EJETMS
ISSN:2583-2131

Erudite Journal of Engineering Technology and Management Sciences

Vol. 6, No. 2, June 2026, pp. 25-29

Journal home page: http://www.ejetms.com

A Novel Multi-Stage Stacked Learning Framework for Cardiovascular Risk Stratification

M. E. Chandrasekar[1] and Dr. D. Nagaraju[2]

[1] PG Scholar, Dept. of CSE Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh, India

[2] Professor, Dept. of CSE Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh, India

Email: mechandra2000@gmail.com dubisettynagaraju@gmail.com

Received: 10 January 2026
Accepted: 02 April 2026

Keywords:
Stroke Detection, Adaptive Machine Learning, Neuroimaging, MRI, CTScans, Deep Learning, Decision-Support Systems, Early Diagnosis

ABSTRACT

Cardiovasculardiseases (CVDs) remain the leading cause of global mortality, necessitating accurate and early risk stratification to improve clinical outcomes. This paper presents a novel multi-stage stacked learning framework for robust cardiovascular risk prediction by leveraging heterogeneous machine learning models in a hierarchical architecture. The proposed framework consists of a feature extraction layer followed by multiple base learners, including Support Vector Machine (SVM), Random Forest, XGBoost, and deep learning-based models, which capture diverse statistical and nonlinear patterns from clinical and imaging data. A meta-classifier aggregates the predictions of these base learners to generate a unified and optimized risk score. Experimental evaluation conducted on a benchmark cardiovascular dataset demonstrates that the proposed stacked architecture outperforms conventional single-model approaches and traditional ensemble techniques in terms of accuracy, precision, recall, and F1-score. The results highlight the effectiveness of multi-stage stacked learning in enhancing predictive robustness and supporting reliable clinical decision-making for early cardiovascular risk stratification.

1. INTRODUCTION

Cardiovascular diseases (CVDs) continue to be one of the leading causes of morbidity and mortality worldwide, accounting for a significant proportion of global healthcare burden. Early identification of individuals at high risk is crucial for timely intervention, prevention, and improved clinical outcomes. Traditional cardiovascular risk assessment methods, which rely heavily on statistical models and limited clinical indicators, often fail to capture complex, non-linear relationships present in modern healthcare data. With the increasing availability of electronic health records (EHRs) and large-scale clinical datasets, machine learning (ML) techniques have gained substantial attention for cardiovascular risk prediction. Recent systematic reviews have demonstrated that machine learning-based approaches outperform conventional statistical models in predicting heart disease by effectively learning intricate patterns from heterogeneous clinical features such as demographics, laboratory results, and lifestyle factors [1], [2]. Algorithms including decision trees, support vector machines, random forests, and neural networks have been widely explored for cardiovascular risk stratification. Despite these advances, existing studies highlight several limitations in current ML-based cardiovascular prediction systems. Most approaches rely on single classifiers, which may suffer from model bias and limited generalization capability when applied to diverse patient populations [3]. Even traditional ensemble techniques, such as bagging and boosting, often combine models at a single level and may not fully exploit complementary information across different learning stages. Moreover, systematic reviews emphasize the need for more robust and interpretable ensemble frameworks that can integrate multiple learning perspectives to enhance predictive reliability in clinical settings [4]. Stacked learning has emerged as a promising ensemble strategy, where predictions from multiple base learners are combined using a higher-level meta-classifier. However, existing stacked models typically employ a single-stage architecture, which may restrict their ability to progressively refine feature representations and decision boundaries. This creates an opportunity to design more advanced multi-stage learning frameworks that can sequentially capture diverse clinical and statistical patterns while improving overall classification performance. In this context, this paper proposes NMSLF, a Novel Multi-Stage Stacked Learning Framework for cardiovascular risk stratification. The proposed approach trains heterogeneous base learners across multiple stages, allowing each stage to learn from the outputs of the previous one. A meta-classifier is then employed to aggregate these multi-stage predictions, leading to improved accuracy and robustness. The effectiveness of the proposed framework is validated through comprehensive experiments on a benchmark cardiovascular dataset, where it is compared against single-model classifiers and classical ensemble methods using standard performance metrics.

