ORIGINAL ARTICLE

Smart Prediction: Class Centric Focal XG- Boost for Accurate Diabetes Forecasting

Vandana C. Bavkar * , **Arundhati A. Shinde**

Bhivarabai Sawant College of Engineering & Research, Narhe, Pune, India

*Corresponding Author: Vandana C. Bavkar
Email: vcbavkar_entc@tssm.edu.in

Received: 05 January 2024 / Accepted: 26 July 2024

Abstract

Purpose: Diabetes, resulting from insufficient insulin production or utilization, causes extensive harm to the body. The conventional diagnostic methods are often invasive. The classification of diabetes is essential for effective management. The progression in research and technology has led to additional classification approaches. Machine Learning (ML) algorithms have been deployed for analyzing the huge dataset and classifying diabetes.

Materials and Methods: The classification and the regression of diabetic and non-diabetic are performed using the XGBoost mechanism. On the other hand, the proposed class-centric Focal XG-Boost is applied to elevate the model performance by measuring the similarity among the features. The prediction of the model is based on the classification and regression rates of diabetic and non-diabetic individuals, which are anticipated using applicable and effectual metrics to estimate their working performance.

The dataset used in the Class-Centric Focal XG Boost model is attained using the Arduino Uno Kit. The data collection is done under a sampling rate of 100 Hz. The data are gathered from Bharati Hospital Pathology Laboratories, located in Pune.

Results: The inclusive outcomes of the proposed model with their appropriate Exploratory Data Analysis (EDA) among classification and regression, with the suitable dataset used in the study are exemplified.

Conclusion: The proposed Class-Centric Focal XG Boost model has numerous advantages and is less delicate to the hyperparameters than the conventional XGBoost algorithm. As a part of the real-time application of the Class-Centric Focal XG Boost model, the model can be utilized in other communicable and communicable disease classification and detection.

Keywords: Class-Centric Focal XGBoost; Minimized Error; Diabetes; COVID-19; Machine Learning; Classification.

DOI:

1. Introduction

Reports since December 2019, the pandemic has evolved making informed infections to about 400 million people in 224 countries [1]. These asymptomatic diseases result in distress syndrome, severe symptomatic complications, and effects including mortality. Concurrently, Diabetes Mellitus (DM) is one of the main risk factors leading to hospitalization and Intensive Care Admission (ICU). These diabetes patients are prone to 2-3 times higher risk rates due to COVID-19 than the effects among the general population. Whereby, type-II diabetic patients suffer from higher rates of multi-morbidity, which is self-associated with the COVID-19 severity [2]. Current pieces of evidence suggest that COVID-19 infections enhance the risk rates of developing diabetes. On the other hand, COVID-19, being curable with severe steroids can result in negative impacts on the diabetes itself, resulting in the worsening of hyperglycemia by increasing insulin resistance and simultaneously reducing the β-cell sections [3]. Though many cases remain undiagnosed, patients suffer from new symptoms during the initial phases of the disease, resulting in a huge challenge for early detection and diagnosis [4]. One of the main advantages of providing the appropriate treatments to patients at their early stage can prohibit expensive treatments and inhibit their case of mortality [5].

The existing study has been implemented with ML algorithms such as SVM, DT, and RF for classifying diabetes with other supervised algorithms [6]. Whereas, the conventional approach has been employed to classify diabetic retinopathy, using the 2D wavelet transform method to prevent the information loss task from the data and to reveal the characteristic features [7]. Some effective pre-processing techniques have been carried out to enhance the visuals of the funds images. The use of CNN with Singular values for decomposition for performing classification has been stated in prior research [8]. CNN has been employed for segmentation level for the evaluation of DR probabilities in the early stage and has been stated in the existing research [9]. Data mining methods such as the KNN and SVM have been used in the classification of DM and attained 90.2 % of accuracy [10]. Type-II diabetes prediction has been carried out in the suggested approach using the RF model with different test samples and attained an overall AUC rate of 0.90 [11]. ML techniques have been applied for diabetes prediction using the PIMA dataset, for early diagnosis of diabetes [12]. Whereas, Type-II diabetes is identified using the Deep NN classifier and the test has been carried out in the PIMA dataset bagged with extra trees and RF for the feature selection [13]. The conventional approach intensifies the Decision Support Systems, known as DSS, upon the clinical decision-making. FCNN has been adapted for the detection and classification of diabetes patients [14].

People with high diabetes are at the maximum risk for severe complications if they are affected by COVID-19. By predicting glucose levels effectively, the healthcare professionals are able to manage diabetes patients more effectively and can help reduce the risk of severe COVID-19 complications [15]. The traditional approaches deployed to identify diabetes using AI approaches direct to severe complications. Henceforth, it is essential to predict the diabetes accurately [16]. The current research deploys a smart prediction tool to forecast glucose levels in patients with diabetes. The present research contributes to making precise predictions of sugar levels, blood, and diabetes risk that can direct to earlier detection of the diseases. Moreover, by understanding the individual variations in glucose levels, health professionals provide more personalized treatment plans for patients. Subsequently, the early and precise predictions avoid the complications of diabetes. Additionally, the class–Centric Focal XGBoost mechanism exemplifies the robustness of advanced ML techniques in the healthcare domain. The current research has paved the way for better predictions that can diminish the need for extensive medical tests and treatments by saving more time and cost consumption for the patients and the healthcare sectors.

