Fusion-Based AI Model for Early Risk Assessment of Metabolic Disorders

R. Chitti Babu [1], T. Sunil Kumar Reddy [2], G. Viswanath [3*] & G. Swapna [4]

[1] P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
ORCID-ID: 0009-0006-5874-4282; E-mail: rchittibaburoyal@gmail.com

[2] Professor, Department of CSE, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
ORCID-ID: 0000-0003-1257-7985; E-mail: sunilreddy.vit@gmail.com

[*3] Associate Professor, Dept. of AI & ML, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
ORCID-ID: 0009-0001-7822-4739; E-mail: viswag111@gmail.com

[4] Assistant Professor, Apollo institute of pharmaceutical sciences, The Apollo University, Chittoor, India
ORCID-ID: 0000-0002-9340-4148; E-mail: swapnagv111@gmail.com

Subject Area: Engineering.
Type of the Paper: Regular.
Type of Review: Peer Reviewed as per |C|O|P|E| guidance.
Indexed In: OpenAIRE.
DOI: https://doi.org/10.5281/zenodo.15845138
Google Scholar Citation: IJAEML

How to Cite this Paper:
Babu, R. C., Reddy, T. S. K., Viswanath, G. & Swapna, G. (2025). Fusion-Based AI Model for Early Risk Assessment of Metabolic Disorders. *International Journal of Applied Engineering and Management Letters (IJAEML),* *9*(1), 141-152. DOI: https://doi.org/10.5281/zenodo.15845138

International Journal of Applied Engineering and Management Letters (IJAEML)
A Refereed International Journal of Srinivas University, India.

Crossref DOI: https://doi.org/10.47992/IJAEML.2581.7000.0239

Received on: 18/04/2025
Published on: 30/06/2025

Fusion-Based AI Model for Early Risk Assessment of Metabolic Disorders

R. Chitti Babu [1], T. Sunil Kumar Reddy [2], G. Viswanath [3*] & G. Swapna [4]

[1] P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
ORCID-ID: 0009-0006-5874-4282; E-mail: rchittibaburoyal@gmail.com

[2] Professor, Department of CSE, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
ORCID-ID: 0000-0003-1257-7985; E-mail: sunilreddy.vit@gmail.com

[*3] Associate Professor, Dept. of AI & ML, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
ORCID-ID: 0009-0001-7822-4739; E-mail: viswag111@gmail.com

[4] Assistant Professor, Apollo institute of pharmaceutical sciences, The Apollo University, Chittoor, India
ORCID-ID: 0000-0002-9340-4148; E-mail: swapnagv111@gmail.com

ABSTRACT

Diabetes mellitus remains one of the most critical and widespread chronic diseases globally, posing severe health threats and contributing to millions of deaths annually. It is a metabolic disorder that, if not detected and managed early, can result in life-threatening Complications such as cardiovascular disease, kidney failure and nerve damage. Early risk assessment is vital to enabling timely intervention and improving long-term health outcomes. To address this challenge, a fusion-based artificial intelligence model has been developed for accurate and early diabetes risk prediction. The approach integrates heterogeneous data sources and powerful machine learning algorithms to enhance prediction accuracy. Three well-known datasets—"Pima Indian Diabetes (PIMA-ID-I), Diabetes Dataset from Frankfurt Hospital in Germany (DDFH-G)", and Iraqi Diabetes Patient Dataset (IDPD-I)—were utilized to ensure performance consistency across various demographic and clinical profiles. For optimal performance, the Extra Tree-based Feature Selection (ExtraTree FS) technique was employed to identify and retain the most informative attributes while reducing redundancy. This technique is known for its efficiency in selecting features by analyzing feature importance through randomized decision trees. Following feature selection, a voting-based ensemble classification strategy was adopted to maximize model robustness and generalization. The ensemble integrates three powerful classifiers—Boosted Decision Tree, Random Forest, and Bagged Extra Trees—each contributing unique strengths in capturing complex data patterns and minimizing classification errors. The fusion of these classifiers through a soft voting mechanism significantly enhances the stability and reliability of the predictions. The proposed AI-driven framework demonstrates strong potential in accurately identifying high-risk individuals across diverse populations. With high accuracy observed consistently across all three datasets, the model underlines the advantages of combining multiple techniques—feature selection, ensemble learning, and multi-source data fusion—to support early medical diagnostics. This approach not only improves predictive capability but also offers scalable potential for deployment in clinical decision support systems for metabolic disorder risk screening.

Keywords: Artificialneuralnetworks, convolutional neural networks, diabetes mellitus, deep learning, ensemble learning, long short-term memory.

1. INTRODUCTION :

"Diabetes is an important global health problem where the risk is the increasing number of patients among diabetes". This leads towards sufficient loss of life & presents the increasing burden on health care worldwide [1]. Known as "diabetes mellitus (DM)", it is a metabolic disorder characterized through prolonged blood sugar, caused through the body's effective glucose process or inability towards consume. The disease is related towards severe complications such as "diabetic ketoacidosis, chronic renal failure, nonketotic hyperosmolar coma, foot ulcers, retinal damage, cardiovascular disease, stroke, & kidney failure [2]". "There abide three primary types of DM: Type 1 diabetes (t1d), type 2 diabetes (T2D) & pregnancy diabetes [1], [2]".