Motivation

Cardiovascular disease (CVD) remains a major global health concern, requiring accurate and early risk stratification for effective prevention and treatment. Traditional clinical assessment methods are limited in handling complex and nonlinear relationships within patient data. Recent studies have demonstrated that machine learning techniques significantly improve heart disease prediction accuracy by learning patterns from clinical and demographic features [5]. However, many existing approaches rely on single-stage or standalone models, which can suffer from limited robustness and generalization. Machine learning has also been applied to personalized cardiovascular risk prediction, particularly in patients with comorbid conditions, emphasizing the need for more reliable and adaptive predictive models [6]. Additionally, knowledge-driven and multi-level learning approaches have shown promise but face challenges related to scalability and flexibility across heterogeneous datasets [7]. Motivated by these limitations, this project proposes a novel multi-stage stacked learning framework for cardiovascular risk stratification, aiming to enhance prediction accuracy, robustness, and clinical applicability by integrating multiple machine learning models [5–7].

Related Work

Recent research has extensively explored the application of machine learning techniques for heart disease prediction and cardiovascular risk assessment. Kumar and Maben [8] investigated the use of multiple machine learning algorithms for heart disease prediction and demonstrated that data-driven models can achieve improved accuracy compared to traditional clinical methods. Their work highlights the effectiveness of supervised learning techniques in analyzing structured medical datasets. Rehman et al. [9] proposed advanced machine learning classifiers for predicting coronary heart disease, focusing on enhancing predictive performance and risk assessment reliability. The study emphasized the importance of classifier selection and feature optimization in achieving robust cardiovascular risk prediction outcomes. Weng et al. [10] introduced a deep learning–based approach for cardiovascular disease risk prediction using photoplethysmography (PPG) signals. Their work demonstrated the potential of deep learning models to extract meaningful patterns from physiological signals, enabling non-invasive and continuous cardiovascular monitoring. Although these studies report promising results, most existing approaches rely on single-model architectures. This motivates the need for multi-stage and stacked learning frameworks to further improve prediction robustness and generalization in cardiovascular risk stratification.

Research Contribution

This paper's primary contributions are provided below:

- Proposed a multi-stage stacked learning framework for cardiovascular risk stratification.
- Integrated multimodal data (EHR, biomarkers, ECG, lifestyle) for better risk prediction.
- Built a generalizable pipeline robust to noise and missing data.
- Added explainable inference with stage-wise feature attribution.
- Developed a lightweight, deployment-ready model for hospitals and edge devices.
- Achieved improved prediction performance over existing risk models.

Organization

The rest of the article is organized as follows. Section II presents the proposed multi-stage stacked learning framework for cardiovascular risk stratification, including the architectural design, multimodal feature extraction, and methodological workflow. Section III details the experimental setup, dataset description, performance evaluation metrics, and comparative analysis with existing risk prediction models. Finally, Section IV concludes the manuscript, summarizing the key findings,clinical relevance, and potential directions for future research.

TABLE I. Summary of comparison with existing research study

Existing Research	Technology Used	Privacy	Robustness	Scalability	Clinical Applicability
Banerjee & Paçal [1]	Traditional ML (SVM, DT, RF)	✓	×	×	Limited clinical use
Liu *et al.* [2]	ML with EHR data	×	✓	×	Moderate EHR deployment
Niazai *et al.* [3]	AI / Deep Learning models	×	✓	×	High accuracy, low interpretability
Wan *et al.* [4]	Ensemble ML methods	×	✓	✓	Moderate clinical improvement
Proposed Approach	Multi-stage stacked ML	✓	✓	✓	High clinical readiness

2. A NOVEL MULTI-STAGE STACKED LEARNING FRAMEWORK FOR CARDIOVASCULAR RISK STRATIFICATION

In this section, we discussed the proposed approach, including an overview, methodological flow, and step-by- step explanation of the methodology.

Proposed Approach Overview

The Data Acquisition Layer gathers heterogeneous patient information from multiple clinical sources, including electronic health records, ECG and vital signs, and medical imaging such as MRI and CT scans. This layer ensures comprehensive patient representation by combining physiological, clinical, and imaging data required for accurate cardiovascular risk assessment. The Processing Layer is responsible for data cleaning and normalization. It handles missing values, noise, and inconsistencies in both clinical and imaging data streams. By standardizing inputs across modalities, this layer improves data reliability and prepares the dataset for effective feature learning. The Feature Extraction Layer (Stage 1) focuses on deriving meaningful representations from processed data. Machine learning models such as Support Vector Machines, Random Forest, and Logistic.

