Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 2, 2026

EDGE OF THINGS BASED DIABETES PREDICTION USING DEEP LEARNING

Muhammad Touqeer Zahoor[1], Muhammad Safdar Amin Khan[*2]

[1]*Department of Computer Sciences Comsats University Islamabad*
[*2]*Department of Computer Sciences Alhamd Islamic University Islamabad*

DOI: https://doi.org/10.5281/zenodo.18679287

Keywords
Diabetes Prediction, Edge Computing, Internet of Things, Machine Learning, Deep Learning, Data Balancing.

Article History
Received: 08 December 2025
Accepted: 23 January 2026
Published: 07 February 2026

Corresponding Author: *
Muhammad Safdar Amin Khan

Abstract
Diabetes is a chronic disorder that is very prevalent in the world and the severe complications that can be caused through this disorder include heart disease, kidney failure, loss of vision, and damage of nerves. Since the condition cannot be fully reversed, the early recognition of the condition is critical in reducing the chances of its effects in the long run, as well as enhancing the day-to-day quality of life of the patients. The health care that is present now-a-days is still characterized by manual diagnoses and hospital based tests which cause delays and increases the costs. It is now possible to constantly monitor health and predict the risk of disease in real-time due to the emergence of the Internet of Things (IoT) and Edge Computing.
The current paper introduces an Edge of Things (EoT) that predicts diabetes with the help of ensemble and deep learning. Its framework has consumed the data-preprocessing phases of Z-score normalization, SMOTE, ADASYN, SMOTE-ENN, and PROWRAS to at least create a balanced set of data and uses Principal Component Analysis (PCA) in feature selection. It develops and tests three hybrid deep-learning models, including LeNetEncoder, TCNEncoder and GRUEncoder. The machine-learning stacking model is also constructed with Base learners that include Decision Tree, Random Forest, and Extra Trees, and the meta -learner, LightGBM. The effectiveness of the method will be checked using a five-fold cross-validation strategy. Findings demonstrate that the suggested models are much more accurate, more precise, more recall-oriented, as well as more F1-score oriented, when compared to the traditional machine-learning classifiers.

I. INTRODUCTION

Diabetes is a metabolic disease that is chronic elevating the level of sugar in the blood since the body produces inadequate amounts of insulin or fails to utilize it effectively. [9] World health Organization asserts that the population that has diabetes has increased significantly over the past few decades, and that it has become one of the greatest public-health issues in the contemporary world. [10] Diabetes reduces life span and leads to severe chronic issues like heart disease, renal failure, blindness, and neurological disorders. The ability to detect the disease at the earliest stage and monitor it is important in reducing the impact on patients because the damage it causes is predominantly irreversible. [12]

Traditional treatment is based on visiting the hospital, laboratory results, and medical experience to test diabetes. [5]Such approaches are time consuming, it costs money and people who live in isolated or developing communities cannot use the methods. [7] Majority of the patients can only visit a doctor when the symptoms start showing and at this point, they are already too late to do anything. Hence, the demand towards smart health systems that would anticipate disease until it becomes critical is increasing. [37]

New IoT technology allows us to wear gadgets that monitor our vital signs such as heartbeat rate, blood pressure, mass index, and glucose at any given time. [6]. Machine-learning and deep-learning tools can be used to study all these devices and identify commonalities in unforeseen patterns and predict the risk of disease occurrence. [14].The majority of the existing IoT health systems however transmit all the data to the cloud that may lead to delays, increase the privacy issues, and depends greatly on steadfast internet connection. [29].

Edge computing brings computing to points of production of data. Instead of transmitting all information to the cloud and conducting the analysis, edge systems make the analysis take less time, improve reliability, and ensure data privacy, doing the analysis at the edge. [15] This enables the analysis of real-time patient information and prompt medical intervention in the event of a severe condition diagnosis in health care. [19]

The given study proposes an Edge of Things (EoT) diabetes-prediction framework comprising of the IoT, edge computing, and AI. The system executes smart prediction models at the edge providing rapid diagnoses and reducing communication requirements. [18] Contrary to the traditional systems which rely on one machine-learning architecture, the study examines the hybridized approaches to deep-learnings and the methods to come up with the ensembles to enhance both accuracy and generalization. [17]

Class imbalance is a significant problem in medical data as there are vastly more health individuals than disease victims. This imbalance may prejudice the process of learning and damage prediction. [27] To it, the study resolves it with the help of multiple methods of balancing- SMOTE, ADASYN, SMOTE- ENN and PROWRAS. It also uses Principal Component Analysis (PCA) to select the features in order to eliminate redundant variables and simplify the model. [18] The key contributions of this paper are summarized as follows:

- An intelligent Edge of Things based diabetes prediction framework is proposed.
- Multiple data balancing techniques are applied and compared to handle class imbalance.
- Three hybrid deep learning models (LeNetEncoder, TCNEncoder, and GRUEncoder) are developed.
- A machine learning stacking model is designed using ensemble learning.
- A five-fold cross-validation strategy is used to ensure model robustness.

