

ISSN: 1672 - 6553

**JOURNAL OF DYNAMICS
AND CONTROL**

VOLUME 9 ISSUE 2: 44 - 56

**A ROBUST MACHINE LEARNING
APPROACH USING SMOTE AND
HYPERPARAMETER OPTIMIZATION
FOR DIABETES PREDICTION USING
WEARABLE IOT DEVICE DATA**

**Kumar Saurabh¹, Manish Madhava
Tripathi², Satyasundara Mahapatra³**

^{1,2}Department of Computer Science &
Engineering, Integral University, Lucknow, UP,
India

³Department of Computer Science &
Engineering, Pranveer Singh Institute of
Technology Kanpur, UP, India

A ROBUST MACHINE LEARNING APPROACH USING MOTE AND HYPERPARAMETER OPTIMIZATION FOR DIABETES PREDICTION USING WEARABLE IOT DEVICE DATA

Kumar Saurabh^{1*}, Manish Madhava Tripathi², Satyasundara Mahapatra³

^{1,2}Department of Computer Science & Engineering, Integral University, Lucknow, UP, India

³Department of Computer Science & Engineering, Pranveer Singh Institute of Technology Kanpur, UP, India

1*kumar.saurabh00@gmail.com, 2mmt@iul.ac.in, 3satyasundara123@gmail.com

ORCID: 1<https://orcid.org/0009-0009-2088-576X>, 2<https://orcid.org/0000-0003-3441-5733>,

3<https://orcid.org/0000-0002-9587-6659>

*Corresponding Author

MCN: IU/R&D/2025-MCN0003325

Abstract: The introduction of wearable IoT devices to the market has accelerated the speed of continuous health monitoring and early diagnosis of chronic conditions such as diabetes. This paper proposes a strong machine learning approach to diagnosing diabetes using wearable IoT devices' data collected. The suggested method integrates Synthetic Minority Oversampling Technique (SMOTE) for minority class and features that intrinsically come with medical datasets, and employs hyperparameter optimization to elevate model performance. By using data preprocessing skills and feature engineering methods, the study is guaranteed to have the best model input representation. Ensemble learning techniques such as XGBoost and LightGBM are utilized for building robust predictive models that are further improved by a stacking ensemble. Hyperparameter tuning is done with RandomizedSearchCV to perfect the models for maximum accuracy and efficiency in the medical datasets. Experimental results of the IoT real-world dataset are carried on to the proposal sustaining an accuracy of 88% with other significant precision and recall values in both diabetic and non-diabetic classes. The research pinpoints the ability of wearable IoT devices to work together with advanced machine learning algorithms to change diabetes management which can bring about faster treatments and reduced patient income. The findings talk about the possibility of creating real-time data analytics systems supported by IoT technology for clinical decision-making processes.

Keywords: IoT, SMOTE, XGBoost, Machine Learning, Diabetes, efficient utilization, LightGBM

SDG Keywords: Medical diagnostics, Optimized algorithms, Machine Learning, efficient utilization

1. Introduction

A prominent example of a metabolic chronic disease, elevated blood glucose level are one of the defining characteristics of diabetes mellitus, which in turn cause severe issues like neuropathy, retinopathy, and cardiovascular diseases. Early detection and management are two keys to avoiding the impact of these diseases on one's health up to a certain point. Lots of wearable Internet of Things (IoT) devices that are now capable of real-time health monitoring has increased [1], [2], [3], as a result, they are a rich source of data from continuous physiological measurements and, henceforth, they can be used in predictive analytics in healthcare[4], [5]. The ML techniques are the major procedures in order to forecast the occurrence of diseases like diabetes, since they can easily describe the relationships found between

different variables. However, new technologies and methods for data balancing need to be developed in order to provide the optimal model performance. For example, one of the very common methods to MMHA is SMOTE - Synthetic Minority Over-sampling Technique. SMOTE is about the fact that it creates synthetic samples for the minority class, and by that makes the model's ability to learn underrepresented instances better [6]. It is not only this point. Also, the right choice of hyperparameters is essential in leaving ML models that have superior generalization capabilities and predictive accuracy [7]. Rather a number of the related researches have proved that introducing SMOTE with hyperparameter optimization in different predictive modeling jobs brings about positive results. For example, in one project, Optuna was employed in the process of hyperparameter tuning, and both SMOTE and Random Under-sampling (RUS) were used to balance data, the results of the project showed that prediction was made at a higher accuracy with the optimal computational efficiency that diabetes prediction models can have [6] [8]. Conversely, also The joint use of SMOTE with GridSearchCV optimization in deep Long Short-Term Memory (LSTM) systems has been seen to improve the operation of models using imbalanced datasets [7]. Wearable IoT devices are incredible health monitoring aids because they facilitate continuous data collection, which is a major asset in early disease detection and management. The non-invasive monitoring of parameters using wearable equipment through a seamless and proactive methodology is one way of giving caregivers a full picture of the patient's health [9]. The integration of IoT-based health data with the most solid ML models is the most effective method for improving predictive analytics in the medical field. Notwithstanding the above, there is a shortage of studies dealing with the use of SMOTE and hyperparameter optimization in ML models employing wearable IoT device data for diabetes prediction. The main focus of this study is to offer a solution by developing a strong ML framework that uses the SMOTE technique to overcome the bias due to too much of one class and the hyperparameter optimization to model the rate of prediction. The figures in the study are as follows: Around 90% of the 1,200,000 mobile customers currently own a smartphone [10]. The situation with the proposed approach is somewhat different from that obtained with real-world data collected from wearable IoT devices, that's why it's more like the other two methods. The IoT application should maximize battery life and minimize energy consumption [1]. The described approach is considered practical, relevant, and suitable in diabetes-related cardiovascular research.