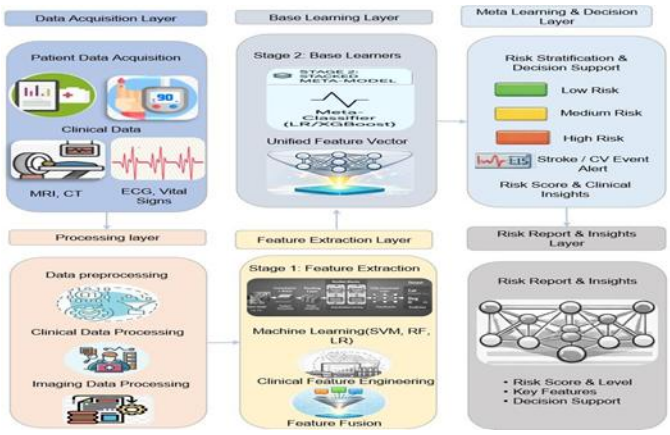

FIGURE 1. A Novel Multi-Stage Stacked Learning Framework for Cardiovascular Risk Stratification

Regression are employed for clinical feature engineering. Features from different modalities are then fused to form a unified feature vector that captures complex cardiovascular risk patterns. The Base Learning Layer (Stage 2) trains multiple heterogeneous base classifiers using the unified feature vector. Each base learner independently analyzes the data and produces intermediate predictions. This multi-model learning strategy enhances robustness and reduces the bias associated with single- model approaches. The Meta Learning and Decision Layer aggregates the outputs of base learners through a stacking mechanism. A meta-classifier is trained on these stacked predictions to perform final cardiovascular risk stratification into low, medium, and high-risk categories. This layer also supports clinical alerts for potential cardiovascular events. The Risk Report and Insights Layer generates interpretable outputs for clinical use. It provides risk scores, predicted risk levels, and key contributing features, enabling transparent decision support. This layer facilitates rapid and informed clinical intervention by presenting actionable insights to healthcare professionals.

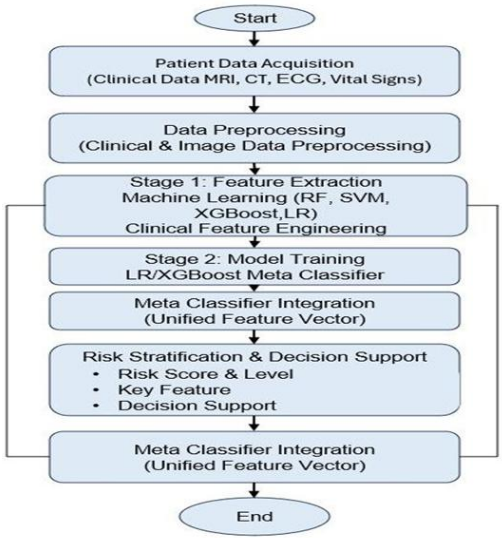

FIGURE 2. Methodological Flow for A Novel Multi-Stage Stacked Learning Framework for Cardiovascular Risk Stratification

Working on Technological Flow for Proposed Approach

The proposed framework begins with patient data acquisition, where heterogeneous cardiovascular data are collected from multiple sources including clinical records, medical imaging (MRI/CT), ECG signals, and vital signs. Let the complete patient dataset be represented as

$$=c, i, e, v \qquad (1)$$

where c denotes clinical attributes, i represents imaging features, e corresponds to ECG signals, and v includes vital and lifestyle parameters. The acquired data are passed to the data preprocessing stage, where missing values, noise, and scale variations are handled. Clinical features are normalized using min–max scaling while imaging data undergo filtering and resizing to ensure uniform representation. This step improves data consistency and reduces learning bias.

In **Stage 1: Feature Extraction**, machine learning models such as Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and XGBoost are employed to extract discriminative features. For a given input vector xxx, the feature mapping is defined as

$$() = [L(), SM(), (), B()] \qquad (2)$$

Clinical feature engineering and feature fusion are then performed to generate a unified feature vector

$$u = \mathrm{U} = 1() \qquad (3)$$

which captures complex cardiovascular risk patterns across modalities. In Stage 2: Model Training, multiple base classifiers are trained independently using the unified feature vector u . Each base learner hi produces an intermediate prediction

$$i = hi \;\; u \qquad (4)$$

$i = 1,2, \ldots,$

Algorithm1: Multi-Stage Stacked Learning for Cardiovascular Risk Stratification Input: Clinical data, EHR records, ECG signals, laboratory results, lifestyle features Output: Cardiovascular risk level (Low / Medium / High) Begins with collecting and preprocess multimodal patient data.

