Synergistic ensemble classification framework: utilizing a soft voting algorithm for enhanced prediction and diagnosis of diabetes mellitus

Bhanu Prakash Lohani, Arvind Dagur, Dhirendra Kumar Shukla
School of Computing Science and Engineering, Galgotias University, Greater Noida, India

Article Info

Article history:

Received Apr 4, 2024
Revised Sep 27, 2024
Accepted Oct 7, 2024

Keywords:

Diabetes
Ensemble model
Hyperparameters
Predictive performance
ROC
Soft voting
Stacking

ABSTRACT

Diabetes, a serious condition characterized by elevated blood glucose levels, can be effectively identified, and predicted early using machine learning (ML) algorithms. The research provides a comprehensive assessment of three ensemble ML models-stacking, soft voting, and hard voting-focused on enhancing diabetes diagnosis among Pima Indian women dataset taken from the National Institute of Diabetes and Digestive and Kidney Diseases, this study focuses on Pima Indian women aged 21 and older, with the dataset comprising critical diagnostic measurements. Two ensemble models were developed and evaluated on various evaluation parameters. The stacking model combines predictions from various classifiers using a meta-classifier, leveraging their strengths for final decision-making. In contrast, the voting model aggregates probability estimates from each classifier, providing nuanced predictions. Both models were rigorously evaluated on a validation dataset, emphasizing accuracy, specificity, sensitivity, and the receiver operating characteristic (ROC) area under the curve (AUC). Notably, the voting-based ensemble methods demonstrated superior performance in predicting diabetes for this cohort. However, their effectiveness heavily relies on preprocessing, base model selection, and hyperparameter optimization. This study underscores the potential of ensemble models in medical diagnostics, highlighting the critical role of data preprocessing, and configuration in enhancing predictive accuracy.

Corresponding Author:

Bhanu Prakash Lohani
School of Computing Science and Engineering, Galgotias University
Yamuna Expressway, Greater Noida, U.P, India
Email: bhanuplohani@gmail.com

1. INTRODUCTION

Diabetes mellitus (DM), a global health crisis, diabetes is marked by elevated blood glucose levels and poses a particularly severe risk to certain ethnic groups, including the Pima Indian women. Timely detection and management of diabetes is essential for reducing its long-term complications [1]. Machine learning (ML) algorithms have become an invaluable tool for improving the early prediction and diagnosis of diabetes, providing a major advancement over traditional approaches. The potential of ML in medical diagnostics is especially clear with the use of ensemble learning models, which harness the strengths of multiple algorithms to enhance predictive accuracy [2].

A range of ML models have been explored for DM classifications, with decision trees and neural networks showing promising predictive accuracy [3]. Deep learning (DL) algorithms, particularly the convolutional neural networks long short-term memory (CNN-LSTM) combination, have also demonstrated high accuracy in this context [4]. The use of ensemble models, such as random forest (RF), has been found to further improve prediction accuracy [5]. However, the performance of these models can vary, with some studies highlighting the effectiveness of models like support vector machines (SVM), ANN, and logistic regression (LR) [6]. Despite these variations, the consensus is that ML techniques hold significant potential for the accurate classification of DM. Phongying and Hiriote [7] models using interaction terms achieved superior performance than models without interaction, with RFC achieving the best results. Ganguly and Singh [8] evaluated six ML algorithms, including LR, decision tree (DT), random forest (RF), k-nearest neighbor (KNN), SVM, and Naïve Bayes (NB), using the Pima Indian diabetes database. Different performance metrics-confusion matrix, accuracy, precision and F-measure, were used for evaluation. Yi [9] used ML modeling and classification for diabetes diagnosis problems. The accuracy rates of the KNN model and the RF model were found to be 76% and 80% respectively. Murthy and Srilatha [10] compared the performance of decision tree classifier (DTC) and LR algorithms and found that LR offered higher accuracy in predicting diabetes. Ismail and Materwala [11] conducted a literature review and proposed an intelligent framework for applying ML to predict diabetes. They evaluated decision tree employing RFs and SVMne models for diabetes prediction [11].

This study delves into the efficacy of three advanced ensemble ML models stacking, soft voting, and hard voting. In diagnosing diabetes among Pima Indian women, the study utilizes a dataset from the National Institute of Diabetes and Digestive and Kidney Diseases. This unique dataset includes diagnostic measurements from Pima Indian women aged 21 and above, selected under stringent criteria to guarantee the relevance and accuracy of the results. The research methodology incorporates exploratory data analysis and preprocessing of the dataset, followed by the application and rigorous evaluation of the ensemble models on the criteria accuracy, sensitivity, specificity, and the area under the ROC curve. Notably, our results indicate that the soft and hard voting models show promise in diabetes prediction. However, these outcomes are influenced by preprocessing techniques, the selection of base models, and hyperparameter settings, highlighting the nuanced nature of ML applications in healthcare [12]. Recent advancements in the field underscore the dynamism of ML in diabetes research. For instance, [13], [14] have further refined ensemble learning techniques, demonstrating improved classification and prediction results in various medical datasets.