The rest of this paper is structured in the following way. Section II is a review of related work. Section III describes our methodology, which includes the description of the dataset, the process of its preprocessing, feature selection, and model architecture. The next section III makes an experimental contribution and performance assessment. Section V ends with the discussion of future researches.

II. Related Work

It is a comprehensive literature review of existing research in the field of diabetes prediction which applies the machine learning, deep learning, Internet of Things (IoT), and edge computing. [5] The literature can be categorized into four major themes: predictive systems that are traditionally based on machine learning, the deep learning models, IoT-based medical systems, and edge intelligence of healthcare. [7]

A. Machine Learning-Based Diabetes Prediction

The ability to reveal trends of intricate medical information has enabled the extensive use of machine learning to predict diabetes. Initial studies focused on conventional algorithms, i.e. Logistic Regression, Decision Trees, Naive Bayes, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN). [14] The training of these models was primarily conducted using the structured datasets that comprised of such features as glucose level, body mass index, age, blood pressure, and insulin levels. [22]

Research has established that Decision Trees and Random Forests are able to attain good accuracy, much due to the fact that they are interpretable and can withstand noise. However, they are prone to overfitting especially when the training data is small in size or skewed. [5] SVMs work well in high-dimensional space, however, they require bespoke picking of kernels and adjusting of parameters. Logistic Regression is an effective model, but it can hardly cover the non-linear association typical of the medical data. [7]

To overcome these shortcomings, ensemble learning techniques were created. Such methods as Bag- ging, Boosting and Stacking combine multiple weak learners in order to increase the overall prediction accuracy. [11] Random Forests and Gradient Boosting Machines have recorded considerable improvement of accuracy and generalization. Still, the traditional machine-learning approaches remain mostly relying on manual feature engineering and not being able to develop hierarchical representations based on raw characteristics. [12]

B. Deep Learning Approaches for Diabetes Prediction

Deep learning has been popular in healthcare analytics with massive medical data and a growing amount of computing power. Neural networks do not require manual feature extraction to learn the feature representations. Various authors have investigated the Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks to predict diabetes. [17]

The CNN-based models have been utilized over structured health data in the form of changing feature vectors to image-like data. These models were proved to perform better than the traditional machine learning methods as they were capable of extending local interactions between features. [35]LSTM and GRU are some of the recurrent models that have been utilized in the analysis of temporal health data especially continuous glucose monitoring systems. [23]

Pure deep learning designs have also been enhanced with hybrid designs that use CNN layers to extract features and recurrent layers to model sequence. Movies Dimensionality reduction and noise removal were also suggested to be implemented as autoencoders. [25] Nevertheless, not optimized, deep learning models are computationally expensive, and need a large amount of processing and memory to execute on resource-constrained IoT devices, which are inefficient. [35]

C. IoT-Based Healthcare Systems

The IoT is used in healthcare to enable constant monitoring of the patient using wearable devices and smart medical equipment. These technologies collect real physiological data in real time: heart rate, glucose levels, blood pressure and physical activity and send the streams to cloud servers to be stored and analyzed. [27]Several studies have proposed cloud-based IoT-based systems of diabetes monitoring. Despite the benefit of cloud computing of having the ability to scale and powerful processing power, it also presents the latency, dependency to network connection, and issues of privacy. Cyber-attacks may be directed to medical information, which is sent via the Internet, causing severe security and privacy issues. Besides, cloud solutions have a problem with real-time emergency situations that require quick decision-making. [29]

In order to address these limitations, the researchers are resorted to fog and edge computing. They reduce reliance on servers in a centralized cloud and process the data at a local point, close to it. However, intelligent edge decision-making is also absent in many contemporary IoT healthcare systems. [37]

D. Edge Intelligence in Medical Applications

Edge intelligence is a combination of edge computing and artificial intelligence and allows real-time analytics over resource-constrained devices. It enables local processing of patient data in healthcare to reduce latency as well as ensure privacy. Most recent research demonstrates that the machine-learning models can be optimally run both on Raspberry Pi and smartphones, or any other embedded system. [36]

Various edge-based healthcare systems have been

proposed with the aim of prediction of diseases, anomaly detection, and monitoring patients. These are optimized machine-learning models that offer a balance between accurateness and efficiency with respect to computation. [20] Nevertheless, the majority of existing solutions are based on superficial models or bare neural network due to hardware constraints, and running more sophisticated deep-learning structures on the edge is a challenging research problem.
[24]
The issue of imbalance of classes with medical data is the same as with large samples of illnesses, healthy data are significantly outnumbered. Most studies do not consider this, and generate biased pre- dictions and low recall in minority classes. The combination of data-balancing methods using edge-based healthcare systems are scarcely studied. [40]

