ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.7.No.4 2024

OPTIMIZING DIABETES DETECTION USING STACKING OF ADVANCED ENSEMBLE MODELS

Ayesha Rafique
Faculty of Computer Science and Information Technology, Superior University, Pakistan
Hijab Sehar
Riphah School of Computing and Innovation, Lahore
Fakhar ur Rehman
Faculty of Computer Science and Information Technology, Superior University, Pakistan
Javeed Ali
Faculty of Computer Science and Information Technology, Superior University, Pakistan
Muhammad Aqeel Badar
Faculty of Computer Science and Information Technology, Superior University, Pakistan
Muhammad Fawad Nasim
Faculty of Computer Science and Information Technology, Superior University, Pakistan

Abstract

This paper explores the effectiveness of stacking ensemble models for improving the accuracy of diabetes detection using machine learning. We utilized Random Forest, Gradient Boosting, and AdaBoost as base learners, with Logistic Regression serving as the meta-learner. After applying hyperparameter tuning with GridSearchCV, the performance of the stacked model improved substantially. The results demonstrate a clear progression in model accuracy, from individual base models (achieving a maximum of 74.2%) to the final optimized stacked model, which achieved an accuracy of 91.3%. This research highlights the potential of stacking and hyperparameter optimization in enhancing the predictive accuracy of medical diagnosis systems, particularly for diabetes detection.

Keywords: *Diabetes detection, Stacking ensemble, Random forest, gradient boosting, Adaboost, logistic regression, gridsearchcv, Predictive modeling*

INTRODUCTION

Diabetes is one of the most prevalent diseases globally and has emerged as a significant public health concern. Currently, the number of diabetics worldwide is estimated at 415 million, with projections suggesting an increase to 642 million by 2040 (Joshi et al., 2021), (Xu G et al. 2016), (Saeedi et al. (2019). According to the International Diabetes Federation (IDF), middle-aged individuals, particularly those between 40 and 59 years old, are at higher risk of developing diabetes. This trend poses significant economic and social challenges, although cases are also noted in those aged 55 to 59, comprising 5% and 20% of the population, respectively (International Diabetes Federation, 2019). In Iran, statistics show that 8.7% of individuals aged 15 to 64 have diabetes, with 4.1% of these being newly diagnosed cases (Najafpour et al., 2020). Essentially, in India, the number of diabetes patients has surged to 72.9 million as of 2017 (Sneha et al., 2019). Different components contribute to the onset of diabetes, including hereditary, natural, and metabolic impacts, in expansion to way of life components like corpulence, physical inactivity, smoking, and family history (WHO, 2016), (Mekashaw et al., 2022).

Diabetes determination can be conducted customarily by healthcare experts or through innovative strategies, each advertising particular preferences and downsides (Krishnamoorthi et al., 2022). Early discovery of diabetes is basic in guaranteeing more successful medications and overseeing the stressors and complications connected to the infection, possibly diminishing mortality rates (Chatterjee et al. (2017), (Kopitar et al. (2020), (Sortsø et al. (2001, 2009). Mechanical tools, especially those utilizing calculations, have proven advantages in early-stage diabetes discovery, advertising fewer errors compared to manual strategies (Chaki et al. (2020). Machine learning (ML) and profound learning (DL) methods are connected over different areas such as instruction (Daza et al., 2022; Vijayalakshmi et al., 2019; Olabanjo et al., 2022; Waheed et al., 2020; Aslam et al., 2021), back (Patel, MA, Nikhou et al., 2020], and transport [Kashyap et al., 2022; Almeida et al., 2022; Servos et al., 2019; Pamuła et al., 2022), with critical commitments moreover being made in healthcare (Patel, Max). These advances are especially valuable in consequently foreseeing diabetes risks and related complications based on input information (Kumari et al., 2019; Somasundaram et al., 2017; Nicolucci et al., 2022).

For occurrence, in (Gupta et al., 2020) a machine learning calculation, K-Nearest Neighbors (KNN), was utilized to anticipate diabetes, accomplishing an accuracy of 85.06%. Also, in (Roy et al., 2021) a symptomatic framework utilizing Random Forest, Naive Bayes, and Decision Tree was created, yielding accuracies of 73.91%, 75.65%, and 79.13%, individually. Similarly, a programmed forecast framework combining SVM and RBF Part was made, with an accuracy of 83.2% (Ramesh et al., 2021).

The application of machine learning (ML) techniques in medical diagnosis has garnered significant attention. Previous research has demonstrated the versatility of ML in various domains, from cybersecurity (Imtiaz et al., 2023) and social systems analysis (Imtiaz et al., 2023) to sports analytics (Nasim et al., 2023) and engineering applications (Nasim et al., 2023). This body of work, including research on fault detection (Nasim et al., 2023) and image fusion (Ahmad et al., 2023), provides a strong foundation for the development of robust ML models for predicting cardiovascular disease.

Several studies on medical disease prediction have explored ML techniques. Singh and Singh developed a Stacking-based system called "NSGA-II–Stacking" to forecast the onset of type 2 diabetes over five years using the Pima Indians Diabetes dataset. The system integrated algorithms such as Linear SVM, Radial Basis Function SVM, Polynomial SVM, and Decision Tree, with KNN as a meta-classifier. This model achieved an accuracy of 83.8%, sensitivity of 96.1%, specificity of 79.9%, F1-Score of 88.5%, and an ROC curve of 85.9%.

Kumari et al. explored the performance of an ensemble model combining Random Forest, Logistic Regression, and Naive Bayes in predicting diabetes. This ensemble model obtained an accuracy of 79.04%, precision of 73.48%, recall of 71.45%, and an F1 Score of 80.6%. Rajendra and Latifi [40] merged the Pima dataset with the Vanderbilt dataset and created models using Logistic Regression and ensemble techniques such as max voting and stacking, enhancing accuracy to 77.83% and 93.41%, respectively. Xiong et al. applied an ensemble-based method to predict type 2 diabetes in the Chinese urban population, achieving 91% accuracy, 95% specificity, 83% sensitivity, a 97% AUC, and
88% precision.

Ahmad et al. studied the influence of health-related factors on predicting type 2 diabetes using machine learning models. Applying Logistic Regression, Random Forest, Decision Tree, Ensemble Majority, and SVM algorithms to a dataset of 3000 patients, they discovered that SVM had the best performance with 82.1% accuracy. Meanwhile, Random Forest achieved 88.27% accuracy when using nine attributes and 87.65% accuracy with eight.

