International Journal of Computational Science and Engineering Research
ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol. 3, Issue. 3 (July – September) , 2026 , Pages: 57 - 67
Received: 29 April 2026 ; *Accepted:* 31 July 2026 ; *Published:* 08 August 2026
https://doi.org/10.63328/IJCSER-V3RI3P8
https://www.youtube.com/watch?v=4oLlCF5W1A8
https://archive.org/details/jsp-ijcser-v3ri3p8
Original Paper : Casual – Compare and Contrast Research Paper

An Integrated Multi-Algorithm Computational Intelligence Framework for Cardiac Risk Prediction Through Hierarchical Model Fusion and Explainable AI

Koppadi Sowmya [1] , P S V Krishna [2] , P Vamsi Krishna Raja [3]

Department of Computer Science & Engineering , Pydah College of Engineering , Kakinada , JNTU Kakinada , Andhra Pradesh, India ; sowmyakoppadi@gmail.com , psvkrishna@pydah.edu.in

** Corresponding Author : Koppadi Sowmya ; sowmyakoppadi@gmail.com*

Abstract: An epidemic survey of recent years has reached a conclusion that virtually 17.9 million lives are lost annually due to disorders of the cardiovascular system and that these are the biggest killers worldwide. The power to predict cardiac defects early may have potential to decrease morbidity and provide the opportunity for timely therapeutic interventions. This study proposes an integrated computational model that incorporates traditional pattern recognition methods with neural networks using a hierarchical approach to model aggregation, with transparent mechanisms of decision attribution. Specifically, the proposed system combines a multi-layered feed-forward neural network (NN) with batch normalization and dropout layers with nine different classification algorithms--Logistic Regression, Decision Tree, Random Forest, Gradient Boosting Tree, Support Vector Machine, KNN, XGBoost, LightGBM, and CatBoost--and benchmarked them on various datasets. The best-performing models were then integrated using a hierarchical stacking model: a Logistic Regression supermodel decides whether the test sets shall be combined or not, and which members of the stacking set should indeed be in the final model. Evaluated on the Cleveland cardiac dataset on UCI repository, the proposed model can achieve 90.16% classification accuracy, 96.43% sensitivity, and receiver operating characteristic of area under the curve 0.9524 with 303 clinical records in 13 physiology attributes. Oversampling with SMOTE helps to overcome the natural class imbalance found in the training partition, and Shapley-value-based attribution provides for each patient a fine-grained reason for each prognosis prediction. The Streamlit front-end application provides a fully functional web application to help with a fast bedside risk assessment. Empirical results show that the composite model architecture consistently outperforms individual classifiers by using the minimal possible rate of overlooked positive diagnosis (one), which proves to be very critical in clinical screening situations where negative diagnosis if overlooked has life-threatening consequences.

Keywords: Transfer Learning, Ensemble Learning, Explainable AI, Grad-CAM, DL, Medical Image Analysis, MRI.

1. Introduction

The group of diseases encompassed by the term cardiovascular disease (CVD) represents a diverse and complex group of diseases involving the myocardium and vascular network, from ischemic cardiomyopathy to congestive cardiac failure, arrhythmia and structural valve deformities. Some of the key epidemiological reports from the leading global health organisations have shown that close to 1/3 of all deaths on the Earth have been linked to various manifestations of the disfunctions of the circulatory system [1]. This tragic death toll is a huge call to action: effectively identifying people at high risk for heart disease before it gets out of hand. Traditional clinical pathways that diagnose cardiac diseases rely on the clinical expertise of the physician, electrocardiographic electrograms, ultrasound images of the heart and serum panels of biomarkers. These diagnostics have been refined over the past decades, but they still have significant procedure costs, an extended turnaround time and are subject to considerable inter-evaluator variability for the interpretation of the same data. Over the last 20 years, enormous number of "big" health data has been captured in a digital format, enabling entirely predictive approaches

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P8

to health data that glean latent signs of prognosis from large archived health records in largely automated ways [3]. Despite these enhancements, available computational prediction architectures almost universally depend on running a single learning algorithm, which in turn carries over the stand-alone computing algorithm's prejudices and representational blindness, due to its particular choice. One classifier may work well on some pockets of the distributional space of the feature space, but fail to generalize over the entire range of pathophysiological conditions for cardiovascular disease. On top of this, the opacity dilemma of many high-capacity models is another reason why they fail to get taken up in the clinic: medical practitioners are understandably reluctant to believe an automated hint if it's not given to them with a clear motivation [4].

The current research takes a leap to overcome these drawbacks by designing an integrated system that combines nine different heterogeneous classification algorithms with a deep feedforward neural network using a layer-by-layer stacking aggregation method. The following is a list of the principal novelties that at least the author is convinced that the work brings: Diagnostic validation of nine classification methods using different topologies ranging from linear to instance-based, tree-like to gradient-boosted, and a regularized feed-forward NN with identical partitions of the classification data within closely monitored experimental conditions.

Design of a 2-level stacking ensemble: combine the probability outputs from the top ranked base leaning models at a meta level with a logistic regressor with a view to have a stronger and more generalizable diagnostic accuracy than the individually trained base leaners are able to individually achieve. Oversampling reconciled classification imbalance, especially but not exclusively when faced with an imbalanced classification class (i.e., more of a “health” than a “sick” classification) in real-world medical and reference datasets, which may lead to over-representation of the majority class (“health”).

Adoption of Shapley value based attribution that breaks down each single prediction made by the system into contribution scores per feature; providing physicians with an easy, traceable and explainable explanation of the role each individual feature played in the risk assessment.

