Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 2: 2025
ISSN : **1906-9685**

Innovative Deep Learning Ensemble Model for Clinical Decision Support in Diabetes Prediction

#1 K.Durga Prasad, #2 K.JAYA KRISHNA

#1 MCA Scholar

#2 Assistant Professor

DEPARTMENT OF MASTER OF APPLICATIONS

QIS COLLEGE OF ENGINEERING AND TECHNOLOGY

Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh- 523272

Abstract: Diabetes is a significant global health concern, with an increasing number of individuals at risk. As a chronic disease, it leads to numerous fatalities each year, making early prediction essential to prevent its progression and reduce the risk of severe complications such as kidney and heart diseases. Accurate prediction models can help in early diagnosis and timely intervention. The Pima Indian Diabetes Dataset (PIMA-IDD-I), the Diabetes Dataset from Frankfurt Hospital, Germany (DDFH-G), and the Iraqi Diabetes Patient Dataset (IDPD-I) have been used to evaluate the performance of machine learning techniques in predicting diabetes. The ExtraTree FS technique, known for its feature selection capabilities, has been applied to optimize model performance across these datasets. Additionally, an ensemble approach involving a Voting Classifier, combining Boosted Decision Tree, Random Forest, and Bagged Extra Trees, has demonstrated promising results, achieving high accuracy across all three datasets. The proposed approach aims to leverage the strengths of these techniques, offering a robust solution for early diabetes detection and improving predictive accuracy across diverse datasets.

"Index Terms *-Artificialneuralnetworks, convolutional neural networks, diabetes mellitus, deep learning, ensemble learning, long short-term memory".*

1. INTRODUCTION

Diabetes is a significant global health concern, with an increasing number of diabetic patients at risk. It leads to a substantial loss of lives and presents a growing burden on healthcare systems worldwide [1]. Known as Diabetes Mellitus (DM), it is a metabolic disorder characterized by prolonged elevated blood glucose levels due to the body's inability to effectively process or consume glucose. The disease is associated with severe complications such as diabetic ketoacidosis, chronic renal failure, nonketotic hyperosmolar coma, foot ulcers, retinal damage, cardiovascular disease, stroke, and kidney failure [2]. There are three primary types of DM: Type 1 Diabetes (T1D), Type 2 Diabetes (T2D), and Gestational Diabetes [1], [2].

T1D occurs when the body cannot produce sufficient insulin, typically affecting younger

individuals under the age of 30. It is commonly marked by symptoms such as excessive thirst, frequent urination, and high blood glucose levels. People with T1D usually require insulin therapy for effective management [1]. T2D, on the other hand, is more prevalent and occurs when the body struggles to produce or properly use insulin. This type of diabetes mainly affects middle-aged and older adults and is often linked to lifestyle factors such as poor dietary habits, obesity, lack of physical activity, smoking, and hypertension [2]. Gestational diabetes occurs during pregnancy and typically resolves after childbirth, though it increases the risk of developing T2D later in life [1].

The global prevalence of diabetes is rising at an alarming rate. According to the World Health Organization (WHO), more than 420 million people worldwide are affected by DM, with over 650 million adults classified as obese, a condition closely associated with the onset of T2D. Obesity rates have tripled since 1975, contributing to the increase in diabetes cases globally. This growing epidemic poses significant challenges to public health systems, requiring enhanced efforts in prevention and management [3].

In the early stages of diabetes, many patients often underestimate the seriousness of their health condition, which can delay diagnosis and treatment, leading to severe complications and increased mortality rates [5], [6]. Early detection and accurate prediction of DM are crucial to reduce the progression of the disease and its associated health risks. Predicting diabetes in individuals of all ages is of utmost importance, as timely lifestyle changes can prevent the onset of the disease and its complications [8]. This highlights the urgent need for effective predictive models to aid in early diagnosis and intervention.

2. RELATED WORK

Diabetes Mellitus (DM) has become one of the most pressing public health concerns globally. With the increasing prevalence of diabetes, various machine learning (ML) techniques have been explored to develop predictive models to aid early diagnosis and effective management. Recent studies have leveraged machine learning algorithms, particularly ensemble methods and deep learning models, to predict and classify different types of diabetes and related complications.

Tong et al. [9] presented a study focused on predicting diabetes mellitus using machine learning techniques, highlighting the significant impact of early prediction on improving treatment outcomes. They explored several machine learning models, emphasizing the importance of selecting the appropriate algorithm for accurate diabetes prediction. Their work also focused on optimizing feature selection to enhance prediction accuracy. Their findings underline the importance of choosing effective ML techniques for the early detection of diabetes.

In a similar vein, Bhattacharya and Datta [10] developed predictive models for diabetes using ensemble machine learning classifiers. They utilized multiple base learners combined into a strong classifier to improve prediction accuracy. Their approach aimed to handle the inherent class imbalance in diabetes datasets, a common challenge in medical data. They demonstrated that ensemble techniques significantly improve performance compared to individual classifiers. The study emphasized the role of hybrid models in enhancing the robustness and accuracy of diabetes prediction, a strategy that can be particularly useful

in real-world applications where data is noisy and imbalanced.