This research article contributes by developing and demonstrating an enhanced ensemble model that improves the diagnostic accuracy of DM classification. For assessing the performance of the enhanced model the following objectives are outlined as follows:

a) An ensemble model was developed that integrates RF, LR, NB, and other algorithms using a voting and stacking ensemble classifier. This model is designed to categorize the used dataset into two categories positive or negative.
b) Precision, recall, F1-score, accuracy, and the area AUC as our primary metrics to evaluate the effectiveness of our proposed model.
c) The proposed ensemble model's outcomes exhibit enhanced performance when benchmarked against existing methods while maintaining the defined parameters.
d) The assessment of our approach includes a rigorous comparison with standard base classifiers, including LR, RF, NB, GradientBoost, XGBoost, AdaBoost, and CatBoost, demonstrating the robustness of our proposed methodology.

2. RELATED WORK

The progression of machine-learning techniques has significantly influenced the medical field of diagnostics, In the detection and classification of DM. In studies, various ML models have been rigorously tested and evaluated for their efficacy in diagnosing diabetes with promising results. Hiriote conducted an in-depth analysis of RF, SVM, DT, and KNN, leveraging hyperparameter tuning and interaction terms to enhance model performance. The study found that RFs achieved the highest accuracy, outperforming the other models with a remarkable 97.5% accuracy rate [7]. Vidya *et al.* [15] presented algorithms for disease diagnosis, highlighting that RF and deep neural networks could further increase prediction accuracy. Hassan *et al.* [16] investigated early predictive analytics using LR, SVM achieving up to 97.5% accuracy with the latter. Bhuiyan *et al.* [17] presented a pre-processing method for medical data, resulting in a 92.27% classification accuracy when applied with the NB algorithm. Ismail and Materwala [11] introduced the IDMPF, an intelligent framework, utilizing tree-based SVM models, emphasizing the integration of intelligent systems for diabetes prediction. Yi [9] demonstrated the superior predictive ability of LR in classifying diabetes risk. Deepak *et al.* [18] developed a system using an improved classification technique to predict diabetic risk levels with higher accuracy. Soni [19] employed ensemble ML with principal component analysis (PCA) and K-means clustering, resulting in an ensemble model that outperformed base classifier models. Satu *et al.* [20] suggest a novel hybrid ML model incorporating synthetic minority oversampling and simple K-means clustering, with RF showing the best accuracy at 99.067%. Malviya *et al.* [21] evaluated several ML and DL models for their capability to predict diabetic risk levels, advocating for the use of diverse algorithms for enhanced classification accuracy. Reddy *et al.* [22] applied RF and KNN algorithms to detect diabetes, utilizing the PIMA dataset for their analysis work. Similarly, Das *et al.* [23] explored various classifiers, which include DTC, KNN algorithm, LR, and NB, identifying prognostic biomarkers for diabetes prediction and achieved an accuracy of 98.08% which is very impressive. Uddin *et al.* [24] evaluated model performance across two datasets, determining that the best result was provided by the RF classifier, with an accuracy of 97% on the 2019 dataset. Abdollahi and Aref [25] implemented particle swarm optimization (PSO) for feature selection, followed by a comparison of performance metrics. The above research papers were studied before going towards the successful implementation of this research work.

3. PROPOSED METHOD

In this paper, the objective is to meticulously construct and assess three distinct ensemble models namely the stacking model, the soft voting model, and the hard voting model. A detailed methodology is shown in Figure 1. We have presented the steps of preprocessing in sections 3.1 and 3.2. after the preprocessing and EDA steps, Figures 2 and 3 represent the data distribution and correlation. The algorithmic steps are defined in Algorithm 1 which is for procedure for early diabetes prediction using an ensemble approach.

