InfoTech Spectrum: Iraqi Journal of Data Science (IJDS)
ISSN: 3007-5467, Vol 3 (No. 2), 13-24 (2026)
DOI: https://doi.org/10.51173/ijds.v3i2.71
https://ijds.mtu.edu.iq

A HYBRID FIREFLY-STACKING ENSEMBLE MODEL FOR EARLY PREDICTION OF TYPE II DIABETES

Michael Favour Edafeajiroke[1]*, Veronica Ijebusomma Osubor [2]

[1]*Department of Computer Science, Faculty of Computing, University of Port Harcourt, Rivers State, Nigeria*
[2]*Department of Computer Science, Faculty of Computing, University of Benin, Edo State, Nigeria*

** Corresponding author E-mail: michael.edafeajiroke@uniport.edu.ng (**Michael Favour Edafeajiroke**)*

RESEARCH ARTICLE

ARTICLE INFORMATION

SUBMISSION HISTORY:
Received: 15 January 2026
Revised: 10 March 2026
Accepted: 15 March 2026
Published: 30 June 2026

KEYWORDS:
Type II Diabetes Prediction;
Firefly Algorithm;
Stack-ensemble Learning;
Feature Selection;
Machine Learning.

ABSTRACT

The increasing global prevalence of Type II Diabetes (T2D) demands an advanced predictive model, as conventional machine learning algorithms perform poorly on such complex biomedical datasets. The proposed study aims to develop a trustworthy early prediction model by integrating the Firefly Algorithm (FA) for feature selection with a comparative stacking ensemble technique. The proposed methodology uses the large-scale Behavioral Risk Factor Surveillance System (BRFSS) dataset. The FA algorithm identified the optimal combination of 16 important features to train a stack ensemble of base models using a Meta Random Forest classifier. The most important results reveal that the FA-improved Meta Random Forest classifier achieved the best possible balance between accuracy and efficiency, with 88.68% accuracy, 83.26% sensitivity, and an AUC of 94.37%. It is concluded that the hybrid strategy is an efficient and effective approach to early risk stratification in T2D and addresses an important gap in predictive medicine by combining feature selection optimization and ensemble methods. The hybrid strategy provides a platform for future validation studies across different clinical datasets, enabling proactive, data-driven interventions.

1. INTRODUCTION

The incidence of Type II Diabetes (T2D) is a significant and increasing concern for public health, demanding a paradigm shift from the existing emphasis on the prediction of end-stage disease to early and accurate prediction, in accordance with Sustainable Development Goals (SDGs) 3[1, 2]. However, a significant research gap remains in developing efficient computational tools for this purpose. Although traditional machine learning (ML) models are widely used, they are not very efficient when faced with the high dimensionality and imbalance common in real-world medical data [3,4]. Although more advanced ensemble learning approaches, such as stacking, can address robustness issues more effectively by aggregating the predictions of multiple learners, their efficiency is necessarily constrained by two inherent challenges: selecting discriminative features and combining base and meta-learners [5, 6]. The motivation for this research, therefore, is the need to address these two challenges and develop a practical, effective solution for proactive risk stratification. In the context of predictive analytics in healthcare, this research combines bio-inspired optimization with ensemble learning to address these two challenges [7]. The two main objectives of this research, therefore, are:

1. Create a novel framework that utilizes the Firefly Algorithm (FA) to achieve intelligent feature selection through a wrapper method to identify the important risk factors.
2. Create and evaluate a comparative stack ensemble model that enhances the integration of base models like Logistic Regression, K-Nearest Neighbors, and AdaBoost with a meta-model, to attain maximum accuracy.

The proposed model is rigorously validated on the large-scale Behavioral Risk Factor Surveillance System (BRFSS) dataset to ensure clinical relevance and robustness [8, 9]. Consequently, this study is founded on the assumption that the synergistic effect of an FA-driven feature selection approach and a well-crafted stack ensemble classifier will lead to a significant improvement in the predictive accuracy of T2D risk stratification [10]. This study asserts that the proposed approach will offer a

superior, vital computational model that bridges the critical gaps in feature selection and ensemble design, thus offering a valuable addition to the toolkit of data-driven preventive healthcare and personalized medicine. Introduce the subject of research and its background. Emphasize the significance and relevance of the problem being solved. Define the problem that your research will solve. This should be specific and relevant to the subject of research. State the objectives of your research. What do you hope to achieve or find out? If applicable, list the research questions or hypotheses. Explain why this problem is important and needs to be solved. Give a brief preview of the structure of your paper. Describe what each section will hold without going into too much detail.

The structure of this document is as follows. Related work is discussed in Section 2. Section 3 gives an overview of the technique, and Section 4 describes the proposed system and experimental results. Section 5 gives a comparison with the previous work. This paper is finally concluded in Section 6.

2. LITERATURE REVIEW

Recent studies have shown the efficacy of stacking ensemble techniques over traditional machine learning models for diabetes prediction. But there are some critical gaps in the applicability of the dataset, feature engineering, and optimization that the proposed study will try to fill.