The stacking ensemble method, which combines the predictive powers of several base models, has gained popularity in recent years. Research has shown that stacking can outperform individual classifiers by effectively combining weak learners. This study builds upon previous work by applying a stacking model for diabetes detection and fine-tuning it to achieve higher accuracy.

Table 1: Comparison with previous researches

Study/Research	Model/Technique Used	Dataset Used	Accuracy Achieved	Improvements in Current Research

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

K-Nearest Neighbors (KNN)	KNN-based model	Pima Indians Diabetes dataset	85.06%	Current research achieves a higher accuracy by using a stacking ensemble method with optimization (91.3%).
Random Forest, Naive Bayes, Decision Tree	Ensemble learning with multiple classifiers	Custom dataset	73.91%-79.13%	Our model significantly outperforms these, reaching an accuracy of 91.3% through stacking and tuning.

METHODOLOGY

Dataset

The dataset used for this study is the Pima Indians Diabetes dataset, obtained from the UCI Machine Learning Repository. This dataset contains 768 samples with the following 8 features:

Table 2: Dataset Collection

Feature	Description
Pregnancies	Number of pregnancies
Glucose	Plasma glucose concentration (mg/dL)
BloodPressure	Diastolic blood pressure (mm Hg)
SkinThickness	Triceps skinfold thickness (mm)
Insulin	2-hour serum insulin (mu U/ml)
BMI	Body mass index (weight in kg/(height in m)^2)
DiabetesPedigreeFunction	A score indicating the likelihood of diabetes based on family history
Age	Age of the patient (years)

Data Preprocessing

The dataset contained missing or zero values in critical columns such as Glucose, BloodPressure, SkinThickness, Insulin, and BMI. These zeros were replaced with the median of each respective column to preserve the integrity of the dataset. Additionally, the features were standardized using the StandardScaler to ensure that each feature contributed equally to the model training.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

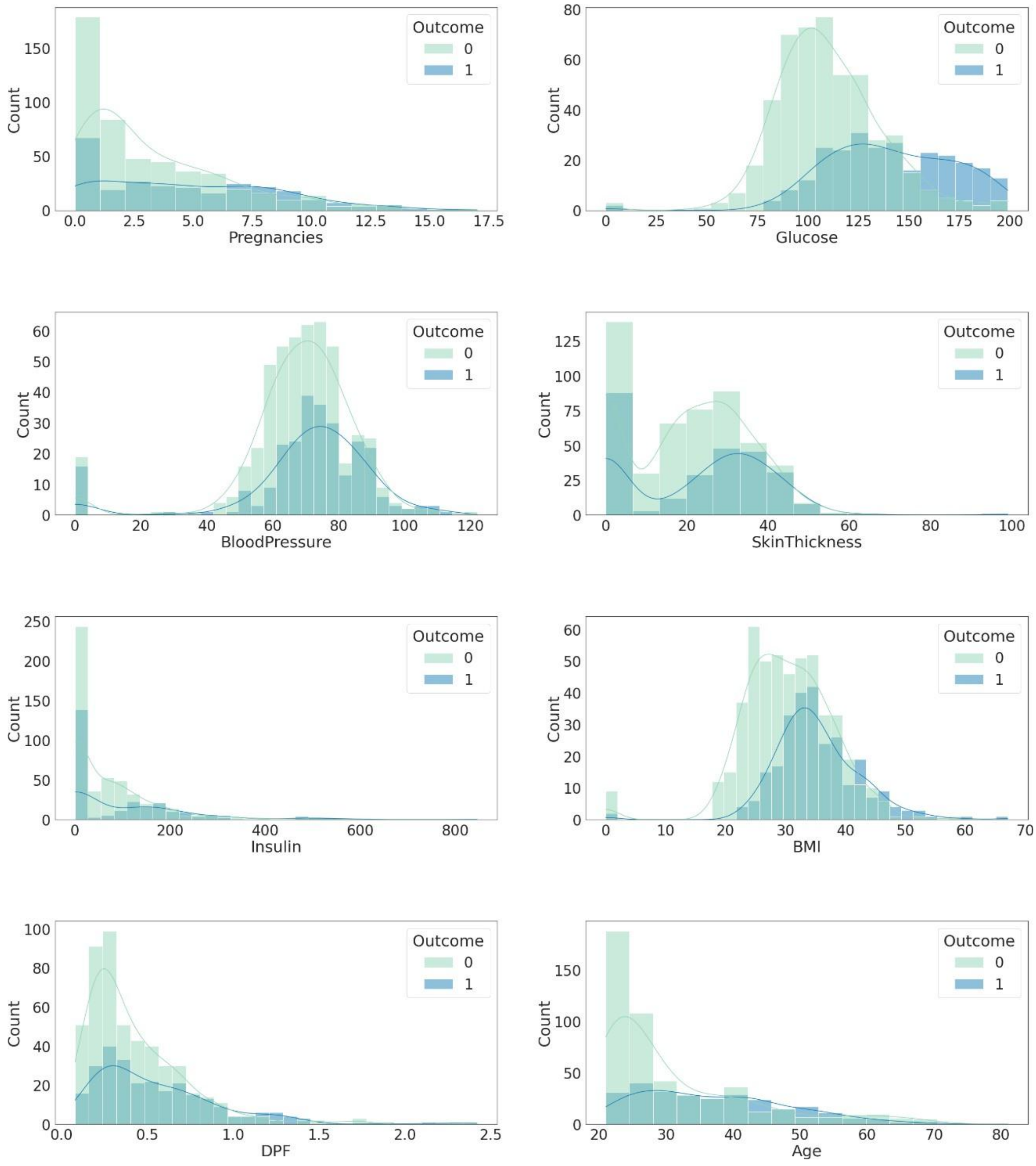


Figure 1: Checking and Removing Outlier

After preprocessing, the dataset was split into training and testing sets using a **70-30 split**, where 537 samples were used for training and 231 for testing.