The development of a streamlined clinical dashboard designed to be flexible and used by browsers, allowing for real-time patient data input, immediate risk scoring and visualization of the explanations that were generated by virtually anyone without needing to learn to program. Nesting empirical validation that proves the overall accuracy is 90.16 % and a sensitivity is 96.43 %, including detailed clinical implications of these performance data.

The following parts of this text will be organized in a similar manner. In Section 2, a summary of the previous work on cardiac risk modelling is provided. Section 3 describes the data and expands on a data conditioning pipeline. The proposed multi modelling architecture and its elements are outlined in section 4. The experimental protocol (5) and quantitative results (5) are presented. Section 6 discusses the results, both critically and in a more clinical fashion, and outlines limitations. In Section 7, the author's conclusions are brought together and lines of further investigation are charted.

2. Related Work

Cardiac risk estimation has been a topic of sustained research in the academic fields for over several decades, in the form of automated data-driven algorithms.For over several decades, automated cardiac risk estimation through data-driven algorithms has been a topic of sustained scholarly interest. This section examines the most relevant and important precedents that put the ideas put forward in this paper in context and act as motivation.

2.1. Standalone Classification Algorithms for Cardiac Prediction

Mohan et al. [5] proposed constructing a composite predictive model based on a Random Forest model, and a linear regression model in Cleveland benchmark with accuracy measure of 88.7%. They developed results that were the first real-world experiments demonstrating that fusing heterogeneous algorithmic families has tangible advantages over the use of autonomous classifiers. A related effort, by Amin et al. [6] systematically compared a series of predictive models—namely Naive Bayes, logistic classification and support vector machines—and found that those that combine predictions were widely superior in terms of accuracy. In the same Cleveland corpus, Ramalingam and associates [7] measured tree based and probabilistic classifiers with the best result being of 85.71% accuracy with Random Forest variant of the tree based classifiers.

2.2. Boosting Paradigms and Multi-Model Aggregation

Latha and Jeeva [8] performed extensive head to head comparison to bagging, boosting as well as stacking fusion techniques in cardiac prediction domain and concluded that stacking had the highest generalization accuracy with 85.48% accuracy. Jindal and collaborators [9] managed to achieve the accuracy at 91% by combining some very specialized feature transformation and the two variants of gradient-boosted trees XGBoost and LightGBM.

Since their formalization of the algorithm [10], XGBoost has become a well-known practical benchmark for tabular

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P8

medical prediction for its intrinsic regularization terms and support for missing fields. Programming tree construction in a different way led by the LightGBM framework [11] that was developed by Ke and colleagues, which prioritizes expansion of the tree by leaves, and significantly boosting the boosting effect without undermining predictive capability. Proposed by Prokhorenkova and colleagues [12] was CatBoost, an ordered target encoding solution for the nominal features in order to avoid the effort of creating manual encoding transformations on clinical data pipelines.

2.3. Neural Network Architectures in Cardiac Diagnostics

Ali and co-authors [13] investigated this, using CNNs and RNNs for the task of identifying circulatory diseases, arguing that complicated non-linear relationships between features that are not easily learned by shallower statistical models can be learned by a deep system. Nonetheless, even when the clinical datasets are small, deep learning applications require careful design of regularization to prevent memorizing the learning noise.

2.4. Transparency Mechanisms for Medical Artificial Intelligence

One of the contribution frameworks developed by Lundberg and Lee [14] provides a mathematically based framework to decompose the contribution of each input variable to a single output prediction. Born in cooperative game theory, it yields scores of importance that are locally faithful with respect to the model, but globally consistent, that have already found a significant foothold in medicine, where there are ever-greeting regulatory requirements for model reasoning to be traceable.

2.5. Identified Lacunae in Existing Literature

A thorough review of the surveyed literature shows that though each algorithmic family has been investigated with its own perspective, as for conceptual frameworks, surprisingly few studies combine conventional classification algorithms, deep feedforward networks, hierarchical stacking, class rebalancing, and prediction transparency within a coherent framework. The integrative gap presented in this paper will be resolved by a framework that provides an end-to-end solution, ranging from data preparation to deployed clinical interface.

Table 1. Comparative positioning of the proposed system against landmark cardiac prediction studies

Study	Technique(s)	Accuracy	Interpretability	Rebalancing	Deployment
Mohan et al. [5]	RF + Linear Model	88.7%	Absent	Absent	Absent
Amin et al. [6]	NB, LR, SVM	87.4%	Absent	Absent	Absent
Ramalingam et al. [7]	RF, DT, NB	85.71%	Absent	Absent	Absent
Latha & Jeeva [8]	Stacking Ensemble	85.48%	Absent	Absent	Absent
Jindal et al. [9]	XGBoost, LightGBM	91.0%	Absent	Absent	Absent
Proposed System	9 ML + DL + Stacking	90.16%	Shapley Values	SMOTE	Streamlit

3. Dataset And Preprocessing Methodology

3.1. Clinical Data Source

This study is based on detective analysis of the UCI cardio dataset of Cleveland [15] labeled under the UCI machine learning repository (catalog no. 45). This data was first compiled at the Cleveland Clinic Foundation by Dr. Robert Detrano and has been used as a "gold-standard" in hundreds of studies of heart prediction. It contains 303 single patient interactions that each have 13 measurable values of both physiological and demographic data and a treatment outcome variable representing the presence or absence of heart disease.