1.1. Motivation

Diabetes is reasoned to be the most hazardous disease worldwide. In the former days, the prediction of diabetes was done by manual methods such as blood tests and imaging tests which led to blunders and the process consumed more time to detect the disorder [17]. It results in late prediction and more time to suggest required treatment [18]. To overcome these complications, the existing AI techniques were applied, but the results which were manipulated did not match with the exact solutions. In order to overcome the mentioned drawbacks, the proposed model initiates to the implementation of efficient ML techniques to diagnose diabetes. Finally, the performance of the proposed model can be evaluated by using multiple metrics.

1.2. Research Problem

Even though the conventional works have established considerable outcomes for the classification of diabetics and regression of glucose levels [19], the prevailing works have lacked the accuracy of the models and also hindered the performance of the model [20]. Moreover, the Class-Centric Focal XGBoost model has engrossed in diminishing the overfitting problems faced in conventional research by using effective algorithms. Subsequently, the combination of diabetics and the prediction of glucose levels before and after COVID-19 has significantly paved the way for the contemporary dimension in the proposed work. Consequently, the Class-Centric Focal XGBoost model works in predicting the presence and absence of diabetic disorder, and entirely the Class-Centric Focal XGBoost model relies on the ranges of glucose present in the individuals. The classification process has been carried out in two different extents of individuals before and after COVID-19. Additionally, the regression process is carried out using the Modified XGBoost algorithm. The overall performance of the model indicates the classification and the regression of DM, which is estimated using performance metrics.

1.3. Research Objectives

The research objectives of the proposed mechanism are deliberated as follows:

- To classify the diabetic disorder using Class-Centric Focal XGBoost model in a patient before and after COVID-19.
- To deploy the regression process in diabetes prediction using the Class-Centric Focal XG-Boost XGBoost model.
- To examine the performance of the Class-Centric Focal XG-Boost model using performance metrics.

1.4. Paper Organization

The proposed research is organized into five sections. In section II, the review of the preceding research will be elucidated along with their pitfalls. Following this, section III will exemplify the proposed methodology with appropriate flow, algorithms, and mathematical derivations. Subsequently, in section IV, the outcomes accomplished through the classification will be established and discussed with conventional works outcomes. Eventually, in section V, the overall research with future recommendations will be concluded.

2. Materials and Methods

According to the reports of the Diabetes Federation, about 382 million people survive with diabetes worldwide. Over the past few decades, the impact of diabetes has been increasing drastically, resulting in a global threat. However, diabetes is one of the largely avertible and can be prohibited from severity at early detection and classification. Thus, a dire need of effectual classification model is needed to produce effectual outcomes and to reduce the future complexity and adversity of diabetes.

With advancements in research, many ML techniques have evolved under various studies for the estimation and classification of diabetes. however, these studies laid back in bringing effectual outcomes in aspects of accuracy and are associated with high false-positive rates. Thus, to overcome the pitfalls, the projected study aims to encounter the classification of diabetic and non-diabetic individuals based on the discrete ranges of glucose rates. The proposed study initially collects the real-time data collected using the sensors and they are loaded to the model for effectual pre-processing approaches.

These are endured with the removal of missing values and procedures of data encoding. Essential values in the data used by the model are assured and data encoding is done in aspects of converting the text characters to the binary forms of data which are easily understood by the model. This is followed by the train and test data split. The entire data used by the model is split in an 80:20 ratio for train and test data correspondingly. The train data is used in training the model and attest data is used in validating the model performance. The Modified XG-Boost encompassing the class-centric focal mechanism is encountered for performing both the classification and the regression of diabetic and non-diabetic. The prediction made by the model upon the classification and the regression rates of the diabetic and non-diabetic individuals are estimated using the effectual and applicable metrics estimating their working performance. The error rates of the model are also estimated to affirm the efficacy of the working model. The complete working of the proposed model is presented in Figure 1.

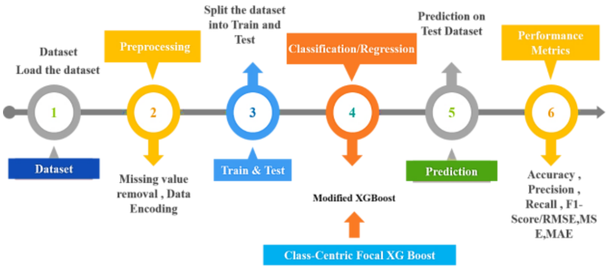

Figure 1. Overall flow of the proposed study

2.1. Dataset Description

The dataset used in the proposed work is obtained using the Arduino Uno kit. This is done under a sampling rate of 100Hz. The data are collected from Bharati Hospital Pathology Laboratories, located in Pune. Some of the essential attributes from the dataset collection are tabulated in Table 1.

The dataset was attained using the Arduino Uno Kit, data are gathered from the Bharati Hospital Pathology Laboratories in Pune. The data are collected using the Near-Infrared (NIR) sensor with a waving of 950 mm and a sampling size of 100 Hz. The Photoplepthysmogram (PPG) signal is considered the primary signal that was used to record over a duration of time. Moreover, the 10-sample average filter has been applied to smooth the signal.

2.1.1. Inclusion and Exclusion Criteria

The inclusion criteria involve individuals of several age groups with or without a diagnosed history of diabetes. The exclusion criteria might comprise individuals with other underlying health conditions that could affect glucose levels or individuals.

2.1.2. Data Splitting

The dataset is separated into training and testing sets in a ratio of 80:20, respectively. In the training set, a portion of data is reserved to tune the hyperparameters and assess the model performance during the training phase. However, external validation is not possible as the proposed work uses a real-time dataset.