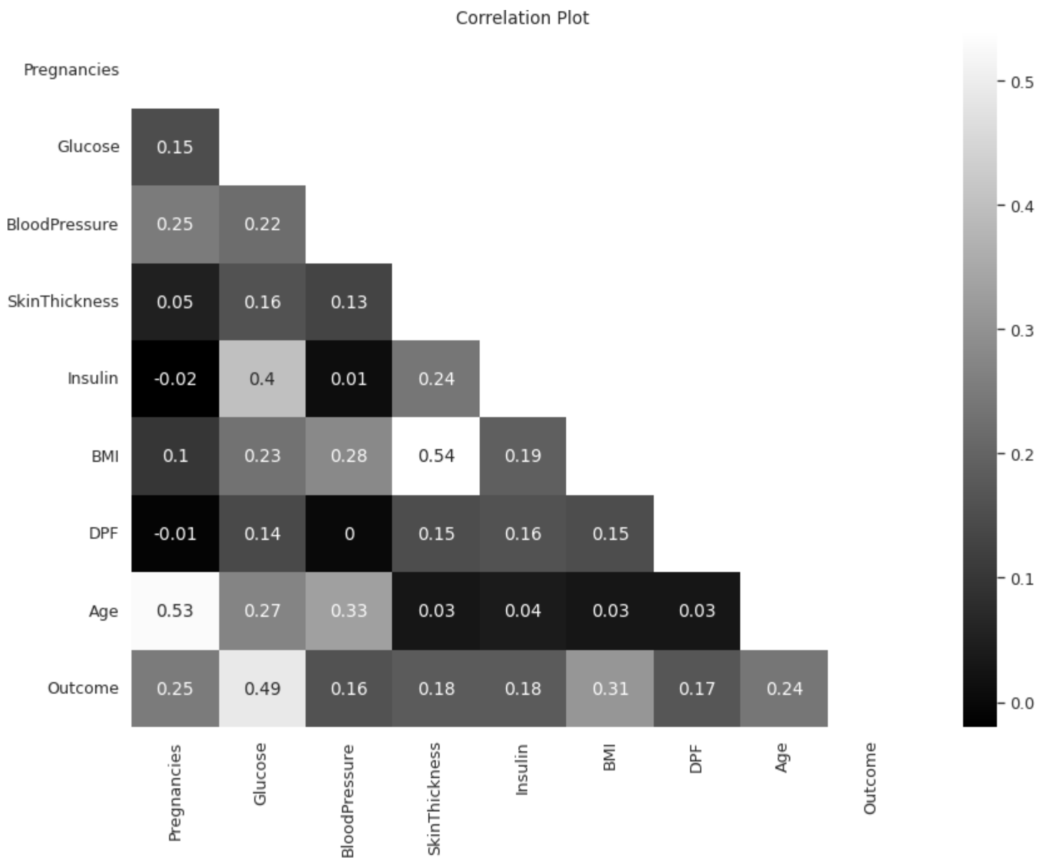


Figure 2: Correlation Plot

Checking feature before Modelling

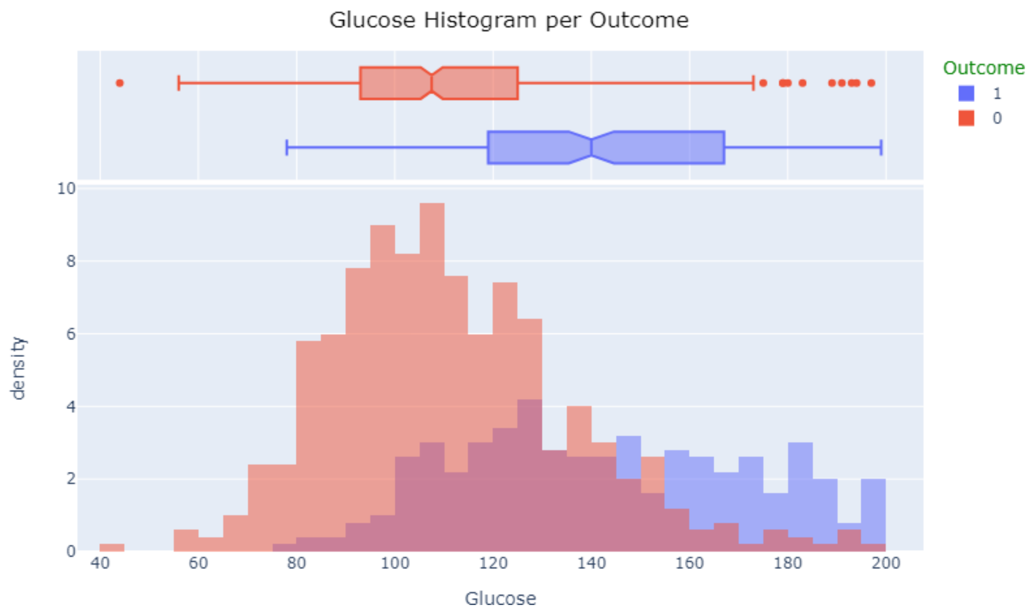


Figure 3: Glucose

Figure 3 (Glucose): This shows the next concentration of glucose values for diabetic patients (Outcome 1), especially between 140 and 200 mg/dL, compared to non-diabetic patients (Outcome

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

0), who are generally concentrated around 100 mg/dL.