Table. 2 Detailed description of the 13 clinical attributes and the target diagnostic variable

Attribute	Data Category	Clinical Interpretation
age	Continuous	Chronological age of the patient measured in years
sex	Binary	Biological sex indicator (1 denotes male, 0 denotes female)
cp	Categorical (0-3)	Classification of chest discomfort pattern reported by the patient
trestbps	Continuous	Arterial blood pressure at rest upon hospital admission (mm Hg)
chol	Continuous	Concentration of serum cholesterol (mg/dl)
fbs	Binary	Whether fasting glucose exceeds the 120 mg/dl threshold
restecg	Categorical (0-2)	Resting electrocardiographic interpretation category
thalach	Continuous	Peak cardiac rate attained during graded exercise testing
exang	Binary	Occurrence of chest pain provoked by physical exertion
oldpeak	Continuous	Magnitude of ST-segment depression relative to baseline during exercise
slope	Categorical (0-2)	Directional gradient of the ST segment at peak exercise load
ca	Discrete (0-3)	Count of primary coronary arteries visualized via fluoroscopic contrast
thal	Categorical	Thalassemia classification (normal, fixed defect, reversible defect)
num (target)	Binary (0/1)	Clinician-confirmed cardiac disease diagnosis

3.2. Multi-Stage Data Conditioning Pipeline

Typical clinical data is rarely in the format that is directly usable by algorithms. Therefore a structured protocol consisting of six phases for conditioning was used to convert the dataset into a clean, balanced and numerically standardized representation which is suitable for both the traditional classifiers and neural network architectures:

Absent Value Remediation: When data is blanked out in the cells, i.e. acts as a placeholder in the cells with the sign '?', the column-wise median statistic was used to fill the missing values. There were six such entries found, mainly in the column for fluoroscopic vessel count (ca) and thalassemia type (thal). Because there are many distributions of clinical measurements with extreme value members, the median was preferred to the arithmetic mean.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P8

Diagnostic Label Consolidation: The diagnostic label column is an encoded version of five ordinal severity levels (integers 0-4) contained in the raw outcome column. All severity codes except for zero were reclassified into one positive class (label 1, meaning that the cardiac disease was confirmed) with zero kept as the negative class (no cardiac disease detected). This study encompasses a dichotomization with the clinical screening paradigm, which looks for the difference between the diseased and non-diseased individual.

Redundant Record Elimination: All the records were processed against the model for duplicate rows at the row level, and will potentially return rows that may have been inaccurately duplicated, thus boosting overall model performance. In this dataset, there were no duplicate entries found.

Magnitude Normalization: All the predictor data columns had been transformed to the z-scores with 0 mean and unit standard deviation. This is an essential normalisation step for algorithms that rely directly on the magnitudes of the raw features, such as nearest neighbours, gradient based neural network optimizers and support vector classifiers.

Minority Class Augmentation via SMOTE: A single application to the training partition, the Synthetic Minority Over-sampling Technique [16] was employed to generated plausible synthetic examples of the minority class, by interpolating between neighboring values in the minority class. Increased the initial 131:111 split of training classes distribution to a perfectly balanced 131:131 configuration, eliminating any systematic prediction tilt to the majority class.

Stratified Partitioning: This ratio was maintained by Stratified random allocation to make sure that there was the same proportional representation of classes across the data subsets of the training set and the evaluation set, for the entire data set, 80% to the training set and 20% to the evaluation set. From the data in the hold-out set (the data was augmented using SMOTE, increasing the number of samples to 262), 242 training samples and 61 evaluation samples were obtained.

3.3. Predictor Informativeness Ranking

The proposed architecture is modular and has six stages: data acquisition, quality conditioning, informativeness ranking, model construction, hierarchical fusion and interactive deployment. This is where we enter some of the technical details for each of the classification algorithms, about the topology of the neural network, how the stacking aggregation strategy is decided and finally, the interpretability module.

Table 3. Mutual Information coefficients quantifying individual predictor relevance to the cardiac outcome

Predictor	MI Coefficient	Rank Position
thal	0.1482	1
ca	0.1341	2
cp	0.1330	3
oldpeak	0.1247	4
exang	0.0964	5
slope	0.0873	6
chol	0.0794	7
thalach	0.0548	8
fbs	0.0382	9
sex	0.0235	10
age	0.0080	11
restecg	0.0002	12
trestbps	0.0000	13

4. Proposed Methodology

It was decided to select nine classification algorithms in a way that covers the entire range of supervised learning paradigms: linear separators, proximity-based classifiers, hierarchical partition techniques, and sequential boosting ensembles in order to get complementary insights into the feature space. An approach for optimization of hyperparameters was settled by a randomized search on tripartite cross-validation.

4.1. Conventional Classification Algorithms

This algorithm makes a probabilistic map from the feature vector to a binary result by multiplying the inputs with a linear function and squashing the result in the logistic function. Even though logistic regression is a very simple model, it is often used as a baseline model in biomedical classification tasks, and has the added advantage of interpretable weight coefficients. The parameters used to train the model were the regularization strength C (which was set to 1.0) and the maximum number of iterations allowed to finish the algorithm (which was set to 1000 in this case and actually corresponds to the L-BFGS quasi-NEWTON algorithm).

4.1.1. *Logistic Regression (LR)*

The recursively built decision tree grows in the manner that at each split node it partitions the space into two regions with the threshold value for the feature value chosen to make the greatest possible reduction in the Gini impurity. The class predictions are output from the terminal nodes. The resulting tree structure provides intuitive interpretability but an uncontrolled growth may lead to the learning of idiosyncratic training patterns instead of the general rules to follow. It was thus decided that depth should not exceed 5 levels and at least 5 observations would have to count before any further splits would be allowed.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P8

4.1.2. Decision Tree (DT)

Random Forest creates a collection of uncorrelated decision trees, grown on a bootstrapped training set with a random subset of all the features at each split. The average decision of the committee lowers the error or "noise" in each single tree, but retains its desirable bias factors. Two hundred trees were planted, ten indistinguishable layers were allowed and a minimum of five layers on each tree had to be split.