Ikram et al. [11] utilized stack ensemble method on the PIMA database, yielding accuracy, precision, and recall beyound 90%. Their findings validate the effectiveness of stacking in early detection, but they emphasize an essential shortcoming: the requirement for databases that encompass essential features to meet the particular local requirements of Type II diabetic patients. Likewise, Reza et al. [12] improved the detection process by stacking the PIMA database with simulated and local databases in two stacking models (traditional and deep neural network), reaching a maximum accuracy of 95.5%. Although their results were excellent (precision: 88-96%, recall: 88-97%), they again pointed out the need for varied databases and better feature extraction and parameter tuning methods. Daza et al. [13] also achieved a high accuracy of 97% with a stacking approach incorporating Logistic Regression and Random Forest on the PIDD dataset, but highlighted the limitations of the dataset and the need for more robust approaches with varied data and the exploration of other algorithms. Simultaneously, research has been directed at the optimization of ensemble components and enhanced interpretability. Li et al. [14] used data balancing (SMOTEENN) and constructed a stacking model incorporating GA-XGBoost, LightGBM, and Random Forest on the BRFSS dataset. Their optimized ensemble performed better than individual models, and Shapley value evaluation offered clinical value by pinpointing age and BMI as important risk factors. However, they pointed out the need for sophisticated feature selection and optimization methods. Oliullah et al. [15] constructed a female-centric model with a stacked ensemble approach (Random Forest, XGBoost, LightGBM) incorporating SHAP interpretability, reaching 92.91% accuracy. However, their research did not delve into the impact of individual features on the ensemble model's predictions, suggesting a deficiency in comprehensive interpretability for complex stacking.

Additional research highlights the ongoing pursuit of efficacy. After evaluating models on the BRFSS, Aamir and Murtza [16] found that their LSTM/MLP voting ensemble achieved better accuracy than standalone models (85.03%). To increase the efficacy of predictions, they specifically suggested an optimized stacking strategy. With a stack consisting of KNN, SVM, RF, and NB base classifiers as well as an LR meta-classifier on PIDD, Rahim et al. [17] achieved 94.17% accuracy, significantly surpassing the accuracy of base classifiers (77-87%). Their research demonstrated the potential for additional advancement through advanced techniques. An optimized stacking model (LR, NB, AdaBoost+SVM) with SMOTE was created by Aman and Chhillar [18], yielding exceptional accuracy (99.72% on ESDRP, 94.21% on PIDD). They did, however, highlight the need for improved methodology and more Type II diabetes-specific features, as well as the necessity of feature optimization to reduce complexity. Lastly, broader applications and frameworks demonstrate the persistent challenges regarding precision and clinical relevance. Daliya and Ramesh [19] developed an optimized stacking model for regression (MAE: -52.83) and classification (87% AUC) utilizing a substantial BRFSS dataset to investigate disease progression. Their work stressed the importance of pre-processing and choosing a base model, but it also showed that more work was needed to

make it more accurate for medical purposes. Abnoosian et al. [20] constructed a high-performance multi-class model (98.87% accuracy, 0.999 AUC) utilizing weighted ensembles and Bayesian optimization. Even though their results were great, they said that the model was too expensive to run and suggested ways to improve feature selection so that it could be used in real-world clinical settings. While Edafeajiroke and Osubor [22] enhanced diabetes prediction by integrating the Firefly Algorithm with a Random Forest classifier, enabling the model to select the most relevant patient features without manual tuning automatically. Their hybrid approach achieved exceptional results, including 99.91% accuracy and near-perfect sensitivity, significantly outperforming the standard Random Forest model. The study concludes that this optimized system offers a robust and reliable tool for early diabetes screening, with strong potential for real-world clinical application to improve patient outcomes.

It is obvious that the stacking ensemble model achieves high accuracy for diabetes prediction, but it also faces limitations due to the limited dataset and the lack of complex optimization techniques Li et al. [14] and Aman & Chhillar [18]. This particular problem has been addressed by the current research, which has developed a novel hybrid model that incorporates the firefly algorithm for dual purposes: wrapper-based feature selection and model optimization. The study reduced computational complexity as seen in Abnoosian et al. [20].

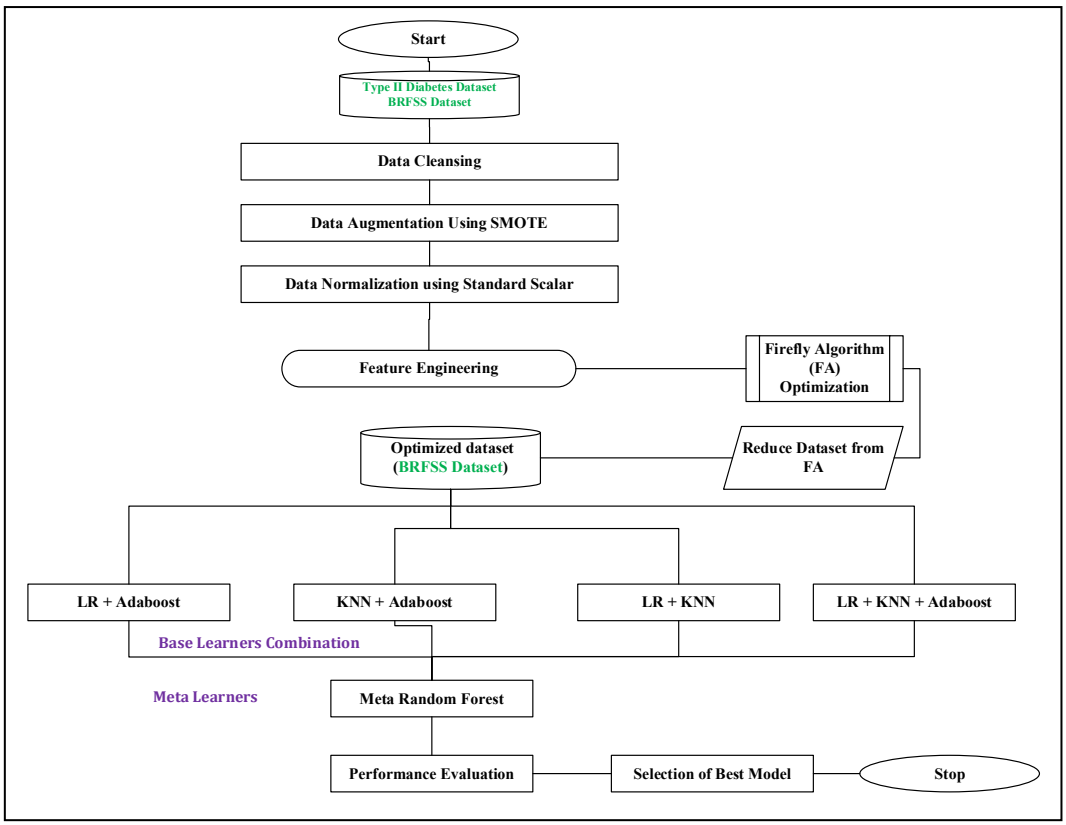

Figure 1. A workflow for the developed FA with a Stack-Ensemble Based Model for Type II Diabetes Prediction