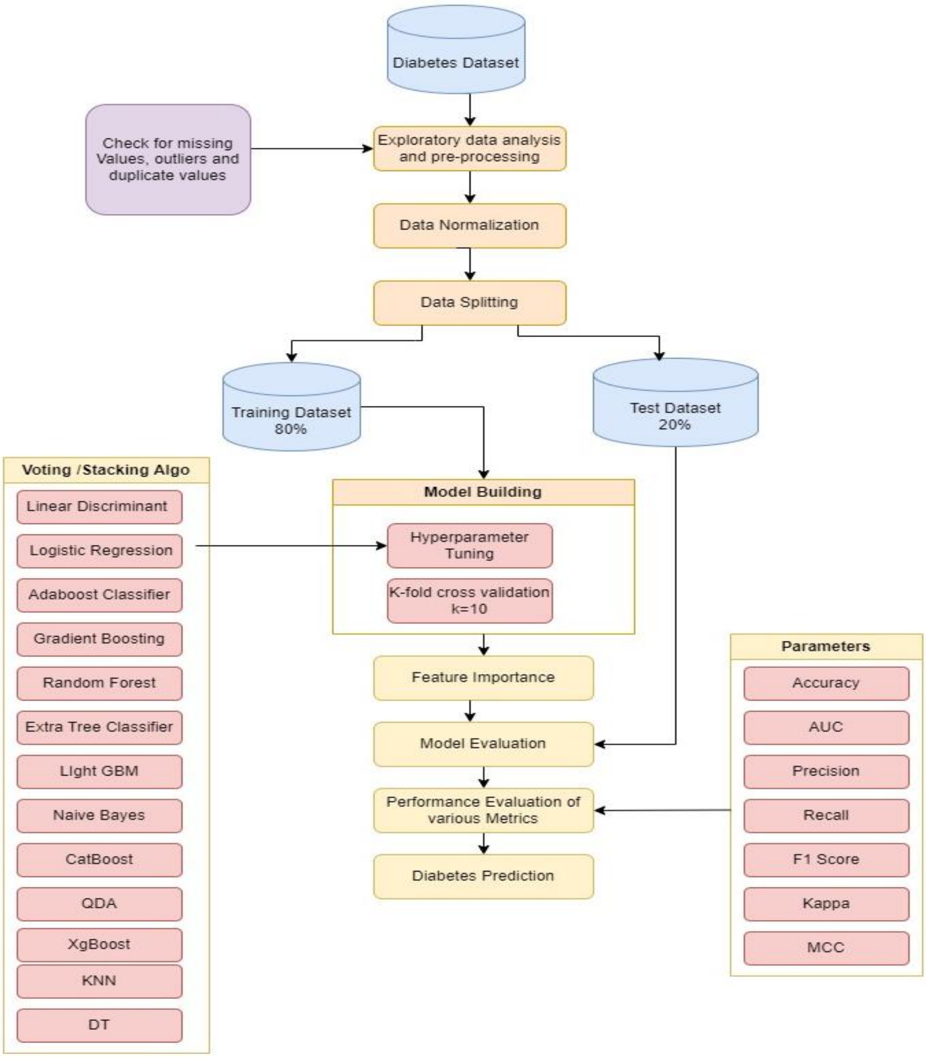

Figure 1. Methodology: flow chart

3.1. Description of the dataset

The data set for the diagnosis of diabetes is taken from the National Institute of Diabetes and Digestive and Kidney Diseases. It encompasses a selection of diagnostic measurements of female patients of the Pima Indian dataset aged 21 years or older. This data set includes 8 features and one outcome. Out of 768 individuals in the data set, 268 are marked with the outcome ‘1’, denoting a diabetes diagnosis, while the rest are labeled as ‘0’, indicating no diabetes. This is a valuable data set that is a resource for constructing predictive models and applying ML techniques to identify patterns and factors significant in the diagnosis of diabetes.

3.2. Data preprocessing step

The dataset underwent a thorough pre-processing regimen to ensure data quality for the subsequent analysis. Initially, the absence of missing values was checked and confirmed that all features were numerical, negating the need for imputation or type conversion. The balance of our target variable and found it to be evenly distributed, eliminating concerns over target imbalance. Observing significant proportions of zero values in the some features, these zeroes were substituted with the respective feature means to avoid distortion in the data distribution. This substitution, along with linear and standard scaling, helped normalize the distribution. However, due to the persistence of skewed distributions in some features, nonlinear scaling using the quantiletransformer to approximate a normal distribution was employed. Further, we identified a notable correlation between ‘Outcome’ and ‘Glucose’, indicating ‘Glucose’ as a potential key predictor. ‘BMI’, ‘Pregnancies’, and ‘Age’ also emerged as important features. A high correlation was observed between ‘SkinThickness’ and ‘BMI’, hinting at possible multicollinearity, which we plan to investigate further if model performance is impacted. Figures 2 and 3 shows the distribution and correlation of different parameters.