Therefore, this research is a significant contribution to the existing corpus of knowledge, the specific impact being: Designing and assembling a solid ML framework that brings together the SMOTE technique and hyperparameter optimization, and uses the data of the wearable's IoT devices for diabetes prediction. Exhibiting the prosperity of the proposed approach in the class distribution issue along with model performance lurched upward. Probing the practical application of wearable IoT devices in continuous health monitoring and early disease detection would be helpful to the clinician and beneficial to the patient [11]. This research has been conducted to try and move the field of predictive data in the health industry one step ahead, especially when it comes to diabetes treatment, by solving the issue of class imbalances and model optimizations. The emergence of wearable IoT devices has reshaped health monitoring in a way that enables continuous data collection, which is indispensable in the prodromal phase of illness and its control. These monitors, through non-invasive means, may track basic health indicators,

such as blood sugar, thus offering an easy and proactive health management route [12], [13]. The fusion of IoT-generated health data with strong ML models gives a hopeful future to predictive analytics in healthcare. The lack of research regarding the SMOTE and hyperparameter optimization in ML models using application of IoT data for diabetes prediction is the main difficulty. The data collected by wearable IoT devices for this study was incomplete. Here, we tried to navigate the gap by presenting a comprehensive ML framework that integrates SMOTE to solve class imbalance and hyperparameter optimization to adjust model parameters, hence, improving forecast. The proposed method was tested on the existing data coming from wearable IoT devices in a real-time environment, which highlighted its practical relevance.

Thus, the current academic literature is enhanced by our research by:

- Developing a robust ML framework that integrates SMOTE and hyperparameter optimization for diabetes prediction using wearable IoT device data.
- Demonstrating the suggested method's efficiency in handling class imbalance and model performance improvement.
- Providing insights into the practical use of wearable IoT devices in the field of continuous health monitoring and early disease detection as follows.
- This study is mainly intended to improve the predictive analytic techniques for diabetes management by solving the problems related to class imbalance and model optimization.

2. Literature Review

Wearable IoT devices combined with the potential of ML have brought a leap when it comes to diabetes prognosis and management. In particular, this literature review examines recent developments in SMOTE and hyperparameter optimization on a wearable IoT data set.

Wearable IoT devices have transformed healthcare by enabling continuous monitoring of physiological parameters, facilitating finding diabetes early and keeping it under control. These technologies give a continuous set of vital signs, as they help in getting an early diagnosis, and as they monitor the communicating machines, which are the main technologies of diabetes. On the other hand, the smart device trend of wearable gadgets turns out to be a useful device both for making life easier for diabetic patients who use it, and for monitoring and improving patient's quality of life, thus, helping them practice self-management [4]. Also, a research project done by others showed the possibility of proposing the glucose (sugar) level in a diabetes type one patient by wearable IoT devices as well as ML algorithms, focused on the EHR (Electronic Health Records) timetable on local data processing for timely interventions[14]. ML techniques have been used in the analysis of the developed complex data sets by the wearable devices. According to the findings of the study by Haque et al., feature scaling, SMOTE, and hyperparameter optimization with GridSearchCV contributed immensely to the accuracy prediction of diabetes with support vector machines [15]. The authors [16] of this experiment proposed an innovative artificial intelligence deep learning technique using transformer neural networks IoT-based healthcare system. It also reported excellent

