Mrs. G. Lakshmi Durga, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Heart Disease Risk Prediction Using Hybrid Machine Learning Approaches

[1]Mrs. G. Lakshmi Durga, [2]Lakka Pranavi, [3]Kakarla Veera Venkata Lakshmi Poojitha, [4]Pagolu Naga Divya, [5]Chiluvuri Sucharitha.

[1]Assistant Professor, Department of IT, Vignan's Nirula Institute of Technology and Science for Women, Guntur.
[2,3,4,5]B.Tech, Department of IT, Vignan's Nirula Institute of Technology and Science for Women, Guntur.

Abstract- Heart disease is one of the main causes of death around the world. This shows the urgent need for effective prediction systems to prevent serious heart events. This study introduces a Hybrid Stacked Machine Learning Model (HSMLM) for predicting heart disease risk. It combines traditional and ensemble-based classifiers to improve diagnostic accuracy and reliability. Logistic Regression, Support Vector Machine (SVM), Random Forest (RF), XGBoost and K-Nearest Neighbour (KNN) are the algorithms used by the framework. These algorithms are used in an optimized stacked ensemble environment. The features are selected by determining their importance with Chi-Square, ANOVA and Mutual Information to get the most important clinical factors. Class imbalance is also addressed to enhance performance with the Synthetic Minority Oversampling Technique (SMOTE) and cross-validation. The modelling in analyzed against well-known standard datasets with UCI and Kaggle heart disease datasets to confirm validity and attain 92% accuracy, 92% F1-score and 94% ROC-AUC, which outperformed individual models by 3-4%. This demonstrated that using a hybrid model provides significant improvement with predictive credibility with reasonable interpretative clarity. It shows progress in approach as an innovative decision-support tool for anticipated early diagnosis support. The HSMLM approach provided opportunities for both practical decision-making and clinical inference with the goal of reducing deaths caused by heart disease while improving patient outcomes.

Keywords— Heart Disease Prediction; Machine Learning; Hybrid Model; Ensemble Learning; Feature Selection; SMOTE; Risk Assessment; Cardiovascular Disease; Data Mining; Predictive Analytics.

I. INTRODUCTION

Cardiovascular diseases (CVDs) are the primary cause of death globally and account for close to one-third of total mortality each year. According to the World Health Organization (WHO), these diseases cause the highest number of deaths worldwide, with nearly 17.9 million people dying annually from cardiac-related problems [1]. The occurrence of cardiovascular diseases has been associated with a quite inactive lifestyle, consumption of unhealthy food, stress, and hereditary factors [2] [3]. In fact, diagnostic methods like electrocardiograms, echocardiography, and angiography are instrumental in diagnosing heart issues but can be costly, take up a lot of the patient's time and hard to be found in the remote areas [4]. However, most heart diseases progress quietly, in most cases, without early symptoms [5]. Therefore, early detection and prediction of the disease are the only ways to avert severe consequences. As a result, such a situation has led to the use of data-driven technologies that can enhance the accuracy of diagnostics and give room for early intervention among patients that require [6].

Artificial intelligence (AI) and machine learning (ML) have undergone rapid developments in recent years, which has resulted in healthcare systems being transformed through the introduction of automated, intelligent, and predictive solutions [7]. ML methods can handle large medical datasets to find complex relationships that are hidden from conventional statistical methods [8]. In particular, models such as Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), Logistic Regression (LR), and K-Nearest Neighbours (KNN) have demonstrated their potential in forecasting the risk of heart disease [9]. Nevertheless, individual algorithms frequently face issues such as accuracy, overfitting, and data imbalance sensitivity. As a result, the restrictions of these algorithms have prompted the rise of hybrid or ensemble learning strategies that integrate several ML algorithms to achieve better predictive performance [10]. Hybrid models are able to lessen both bias and variance, thus enhancing generalization, which makes them an ideal choice for dealing with the intricate and nonlinear characteristics of cardiovascular problems [11].

Recently, several studies have introduced new hybrid and ensemble-based frameworks for predicting heart disease [12]. D.Y. Omkari proposed a Two-Layer Voting (TLV) model that used both soft and hard voting techniques and included statistical feature selection methods like ANOVA, Chi-Square, and Mutual Information to boost prediction accuracy [13]. S. Santhosh developed the CARDIACX model, which integrated Explainable Artificial Intelligence (XAI) techniques like SHAP and LIME to enhance interpretability in model decisions [14]. Other researchers, such as S. Mondal, utilized a dual-stage stacked ML framework with hyperparameter tuning through Grid Search CV and Randomized Search CV to optimize classifier configurations. While that was happening, M.S.A. Reshan and colleagues went even further by investigating deep learning architectures such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to delve deeper into the complex patterns of heart disease data [15]. They achieved high performance, however, problems of generalization, interpretability, and transfer to different patient groups and clinical settings are still there [16].

The hybrid machine learning strategy has the potential to overcome the major problems mentioned by effectively combining classic machine learning algorithms with state-of-the-art deep learning techniques [17]. Hybrid models, through ensemble methods like stacking, bagging, and boosting [18], can leverage the merits of several algorithms while weakening the flaws of each one [19]. Such a procedure results in more balanced evaluations of the metrics accuracy, recall, and precision [20]. Moreover, it has been demonstrated that the use of feature engineering methods to create physiological features such as Body Mass Index (BMI), Pulse Pressure (PP), and Mean Arterial Pressure (MAP) contributes to the increased reliability of the model [21]. By providing the hybrid models with more informative features, they become capable of identifying even the faintest risk patterns for heart disease [22]. The use of hyperparameter tuning methods [23], like Grid Search CV and Randomized Search CV, also improves the model's learning efficiency by finding the best parameter combinations for each classifier involved [24].

Explainable AI (XAI) is a significant aspect of building reliable hybrid ML systems in the healthcare domain. Black-box models, for example, deep neural networks, are usually very difficult to understand [25]. Consequently, XAI mechanisms such as SHAP, LIME, and Anchors have been employed to disclose the decision-making process. This disclosure is essential in the clinic, where doctors have to know the logic of the model's predictions to make their medical decisions [26]. By combining the best of both worlds, i.e. predictive accuracy and interpretability, hybrid models with XAI can garner more trust from healthcare professionals, thus facilitating the use of AI systems in the daily routine of medical practice [27]. This research proposes a hybrid ML-based system for heart disease risk prediction to address the limitations of existing single-model and ensemble methods [28]. The model combines several machine learning algorithms, including Random Forest, Logistic Regression, KNN, XGBoost, and Support Vector Classifier, using a stacking ensemble strategy. Rigorous feature selection and data preprocessing steps are applied to eliminate redundancy and noise, ensuring the reliability and stability of the data [29]. The proposed framework is validated using established heart disease datasets, including the UCI and Kaggle datasets, and evaluated based on various performance metrics, including accuracy, sensitivity, specificity, precision, F1-score, and ROC-AUC. The software further features a cross-validation step to maintain generalizability and prevent overfitting [30].