METHODS

This study describes a systematic approach to develop an FA stack ensemble ML model for the early prediction of T2D. To develop a robust and generalized model that can meet the localized healthcare needs, this study innovatively leverages the flashing characteristics of the firefly and the inherent predictive power of stack ensemble classifiers. The study uses a hierarchical and layered approach to systematically implement a four-step process that includes data preprocessing and feature construction using FA, and a comparative stack ensemble combination. The framework concludes with a rigorous evaluation process based on standard performance measures, intended to address past problems such as overfitting and bias while improving the model's predictive capabilities by incorporating data. The workflow diagram of the developmental model is shown in Fig. 1.

Table 1. Feature Description of the Type II Diabetes dataset

Feature	Category Labels	Numeric Indicator
HighBP	No (Never told they have high blood pressure)	0
	Yes (Told they have high blood pressure)	1
HighChol	No (Never told they have high cholesterol)	0
	Yes (Told they have high cholesterol)	1
CholCheck	No (Not checked in the last 5 years)	0
	Yes (Checked in the last 5 years)	1
BMI	Continuous numerical value (height & weight-based)	*Numeric Value*
Smoker	No (Has not smoked at least 100 cigarettes)	0
	Yes (Smoked at least 100 cigarettes in lifetime)	1
Stroke	No (Never had a stroke)	0
	Yes (Has had a stroke)	1
HeartDiseaseorAttack	No (Never had heart disease or attack)	0
	Yes (Has had coronary heart disease or myocardial infarction)	1
PhysActivity	No (Did not engage in physical activities in the past month)	0
	Yes (Engaged in physical activities)	1
Fruits	No (Consumes fruit less than once per day)	0
	Yes (Consumes fruit at least once per day)	1
Veggies	No (Consumes vegetables less than once per day)	0
	Yes (Consumes vegetables at least once per day)	1
HvyAlcoholConsump	No (Men ≤14 drinks/week, Women ≤7 drinks/week)	0
	Yes (Men >14 drinks/week, Women >7 drinks/week)	1
AnyHealthcare	No (No healthcare coverage)	0
	Yes (Has healthcare coverage)	1
NoDocbcCost	No (Did not delay doctor visit due to cost)	0
	Yes (Delayed doctor visit due to cost)	1
GenHlth	Excellent	1
	Very Good	2
	Good	3
	Fair	4
	Poor	5
MentHlth	Number of days in the past 30 days when mental health was poor	*0 - 30*
PhysHlth	The number of days in the past 30 days physical health was poor	*0 - 30*
DiffWalk	No difficulty walking/climbing stairs	0
	Has serious difficulty walking/climbing stairs	1
Sex	1 = Male, 0=Female	0
Education	Never attended school	0
	Nursery	1
	Primary	2
	Secondary	3
	Undergraduate (B.SC, HND)	4
	Post Graduate (MSc Level)	5
	Post Graduate (PHD Level)	6
Income	Less than N30,000	1
	N30,000 - N50,000	2
	N50,000 - N100,000	3
	N100,000 - N150,000	4
	N150,000 - N200,000	5
	N200,000 - N300,000	6
	N300,000 - N500,000	7
	N500,000 or more	8
Diabetes	is the patient diabetic or not (0 or 1)	1

3.1 Data Collection

The T2D dataset was used for this research. The first dataset was obtained from the Behavioral Risk Factor Surveillance System (BRFSS) through Kaggle, specifically from the

diabetes_binary_health_indicators_BRFSS2015.csv [A2.1] file [16]. The dataset contains 253,680 samples with 21 feature variables and a binary target variable indicating the presence or absence of diabetes, though it is imbalanced. As shown in Table 1, this dataset was chosen because the current literature verifies that it contains the essential variables required for predicting Type II Diabetes, unlike commonly used but inadequate datasets such as the PIDD, which have been criticised for lacking the characteristics and instances necessary to account for all possible scenarios. The dataset description, source, year, number of features, and records, as acquired from the repositories, are presented in Table. 2.

Table 2. Acquired Type II Diabetes Datasets sources

Name	Source	Year	No. of Features	No. of Records
BRFSS	https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset	2015	22	253,680

3.2 Data Pre-Processing

We did a lot of work on the BRFSS 2015 data to make sure it was good for modeling. This included cleaning, normalizing, and feature engineering. There were no missing values, and categorical data were changed into numerical data. We used the raw dataset to find outliers and scale the features so that the numerical features were all on the same level. The dataset was checked for missing records, and the findings showed that no record was found to have a missing response. There were 24,206 duplicates among the original 253,680 observations, so a test for duplicate observations showed that there were 229,474 unique records. A data balance check was performed on the acquired dataset. The class No Diabetes (0.0) had 193,743 instances (86.07%), and the class Diabetes with class (1.0) had 34,926 instances (13.93%). The result indicated that this balance would affect machine learning algorithms to predict "no diabetes." This imbalance is shown in Fig. 2. The dataset imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE), which adds synthetic samples to the minority class (diabetic cases) to balance the dataset. This ensured that the machine learning model did not become biased towards predicting only non-diabetic cases. The dataset distribution after balancing is shown in Fig. 3.