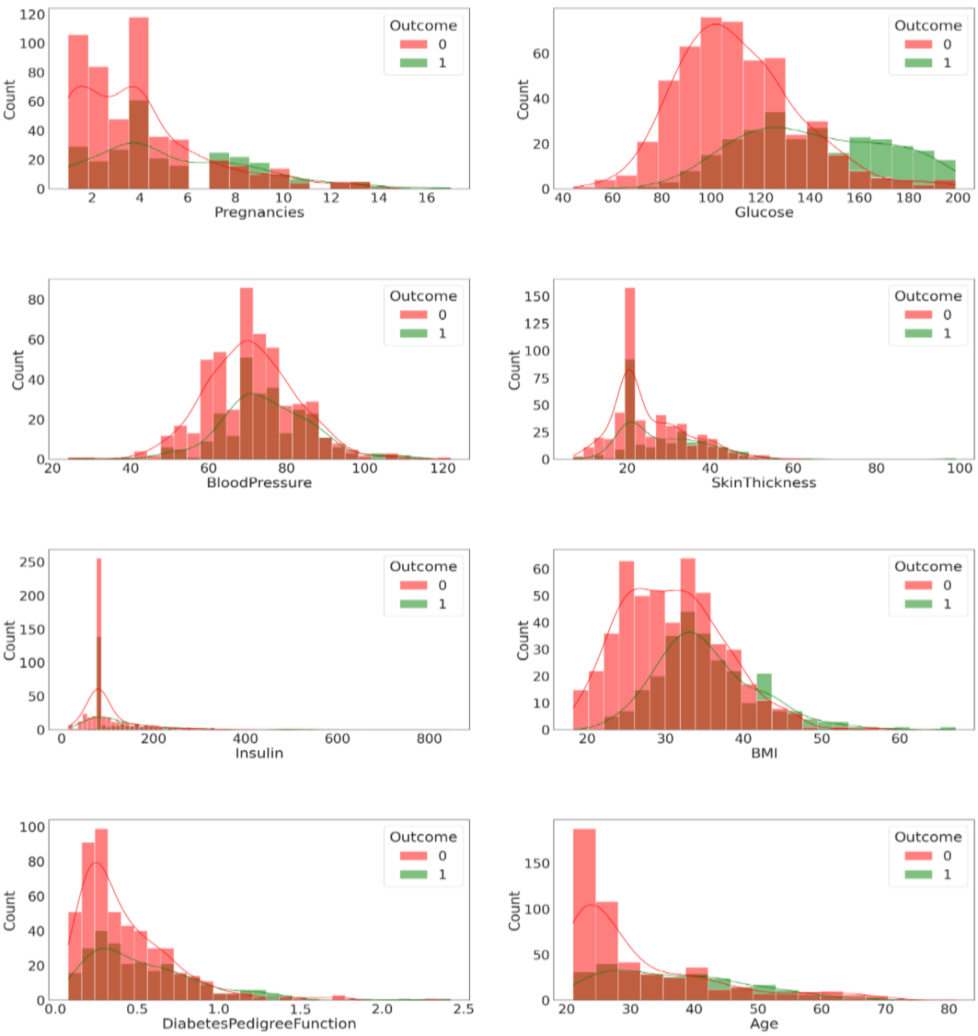

Figure 2. Distribution of different parameters

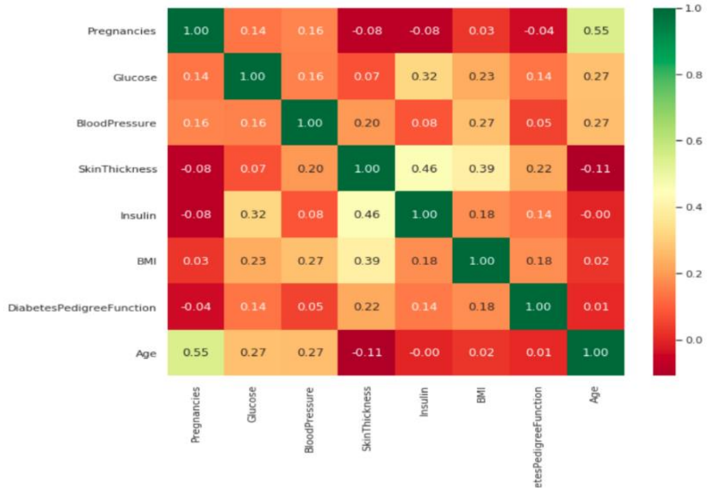

Figure 3. Correlation of different parameters

Algorithm 1. Diabetes prediction using ensemble approach

Input: A PIMA data set containing X features (x_1 to x_8) and one target variable, y.
Output: Prediction of whether the patient is diabetic or nondiabetic.
Steps:

1. Data Preprocessing
   - Collect the data set and load the CSV file.
   - Data Cleaning
2. Feature Extraction and Correlation analysis
   - Correlation analysis using Pearson correlation analysis (Heatmap)
   - Select the higher correlated feature
3. Data Splitting
   - Split the data set (Criteria 80:20 train test ratio)
4. Model Evaluation and Training
   - Model Evaluation done for ML classifier LR, XGBOOST, RF, Schocastic Gradient Descend, Adaboost, Gboost, KNN, and SVM
   - Calculate prediction and find the performance matrices
   - Calculate the metrics as follows:
   - \- Accuracy = (correct predictions / total predictions) * 100
   - \- Precision = (true positives/(true positives + false positives))
   - \- Recall = (true positives / (true positives + false negatives))
   - \- F1-score = 2 * ((Precision * Recall) / (Precision + Recall))
5. Voting Classifier

Assign a weight to each classifier. These weights often represent confidence in the model's predictions, which can be determined by (P_v)

   - Accuracy on a validation dataset.
   - Cross-validation scores.
   - Domain-specific knowledge or heuristics.