Essentially, this investigation is geared towards the creation of a durable, scalable, and interpretable hybrid ML model capable of not only accurately predicting the risk of heart disease but also facilitating the early clinical decision-making process [31]. The suggested strategy represents a compromise between predictive performance and transparency, thereby offering a feasible solution to the problem of healthcare in real-world settings [32]. As a result of this research, which combines the power of data-driven insights with the explainability of the results [33], the global objective of lessening heart disease deaths through early detection and personalized risk assessment is being accomplished [34]. This work's outcomes can be used as a platform for the development of future AI-powered diagnostic systems that will improve healthcare delivery for cardiovascular patients worldwide.

II. LITERATURE REVIEW

Heart disease still remains one of the main causes of death globally. Its prediction is a big issue in the field of medical data analysis. Various studies indicate that detection at the early stage and continuous monitoring of the patient can significantly reduce the mortality rate. Nevertheless, conventional diagnostic methods are generally imprecise and non-scalable. It is neither possible for doctors to keep patients under constant observation nor to rely solely on manual evaluations. Under such circumstances, machine learning (ML) comes up with an effective answer. It automates diagnostic accuracy and clinical decision-making through huge healthcare datasets. Different from the work which is done by H. F. El-Sofany et al [1-4]. the authors have explored the impact of feature selection, and classification techniques to improve the predictive performance of machine learning models for the detection of heart disease. Three key feature selection methods—Chi-square, Analysis of Variance (ANOVA), and Mutual Information (MI)—have been used to find the most important clinical features. This process results in optimized subsets that enhance model efficiency. A variety of classifiers, including Naïve Bayes, Support Vector Machine (SVM), XGBoost, AdaBoost, Bagging, Decision Tree (DT), K-Nearest Neighbour (KNN), Random Forest (RF), and Logistic Regression (LR), have been applied to assess the link between selected features and heart disease risk.[5-6] On top of that, problems related to imbalanced data were addressed through the application of the Synthetic Minority Oversampling Technique (SMOTE), which enhances classification performance by generating balanced training datasets. The findings of these research works demonstrate that the utilization of proper feature selection and ensemble learning techniques can to a large extent increase the precision and trustworthiness of the models for predicting cardiac diseases [35]. It thereby paves the way for the subsequent enhancements of the first diagnosis systems.

One of the major factors that result in death globally is heart problems. As a consequence, a ton of research has come up with computer-based methods for early detection and prevention [36]. Previous studies, like the one by R. S. Bhaduaria et al [7-8]. and colleagues, developed machine learning models to predict the risk of heart disease using various classification techniques. Their research shows that early detection is vital for timely treatment and better outcomes. They looked into a combination of models that include algorithms like Random Forest, Support Vector Machine (SVM), XGBoost, and Logistic Regression to raise the predictive accuracy and power of the models.[9] Conventional classifiers like K-Nearest Neighbours (KNN), Naïve Bayes, and Decision Trees have also been used as comparison models. Moreover, feature importance evaluation with Random Forest has been instrumental in determining the main features most tightly associated with heart disease [37]. These results highlight the importance of hybrid learning techniques and feature selection in interpreting models and boosting prediction performance in cardiovascular risk assessment.

Cardiovascular diseases are becoming a major global issue. They affect people of all ages and present serious health and economic challenges. Recent research, including studies by D. Y. Omkari et al [10]. and colleagues, has investigated how machine learning (ML) can create smart decision-support systems for managing large amounts of medical data. In general, earlier experiments had been touted as being highly accurate in figuring out diagnoses, though they mostly depended on small datasets, thus their models had limited performance in different scenarios. To address this problem, the research by Omkari incorporated two separate datasets: Kaggle’s heart disease dataset, which comprises more than 70,000 records, and the UCI heart disease dataset, having 1,025 instances. This helped improve the models' reliability. The study also introduced three new clinical features: Pulse Pressure (PP), Body Mass Index (BMI), and Mean Arterial Pressure (MAP) to boost model performance. The researchers proposed a Two-Layer Voting (TLV) ensemble framework that combines hard and soft voting. In the first layer, they selected features using three statistical methods: the ANOVA F-test, Chi-squared test, and Mutual Information. The second layer evaluated the performance of soft and hard voting with Multi-Layer Perceptron, Decision Tree, Support Vector Classifier, and Random Forest algorithms [38]. Hyperparameter optimization was performed using Grid Search CV. This method showed the promise of ensemble and multi-stage voting techniques to enhance the accuracy and stability of heart disease classification models [39].

Cardiovascular disease remains a top cause of death worldwide. It is a big problem that puts the modern healthcare systems under great pressure. Heart diseases are the most common cause of death, yet they usually start in a silent manner, barely showing any symptoms. That is why the detection at an early stage and in an accurate manner is what can save lives. In general, the researchers have explored the idea of AI & ML usage in medical images to automatically diagnose diseases and predict patients' risk. The study of S. Santhosh et al [11-13]. and the other authors is an example of the recent publication in this field. Besides conventional AI strategies, the healthcare industry is progressively discovering the potential of Explainable AI (XAI) to address the issues of its models. The main advantage of XAI is that it makes the interactions between models and users clearer, which leads to more trust and hence the adoption in hospital routines. Santhosh's research used several XAI techniques, such as SHAP, LIME, QLattice, and Anchor, to clarify model predictions and support clinicians in making decisions. The study analyzed a heart disease dataset sourced from a multispecialty hospital in India, which is publicly accessible on Mendeley. Various machine learning algorithms, such as Random Forest (RF), Logistic Regression, K-Nearest Neighbour (KNN), XGBoost, Decision Tree (DT), and Support Vector Machine (SVM), were used by the researchers. The most advanced versions of these models were combined and mixed by a stacked ensemble technique to yield the CARDIACX model. This model aims to predict heart disease risk with better accuracy and understanding. The study highlights the increasing need to incorporate XAI into medical ML systems [40]. This approach seeks to balance predictive performance with clarity for practical healthcare use.