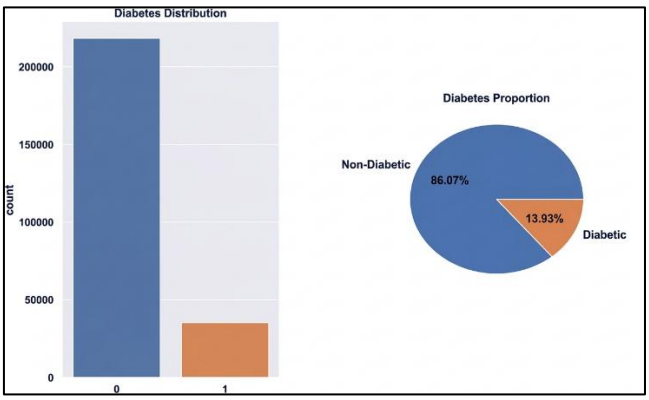

Figure 2. Dataset Distribution Before Data Balancing

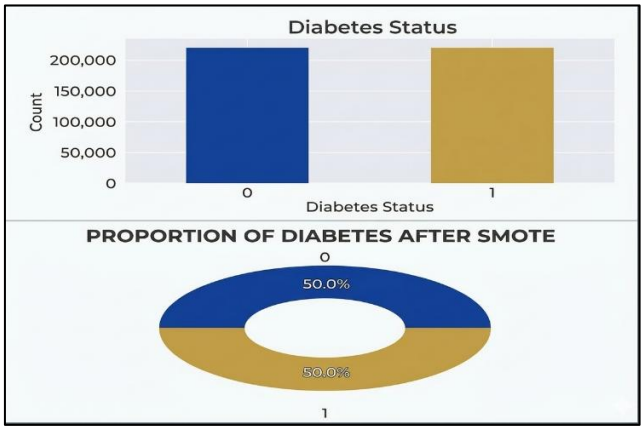

Figure 3. Dataset Distribution Before Data Balancing

Before normalization, the numerical feature varies greatly in terms of BMI, physical health days, and mental health days, with an average BMI of 30. The data is widely dispersed, as indicated by the large standard deviations, and thus, there is a need for normalization to enable fair comparison in modeling. This is evident from Fig. 4.

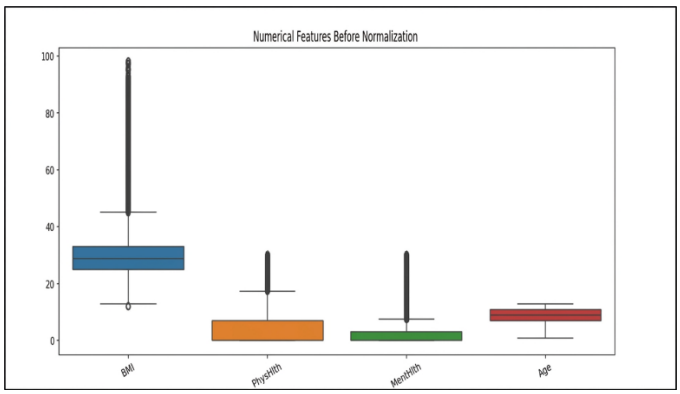

Figure 4. Numerical features before Normalization

3.3 Feature selection

The Firefly Algorithm (FA) was utilized for feature selection because of its effectiveness in wrapper-based feature selection and optimization of hyperparameters, as this addresses the important research gaps in feature engineering and model efficiency identified in previous research [14], [18]. Although deep learning models are effective in feature learning from complex patterns, they require a much larger dataset and computational power. Therefore, this optimization approach is more feasible in terms of feature interpretation in clinical risk stratification. The FA was first initialized by assigning typical default values to its critical parameters, such as population size and attractiveness, depending on the size of the 21-feature BRFSS dataset. These parameters were then optimized using a higher-level optimization technique, such as Grid Search, which performed several trials of the FA to test different parameter settings. The best setting was determined based on which set allowed the FA to select feature subsets that resulted in the highest F1-score when tested with a light-weight classifier.

3.4 Stack-Ensemble Learning Phase

Four different stack-ensemble models were developed, combining various base learners such as Logistic Regression, KNN, and AdaBoost, with a Meta Random Forest serving as the combination of the simpler stack-ensemble methods. The feature derived from the firefly algorithm was used to create the models. The models were trained to determine the best-performing setting for a comparative stack predictive model for Type II diabetes.

Table 3. Performance Evaluation Metrics

Measure	Formula	
Precision	$\frac{TP}{TP + FP}$	... (1)
Sensitivity	$\frac{TP}{TP + FN}$	... (2)
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	... (3)
Specificity	$\frac{TN}{TN + FP}$	... (4)

3.5 Performance Evaluation Parameters

The performance of the model was measured using a variety of metrics, which were calculated from the confusion matrix [21]. These metrics include F1-score and AUC-ROC. In the context of a medical screening task, such as diabetes detection, high sensitivity is crucial to ensure that no diabetic individuals are missed. From Table. 3, the metrics used in this study, where a balanced assessment of both positive and negative instances is needed, accuracy is the most appropriate metric. This is because literature has it that in medical screening, accuracy is the best metric to use.

After all, it offers a fair assessment of the overall correctness of the model, ensuring that the model has minimized both false negatives and false positives. Precision, sensitivity, specificity, and accuracy are metrics that assess the correctness of positive predictions, the identification of actual positives, the identification of actual negatives, and the overall number of correct predictions, respectively.

4. RESULTS

All the implementations were done using Python in a Jupyter Notebook environment hosted on the Google Colab platform. The simulation was run on a Dell Latitude E7470 notebook computer with an Intel Core i7-6600U processor and 16GB of RAM.