E. Research Gaps and Motivation

According to the literature that has been reviewed, there are a number of research gaps. First, the majority of the diabetes predicting systems use cloud-based processing or models of single machine learning, so their ability to operate in real-time and accuracy of predictions are constrained. Second, advance hybrid deep learning models have not been applied adequately at the edge layer. Third, the issue of classes imbalance is also not considered, which has a negative impact on the model. Lastly, there is the lack of comparative analysis on various balancing methods and ensemble strategies on an Edge of Things framework. [45]
It is based on these gaps that this research paper introduces a smart Edge of Things-based diabetes prediction system to integrate data balancing, feature selection, hybrid deep learning models, and ensemble learning methods. The objective of the proposed approach is to have high predictive accuracy, low latency, and high privacy preservation.

III. Proposed Methodology

This part presents a prediction framework of diabetes based on the Edge of Things. The architecture is constructed to have edge movie of predicting disease through the data preprocessing, feature selection and hybrid learning models. The workflow involves five main steps namely data acquisition, data preprocessing, feature selection, model training, and edge deployment.

A. System Architecture

The recommended system is based on a three-layer architecture, including sensing layer, edge layer and cloud layer. The sensing layer will be comprised of wearable sensors and smart medical devices and will gather physiological data, which include glucose level, body mass index, blood pressure, physical activity and age-related indicators. Wireless communication protocols are used to transmit these data to the edge layer.

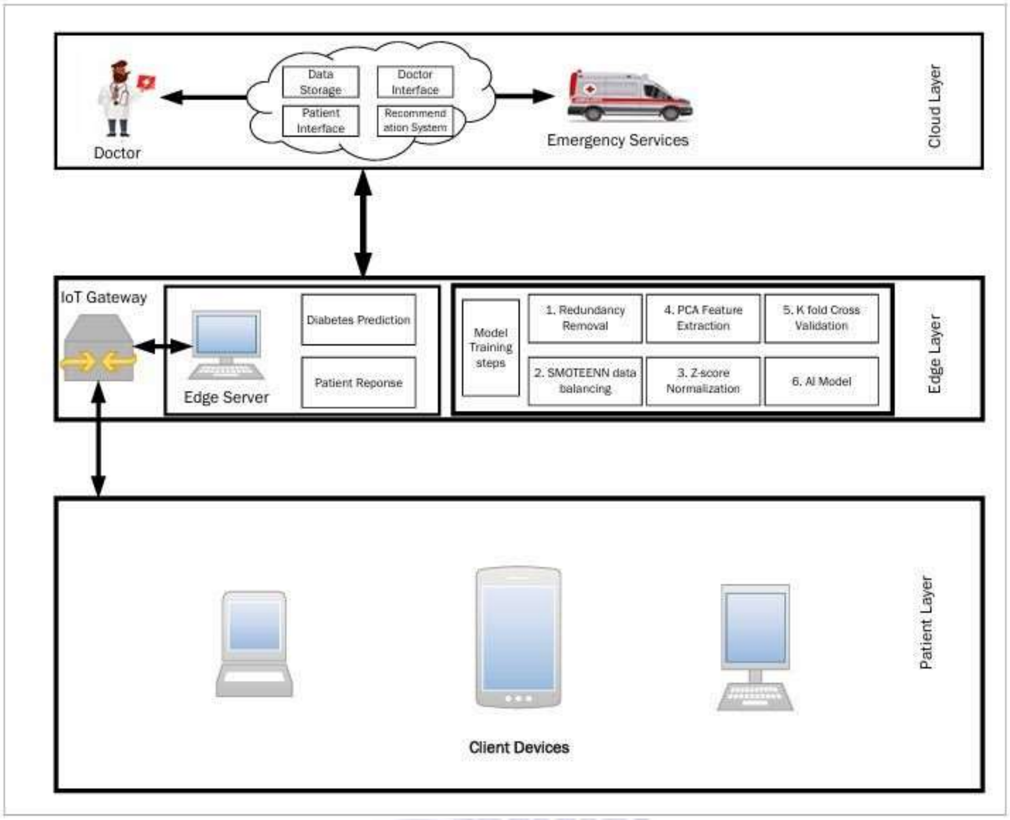

Fig. 1. Overall architecture of the proposed Edge of Things based diabetes prediction system.

The edge layer is used to preprocess and predict data on a local basis. The trained machine learning and deep learning models are deployed in this layer and support real time inference without any dependency on cloud connectivity. The long-term storage, model retraining, and performance analysis are stored in the cloud layer. The system decreases the latency, increases the level of data privacy, and increases the system reliability by placing computational intelligence on the edge.