The selection, eligibility, and generalizability of the sample set for the Class-Centric Focal XGBoost model are utilized in classifying diabetic and non-diabetic individuals which are essential for precise model performance. The selection of the sample set confirms the relevant healthy records. A balanced sample size is deployed to represent both diabetic and non-diabetic individuals. This selection process enhances the Class-Centric Focal XGBoost model reliability in forecasting health outcomes across various diverse populations. The Class-Centric Focal XGBoost model confirms the comprehensive data that involves capturing the relevant features like blood glucose levels, insulin sensitivity, BMI, and age. This eligibility criteria confirm the Class-Centric Focal XGBoost model accuracy and generalizability in predicting diabetic and non-diabetic outcomes. With the aid of cross-validation techniques, the Class-Centric Focal XGBoost model performance is accessed and ensured that the model is not overfitting to a particular dataset. Moreover, the Class-Centric

Table 1. Attributes of the Dataset

Attribute	Description
Sensor for data collection	Data were collected using the NIR sensor
Essentials for data collection	950mm wavelength of transmitter and receiver with timer routing
Signal for data	PPG signal at a time duration of 1 minute
Filer for signal	Avg. filter of 10 samples is used in smoothening the signal
Sampling Rate	100 Hz for 10 MS is adapted

Focal XGBoost model is flexible to wider the real-world scenarios and confirm the accurate diabetic and non-diabetic classifications.

2.2. Pre-Processing

Pre-processing steps were applied to ensure data quality and compatibility with the model. This included the application of a 10-sample average filter to smooth the PPG signal and the removal of any missing values. Additionally, data encoding techniques were employed to convert textual characters into binary forms suitable for model input.

2.3. XG-Boost Algorithm

The XG-Boost is one of the algorithms used extensively in the construction of additive models. XG-Boost is one of the Gradient Boost (GB) algorithms that completely relies upon the DT classifier. These are fast, effective, and with higher rates of scalability. The features are selected in the feature engineering which are then used in the classification and are estimated. The feature importance provided by the XG-Boost is to filter the effectiveness of the features in aspects of increasing the accuracy of the overall model performing classification. In XG-Boost, each of the individual trees is created using the multiple cores, and the data are organized in aspects of minimizing the time of classification. The deceased time of training enhances the model performance.

The action of sparse aware implementation which is used in handling of missing values also ensures the block structure to the parallelization of the entire tree construction and the training of the model which further boosts the previous fitted model to the data. Thus, XG-Boost is one of the dominant structures for both the structured and tabular forms of the dataset which performs both classification and regression for predicting the modeling problems.

Although XG-Boost has several advantages such as reduced complexity, increased evaluation rates, and faster performance, the model is devoid of making less number of tree generation, where the terms of classification are done rapidly and with less hierarchy rates. Also, the model performs the sequential form of generation which ends up with a higher error rate, which sequentially reduces each of the model functions. Thus, to overcome the pitfall the conventional XG-Boost is modified to compute various measures and support the model learning for classification.

2.4. Proposed Class-Centric Focal XG-Boost

The proposed class-centric Focal XG-Boost is used in elevating the model performance by measuring the similarity among the features, by only measuring the class-based fitness levels of the features for each of the features. These are oriented to numerical and the class oriented categorical fitness which are used in measuring the suitable features for effective classification.

The ability of the proposed Class-Centric Focal XG Boost model aids in reducing the false positives in diabetic prediction thereby enhancing the classification accuracy. The complete working of the proposed XG-Boost is pictured in Figure 2.

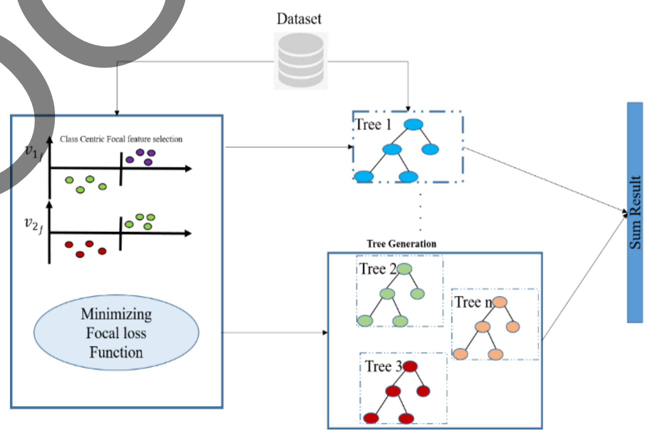

Figure 2. Working Architecture of the proposed Class-Centric Focal XG Boost

The features are classified depending on their types such as numerical, binary, and categorical. The prediction of diabetes was likely performed using various features extracted from the collected data. These features could include physiological parameters related to glucose metabolism, as well as other relevant health metrics. The features such as mean-kte, var-kte, iqr-kte, skew-kte, mean-pks, iqr-pks, vare-pks, skew-kte, and glucose levels are included. For each of the features, the histogram value is computed and a set of unique values are returned. The fitness score generated aids in selecting the optimal features for the generation of trees and for further classification. The complete working algorithm of the

Class-Centric Focal XG-Boost is tabulated Algorithm-I.

