89, recall to 0.88, F1-score to 0.88, and AUC rises to 0.91. The third experiment incorporates an Attention Mechanism into the model, further enhancing the model's focus on important features. This experiment results in a precision of 0.91, recall of 0.89, F1-score of 0.91, and an AUC of 0.93. Finally, in the fourth experiment, we applied a Hybrid Loss Function combining cross-entropy and focal loss, which yielded the highest performance across all metrics: a precision of 0.98, recall of 0.95, F1-score of 0.97, and an AUC of 1.00. These results demonstrate that each incremental improvement contributed significantly to the overall performance of the DBN model, culminating in near-perfect classification with the hybrid loss function setup. The performance comparison of each increment is also depicted in Fig. 7.

5 Discussion

The results demonstrate that our Deep Belief Network (DBN) model with enhancements such as attention mechanisms, GAN-based data augmentation, and a hybrid loss function significantly outperforms traditional machine learning models, such as Decision Trees, Random Forests, and k-Nearest Neighbors, in the context of early diabetes prediction. One of the major challenges identified in related works is the issue of class imbalance, which can result in poor performance for minority class predictions, such as diabetes-positive cases. While traditional models tend to overlook or misclassify these instances, our model effectively addresses this by generating synthetic samples using Generative Adversarial Networks (GANs) to augment the minority class, thus achieving a better balance in the dataset. As a result, the model demonstrated higher precision, recall, and F1-scores, showing significant improvement in correctly identifying diabetes-positive patients while reducing false negatives.

A key strength of our approach lies in the use of feature prioritization through an attention mechanism, rather than traditional feature selection methods. While feature selection reduces the dimensionality of the dataset by keeping only the most relevant features, it risks discarding potentially valuable information. In contrast, our attention mechanism retains all features but assigns different importance to each one based on its contribution to the classification task. For example, features like polyuria and polydipsia were dynamically weighted higher due to their stronger correlation with diabetes outcomes, while less informative features had a reduced influence. This allows the model to make more informed decisions by considering all features, rather than eliminating any potentially valuable input. The flexibility of feature prioritization enables the model to capture subtle patterns in the data that would otherwise have been missed with strict feature selection, leading to improved predictive accuracy.

Furthermore, feature prioritization enhances the interpretability of the model. By adjusting the importance of features rather than excluding them entirely, the attention mechanism provides insights into the decision-making process, highlighting which features contribute most to the model's predictions. This transparency is particularly crucial in the medical field, where understanding the factors influencing a diagnosis is essential for building trust in machine learning models. While traditional feature selection might obscure valuable information by excluding features, our approach retains all relevant data and clearly indicates which features are driving the classification decisions. This ensures that the model is not only more accurate but also more interpretable and clinically useful, providing healthcare professionals with a clearer understanding of the factors leading to a prediction.

Results also show that that sex, polyuria, polydipsia, sudden weight loss, weakness, polyphagia, muscle stiffness and alopecia are the strongest indicators in the dataset while age, itching and obesity are deemed by the voting ensemble model to have no significant contribution to the diabetes status of the patients. This study finds its significance in the fact that the deep learning model developed in this work can assist medics and patients in creating awareness on the early predictors of diabetes mellitus. One rather shocking discovery in this study is the fact that even though diabetes affects older people the more, our feature rankers disqualified it as a possible threat of diabetes. Early detection is advantageous in the sense that it can help shape lifestyle, dietary and sleeping patterns. Studies have also

shown that early and intensive intervention, not only prevents beta-cell dysfunction but also informs on the potential associated cardiovascular risk factors before reaching the blood glucose thresholds currently set for diagnosing Type II diabetes. It has also been established in literature that early treatment combined with metformin-vildagliptin provides relevant improvements in long-term glycaemic control and can positively affect the disease's progression. Hence, the importance of this study [86–88].