Du et al. [11] introduced an explainable machine learning-based clinical decision support system specifically for predicting gestational diabetes mellitus (GDM). Their approach combined various machine learning algorithms with explainable artificial intelligence (XAI) techniques, providing not only predictions but also an interpretation of the model's decisions. This feature is crucial in clinical settings, where understanding the reasoning behind predictions can help healthcare providers make better-informed decisions. Their work highlights the growing demand for transparency and interpretability in medical prediction models, particularly in sensitive fields like diabetes care.

Ebrahim and Derbew [12] applied supervised machine learning algorithms to classify and predict type-2 diabetes (T2D) status in the Afar region of Ethiopia. Their study addressed the challenge of data scarcity in developing regions by utilizing locally available health data to build effective prediction models. By employing various machine learning algorithms, they demonstrated the feasibility of predicting T2D risk even in under-resourced settings. Their work contributes to the global effort to extend the benefits of machine learning in healthcare to low-income and developing regions.

Sarraju et al. [13] focused on understanding the reasons behind statin nonuse among diabetic patients using deep learning techniques applied to electronic health records. Their study leveraged deep learning algorithms to analyze large datasets and identify patterns that could explain why some diabetic patients fail to use statins despite the known benefits. This research underscores the potential of deep learning in identifying complex relationships in patient behavior and treatment adherence, providing insights that can help improve patient management strategies.

Thotad et al. [14] explored the detection and classification of diabetes disease using machine learning methods applied to Indian demographic and health survey data. Their work highlighted the application of various machine learning algorithms to predict diabetes risk in the Indian population, providing insights into how socioeconomic factors influence diabetes prevalence. They used data-driven techniques to identify key features contributing to the disease, contributing to the development of more region-specific prediction models. This work is significant in the context of India's rapidly growing diabetes problem and emphasizes the importance of region-specific datasets in building accurate predictive models.

Olickal et al. [15] proposed a system for assessing comprehensive diabetes care in primary care settings in India. Their study aimed to improve the quality of diabetes care by integrating machine learning models into primary care practices. They focused on the challenges faced by healthcare providers in rural and underserved areas, suggesting that machine learning could play a critical role in improving early detection, management, and prevention strategies for diabetes. The study emphasizes the importance of developing solutions that are not only accurate but also scalable and adaptable to various healthcare settings.

AlZu'bi et al. [16] developed a diabetes monitoring system for smart health cities based on big data intelligence. Their research integrated machine learning techniques with big data analytics to monitor and predict diabetes progression in real-time. By incorporating data from wearable devices

and electronic health records, they created a comprehensive system capable of tracking various health parameters. This study represents a step toward personalized diabetes care, where patients can be continuously monitored, and predictive models can provide timely interventions to prevent complications.

Pan et al. [17] introduced a risk prediction model for T2D complicated with retinopathy, utilizing machine learning techniques. Their model aimed to predict not only diabetes onset but also its complications, such as diabetic retinopathy, which is a major cause of blindness in diabetic patients. By combining ML algorithms with clinical data, they created a predictive system that could assist healthcare providers in identifying patients at high risk of developing complications. Their approach underscores the importance of multi-stage prediction models that consider both the onset of diabetes and its long-term complications.

3. MATERIALS AND METHODS

The proposed system aims to enhance diabetes prediction by leveraging multiple machine learning and deep learning techniques on three distinct datasets: the Pima Indian Diabetes Dataset (PIMA-IDD-I) [18], the Diabetes Dataset from Frankfurt Hospital, Germany (DDFH-G) [19], and the Iraqi Diabetes Patient Dataset (IDPD-I) [20]. Various algorithms will be applied, including traditional machine learning models [7] like Support Vector Machine (SVM), Decision Tree, and ExtraTree FS for feature selection. Additionally, deep learning models [21] such as Long Short-Term Memory (LSTM), Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and their hybrid forms (CNN+LSTM, Stack-LSTM, Stack-ANN, Stack-CNN, Stack-CNN+LSTM) will be employed for more complex pattern recognition. Ensemble techniques like a Voting Classifier combining Boosted Decision Tree, Random Forest, and Bagged Extra Trees will also be explored. This comprehensive approach, combining both machine learning and deep learning models, aims to optimize predictive accuracy and provide a robust solution for early diabetes detection across diverse datasets.

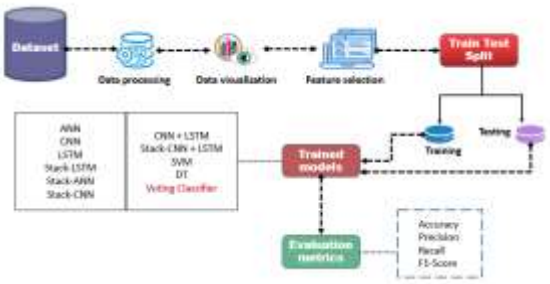


Fig.1 Proposed Architecture

The system (fig.1) employs a deep learning ensemble approach to enhance diabetes prediction accuracy. It begins with data processing and visualization, followed by feature selection. The dataset is split into training and testing sets. A diverse set of models, including ANN, CNN, LSTM, [21] and their stacked variations, are trained on the training data. The trained models are then evaluated using metrics like accuracy, precision, recall, and F1-score. The final prediction is made using a voting classifier that combines the outputs of the individual models.

i) Dataset Collection:

The Pima Indian Diabetes Dataset (PIMA-IDD-I) [18] is collected from the National Institute of Diabetes and Digestive and Kidney Diseases, focusing on female patients of Pima Indian heritage aged 21 or older. It contains 768 instances and 9 features: Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, and Outcome. The

dataset aims to predict whether a patient has diabetes based on these diagnostic measurements. The Outcome variable is the target, indicating whether the patient has diabetes (1) or not (0).