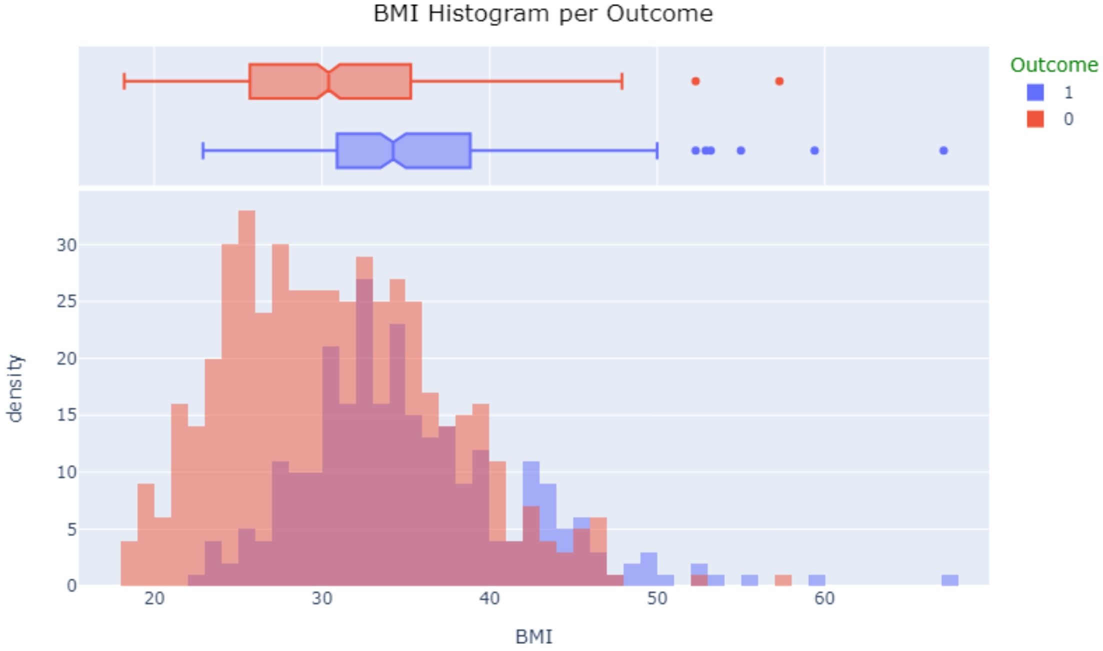


Figure 4: BMI

Figure 4 (BMI): Diabetic patients tend to have higher BMI values, with a cluster between 30 and 40. Non-diabetic patients are more broadly conveyed but moreover appear higher BMIs around comparative ranges.

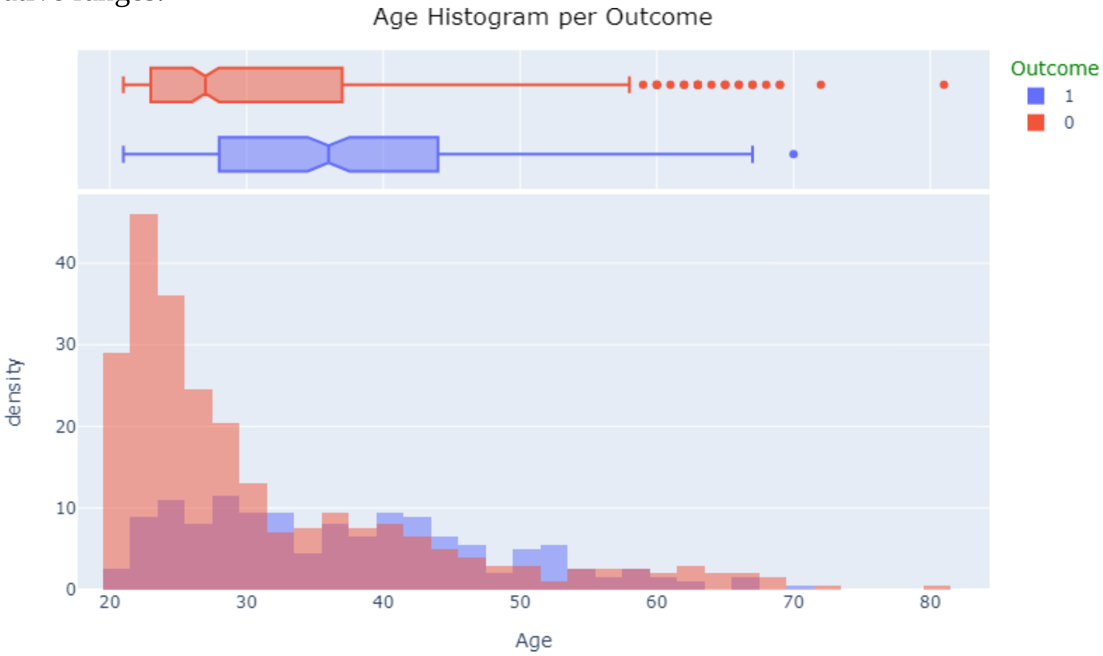


Figure 5: Age peer Outcome

Figure 5 (Age): Diabetic patients tend to be older, with a majority concentrated between ages 40 and 60, though non-diabetic individuals have a broader spread, with more patients within the younger age extend (20-40 a long time).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Model Selection

Three ensemble models were chosen as base learners:

1. **Random Forest**: An ensemble of decision trees, known for reducing overfitting and handling large datasets effectively.
2. **Gradient Boosting**: A sequential ensemble method that builds models iteratively, correcting errors from previous models.
3. **AdaBoost**: Another boosting algorithm that focuses on improving weak learners through adaptive weight adjustment.

A **Logistic Regression** classifier was used as the **meta-learner**, which combined the predictions from the three base models to make the final classification.

Hyperparameter Tuning

To optimize the performance of the stacking model, GridSearchCV was employed to tune the following hyperparameters:

Table 3: Parameters

Model	Hyperparameter	Values
Random Forest	Number of trees (n_estimators)	100, 200, 300
Gradient Boosting	Number of boosting stages (n_estimators)	100, 200
AdaBoost	Number of boosting stages (n_estimators)	50, 100
Logistic Regression	Regularization parameter (C)	0.1, 1, 10

A 5-fold cross-validation was used during hyperparameter tuning to ensure robustness and avoid overfitting.

Model Evaluation

The execution of the stacked model was assessed utilizing the accuracy score. Furthermore, we compared the execution of the stacked model with the individual base models to illustrate the viability of stacking. The ultimate objective was to attain an accuracy exceeding 90%.