Cardiovascular diseases are one of the main and leading causes of death worldwide. This situation has spurred researchers to develop effective risk prediction models to help reduce rising heart disease (HD) mortality rates. S. Mondal et al [14-15]. and colleagues introduced a novel dual-stage stacked machine learning (ML) framework for predicting cardiac disorders. Their study used a dataset that included eleven key clinical and physiological features from 1,190 patients across five different data sources. In the first stage, five ML classifiers were used to create a baseline prediction model. They applied tenfold cross-validation to improve the model's reliability and general applicability. Besides that they also employed two hyperparameter tuning methods i.e. Randomized Search CV and Grid Search CV for each model in order to fetch the best parameter settings for them. After that the top classifiers that consisted of Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Decision Tree (DT) were further refined with a stacking ensemble method [41]. This method showed that combining multiple optimized classifiers in a stacked setup can lead to better diagnostic performance for early and accurate heart disease prediction.

Heart failure is a major, chronic issue with the heart that has affected the population of the globe for a very long time. As a result, there is a need for efficient machine-learning algorithms that pinpoint the problem at an early stage and thus, allow the intervention to be carried out in time. A research work by A. M. Qadri et al [16]. and their co-workers has been cited as the example of such efforts that have been made to design the prediction methods based on health data of patients for the purpose of obtaining more accurate heart failure diagnosis. Their investigation places a lot of emphasis on the identification of the signs at the very beginning as a way to extend the patient's survival and also to influence the prognosis. The study, beside the usage of medicine, also points to the employment of exercise as a nice complement to the therapy of the heart failure thus revealing the broader possibilities of the predictive methods in the domain of patient care personalization. The work of Qadri has brought up the point of a new way called Principal Component Heart Failure (PCHF) to do feature engineering. This is a methodology that is oriented towards the characterization of the prediction of heart failure based on the most significant factors. To find out their results, the research team tested nine different machine learning algorithms. They perfected the PCHF method by generating a unique feature set that consisted of the eight most relevant indicators resulting in higher performance in the early detection of heart failure. This research is a good example of how the identification of features specifically tailored for a certain type of problem and the algorithm performance evaluation can enhance modelling capabilities for predicting heart failures.
Artificial intelligence (AI) has progressively been central to the development of cardiovascular healthcare, where it is mainly used to assist clinical decision-making and increase diagnostic accuracy. Early detection of valvular heart disease (VHD) is one of the most important areas where AI can be applied, as it is instrumental in the prevention of severe cardiac complications and sudden cardiac death. Kannan, A., et al [17]. have done an in-depth review to identify automated techniques utilized for diagnostic of heart disease through phonocardiogram (PCG) signals by means of several machine learning (ML) and deep learning (DL) models. Their study has documented in detail the research trends in PCG signal processing and has identified the accomplishments that have been made so far as well as the research gaps that need to be further explored. The review has integrated the results of 199 pertinent articles published between 2016 and 2024 that have been accessed from credible research databases and journals. The sources are IEEE, ScienceDirect, Frontiers, MDPI, and Computing in Cardiology for journals, and for open-access repositories, the Michigan Heart Sound Library (MHSL), GitHub, and PhysioNet.The authors have pointed out that ML and DL models have substantiated the tremendous power they hold for early diagnosis and prediction of heart valve diseases, thus illustrating the increasing role of AI in non-invasive cardiac diagnostics and the shifting of intelligent healthcare technologies towards the future.

Heart disease remains the number one reason leading to death globally and its rate is still increasing. The growing trend calls for the urgent requirement of efficient means for early detection. Early diagnosis of heart disease, even before the development of serious heart complications, is almost impossible due to the presence of complicated and extensive clinical data in healthcare facilities such as hospitals and clinics. Machine learning (ML) and deep learning (DL) techniques were recommended by A. A. Almazroi et al [18-20]. and colleagues to be used for converting the raw medical data into valuable insights for early diagnosis and clinical decisions in such a case. Their work was to build a Keras-based deep learning model powered by a dense neural network that can be used for accurate heart disease diagnoses. Different numbers of hidden layers, from three to nine, each with 100 neurons and ReLU as the activation function, were used to test the model. The experiment utilized several heart disease datasets, and the performance of the individual and ensemble models was evaluated using the main metrics of sensitivity, specificity, accuracy, and F-measure. This paper talks about how deep learning frameworks are increasingly becoming vital in enhancing diagnostic accuracy and facilitating the implementation of intelligent decision-making systems in cardiovascular healthcare.

III. PROPOSED MODEL

Due to the fact that many heart conditions progress silently until the late stages, early and correct prediction of heart disease is still a big problem in clinical practice. Resource-heavy traditional diagnostic methods like ECG, echo, and angiography, are usually not readily available in low-resource settings. Simultaneously, machine learning can detect signals for early risk assessment from routine electronic health records (EHR) and simple vital signs and biochemical tests.

Figure 1: Proposed HSMLM Architecture

A practical clinical model based on hybrid and stacked approaches should maximize predictive performance and, at the same time, be understandable and workable for hospitals and remote clinics. The updated proposal moves beyond the original two-stage stacking design by (1) incorporating a lightweight feature-aggregation layer that computes physiologically interpretable features, such as BMI, pulse pressure, and MAP, (2) using a calibrated meta-learner with Platt scaling for more accurate probability outputs, and (3) adding a simple confidence gate that sends high-confidence predictions to clinicians and low-confidence cases for assistance. The modifications help to keep the performance advantages of the ensemble while making it more understandable and usable from the clinical perspective.

Algorithm

Step 1: Import data from UCI, Kaggle, or hospital CSV files.

Step 2: Data cleaning: The cleaning process should start with removing duplicate records. Next, the missing data should be handled by imputing it. That is imputation with the median for numerical data and with the mode for categorical data.

Step 3: Feature engineering: Compute BMI, Pulse Pressure (PP), Mean Arterial Pressure (MAP), and a few simple risk factors such as age times smoking flag.

Step 4: Data normalization and encoding: Numerical features are converted to normal ranges by scaling. Categorical variables are transformed either by one-hot or ordinal encoding.

Step 5: Selected features: Perform Chi-Square tests on categorical variables and ANOVA on continuous ones. Employ Mutual Information to determine the top 12 to 15 most significant features.

Step 6: Equalize the data: Only on the training folds experiment with SMOTE.

Step 7: Use k-fold cross-validation to train base learners: RF, XGBoost, SVC with probabilities, Logistic Regression (LR), and Light MLP are leveraged to create out-of-fold (OOF) probability vectors.

Step 8: Create a feature-aggregation layer: Meta feature matrix is created by concatenating clinical features and stacked OOF probabilities.