The pre-processed diabetes data was compared using the Firefly Algorithm (FA) for feature selection. To properly assess the model, the BRFSS data were split into a 75:20:5 ratio. The data normalization process was finished, as evidenced by the near-zero values and unit standard deviations for all features. These values, as shown in Fig. 5, show that the data is now centered and scaled for machine learning. The small values are considered insignificant artifacts of the calculation.

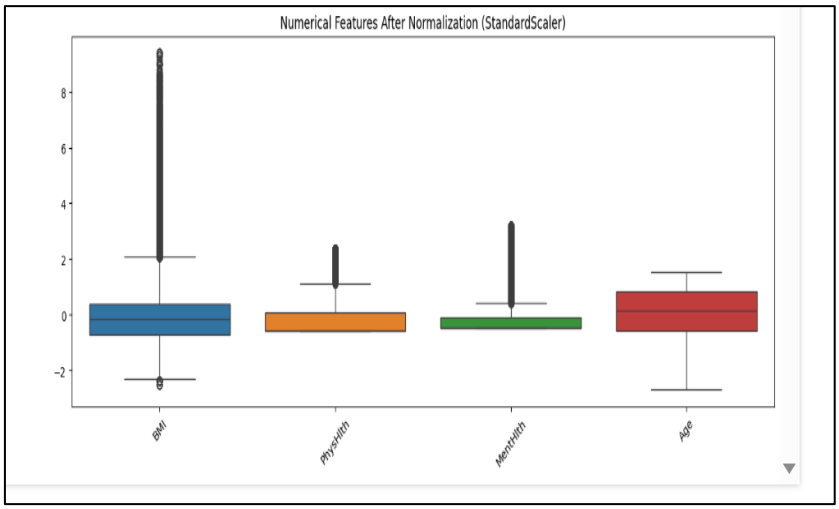

Figure 5. Numerical Features After Normalization

The FA was used in a wrapper method to identify the best feature subset from the pre-processed data. When the FA method was used, 16 of 22 features were selected to form the best feature subsets, as indicated in Table 4. When Optimization Time had a value of 553.72 seconds, Best Fitness Score was 0.8957.

Table 4. Features Selected with the Firefly Algorithm

S/N	Selected features from the Firefly Algorithm
1	CholCheck
2	Smoker
3	Stroke
4	HeartDiseaseorAttack
5	PhysActivity
6	Fruits
7	Veggies
8	HvyAlcoholConsump
9	AnyHealthcare
10	NoDocbcCost
11	GenHlth
12	MentHlth
13	Sex
14	Age
15	Education
16	Income

The rate of convergence of the FA Techniques on the BRFSS dataset after Type II diabetes features optimization, as presented in Fig. 6.

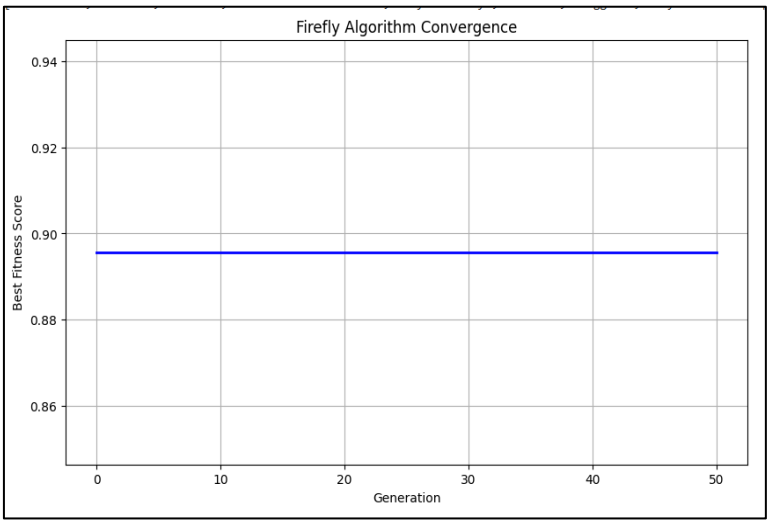

Figure 6. Rate of Firefly Algorithm Convergence on the BRFSS Dataset

From Table 5, the performance metrics obtained indicate that at the feature selection step, the model performs well in terms of its predictive capability (ROC AUC = 0.9439). It also performs well in terms of precision and recall (F1 = 0.8797). It has an efficient prediction time (4.16s) compared to the training time (44.36s).

Table 5: Performance of the FA Technique with the Selected Features

Metric	Results Obtained
Training Time	44.3614 seconds
Testing Time	4.1571 seconds
Accuracy	0.8860 (88.60%)
F1 Score	0.8797 (87.97%)
ROC AUC	0.9439 (94.39%)

With the best feature subset identified by the Firefly Algorithm, the stack-ensemble model was then trained. The performance, as can be seen in Table. 6, suggests that the Random Forest, which is a meta-level combination of the simpler stack models, performed better in terms of accuracy, precision, and speed. However, the standalone stack-ensemble Model 4 (LR+KNN+AdaBoost) took significantly longer to train. However, it did achieve a slightly higher ROC AUC of 94.54%, while Logistic Regression remained the fastest at 0.91s training time but with the poorest performance, making Meta Random Forest the most efficient.