Table 7 provides a comparative analysis of recent studies on diabetes prediction using machine learning, highlighting the distinct methodologies and outcomes achieved in prior research. While studies such as Palabaş Tugba et al. [89] and Taminul Islam et al. [70] employed ensemble techniques to enhance model accuracy, our proposed model goes a step further by integrating ensemble feature selection, a hybrid loss function, and attention mechanisms to improve classification on imbalanced datasets. Compared to Yadu Shweta et al. [90] and Darmawan Irfan et al. [91], who utilized Random Forest and multiple model comparisons, our DBN leverages synthetic data generated through Generative Adversarial Networks (GANs) to address class imbalance, resulting in better performance. Furthermore, studies like Tahsin Mohammad et al. and Taser Pelin Yildirim [94] explored stacking and boosting approaches, which our DBN outperformed by using advanced feature selection and hybrid loss functions. With a recall of 95% and F1-score of 1, our model demonstrates superior accuracy and robustness compared to models such as those by Akter Laboni [93] and Ma Juncheng [95], who focused on feature importance and neural network-based predictions. This justifies the effectiveness and novelty of our approach, establishing our model as a valuable tool for early diabetes risk prediction.

6 Conclusion and future work

In this paper, we presented a novel approach to early diabetes prediction using a Deep Belief Network (DBN) enhanced with attention mechanisms, GAN-based data augmentation, and a hybrid loss function. Our model effectively addresses the challenges of class imbalance, feature relevance, and classification accuracy, which are critical in medical diagnosis tasks. The use of GANs for synthetic data generation helped balance the dataset, improving the model's ability to correctly classify minority cases. The attention mechanism dynamically prioritized the most relevant features, allowing the model to focus on key indicators like polyuria and polydipsia, while still considering other features in the dataset. Additionally, the hybrid loss function, combining cross-entropy and focal loss, ensured that the model maintained high performance, especially in handling harder-to-classify instances.

Future research on diabetes prediction models could benefit from integrating a broader range of data sources, such as genetic profiles, lifestyle factors, and environmental influences, to increase model robustness and predictive accuracy. Enhancing model interpretability through techniques like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) would also make the models more transparent and actionable in clinical settings, providing healthcare professionals with insights into the key factors driving predictions. Additionally, testing the model across diverse demographic groups and larger datasets would help establish generalizability and ensure its applicability across various populations. Developing mobile applications that incorporate these predictive models could enable real-time, personalized health insights, supporting early intervention efforts and empowering individuals to manage their risk more proactively.

Author contribution All authors contributed equally to this work.

Funding No funding was received for this study.

Data availability Data and code will be made available upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent Not applicable.

Research involving human and/or animals Not applicable.

References

1. Preston EV et al (2020) Climate factors and gestational diabetes mellitus risk–a systematic review. Environ Health 19(1):1–19