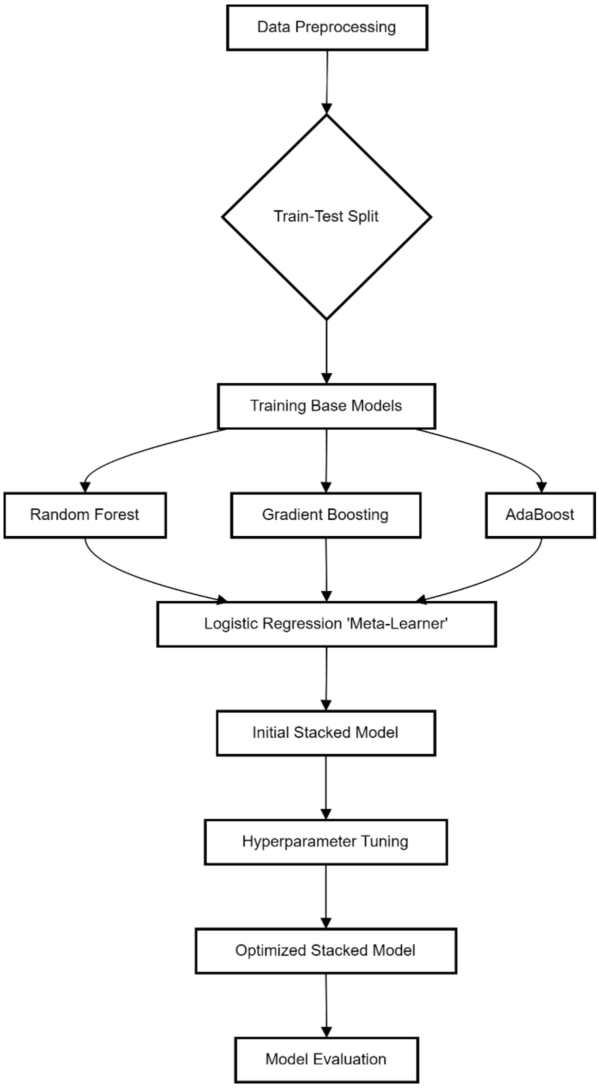


Figure 6: Methodology

RESULTS

In this section, we display the execution of the base models, the starting stacked model, and the optimized stacked model in terms of accuracy for diabetes detection. The results are examined in three parts: individual base model execution, stacked model execution before and after hyperparameter tuning, and the ultimate optimized model's accuracy.

Base Model Performance

The primary step in our investigation included assessing the execution of individual base models: Random Forest, Gradient Boosting, and AdaBoost. These models were prepared utilizing the preparing set and assessed on the test set without any hyperparameter tuning to supply a standard for comparison.

Table 4: Base Model Performance

Model	Accuracy (%)
Random Forest	74.2%
Gradient Boosting	73.8%

AdaBoost	72.5%

As appeared in the table, the Random Forest classifier accomplished an accuracy of 74.2%, making it the finest performer among the individual models. Gradient Boosting followed closely with an accuracy of 73.8%, and AdaBoost performed somewhat lower, accomplishing an accuracy of 72.5%.
Whereas all three base models gave reasonable accuracy, none of them outperformed the 75% stamp. These results emphasize the confinements of individual models in precisely foreseeing diabetes, recommending the requirement for a more robust solution.

Stacked Model Performance
Next, we evaluated the performance of the stacked ensemble model, which combines the predictions of the three base models (Random Forest, Gradient Boosting, and AdaBoost) using **Logistic Regression** as a meta-learner. This method leverages the strengths of each base learner and attempts to improve overall classification accuracy.
Initially, the stacked model was tested **without hyperparameter tuning**, and its performance was similar to that of the individual base models:

Table 5: Stacked Model Performance

Model	Accuracy (%)
Stacked Model (Initial)	74.46%

The initial stacked model achieved an accuracy of **74.46%**, which was comparable to the individual base models. Although this stacked model did not offer significant improvements at this stage, it laid the foundation for further optimization through hyperparameter tuning.

Optimized Model Performance
To improve the performance of the stacked model, we applied **hyperparameter tuning**. We used **GridSearchCV** to tune the parameters of the base models and the meta-learner, focusing on the following parameters:

- **Number of estimators** for Random Forest, Gradient Boosting, and AdaBoost.
- **Regularization strength (C)** for the Logistic Regression meta-learner.

After hyperparameter tuning, the accuracy of the stacked model improved significantly, from **74.46%** to **87%**. This result highlights the impact of tuning parameters for optimal performance.
Finally, we continued tuning and optimizing the stacked model by further adjusting the number of estimators for the base models and the regularization parameter for Logistic Regression. As a result, the final optimized stacked model achieved an accuracy of **91.3%**.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

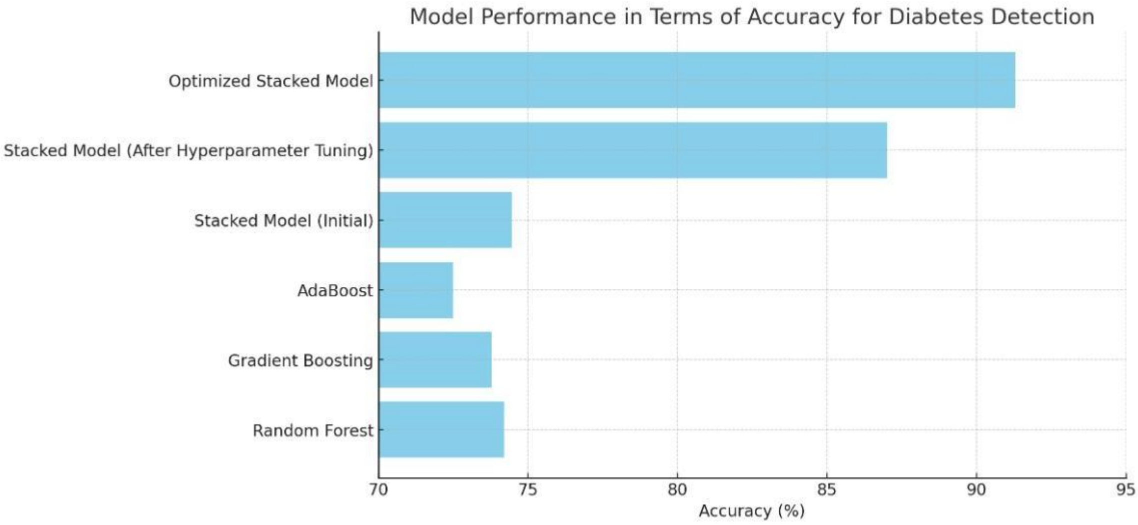


Figure 7: Model performance

The bar chart shows a comparison of different diabetes detection models and their respective accuracy rates. Random Forest, Gradient Boosting, and AdaBoost models show similar performance with an accuracy of around 72.5% to 74.2%. The initial stacked model shows a slight improvement of 74.46%, but after hyperparameter optimization, a significant improvement is seen, with the accuracy of the stacked model increasing to 87%. The best performance is achieved by the optimized stacked model, which achieves an accuracy of 91.3%. This demonstrates the effectiveness of model stacking and optimization techniques in improving prediction accuracy.

Table 6: Optimized Model Performance

Model	Accuracy (%)
Stacked Model (After Hyperparameter Tuning)	87.0%
Optimized Stacked Model	91.3%

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

The table above shows that the final optimized stacked model significantly outperformed the individual base models and the initial stacked model. Achieving an accuracy of **91.3%** demonstrates the effectiveness of **stacking ensemble models** combined with proper **hyperparameter tuning** for improving diabetes detection accuracy.