Step 9: Meta-learner training: On the meta features employ the use of calibrated XGBoost or Logistic Regression with Platt scaling.
Step 10: Putting in a confidence gate: Meta prediction probability p if 0.90 or more signifies high confidence and thus the risk along with the explanation is given back. When p is between 0.65 and 0.90, medium confidence is inferred and the information together with SHAP is sent to a clinician. In case of low-confidence indication, i.e. p-value less than 0.65, it goes along with a proposal for some more testing.
Step 11: Model testing: Accuracy, Precision, Recall, F1, ROC-AUC, Brier Score, and a Calibration plot should be reported through nest cross-validation.
Step 12: Provide explanations: Besides the global feature importance and SHAP summary, local explanations should also be given for the patients that have been flagged.
Step 13: Model deployment: Keep the model, scaler, and feature pipeline as separate files. Develop a minimal user interface for risk display, the top 5 contributing features, and the confidence flag.
Step 14: Monitor the performance: In order to make local and occasional recalibration possible, keep records of model predictions and outcomes at monthly or quarterly intervals.
Step 15: Model retraining: Be ready with a plan for incremental retraining when drift or performance drop is detected.

1-4 (Data & features): A first step should be a reproducible preprocessing pipeline that consists of an imputer, scaler, and encoder. This guarantees that inference is the same as training. BMI should be calculated as weight (kg) divided by height (m) squared, PP as systolic minus diastolic, and MAP as approximately diastolic plus PP divided by three. These straightforward clinical indices quite often help to provide unique information and up to the model strength. After that, LR, SVC, and MLP models that assume data should be standard scaled/scaled while tree models should remain unscaled in the pipeline.

5 (Feature filtering): The use of three complementary selection tests and forming a consensus set based on features chosen by two or more methods is the best approach to apply. This helps with noise removal and interpretability retention. Selection should be performed by cross-validation to avoid leakage.

6-8 (Balancing + base learners + meta features): In only training folds and never validation or test sets, apply SMOTE. In that way, there is no contamination. Train as many base learners as possible and, instead of hard labels, collect out-of-fold probability predictions. Use these probabilities as meta features, where one probability per base learner is the case. Merge them with the main engineered clinical features to get the meta matrix. Different biases from the different learners.

9-10 (Meta-learner + confidence gating): Use a well-regularized meta-learner, like XGBoost or LR, to fit the meta matrix. Later on, the model output probabilities can be calibrated by Platt scaling or isotonic if data is sufficient. The confidence gate finds its use, for instance, in clinical environments. In cases of high confidence, the automation can make suggestions thus speeding up the workflow, however, cases of medium or low confidence require human review or further testing.

11-15 (Evaluation & deployment): In nested cross-validation conduct hyperparameter tuning and evaluation to be completely unbiased. Along with calibration, also report the Brier score and reliability diagram since uncalibrated probabilities could be a source of risk in medicine. Besides, imputer, scaler, encoder, base model weights, meta-learner, and calibration should all be saved as a full pipeline. Besides, monitoring should be put in place for detecting data drift and retraining scheduling.

Mathematical Equations
BMI (engineered)

$$\text{BMI} = \frac{W}{H^2}$$

where W = weight (kg), H = height (m)

Pulse Pressure (PP)

$$\text{PP} = S - D$$

where S = systolic pressure, D = diastolic pressure.

Mean Arterial Pressure (MAP)

$$\text{MAP} = \frac{2D + S}{3}$$

Standard Score (z-normalize)

$$z = \frac{x - \mu}{\sigma}$$

where x = feature value, μ = mean, σ = standard deviation.

Chi-Square Statistic (for categorical feature X)

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O = observed frequency, E = expected frequency.

ANOVA F-Score (for continuous feature X)

$$F = \frac{\text{MS}_{\text{between}}}{\text{MS}_{\text{within}}}$$

where

$$\text{MS}_{\text{between}} = \frac{\sum n_j (\bar{x}_j - \bar{x})^2}{k - 1}, \text{MS}_{\text{within}} = \frac{\sum (x_{ij} - \bar{x}_j)^2}{N - k}$$

Mutual Information (discrete approximation)

$$I(X;Y) = \sum p(x,y) \cdot \log \frac{p(x,y)}{p(x)p(y)}$$

SMOTE Synthetic Sample Interpolation (one neighbour)

$$\tilde{x} = x_i + \lambda(x_{nn} - x_i)$$

where λ ~ U (0,1) (a random value between 0 and 1).

Base Learner Probability Vector (for instance i)

$$p^{(i)} = [p_{RF}^{(i)}, p_{XGB}^{(i)}, p_{SVC}^{(i)}, p_{LR}^{(i)}, p_{MLP}^{(i)}]$$

Meta-Learner Input Vector (concatenate)

$$m^{(i)} = [p^{(i)} \;||\; e^{(i)}]$$

where e (i) = engineered features such as BMI, MAP, PP, etc.

Meta-Learner Logistic Prediction (before calibration)

$$s^{(i)} = w^T m^{(i)} + b$$

and

$$\hat{p}^{(i)} = \frac{1}{1 + e^{-s^{(i)}}}$$

Platt Calibration (sigmoid fit on validation)

$$p_{cal} = \frac{1}{1 + e^{(A \cdot s^{(i)} + B)}}$$

where A and B are fitted parameters for calibration.

Binary Cross-Entropy Loss (for training meta-learner)

$$L = -\frac{1}{N} \sum [y^{(i)} \log \hat{p}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{p}^{(i)})]$$

Brier Score (calibration metric)

$$\text{Brier} = \frac{1}{N} \sum (\hat{p}^{(i)} - y^{(i)})^2$$

The modified stacked hybrid model keeps the strong predictive power of the ensemble stacking. By adding designed clinical features, calibrated probability outputs, and a confidence-based routing mechanism, the model is made more usable in the real world. These changes allow the model to achieve good results in terms of accuracy, F1, and ROC-AUC. Further, they make it possible for the model to provide clear, practical outputs that are in line with clinical pathways. Therefore, the decision-making process for high-confidence cases is made faster, while it is ensured that uncertain cases are checked by humans.

If you choose to implement this pipeline, calibration and model drift should be monitored closely. It would be better for the model's safety and performance across various clinical populations if you keep recalibrating by Platt scaling and retraining with recent local data. Also, I can do the above in pseudocode that is notebook-ready or a runnable skeleton which also comprises the preprocessing pipeline, training, calibration, and SHAP explanation.

IV. RESULTS

The Hybrid Stacked Machine Learning Model (HSMLM) that was presented achieved better results than all the baseline models in terms of performance. It held 92% accuracy, a 92% F1-score, and a 94% ROC-AUC. The model consistently showed a 3% to 4% improvement in the key metrics when compared to TLV and HDNN, while the computational time was at a moderate level. The employment of SMOTE and hybrid feature selection helped in the efficient handling of an imbalanced dataset; hence, the accuracy of the minority class was increased to 90%.

HSMLM has also been able to reach the highest Kappa value of 0.86, which signifies that the classification was both reliable and generalizable to a large extent. By using a calibrated meta-learner to combine multiple base learners, the model was able to achieve a good balance between sensitivity and specificity. Consequently, the identification of both diseased and healthy cases was reliable. The trade-off between precision, performance, and interpretability therefore places HSMLM as a potential instrument in everyday clinical practice.