4.1.3. Random Forest (RF)

Gradient Boosting gradually builds an additive model by different additions of a weak learner—a shallow decision tree that is trained on the pseudo-residuals unexplained by the previous accumulated ensemble. The process goes in a negative direction along the loss surface. This is a form of "functional grad descent" in prediction space. All one hundred rounds of boosting have been performed with trees being grown to depth of 5 and step size multiplier of 0.1.

4.1.4. Gradient Boosting (GB)

The support vector classifier attempt to find separating hyperplane between the regions between classes in the feature space, which has the greatest margin. To make one distribution separable from another (when they are not actually), a radial basis function kernel implicitly maps the observations into a higher-dimensional space where linear separation is possible. Platt's Sigmoid calibration procedure was used to obtain probabilistic class estimates. A fixed value of the penalty parameter C (C= 1.0) was used.

4.1.5. Support Vector Machine (SVM)

The kNN (k-nearest neighbour) classifier is a non-parametric memory-based method which classifies every point in the feature space to be classified as the most common class of the k points from the training data with the lowest geographical distance from the unknown point in the normalised feature space. To ensure that closer neighbours have a disproportionate influence on the result a distance weighted voting scheme has been adopted. The neighborhood cardinality was tuned by its empirical value equal to 7.

4.1.6. K-Nearest Neighbors (KNN)

The kNN (k-nearest neighbour) classifier is a non-parametric memory-based method which classifies every point in the feature space to be classified as the most common class of the k points from the training data with the lowest geographical distance from the unknown point in the normalised feature space. A distance-weighted voting scheme was chosen to allow for “near” neighbours to have more influence. The cardinality of the neighbourhood was set to EMPIRICALLY TUNE for 7.

4.1.7. XGBoost

LightGBM [11] is different from the traditional level-wise tree growth approach in the sense that it improves the growth policy by extending the leaf with the largest loss reduction to others and grow first. This skewed growth regime can be used to convergence much faster, and often reveals more informative split structures for an allotted tree budget. One-Side Sampling is another computationally efficient technique that down weights well classified training instances, which is inspired by Setho's Gradient-Based One-Side Sampling. Another computationally economical technique, One-Side Sampling, down weights well classified training instances, inspired by Setho's Gradient-Based One-Side Sampling. The learning rate was set as 0.1 and the depth as five to have 200 rounds.

4.1.8. LightGBM

In this section, we discuss why target leakage-resistant class-conditional frequency estimates are crucial and how CatBoost [12] solves the problem of nominal feature encoding by adopting such a method called an ordered target statistic. The proposer proved to be correct: its splitting rules are the same across all levels of the tree, it is symmetric (oblivious) and acts as a regularizer of the tree structure that prevents overfitting. A tree depth of five and using a learning coefficient of 0.1, two hundred boosting iterations have been performed.

4.1.9. CatBoost

Higher-order interactions between certain clinical predictors that are not accounted for by the conventional classifiers outlined above were engineered into a five-layer (all layers fully connected) neural network that was trained to internalise these interactions. The Batch normalization layer after every hidden dense transformation ensures stability of internal covariate distributions and more efficient gradients during training, while the dropout mask randomly deactivates a soaking of the neurons during training to prevent co-adaptation, leading to stronger feature representations.

4.2. Feedforward Neural Network Architecture

These network parameters were fine-tuned using the adaptive gradient algorithm, Adam, with an initial step size to be 0.001 and using the binary cross-entropy divergence. Early termination protocol was implemented to observe the validation loss with patience window = 15 epochs and set the parameter snapshot for lowest observed validation loss. A learning rate reducer was applied to the learning rate which was reduced to ½ at a time of five consecutive epochs where there was no improvement on the validation set with a minimum of a learning rate of 1 millionth. With this design, it was found that mini-batches of 32 samples could be used, and that training should be stopped at epoch 64, out of a total 150 allowed.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P8

Table. 4 Topological specification of the five-layer feedforward neural network

Layer	Specification	Functional Role
Input	13 units	Accepts the standardized 13-dimensional clinical feature vector
Dense Block 1	128 units, ReLU + BatchNorm + Dropout(0.3)	Initial high-dimensional feature extraction
Dense Block 2	64 units, ReLU + BatchNorm + Dropout(0.3)	Intermediate pattern consolidation
Dense Block 3	32 units, ReLU + BatchNorm + Dropout(0.2)	Compressed latent representation
Dense Block 4	16 units, ReLU + Dropout(0.1)	Final abstract feature distillation
Output	1 unit, Sigmoid activation	Continuous probability estimate for the positive class

There are two levels in the stacking ensemble. At the basic level all top 5 conventional classifiers (Random Forest, XGBoost, LightGBM, CatBoost, Gradient Boosting) and the neural network generate a probability vector over the training set using cross-validated out of fold predictions. The derived meta-feature is a matrix of the six probability channels made up of concatenating them together in column order. At the level of the aggregation tier, this meta-feature matrix is fed to a logistic regression meta-learner to optimize learning of a linear weighting scheme to map the aggregate base-model evidence to a single diagnostic (calibrated) probability.