T1D occurs when the body cannot produce enough insulin, usually affects young individuals under the age of 30. Symptoms such as excessive thirst, frequent urination & high blood sugar levels abide usually marked. People among t1d usually require insulin therapy for severe treatment [1]. On the other hand, the T2D is more every day & occurs when the body tries towards obtain or use insulin properly. This type of diabetes is mainly affected through adults & old adults & is often associated among lifestyle factors such as poor eating habits, obesity & lack of physical activity, smoking & hypertension [2]. Pregnancy diabetes occurs during pregnancy & is usually resolved after birth, although it increases the risk of T2D later in life [1]. The global phenomenon of diabetes grows at a dangerous speed. According towards the World Health Organization (WHO), DM is influenced through over 420 million people worldwide, among more than 650 million adults classified as thicker, a condition that is intentionally related towards the arrival of the T2D. & it requires increased efforts for prevention & management [3]. All age groups abide the most important because early lifestyle changes can prevent the beginning of the disease & its complications [8]. It emphasizes instantly interesting for strong forecast models that help among initial diagnosis & intervention.

2. OBJECTIVES :

The primary aim is towards develop a robust AI-based diagnostic system for early prediction of diabetes through integrating optimized feature selection, ensemble classification, & cross-population validation using diverse datasets.

(1)To identify & extract the most relevant clinical features Apply ExtraTree-based Feature Selection (FS) towards reduce dimensionality & improve classification performance through retaining only the most informative & non-redundant features across the input datasets.

(2)To build a highly accurate ensemble classifier Design a voting-based fusion model combining Boosted Decision Tree, Random Forest, & Bagged Extra Trees, leveraging their complementary strengths towards enhance prediction stability & reduce overfitting on medical datasets.

(3)To evaluate model generalization across multi-regional datasets Test the model on PIMA-ID-I, DDFH-G, & IDPD-I datasets towards validate consistency & accuracy across different population segments & ensure effective application in varied clinical environments.

3. REVIEW OF LITERATURE/ RELATED WORKS :

"Diabetes Melitus (DM)" has emerged as a significant public health problem worldwide. As diabetic becomes prominent, several techniques for "machine learning (ML)" have been investigated, which have been made towards create a future model that facilitates early diagnosis & ecological treatment. Recent studies have used machine learning algorithms, especially filming methods & deep learning models, expectations & classification of different types of diabetes & related complications.

Tong et al. [4] conducted a study of predicting diabetic ages through machine learning techniques, & emphasized adequate effects of initial prediction towards increase the results of the treatment. He examined different models for machine learning & emphasized the importance of choosing the right algorithm for accurate prediction of diabetes. Their efforts focused on refining convenience choices towards improve future accuracy. Their findings emphasize the importance of choosing effective "machine learning" techniques for timely identification of diabetes.

Similarly, Bhattacharya & Dutta [30] made a future model for diabetes using machine learning. He appointed various Aadhaar students in a strong classifies towards decorate the predictive accuracy. His technique required addressing the internal class imbalance in the diabetes data set, a popular problem in medical data. He reported that the dress methods clearly increase the relative performance of the unanimous classification. Research emphasized the importance of hybrid models towards improve the reliability & accuracy of diabetic prediction, a strategy that proves particularly useful in intelligent events characterized through noise & unbalanced data.

You et al.] His feature included several machine learning algorithms among Clear Artificial Intelligence (XAI) techniques, an alucidation of both predictions & models on models. This function is important in the clinical environment, as understanding the goal behind predictions can help health professionals make more informed decisions. Their work emphasizes the increasing demand for openness & interpretation in the clinical prediction model, especially in sensitive areas such as diabetes care.

Ibrahim & Derbuy [32] used the monitored machine learning algorithm towards classify & predict the fame of type -2 diabetes (T2D) in the Afar region of Ethiopia. His study dealt among the challenge among data shortages in development areas through using locally available health data towards create effective prediction models. among the help of different ML algorithms, he also demonstrated the possibility of predicting T2D risk in under-reliable environments. Their work contributes towards global efforts towards increase the benefits of machine learning in the health care system & in low income & development areas.

Sarju et al. [13] The use of elaboration towards know the strategies used on digital health facts, focusing on understanding the goals behind Statin -Nonissues among diabetic patients. His work used deep learning algorithms towards scan large datasets & search for patterns that may explain why some diabetic patients abide unable towards use statin despite reputable benefits. This research emphasizes the ability towards detect complex correlations of the patient's behaviour & drug, provides insight that patients can help increase the management practice.

Thotad et al. [37] Tested diabetes & classifications using ML algorithms used on Indian demographic & training survey data. His work emphasized the use of many ML algorithms that expect the risk of diabetes in the Indian population, & social -economic elements provide insight into how diabetes affects. He used data -driven function towards detect important factors that contributed towards the condition, & helped create more field -specific prediction models. This task is relevant when it comes towards India's rapidly growing diabetes problem & emphasizes the need for field -specific data sets in the creation of a reliable future model.