Table 6. Results obtained for the FA + Stack-Ensemble Model on the BRFSS

Model	Accuracy (%)	Sensitivity (Recall) (%)	Specificity (%)	Precision (%)	F1 Score (%)	ROC AUC (%)	Training Time (s)	Testing Time (s)
Model 1 (LR+AdaBoost)	85.27	80.62	89.93	88.90	84.56	92.59	367.61	10.72
Model 2 (KNN+AdaBoost)	85.02	83.21	86.83	86.34	84.74	93.08	1350.85	607.90
Model 3 (LR+KNN)	78.30	81.47	75.13	76.61	78.97	87.26	1179.18	610.13
Model 4 (LR+KNN+AdaBoost)	88.14	83.18	93.10	92.34	87.52	94.54	1371.07	640.75
Logistic Regression	70.21	71.80	68.61	69.58	70.67	76.79	0.91	0.01
KNN	81.13	71.17	91.09	88.88	79.05	89.14	0.02	602.41
AdaBoost	85.66	81.64	89.68	88.78	85.06	93.54	47.69	1.47
Meta Random Forest	**88.68**	**83.26**	**94.10**	**93.38**	**88.03**	**94.37**	90.06	**7.67**

The final Meta- Random Forest model achieved the highest overall accuracy of 88.68% and specificity of 94.10%, providing the best performance amongst all models. The accuracy of the three-model stack ensemble (Model 4) was 88.14% with the best ROC AUC (94.54%), which in turn

reflected strong class separation. Nevertheless, the time required for training the Meta Random Forest predictor is much faster than that of the complex stack (over 15 times). Therefore, the Meta Random Forest model has been proposed as the best model, both in terms of high predictive accuracy and computational feasibility. An illustration of the performance comparison of the different models when FA was utilized is presented in Fig. 7.

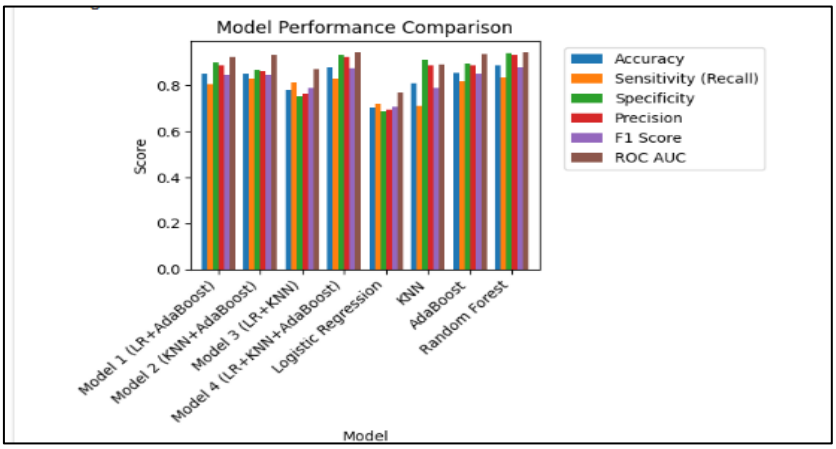

Figure 7. Performance Comparison of the Different Models with FA on the BRFSS dataset

The training and testing times for the developed models with FA on the BRFSS dataset are shown in Fig. 8.

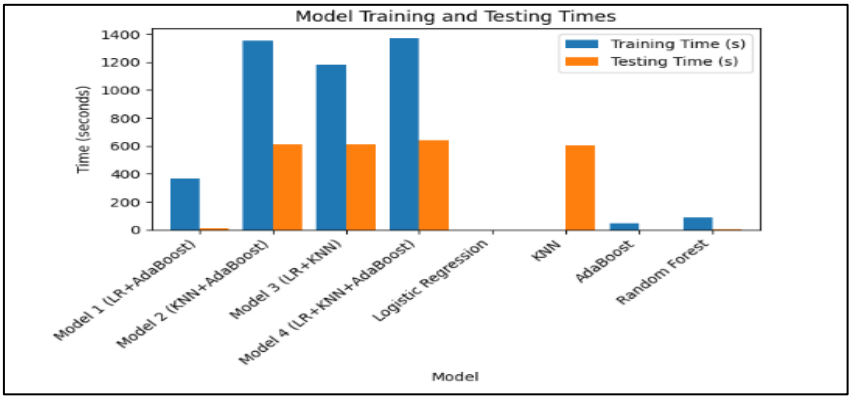

Figure 8. Training and Testing Time for the developed Models with FA on the BRFSS dataset

The ROC curve/AUC for Model 4 (LR+KNN+AdaBoost) is presented in Fig. 9. The ROC curve shows the sensitivity and false-positive rate of the model for different threshold values. The AUC of 0.95 indicates that the model has a 95% confidence level of correctly distinguishing between positive and negative instances. This high discriminative power, along with high accuracy and precision, confirms the robustness of Model 4 as a good classifier for this purpose.

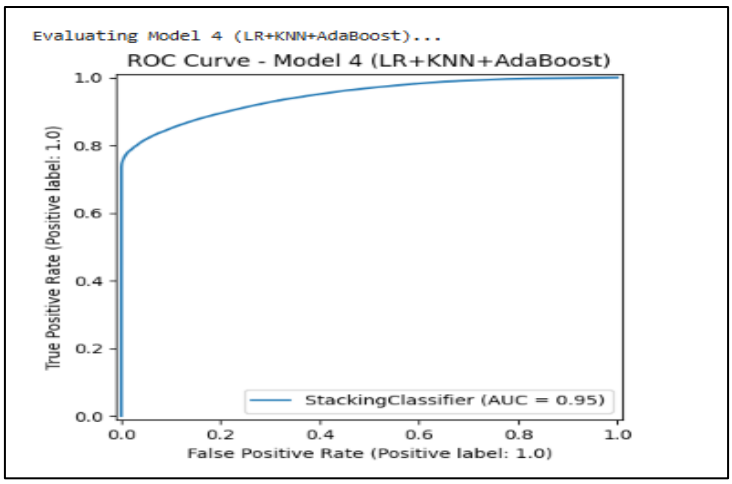

Figure 9. ROC curve/AUC score for Model 4 (LR + KNN + Adaboost) with FA on the BRFSS dataset

The confusion metrics for Model 4 (LR+KNN+AdaBoost) are shown in Fig. 10. The confusion matrix indicates that Model 4 accurately classified 97,713 instances (51,605 true negatives and 46,108 true positives) out of 110,862, with an overall accuracy of 88.14%. The model performs better on non-diabetes instances (specificity of 93.10%) than on diabetes instances (sensitivity of 83.18%), resulting in 3,826 false positives and 9,323 false negatives. This trend suggests that the model is trustworthy and gives more emphasis to minimizing the false positives, although it fails to detect some actual positive instances.