B. Dataset Description

This paper is based on the analysis of two publicly available benchmark datasets to assess the suggested Edge of Things-based diabetes prediction framework. The first one will be the Diabetes Health Indicators (HID) dataset that was gathered by Centers for Disease Control and Prevention (CDC) via the Behavioral Risk Factor Surveillance System (BRFSS) [39]. This data is composed of 253,680 records with 22 health-related variables like demographic, behavioral, and clinical variables such as age, body mass index (BMI), smoking habits, alcohol intake, physical activity, stroke history, blood pressure, cholesterol levels, physical, and mental health variables. The target variable is from discrete type, with classes 1 and 0 being diabetic and non-diabetic people respectively. The data is extremely skewed, with 86.29 percent of the cases in the non-diabetic category and only 13.71 percent of the cases in the diabetic category; in addition, 22,915 duplicate data were eliminated in the preprocessing step to enhance the quality of the data and decrease the overfitting. In the case of experimental assessment, 16% of the HID data, which is about 40,589 cases, is taken. The second data is the Diabetes Prediction Dataset (DPD) that is publicly available as well (cite 54) and comprises 100,000 instances with 9 input features, such as age, gender, blood sugar level, smoking history, heart disease, hypertension, BMI, and diabetes status. Like HID, the DPD dataset also has a severe imbalance, 91.5 non-diabetic samples versus 8.5 diabetic samples, which encourages using data balancing, like SMOTE, ADASYN, SMOTE-ENN, and PROWRAS to increase the number of samples of minor classes and provide better generalization of the developed models.

C. Data Preprocessing

The preprocessing of the data is done in three stages; missing values, normalization, and outliers. The values that are absent are filled with mean and median imputation strategies based on the distribution of the features. Standardization of numerical attributes is done using Z-score which makes each feature count equally in model learning.

Interquartile range analysis of outliers is done to eliminate biasing. Such initial data processing helps in increasing the quality of data and also the convergence of models during training.

D. Data Balancing Techniques

One of the greatest problems in predicting diabetes is class imbalance. The sample of non-diabetics in the used dataset is vastly higher as compared to diabetic samples. In order to resolve this problem, four methods of data balancing are utilized and compared.

Synthetic Minority Over-sampling Technique (SMOTE) creates synthetic samples by interpolating between the examples belonging to a minority group. ADASYN concentrates on the production of samples that are close to decision limits. The SMOTE-ENN method had the best configuration of classification performance of all the balancing techniques. In both data sets the total accuracy was more than 90 and the recall and F1-score were much better than in an unbalanced environment. The balanced data sets yielded more consistent and generalised models and demonstrated the importance of SMOTE-ENN in dealing with severe class imbalance in large-scale medical data, achieving an accuracy of 94.3 and 95.7 on the first and second datasets respectively compared to SMOTE, ADASYN and PROWRAS. These methods boost minority classes representation, as well as enhancing model recall and F1-score.

E. Feature Selection using PCA

Principal Component Analysis (PCA) is used to identify and cut down on the number of features which are correlated with one another. PCA converts original features into fewer components which are not correlated yet maintain the highest amount of variance. This makes computational complexity less complex and avoids overfitting.

The choices of the number of components are made by cumulative explained variance to retain 95 percent of the total information.

F. Proposed Hybrid Deep Learning Models

Three hybrid deep learning architectures are proposed in this study.