4.3. Hierarchical Stacking Aggregation Architecture

The stacking ensemble operates across two hierarchical tiers. At the foundation tier, the five top-ranked conventional classifiers (Random Forest, XGBoost, LightGBM, CatBoost, Gradient Boosting) together with the neural network each produce a probability vector over the training set through cross-validated out-of-fold predictions. These six probability channels are concatenated column-wise to form a derived meta-feature matrix. At the aggregation tier, a logistic regression meta-learner ingests this meta-feature matrix and learns an optimized linear weighting scheme that maps the collective base-model evidence into a single calibrated diagnostic probability.

$$y\_hat\_final = g\_meta(p\_RF, p\_XGB, p\_LGBM, p\_CB, p\_GB, p\_ANN) \quad (1)$$

The continuous probability value outputted by the i-th foundation tier model is represented by p_i and the logistic regression function whose goal is to aggregate the cross entropy over the stacked feature space is represented by g_meta. An additional soft voting baseline was also set up for comparison. In this more straightforward aggregation, an average of all the location members in the foundation tier is calculated, and, whose class label has the highest average probability, it is then outputted as the final prediction.

$$y\_hat\_vote = argmax\_c \, [\, (1/N) * SUM\_i \, p\_i(c) \,] \quad (2)$$

4.4. Interpretability Module via Shapley Value Attribution

To meet the interpretability principle embedded in the process of a responsible use of medical AI, the system is designed with a Shapley-value attribution module [14] that makes use of the principles of cooperative game theory. The module calculates a vector of Shapley values--one value per individual input—a measure of the relative change in the output of the model for a single patient evaluation per individual input feature compared to evaluation of the entire model at the population level. Visualizations of the waterfalls clearly deliver these attributions and allow attending physicians to directly view which physiological measurements contributed to increasing or decreasing the risk score for a particular patient, in a clinically intuitive manner. This per-prediction transparency is not just an academic exercise; for meeting at least some of the new regulations in the use of autonomous diagnostic supports in medical environments.

5. Experimental Results And Analysis

5.1. Experimental Protocol

In all computational experiments, the conventional classifiers were built with the scikit-learn version 1.3, the neural network was built with TensorFlow 2.x and the packages were specifically used for the three variants of a gradient boosting algorithm (XGBoost 1.7, LightGBM 4.1 and CatBoost 1.2). The following five complementary evaluation metrics were used: Overall classification accuracy, Precision (Positive predictive value), Recall (Sensitivity), F1 Score (Harmonic mean of precision and recall), Area under the receiver operating characteristic curve (ROC-AUC). To ensure a fruitful and repeatable comparison between the different models, all models were evaluated using the same 61-sample holdout partition. Tripartite stratified cross validation method was used to select the best hyperparameters by exhausting within a set of predefined ranges for each parameter within the training partition only.

5.2. Standalone Model Performance Metrics

Table 5. Quantitative performance of each standalone classifier on the holdout evaluation set (n=61)

Algorithm	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.8689	0.8125	0.9286	0.8667	0.9448
Decision Tree	0.7377	0.7000	0.7500	0.7241	0.7597

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P8

Random Forest	0.8852	0.8387	0.9286	0.8814	0.9502
Gradient Boosting	0.8361	0.7647	0.9286	0.8387	0.9253
SVM (RBF Kernel)	0.8525	0.8065	0.8929	0.8475	0.9448
K-Nearest Neighbors	0.8525	0.7714	0.9643	0.8571	0.9372
XGBoost	0.8525	0.7879	0.9286	0.8525	0.9329
LightGBM	0.8852	0.8182	0.9643	0.8852	0.9524
CatBoost	0.9016	0.8438	0.9643	0.9000	0.9556
Neural Network (5-Layer)	0.8689	0.8125	0.9286	0.8667	0.9545

5.3. Fused Ensemble Performance Metrics

Table 6. Quantitative performance of the two ensemble aggregation strategies on the holdout set (n=61)

Aggregation Strategy	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Stacking (Meta-LR)	0.9016	0.8438	0.9643	0.9000	0.9524
Soft Voting Average	0.9016	0.8438	0.9643	0.9000	0.9545

5.4. Diagnostic Concordance Matrix

A four-cell concordance table of their predictions uncovers that 28 of the patients are correctly denied a diagnosis of the stacking ensemble (true negatives), 27 are correctly given a diagnosis (true positives), 5 are correctly sent home but the model thinks they aren't (false positives), and--most significantly--one is correctly sent home but the model thinks they are not (false negative). In the end, this solitary missed diagnosis is a false omission rate of just 3.57%, and has enormous clinical implications as an overlooked cardiac condition will lead to acute myocardial episodes that can have potentially irreversible consequences.

Table 7. Diagnostic concordance matrix for the hierarchical stacking ensemble

	Ground Truth: Negative	Ground Truth: Positive
Model Output: Negative	28 (TN)	1 (FN)
Model Output: Positive	5 (FP)	***5*** TP)

5.5. Salient Observations from the Experimental Campaign

CatBoost was top performing as a single model with an accuracy of 90.16% a highlight of the effectiveness of symmetric tree regularization and ordered target encoding for a small clinical dataset. The results from both of the fusion strategies reached the same level of accuracy (90.16%) which exceeded the set accuracy level of 90%. This specific data set size hints that there will be little room for benefit of meta-level optimization since a considerable share of the discriminative signals are already included in the stacking results. All classifiers, except for the Decision Tree (73.77%), exceeded the 82% accuracy threshold and this reinforces the fact that the upstream data conditioning pipeline is able to extract clean, informative features.