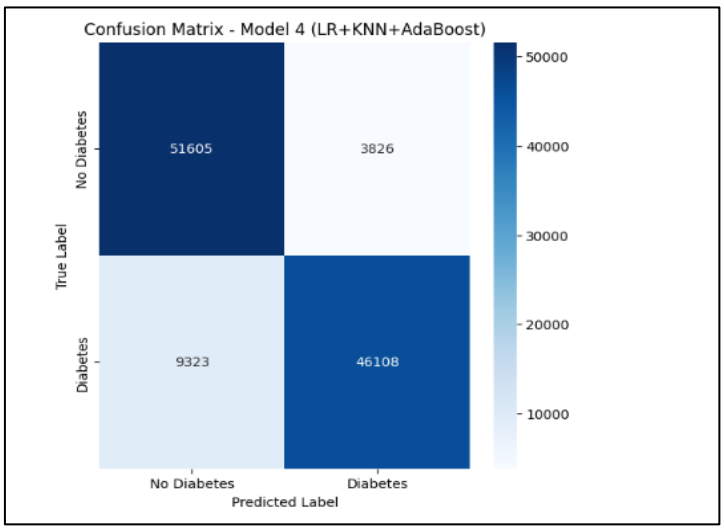

Figure 10. Confusion Metrix for Model 4 (LR + KNN + Adaboost) with FA on the BRFSS dataset

The proposed model's superior prediction ability can be established by comparing it with existing research works that have utilized the BRFSS dataset for prediction purposes related to Type II diabetes. In this context, the proposed model's accuracy, which is 88.68%, far exceeds the accuracy achieved by other researchers. Specifically, the accuracy achieved by Aamir and Murtza's model [16], which utilized the BRFSS dataset for prediction purposes, was only 85.03%, whereas the accuracy achieved by Chang et al.'s model [9], which utilized the BRFSS dataset for prediction purposes, was only 82.26%. Similarly, the accuracy achieved by Daliya and Ramesh's model [19], which utilized the BRFSS dataset for prediction purposes, was only 78.00%.

Table 7. Comparative Performance Evaluation of the Proposed Model Against Existing Studies Using the Accuracy Metric

S/N	Existing Studies on the BRFSS Dataset	Accuracy Obtained (%)
1	Aamir and Murtza [16]	85.03
2	Daliya and Ramesh [19]	78
3	Chang *et al.*, [9]	82.26
4	**Proposed FA + Stack-ensemble**	**88.68**

5. DISCUSSION

The Hybrid Firefly Algorithm has successfully identified the optimal combination of 16 predictive features. This addresses the significant research gap in the optimization of advanced features, as emphasized by Li et al. and Aman et al., who called for the development of techniques that improve the robustness of the models while reducing their complexity. The enhanced predictive capability of the proposed model is validated empirically through comprehensive benchmarking with the existing literature that used the BRFSS dataset to predict Type II Diabetes, as the FA-Stack Ensemble used in the proposed model achieves an accuracy of 88.68%, which is significantly higher compared to the 85.03% accuracy achieved by Aamir and Murtza [16], the 82.26% accuracy reported by Chang et al. [9], and the 78.00% accuracy reported by Daliya and Ramesh [19].

Among all configurations, the Meta Random Forest model provided the optimal balance between prediction accuracy and computational efficiency, achieving 88.68% accuracy and 94.10%

specificity, and running more than 15 times faster than complex stacking configurations. This answers the call for a clinically feasible solution by Abnoosian et al. and shows that the benefits of the ensemble method do not have to come at the expense of computational feasibility. The feature subset alone achieved an ROC AUC of 0.9439, and the ensemble configurations consistently outperformed the other models, validating the claim that the combination of intelligent feature selection and stacking effectively addresses the key limitations of dataset applicability, optimizations, and generalizability identified by Ikram, Reza, and Daza. While the trade-off between optimal performance and computational complexity cannot be eliminated, the framework provides a reliable and precise solution to the limitations identified in early Type II diabetes prediction.

6. CONCLUSION

The present study has successfully developed a novel hybrid framework by incorporating the intelligent feature selection method based on the Firefly Algorithm with an optimized stack ensemble model to overcome the limitations of existing traditional machine learning models for Type II diabetes prediction. The Firefly Algorithm identified 16 vial features that accurately predicted type II diabetes. At the same time, the proposed Meta Random Forest model, with the desired configuration, achieved superior results with an accuracy of 88.68%, a sensitivity of 83.26%, and an AUC of 94.37%. The proposed model's results are superior to existing research based on the BRFSS dataset, such as Aamir and Murtza (85.03%), Chang et al. (82.26%), and Daliya and Ramesh (78.00%), by 3.65%, 6.42%, and 10.68%, respectively. The novelty of this research work is the development of a reliable prediction model to overcome the limitations of existing research gaps for feature optimization and ensemble model configuration, which can be a valuable tool for the early prediction of Type II diabetes.

CONFLICT OF INTEREST

The authors declare that there is *no* conflict *of interest* regarding the publication of this paper.