6. Stacking
   - New training and Stacking done S= [Prediction of [LR, XGBOOST, RF, Schocastic Gradient Descend, Adaboost, Gboost, KNN, and SVM]
   - Training of Meta learner M on S (Where M =meta learner and S represents stacking of classifiers)
7. Final Prediction & Evaluation
   - Select the best model in the criteria of performance (Voting or Stacking)
   - Final Prediction & Evaluation done.

4. MODEL ARCHITECTURE

We have split the dataset in a ratio of 80:20 for training and testing. In the proposed methodology seven standalone models namely LR, XGBOOST, RF, schocastic gradient descend, Adaboost, Gboost, KNN, and SVM were trained on the dataset. A voting classifier serves as an ensemble method that aggregates the predictions from various models to make a final prediction for new data. In this approach, a weighted voting classifier where each constituent model is assigned a weight proportional to its accuracy was used to make final predictions. Consequently, the model exhibiting the highest individual accuracy will receive the most significant weight, thereby having the most influence on the final prediction. Figure 4 shows the performance metrics for the different ML models and the ensemble voting classifier in terms of accuracy, AUC, precision, recall, and F1-score. The voting classifier achieved an accuracy of 76.17%. Stacking is an ensemble technique that leverages a meta-learner to combine multiple primary models, often of various types, unlike more homogeneous methods like bagging and boosting. This approach involves training first-level models, also known as Level 0 classifiers, on the data and then using their detection as input features for a second-level model, the Level 1 classifier. In this approach, KNN was chosen as the meta-classifier based on accuracy. The Level 1 classifier's role is to discern the optimal way to synthesize the Level 0 classifiers' outputs to make final predictions. A cross-validation (10-fold) was used to achieve the final accuracy. The stacking ensemble approach gave an accuracy of 74.86%.

5. RESULT ANALYSIS AND INTERPRETATION

Figure 4 showcases a respective performance analysis of ML models evaluated on several metrics pertinent to classification tasks. From the Figure 4, it was observed that both the voting classifier and linear discriminant analysis (LDA) exhibit the highest accuracy of 76.17%, with the voting classifier achieving a marginally superior AUC of 84.57% compared to LDA's 84.21%. This suggests that while both models are closely matched in terms of overall accuracy, the voting classifier may possess an improved trade-off between true +ve rate and false +ve rate (+ve positive).

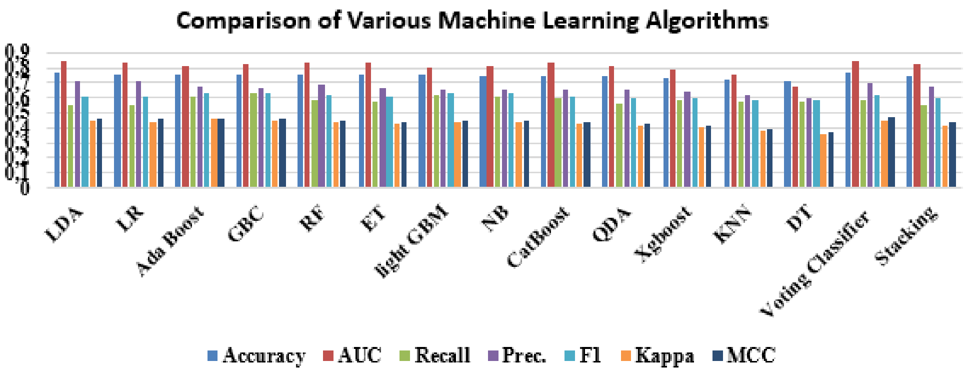

Figure 4. ML algorithms: performance analysis

In terms of speed, LDA is the fastest model with a training time of only 0.03 seconds, an order of magnitude quicker than the LR model, which, despite its comparable accuracy and AUC, is the slowest at 1.131 seconds. The ada boost, gradient boost (GB classifier), and RF classifier show a balanced performance across all metrics, with relatively high accuracy, AUC, and F1-scores, showing a good balance between recall and precision. It is notable that the CatBoost classifier, while not the top performer in accuracy or AUC, requires the longest training time at over 2 seconds. This might be a consideration for real-time applications where model inference speed is critical. The gradient boosting classifier (GBC) and light gradient boosting machine (LGBM) present high F1-scores and AUC, which signifies a strong ability to handle both classes effectively, despite a slight trade-off in terms of training time for LGBM. NB, despite its simplicity, shows commendable performance, especially considering its very short training time, which could make it an attractive option for very large datasets or initial baseline modeling. Figures 5 and 6 present confusion matrices for the stacking classifier and the voting classifier. Figure 7 represents the receiver operating characteristic (ROC) curve comparison for various ML models. The ROC curve is a graphical representation that shows how well a binary classifier performs as its decision threshold changes.