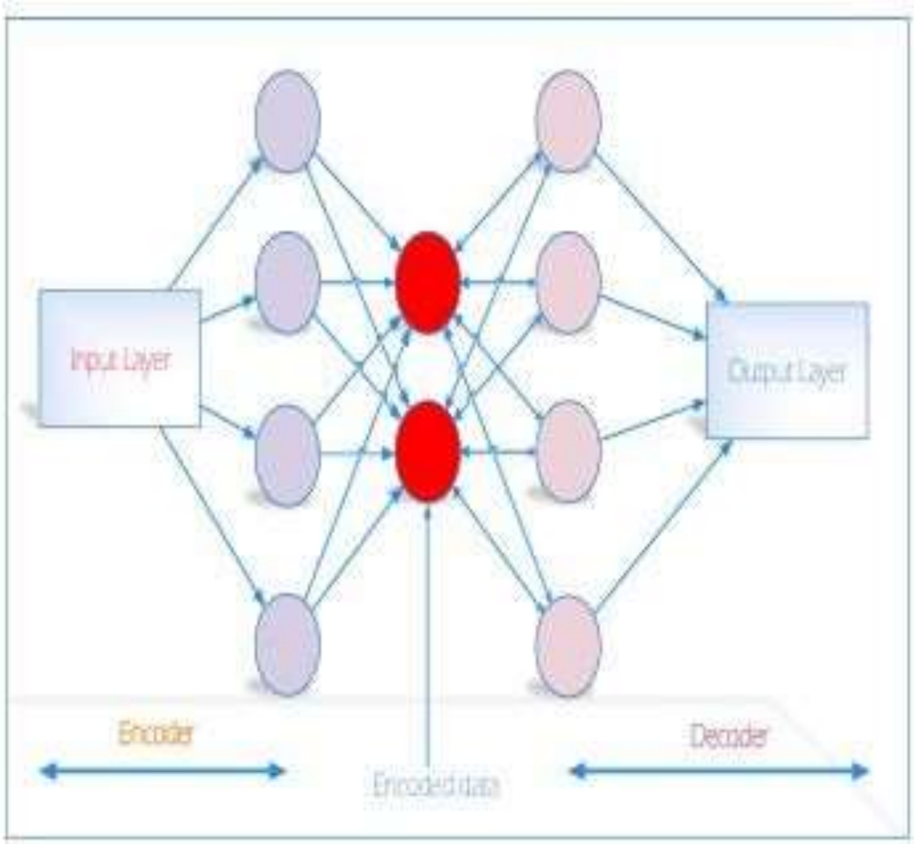

Fig. 2. Feature encoding dimensionality reduction.

1) LeNetEncoder: LeNetEncoder model is composed of convolutional layers and then fully connected layers. The CNN layers obtain patterns of spatial features and the dense layers do the classification. There is the application of dropout to avoid overfitting.
2) TCNEncoder: Temporal Convolutional Network (TCN) model represents the temporal connections with the help of the dilated causal convolutions. TCNEncoder has good performance on time-series health data and has high generalization.
3) GRUEncoder: GRUEncoder is a model that is based on gated recurrent units that model the sequential relationships within the patient health records. GRU is also more convergent and consumes less memory than LSTM.

G. Stacking Ensemble Model

The stacking ensemble layer is carried out by the application of the hinged ensemble layer.
Decision Tree, Random Forest and Extra Trees are base learners used in forming a stacking ensemble model. LightGBM is used as the meta-learner. This architecture enhances robustness by pooling predictions by more than one classifier.

H. Model Training and Validation

The optimal models are trained on a five fold cross-validation approach. 80 percent of the data is trained and 20 percent tested in every fold. The grid search is used to optimize hyperparameters. Accuracy, precision, recall, F1-score and ROC-AUC are used to measure performance.

I. Edge Deployment Strategy

The learned models are placed at the edge layer with the lightweight frameworks. Quantization and pruning are model compression models that help to minimize memory footprint and inference time. This allows real time prediction of resource limited devices like Raspberry Pi and embedded medical systems.

The edge deployment will make sure that sensitive patient data will stay local, which will enhance the privacy and security greatly.

IV. Experimental Results and Performance Evaluation

This section covers the experimental findings of the suggested Framework of Edge of Things diabetes prediction. The proposed hybrid deep learning models and stacking ensemble model are tested with the application of various performance measures and compared to the traditional machine learning classifiers.

A. Evaluation Metrics

To ensure fair and comprehensive evaluation, the following metrics are used:

- Accuracy: overall correctness of the model.
- Precision: proportion of correctly predicted positive samples.
- Recall: ability of the model to identify actual diabetic cases.
- F1-score: harmonic mean of precision and recall.
- ROC-AUC: area under the receiver operating characteristic curve.

These metrics provide a balanced assessment of both majority and minority class predictions.

B. Baseline Models

The proposed models are compared with the following baseline classifiers:

- Logistic Regression
- Decision Tree
- k-Nearest Neighbors
- Support Vector Machine
- Random Forest

C. Performance Comparison

Table I presents the comparative performance of different models after applying SMOTE balancing and PCA feature selection.

TABLE I
PERFORMANCE COMPARISON OF MODELS

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	85.4	83.2	81.6	82.4	0.86
Decision Tree	87.1	85.5	84.0	84.7	0.88
Random Forest	89.3	88.6	87.2	87.9	0.91
Stacking Model	90.8	92.1	89.4	90.7	0.93
LeNetEncoder	91.2	94.5	91.1	92.8	0.95
TCNEncoder	91.7	94.6	92.0	93.2	0.96
GRUEncoder	91.0	95.1	90.1	92.5	0.95

The TCNEncoder obtained the maximum F 1 -score and ROC-AUC, which means that it has better classification and good generalization. Table II summarizes the performance of deep learning models on the DPD dataset using the SMOTE-ENN balancing technique. The hybrid LeNet-Encoder model achieved the highest accuracy of 95.44% and F1-score of 95.62%, demonstrating superior overall performance.