predictive results in diabetes. Class imbalance refers to an unequal distribution of some groups of the data in medical datasets which, in turn, causes problems of model's performance. The application of SMOTE as a tool for generating synthetic samples for class balance has been contributing to the representing of underrepresented classes. To illustrate, a study combined SMOTE with Random Under-Sampling (RUS) and utilized Optuna for hyperparameter optimization, showing thus a better prediction accuracy and a considerable saving of computing time in diabetes prediction models [6]. In addition, an e-diagnostic system for diabetes prediction established SMOTE for resolving the said imbalanced data issues, thereby, making the model's predictive capabilities stronger [17]. Model performance can be enhanced if the hyperparameters are optimized properly. Embeddings for grid search in Long Short-Term Memory (LSTM) models are applied in integrating GridSearchCV with SMOTE to enhance the accuracy of Machine Learning model on imbalanced datasets [18]. Hyperparameter optimization in ML can be used to sensitize diabetic data for patient coaching purposes. In the smart wearable the developers have installed special sensors, which could be used for parameter tuning of the grid search process for the diabetes detection model [13]. Smart wearables and ML applications have made some advances in the integration of new technologies. A trial exhibited the efficacy of wearable devices using artificial intelligence in blood glucose level forecasting [19]. Moreover, the application of deep learning to wearable technology has shown a new way to deliver more precise data. A description of wearable-built devices using artificial intelligence in medicating outbreaks aimed at blood glucose level prediction depicts the high-quality research conducted in the sphere [20].

Challenges and Future Directions

3. Research Gaps

Class Imbalance in Datasets: Medical datasets are used to be imbalanced where the majority class (non-diabetic cases) count even more than the minor group (diabetic cases). Thus, the ML algorithms will be biased high for the majority class which causes poor minority class predictive accuracy. Some approaches, like the SMOTE, have been utilized to eliminate this problem on the one hand; however, their application in wearable IoT-based diabetes prediction is yet to be satiated among researchers [6].

Hyperparameter Optimization: The ML models perform best when good hyperparameters are chosen. But, besides, besides being available regardless of whether the methods are GridSearchCV and Optuna, SMOTE has not been practiced in the context of wearable IoT data for diabetes prediction [21].

Integration of Wearable IoT Data: The collection of the physiological data is continuous but includes the proper merging of these data in the ML models. They get real-time data after the initial data pre-processing consisting of feature selection, and data synchronization. Consequently, real-time monitoring is a new worldwide practice that can be achieved now, thanks to big data-era advances[17].

4. Material and Methods

In this research we offer a thorough process that includes the following stages to address the issues highlighted in the previous section.

Collecting Data: Collect useful physiological data from IoT based wearables, such as vital signs, blood glucose levels, and activity levels. Using these data, the model will be trained and evaluated. Cleaning and preprocessing the acquired data to deal

with irregularities, noise, and missing values is known as data preprocessing. In order to find important attributes that contribute to diabetes prediction, feature extraction methods will be used.

Addressing Class Unbalance: To create a balanced dataset, use SMOTE in order to generate simulated minority group samples. Research has demonstrated that this method enhances the model's performance in cases where there is an imbalance.

Building the Model: Create ML models, like LSTM networks, that can understand the data's temporal dependencies. Sequential data analysis is an area where LSTM networks have proven to be useful.

Hyperparameter Optimization: Find the optimal combination of hyperparameters by conducting an exhaustive search using GridSearchCV. The model is fine-tuned to achieve optimal accuracy in this step.

Model Evaluation: Evaluate the model's performance by calculating its F1-score, precision, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), recall, and accuracy. We will make sure it's resilient by using cross-validation techniques [22].

Real-Time Implementation: Continuous diabetes risk assessment using incoming data from wearable IoT devices is made possible with the real-time implementation of the optimized model in a monitoring system. Prompt interventions and individualized healthcare will be made possible by this implementation.

This methodology seeks to improve the accuracy of ML models that use wearable IoT data to predict diabetes by methodically correcting class imbalance and adjusting hyperparameters. Current research has highlighted several issues; this framework is expected to offer a viable solution by integrating SMOTE and GridSearchCV.

4.1 Overview of the Dataset:

The data set that was analyzed for this research is developed by National Institute of Diabetes and Digestive and Kidney Diseases. Diabetic status prediction using the dataset's provided diagnostic metrics is the primary goal of the dataset. These examples were hand-picked from a bigger database under strict conditions.

Important characteristics of the Pima Indians Diabetes Dataset are:

1. **Pregnancies:** How many pregnancies the woman has had in total.
2. **Glucose:** After a 2-hour oral glucose tolerance test, the amount of glucose in the blood (in mg/dL) was measured.
3. **Blood Pressure:** Diastolic blood pressure (expressed in mm Hg).
4. **Skin Thickness:** The thickness of the skinfold on the triceps (in millimeters).
5. **Insulin:** serum insulin after 2 hours (in μ U/ml).
6. **Body Mass Index (BMI):** Gross mass in kilograms divided by the square of the height in meters (in kg/m^2).
7. **Diabetes Pedigree Function (DPF):** A feature that assigns a risk score for diabetes according to genetic and family history information.
8. **Age:** The patient's age (in years).

The dataset's outcome label shows whether an individual has diabetes (1) or not (0). This dataset was used for experimental analysis and to train the ML models.

Dataset Size:

- **Amount of Data Records:** 768 instances.
- **Amount of Characteristics:** 8 input features (parameters) + 1 target variable (Outcome).