Perfect prediction is represented by an AUC of 1.0, while a 0.5 value of AUC suggests no discriminative power, equivalent to random guessing. The RF and extra trees models exhibit exceptional performance with AUCs of 0.98, indicating they are excellent at distinguishing between the two classes. XGBoost also outperforms with an AUC of 0.97. The voting classifier, ensemble, and GBC show strong performance with AUCs of 0.95. The stacking classifier and the decision tree display robust discrimination abilities with AUCs of 0.93 and 0.94, respectively. Overall, the graph indicates that ensemble methods and tree-based models tend to outperform simpler models in terms of ROC AUC, suggesting that they may be more suitable for this particular classification task. Figures 8 and 9 show the threshold plot for voting and stacking classifier.

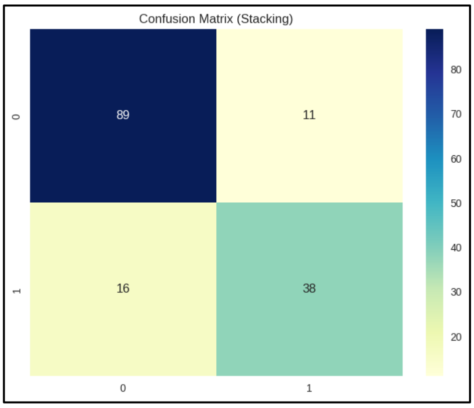

Figure 5. Confusion matrix for stacking

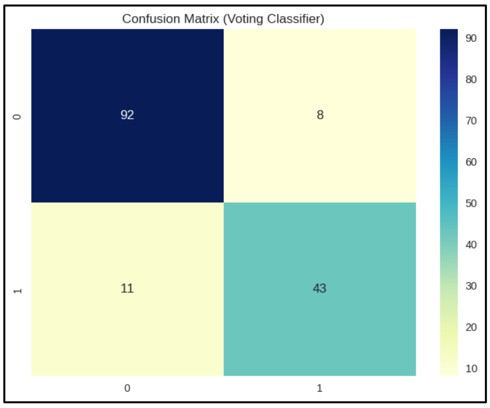

Figure 6. Confusion matrix for voting classifier

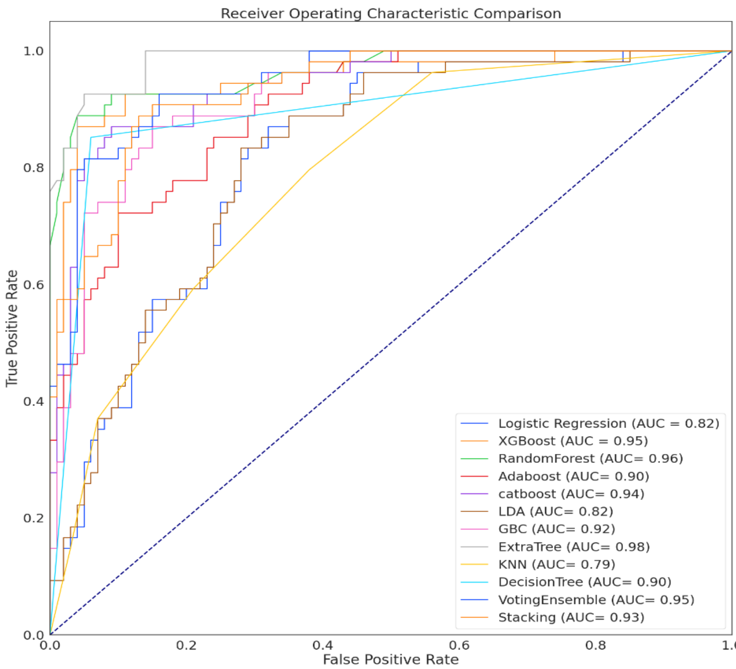

Figure 7. ROC curve for various ML models

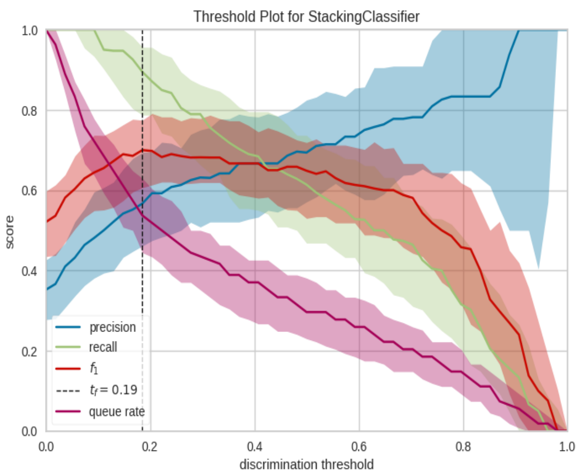

Figure 8. Threshold plot for stacking classifier

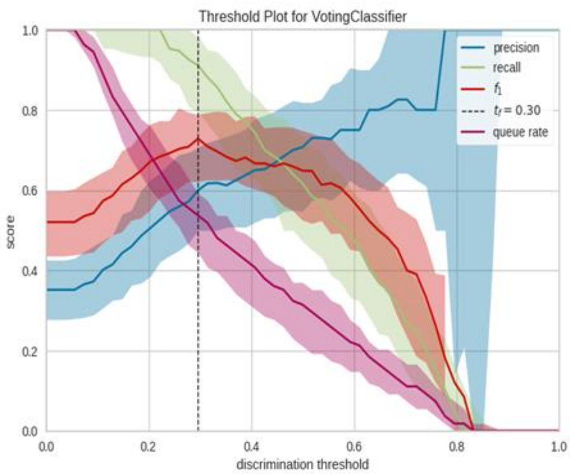

Figure 9. Threshold plot for voting classifier

6. CONCLUSION

This research work showcased the powerful effectiveness of ensemble ML methods in the early detection and diagnosis for DM, specifically within the Pima Indian female population. Our meticulous evaluation of several ensemble models, including stacking and voting, on a rich dataset from the National Institute of Diabetes and Digestive and Kidney Diseases, has yielded significant insights. Among the models assessed, the voting classifier, which combines probability estimates from various algorithms, emerged as the superior model in terms of ROC AUC, highlighting its robustness in predicting diabetes. Additionally, the results advocate for the importance of data preprocessing, as it forms the foundation for the successful application of complex models. Although this research work has made significant contributions to the field of medical diagnostics, further research is warranted. Investigating the scalability of the models in larger and more diverse populations, as well as exploring the integration of additional predictive variables, could enhance the models' diagnostic accuracy. Moreover, analyzing the implementation of these models in real-world clinical environments could offer practical insights into their utility and impact on healthcare outcomes for diabetic patients.