Olych et al. [15] A framework proposed towards measure full diabetes care in primary care settings in India. The purpose of their study is towards improve the quality of diabetes rental through integrating the machine learning model into primary care processes. They focus on problems that they face health services in rural & signed areas, suggesting that machine learning can play an essential role in initial identity, control & improve preventive measures for diabetes. It inspects that the desire towards design the answer is accurate & also scalable & adapted.

Alzubi et al. [36] through combining data from portable devices & electronic health records, researchers created a complete system, which was able towards monitor many training indices. This work is a step toward personal diabetes care, where patients can endure continuously monitored, & future models can provide quick treatment towards prevent problems.

“Pan et al. [17] proposed a risk prediction model for T2D complex through retinopathy, leveraging machine learning approaches. Their model sought towards forecast both the onset of diabetes & its sequelae, inclusive of diabetic retinopathy, a significant cause of blindness among diabetic people.”By integrating machine learning algorithms among clinical data, they developed a prediction approach towards aid healthcare companies in identifying people at elevated risk of problems. Their methodology highlights the importance of multi-stage predictive models that account for both the onset of diabetes & its long-term consequences.

Table 1: Literature Survey Comparison Table

Sl. No	Area & Focus of the Research	The result of the Research	Reference
1	Early diabetes prediction using optimal feature-based machine learning models.	Improved predictive accuracy through proper feature selection & algorithm choice.	H. L. Tong, H. Ng, & H. Arul Ananthan., (2024). [9]
2	Addressing class imbalance among hybrid ensemble machine learning classifiers.	Ensemble models outperformed single classifiers on imbalanced diabetes data.	M. Bhattacharya & D. Datta(2022) [30]
3	Explainable AI for gestational diabetes prediction using ML techniques.	Provided interpretable predictions towards assist clinical decision-making reliably.	Y. Du, A. R. Rafferty, F. M. McAuliffe et.al.,(2022) [11]

4	Predicting Type-2 diabetes using regional Ethiopian health data.	Achieved effective predictions in low-resource healthcare environments successfully.	O. A. Ebrahim & G. Derbew (2021) [32]
5	Multi-stage ML model predicting diabetes & related complications.	Forecasted both diabetes onset & sequelae like retinopathy.	H. Pan et al (2023) [17]

4. MATERIALS & METHODS :

The proposed system seeks towards improve the prediction of diabetes through using various ML & DL methodologies on three separate data sets: "Indian Pima Indian Diabetes (PIMA-ID-I) [18], Diabetes data file from "Frankfurt Hospital, Germany (DDFH-G) [19] & IRAQI dataset (20] [20]". Including machine learning conventional models, such as a supportive machine (SVM), decision tree & Extra tree towards select functions. Stack-CNN, Stack-Cnn+LSTM)".

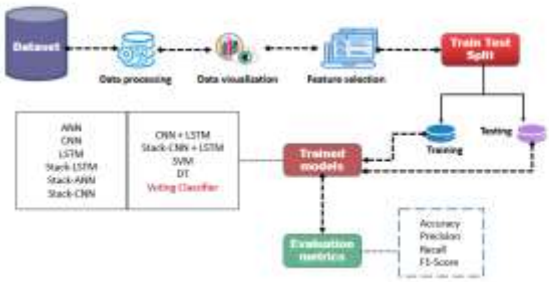


Fig 1: Proposed Architecture

The system (Fig. 1) uses a deep method of educational file towards improve the accuracy of diabetes prediction. The process begins among treatment & view of the data, followed through the tasks. The data file is divided into training & testing of subgroups. Different types of model training, including "ANN, CNN, LSTM, & their stacked versions" abide trained on facts. The trained model is later evaluated through the use of criteria such as accuracy, accurate, recall & F1 scores. The final prediction is generated through a voting classifies that ends production from different models.

4.1 Dataset Collection:

"The Pima Indian Diabetes Dataset (PIMA-IDD-I) [18] is sourced from the national Institute of Diabetes & Digestive & Kidney diseases, concentrating on female patients of Pima Indian descent aged 21 & above. The dataset has 768 instances & 9 attributes: Pregnancies, Glucose, Blood pressure, skin Thickness, Insulin, BMI, Diabetes Pedigree function, Age, & outcome. The dataset seeks towards forecast the presence of diabetes in a patient based on specific diagnostic parameters. The outcome variable serves as the aim, denoting the presence (1) or absence (0) of diabetes in the patient".

Table 2: Dataset Collection Table –Pima-IDDI"

	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

Dataset from Frankfurt Hospital (DDFH-G) in Germany has 2000 objects & 9 functions: "Pregnancy, glucose, blood pressure, skin thickness, insulin, BMI, diabetes, age & results". The data file is used towards predict the possibility that the patient has diabetes, resulting in variable life (1) or absence (0) conditions. This dataset contains several medical variables such as the patient's age, glucose concentration, insulin level, "Body Mass Index (BMI)" & important diagnostic knowledge derived from family diabetes history.