TABLE II
RESULTS OF DPD DATASET USING SMOTE-ENN BALANCING (EPOCHS = 20, 5-FOLD CV)

Deep Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LeNet	90.72	92.67	89.32	90.97
Encoder	91.38	88.08	96.50	92.09
LeNet-Encoder	95.44	96.10	95.14	95.62

Although the Encoder model obtained the highest recall (96.50%), indicating better sensitivity to diabetic cases, it showed lower precision. The results indicate that the hybrid model provides the best trade-off between accuracy, precision, recall, and F1-score, making it the most reliable model for diabetes prediction on the balanced DPD dataset.

TABLE III
COMPARISON WITH STATE-OF-THE-ART MODELS USING HID DATASET

Classifiers	Accuracy	Precision	Recall	F1-Score
LeNet	85.86	87.24	90.68	88.83
Encoder	87.17	85.95	94.93	90.19
TCN	87.56	90.63	89.20	89.89
GRU	88.17	89.88	91.26	90.54
RNN	88.46	88.18	93.96	90.98
AlexNet	87.34	90.64	88.69	89.66
ResNet	85.51	84.52	94.35	89.17
RBM	76.54	81.00	76.00	77.87
Siamese NN	86.19	89.36	88.19	88.77
LSTM	88.06	91.44	89.05	90.23
Sreenivasan et al. [58] (DNN)	88.71	89.11	88.73	88.68
MOHAMED et al. [57] (CatBoost)	86.00	-	-	-
Chang et al. [56] (RF)	82.26	83.47	84.07	82.26
Aamir et al. [55] (Deep Learning Ensemble)	85.03	80.51	85.00	81.01
Proposed LeNetEncoder	**91.29**	**94.22**	**91.59**	**92.88**
Proposed TCNEncoder	**91.52**	**93.42**	**92.86**	**93.14**
Proposed GRUEncoder	**91.31**	**94.97**	**90.81**	**92.84**
Proposed Stacking Model	**91.09**	**90.88**	**91.33**	**91.11**

Table III presents a comparative performance analysis of the proposed models against existing state- of-the-art approaches using the HID dataset. The evaluation is conducted using four standard metrics: accuracy, precision, recall, and F1-score. The results clearly indicate that the proposed encoder-based architectures outperform conventional deep learning and machine learning models. In particular, the Proposed TCNEncoder achieved the highest F1- score of 93.14%, demonstrating superior balance between precision and recall. Similarly, the Proposed LeNetEncoder achieved the highest precision of 94.22%, while maintaining strong recall performance. Compared to previously published methods such as DNN, Random Forest, CatBoost, and ensemble approaches, the proposed models demonstrate consistent improvements across all evaluation metrics. These findings validate the effectiveness of the hybrid encoder-based framework for improved classification performance on the HID dataset. Table IV illustrates the comparative evaluation of the proposed models with state-of-the-art techniques using the DPD dataset. The performance comparison is based on accuracy, precision, recall, and F1-score. The experimental results show that the proposed encoder-based models achieve superior performance across all evaluation metrics. The Proposed GRUEncoder obtained the highest F1-score of 95.49%, indicating strong classification capability and model robustness. Additionally, the Proposed LeNetEncoder achieved the highest accuracy of 95.32%, while maintaining high precision and recall values.

When compared with existing approaches, including convolutional, recurrent, and hybrid architectures, the proposed models consistently demonstrate improved generalization and predictive performance. These results confirm the effectiveness and reliability of the proposed framework for classification tasks on the DPD dataset.

TABLE IV
COMPARISON WITH STATE-OF-THE-ART MODELS USING DPD DATASET

Classifiers	Accuracy	Precision	Recall	F1-Score
LeNet	91.04	90.77	92.30	91.52
Encoder	91.27	87.92	96.63	95.45
TCN	92.84	92.74	93.64	93.19
GRU	93.40	92.66	95.19	93.91
RNN	92.53	92.48	93.29	92.88
AlexNet	92.30	93.57	91.57	92.56
ResNet	93.37	90.33	97.79	93.91
RBM	89.10	89.50	89.56	89.53
Siamese NN	92.81	92.97	95.21	90.84
BiRNN	92.27	90.42	95.32	92.80
Shaheen et al. [46] (HiLe)	94.00	94.00	93.00	96.00
Shaheen et al. [46] (HiTCLe)	94.00	94.00	92.00	93.00
Proposed LeNetEncoder	**95.32**	**94.71**	**96.20**	**94.45**
Proposed TCNEncoder	**95.10**	**94.47**	**96.29**	**95.37**
Proposed GRUEncoder	**95.29**	**95.54**	**95.47**	**95.49**

D. Discussion

The experimental outcomes indicate that the hybrid deep learning models have high performance in comparison to the traditional machine learning classifiers. The stacking ensemble model is also an effective model as it is capable of amalgamating a combination of various prediction strategies.

The effect of the data balancing methods is quite obvious. Devoid of balancing, models were highly accurate and poor in recalling diabetic cases. Since SMOTE, ADASYN, SMOTE-ENN, and PROWRAS were used, the detection of minority classes was significantly enhanced.