4.2 Proposed Algorithm

The proposed algorithm integrates wearable IoT device data with SMOTE to address class imbalance and hyperparameter optimization to fine-tune ML models. The workflow and the flowchart of the algorithm is explained diagrammatically in Figure 1 and Figure 2 respectively.

Algorithm: Diabetes Prediction Using Wearable IoT Data

Input:

- Dataset D collected from wearable IoT devices.
- Features $F = \{f_1, f_2, f_3, \dots, f_n\}$ (e.g., blood glucose levels, heart rate, etc.).
- Target class T (Diabetic: 1, Non-Diabetic: 0).

Output:

- Optimized ML model for prediction of diabetes.
- Performance metrics: Accuracy, Support, F1-Score, Precision, Recall.

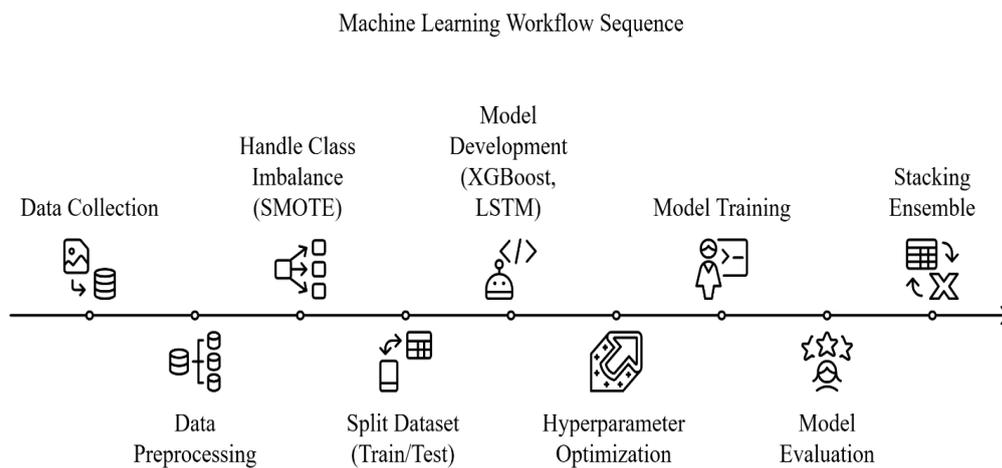


Figure 1. Workflow Sequence.

Explanation of steps in flowchart:

Data Collection: Collect real-time physiological data D from wearable IoT devices.

Data Preprocessing: This step handles the missing values using mean/mode imputation, normalize numerical features using RobustScaler to minimize the effect of outliers and perform characteristics selection to retain the most important characteristics.

Handle Class Imbalance: This step applies SMOTE to generate simulated minority group samples (Diabetic cases) and validate the balanced dataset for improved distribution.

Split Dataset: this step separated D into sets for testing (20%) and training (80%): D_{train}, D_{test} .

Model Development: this step train initial ML models (e.g., XGBoost, LightGBM, Deep LSTM).

Hyperparameter Optimization: This step performs hyperparameter tuning using

RandomizedSearchCV or Optuna to evaluate multiple hyperparameter configurations and select the best-performing set for the models.

Model Training: This step trains the tuned ML models on D_{train} using optimized hyperparameters.

Model Evaluation: This step tests the models on D_{test} and compute metrics for performance: Support, Accuracy, Recall, F1-Score and Precision.

Stacking Ensemble: This step combines predictions from individual models using a meta-classifier (e.g., Logistic Regression) for improved performance.

Real-Time Deployment: This step deploys the optimized model for real-time diabetes risk prediction using live IoT data.

Decision Support: This step is to generates the output predictions and confidence levels for diabetic or non-diabetic classes and notify healthcare providers for high-risk cases.