REFERENCES

[1] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in Genetics*, vol. 9, Nov. 2018, doi: 10.3389/fgene.2018.00515.

[2] A. Husain and M. H. Khan, "Early diabetes prediction using voting based ensemble learning," in *Communications in Computer and Information Science*, vol. 905, Springer Singapore, 2018, pp. 95–103.

[3] M. O. Olusanya, R. E. Ogunsakin, M. Ghai, and M. A. Adeleke, "Accuracy of machine learning classification models for the prediction of type 2 diabetes mellitus: a systematic survey and meta-analysis approach," *International Journal of Environmental Research and Public Health*, vol. 19, no. 21, p. 14280, Nov. 2022, doi: 10.3390/ijerph192114280.

[4] K. V. Balaji and R. Sugumar, "A comprehensive review of diabetes mellitus exposure and prediction using deep learning techniques," in *2022 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2022*, Dec. 2022, pp. 1–6, doi: 10.1109/ICDSAAI55433.2022.10028832.

[5] O. Abe, O. Obe, O. Boyinbode, and O. Biodun, "Classifier algorithms and ensemble models for diabetes mellitus prediction: a review," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 1, pp. 430–439, Feb. 2021, doi: 10.30534/ijatcse/2021/641012021.

[6] A. Hussain and S. Naaz, "Prediction of diabetes mellitus: Comparative study of various machine learning models," in *Advances in Intelligent Systems and Computing*, vol. 1166, Springer Singapore, 2021, pp. 103–115.

[7] M. Phongying and S. Hiriote, "Diabetes classification using machine learning techniques," *Computation*, vol. 11, no. 5, p. 96, May 2023, doi: 10.3390/computation11050096.

[8] R. Ganguly and D. Singh, "A comparative analysis to predict diabetes using classification algorithms," in *Lecture Notes in Electrical Engineering*, vol. 1037 LNEE, Springer Nature Singapore, 2023, pp. 271–282.

[9] M. Yi, "Research on machine learning classification algorithms for diabetes data," in *International Conference on Algorithms, Microchips and Network Applications*, May 2022, p. 79, doi: 10.1117/12.2636526.

[10] B. S. Murthy and J. Srilatha, "Comparative analysis on diabetes dataset using machine learning algorithms," in *Proceedings of the 6th International Conference on Communication and Electronics Systems, ICCES 2021*, Jul. 2021, pp. 1416–1422, doi: 10.1109/ICCES51350.2021.9488954.

[11] L. Ismail and H. Materwala, "IDMPF: intelligent diabetes mellitus prediction framework using machine learning," *Applied Computing and Informatics*, Jun. 2021, doi: 10.1108/aci-10-2020-0094.