Algorithm-I Class-Centric Focal XG-Boost
Data: Dataset and hyperparameters
Initialize $f_o(x)$;
For $k = 1,2, \ldots .\, M$ do //M − number of base learners
Preprocessed data
$F \rightarrow$ Feature list
Find Feature Types $\mathrm{FType} = \sum_{k=1}^{size(F)} F(k).\mathrm{Type}$
Initialize Feature set F_s
For each feature f
If Ftype == categorical, then

compute class oriented categorical Fitness measure (CC).
$CC = size\ (Fhist) > \left(\frac{1}{8} * size(Pds)?\, 1: 0\right)$
If $CC > 0$
$F_s = FS\ F$
$hgs = Hgs \cup F$
End
Else
Add to the set $F_s = FS \cup F$
End
Calculate $s_k = \frac{\partial L(y,f)}{\partial f}$;
Calculate $v_k = \frac{\partial^2 L(y,f)}{\partial f^2}$;
Determine the binary splits with maximized gain
$A = \frac{1}{2}\left[\frac{S_{Left}^2}{V_{Left}} + \frac{S_{Right}^2}{V_{Right}} - \frac{S^2}{V}\right]$;
Determine the leaf weights $w^* = -\frac{S}{V}$;
Determine the base learner $\hat{b}(x) = \sum_{j=1}^{T} wI$;
Add trees $f_k(x) = f_{k-1}(x) + \hat{b}(x)$;
End
Result: $f(x) = \sum_{k=0}^{M} f_k(x)$

The proposed algorithm is then carried out for the model training which is trained using the selected features. This is followed by the procedure of minimizing the function of focal loss upon the provided training tree. The minimization of the focal loss in the model is carried out by making the binary form of the dataset to reduce the ranges of mismatching and the training and testing samples, which are generally associated with affecting accuracy rates. Thus, the proposed XG-Boost is modified in aspects of the boosting, which enhances the model performance to a greater extent. The complete working of the proposed Class-Centric Focal XG Boost is pictured in Figure 3.

Finally, the model performing the classification and the regression are estimated by adding the new tree to the model, for Ensembling and enhancing the classification rates of the model. The overall model is evaluated with suitable test features which are estimated using the applicable performance metrics establishing the model capability for performing classification and regression. The complete outcomes from the model after the evaluation are presented in the following sections.

3. Results and Discussion

The overall outcomes of the proposed model with their appropriate EDA among classification and regression, with the suitable dataset used in the study are deliberated in the following section. Moreover, the comparison among the models have also been made in the following sections.

3.1. Exploratory Data Analysis (EDA)

The EDA is employed in the case of authorizing the data used by acclimatizing various methodologies for visualization. EDA is utilized in defining the patterns to legalize the expectations by implementing a graphical representations and statistical summaries. Moreover, EDA offers data that helps in considering the dataset. The suitable EDA from the model

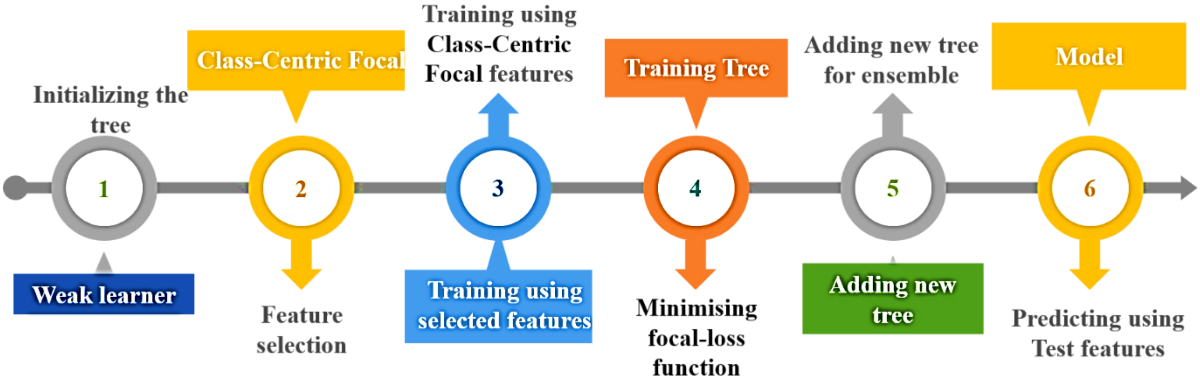

Figure 3. Overall flow of the proposed Class-Centric Focal XG Boost

performance performing the classification and the regression of diabetes for both the population count before and after COVID-19 are presented in the corresponding section.

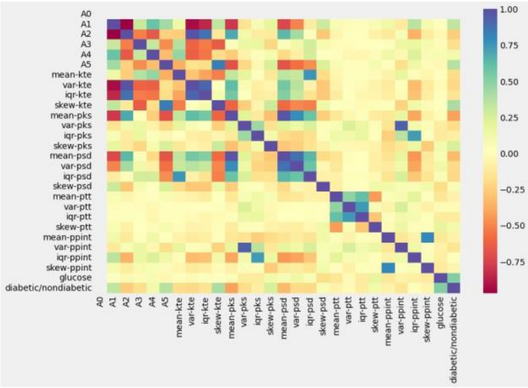

Figure 4. Heat Map of proposed model before COVID-19

Heat map figured in Figure 4 is used as a primary graphical representation for the data used by the model and their performance. These are used in examining several forms of analysis and often use a color coding for representing various values. The proposed model for performing the classification and regression of the diabetic and non-diabetic depending on the regression values of glucose is presented using various attributes.