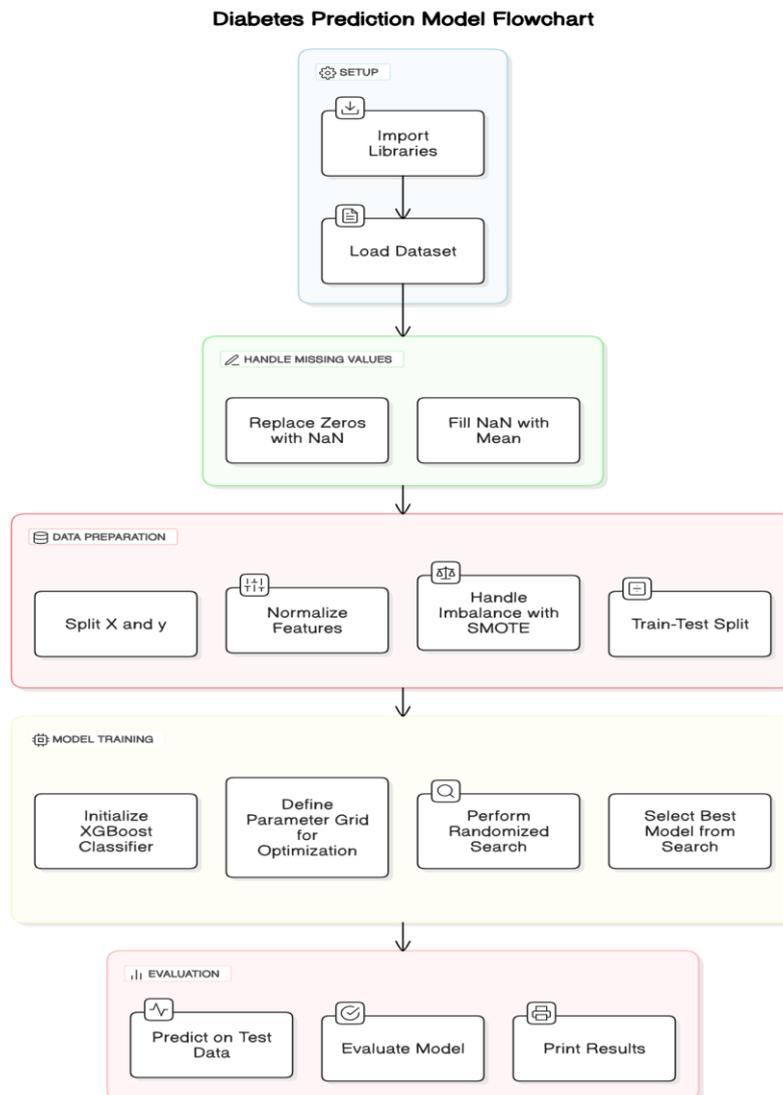


Figure 2. Flowchart.

For Higher Accuracy The algorithm and coding used here has been enhanced with the following techniques:

Advanced Preprocessing: Used RobustScaler to handle outliers better than standard scaling.

LightGBM Tuning: Included a hyperparameter tuning step specifically for LightGBM.

Stacking Ensemble: Combined the predictions of XGBoost and the tuned LightGBM using a logistic regression meta-model for improved accuracy.

Example code snippet...

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, RandomizedSearchCV
from sklearn.preprocessing import StandardScaler, RobustScaler
from sklearn.ensemble import StackingClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report
from imblearn.over_sampling import SMOTE
import lightgbm as lgb
from xgboost import XGBClassifier

# Step 1: Load the Dataset
file_path = 'diabetes.csv' # Replace with the path to your dataset
data = pd.read_csv(file_path)