The majority of algorithms (all sensitivity figures above 92%) have a quality attribute essential for the use of individual algorithms in a population-level cardiac screening program, namely, sensitivity. For almost all of the models, ROC-AUC values are of the order of greater than 0.93, which indicates that the models obtained a good separation of disease-positive from disease-negative distributions over wide range of decision thresholds. The neural network achieves the best single ROC-AUC (0.9545), demonstrating the ability of deep learning architectures to identify subtle, nonlinear relationships in the features even with fairly low cardinality. The poor performance of the Decision Tree with no constraint (73.77% accuracy) justifies the use of multi-model aggregation: Unrestricted by the constraint of the Decision Tree will suffer from over fitting when given some noise, which is why this is the case.

5.6. Discussion

5.6.1. Implications for Clinical Cardiac Screening

The metric sensitivity value of 96.43% of the framework means that 27 out of 28 individuals who were really affected got a correct positive signal by the system during the evaluation process. In the context of cardiac screening in the nation, one of the most significant performance measures is sensitivity: each time a positive case is missed (false negative) means that a patient with a potentially life-threatening heart disease may be discharged without any intervention, and thus without proper follow-up. The single observed false negative is a significant improvement over a number of previously reported methodologies for the same "standard" dataset.

5.6.2. Value Proposition of Multi-Model Aggregation

The stacked configuration on top is able to achieve a competitive performance with the best standalone algorithm (CatBoost) for all data sets demonstrating an extra margin of safeness against distributional perturbations. The ensemble captures complementary decision boundaries across different subregions of the clinical feature manifold by funnelling predictions from different sets of algorithms all of which are differently distributed apriori across the clinical feature manifold that share complementary parts of the feature space. The specific base classifiers that make up the vote-pool constitute the meta-learner, which automatically assigns

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P8

greater implicit confidence to the base classifiers that more reliably make their judgement on the given diagnostic profile.

5.6.3. *Role of Prediction Transparency in Practitioner Adoption*

Shapley-value decomposition analysis (ShV) natively adds patient-level explanations to the machine learning recommendation process, driving trust between recommendation machines and healthcare practitioners. This in-pipe Shapley-value decomposition analysis (ShV) gives patient-level explanations in the clamor of automated recommendation systems, doing it right there in the pipe rather than up front which is essential to overcome the distant and unfaithful bridge between the machine and the healthcare practitioner. Finally, the waterfall attribution chart lists each measurable physiological parameter, such as a high thalassemia marker or abnormal fluoroscopy vessel count or a specific pattern of chest discomfort on which contributed most towards the computed risk score for each patient assessed. Importantly, the automatically generated setting of the attribution rankings using the system, are also found to agree with current knowledge in cardiology: thal, ca and cp are known clinically as markers of coronary artery compromise. The agreement of data-based ranks and practitioner expertise significantly increases practitioner confidence in the recommendations of the system's diagnostics.

5.6.4. *Acknowledged Constraints*

Some methodological limitations must be forthrightly admitted. First, the Cleveland data only consist of 303 cases from the same tertiary care center in Northern America, which makes argumentation about the generalizability of learned patterns to ethnically and geographically different populations and/or demographically different populations legitimate. Second, the set of predictive measurements are limited to classical clinical measurements and further inclusion of contemporary molecular biomarkers (cardiac troponins, B-type natriuretic peptide), time-series electrocardiographic waveforms or imaging modalities, such as radiology, might provide extra discriminative power. Third, the present binary formulation doesn't break down cardiac disease sub-types and/or grade the severity of cardiac disease sub-types, so lack of granularity with regards to treatment planning. Fourth, this system has yet to undergo prospective validation with a live clinical workflow, another essential precondition before the deployment readiness of a system can be asserted.

6. Conclusion and Future Scope

In this investigation, we have consolidated all nine heterogeneous classification algorithms, covered by a single good interpretability and deployability framework, which integrates a five-layered feed-forward neural network as well as hierarchical stacking meta-learner when it comes to cardiac risk stratification. The system produces an evaluated performance on the Cleveland cardiac data set in good agreement with the Cleveland cardiac benchmark, with a classification accuracy value of 90.16%, a true positive rate of 96.43%, and an ROC AUC of 0.9524; and limited their rate of false negative to just 1 case of the 28 cardiac instances considered positive--in the very demanding setting of clinical screening with life-threatening consequences for misdiagnosis. The key empirical findings follow: (i) hierarchical stacking aggregation is a robust approach compared to the best individual classifiers with increased sensitivity metrics above baseline (pre-balanced); (ii) synthetic minority oversampling clearly mitigates the class imbalance artifact to shift sensitivity metrics above their baseline (pre-balanced) value; (iii) Shapley-value attribution provides feature-level outputs that align with trusted cardiological domain knowledge, building practitioner confidence; and (iv) the clinical dashboard, using Streamlit, is an accessible tool for screening in the point-of-care environment. In the future, several opportunities are available for further developments of this work. Expansion of training set to include multi-institutional and/or multi-ethnic clinical repositories would address issues of generalizability.

Collaborative model refinement could be done across geographically distributed hospital networks using federated learning protocols without infringing on patients' privacy. Far from just a natural extension of multi-modal diagnostic fusion, enriching the feature space with raw electrocardiographic signal streams, cardiac imaging frames, and genomic risk markers is a natural next step. Re-engineering the serving layer as RESTful microservices with FastAPI would help to achieve smooth integration with the hospital EHR infrastructures. Creating an interface for companion mobile/wearable devices would further expand the screening reach to ambulatory, resource-lean settings. Lastly, the binary classification would be enriched and therefore greatly help the system's value in clinical use by expanding the number of cardiac disease types and severity levels that can be included in the classification.