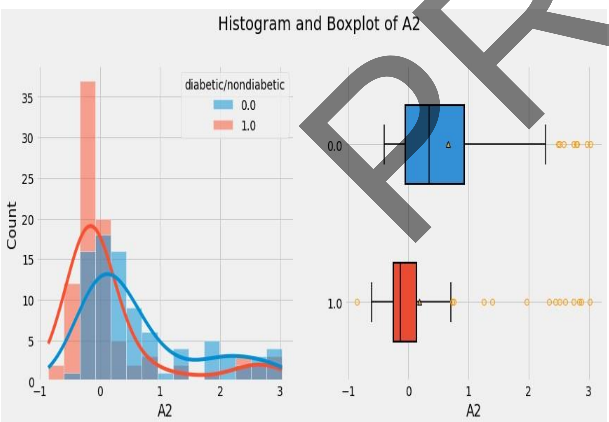

Figure 5. Histogram and Boxplot of A2 - before COVID-19

Figures 5 and 6 represent the histogram plot and the boxplot which are used as the graphical representation for a grouped frequency distribution present with continuous class. Whereas, the box plot is used in deriving the summary of the set of data used by the model for performing tasks. The figures represent the presence and absence of diabetes in an individual by considering the glucose rate and projecting the

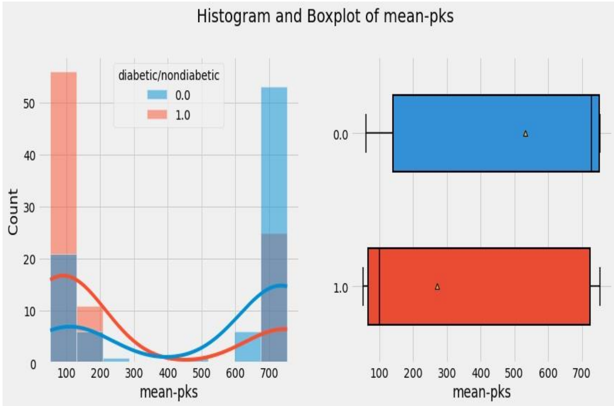

Figure 6. Histogram and Boxplot of mean-PKs

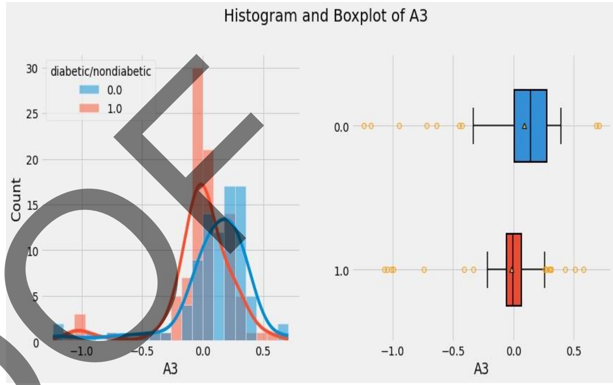

Figure 7. Histogram and Boxplot of A3 - before COVID-19

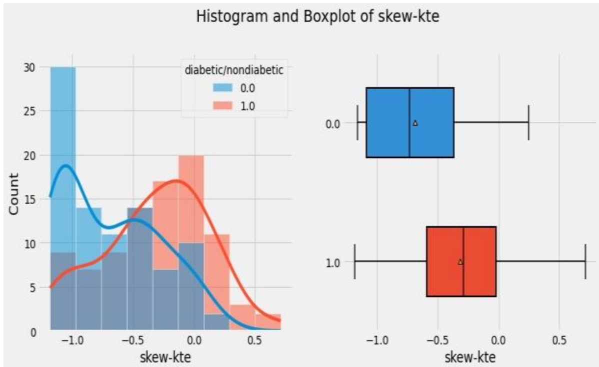

Figure 8. Histogram and Boxplot of Skew-Kte

classification as 1 and 0 indicating diabetic and non-diabetic in blue and red shades.

Figure 7 indicates the same histogram and the box-plot representation of the model performing the classification of the diabetic and non-diabetic, whereas Figure 8, the skew-Kte, commonly known as Kaiser Teager Energy. These are used in affirming the efficacy of the working model performing the classification of diabetes. Finally, Figure 9 represents the same histo and box-plot for the proposed model with respect to the A1 attribute.