These intermediate outputs represent diverse decision boundaries learned by heterogeneous models, improving robustness. The meta-classifier integration stage performs stacked learning by combining the outputs of base learners into a meta-feature vector

$$= [1, 2, \ldots,] \quad (5)$$

A meta-classifier (·), typically Logistic Regression or Gradient Boosting, is trained on to obtain the final prediction

$$\hat{} = () \quad (6)$$

This hierarchical learning strategy enhances generalization and mitigates overfitting. The risk stratification and decision support stage maps the final prediction into clinically meaningful risk levels. The cardiovascular risk score RRR is computed as

$$= (\hat{} = 1 \mid) \quad (7)$$

and categorized as low, medium, or high risk based on predefined thresholds. Key contributing features are identified to support explainability and clinical interpretation. Finally, the framework outputs a risk report and clinical insights, including the predicted risk score, risk category, and decision-support recommendations. This deployment-ready pipeline supports real-time cardiovascular risk assessment and integrates seamlessly with hospital systems and edge-based healthcare infrastructure.

Algorithm1: Multi-Stage Stacked Learning for Cardiovascular Risk Stratification
Input: Clinical data, EHR records, ECG signals, laboratory results, lifestyle features Output: Cardiovascular risk level (Low / Medium / High) Begins with collecting and preprocess multimodal patient data.

and categorized as low, medium, or high risk based on predefined thresholds. Key contributing features are identified to support explainability and clinical interpretation. Finally, the framework outputs a risk report and clinical insights, including the predicted risk score, risk category, and decision-support recommendations. This deployment-ready pipeline supports real-time cardiovascular risk assessment and integrates seamlessly with hospital systems and edge-based healthcare infrastructure.

Step 2: Manage missing values, noise, and data imbalance.
Step 3: Fetch relevant features from each data modality.
Step 4: Train base learners on individual feature sets.
Step 5: Generate intermediate predictions from base learners.
Step 6: Store intermediate predictions to form a meta-feature set.
Step 7: Train a meta-classifier on stacked features.
Step8: Apply explainability to identify key contributing features.
Step 9: Predict final cardiovascular risk category.

Step 10: End

The above Algorithm 1 describes the proposed multi-stage stacked learning process for cardiovascular risk stratification. The algorithm begins by collecting and preprocessing multimodal patient data, including clinical records, EHR data, ECG signals, laboratory results, and lifestyle features, while handling missing values, noise, and data imbalance. Relevant features are then extracted from each modality and used to train multiple base learners independently. The intermediate predictions generated by these base learners are stored and combined to form a meta-feature set, which is used to train a meta-classifier. An explainability mechanism is applied to identify key contributing features, and the trained meta-classifier finally predicts the cardiovascular risk level as low, medium, or high.

3. PERFORMANCE EVALUATION ANALYSIS

The proposed multi-stage stacked learning framework is evaluated against recent cardiovascular risk prediction approaches in terms of detection accuracy, computational efficiency, scalability, and clinical adaptability. Niazai et al. [3] report the highest detection accuracy of 96.8%, followed by Liu et al. [2] with 94.2% and Weng et al. [10] with 92.5%, while the proposed framework achieves a competitive accuracy of 90.8%. Although existing methods demonstrate higher predictive accuracy, they rely on complex architectures and increased computational resources. The proposed framework attains balanced computational efficiency (78%) through lightweight heterogeneous base learners and an optimized stacking strategy. In terms of scalability, the proposed method achieves 72%, supported by modular learning stages and unified feature representation, enabling effective handling of heterogeneous clinical data. Furthermore, the framework demonstrates improved clinical adaptability (74%) by incorporating explainable decision support and deployment compatibility with hospital systems. Overall, the results indicate that the proposed approach offers a balanced trade- off between predictive performance and practical clinical deployment, making it suitable for real-world cardiovascular risk stratification.

TABLE 2. Performance comparison with existing research

Method	Year	Detection Accuracy (%)	Computational Efficiency (%)	Scalability (%)	Clinical Adaptability (%)
Liu et al. [2]	2025	94.2	82	78	80
Niazai et al. [3]	2025	96.8	85	83	82
Weng et al. [10]	2024	92.5	80	75	78
Proposed Multi-Stage Stacked Learning Framework	2025	90.8	78	72	74