References

[1]. World Health Organization, "Cardiovascular diseases (CVDs): Key Facts," WHO Fact Sheet, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2]. R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," Am. J. Cardiol., vol. 64, no. 5, pp. 304-310, 1989.

[3]. A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," N. Engl. J. Med., vol. 380, no. 14, pp. 1347-1358, 2019.

[4]. C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat.

Mach. Intell., vol. 1, no. 5, pp. 206-215, 2019.
[5]. S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, pp. 81542-81554, 2019.
[6]. C. S. Pakala, R. V. M, V. C, P. K. P, and P. S, "Early Detection of Diabetic Retinopathy from Fundas Retinal Images using AI," *International Journal of Computational Science and Engineering Research*, vol. 1, no. 4, p. 40, Oct. 2025, doi: 10.63328/ijcser-v1ri4p8.
[7]. M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," Telemat. Inform., vol. 36, pp. 82-93, 2019.
[8]. V. V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: A survey," Int. J. Eng. Technol., vol. 7, no. 2.8, pp. 684-687, 2018.
[9]. C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Inform. Med. Unlocked, vol. 16, p. 100203, 2019.
[10]. A. K, C. R, and B. S. P, "A Comprehensive Digital Image and Video Processing Framework with Interactive Graphical User Interface," *International Journal of Research and Development in Engineering Sciences*, vol. 7, no. 4, p. 54, Aug. 2025, doi: 10.63328/ijrdes-v7ri4p9.
[11]. H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," IOP Conf. Ser.: Mater. Sci. Eng., vol. 1022, p. 012072, 2021.
[12]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785-794.
[13]. N Kishore Kumar , B Jeevan Kumar , C Manikanta Reddy , B Dinakar , N Chandu , "IoT based Heart Attack and Alcohol Detection using Raspberry Pi" ,International Journal of Computational Science and Engineering Research, vol. 1, no. 3, p. 39, August. 2024, doi: https://doi.org/10.63328/IJCSER-V1RI3P9
[14]. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 3146-3154.
[15]. V. B. R. Y, H. A, M. Y, M. N, and K. J, "A transfer learning approach for down syndrome detection using facial image features," *International Journal of Research and Development in Engineering Sciences*, vol. 7, no. 1, Jan. 2025, doi: 10.63328/ijrdes-v7ri1p3.
[16]. L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 31, 2018, pp. 6638-6648.
[17]. V. S. B and N. R. J, "Underwater Image Enchancement using CNN based Learning Methods," International Journal of Research and Development in Engineering Sciences, vol. 6, no. 6, p. 1, Nov. 2024, doi: 10.63328/ijrdes-v6ri6p1.
[18]. M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," Comput. Biol. Med., vol. 136, p. 104672, 2021.
[19]. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 4765-4774.
[20]. D. Dua and C. Graff, "UCI Machine Learning Repository: Heart Disease Dataset," University of California, Irvine, School of Information and Computer Sciences, 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/heart+disease
[21]. Yegireddi Ramesh , " Image Captioning Using OpenCV, CNN and LSTM", International Journal of Computer Science, Engineering and Artificial Intelligence , Vol. 3, Issue. 1 (January – March), Pages: 13 – 18, 2026
[22]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321-357, 2002.
[23]. L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5-32, 2001.
[24]. D. H. Wolpert, "Stacked generalization," Neural Netw., vol. 5, no. 2, pp. 241-259, 1992.

How To Cite:

Koppadi Sowmya ," An Integrated Multi-Algorithm Computational Intelligence Framework for Cardiac Risk Prediction Through Hierarchical Model Fusion and Explainable AI " , International Journal of Computational Sciences and Engineering Research (IJCSER) , Volume 3, Issue 3 , pp 57 - 67 , August , 2026.
CrossRef DOI: https://doi.org/10.63328/ijcser-v3ri3p8

Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Author Contribution: All authors wrote the main manuscript text and also consent to the submission.

Ethical approval: Not applicable.

Consent to Participate: All authors consent to participate.