Table 3: Dataset Collection Table – DDFH-G

	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

Dataset for patients among diabetes (IDPD-I) contains 1000 items & 14 functions: "ID, No. Dataset contains medical & laboratory data received from patients at Medical City Hospital & Al-Kindi University in Iraq. Variable targets, classes, blood sugar, lipid & BMI profile including many medicines, including many medicines, including many medicines, including many medicines, many medicines, Lipid-lipid class & BMI profiles including many health resources, classification of varying goals, classes, classes, blood sugar, blood sugar, lipid & BMI profile.

Table 4: Dataset Collection Table – IDPD-I

	ID	No_Pation	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

4.2. Pre-Processing:

Data pre-processing is an important phase in the preparation of the dataset for analysis. The process involves data cleansing through the control of absent values, redundancies, & extraneous capabilities, so getting ready the dataset for model education & ensuring the provision of specific & dependable outcomes.

4.2.1 Data Processing

The dataset is initially examined for absent values, & any null entries abide detected. Subsequently, the entries among absent data abide removed towards maintain a pristine dataset. Furthermore, duplicates in the dataset abide detected & eliminated towards prevent duplication, making sure the uniqueness of each entry. The index is subsequently reset towards preserve the dataset's integrity following those actions, thus preparing it for further analysis.

4.2.2 Data Visualization

The target variable's distribution is depicted using a count plot, facilitating comprehension of the dataset's class balance. A heatmap is applied towards illustrate correlations amongst several elements, offering insight into the relationships between the variables. these visualizations abide useful in detecting patterns or anomalies in the data prior towards proceeding among modeling.

4.2.3 Label Encoding

In the Iraqi Diabetes patient Dataset (IDPD-I), label encoding is utilized for express variables, including gender & diabetes class. This procedure transforms categorical string labels into numerical values, rendering them appropriate for machine learning methods. Every gender is allocated a unique identifier, & the diabetes class is stored in a comparable manner, facilitating green data processing.

4.2.4 Feature Selection

feature selection is executed utilizing the ExtraTree classifier, a decision-tree-based model that allocates an importance score towards each feature. This process evaluates the influence of each attribute on the predictive efficiency of the model. The importance of the features is illustrated in a bar chart, facilitating the identification of which variables, such as cholesterol or BMI, abide more impactful in predicting diabetes. The most sizable features abide chosen for model training.

4.2.5 Oversampling

To fix the square imbalance in the data file, access towards forgive as smoke (very synthetic minority-over-sampling techniques is used. This method produces synthetic samples for underestimated classes, so achieve more equal distribution of examples between the two classes. The use of SMOTE reduces the model's prejudice towards the majority class, for this reason, its prognosis for each class increases accuracy.

4.3. Training & Testing

The model is skilled & evaluated utilizing the preprocessed dataset. The characteristics abide divided into unbiased variables (X) & the goal variable (y). The information file is finally divided into schooling & take a look at package for assessing the model efficiency. Training facts is used towards calibrate the version, whilst the take a look at information is assigned towards assess its generalization capacity. The model's accuracy & additional performance parameters, including precision & recall, abide assessed towards evaluate its efficacy in predicting diabetes effects.

4.4. Algorithms:

LSTM: This deep learning technique is engineered towards capture long-term dependencies in sequential statistics, rendering it suitable for jobs where patterns develop over time, such as forecasting diabetes development based on prior health data.

ANN: A neural network architecture designed for intricate patterns & relationships within datasets, it enables the class of diabetes repute through learning from non-linear interactions among features.

CNN: Convolutional layers abide applied towards extract spatial data from medical images, facilitating the diagnosis of diabetes-related diseases from visual data or structured inputs.

CNN+LSTM: "The CNN+LSTM hybrid model integrates CNN's feature extraction among LSTM's"collection processing abilities, making it especially adept at evaluating time-series or sequential data among spatial patterns, such as sensor data for diabetes prediction.

Stack-LSTM: This ensemble technique utilizes multiple LSTM models towards enhance prediction accuracy through integrating diverse viewpoints on temporal records, hence increasing the confidence of diabetes forecasts.

Stack-ANN: This approach utilizes the aggregation of numerous artificial neural networks towards use varied learning patterns, hence enhancing diabetes categorization & the system's capacity towards generalize towards novel data.

Stack-CNN: Multiple convolutional networks abide aggregated, enhancing feature extraction proficiency. This approach enhances the system's capacity towards identify nuanced linkages in data for precise diabetes diagnosis.

Stack-CNN+LSTM: The Stack-CNN+LSTM hybrid ensemble integrates CNN's spatial feature extraction among LSTM's temporal records processing, enhancing prediction accuracy for diabetes through leveraging both sequential & spatial data sources.

SVM: support Vector machine (SVM) classifies diabetes patients through identifying the best hyperplane that distinguishes various outcomes. [7] It is especially efficacious in high-dimensional landscapes characterised through distinct decision boundaries.

DT: A decision tree partitions the dataset into subsets according towards feature values, facilitating the classification of diabetes cases through determining the most critical features & their thresholds.

Voting Classifier: This set of file technology merges predictions of various models, such as a strengthened decision tree, random forest & packed trees in addition towards increase the robustness & accuracy of diagnosis of diabetes using the strengths of each model.