2. Wang P et al (2020) Seasonality of gestational diabetes mellitus and maternal blood glucose levels: evidence from Taiwan. Medicine 99(41):e22684
3. Boiko M, Ovchinnikova R, Shabrina A (201) The role of hormones in the human body. in Человек. Общество. Культура. Социализация
4. Hauge-Evans AC (2021) Sugar, dogs, cows, and insulin—the story of how diabetes stopped being deadly. Front Young Minds. https://doi.org/10.3389/frym.2021.585489
5. Padhi S, Nayak AK, Behera A (2020) Type II diabetes mellitus: a review on recent drug based therapeutics. Biomed Pharmacother 131:110708
6. Sarwar A et al (2020) Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. Int J Inf Technol 12:419–428
7. Eizirik DL, Pasquali L, Cnop M (2020) Pancreatic β-cells in type 1 and type 2 diabetes mellitus: different pathways to failure. Nat Rev Endocrinol 16(7):349–362
8. Padhi S, Dash M, Behera A (2021) Nanophytochemicals for the treatment of type II diabetes mellitus: a review. Environ Chem Lett 19(6):4349–4373
9. Zohair M et al (2024) A model fusion approach for severity prediction of diabetes with respect to binary and multiclass classification. Int J Inf Technol 16(3):1955–1965
10. Lee KW et al (2020) Neonatal outcomes and its association among gestational diabetes mellitus with and without depression, anxiety and stress symptoms in Malaysia: A cross-sectional study. Midwifery 81:102586
11. Yang Q-Q et al (2021) The association between diabetes complications, diabetes distress, and depressive symptoms in patients with type 2 diabetes mellitus. Clin Nurs Res 30(3):293–301
12. kumari, M. and P. Ahlawat, (2021) DCPM: an effective and robust approach for diabetes classification and prediction. Int J Inform Technol 13:1079–1088
13. Kopitar L et al (2020) Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Sci Rep 10(1):1–12
14. Gadekallu TR et al (2020) Early detection of diabetic retinopathy using PCA-firefly based deep learning model. Electronics 9(2):274
15. Yang H et al (2020) New perspective in diabetic neuropathy: from the periphery to the brain, a call for early detection, and precision medicine. Front Endocrinol 10:929
16. Sungheetha A, Sharma R (2021) Design an early detection and classification for diabetic retinopathy by deep feature extraction based convolution neural network. J Trends Comput Sci Smart Technol (TCSST) 3(02):81–94
17. Tofte N et al (2020) Early detection of diabetic kidney disease by urinary proteomics and subsequent intervention with spironolactone to delay progression (PRIORITY): a prospective observational study and embedded randomised placebo-controlled trial. Lancet Diabetes Endocrinol 8(4):301–312
18. Hasan DA, et al. (2021) Machine learning-based diabetic retinopathy early detection and classification systems-a survey. In: 2021 1st Babylon International Conference on Information Technology and Science (BICITS). 2021. IEEE
19. Ben-Israel D et al (2020) The impact of machine learning on patient care: a systematic review. Artif Intell Med 103:101785
20. Peiffer-Smadja N et al (2020) Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 26(5):584–595
21. Bernabe-Ortiz A et al (2018) Diagnostic accuracy of the Finnish Diabetes Risk Score (FINDRISC) for undiagnosed T2DM in Peruvian population. Prim Care Diabetes 12(6):517–525
22. Boulton AJ et al (2008) Comprehensive foot examination and risk assessment: a report of the task force of the foot care interest group of the American Diabetes Association, with endorsement by the American Association of Clinical Endocrinologists. Diabetes Care 31(8):1679–1685
23. Gray L et al (2012) Implementation of the automated Leicester Practice Risk Score in two diabetes prevention trials provides a high yield of people with abnormal glucose tolerance. Diabetologia 55(12):3238–3244
24. Coetzee A et al (2019) The prevalence and risk factors for diabetes mellitus in healthcare workers at Tygerberg hospital, Cape Town, South Africa: a retrospective study. J Endocrinolo, Metabol Diabetes South Afr 24(3):77–82
25. El Jerjawi NS, Abu-Naser SS (2018) Diabetes prediction using artificial neural network. Inte J Adv Sci Technol 121:55–64
26. NirmalaDevi M, Alias Balamurugan SA, Swathi U (2013) An amalgam KNN to predict diabetes mellitus. In: 2013 IEEE international conference on emerging trends in computing, communication and nanotechnology (ICECCN).. IEEE
27. Alehegn M, Joshi RR, Mulay P (2019) Diabetes analysis and prediction using random forest, KNN, Naïve Bayes And J48: an ensemble approach. Int J Sci Technol Res 8(9):1346–1354
28. Brown GC et al (2000) Quality of life associated with diabetes mellitus in an adult population. J Diabetes Complic 14(1):18–24
29. Tabaei BP, Herman WH (2002) A multivariate logistic regression equation to screen for diabetes: development and validation. Diabetes Care 25(11):1999–2003
30. Parthiban G, Rajesh A, Srivatsa S (2011) Diagnosis of heart disease for diabetic patients using naive bayes method. Int J Comput Appl 24(3):7–11
31. Xu W et al (2017) Risk prediction of type II diabetes based on random forest model. in 2017 third international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB). IEEE
32. Al Jarullah AA (2011) Decision tree discovery for the diagnosis of type II diabetes. In; 2011 International conference on innovations in information technology. IEEE
33. Kumari VA, Chitra R (2013) Classification of diabetes disease using support vector machine. Int J Eng Res Appl 3(2):1797–1801
34. Annamalai R, Nedunchelian R (2023) Design of optimal bidirectional long short term memory based predictive analysis and severity estimation model for diabetes mellitus. Int J Inf Technol 15(1):447–455
35. Miotto R et al (2016) Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 6(1):1–10
36. Pham T et al (2017) Predicting healthcare trajectories from medical records: a deep learning approach. J Biomed Inform 69:218–229
37. Tripathi G, Kumar R (2020) Early prediction of diabetes mellitus using machine learning. in 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE
38. Parte R et al (2019) Non-invasive method for diabetes detection using CNN and SVM classifier. Int J Res Eng, Sci Manage 2:659–661
39. Swapna G, Soman K, Vinayakumar R (2020) Diabetes detection using ecg signals: An overview. Deep Learning Techniques for Biomedical and Health Informatics, pp. 299–327
40. Hu J et al (2022) Raman spectrum classification based on transfer learning by a convolutional neural network: Application to pesticide detection. Spectrochim Acta Part A Mol Biomol Spectrosc 265:120366
41. Al-Smadi M et al (2021) A transfer learning with deep neural network approach for diabetic retinopathy classification. Int J Electr Comput Eng 11(4):3492