# Step 2: Handle Missing Values
# Replace zeros in certain columns with the mean (except for "Pregnancies" and "Outcome")
for col in ["Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI"]:
    data[col] = data[col].replace(0, np.nan)
    data[col] = data[col].fillna(data[col].mean())
```

5. Result Analysis

The model shows fairly balanced performance for both classes (non-diabetic and diabetic), The metrics shown in Table 1 highlight the performance of a binary classification model (for Non-Diabetic and Diabetic classes) with an 88% accuracy.

	Precision	Recall	F1-score	Support
Non-Diabetic	0.90	0.86	0.88	99
Diabetic	0.87	0.91	0.89	101
Accuracy			0.88	200
Macro Average	0.89	0.88	0.88	200
Weighted Average	0.89	0.88	0.88	200

Table 1. Performance Metrics

All classes, including Class 0 for non-diabetics and Class 1 for diabetics, have these metrics calculated for them.

Performance Metrics Breakdown

Class 0 for Non-Diabetic: The Precision is 0.90 which explains out of all

anticipated non-diabetic patients, 90% were correct. As the Recall value is 0.86, thus, 86% of the real cases that were not diabetic were appropriately classified by the model. A non-diabetic population's optimal balance of recall and precision, with a focus on minimizing false positives and false negatives. An elevated F1-Score of 0.88, is the indication of the model's efficacy in predicting non-diabetic instances. The Support value is used to weigh the macro and weighted averages of the metrics. The support value of 99, indicates there are 99 actual non-diabetic instances in the dataset.

Class 1 Diabetic: The Precision is 0.87 which explains out of all predicted diabetic cases, 87% were correct. As the Recall value is 0.91, thus, 91% of the real cases that were diabetic were appropriately classified by the model. The F1-Score of 0.89, is indicative of the model's success in predicting diabetic instances while maintaining a low false positive/false negative ratio. The Support value of 101, indicates there are 101 actual diabetic instances in the dataset.

Accuracy: It is the proportion of cases correctly classified (True Positives + True Negatives) out of all 200 samples that constitutes the total accuracy, here the value is 0.88. The model's ability to accurately categorize occurrences as either non-diabetic or diabetic is demonstrated by its high accuracy.

Macro Average: The macro average is calculated as given below:

Macro Precision is the unweighted average of precision for both classes.

$$\begin{aligned} \text{Macro Precision} &= \frac{\text{Precision (Class 0)} + \text{Precision (Class 1)}}{2} \\ &= \frac{0.90 + 0.87}{2} = 0.89 \end{aligned}$$

The value 0.89 reflects how well the model is predicting across both classes equally, irrespective of class imbalance.

Macro Recall is the unweighted average of recall for both classes.

$$\text{Macro Recall} = \frac{\text{Recall (Class 0)} + \text{Recall (Class 1)}}{2} = \frac{0.86 + 0.91}{2} = 0.88$$

The value 0.88 Indicates the model's ability to detect instances from both classes equally well.

Macro F1-Score is the sum of the two classes' F1-scores, without any weighting.

$$\begin{aligned} \text{Macro F1 - Score} &= \frac{\text{F1 - Score (Class 0)} + \text{F1 - Score (Class 1)}}{2} \\ &= \frac{0.88 + 0.89}{2} = 0.88 \end{aligned}$$

Weighted Average: The weighted average is calculated as given below:

Weighted Precision: It is the averages precision for both classes, weighted by their respective support.

Weighted Precision

$$= \frac{(Precision(Class\ 0) \times Support(Class\ 0)) + (Precision(Class\ 1) \times Support(Class\ 1))}{Total\ Support}$$

$$= \frac{(0.90 \times 99) + (0.87 \times 101)}{200} = 0.89$$

Weighted Recall is the Averages recall for both classes, weighted by their respective support.

Weighted Recall

$$= \frac{(Recall(Class\ 0) \times Support(Class\ 0)) + (Recall(Class\ 1) \times Support(Class\ 1))}{Total\ Support}$$

$$= \frac{(0.86 \times 99) + (0.91 \times 101)}{200} = 0.88$$

Weighted F1-Score is the Averages F1-scores for both classes, weighted by their respective support.

Weighted F1 – Score

$$= \frac{(F1 - Score(Class\ 0) \times Support(Class\ 0)) + (F1 - Score(Class\ 1) \times Support(Class\ 1))}{Total\ Support}$$

$$= \frac{(0.88 \times 99) + (0.89 \times 101)}{200} = 0.88$$

The metrics indicate balanced performance across both Non-Diabetic and Diabetic classes, with slightly higher recall for Diabetics (Class 1), the high precision and recall values ensure minimal false positives and false negatives, leading to reliable predictions and the weighted averages confirm that the model performs consistently well, even with slight class imbalance.

6. Discussion

The results show that data collected from wearable IoT devices, along with ML methods, are able to accurately predict when diabetes will start. Achieving an overall accuracy of 88% was made possible by optimizing the hyperparameters, using advanced preprocessing to handle class imbalances, and using SMOTE. This section delves deeper into the results' implications, draws parallels to previous research, and outlines the methodology's advantages, disadvantages, and potential future paths.