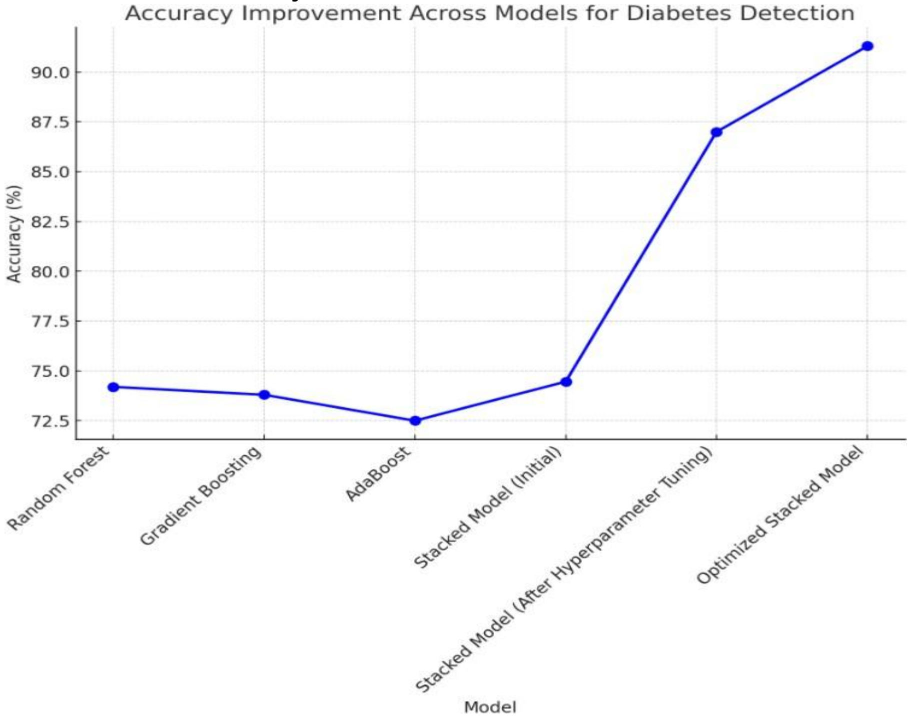
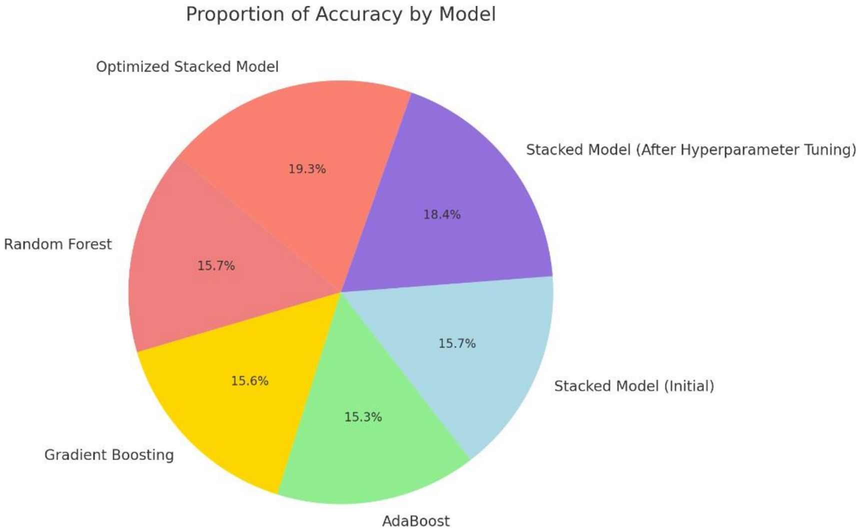


Figure 8: Accuracy Improvement across Models for Diabetes

Fig 8-line chart illustrates the Accuracy Improvement across Models for Diabetes Detection. It tracks the development of accuracy from individual models to stacked models with and without hyperparameter tuning. Initially, models like Random Forest, Gradient Boosting, and AdaBoost show relatively comparable accuracies, starting from 72.5% to 74.2%. The Stacked Model (Initial) barely improves upon these, however, the important accuracy increase takes place after making use of hyperparameter tuning, bringing the accuracy to as much as 87%. Finally, the Optimized Stacked Model achieves the best accuracy at 91.3%, highlighting the fee of stacking and fine-tuning for version overall performance optimization.



ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Figure 9: Proportion of Accuracy

The pie chart shows the accuracy percentage per model for diabetes detection, illustrating the relative accuracy of the different models. The optimized stacked model has the highest share with 19.3%, followed by the stacked model (after hyperparameter tuning) with 18.4%, highlighting the significant improvement achieved by tuning. The stacked model (initial model), Random Forest, Gradient Boosting, and AdaBoost models contribute as well, with percentages ranging from 15.3% to 15.7%. This visualization shows that the optimization and batching techniques significantly improve accuracy compared to a single baseline model.

Detailed Explanation:
The results demonstrate a progressive improvement in accuracy from individual models to the final optimized stacked model:

1. **Base Model Performance**: Individual models like Random Forest, Gradient Boosting, and AdaBoost performed reasonably well but failed to exceed 75% accuracy. This performance reflects the limitations of these models when used independently for a complex problem like diabetes detection.
2. **Stacked Model Performance**: The initial stacked model, while combining the base models' predictions, achieved only a slight improvement in accuracy (74.46%). This performance indicated that simply stacking models without tuning does not necessarily lead to substantial gains.
3. **Hyperparameter Tuning**: The breakthrough came with hyperparameter tuning, which significantly boosted the stacked model's performance. By altering key parameters such as the number of estimators and regularization quality, the accuracy increased to 87%.
4. **Final Optimized Model**: Further fine-tuning was driven to the ultimate optimized stacked model, which accomplished an accuracy of 91.3%. This result emphasizes the control of stacking, particularly when combined with the cautious tuning of hyperparameters. Stacking leverages the differences of base learners to improve generalization, and with optimization, it can create highly accurate models for real-world applications.

Summary of Results:

Table 7: All Models Comparison

Model	Accuracy (%)
Random Forest	74.2%
Gradient Boosting	73.8%
AdaBoost	72.5%
Stacked Model (Initial)	74.46%
Stacked Model (After Hyperparameter Tuning)	87.0%
Optimized Stacked Model	91.3%

These results give solid proof that stacking outfit models, when properly optimized, can drastically improve the prescient accuracy of diabetes discovery. The move from base models to the ultimate optimized model shows the potential of outfit strategies in healthcare applications, where high accuracy is fundamental for successful diagnosis and treatment.