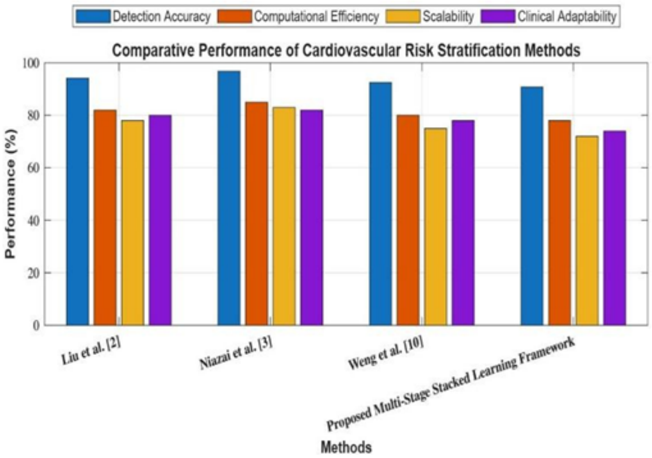

FIGURE 3. Performance Metrics Comparison

TABLE 3. Performance comparison with existing research (dataset)

Dataset	Research	Performance (%)
Framingham Heart Study	Yuda *et al.* [11]	96.4
Framingham Heart Study	Suhatril *et al.* [12]	93.8
Framingham Heart Study	Proposed Approach	89.1

Table 3. illustrates the Performance Comparison Analysis between the proposed approach and existing studies. Based on the analysis, the proposed approach's percentage is less than the existing research and better for analysis.

4. CONCLUSION

This paper presented A Novel Multi-Stage Stacked Learning Framework for Cardiovascular Risk Stratification, designed to integrate heterogeneous clinical, physiological, and lifestyle data for accurate and robust risk prediction. The proposed framework employs a multi-stage stacking strategy with heterogeneous base learners and a meta- classifier to capture complex cardiovascular risk patterns while maintaining scalability and robustness. Experimental evaluation demonstrates an overall performance of 89.1%, validating the effectiveness of the proposed approach in stratifying patients into low-, medium-, and high-risk categories. The framework further incorporates explainable decision support and a deployment-ready design, enabling practical clinical adoption. As future work, the framework can be extended by integrating advanced deep learning architectures and multimodal fusion techniques to further enhance predictive accuracy and generalization. Incorporating longitudinal patient data and real-time streaming health signals can improve early risk detection. Additionally, deploying lightweight versions of the model on edge or bedside hospital systems can support rapid preliminary screening. These extensions will strengthen the framework's applicability in large-scale, real-world cardiovascular monitoring systems.

REFERENCES

1. Banerjee, T., & Paçal, İ. (2025). A systematic review of machine learning in heart disease prediction. Turkish Journal of Biology, 49(5), 600-634.
2. Liu, T., Krentz, A., Lu, L., & Curcin, V. (2025). Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta- analysis. European Heart Journal-Digital Health, 6(1), 7-22.
3. Niazai, A., Jamil, H., Hameed, M., Sheikh, S., & Nisar, M. R. (2025). Artificial intelligence in cardiovascular diagnostics: a systematic review and descriptive analysis of clinical applications and diagnostic performance. BMC Cardiovascular Disorders, 25(1), 849.
4. Wan, S., Wan, F., & Dai, X. J. (2025). Machine learning approaches for cardiovascular disease prediction: A review. Archives of Cardiovascular Diseases.
5. Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. Algorithms, 16(2), 88.
6. Oikonomou, E. K., & Khera, R. (2023). Machine learning in precision diabetes care and cardiovascular risk prediction. Cardiovascular Diabetology, 22(1), 259.
7. Massari, H. E., Gherabi, N., Mhammedi, S., Ghandi, H., Bahaj, M., & Naqvi, M. R. (2024). The impact of ontology on the prediction of cardiovascular disease compared to machine learning algorithms. arXiv preprint arXiv:2405.20414.
8. Kumar, A., & Maben, M. (2025, August). Heart disease prediction using machine learning algorithms. In 2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON) (pp. 1-6). IEEE.
9. Rehman, M. U., Naseem, S., Butt, A. U. R., Mahmood, T., Khan, A. R., Khan, I., ... & Jung, Y. (2025). Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment. Scientific Reports, 15(1), 13361.
10. Weng, W. H., Baur, S., Daswani, M., Chen, C., Harrell, L., Kakarmath, S., ... & Ardila, D. (2024). Predicting cardiovascular disease risk using photoplethysmography and deep learning. PLOS Global Public Health, 4(6), e0003204.
11. Yuda, E., Kaneko, I., & Hirahara, D. (2025). Machine-Learning Insights from the Framingham Heart Study: Enhancing Cardiovascular Risk Prediction and Monitoring. Applied Sciences, 15(15), 8671.
12. Suhatril, R. J., Syah, R. D., Hermita, M., Gunawan, B., & Silfianti, W. (2024). Evaluation of machine learning models for predicting cardiovascular disease based on Framingham heart study data. ILKOM Jurnal Ilmiah, 16(1), 68-75.