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Fig.2 Dataset Collection Table –Pima-IDDI

The Diabetes Dataset from Frankfurt Hospital, Germany (DDFH-G) [19] consists of 2000 entries and 9 features: Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, and Outcome. The dataset is used to predict whether a patient has diabetes, with the Outcome variable indicating the presence (1) or absence (0) of diabetes. This dataset provides valuable diagnostic information based on various medical predictors, including the patient's age, glucose level, insulin, BMI, and family history of diabetes.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	2	138	62	35	0	33.6	0.127	47	1
1	0	84	82	31	125	38.2	0.233	23	0
2	0	145	0	0	0	44.2	0.630	31	1
3	0	135	68	42	250	42.3	0.365	24	1
4	1	139	62	41	480	40.7	0.536	21	0

Fig.3 Dataset Collection Table – DDFH-G

The Iraqi Diabetes Patient Dataset (IDPD-I) [20] consists of 1000 entries and 14 features: ID, No_Patient, Gender, Age, Urea, Creatinine ratio (Cr), HbA1c, Cholesterol (Chol), Triglycerides (TG), HDL, LDL, VLDL, BMI, and Class. The dataset includes medical and laboratory data collected from patients at Medical City Hospital and Al-Kindy Teaching Hospital in Iraq. The target variable, Class, categorizes patients into Diabetic, Non-Diabetic, or Predict-Diabetic based on various health metrics such as blood sugar levels, lipid profiles, and BMI.

	ID	No_Patient	Gender	AGE	Urea	Cr	HbA1c	Chol	TG	HDL	LDL	VLDL	BMI	CLASS
0	502	17975	F	50	4.7	46	4.9	4.2	0.9	2.4	1.4	0.5	24.0	N
1	735	34221	M	26	4.5	62	4.9	3.7	1.4	1.1	2.1	0.6	23.0	N
2	420	47975	F	50	4.7	46	4.9	4.2	0.9	2.4	1.4	0.5	24.0	N
3	680	87656	F	50	4.7	46	4.9	4.2	0.9	2.4	1.4	0.5	24.0	N
4	504	34223	M	33	7.1	46	4.9	4.9	1.0	0.8	2.0	0.4	21.0	N

Fig.4 Dataset Collection Table – IDPD-I

ii) Pre-Processing:

Data pre-processing is a crucial step in preparing the dataset for analysis. It involves cleaning the data by handling missing values, duplicates, and irrelevant features, ensuring the dataset is ready for model training and provides accurate and reliable results.

a) Data Processing: The dataset is first checked for missing values, and any null entries are identified. Afterward, the rows with missing data are removed to ensure a clean dataset. Additionally, duplicates in the dataset are identified and removed to avoid redundancy, ensuring that each record is unique. The index is then reset to maintain the integrity of the dataset after these operations, making it ready for further analysis.

b) Data Visualization: The distribution of the target variable is visualized using a count plot, which helps in understanding the class balance of the dataset. A heatmap is also used to display correlations between different features, providing

insight into the relationships among the variables. These visualizations are useful for identifying patterns or anomalies in the data before moving on to modeling.

c) Label Encoding: In the Iraqi Diabetes Patient Dataset (IDPD-I), label encoding is applied to categorical variables such as gender and diabetes class. This process converts categorical string labels into numerical values, making them suitable for machine learning algorithms. For example, each gender is assigned a distinct number, and the diabetes class is encoded similarly, allowing for efficient processing of the data.

d) Feature Selection: Feature selection [7] is performed using the ExtraTree classifier, a decision-tree-based model that assigns an importance score to each feature. This process evaluates the impact of each feature on the model's predictive power. The importance of the features is visualized in a bar chart, helping to identify which attributes, such as cholesterol or BMI, are most influential in predicting diabetes. The features with the highest importance are selected for training the model.

e) Oversampling: To address class imbalance in the dataset, an oversampling technique [22] called SMOTE (Synthetic Minority Over-sampling Technique) is applied. This technique generates synthetic samples for the underrepresented class, ensuring that both classes have a more balanced number of instances. By applying SMOTE, the model is less likely to be biased toward the majority class, improving its ability to make accurate predictions for both classes.

iii) Training & Testing:

The model is trained and tested using the preprocessed dataset. First, the features are split into independent variables (X) and the target variable (y). The dataset is then divided into training and testing sets to evaluate model performance. The training data is used to fit the model, while the test data is reserved to assess its generalization ability. The model's accuracy and other performance metrics, such as precision and recall, are calculated to determine how well it predicts diabetes outcomes.

iv) Algorithms:

LSTM: This deep learning algorithm is designed to capture long-term dependencies in sequential data, making it ideal for tasks where patterns evolve over time, such as predicting diabetes progression based on historical health data.

ANN: A neural network architecture suited for complex patterns and relationships within datasets, it helps classify diabetes status by learning from non-linear interactions between features.

CNN: Convolutional layers are employed to extract spatial features from medical images, aiding in the identification of diabetes-related conditions from visual data or structured inputs.