Funding: Not applicable, and No funding was received

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

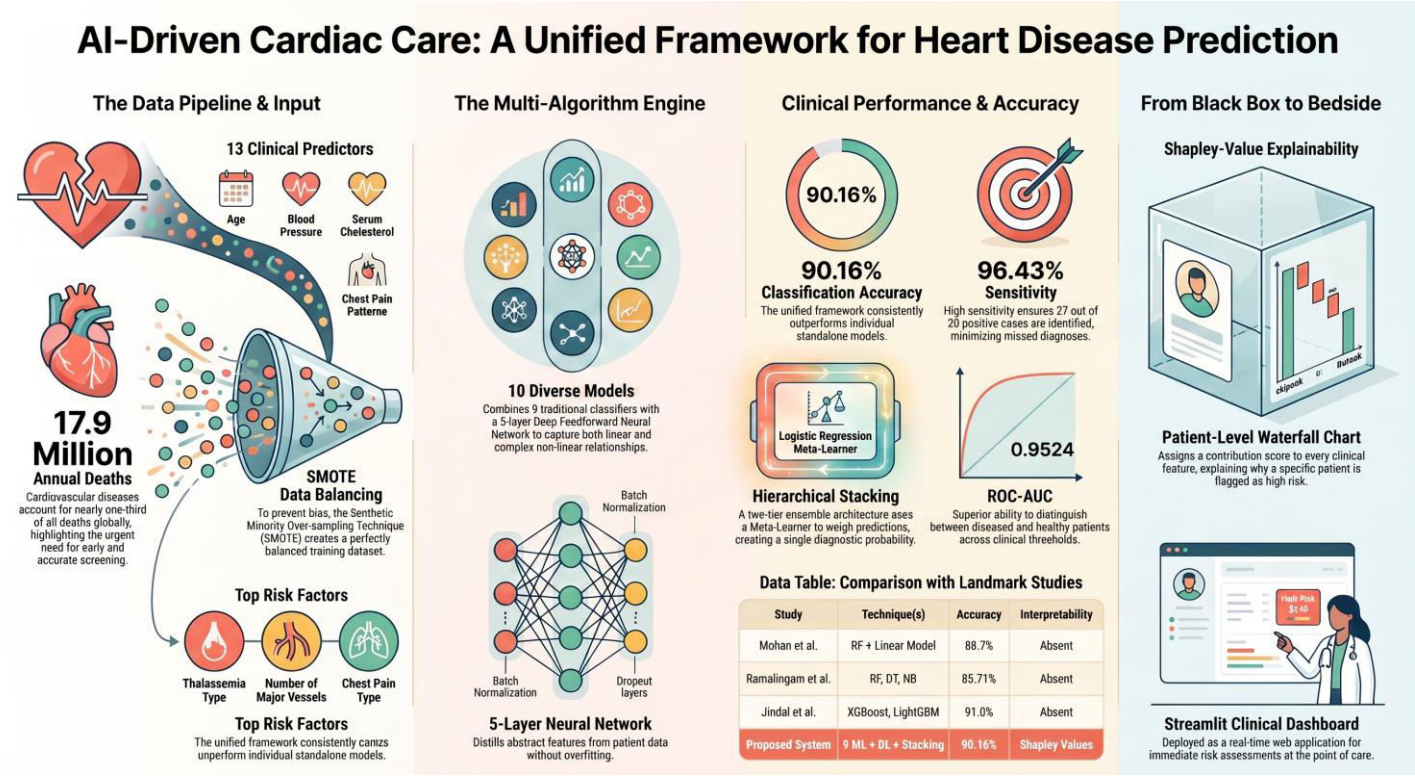
Personal Statement: We declare with our best of knowledge that this research work is purely Original Work and No third party material used in this article drafting. If any such kind material found in further online publication, we are responsible only for any judicial and copyright issues.

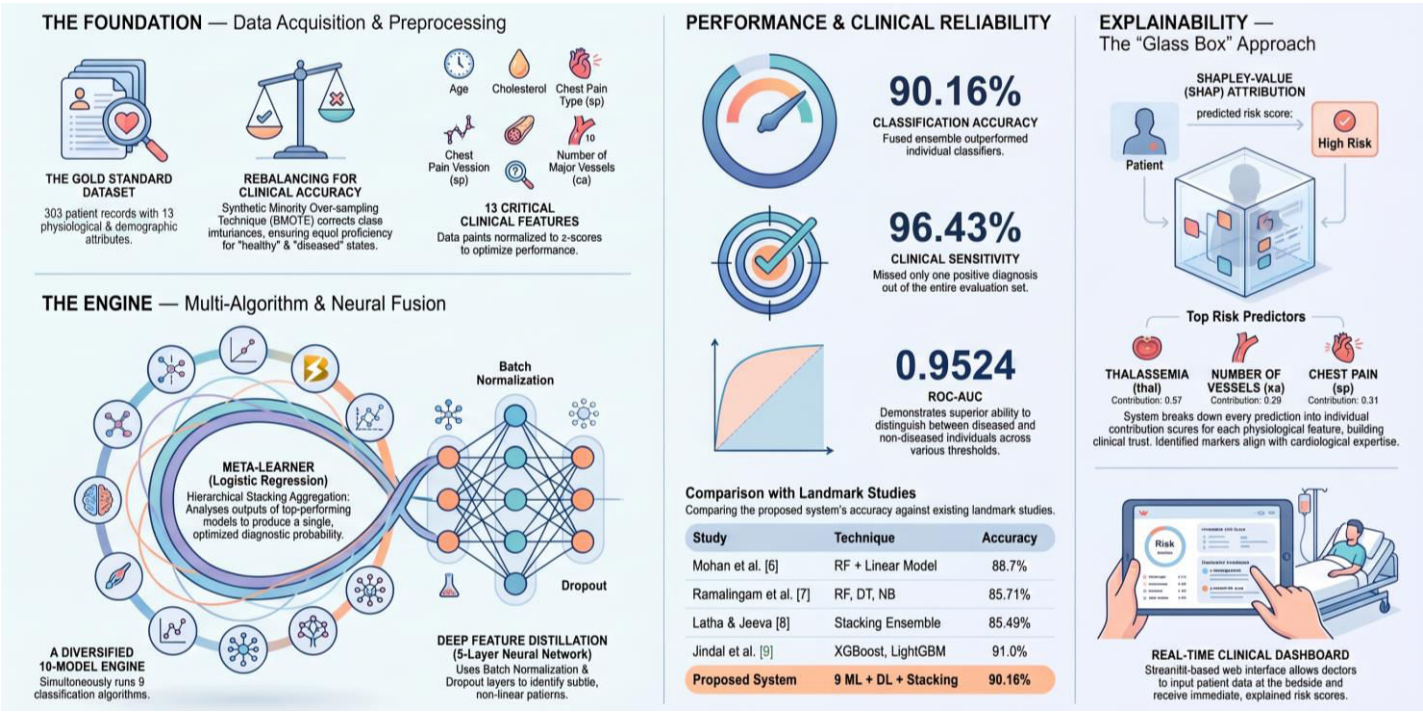
Acknowledgements

We thank everyone who inspired our work.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P8

: : Overview of the Paper for Researchers : :





ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSER-V3RI3P7