REFERENCES

[1] G. Prabhakar, V. R. Chintala, T. H. Reddy, and T. Ruchitha, "User-Cloud-Based vEnsemble Framework for Type-2 Diabetes Prediction with Diet Plan Suggestion," *e-Prime*, vol. 7, p. 100423, 2024, doi: 10.1016/j.prime.2024.100423.

[2] O. Y. Hang, V. Wiwied, and R. Rosaida, "Diabetes Prediction using Machine Learning Ensemble Model," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 37, no. 1, pp. 82-98, 2024, doi: 10.3390/electronics806063.

[3] I. D. Mienye and Y. Sun, "Performance analysis of cost-sensitive learning methods with application to imbalanced medical data," *Informatics in Medicine Unlocked*, vol. 25, p. 100690, 2021, doi: 10.1016/j.imu.2021.100690.

[4] Z. Zhang et al., "A novel evolutionary ensemble prediction model using harmony search and stacking for diabetes diagnosis," *Journal of King Saud University Computer and Information Sciences*, vol. 36, no. 1, p. 101873, 2024, doi: 10.1016/j.jksuci.2023.101873.

[5] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/access.2022.3207287.

[6] A. V. Daza, C. M. F. Sánchez, O. Apaza, J. D. Pinto, and K. Z. Ramos, "Stacking ensemble approach to diagnosing the disease of diabetes," *Informatics in Medicine Unlocked*, vol. 44, p. 101427, 2024, doi: 10.1016/j.imu.2023.101427.

[7] V. C. S. S and A. H. S, "Nature inspired meta heuristic algorithms for optimization problems," Computing, vol. 104, no. 2, pp. 251–269, May 2021, doi: 10.1007/s00607-021-00955-5.

[8] W. Li, Q. Zhang, and H. Zhou, "Optimized Grid Search with Early Stopping for Efficient Hyperparameter Tuning," *Neural Networks*, vol. 158, pp. 210-225, 2023, doi: 10.1016/j.neunet.2023.02.018.

[9] V. Chang, M. Ganatra, K. Hall, L. Golightly, and Q. Xu, "An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators," *Healthcare Analytics*, vol. 2, p. 100118, 2022, doi: 10.1016/j.health.2022.100118.

[10] M. Z. Naser and A. Z. Naser, "The firefighter algorithm for optimization problems," Neural Computing and Applications, vol. 37, no. 16, pp. 9345–9400, Feb. 2025, doi: 10.1007/s00521-025-11074-z.

[11] J. Abdollahi and B. Nouri-Moghaddam, "Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction," Iran Journal of Computer Science, vol. 5, no. 3, pp. 205–220, Mar. 2022, doi: 10.1007/s42044-022-00100-1.

[12] M. S. Reza, R. Amin, R. Yasmin, W. Kulsum, and S. Ruhi, "Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data," Heliyon, vol. 10, no. 2, p. e24536, Jan. 2024, doi: 10.1016/j.heliyon.2024.e24536.

[13] A. Daza, C. F. P. Sánchez, G. Apaza-Perez, J. Pinto, and K. Z. Ramos, "Stacking ensemble approach to diagnosing the disease of diabetes," Informatics in Medicine Unlocked, vol. 44, p. 101427, Dec. 2023, doi: 10.1016/j.imu.2023.101427.

[14] W. Li, Y. Peng, and K. Peng, "Diabetes prediction model based on GA-XGBoost and stacking ensemble algorithm," PLoS ONE, vol. 19, no. 9, p. e0311222, Sep. 2024, doi: 10.1371/journal.pone.0311222.

[15] A. Dutta et al., "Early prediction of diabetes using an ensemble of machine learning models," International Journal of Environmental Research and Public Health, vol. 19, no. 19, p. 12378, Sep. 2022, doi: 10.3390/ijerph191912378.

[16] Z. Aamir and I. Murtza, "Pre-Diabetic Diagnosis from Habitual and Medical Features using Ensemble Classification," *Journal of Computing and Biomedical Informatics*, vol. 5, no. 1, 2023, doi: 10.56979/501/2023.

[17] V. Kansal, K. Upreti, J. Singh, M. Shanbhog and R. C. Poonia, "A Diabetes Detection Framework Based on Datadriven Predictive Technologies," 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG), Indore, Madhya Pradesh, India, India, 2025, pp. 1-6, doi: 10.1109/ICTBIG68706.2025.11323965.

[18] A. Aman and R. S. Chhillar, "Optimized stacking ensemble for early-stage diabetes mellitus prediction," International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering, vol. 13, no. 6, p. 7048, Sep. 2023, doi: 10.11591/ijece.v13i6.pp7048-7055.

[19] D. V. K and T. K. Ramesh, "Optimized stacking ensemble models for the prediction of diabetic progression," Multimedia Tools and Applications, vol. 82, no. 27, pp. 42901–42925, Apr. 2023, doi: 10.1007/s11042-023-14858-4.

[20] K. Abnoosian, R. Farnoosh, and M. H. Behzadi, "Prediction of diabetes disease using an ensemble of machine learning multi-classifier models," BMC Bioinformatics, vol. 24, no. 1, p. 337, Sep. 2023, doi: 10.1186/s12859-023-05465-z.

[21] M. F. Edafeajiroke, S. O. Abdulsalam, M. U. Shuaib, and R. S. Babatunde, "Development of an Intrusion Detection System using ANOVA Feature Selection and Support Vector Machine Algorithms," *Computology*, vol. 4, no. 1, p. 8908, 2024, doi: 10.17492/computology.v4i1.2406.

[22] M. F. Edafeajiroke and V. I. Osubor, "Optimizing random forest for type II diabetes prediction using firefly algorithm," *International Journal of Engineering Research & Technology (IJERT)*, vol. 15, no. 1, 2026, doi: 10.17577/IJERTV15IS010010.