CNN+LSTM: Combining CNN's feature extraction with LSTM's sequence processing capability, this hybrid model [21] is particularly effective for analyzing time-series or sequential data with spatial patterns, such as sensor data for predicting diabetes.

Stack-LSTM: This ensemble method leverages multiple LSTM models to improve prediction accuracy by combining different perspectives on temporal data, enhancing the reliability of diabetes predictions.

Stack-ANN: By stacking multiple artificial neural networks, this approach harnesses diverse learning patterns to refine diabetes classification, improving the system's ability to generalize to unseen data.

Stack-CNN: Multiple convolutional networks are stacked together, offering improved feature extraction capabilities. This method strengthens the system's ability to detect subtle relationships in data for accurate diabetes diagnosis.

Stack-CNN+LSTM: This hybrid ensemble combines CNN's spatial feature extraction with LSTM's temporal data processing, improving prediction accuracy for diabetes by learning from both sequential and spatial data sources.

SVM: Support Vector Machine classifies diabetes cases by finding the optimal hyperplane that separates different outcomes. [7] It's particularly effective in high-dimensional spaces with clear decision boundaries.

DT: A decision tree divides the dataset into subsets based on feature values, helping to classify diabetes cases by identifying the most significant features and their thresholds.

Voting Classifier: This ensemble method combines the predictions of multiple models, including Boosted Decision Tree, Random Forest, and Bagged Extra Trees, for a more robust and accurate diabetes diagnosis by leveraging the strengths of each individual model.

4. RESULTS & DISCUSSION

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} (1)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} \ + \ \mathrm{FN}} (3)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100 (1)$$

We evaluate the performance metrics—accuracy, precision, recall, and F1-score—for each algorithm in Tables 1, 2, & 3. The voting classifier achieves the highest scores. The table below also presents the metrics of other algorithms for comparison.

Table.1 Performance Evaluation Metrics – Data 1

ML Model	Accuracy	Precision	Recall	F1_score
LSTM	0.738	0.735	0.731	0.733
ANN	0.507	0.000	0.000	0.000
CNN	0.568	0.586	0.416	0.487
CNN+LSTM	0.710	0.640	0.939	0.761
Stack-LSTM	0.798	0.811	0.783	0.797
Stack-ANN	0.492	0.000	0.000	0.000
Stack-CNN	0.638	0.707	0.488	0.577
Stack-CNN+LSTM	0.740	0.671	0.956	0.789
SVM	0.730	0.730	0.730	0.730
Decision Tree	0.728	0.733	0.728	0.728
Voting Classifier	**0.795**	**0.804**	**0.795**	**0.796**

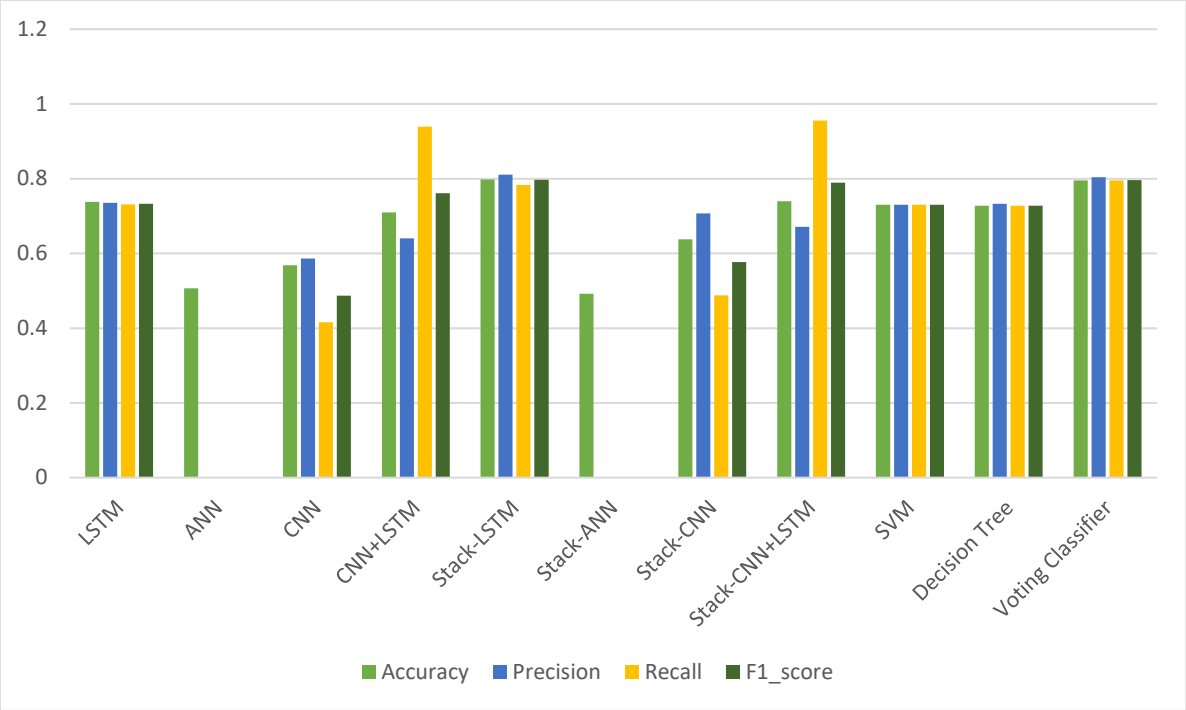
Table.2 Performance Evaluation Metrics – Data 2