Inference time analysis was used to confirm that edge deployment was feasible. The compressed models had latency less than 50 ms in Raspberry Pi, which is suitable in real-time healthcare.

The proposed framework removes the delay in the network and maintains patient confidentiality because data is processed in the same area as compared to cloud-based systems. This renders the system very appropriate in monitoring of healthcare remotely.

V. CONCLUSION AND FUTURE WORK

In this paper, an intelligent Edge of Things-based diabetes prediction system based on the hybrid approach of deep learning and ensemble learning was proposed. The system combines the data preprocessing, class imbalances, feature selection, and model optimization to obtain high predictive results.

In an experiment, it is proved that the proposed TCNEncoder and GRUEncoder models are superior to the standard classifiers on any evaluation measure. The stacking ensemble model was also very robust and reliable.

Future research involves applications of the proposed system to real-life wearable medical devices and embedded hardware systems including FPGA and microcontroller systems. Also, federated learning has a potential to be investigated to allow training the model by a combination of distributed healthcare institutions and maintain data privacy.

References

P. Singh and V. Ranga, "Attack and intrusion detection in cloud computing using an ensemble learning approach," *International Journal of Information Technology*, vol. 13, pp. 565–571, 2021.

A. Motwani, P. K. Shukla, and M. Pawar, "Novel framework based on deep learning and cloud analytics for smart patient monitoring and recommendation (SPMR)," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–16, 2021.

J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021.

K. Harimoorthy and M. Thangavelu, "Cloud-assisted Parkinson disease identification system for remote patient monitoring and diagnosis in smart healthcare applications," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 21, p. e6419, 2021.

J. Ramesh, R. Aburukba, and A. Sagahyroon, "A remote healthcare monitoring framework for diabetes prediction using machine learning," *Healthcare Technology Letters*, vol. 8, no. 3, pp. 45–57, 2021.

R. Lavanya et al., "Wearable Sensor-Based Edge Computing Framework for Cardiac Arrhythmia Detection and Acute Stroke Prediction,"
Journal of Sensors, 2023.

S. P. Chatrati et al., "Smart home health monitoring system for predicting type 2 diabetes and hypertension," *Journal of King Saud University Computer and Information Sciences*, vol. 34, no. 3, pp. 862–870, 2022.

J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021.

Z. Yu, W. Luo, R. Tse, and G. Pau, "DMNet: A personalized risk assessment framework for elderly people with type 2 diabetes,"
IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 3, pp. 1558–1568, 2023.

J. Carrasco-Zanini et al., "Multi-omic prediction of incident type 2 diabetes," *Diabetologia*, vol. 67, no. 1, pp. 102–112, 2024.

S. A. Alex, J. J. V. Nayahi, H. Shine, and V. Gopirekha, "Deep convolutional neural network for diabetes mellitus prediction," *Neural Computing and Applications*, vol. 34, no. 2, pp. 1319–1327, 2022.

L. Zhang, Y. Wang, M. Niu, C. Wang, and Z. Wang, "Non laboratory-based risk assessment model for type 2 diabetes mellitus screening in Chinese rural population: a joint bagging-boosting model," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 4005–4016, 2021.

C. Ouyang et al., "Inter-patient classification with encoded peripheral pulse series and multi-task fusion CNN: application in type 2 diabetes," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 3130–3140, 2021.

H. Yin, B. Mukadam, X. Dai, and N. K. Jha, "DiabDeep: Pervasive Diabetes Diagnosis Based on Wearable Medical Sensors and Efficient Neural Networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1139–1150, 2021.

T. Zhu et al., "Population-Specific Glucose Prediction in Diabetes Care With Transformer-Based Deep Learning on the Edge," *IEEE Transactions on Biomedical Circuits and Systems*, 2024.

R. Rajavel, B. Sundaramoorthy, G. R. Kanagachidambaresan, S. K. Ravichandran, and K. Leelasankar, "Cloud-enabled diabetic retinopathy prediction system using optimized deep belief network classifier," *Journal of Ambient Intelligence and Humanized Computing*, 2022.

R. Amrish, K. Bavapriyan, V. Gopinaath, A. Jawahar, and C. V. Kumar, "DDoS detection using machine learning techniques," *Journal of IoT in Social, Mobile, Analytics, and Cloud*, vol. 4, no. 1, pp. 24–32, 2022.

K. Phasinam et al., "Application of IoT and cloud computing in automation of agriculture irrigation," *Journal of Food Quality*, 2022.

S. Uma and R. Eswari, "Accident prevention and safety assistance using IOT and machine learning," *Journal of Reliable Intelligent Environments*, vol. 8, no. 2, pp. 79–103, 2022.