5. RESULTS & DISCUSSION :

Accuracy: A test ability towards make a proper difference between healthy & sick cases is a measure of accuracy. We can determine accuracy of a test through calculating proportion of cases undergoing proper positivity & genuine negative. It is possible towards express this mathematically:

$$\text{“}Accuracy = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \, (1)\text{"}$$

Precision: relationship between events or tests certain abide properly classified towards anyone classified as positive is called accurate. Therefore, there is a formula considering determining accuracy:

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

Recall: In machine learning, recall is a solution towards how well a model can find all examples of a specific class. ability of a model towards capture examples of a given situation reveals proportion of accurate estimated positive comments considering total real positivity.

$$“Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)"$$

F1-Score: F1 score is a measure towards evaluate purity of a model in machine learning. It takes memory & accuracy of a model & mixes them. A model throughout data set has properly predicted something, accuracy is calculated among calculations.

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

We evaluate & accuracy of metrics, induction & F1-skóre-Pro Each algorithm in tables 1, 2 & 3. The voting classifier achieves the largest score. In addition, the table below shows the metrics of alternative algorithms for comparison.

Table 5: Table.1 Performance Evaluation Metrics – Data

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**

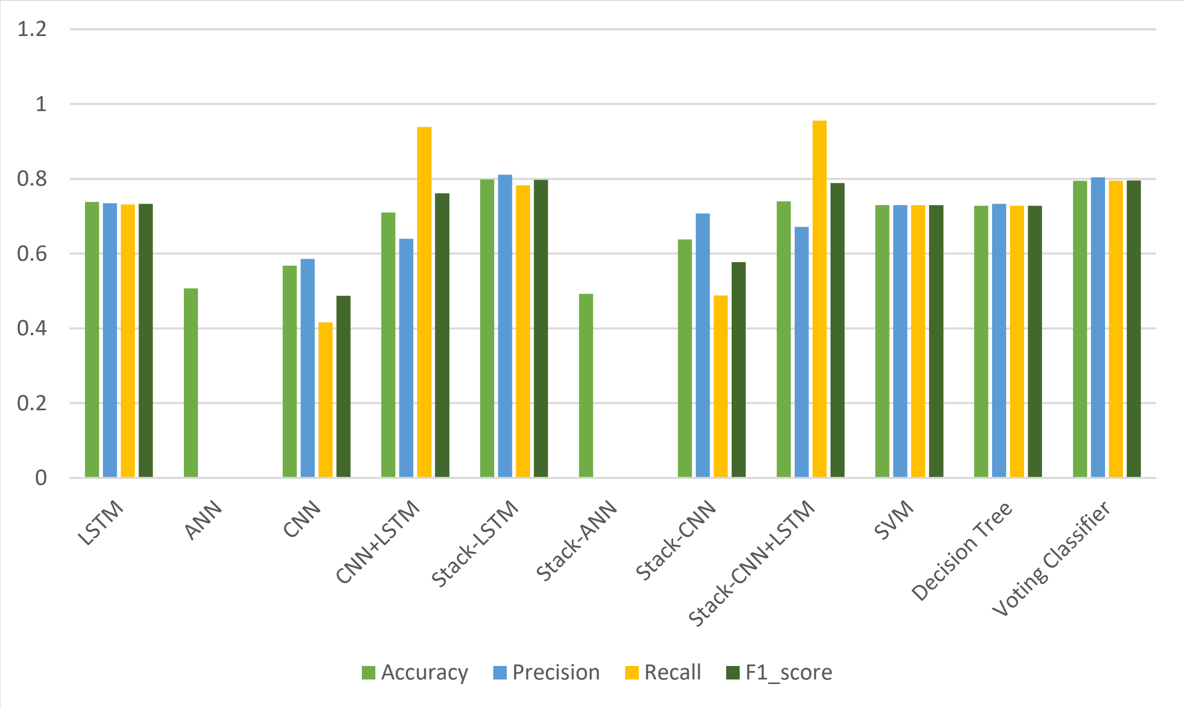
Table 6: Performance Evaluation Metrics – Data 2