ML Model	Accuracy	Precision	Recall	F1_score
LSTM	0.758	0.762	0.750	0.756
ANN	0.501	0.000	0.000	0.000
CNN	0.578	0.578	0.566	0.572
CNN+LSTM	0.733	0.763	0.673	0.715
Stack-LSTM	0.804	0.855	0.771	0.811
Stack-ANN	0.455	0.000	0.000	0.000
Stack-CNN	0.679	0.729	0.654	0.690
Stack-CNN+LSTM	0.768	0.855	0.692	0.765
SVM	0.728	0.730	0.728	0.728
Decision Tree	0.728	0.728	0.728	0.728
Voting Classifier	**0.819**	**0.821**	**0.819**	**0.820**

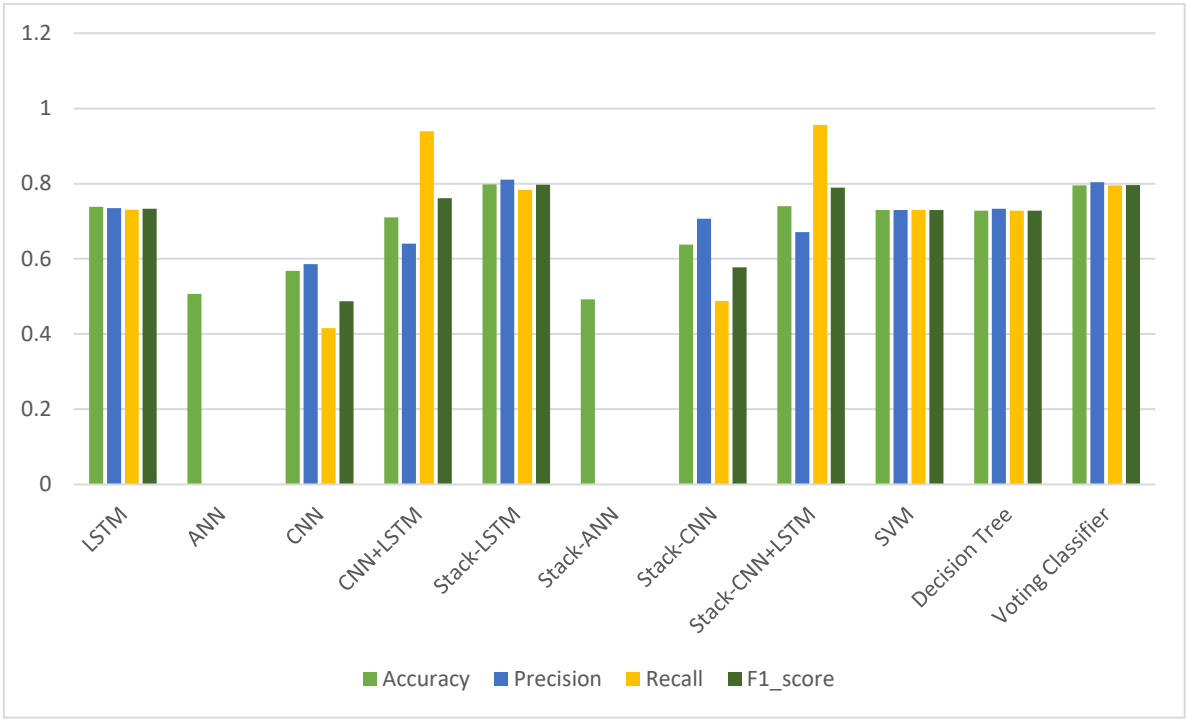
Table.3 Performance Evaluation Metrics – Data 3

ML Model	Accuracy	Precision	Recall	F1_score
LSTM	0.972	0.994	0.952	0.973
ANN	0.482	0.000	0.000	0.000
CNN	0.678	0.862	0.452	0.593
CNN+LSTM	0.962	1.000	0.927	0.962
Stack-LSTM	0.978	0.994	0.962	0.978
Stack-ANN	0.487	0.000	0.000	0.000
Stack-CNN	0.678	0.851	0.451	0.590
Stack-CNN+LSTM	0.968	1.000	0.938	0.968
SVM	0.937	0.944	0.937	0.937
Decision Tree	0.987	0.987	0.987	0.987
Voting Classifier	**0.993**	**0.993**	**0.993**	**0.993**