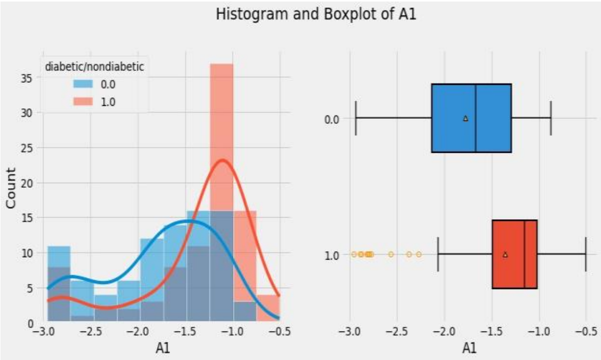

Figure 9. Histogram and Boxplot of A1- before COVID-19

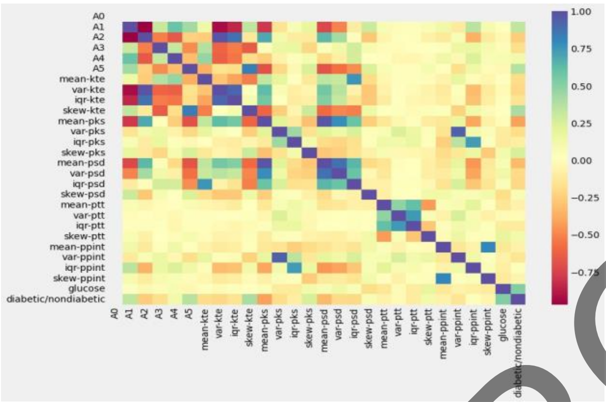

Figure 10. Heat Map of proposed model - COVID-19

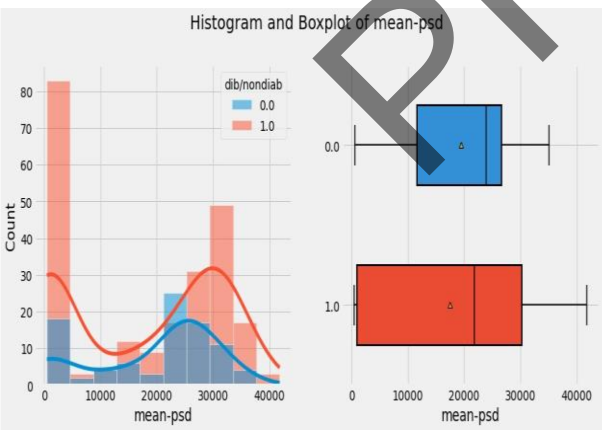

Figure 11. Histogram and Boxplot of mean-Psd

[Figure 10](#) and [Figure 11](#) indicate the heat map for the same model performing the classification among diabetic and non-diabetic considering various other essential attributes, after the COVID-19 pandemic. The model has been implanted for classifying the diabetic presence and absence after the COVID-19 period to a population of 319. Whereas, the histogram and the box plot for the same model performing classification and regression for diabetic and non-diabetic are indicated by considering the mean-Psd rates.

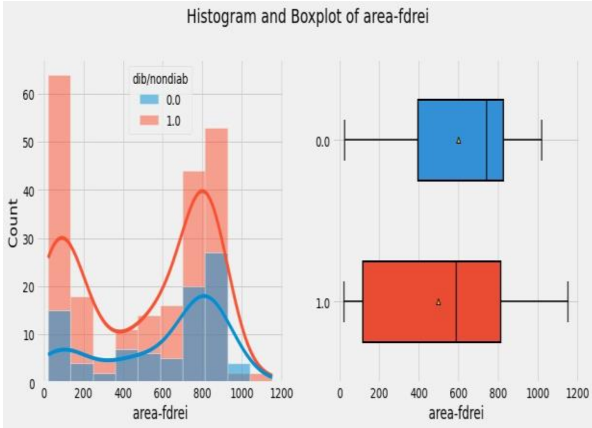

Figure 12. Histogram and Boxplot of area-fdrei

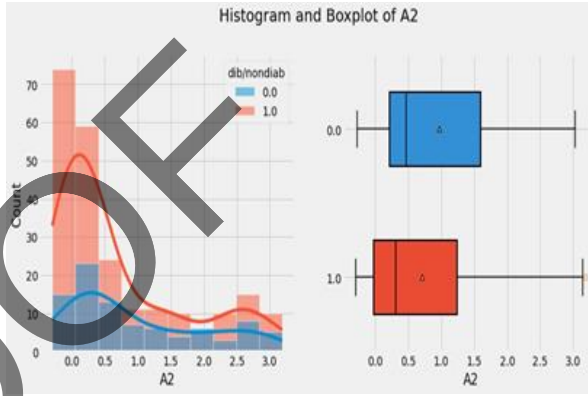

Figure 13. Histogram and Boxplot of A2 - after COVID-19

Figure 12 and Figure 13, concurrently, the Histogram and Boxplot for the proposed model by considering the area-fdrei and the A2 attributes are pictured. These are used for evaluating the wellness of the working model performing the classification and regression.

Figure 14 represents the count plot for the proposed model. The model is designed for performing both the classification and regression of the presence and absence of diabetes, depending on the glucose rates and various other essential attributes such as mean, variance, and skew rates for PKs, Ptt. The blue graph indicates the individuals with diabetes, and the red indicates the individuals with no diabetes.

From the outcomes, each of the EDA indicates the effective working of the proposed model by

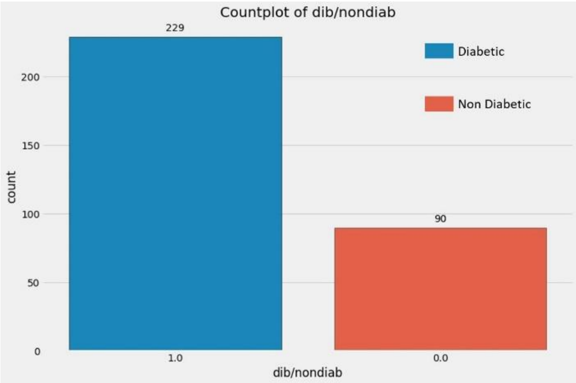

Figure 14. Count plot of the proposed model

considering each of the attributes and their effectual working for performing diabetic classification before and after the COVID-19 period. The following section deliberates the outcomes based on the model working with their error rates.

3.2. Experimental Outcomes

The corresponding section encounters the outcomes of the proposed model based on their metrics such as accuracy, affirming the model affinity, precision rates, and recall scores, followed by the F1-score of the proposed model.