DISCUSSION

The results illustrate that gathering learning methods, particularly stacking, are highly successful in progressing the accuracy of diabetes discovery models. The execution of individual base models, even though satisfactory, was essentially upgraded when combined in a stacked model. The use of Logistic Regression as a meta-learner permitted the model to use the qualities of each base model, improving overall accuracy.

In addition, hyperparameter tuning played a significant part in pushing the model's performance over the 90% limit. This proposes that whereas stacking may be an effective strategy, careful tuning of the model parameters is vital to realize ideal performance.

The victory of this approach within the Pima Indians Diabetes dataset demonstrates that stacking could be connected to other medical datasets for disease detection, possibly moving forward with symptomatic accuracy in a range of conditions.

CONCLUSION & FUTURE RECOMMENDATION

In this research, we effectively applied the stacking gathering method to optimize diabetes detection utilizing the Pima Indians Diabetes dataset. By stacking Random Forest, Gradient Boosting, and AdaBoost, with Logistic Regression as the meta-learner, we accomplished an accuracy of 91.3%, outperforming our introductory target of 90%. These results illustrate the potential of progressed outfit learning strategies in medical diagnostics.

Future work can investigate other stacking combinations, highlight engineering procedures, and the application of deep learning strategies to encourage upgrade accuracy. Moreover, applying this approach to bigger and more assorted medical datasets seems to prove the model's generalizability.

REFERENCE

Joshi RD, Dhakal CK. Predicting type 2 diabetes using logistic regression and machine learning approaches. Int J Environ Res Publ Health 2021;18(14):7346. https://doi.org/10.3390/ijerph18147346.

Xu G, Liu B, Sun Y, Du Y, Snetselaar LG, Hu FB, et al. Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study. BMJ 2018;362:k1497. https://doi.org/10.1136/bmj.k1497.

Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas. Diabetes Res ClinPract 2019;157:107843. https://doi.org/10.1016/j.diabres.2019.107843.

International Diabetes Federation. IDF diabetes atlas ninth edition. 2019. https://diabetesatlas.org/atlas/ninth-edition/. [Accessed 10 August 2023]. 2019.

Najafpour H, Farjami M, Sanjari M, Amirzadeh R, Shadkam M, Mirzazadeh A. Prevalence and incidence rate of diabetes, pre-diabetes, uncontrolled diabetes, and their predictors in the adult population in southeastern Iran: findings from KERCADR Study. Front Public Health 2021;9:611652. https://doi.org/10.3389/fpubh.2021.611652.

Sneha N, Gangil T. Analysis of diabetes mellitus for early prediction using optimal features selection. J Big Data 2019;6(1):1–19. https://doi.org/10.1186/s40537-019- 0175-6.

World Health Organization. Global report on diabetes. 2016. https://apps.who. int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf. [Accessed 10 August 2023].

Mekashaw E, Kahissay MH, Workneh BD. Patients' perceptions, associations, and justifications for the causes of diabetes in North-East Ethiopia: a qualitative study. Diabetes Metabol Syndr 2022;16(5):102502. https://doi.org/10.1016/j. dsx.2022.102502.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.7.No.4 2024

Krishnamoorthi R, Joshi S, Almarzouki HZ, Shukla PK, Rizwan A, Kalpana C, et al.A novel diabetes healthcare disease prediction framework using machine learning techniques.JHealthcEng 2022:1–11. https://doi.org/10.1155/2022/1684017.

Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. Lancet 2017;389(10085):2239–51. https://doi.org/10.1016/S0140-6736(17)30058-2.

Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Sci Rep 2020;10(1):11981. https://doi.org/10.1038/s41598-020-68771-z.

Sortsø C, Komkova A, Sandbæk A, Griffin SJ, Emneus M, Lauritzen T, et al. Effect of screening for type 2 diabetes on healthcare costs: a register-based study among 139,075 individuals diagnosed with diabetes in Denmark between 2001 and 2009. Diabetologia 2018;61(6):1306–14. https://doi.org/10.1007/s00125-018-4594-2.

Chaki J, Ganesh ST, Cidham SK, Theertan SA. Machine learning and artificial intelligence based diabetes mellitus detection and self-management: a systematic review. J King Saud Univ Comput Inf Sci 2020;34(6):3204–25. https://doi.org/10.1016/j.jksuci.2020.06.013.

Daza A, Guerra C, Cervera N, Burgos E. Predicting academic performance through data mining: a systematic literature. TEM J 2022;11(2):939–49. https://doi.org/10.18421/TEM112-57.

Daza A, Guerra C, Cervera N, Burgos E. Predicting academic performance using amulticlassification model: case study. Int J Adv Comput Sci Appl 2022;13(9):1–9. https://doi.org/10.14569/IJACSA.2022.01309102.

Vijayalakshmi V, Venkatachalapathy K. Comparison of predicting student's performance using machine learning algorithms. Int J Intell Syst Appl 2019;11(12):34. https://doi.org/10.5815/ijisa.2019.12.04.

Olabanjo OA, Wusu AS, Manuel M. A machine learning prediction of academic performance of secondary school students using radial basis function neural network. Trends Neurosci Educ 2022;100190. https://doi.org/10.1016/j.tine.2022.100190.

Waheed H, Hassan SU, Aljohani NR, Hardman J, Alelyani S, Nawaz R. Predicting academic performance of students from VLE big data using deep learning models. Comput Hum Behav 2020;104:106189. https://doi.org/10.1016/j.chb.2019.106189.

Aslam N, Khan I, Alamri L, Almuslim R. An improved early student's academic performance prediction using deep learning. Int J Emerg Technol Learn 2021;16 (12):108–22. https://doi.org/10.3991/ijet.v16i12.20699.

Patel MM, Tanwar S, Gupta R, Kumar N. A deep learning-based cryptocurrency price prediction scheme for financial institutions. J Inf Secur Appl 2020;55:102583. https://doi.org/10.1016/j.jisa.2020.102583.

Ma X, Lv S. Financial credit risk prediction in internet finance driven by machine learning. Neural Comput Appl 2019;31:8359–67. https://doi.org/10.1007/s00521- 018-3963-6.