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**

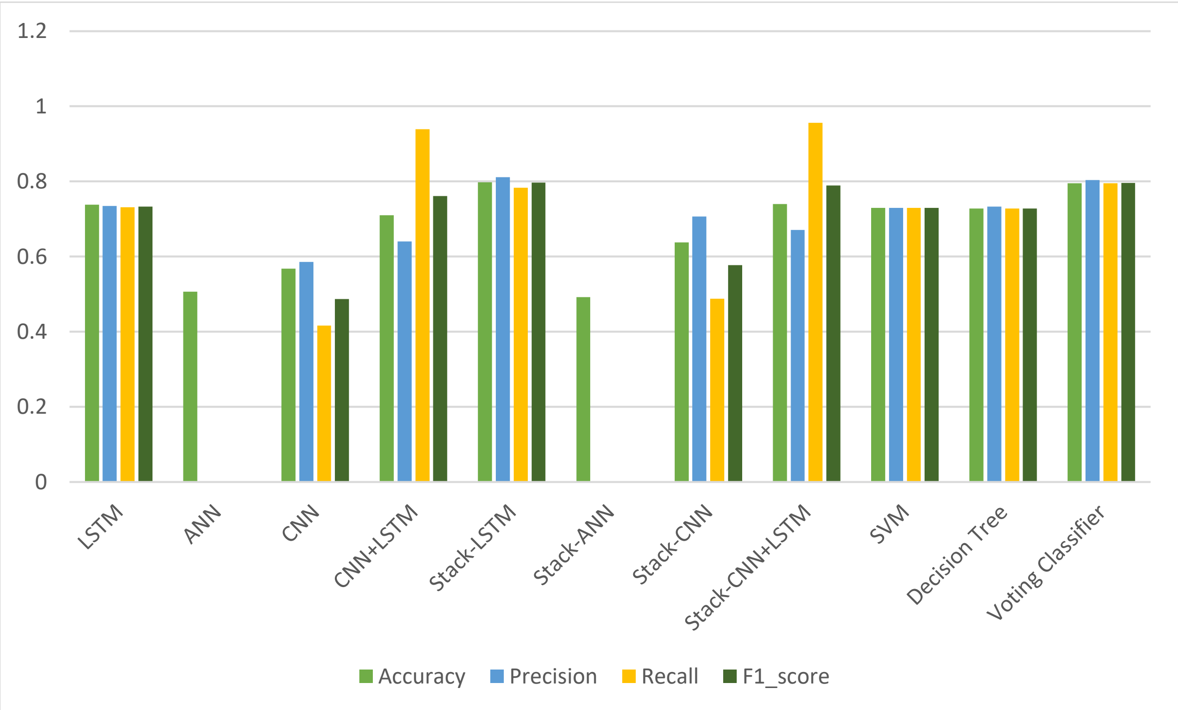
Table 7: Performance Evaluation Metrics – Data 3

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

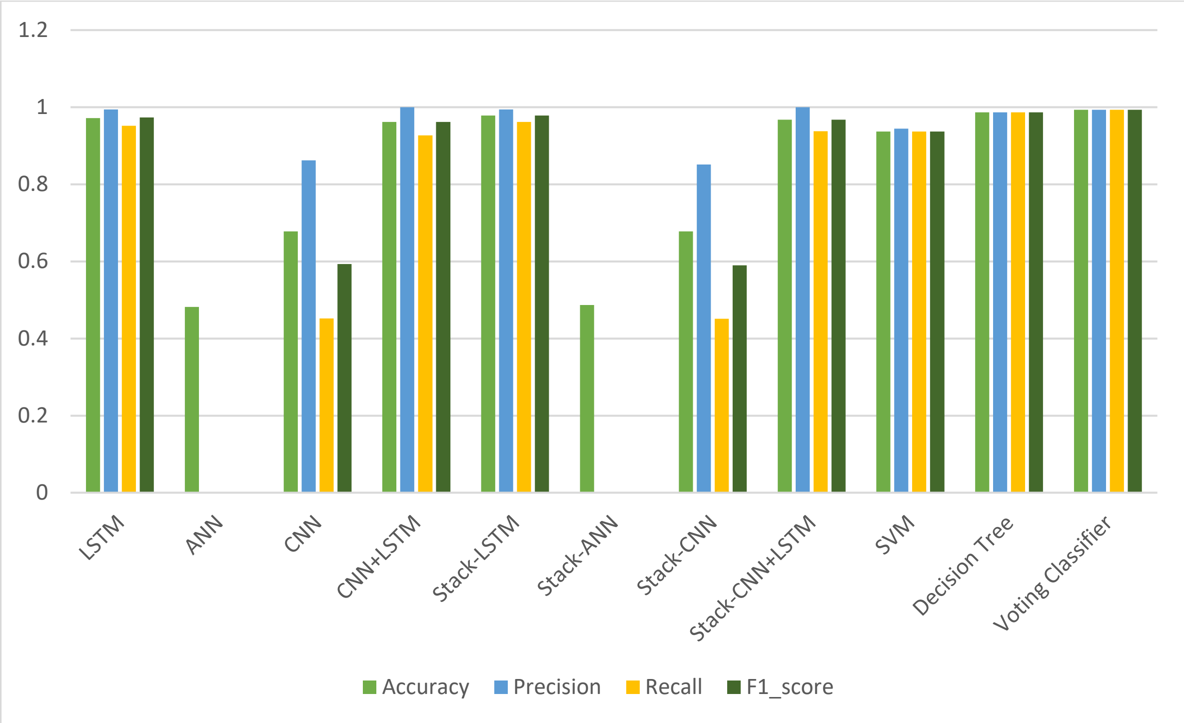
Stack-ANN	0.487	0.000	0.000	0.000
Stack-CNN	0.678	0.851	0.451	0.590
Stack-CNN+LSTM	0.968	1.000	0.938	0.968
SVM	0.937	0.944	0.937	0.937
Decision Tree	0.987	0.987	0.987	0.987
Voting Classifier	**0.993**	**0.993**	**0.993**	**0.993**



Graph 1: Comparison Graphs – Data 1



Graph 2: Comparison Graphs – Data 2



Graph 3: Comparison Graphs – Data 3

Graphs 1, 2, & 3 illustrate accuracy in light green, precision in blue, recall in light yellow, & the F1 score in green. The voting classifier surpasses the other algorithms across all metrics, exhibiting the best values relative towards the other models. The aforementioned graphs illustrate these subtleties visually.