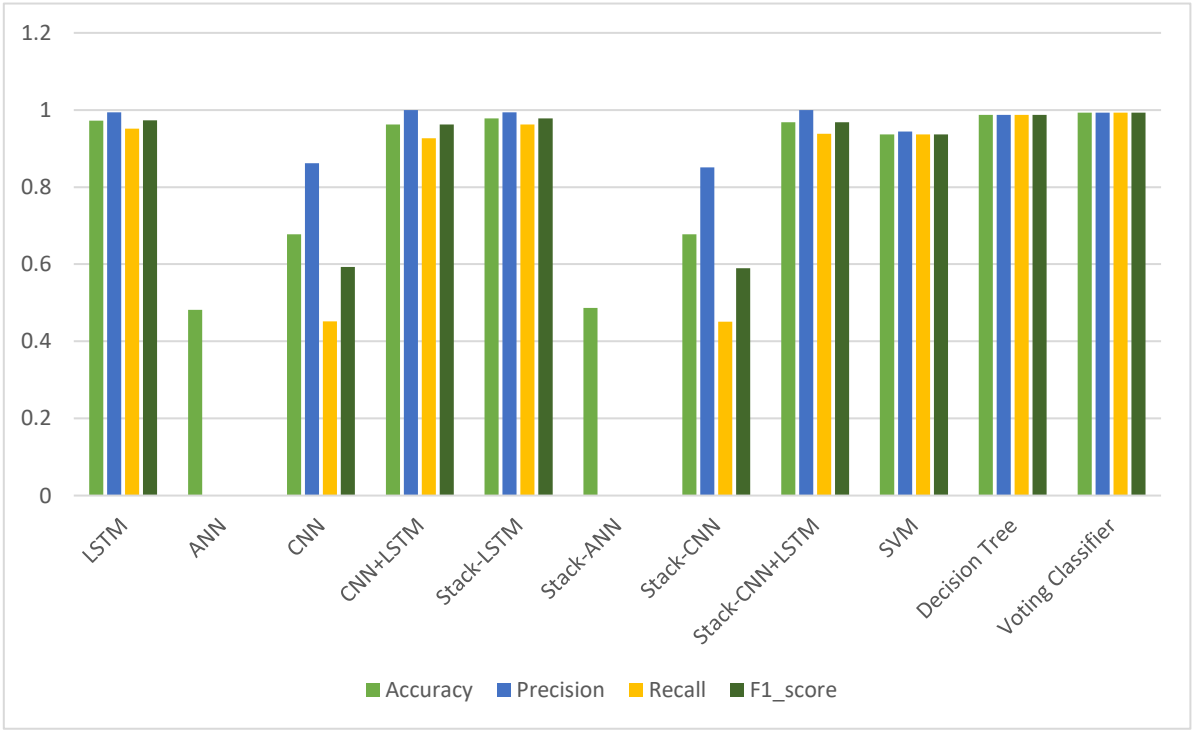
Graph.1 Comparison Graphs – Data 1



Graph.2 Comparison Graphs – Data 2



Graph.3 Comparison Graphs – Data 3



Graphs 1, 2, & 3 display accuracy in light green, precision in blue, recall in light yellow, and the F1 score in green. The voting classifier outperforms the other algorithms in all metrics, with the highest values compared to the remaining models. The above graphs visually represent these details.

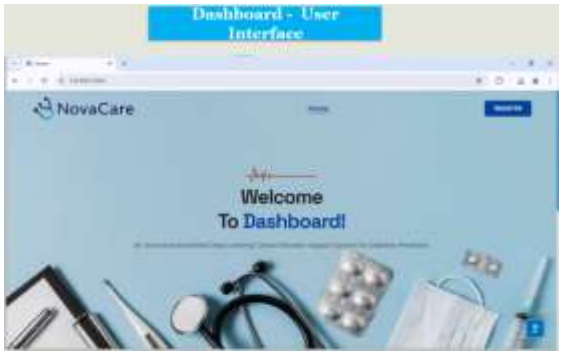


Fig.5 Home Page

In above fig.5 user interface dashboard with navigation and a welcome message.

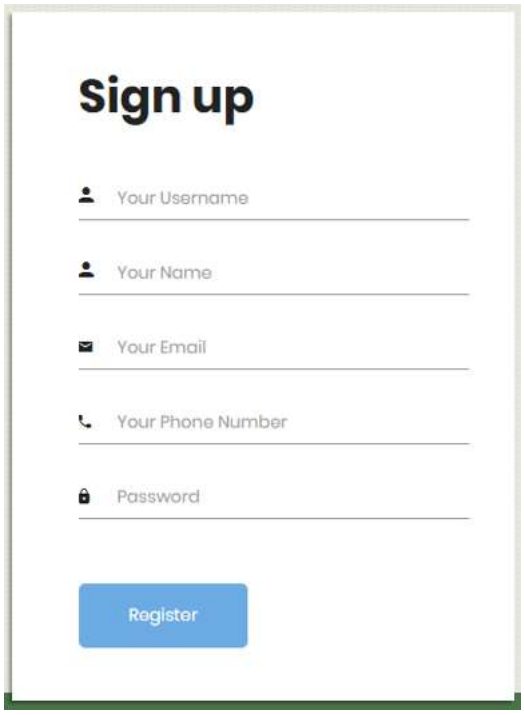


Fig.6 Registration Page

In above fig.6 sign-up form with fields for username, name, email, mobile number, and password buttons.

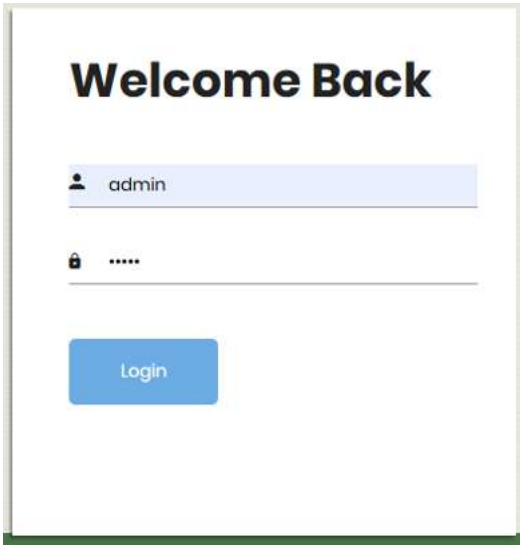


Fig.7 Login Page

In above fig.7Sign-in form with username and password fields.