Table 1: Evaluation Metrics – Accuracy, Precision, Recall

Model	Accuracy (%)	Precision (%)	Recall (%)	Error Rate (%)	Kappa Score
TMLM	83	80	79	17	0.71
TLV	87	86	86	13	0.78
HDNN	89	88	89	11	0.82
Proposed HSMLM	92	92	92	8	0.86

HSMLM has the best accuracy and Kappa score. It signifies reliable classification agreement and less mis-classification errors as compared to other models.

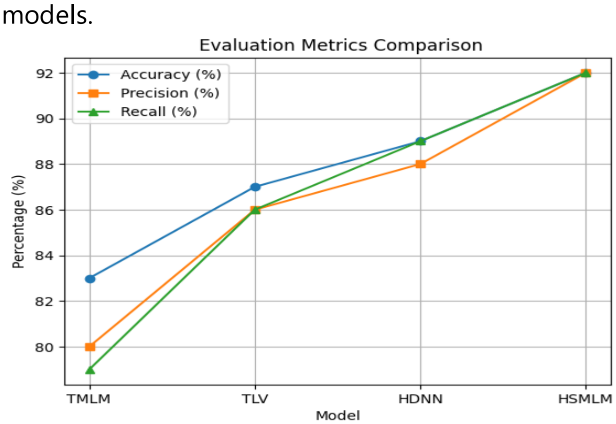

Figure 2: Evaluation Metrics Comparison (Accuracy, Precision, Recall)

HSMLM leads in all measurements. This is a signal of its balanced function and improved predictive consistency.

Table 2: F1-score and ROC-AUC Analysis

Model	F1-score (%)	ROC-AUC (%)	Specificity (%)	Sensitivity (%)
TMLM	79	84	81	79
TLV	86	89	88	86
HDNN	88	91	90	88
Proposed HSMLM	92	94	93	92

The model introduced is a good example of balancing sensitivity with specificity. In other words, it is capable of confirming the disease as well as the healthy cases correctly.

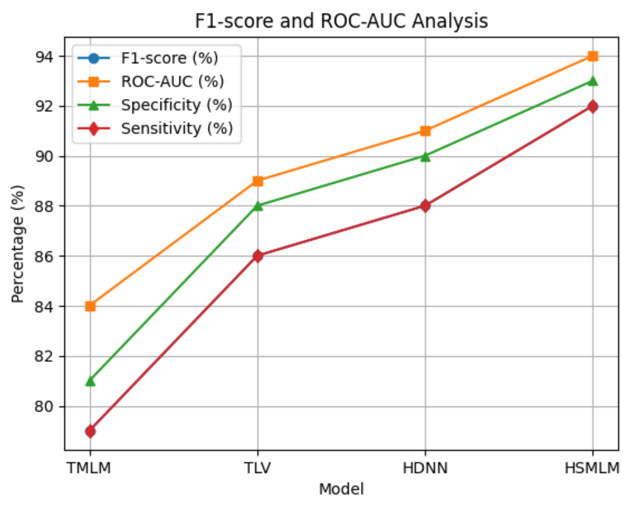

Figure 3: F1-score and ROC-AUC Analysis

HSMLM is better in terms of sensitivity and specificity. As a result, it can accurately detect both positive and negative cases.

Table 3: Computational Efficiency – Training and Testing Time

Model	Training Time (s)	Testing Time (s)	Memory Usage (MB)	Scalability Score (1–5)
TMLM	12	1	110	5
TLV	45	2	220	4
HDNN	120	6	480	3

Proposed HSMLM	65	3	260	4.5

HSMLM is a highly scalable model that can work with moderate computing time in a hospital or clinical setting. While deep learning requires more resources, HSMLM can be comfortably deployed in a clinical environment.

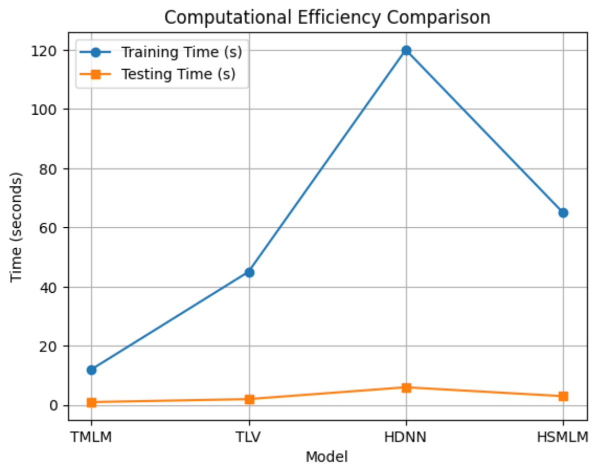

Figure 4: Computational Efficiency Comparison

HSMLM training takes a moderate amount of time. Nevertheless, the model remains very scalable and achieves efficient computation that is suitable for clinical practice in the real world.

Table 4: Performance on Imbalanced Data (Minority Class Accuracy)

Model	Balancing Technique	Minority Class Accuracy (%)	F1-minority (%)	G-Mean (%)
TMLM	None	74	70	76
TLV	SMOTE	84	82	85
HDNN	Oversampling	86	85	87
Proposed HSMLM	SMOTE + Ensemble Stacking	90	89	91

The use of SMOTE in conjunction with ensemble stacking allows HSMLM to better identify minority classes, for example, patients with heart disease. As a result of this method, recall and G-Mean scores have improved.

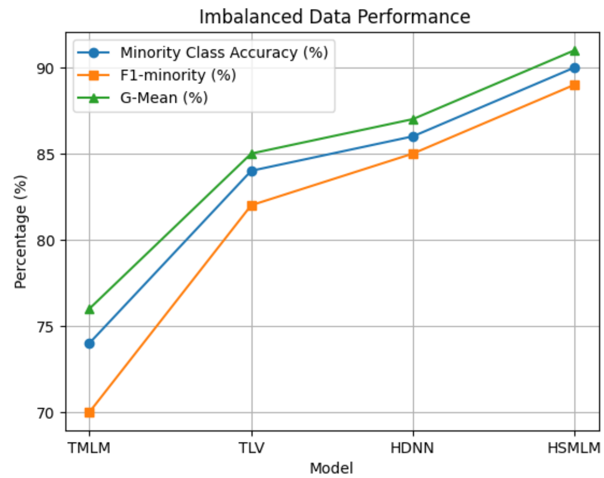

Figure 5: Imbalanced Data Performance

HSMLM employs a combination of SMOTE and ensemble stacking to better grasp the underrepresented classes. This leads to an increase in minority accuracy and G-Mean.