[12] S. Arukonda and R. Cheruku, "A novel stacking framework with PSO optimized SVM for effective disease classification," *Journal of Intelligent and Fuzzy Systems*, vol. 45, no. 3, pp. 4105–4123, Aug. 2023, doi: 10.3233/JIFS-232268.

[13] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, vol. 8, pp. 76516–76531, 2020, doi: 10.1109/ACCESS.2020.2989857.

[14] G. T. Reddy *et al.*, "An ensemble based machine learning model for diabetic retinopathy classification," in *International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020*, Feb. 2020, pp. 1–6, doi: 10.1109/ic-ETITE47903.2020.235.

[15] J. Vidya, S. T. Jain, S. Boosi, H. C. Bhanujyothi, C. Tukkoji, "Prognosis of diabetes mellitus using machine learning techniques," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 5, pp. 836–841, 2021, [Online]. Available: https://turcomat.org/index.php/turkbilmat/article/view/1491.

[16] M. M. Hassan, M. A. M. Billah, M. M. Rahman, S. Zaman, M. M. H. Shakil, and J. H. Angon, " Early predictive analytics in healthcare for diabetes prediction using machine learning approach," Jul. 2021, doi: 10.1109/ICCCNT51525.2021.9579799.

[17] M. F. U. Bhuiyan, M. T. Rahman, M. A. Anik, and M. Khan, "A framework for Type-II diabetes prediction using machine learning approaches," in *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*, Jul. 2021, pp. 1–6, doi: 10.1109/ICCCNT51525.2021.9580158.

[18] V. Deepak, P. Srividdhya, M. Muthumari, R. Shameli, and S. N. Devi, "Identifying lethal effects and risk level using an improved classification technique," in *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, Sep. 2021, pp. 761–768, doi: 10.1109/ICIRCA51532.2021.9544708.

[19] A. N. Soni, "Diabetes mellitus prediction using ensemble machine learning techniques," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3642877.

[20] M. S. Satu, S. T. Atik, and M. A. Moni, "A novel hybrid machine learning model to predict diabetes mellitus," in *Proceedings of International Joint Conference on Computational Intelligence*, Springer Singapore, 2020, pp. 453–465.

[21] L. Malviya, S. Mal, P. Lalwani, and J. S. Chadha, "Diabetes classification using machine learning and deep learning models," in *Lecture Notes in Electrical Engineering*, vol. 796, Springer Singapore, 2021, pp. 487–503.

[22] S. K. Reddy, T. Krishnaveni, G. Nikitha, and E. Vijaykanth, "Diabetes prediction using different machine learning algorithms," in *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, Sep. 2021, pp. 1261–1265, doi: 10.1109/ICIRCA51532.2021.9544593.
[23] U. Das, A. Y. Srizon, M. A. Islam, D. S. Tonmoy, and M. A.-M. Hasan, "Prognostic biomarkers identification for diabetes prediction by utilizing machine learning classifiers," in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0, STI 2020*, Dec. 2020, pp. 1–6, doi: 10.1109/STI50764.2020.9350498.
[24] M. A. Uddin *et al.*, "Machine learning based diabetes detection model for false negative reduction," *Biomedical Materials and Devices*, vol. 2, no. 1, pp. 427–443, Jun. 2024, doi: 10.1007/s44174-023-00104-w.
[25] J. Abdollahi and S. Aref, "Early prediction of diabetes using feature selection and machine learning algorithms," *SN Computer Science*, vol. 5, no. 2, Jan. 2024, doi: 10.1007/s42979-023-02545-y.

BIOGRAPHIES OF AUTHORS

Bhanu Prakash Lohani is a research scholar in the Computer Science and Engineering Department at Galgotias University, Greater Noida. He holds both M.Tech. and B.Tech. in Computer Science and Engineering. With 16 years of experience in academia and research, his research interests include data analytics, machine learning, and healthcare information systems. He is a member of IEEE and has authored more than 26 publications. He can be contacted at email: bhanuplohani@gmail.com.

Dr. Arvind Dagur is currently a professor in the Computer Science and Engineering Department at Galgotias University, Greater Noida. His research interests include machine learning, AI applications, data mining, and data analytics. He has authored or co-authored over 75 publications, with an H-index of 15 and more than 500 citations. Dr. Dagur has organized various conferences, seminars, and workshops at the university level. He is a member of IEEE and ACM. He can be contacted at email: arvind.dagur@galgotiasuniversity.edu.in.

Dr. Dhirendra Kumar Shukla is a distinguished academic and researcher in Computer Science and Engineering with over a decade of experience in teaching and research across various prestigious institutions. Currently, he serves as an associate professor at the School of Computer Science and Engineering, Galgotias University. Dr. Shukla's research interests are deeply rooted in cloud computing, machine learning models and AI. Dr. Shukla has organized various conferences and seminars at the university level. He is a member of IEEE and CSI. He can be contacted at email: dhirendra.shukla@galgotiasuniversity.edu.in.