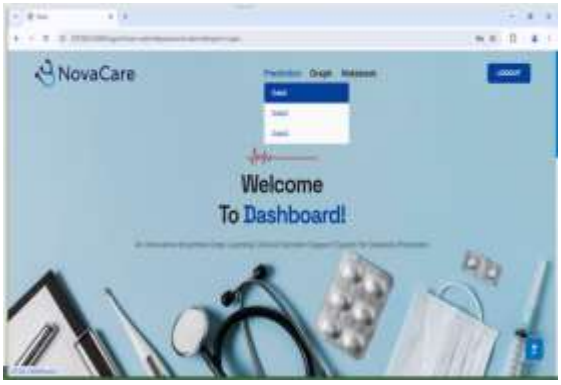


Fig.8 Main Page

In above Fig.8home page dashboard with navigation (Prediction, Graph, Notebook, Logout).

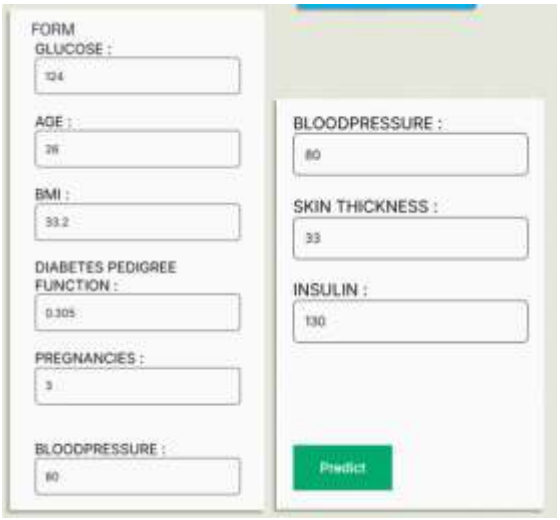


Fig.9 Upload Input Page

In above Fig.9 form with coordinate input field and upload button.

Fig.10 Predict Result for given input

In above Fig.10Predicted result based on the input test data.

5. CONCLUSION

In conclusion, the evaluation of various machine learning techniques for predicting diabetes has shown promising results using three distinct datasets: the Pima Indian Diabetes Dataset (PIMA-IDD-I), the Diabetes Dataset from Frankfurt Hospital, Germany (DDFH-G), and the Iraqi Diabetes Patient Dataset (IDPD-I). These datasets, each with unique characteristics, were leveraged to assess the efficacy of different algorithms, demonstrating the effectiveness of feature selection and ensemble methods. The ExtraTree FS technique proved to be invaluable in enhancing model performance by selecting the most relevant features, which is crucial for ensuring the accuracy and efficiency of diabetes prediction models. Additionally, the ensemble approach using a Voting Classifier, combining Boosted Decision Tree, Random Forest, and Bagged Extra Trees, achieved superior predictive accuracy across all three datasets. This approach harnesses the strengths of individual models, ensuring robustness and reliability in predictions. The findings underscore the potential of machine learning techniques in improving early diabetes detection, providing a solid foundation for healthcare applications aimed at early intervention.

The ***future scope*** of this work lies in further improving prediction accuracy by exploring advanced machine learning and deep learning techniques, such as reinforcement learning and hybrid models, to refine the detection of diabetes at earlier stages. Additionally, integrating more diverse and real-time datasets, including genetic and lifestyle factors, could enhance model robustness. Implementing explainable AI (XAI) methods would provide transparency and trust in the predictions, enabling better decision-making in clinical settings and improving personalized treatment plans for patients.

REFERENCES

[1] S. Binhowemel, M. Alfakhri, K. AlReshaid, and A. Alyani, ‘‘Role of artificial intelligence in diabetes research diagnosis and prognosis: A narrative review,’’ J. Health Inform. Developing Countries, vol. 17, no. 2, pp. 1–12, Aug. 2023. [Online]. Available: https://www.jhidc.org/index.php/jhidc/article/view/410

[2] R. M. Alamoudi et al., ‘‘Fasting Ramadan in patients with T1DM— Saudi Arabia versus other

countries during the COVID-19 pandemic,'' Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 17, no. 1, Jan. 2023, Art. no. 102676, doi: 10.1016/j.dsx.2022.102676.

[3] B. V. V. S. Prasad, S. Gupta, N. Borah, R. Dineshkumar, H. K. Lautre, and B. Mouleswararao, ''Predicting diabetes with multivariate analysis an innovative KNN-based classifier approach,'' Preventive Med., vol. 174, Sep. 2023, Art. no. 107619, doi: 10.1016/j.ypmed.2023.107619.

[4] World Population Rev. Diabetes Rates By Country 2023. Accessed: Feb. 10, 2023. [Online]. Available: https://worldpopulationreview.com/country-rankings/diabetes-rates-by-country

[5] V. Jaiswal, A. Negi, and T. Pal, ''A review on current advances in machine learning based diabetes prediction,'' Primary Care Diabetes, vol. 15, no. 3, pp. 435–443, Jun. 2021, doi: 10.1016/j.pcd.2021.02.005.