42. Spänig S et al (2019) The virtual doctor: an interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. Artif Intell Med 100:101706
43. Nguyen BP et al (2019) Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. Comput Methods Programs Biomed 182:105055
44. Ryu KS et al (2020) A deep learning model for estimation of patients with undiagnosed diabetes. Appl Sci 10(1):421
45. Prabhu P, Selvabharathi S (2019) Deep belief neural network model for prediction of diabetes mellitus. In: 2019 3rd international conference on imaging, signal processing and communication (ICISPC). IEEE
46. Wanda P, Hiswati ME (2024) Belief-DDoS: stepping up DDoS attack detection model using DBN algorithm. Int J Inf Technol 16(1):271–278
47. Jiang L et al (2023) Diabetes risk prediction model based on community follow-up data using machine learning. Prev Med Rep 35:102358
48. Islam MR et al (2023) A comparative approach to alleviating the prevalence of diabetes mellitus using machine learning. Comput Methods Progr Biomed Update 4:100113
49. Lugner M et al (2024) Identifying top ten predictors of type 2 diabetes through machine learning analysis of UK Biobank data. Sci Rep 14(1):2102
50. Li L et al (2023) Machine learning for predicting diabetes risk in western China adults. Diabetol Metab Syndr 15(1):165
51. Sisodia D, Sisodia DS (2018) Prediction of diabetes using classification algorithms. Proced Comput Sci 132:1578–1585
52. Suryadevara CK (2023) Diabetes risk assessment using machine learning: a comparative study of classification algorithms. IEJRD-Int Multidiscip J 8(4):10
53. Lu H et al (2022) A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. Appl Intell 52(3):2411–2422
54. Kopitar L et al (2020) Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Sci Rep 10(1):11981
55. Birjais R et al (2019) Prediction and diagnosis of future diabetes risk: a machine learning approach. SN Appl Sci 1:1–8
56. Bhat SS et al (2022) Prevalence and early prediction of diabetes using machine learning in North Kashmir: a case study of district bandipora. Comput Intell Neurosci 2022(1):2789760
57. Lama L et al (2021) Machine learning for prediction of diabetes risk in middle-aged Swedish people. Heliyon 7(7):e07419
58. Bhat SS et al (2023) A risk assessment and prediction framework for diabetes mellitus using machine learning algorithms. Healthc Anal 4:100273
59. Bhat SS, et al. (2022) Analysis of diabetes mellitus using machine learning techniques. In 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT). IEEE
60. Swain A, Mohanty SN, Das AC (2016) Comparative risk analysis on prediction of diabetes mellitus using machine learning approach. in 2016 international conference on electrical, electronics, and optimization Techniques (ICEEOT). IEEE
61. Prasetyo SY, ZN Izdihar, Nabiilah GZ (2024) Analyzing machine learning approaches for diabetes risk prediction: comparative performance Assessment using BRFSS Data. in 2024 7th International Conference on Informatics and Computational Sciences (ICICoS). IEEE
62. Lai H et al (2019) Predictive models for diabetes mellitus using machine learning techniques. BMC Endocr Disord 19:1–9
63. Samet S, Laouar MR, Bendib I (2021) Diabetes mellitus early stage risk prediction using machine learning algorithms. In: 2021 International Conference on Networking and Advanced Systems (ICNAS). IEEE
64. Zarar M, Wang Y (2023) Early stage diabetes prediction by approach using machine learning techniques
65. Khanam JJ, Foo SY (2021) A comparison of machine learning algorithms for diabetes prediction. Ict Express 7(4):432–439
66. Kaluarachchi K, Premachandra K, Dissanayake R (2023) Comparison of machine learning techniques used in type II diabetes risk prediction
67. Akther, M., et al. A comparative study on different machine learning techniques in diabetes risk assessment. In: 2023 3rd International Conference on Intelligent Technologies (CONIT). IEEE
68. Uddin MJ et al (2023) A comparison of machine learning techniques for the detection of type-2 diabetes mellitus: experiences from bangladesh. Information 14(7):376
69. Islam M et al (2020) Likelihood prediction of diabetes at early stage using data mining techniques. Computer vision and machine intelligence in medical image analysis. Springer, pp 113–125
70. Islam T, et al. (2023) Early-stage diabetes risk prediction using supervised machine learning algorithms. In: 2023 2nd International Conference on Futuristic Technologies (INCOFT). IEEE
71. Xu Y, Nie Y (2024) Diabetes prediction based on support vector machine model. Highlights Sci, Eng Technol 102:311
72. Nipa N et al (2024) Clinically adaptable machine learning model to identify early appreciable features of diabetes. Intell Med 4(01):22–32
73. Dong Q, Gong S, Zhu X (2018) Imbalanced deep learning by minority class incremental rectification. IEEE Trans Pattern Anal Mach Intell 41(6):1367–1381
74. Johnson JM, Khoshgoftaar TM (2019) Survey on deep learning with class imbalance. Journal of big data 6(1):1–54
75. Thabtah F et al (2020) Data imbalance in classification: experimental evaluation. Inf Sci 513:429–441
76 Drummond C, Holte RC (2005) Severe class imbalance Why better algorithms aren't the answer. European Conference on Machine Learning. Springer, Berlin
77. Shao S, Wang P, Yan R (2019) Generative adversarial networks for data augmentation in machine fault diagnosis. Comput Ind 106:85–93
78. Biswas A et al (2023) Generative adversarial networks for data augmentation. Data Driven Approaches on Medical Imaging. Springer, pp 159–177
79. Gui N, Ge D, Hu Z (2019) AFS: An attention-based mechanism for supervised feature selection. In: Proceedings of the AAAI conference on artificial intelligence
80. Bosman AS, Engelbrecht A, Helbig M (2020) Visualising basins of attraction for the cross-entropy and the squared error neural network loss functions. Neurocomputing 400:113–136
81. Ma Y et al (2020) Low-dose CT image denoising using a generative adversarial network with a hybrid loss function for noise learning. IEEE Access 8:67519–67529
82. Goceri E (2024) GAN based augmentation using a hybrid loss function for dermoscopy images. Artif Intell Rev 57(9):1–19
83 de Rosa GH et al (2021) Enhancing anomaly detection through restricted Boltzmann machine features projection. Int J Inform Technol 13:49–57
84. Diqi M, Hiswati ME, Nur AS (2022) StockGAN: robust stock price prediction using GAN algorithm. Int J Inf Technol 14(5):2309–2315
85. Bonett DG, Wright TA (2000) Sample size requirements for estimating Pearson Kendall and Spearman correlations. Psychometrika 65(1):23–28
86. Gómez-Peralta F et al (2020) When does diabetes start? Early detection and intervention in type 2 diabetes mellitus. Rev Clín Española (English Edition) 220(5):305–314