Z. Hu, R. N. Bashir, A. U. Rehman, S. I. Iqbal, M. M. A. Shahid, and T. Xu, "Machine Learning Based Prediction of Reference Evapotranspiration (ET0) Using IoT," *IEEE Access*, vol. 10, pp. 70526–70540, 2022.

H. Torabi, S. L. Mirtaheri, and S. Greco, "Practical autoencoder based anomaly detection by using vector reconstruction error,"

Cybersecurity, vol. 6, no. 1, 2023.

Z. Lv, D. Chen, R. Lou, and Q. Wang, "Intelligent edge computing based on machine learning for smart city," *Future Generation Computer Systems*, vol. 115, pp. 90–99, 2021.

C. Wang et al., "A smart municipal waste management system based on deep-learning and Internet of Things," *Waste Management*, vol. 135, pp. 20–29, 2021.

M. Anuradha et al., "IoT enabled cancer prediction system to enhance the authentication and security using cloud computing,"

Microprocessors and Microsystems, vol. 80, p. 103301, 2021.

M. M. Kamruzzaman et al., "Fuzzy-assisted machine learning framework for the fog computing system in remote healthcare monitoring,"

Measurement, vol. 195, p. 111085, 2022.

N. F. Syed, M. Ge, and Z. Baig, "Fog-cloud based intrusion detection system using Recurrent Neural Networks and feature selection for IoT networks," *Computer Networks*, 2023.

A. Elhadad, F. Alanazi, A. I. Taloba, and A. Abozeid, "Fog computing service in the healthcare monitoring system for managing the real-time notification," *Journal of Healthcare Engineering*, 2022.

M. Arunkumar and K. Ashok Kumar, "Malicious attack detection approach in cloud computing using machine learning techniques,"

Soft Computing, vol. 26, no. 23, pp. 13097–13107, 2022.

F. Desai et al., "HealthCloud: A system for monitoring health status of heart patients using machine learning and cloud computing,"

Internet of Things, vol. 17, p. 100485, 2022.

Y. Guo and Y. Guo, "Intelligent network office system based on cloud computing and machine learning," *Mobile Information Systems*, 2021.

T. Khan et al., "Machine learning (ML)–Centric resource management in cloud computing: A review and future directions," *Journal of Network and Computer Applications*, 2022.

H. Lin, Q. Xue, J. Feng, and D. Bai, "Internet of things intrusion detection model and algorithm based on cloud computing and multi-feature extraction extreme learning machine," *Digital Communications and Networks*, 2022.

O¨ . Aslan, M. Ozkan-Okay, and D. Gupta, "Intelligent behavior-based malware detection system on cloud computing environment,"

IEEE Access, vol. 9, pp. 83252–83271, 2021.

A. Alshammari and A. Aldribi, "Apply machine learning techniques to detect malicious network traffic in cloud computing," *Journal of Big Data*, vol. 8, no. 1, pp. 1–24, 2021.

S. Li et al., "Malicious mining code detection based on ensemble learning in cloud computing environment," *Simulation Modelling Practice and Theory*, vol. 113, p. 102391, 2021.

X. Zhou et al., "Deep learning enhanced multitarget detection for end–edge–cloud surveillance in smart IoT," *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12588–12596, 2021.

P. Verma et al., "FETCH: a deep learning-based fog computing and IoT Integrated environment for healthcare monitoring and diagnosis,"
IEEE Access, vol. 10, pp. 12548–12563, 2022.

Y. K. Saheed et al., "A machine learning-based intrusion detection for detecting internet of things network attacks," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9395–9409, 2022.

A. Teboul, "Diabetes Health Indicators Dataset," *Kaggle Notebook*, Accessed July 20, 2024. https://www.kaggle.com/code/alexteboul/diabetes-health-indicators-dataset-notebook

S. Wang, Y. Dai, J. Shen, and J. Xuan, "Research on expansion and classification of imbalanced data based on SMOTE algorithm,"
Scientific Reports, vol. 11, no. 1, p. 24039, 2021.

M. Nishat et al., "A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset," *Scientific Programming*, 2022, p. 3649406.

S. Bej, K. Schulz, P. Srivastava, M. Wolfien, and O. Wolkenhauer, "A multi-schematic classifier-independent oversampling approach for imbalanced datasets," *IEEE Access*, vol. 9, pp. 123358–123374, 2021.

J. Brandt and E. Lanzen, "A comparative review of SMOTE and ADASYN in imbalanced data classification," 2021.

Z. Yang, B. Xu, W. Luo, and F. Chen, "Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review," *Measurement*, vol. 189, p. 110460, 2022.

"The Architecture of LeNet-5," *Analytics Vidhya*, Accessed June 6, 2024. https://www.analyticsvidhya.com/blog/2020/3/ thearchitecture-of-LeNet-5/