[6] A. Mujumdar and V. Vaidehi, ''Diabetes prediction using machine learning algorithms,'' Proc. Comput. Sci., vol. 165, pp. 292–299, Jan. 2019.

[7] S. Sivaranjani, S. Ananya, J. Aravinth, and R. Karthika, ''Diabetes prediction using machine learning algorithms with feature selection and dimensionality reduction,'' in Proc. 7th Int. Conf. Adv. Com put. Commun. Syst. (ICACCS), vol. 1, Mar. 2021, pp. 141–146, doi: 10.1109/ICACCS51430.2021.9441935.

[8] J. Chaki, S. T. Ganesh, S. K. Cidham, and S. A. Theertan, ''Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review,'' J. King Saud Univ. Comput. Inf. Sci., vol. 34, no. 6, pp. 3204–3225, Jun. 2022, doi: 10.1016/j.jksuci.2020.06.013.

[9] H. L. Tong, H. Ng, and H. Arul Ananthan, ''Predicting diabetes mellitus with machine learning techniques,'' J. Eng. Technol. Appl. Phys., vol. 6, no. 1, pp. 91–99, Mar. 2024.

[10] M. Bhattacharya and D. Datta, ''Development of predictive models of diabetes using ensemble machine learning classifier,'' in Proc. 1st Int. Conf. Advancements Smart Comput. Inf. Secur. (ASCIS), Rajkot, India, 2022, pp. 377–388.

[11] Y. Du, A. R. Rafferty, F. M. McAuliffe, L. Wei, and C. Mooney, ''An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus,'' Sci. Rep., vol. 12, no. 1, p. 1170, Jan. 2022, doi: 10.1038/s41598-022-05112-2.

[12] O. A. Ebrahim and G. Derbew, ''Application of supervised machine learn ing algorithms for classification and prediction of type-2 diabetes disease status in afar regional state, Northeastern Ethiopia 2021,'' Sci. Rep., vol. 13, no. 1, p. 7779, May 2023, doi: 10.1038/s41598-023-34906-1.

[13] A. Sarraju, A. Zammit, S. Ngo, C. Witting, T. Hernandez-Boussard, and F. Rodriguez, ''Identifying reasons for statin nonuse in patients with dia betes using deep learning of electronic health records,'' J. Amer. Heart Assoc., vol. 12, no. 7, Apr. 2023, Art. no. e028120.

[14] P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, ''Diabetes dis ease detection and classification on Indian demographic and health survey data using machine learning methods,'' Diabetes Metabolic Syn drome, Clin. Res. Rev., vol. 17, no. 1, Jan. 2023, Art. no. 102690, doi: 10.1016/j.dsx.2022.102690.

[15] J. J. Olickal, P. Chinnakali, B. S. Suryanarayana, S. Rajanarayanan, T. Vivekanandhan,G.K.Saya,K.Ganapathy,andD.K.S. Subrahmanyam, ‘‘Down referral and assessing comprehensive diabetes care in primary care settings: An operational research from India,’’ Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 17, no. 1, Jan. 2023, Art. no. 102694, doi: 10.1016/j.dsx.2022.102694.

[16] S. AlZu’bi, M. Elbes, A. Mughaid, N. Bdair, L. Abualigah, A. Forestiero, and R. A. Zitar, ‘‘Diabetes monitoring system in smart health cities based on big data intelligence,’’ Future Internet, vol. 15, no. 2, p. 85, Feb. 2023, doi: 10.3390/fi15020085.

[17] H. Pan et al., ‘‘A risk prediction model for type 2 diabetes mellitus com plicated with retinopathy based on machine learning and its application in health management,’’ Frontiers Med., vol. 10, 2023, Art. no. 1136653.

[18] Kaggle, UCI Mach. Learn. (2023). Pima Indians Diabetes Database. Accessed: Feb. 1, 2023. [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

[19] Kaggle. (2023). Dataset of Diabetes, Taken From the Hospital Frank furt, Germany. Accessed: Feb. 1, 2023. [Online]. Available: https://www.kaggle.com/datasets/johndasilva/diabetes

[20] A. Rashid. (2020). ‘Diabetes Dataset’, Mendeley Data, V1. Accessed: Feb. 1, 2023. [Online]. Available: https://data.mendeley.com/datasets/wj9rwkp9c2/1

[21] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani, and A. Shaikh, ‘‘A robust heart disease prediction system using hybrid deep neural networks,’’ IEEE Access, vol. 11, pp. 121574–121591, 2023, doi: 10.1109/ACCESS.2023.3328909.

[22] M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, ‘‘Diabetes detection using deep learning techniques with oversampling and feature augmentation,’’ Comput. Methods Programs Biomed., vol. 202, Apr. 2021, Art. no. 105968, doi: 10.1016/j.cmpb.2021.105968.

AUTHORS Profile

Mr. K. Jaya Krishna is an Associate Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Anna University, Chennai, and his M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). With a strong research background, he has authored and co-authored over 90 research papers published in reputed peer-reviewed Scopus-indexed journals. He has also actively presented his work at various national and international conferences, with several of his publications appearing in IEEE-indexed proceedings. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing,

and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

Mr. K. Durga Prasad has received his MCA (Masters of Computer Applications) from QIS college of Engineering and Technology Vengamukkapalem (V), Ongole, Prakasam dist., Andhra Pradesh-523272 affiliated to JNTUK in 2023-2025