Table 2. Performance Metrics of the Proposed Model

COVID-19 patient	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Before	92	92	92	92
After	95	93	95	94

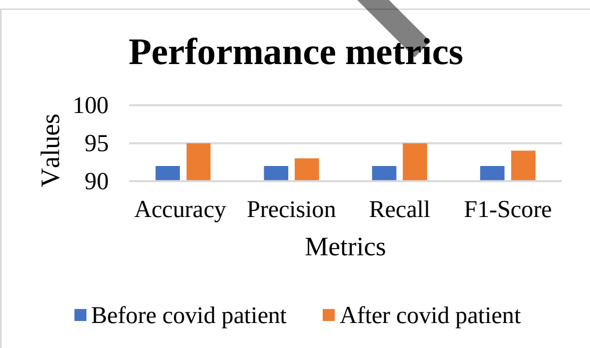

Figure 15. Graphical Representation of the proposed model

From Figure 15 and Table 2, the proposed model establishes their effectiveness and efficacy of the model performing the classification of the diabetic and non-diabetic depending upon various attribute considerations. The overall accuracy rates of the model for the model performing classification before and after COVID-19 are 92% and 95%, respectively. Accuracy being one of the vital attributes used in measuring the efficacy of the model affirms the model's capability upon classification. Whereas, precision used in affirming the overall measure of the quality of the model being proposed for the task affirms the model quality at a rate of 92% and 93% for the classification before and after COVID-19, respectively. Followed by, the Recall used in the performance metrics, for measuring the completeness of the overall positive prediction, made by the model. Correspondingly, the model has made an effectual outcome in the range of the same 92% and 95%, respectively. Finally, the F1-score evaluating the model efficacy are in rates of 92% and 94%, respectively. Each of the attributes at a higher rate affirms the effective working of the model. Table 3 represents the error obtained by the proposed model before and after the patient is affected by COVID-19.

Table 3. Error Rates of the Proposed Model

	MAE	RMSE
Before COVID-19 patient	0.374	0.514
After COVID-19 patient	0.334	0.478

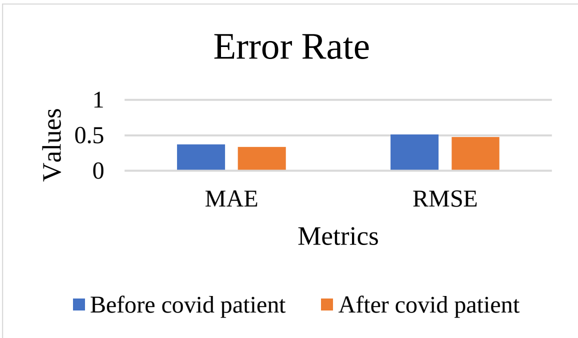

Figure 16. Graphical Representation of Error Rates of the Proposed Model

From Figure 16, minimal error rates for the model performing the regression affirm the effectual working of the model. The error rates of the proposed model are notably lower for both the before and after COVID-19 period diabetics prediction. The MAE rates of the proposed model are 0.374 and 0.334 respectively, being 0.04 times lower than the model after COVID-19 prediction. Whereas, the RMSE rates of the model are 0.514 and 0.478 are the error rates of tem doe before and after COVID-19 period prediction. About, 0.036 times, the model has less error than the regression made by the model affirming the efficacy of the model.

From the outcomes of both the performance metrics, both the classification and the regression made by the proposed model are effective and are suitable for making effectual predictions with less error rates comparatively.

3.3. Discussion

The most widespread form of NCD worldwide is diabetes diseases. The early identification of the diseases utilizes the AI-based approaches and it leads to complications. In order to overcome the complications, the IoMT techniques are deployed by correlating with the most reliable ML approach named Class-Centric Focal XG-Boost, respectively [21]. The Class-Centric Focal XG-Boost model is a promising tool for diminishing the error rates during regression and classification. Moreover, the proposed model engrosses over the proximal features and maintains the balance for better classification and regression outcomes [22]. The conventional research [23] engrossed on diabetes prediction through the real-time dataset gathered from the Shahid Beheshti University of Medical Sciences in Tehran. The data are analyzed using the DT model and attained an accuracy of 87%. The prior research [24] focused on diabetes anticipation through the real-time dataset gathered from the patient demographics and clinical data and patient demographics, clinical data, and blood test results. The CatBoost classifiers are used in the prediction model and attained an accuracy of 92%. However, the conventional models have attained moderate accuracy. Moreover, the sample size is relatively small and the data gathering phase was carried out during the COVID-19 crisis, and the data cleaning and pre-processing phase took a long time. These are the limitations. To resolve the limitation, the Class-Centric Focal XG-Boost model shows outstanding performance in terms of accurately forecasting the prediction of diabetes. The proposed model will be enhanced by its robustness and performance across the diverse populations of the datasets. Moreover, additional features and data sources are included to improve the proposed model's predictive reliability and accuracy. The longitudinal studies will be conducted to examine the long-term effectiveness and clinical utility of the proposed model in the healthcare domain are the future recommendations of the research.

4. Conclusion

The proposed study used Class-Centric Focal XG-Boost for classification and regression. The model is constructed for performing the classification among diabetic and non-diabetic individuals. This was done by effectual consideration of discrete ranges of glucose rate of individuals. The model affirmed its efficacy upon each of the performance metrics with minimal error rates. The proposed Class-Centric Focal XG-Boost model demonstrated notable accuracy rates of 92% and 95% for classifying diabetic and non-diabetic individuals before and after the COVID-19 pandemic, respectively. The model also exhibited minimal error rates in regression, with MAE rates of 0.374 and 0.334, and RMSE rates of 0.514 and 0.478 for predictions before and after the COVID-19 period, respectively. Moreover, the present model emphasized the classification of diabetes disorder. The regression process in the diabetic prediction is accomplished and the performance of the model is evaluated using the performance metrics highlighted in the research. The findings of the study stated that the proposed model offered an effective tool for diabetes classification, with potential applications in healthcare settings for early detection and personalized treatment. The model's high accuracy and minimal error rates contribute healthcare professionals in making informed decisions, thereby improving patient outcomes and reducing the burden on healthcare systems.