87. Gilmer TP, O'Connor PJ (2010) The growing importance of diabetes screening., Am Diabetes Assoc. 1695–1697
88. Sabariah MMK, Hanifa SA, Sa'adah MS (2014) Early detection of type II Diabetes Mellitus with random forest and classification and regression tree (CART). In: 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA). IEEE
89. Palabaş T (2024) EARLY-STAGE DIABETES RISK PREDICTION USING MACHINE LEARNING TECHNIQUES BASED ON ENSEMBLE APPROACH. Eskişeh Tek Üniv Bilim ve Teknol Derg-C Yaşam Bilimleri Ve Biyoteknol 13(2):74–85
90. Yadu S, Chandra R, Sinha VK (2024) Comparing different machine learning techniques in predicting diabetes on early stage. Eng Proceed 62(1):20
91. Darmawan I, Gunawan RI, Rahmatulloh A (2017) Model accuracy test for early stage of diabetes risk prediction with data science approach. In: International Conference on Engineering, Applied Sciences and System Modeling. Springer
92. Tahsin MS, et al. (2022) Predictive analysis & brief study of early-stage diabetes using multiple classifier models. In: 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE
93. Akter L (2022) Diabetes mellitus prediction and feature importance score finding using extreme gradient boosting. Proceedings of International Conference on Emerging Technologies and Intelligent Systems: ICETIS 2021. Springer, Cham, pp 643–654
94. Taser PY (2021) Application of bagging and boosting approaches using decision tree-based algorithms in diabetes risk prediction. in Proceedings.. MDPI
95. Stone M et al (2020) Artificial intelligence (AI) in strategic marketing decision-making: a research agenda. Bottom Line 33(2):183–200