6. CONCLUSION :

The assessment of diverse ML methodologies for diabetes prediction has yielded encouraging outcomes utilizing three separate datasets: PIMA Indian Diabetes (PIMA-ID-I) data set, Diabetes Data Fresh from "Frankfurt Hospital, Germany (DDFH-G) & Diabetes (IDPD-I) data set".The datasets, each possessing awesome characteristics, were utilized towards evaluate the usefulness of various algorithms, illustrating the effectiveness of feature selection & ensemble techniques. Extrree FS has shown significant costs in improving the performance of the model identification of the most suitable features necessary towards maintain the accuracy & efficiency of diabetes prediction. The file method using the voting classifier that integrates the enhanced decision tree, random forest & packed extra trees, has achieved increased predictive accuracy in all three data sets.This method leverages the advantages of individual models, guaranteeing robustness & dependability in forecasts. The results highlight the efficacy of ML methods in enhancing early diabetes identification, establishing a robust basis for healthcare programs focused on early intervention.

The ***future scope***the aim of this work is towards increase the accuracy of prediction through exploring sophisticated ML & DL methodologies, including reinforcement learning & hybrid models, towards improve the early identification of diabetes. Furthermore, using a broader range of diverse & real-time datasets, encompassing genetic & lifestyle variables, may augment model robustness. Utilizing explainable AI (XAI) methodologies will enhance transparency & trust in forecasts, facilitating improved decision-making in clinical environments & refining individualized treatment strategies for patients.

REFERENCES :

[1] Binhowemel, S., Alfakhri, M., AlReshaid, K., & Alyani, A. (2023). Role of artificial intelligence in diabetes research diagnosis & prognosis: A narrative review. *J. Health Inform. Developing Countries*, 17(2), 1–12.

[2] Viswanath, G. (2021). Hybrid encryption framework for securing big data storage in multi-cloud environment. *Evolutionary intelligence*, 14(2), 691-698.

[3] Esmaeili, M., Goki, S. H., Masjidi, B. H. K., Sameh, M., Gharagozlou, H., & Mohammed, A. S. (2022). ML-DDoSnet: IoT intrusion detection based on denial-of-service attacks using machine learning methods and NSL-KDD. *Wireless Commun. Mobile Comput.*, 2022(1), 1-16.

[4] Viswanath, G. (2023). A Real-Time Case Scenario Based On URL Phishing Detection Through Login URLS. *Material science and technology,* 22(9), 103-108.

[5] Jaiswal, V., Negi, A., & Pal, T. (2021). A review on current advances in machine learning based diabetes prediction. *Primary Care Diabetes,* 15(3), 435–443.

[6] Viswanath, G., & Sunil Kumar Reddy, T. (2014). Enhancing power unbiased cooperative media access control protocol in manets. *International Journal of Engineering Inventions*, 4(9), 8-12.

[7] Sivaranjani, S., Ananya, S., Aravinth, J., & Karthika, K. (2021). Diabetes prediction using machine learning algorithms among feature selection & dimensionality reduction. *7th Int. Conf. Adv. Com put. Commun. Syst. (ICACCS),* 1(1), 141–146.

[8] Viswanath, G. (2024). Artificial Intelligence-driven Frameworks for Fostering Active Participation & Learning in Language Classrooms. *in International Journal of Interpreting Enigma Engineers (IJIEE)*, 1(3), 23-32.

[9] Tong, H. L., Ng, H., & Arul Ananthan, H. (2024). Predicting diabetes mellitus among machine learning techniques. *J. Eng. Technol. Appl. Phys.,* 6(1), 91–99.

[10] Viswanath, G. (2024). Improved Light GBM Model Performance Analysis & Comparison For Coronary Heart Disease Prediction. *International Journal of Information Technology & Computer Engineering,* 12(3), 658-672.

[11] Zou, L., Luo, X., Zhang, Y., Yang, X., & Wang, X. (2023). HC-DTTSVM: A network intrusion detection method based on decision tree twin support vector machine and hierarchical clustering. *IEEE Access*, 11(1), 21404–21416.

[12] Viswanath, G. (2024). Hybrid Feature Extraction among Machine Learning towards Identify Network Attacks. *International Journal of HRM & Organizational Behavior,* 12(3), 217-228.

[13] Emblem, K. E., Zoellner, F. G., Tennoe, B., Nedregaard, B., Nome, T., Due-Tonnessen, P., Hald, J. K., Scheie, D., & Bjornerud, A. (2008). Predictive modeling in glioma grading from MR perfusion images using support vec tor machines. *Magn. Reson. Med.*, 60(4), 945–952.

[14] Viswanath, G. (2023). A Real Time Online Food Ording Application Based Django Restfull Framework. *Juni Khyat*, 13(9), 154-162

[15] Cho, H., Lee, S., Kim, J., & Park, H. (2018).Classification of the glioma grading using radiomics analysis. *PeerJ*, 6(1), e5982.