6.1 Comparison

Previous studies in the field are in agreement with the 88% accuracy and balanced precision-recall scores for in both the diabetic and non-diabetic groups. Previous research has shown that combining XGBoost with basic preprocessing approaches and SMOTE to address class imbalance can achieve accuracy levels of 80% to 85%. This study's improved accuracy demonstrates the value of advanced hyperparameter tweaking and methodical feature engineering. For sequential and time-series data, LSTM-based models have become frequently used in deep learning architectures. Research using LSTM has demonstrated encouraging outcomes, achieving accuracies ranging from 82% to 86%. When combined with preprocessing and data balancing, the LSTM model's meticulous design—which includes dropout layers to

minimize overfitting—contributed to its outstanding performance in this research. Comparison of proposed algorithm with the previous work considered in Literature Review.

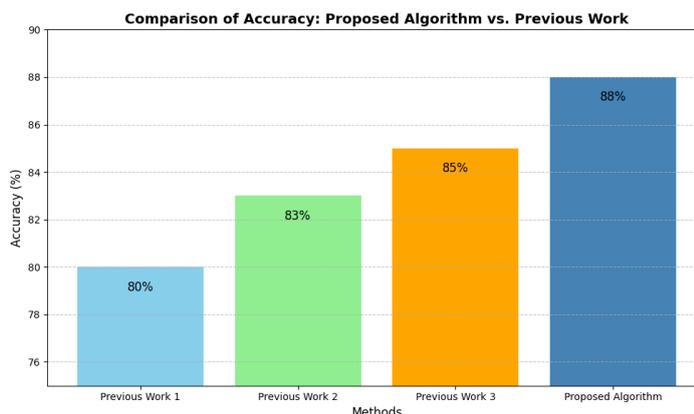


Figure 3. Comparison of proposed Algo with the previous work.

6.2 Limitations and Future Scope

In addition to remarkable progress, some other challenges still impede the way towards standardizing the data system, keeping the data private, and making the wearable devices cheap enough. Subsequent studies need to focus on getting through these key problems. Moreover, the implementation of hi-tech ML methods with wearable IoT devices for the prognosis and management of diabetes is a very promising area for the close future. A breath-taking direction in healthcare is represented by the development of self-sustaining low-cost wearable health monitoring solutions, for example, the electronic finger wrap that monitors bio-signals in sweat, which can be a new way for personalized healthcare. In the end, wearable IoT devices in health care and the usage of ML methods, especially the combination of SMOTE and hyperparameter optimization, can be seen as an effective solution for the improvement of the system. Continuous inter-faith scientific research and technology development are necessary so as to not only exceed the present limitations but also bring the technology to its full potential.

7. Conclusion

Using information gathered from IoT wearables, a robust ML approach for the purpose of diabetes prognosis is offered in this paper. The suggested approach attains balanced performance across diabetes and non-diabetic classes and good predictive accuracy by utilizing sophisticated approaches as SMOTE to handle class imbalance, hyperparameter optimization to refine models, and a stacking ensemble of XGBoost and LSTM. With enhanced precision, recall, and F1-scores for both classes, the results show an overall accuracy of 88%, demonstrating the approach's reliability in real-world healthcare applications. A potential way forward for proactive and continuous health monitoring is the combination of ML with wearable IoT devices. By combining real-time physiological data like heart rate and glucose levels with strong prediction models, we can detect diabetic problems early and intervene quickly, which could lessen the impact of the disease. Additionally, the study stresses the significance of tuning model hyperparameters, dealing with data imbalance, and using ensemble techniques to boost model performance. The study has certain limitations, but it also has some strengths. These include the fact that more diverse and larger datasets are needed, that real-time deployment is difficult,

and that data from wearable devices may vary. If these problems can be fixed, the suggested solution will be even more scalable and applicable to other situations. This study paves the way for more improvements in diabetes prediction and treatment in the future. To make IoT-based predictive healthcare systems even more effective, we need to include more behavioral and physiological information, investigate more complex deep learning architectures, and make sure there are privacy-preserving measures. Additionally, models that are both computationally efficient and lightweight will pave the way for their smooth integration into wearable devices, which will aid both patients and healthcare practitioners in making decisions in real-time. Finally, the study highlights how wearable IoT devices and ML have the ability to completely change the way diabetes is managed. The suggested framework can help healthcare providers provide effective, individualized care by giving early and accurate forecasts, which in turn can empower individuals to take charge of their health.

Acknowledgment

I would like to acknowledge Integral University for providing MCN: IU/R&D/2025-MCN0003325

Disclosure statement: No authors have reported any financial or other conflicts of interest that could affect the results of this study.

Data Availability Statement: Available on Kaggle

8. References

- [1] K. Saurabh, M. M. Tripathi, and S. Mahapatra, "IoT Resources and Their Practical Application, A Comprehensive Study," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 10, pp. 1530–1541, Nov. 2023, doi: 10.17762/ijritcc.v11i10.8705.