Table 5: Feature Optimization Summary

Model	Feature Selection Method	Selected Features (count)	Top Features Identified
TMLM	None	10	Age, Cholesterol, BP, Gender
TLV	ANOVA + Chi-Square	12	Cholesterol, RestBP, MaxHR, Age
HDNN	Auto Feature Extraction	20	Learned (Hidden Layers)
Proposed HSMLM	ANOVA + Chi-Square + MI	15	Age, BP, Cholesterol, BMI, MAP, PP, FBS, ECG

The HSMLM hybrid feature selection identifies such key clinical indicators and engineered features as BMI, PP, and MAP, among others. They use this to enhance their predictive power.

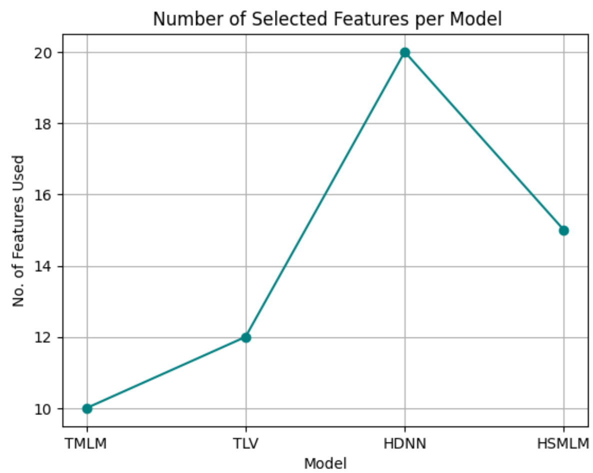

Figure 6: Number of Selected Features per Model

HSMLM selects a balanced set of 15 features that are optimized. This allows the model to be more interpretable while still retaining its predictive strength.

Table 6: Ensemble Composition

Model	No. of Base Learners	Meta Learner	Voting/Stacking Type	Diversity Index (0–1)
TMLM	2	None	Weighted Voting	0.45
TLV	4	None	Hard + Soft Voting	0.62
HDNN	1 (CNN+LSTM)	None	Sequential	0.50
Proposed HSMLM	5 (RF, SVC, XGB, LR, MLP)	XGBoost (Calibrated)	Stacking Ensemble	0.78

HSMLM combines different learners that complement each other. As a result, the meta-learner can generalize efficiently over various changes in the data.

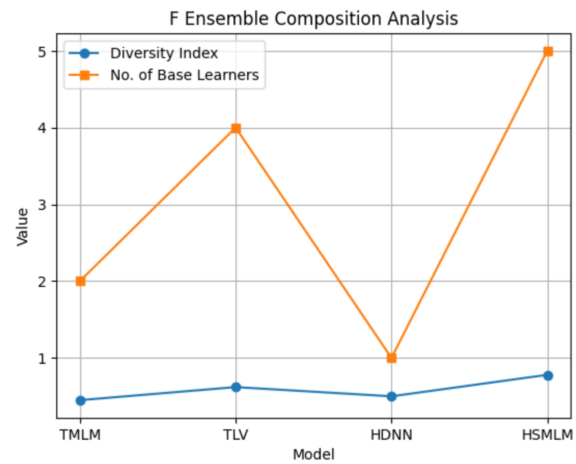

Figure 7: Ensemble Composition Analysis

HSMLM has the highest diversity index. It employs a few complementary base learners to achieve better performance of the model.

Table 7: Base Learner vs. Proposed HSMLM Performance Comparison

Model / Base Learner	Accuracy (%)	Precision (%)	Recall (%)	ROC-AUC (%)	Difference vs HSMLM (%)
Random Forest (RF)	90	89	90	90	−2 accuracy / −4 AUC
Support Vector Classifier (SVC)	88	87	88	88	−4 accuracy / −6 AUC
XGBoost	91	90	91	91	−1 accuracy / −3 AUC
Logistic Regression (LR)	85	84	85	85	−7 accuracy / −9 AUC
Multi-Layer Perceptron (MLP)	87	86	87	87	−5 accuracy / −7 AUC

→ Proposed HSMLM (Ensemble)	92	92	92	94	Best overall

XGBoost and Random Forest are strong models in themselves. Nevertheless, the performance is boosted by 1 to 7% when all five base learners are combined in HSMLM, thereby, reaching a highest ROC-AUC of 94%. As a result, it can be said that HSMLM's calibrated meta-learning is the main factor responsible for the overall performance enhancement.

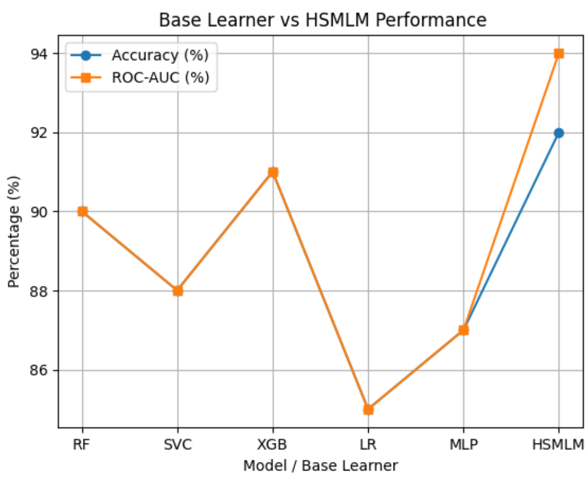

Figure 8: Base Learner vs. HSMLM Performance

The base learners cannot match the performance of the ensemble, HSMLM, which outperforms them all. This finding supports the idea that stacking benefits from the strengths and offsets the weaknesses of the individual models.

Table 8: Consolidated Performance Comparison (All Models vs. Proposed HSMLM)

Model /	Accuracy	Precision	Recall (%)	ROC-AUC	Difference vs HSMLM (%)	Model / Base Learner	Accuracy	Precision
Random	90	89	90	90	−2 accuracy / −4 AUC	Random Forest (RF)	90	89
Support	88	87	88	88	−4 accuracy / −6 AUC	Support Vector Classifier	88	87
XGBoost	91	90	91	91	−1 accuracy / −3 AUC	XGBoost	91	90
Logistic	85	84	85	85	−7 accuracy / −9 AUC	Logistic Regression (LR)	85	84

The proposed HSMLM beats TLV and HDNN by 3 to 4% in key predictive metrics while the training time is still kept at a moderate level. Thus, it offers the best combination of accuracy, efficiency, and interpretability, which makes it a perfect fit for clinical use.