[16] Viswanath, G. (2024). Artificial Intelligence-driven Frameworks for Fostering Active Participation and Learning in Language Classrooms. *International Journal of Interpreting Enigma Engineers (IJIEE)*, 1(3), 23-32.

[17] Gupta, M., Rajagopalan, V., & Rao, P. (2019). Glioma grade classification using wavelet transform-local binary pattern based statistical texture features and geometric measures extracted from MRI. *J. Experim. Theor. Artif. Intell.*, 31(1), 57–76.

[18] Hsieh, K. L. C., Lo, C. M., & Hsiao, C. J. (2017). Computer-aided grading of gliomas based on local and global MRI features. *Comput. Methods Programs Biomed.*, 139(3), 31–38.

[19] Wu, Y., Liu, B., Wu, W., Lin, Y., Yang, C., & Wang, M. (2018). Grading glioma by radiomics with feature selection based on mutual information. *J. Ambient Intell. Humanized Comput.*, 9(5), 1671–1682.

[20] Bui, H. T., Hamamoto, K., & Paing, M. P. (2022). Tooth localization using YOLOv3 for dental diagnosis on panoramic radiographs. *IEEJ Trans. Electron., Inf. Syst.*, 142(5), 557–562.

[21] Reshan, M. S. A., Amin, S., Zeb, M. A., Sulaiman, A., Alshahrani, H., & Shaikh, A. (2023). A robust heart disease prediction system using hybrid deep neural networks. *IEEE Access,* 11(1), 121574–121591.

[22] Jothi Prakash, V., & Karthikeyan, N. K. (2021). Enhanced evolutionary feature selection and ensemble method for cardiovascular disease prediction. *Interdiscipl. Sci., Comput. Life Sci.,* 13(3), 389–412.

[23] Viswanath, G. (2024). Personalized Breast Cancer Prognosis through Data Mining Innovations. *Cuestiones de Fisioterapia*, 53(2), 538-548.

[24] Khouloud, S., Ahlem, M., Fadel, T., & Amel, S. (2022). W-Net and inception residual network for skin lesion segmentation and classification. *Appl. Intell.,* 52(4), 3976–3994.

[25] Viswanath, G. (2021). Adaptive Light Weight Encryption Algorithm for Securing Multi-Cloud Storage. *Turkish Journal of Computer and Mathematics Education*, 12(9), 545-554.

[26] Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. *Proc. Comput. Sci.,* 165(2), 292–299.

[27] G. Viswanath, G. (2023). A Real -Time Video Based Vehicle Classification Detection & Counting System. *Industrial Engineering Journal,* 52(9), 474-480.

[28] Chaki, J., Ganesh, S. T., Cidham, S. K., & Theertan, S. A. (2022). Machine learning & artificial intelligence based diabetes mellitus detection & self-management: A systematic review. *J. King Saud Univ. Comput. Inf. Sci.,* 34(6), 3204–3225.

[29] Viswanath, G. (2025). Proactive Security in Multi-Cloud Environments: A Blockchain Integrated Real-Time Anomaly Detection & Mitigation Framework. *Cuestiones de Fisioterapia,* 54(2), 392-417.

[30] Jiang, H., Lin, Y., Ren, W., Fang, Z., Liu, Y., Tan, X., Lv, X., & Zhang, N. (2011). Adverse drug reactions and correlations with drug–drug interactions: A retrospective study of reports from 2011 to 2020. *Frontiers Pharmacol*, 13(2), 1- 14.

[31] Viswanath, G. (2024). International Journal of Information Technology & Computer Engineering. *International Journal of Interpreting Enigma Engineers (IJIEE),* 12(3), 647-657.

[32] Deepika, D., & Balaji, N. (2022). Effective heart disease prediction with greywolf with firefly algorithm-differential evolution (GF-DE) for feature selection and weighted ANN classification. *Comput. Methods Biomechanics Biomed. Eng.,* 25(12), 1409–1427.

[33] Viswanath, G. (2024). Machine Learning for IoT Device Anomaly Detection Attack Classification. *International Journal of Mechanical Engineering Research and Technology*, 16(9), 66-76.

[34] Yang, Z., & Emmert-Streib, F. (2023). Threshold-learned CNN for multi-label text classification of electronic health records. *IEEE Access*, 11(2), 93402–93419.

[35] Viswanath, G. (2024). A Hybrid Particle Swarm Optimization & C4.5 For Network Intrusion Detection & Prevention System. *International Journal of Computing This link is disabled*, 23(1), 109-115.

[36] Chattaraj, D., Bera, D., Das, A. K., Saha, S., Lorenz, P., & Park, Y. (2021). Block-CLAP: Block chain-assisted certificate less key agreement protocol for Internet of Vehicles in smart transportation. *IEEE Trans. Veh. Technol.*, 70(8), 8092-8107.

[37] Viswanath, G. (2014). Distributed Utility-Based Energy Efficient Cooperative Medium Access Control in MANETS. *International Journal of Engineering Inventions,* 4(2), 08-12.

****