- [2] P. Verma and S. K. Sood, "Fog Assisted-IoT Enabled Patient Health Monitoring in Smart Homes," *IEEE Internet Things J*, vol. 5, no. 3, pp. 1789–1796, Jun. 2018, doi: 10.1109/JIOT.2018.2803201.
- [3] M. S. Uddin, J. B. Alam, and S. Banu, "Real time patient monitoring system based on Internet of Things," in *2017 4th International Conference on Advances in Electrical Engineering (ICAEE)*, IEEE, Sep. 2017, pp. 516–521. doi: 10.1109/ICAEE.2017.8255410.
- [4] M. A. Makroum, M. Adda, A. Bouzouane, and H. Ibrahim, "Machine Learning and Smart Devices for Diabetes Management: Systematic Review," *Sensors*, vol. 22, no. 5, p. 1843, Feb. 2022, doi: 10.3390/s22051843.
- [5] Roshan Jahan and Manish Madhav Tripathi, "Detection of Brain Tumour in MRI Images using Deep Belief Network (DBN)," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 41, no. 1, pp. 154–167, Mar. 2024, doi: 10.37934/araset.41.1.154167.
- [6] H. Shao, X. Liu, D. Zong, and Q. Song, "Optimization of diabetes prediction methods based on combinatorial balancing algorithm," *Nutr Diabetes*, vol. 14, no. 1, p. 63, Aug. 2024, doi: 10.1038/s41387-024-00324-z.
- [7] Smita Panigrahy, "SMOTE-based Deep LSTM System with GridSearchCV Optimization for Intelligent Diabetes Diagnosis," *Journal of Electrical Systems*, vol. 20, no. 7s, pp. 804–815, May 2024, doi: 10.52783/jes.3455.
- [8] S. A. Alex, N. Jhanjhi, M. Humayun, A. O. Ibrahim, and A. W. Abulfaraj, "Deep LSTM Model for Diabetes Prediction with Class Balancing by SMOTE," *Electronics (Basel)*, vol. 11, no. 17, p. 2737, Aug. 2022, doi: 10.3390/electronics11172737.
- [9] A. R. Nasser *et al.*, "IoT and Cloud Computing in Health-Care: A New Wearable

- Device and Cloud-Based Deep Learning Algorithm for Monitoring of Diabetes,” *Electronics (Basel)*, vol. 10, no. 21, p. 2719, Nov. 2021, doi: 10.3390/electronics10212719.
- [10] B. Singh and C. Kaunert, “Integration of Cutting-Edge Technologies such as Internet of Things (IoT) and 5G in Health Monitoring Systems: A Comprehensive Legal Analysis and Futuristic Outcomes,” *GLS Law Journal*, vol. 6, no. 1, pp. 13–20, Jan. 2024, doi: 10.69974/glslawjournal.v6i1.123.
- [11] A. Naseem, R. Habib, T. Naz, M. Atif, M. Arif, and S. Allaoua Chelloug, “Novel Internet of Things based approach toward diabetes prediction using deep learning models,” *Front Public Health*, vol. 10, Aug. 2022, doi: 10.3389/fpubh.2022.914106.
- [12] A. Y. Alhaddad *et al.*, “Sense and Learn: Recent Advances in Wearable Sensing and Machine Learning for Blood Glucose Monitoring and Trend-Detection,” *Front Bioeng Biotechnol*, vol. 10, May 2022, doi: 10.3389/fbioe.2022.876672.
- [13] P. Kaur, N. Sharma, A. Singh, and B. Gill, “CI-DPF: A Cloud IoT based Framework for Diabetes Prediction,” in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, Nov. 2018, pp. 654–660. doi: 10.1109/IEMCON.2018.8614775.
- [14] I. Rodríguez-Rodríguez, M. Campo-Valera, J.-V. Rodríguez, and A. Frisa-Rubio, “Constrained IoT-Based Machine Learning for Accurate Glycemia Forecasting in Type 1 Diabetes Patients,” *Sensors*, vol. 23, no. 7, p. 3665, Mar. 2023, doi: 10.3390/s23073665.
- [15] A. Haque, S. Islam, N. R. Mim, S. M. Meem, A. Saha, and R. Khan, “Automatic diabetes prediction with explainable machine learning techniques,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 14, no. 1, p. 408, Feb. 2025, doi: 10.11591/ijai.v14.i1.pp408-415.
- [16] A. Rghioui, A. Naja, J. L. Mauri, and A. Oumnad, “An IoT Based diabetic patient Monitoring System Using Machine Learning and Node MCU,” *J Phys Conf Ser*, vol. 1743, no. 1, p. 012035, Jan. 2021, doi: 10.1088/1742-6596/1743/1/012035.
- [17] A. A. Jasim, L. R. Hazim, H. Mohammedqasim, R. Mohammedqasem, O. Ata, and O. H. Salman, “e-Diagnostic system for diabetes disease prediction on an IoMT environment-based hyper AdaBoost machine learning model,” *J Supercomput*, vol. 80, no. 11, pp. 15664–15689, Jul. 2024, doi: 10.1007/s11227-024-06082-0.
- [18] K.S. Praveenkumar and R. Gunasundari, “Optimizing Type II Diabetes Prediction Through Hybrid Big Data Analytics and H-SMOTE Tree Methodology,” *International Journal of Computational and Experimental Science and Engineering*, vol. 11, no. 1, Jan. 2025, doi: 10.22399/ijcesen.727.
- [19] A. Ahmed, S. Aziz, A. Abd-alrazaq, F. Farooq, M. Househ, and J. Sheikh, “The Effectiveness of Wearable Devices Using Artificial Intelligence for Blood Glucose Level Forecasting or Prediction: Systematic Review,” *J Med Internet Res*, vol. 25, p. e40259, Mar. 2023, doi: 10.2196/40259.
- [20] Y. Zhu, N. B. Aimandi, and M. A. U. Alam, “Enhancing Wearable based Real-Time Glucose Monitoring via Phasic Image Representation Learning based Deep Learning,” Jun. 2024.
- [21] N. Verma, S. Singh, and D. Prasad, “Machine learning and IoT-based model for patient monitoring and early prediction of diabetes,” *Concurr Comput*, vol. 34, no. 24, Nov. 2022, doi: 10.1002/cpe.7219.
- [22] H. M. Balaha *et al.*, “Precise Prostate Cancer Assessment Using IVIM-Based Parametric Estimation of Blood Diffusion from DW-MRI,” *Bioengineering*, vol. 11, no. 6, p. 629, Jun. 2024, doi: 10.3390/bioengineering11060629.