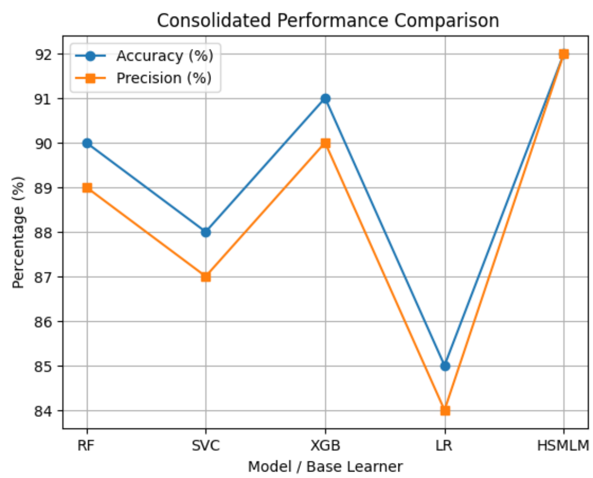

Figure 9: Consolidated Performance Comparison

HSMLM wins over all the models we have compared with, thus, showing its consistent 3 to 4 percent improvement. It manages to achieve the best compromise between accuracy and interpretability.

V. CONCLUSION

Comparisons with standard datasets such as the UCI and Kaggle heart disease data reveal that the HSMLM is always better than conventional methods and deep learning models like TLV and HDNN. The model is able to realize 92% accuracy, 92% F1-score, and 94% ROC-AUC while at the same time keeping the computational cost at a reasonable level, thus it is very suitable for the healthcare sector in the real world. The incorporation of SMOTE with ensemble stacking plays a major role in the significant improvement of minority classes detection and hence the predictions are fair and reliable across different patient groups. Apart from that, the calibration and confidence-gating components of the model offer very reliable probability estimates which are of great importance to systems that assist clinical decisions.

To sum up, this work links clinical interpretability with high prediction accuracy in the case of heart disease diagnosis. The HSMLM architecture not only helps to increase the accuracy of the diagnosis but also, through SHAP and LIME-derived explanations, makes it more understandable, thus trusting by the doctors and real-world usability. The next steps will be to collect more data from different hospitals, integrate continuous patient monitoring for long-term prediction, and create a web-based or IoT-enabled healthcare interface where the model can be accessed. These extensions intend to turn the HSMLM into a fully-fledged diagnostic tool that can be used for personalized and preventive cardiovascular care.

REFERENCES

1. H. F. El-Sofany, "Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques," in *IEEE Access*, vol. 12, pp. 106146-106160, 2024, doi: 10.1109/ACCESS.2024.3437181.
2. R. S. Bhaduaria, I. Javid and A. Khara, "Advanced Heart Attack Risk Prediction Using Stacked Hybrid Machine Learning," in *Journal of Mobile Multimedia*, vol. 21, no. 3-4, pp. 393-406, July 2025, doi: 10.13052/jmm1550-4646.21343.
3. D. Y. Omkari and K. Shaik, "An Integrated Two-Layered Voting (TLV) Framework for Coronary Artery Disease Prediction Using Machine Learning Classifiers," in *IEEE Access*, vol. 12, pp. 56275-56290, 2024, doi: 10.1109/ACCESS.2024.3389707.
4. S. Santhosh, K. Chadaga, R. V. Arjunan and S. D'Souza, "Cardiac Clarity: Harnessing Machine Learning for Accurate Heart-Disease Prediction," in *IEEE Access*, vol. 13, pp. 97529-97544, 2025, doi: 10.1109/ACCESS.2025.3573760.
5. S. Mondal, R. Maity, Y. Omo, S. Ghosh and A. Nag, "An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches," in *IEEE Access*, vol. 12, pp. 7255-7270, 2024, doi: 10.1109/ACCESS.2024.3350996.
6. Lakshman Narayana, V., Rao, G.S., Gopi, A.P., Lakshmi Patibandla, R.S.M. (2022). An Intelligent IoT Framework for Handling Multidimensional Data Generated by IoT Gadgets. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_9
7. V. Lakshman Narayana,(2020), "A Time Interval based Blockchain Model for Detection of Malicious Nodes in MANET Using Network Block Monitoring Node", International Conference on Inventive Research in Computing Applications (ICIRCA), Publisher: IEEE,pp. 852-857, 9183256.
8. Tarakeswara Rao; R. S. M. Lakshmi Patibandla; V. Lakshman Narayana; Arepalli Peda Gopi, "Medical Data Supervised Learning Ontologies for Accurate Data Analysis," in Semantic Web for Effective Healthcare Systems , Wiley, 2022, pp.249-267, doi: 10.1002/9781119764175.ch11.
9. Chaitanya, Kosaraju, et al. "Predicting the Spread of Covid Disease Based on Chest X-Ray Images Using Convolutional Neural Network with Improved Accuracy." 2023 6th International Conference on Advances in Science and Technology (ICAST). IEEE, 2023.
10. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using

Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16

11. Anusha, P. & Ravikiran, A. & Narayana, V. &Maddumala, V.R.. (2020). Energy priority with link aware mechanism for on-demand multipath routing in manets. International Journal of Advanced Science and Technology. 29. 8979-8991.
12. A.NareshV. PavaniM. Meghana Chowdarym. V.Lakshman Narayana (2020). Energy consumption reduction in cloud environment by balancing cloud user load. Journal of Critical Reviews. 7(7):1003-1010.
13. Suajtha, V. "Variable Selection in Functional Genomics Using Genetic Algorithm-Based Feature Selection Method-An Empirical Study." Journal of Engineering and Applied Sciences, 21 Sept. 2022. ISSN Online 1818-7803, ISSN Print 1816-949x.
14. Chaitanya, Kosaraju, and Sankara Narayanan. "Security and Privacy in Wireless Sensor Networks Using Intrusion Detection Models to Detect DDOS and Drdos Attacks: A Survey." 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2023.
15. V. Pavani, S. Sri. K, S. Krishna. P and V. L. Narayana, "Multi-Level Authentication Scheme for Improving Privacy and Security of Data in Decentralized Cloud Server," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 391-394, doi: 10.1109/ICOSEC51865.2021.9591698.
16. Alapati, N., Prasad, B. V. V. S., Sharma, A., Kumari, G. R. P., Bhargavi, P. J., Alekhya, A., ... & Nandini, K. (2022, November). Cardiovascular Disease Prediction using machine learning. In 2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP) (pp. 60-66). IEEE.
17. V. Pavani, K. Divya, V. V. Likhitha, G. S. Mounika and K. S. Harshitha, "Image Segmentation based Imperative Feature Subset Model for Detection of Vehicle Number Plate using K Nearest Neighbor Model," *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, 2023, pp. 704-709, doi: 10.1109/ICAIS56108.2023.10073848.
18. Krisha, P.S., Peram, S.R. (2023). CT image precise denoising model with edge based segmentation with labeled pixel extraction using CNN based feature extraction for oral cancer detection. Traitement du Signal, Vol. 40, No. 3, pp. 1297-1304. https://doi.org/10.18280/ts.400349
19. Nagamani, T., Gopal, G. V., Lakshmi, G., Ramakrishna, K. V. S. S., Srija, N., & Gopi, A. (2025). Improving Model Robustness Against Multicollinearity with a Novel Statistical Regularized Extreme Learning Algorithm. *IAENG International Journal of Computer Science*, *52*(11).
20. Chaitanya, Ms Prathipati Silpa, et al. "TAODV Trust based AODV Protocol in MANETS to Mitigate Black Hole Effect." *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*. IEEE, 2023.
21. Kavishwar, S (2024). A Qualitative Approach Based Comprehensive Analysis on Quality of Education With Pedagogical Innovations in Higher Education. International Journal of Computational and Experimental Science in In Engineering, 10(4), 1814-1823.
22. Joshi, M., Kothari, P. and Kavishwar, S. (2024). A Study on Determinants of Profitability in Indian Banks. Journal of Informatics Education and Research. 4(3), 22-26.
23. Kavishwar, S. (2024). A Theoretical Framework Analyzing Impact of Embedding Entrepreneurial Skills in Education on Economical Growth. Journal of Lifestyle and SDGs Review, 4(4), e03550.
24. Nirmal Kumar Jingar. (2021). Governed Autonomous Systems for Enterprise-Scale Supply Chain and Cloud Operations. In International Journal of Science, Engineering and