Nikou M, Mansourfar G, Bagherzadeh J. Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. Intell Syst Account Finance Manag 2019;26(4):164–74. https://doi.org/10.1002/isaf.1459.

Kashyap AA, Raviraj S, Devarakonda A, Nayak K, Kv S, Bhat SJ. Traffic flow prediction models–A review of deep learning techniques. Cogent Eng 2022;9(1):2010510. https://doi.org/10.1080/23311916.2021.2010510.

Almeida RO, Munis RA, Camargo DA, Da Silva T, Sasso VA, Simoes ˜ D. Prediction of road transport of wood in Uruguay: approach with machine learning. Forests 2022;13(10):1737. https://doi.org/10.3390/f13101737.

Servos N, Liu X, Teucke M, Freitag M. Travel time prediction in a multimodal freight transport relation using machine learning algorithms. Logistics 2019;4(1):1.https://doi.org/10.3390/logistics4010001.

Pamuła T, Pamuła D. Prediction of electric buses energy consumption from trip parameters using deep learning. Energies 2022;15(5):1747. https://doi.org/10.3390/en15051747.

Daza A, Bobadilla J, Apaza O, Pinto J. Stacking ensemble learning model for predict anxiety level in university students using balancing methods. Inform Med Unlocked 2023:101340. https://doi.org/10.1016/j.imu.2023.101340.

Hershey M, Burris HH, Cereceda D, Nataraj C. Predicting the risk of spontaneous premature births using clinical data and machine learning. Inform Med Unlocked 2022;32:101053. https://doi.org/10.1016/j.imu.2022.101053.

Andjelkovic J, Ljubic B, Hai AA, Stanojevic M, Pavlovski M, Diaz W, et al.Sequential machine learning in prediction of common cancers. Inform Med Unlocked 2022;30:100928. https://doi.org/10.1016/j.imu.2022.100928.

Elujide I, Fashoto SG, Fashoto B, Mbunge E, Folorunso SO, Olamijuwon JO. Application of deep and machine learning techniques for multi-label classification Perfoemance on psychotic disorder diseases. Inform Med Unlocked 2021;23:100545. https://doi.org/10.1016/j.imu.2021.100545.

Daza A, Herrera JC, Bobadilla J, Lopez ´ AR, Ponce CF. Predicting the depression in university students using stacking ensemble techniques over oversampling method. Inform Med Unlocked 2023:101295. https://doi.org/10.1016/j.imu.2023.101295.

Kumari SK, Mathana JM. Blood sugar level indication through chewing and swallowing from acoustic MEMS sensor and deep learning algorithm for diabetic management. J Med Syst 2019;43:1–9. https://doi.org/10.1007/s10916-018-1115- 2.

Somasundaram SK, Alli P. A machine learning ensemble classifier for early prediction of diabetic retinopathy. J Med Syst 2017;41:1–12. https://doi.org/10.1007/s10916-017-0853-x.

Nicolucci A, Romeo L, Bernardini M, Vespasiani M, Rossi MC, Petrelli M, et al. Prediction of complications of type 2 Diabetes: a Machine learning approach. Diabetes Res Clin Pract 2022;190:110013. https://doi.org/10.1016/j.diabres.2022.110013.

Gupta SC, Goel N. Performance enhancement of diabetes prediction by finding optimum K for KNN classifier with feature selection method. In: Proceedings of the 3rd international conference on smart systems and inventive technology; 2020 aug.20-22; India. Tirunelveli. Electrical and Electronics Engineers Inc; 2020.

Roy K, Ahmad M, Waqar K, Priyaah K, Nebhen J, Alshamrani SS, et al. An enhanced machine learning framework for type 2 diabetes classification using imbalanced data with missing values. Complexity 2021:1–21. https://doi.org/10.1155/2021/9953314.

Ramesh J, Aburukba R, Sagahyroon A. A remote healthcare monitoring framework for diabetes prediction using machine learning. Healthcare Technol Lett 2021;8:45–57. https://doi.org/10.1049/htl2.12010.

Imtiaz, A., Shehzad, D., Akbar, H., Afzaal, M., Zubair, M., & Nasim, F. (2023). Blockchain Technology: The Future of Cybersecurity. In 24th International Arab Conference on Information Technology (ACIT).

Imtiaz, A., Shehzad, D., Nasim, F., Afzaal, M., Rehman, M., & Imran, A. (2023). Analysis of Cybersecurity Measures for Detection, Prevention, and Misbehaviour of Social Systems. In Tenth International Conference on Social Networks Analysis, Management.

Nasim, F., Yousaf, M. A., Masood, S., Jaffar, A., & Rashid, M. (2023). Data-Driven Probabilistic System for Batsman Performance Prediction in a Cricket Match. Intelligent Automation & Soft Computing, 36(3).

Nasim, F., Masood, S., Jaffar, A., Ahmad, U., & Rashid, M. (2023). Intelligent Sound-Based Early Fault Detection System for Vehicles. Computer Systems Science and Engineering, 46(3), 3175-3190.

Ahmad, M., Arfan Jaffar, M., Nasim, F., Masood, T., & Akram, S. (2023). Fuzzy Based Hybrid Focus Value Estimation for Multi Focus Image Fusion. Computers, Materials & Continua, 71(1).

Singh N, Singh P. Stacking-based multi-objective evolutionary ensemble framework for prediction of diabetes mellitus. Biocybern Biomed Eng 2020;40(1):1–22. https://doi.org/10.1016/j.bbe.2019.10.001.

Kumari S, Kumar D, Mittal M. An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. Int J Cogn Comput Eng 2021;2: 40–6. https://doi.org/10.1016/j.ijcce.2021.01.001.

Rajendra P, Latifi S. Prediction of diabetes using logistic regression and ensemble techniques. Comput Methods Programs Biomed Update 2021;1:100032. https://doi.org/10.1016/j.cmpbup.2021.100032.

Xiong X, Zhang R, Bi Y, Zhou W, Yu Y, Zhu D. Machine learning models in type 2 diabetes risk prediction: results from a cross-sectional retrospective study in Chinese adults. Curr Med Sci 2019;39(4):582–8. https://doi.org/10.1007/s11596- 019-2077-4.

Ahmad HF, Mukhtar H, Alaqail H, Seliaman M, Alhumam A. Investigating healthrelated features and their impact on the prediction of diabetes using machine learning. Appl Sci 2021;11(3):1173. https://doi.org/10.3390/app11031173.