References

1- Aidin Rawshani *et al.*, "Severe COVID-19 in people with type 1 and type 2 diabetes in Sweden: A nationwide retrospective cohort study." Vol. 4(2021).

2- Jason I Chiang *et al.*, "Multimorbidity, mortality, and HbA1c in type 2 diabetes: A cohort study with UK and Taiwanese cohorts." Vol. 17 (No. 5), p. e1003094, (2020).

3- Cyril P Landstra and Eelco JP %J Frontiers in endocrinology de Koning, "COVID-19 and diabetes: understanding the interrelationship and risks for a severe course." Vol. 12p. 649525, (2021).

4- Victor Chang, Jozeene Bailey, Qianwen Ariel Xu, Zhili %J Neural Computing Sun, and Applications, "Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms." Vol. 35 (No. 22), pp. 16157-73, (2023).

5- Nazin Ahmed *et al.*, "Machine learning based diabetes prediction and development of smart web application." Vol. 2pp. 229-41, (2021).

6- Isfafuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam, and Riasat %J Healthcare Technology Letters Khan, "Diabetes prediction using machine learning and explainable AI techniques." Vol. 10 (No. 1-2), pp. 1-10, (2023).

7- Yusuf Bahri Özçelik, Aytaç %J Fractal Altan, and Fractional, "Overcoming nonlinear dynamics in diabetic retinopathy classification: A robust AI-based model with chaotic swarm intelligence optimization and recurrent long short-term memory." Vol. 7 (No. 8), p. 598, (2023).

8- Anas Bilal, Liucun Zhu, Anan Deng, Huihui Lu, and Ning %J Symmetry Wu, "AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning." Vol. 14 (No. 7), p. 1427, (2022).

9- Laxmi Math, Ruksar %J Multimedia Tools Fatima, and Applications, "Adaptive machine learning classification for diabetic retinopathy." Vol. 80 (No. 4), pp. 5173-86, (2021).

10- Abdulhakim Salum Hassan, I Malaserene, and A Anny %J Int. J. Innov. Technol. Explor. Eng Leema, "Diabetes mellitus prediction using classification techniques." Vol. 9 (No. 5), pp. 2080-84, (2020).

11- Jake A Carter, Christina S Long, Beth P Smith, Thomas L Smith, and George L %J Expert Systems with Applications Donati, "Combining elemental analysis of toenails and machine learning techniques as a non-invasive diagnostic tool for the robust classification of type-2 diabetes." Vol. 115pp. 245-55, (2019).

12- Huma Naz, Sachin %J Journal of Diabetes Ahuja, and Metabolic Disorders, "Deep learning approach for diabetes prediction using PIMA Indian dataset." Vol. 19pp. 391-403, (2020).

13- R K Nadesh and K %J International Journal of Cognitive Computing in Engineering Arivuselvan, "Type 2: diabetes mellitus prediction using deep neural networks classifier." Vol. 1pp. 55-61, (2020).

14- Amani Yahyaoui, Jawad Rasheed, Akhtar Jamil, and Mirsat Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques."

15- Kamlesh Khunti *et al.*, "The impact of the COVID-19 pandemic on diabetes services: planning for a global recovery." *The Lancet Diabetes & Endocrinology,* Vol. 10 (No. 12), pp. 890-900, (2022).

16- Munish Bhatia, Simranpreet Kaur, Sandeep K Sood, and Veerawali Behal, "Internet of things-inspired healthcare system for urine-based diabetes prediction." *Artificial Intelligence in Medicine,* Vol. 107p. 101913, (2020).

17- Boon Feng Wee, Saaveethya Sivakumar, King Hann Lim, Wei Kitt Wong, and Filbert H Juwono, "Diabetes detection based on machine learning and deep learning approaches." *Multimedia Tools and Applications,* Vol. 83 (No. 8), pp. 24153-85, (2024).

18- Monalisa Panda, Debani Prashad Mishra, Sopa Mousumi Patro, and Surender Reddy Salkuti, "Prediction of diabetes disease using machine learning algorithms." *IAES International Journal of Artificial Intelligence,* Vol. 11 (No. 1), p. 284, (2022).

19- Saloni Kumari, Deepika Kumar, and Mamta Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier." *International Journal of Cognitive Computing in Engineering,* Vol. 2pp. 40-46, (2021).

20- Nazin Ahmed *et al.*, "Machine learning based diabetes prediction and development of smart web application." *International Journal of Cognitive Computing in Engineering,* Vol. 2pp. 229-41, (2021).

21- Leila Ismail, Huned Materwala, Maryam Tayefi, Phuong Ngo, and Achim P Karduck, "Type 2 diabetes with artificial intelligence machine learning: methods and evaluation." *Archives of Computational Methods in Engineering,* Vol. 29 (No. 1), pp. 313-33, (2022).

22- Rung-Ching Chen *et al.*, "An end to end of scalable tree boosting system." *Sylwan,* Vol. 165 (No. 1), pp. 1-11, (2020).

23- Pooneh Khodabakhsh *et al.*, "Prediction of in-hospital mortality rate in covid-19 patients with diabetes mellitus using machine learning methods." *Journal of Diabetes & Metabolic Disorders,* Vol. 22 (No. 2), pp. 1177-90, (2023).

24- Serpil Ustebay, Abdurrahman Sarmis, Gulsum Kubra Kaya, and Mark Sujan, "A comparison of machine learning algorithms in predicting COVID-19 prognostics." *Internal and Emergency Medicine,* Vol. 18 (No. 1), pp. 229-39, (2023).