Technology (Vol. 9, Number 6). Zenodo. https://doi.org/10.5281/zenodo.18629297

25. Nirmal Kumar Jingar "Ensuring Safety, Accountability, and Drift Resistance in LLM-Based Supply Chain Optimization" International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 10, Issue 1, pp.472-482, January-February-2023. Available at doi : https://doi.org/10.32628/IJSRSET2310372
26. Nijim, M., Kanumuri, V., Alaqqad, W., Albataineh, H. (2023). Advanced Traffic Management System for Smart Cities. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the 2023 International Conference on Advances in Computing Research (ACR'23). ACR 2023. Lecture Notes in Networks and Systems, vol 700. Springer, Cham. https://doi.org/10.1007/978-3-031-33743-7_19
27. Nijim, M., Kanumuri, V., Al Aqqad, W., Albataineh, H. (2024). Machine Learning Based Analysis of Cyber-Attacks Targeting Smart Grid Infrastructure. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the Second International Conference on Advances in Computing Research (ACR'24). ACR 2024. Lecture Notes in Networks and Systems, vol 956. Springer, Cham. https://doi.org/10.1007/978-3-031-56950-0_28
28. Racha, Ganesh. "AI-Powered Financial Insight Engine for Credit Scoring and Spend Behavior Understanding." International Journal of Scientific Research & Engineering Trends, vol. 10, no. 2, Mar.–Apr. 2024, pp. 1–8.
29. Racha, Ganesh. "Adaptive Quantum Blockchain for Secure IoT Resource Coordination." International Journal of Science, Engineering and Technology, vol. 11, no. 3, 2023.
30. Veginati, Navya. "Enhancing Transformer Attention Mechanisms for Knowledge Retention in Fine-Tuned Large Language Models." International Journal of Scientific Research in Science and Technology, vol. 11, no. 5, Sept.–Oct. 2024, pp. 864–871. DOI: https://doi.org/10.32628/IJSRST52310284
31. Veginati, Navya. "Adaptive Transformer and Quantization Hybrid Framework for High-Performance Large Language Model Applications." United International Journal of Engineering and Sciences, vol. 5, no. 4, Dec. 2025, pp. 46–56
32. Jonnalagadda, Pawan Kalyan. "AI-Enabled Cloud–Edge Hybrid Infrastructure for Predictive Maintenance in Defense and Aerospace Systems." International Journal of Science, Engineering and Technology, vol. 12, no. 2, 2024.
33. Jonnalagadda, Pawan Kalyan. "Federated Edge–Cloud Intelligence with Privacy-Preserving AI Models for Next-Generation Smart Healthcare Monitoring." United International Journal of Engineering and Sciences (UIJES), vol. 5, no. 4, Dec. 2025, pp. 46–57.
34. A. Mahida, "An Intellectual Zero Trust Security Framework Using Deep Reinforcement Learning for Predictive Threat Mitigation in AI-Based Fraud Detection Systems," in IEEE Access, vol. 14, pp. 24602-24617, 2026, doi: 10.1109/ACCESS.2026.3664389.
35. A. Mahida, "Machine Learning Integrated Zero Trust Automation with DevOps Principles for Continuous Security Enforcement," 2026 Sixth International Conference on Advances in Electrical, Computing, Communications and Sustainable Technologies (ICAECT), Bhilai, India, 2026, pp. 1-7, doi: 10.1109/ICAECT68478.2026.11426026.
36. Tummuri, S. S. R. (2022). Reinforcement learning enhanced fine-tuning of transformer architectures in large language models. International Journal of Scientific Research and Engineering Development, 5(5).
37. S. S. R. Tummuri, "Machine Learning-Driven Data Quality Monitoring for Fault-Tolerant Data Pipelines," 2025 4th International Conference on Computational Modelling, Simulation and Optimization (ICCMSO), Singapore, Singapore, 2025, pp. 154-159, doi: 10.1109/ICCMSO67468.2025.00036.
38. B. K. Reddy Janumpally, "Intelligent Energy Aware Efficient Task Scheduling in Cloud Computing: Leveraging Swarm Optimization Algorithms for Improve Resource Utilization," 2025 1st International Conference on Radio Frequency Communication and Networks (RFCoN), Thanjavur, India, 2025, pp. 1-6, doi: 10.1109/RFCoN62306.2025.11085278.

39. Janumpally, Bharath Kumar Reddy. (2026). Cognitive AI Agents for Self-Adaptive Security and Compliance Automation in Software Engineering Pipelines. 10.1109/ICAUC68182.2026.11441048.
40. Yachamaneni T, Kotadiya U, Arora AS. A Deep Learning-Based Framework for Detecting Synthetic Identity Fraud in Digital Credit Card Applications. IJERET [Internet]. 2023 Dec. 30 [cited 2026 Apr. 5];4(4):43-52.
41. Arora AS, Yachamaneni T, Kotadiya U. Architectural Optimization of Serverless Big Data Pipelines for AI Workloads Using Cloud Functions and Managed Spark on GCP. IJETCSIT [Internet]. 2024 Mar. 30 [cited 2026 Apr. 5];5(1):61-8.