© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720179

Heart Disease Prediction Using a Hybrid Random Forest–Artificial Neural Network Model: A Replication and Extension of the HRFLM Approach

JYOTI[1], AMANDEEP[2], DHARMENDER KUMAR[3], ANKIT[4], SURAJ SHARMA[5]
[1,4,5]M.Sc. Computer Science, Artificial Intelligence and Data Science, GJUS&T, India.
[2]Assistant Professor, Artificial Intelligence and Data Science, GJUS&T, India
[3]Professor, Artificial Intelligence and Data Science, GJUS&T, India.

Abstract- ***The leading cause of death worldwide is cardiovascular disease, which is motivating screening using non-invasive machine learning on clinical parameters collected routinely. The present work paper reviews eight studies representative and heart disease prediction most relevant, highlights the major research gaps they leave open, and replicates a well-known hybrid baseline the Hybrid Random Forest with Linear Model (HRFLM) proposed by Mohan et al., and finally presents an improved Hybrid RF-ANN. The proposed model rescaling clinical input feature by their importance score derived from Random-Forest and feeding those features to two hidden layers NN. The Hybrid RF-ANN, whose code is published, attained an accuracy of 90.16% in the study, a significant 9.83 percentage point improvement over the accuracy of HRFLM, while the Hybrid RF-ANN achieved the highest ROC-AUC of all ten studied models (0.977), which is an enhancement of 0.0787 in ROC-AUC against HRFLM.***

Keywords: Heart disease prediction, Hybrid RF-ANN, HRFLM, Random Forest, ensemble learning, ROC-AUC.

I. INTRODUCTION

Cardiac disease continues to be one of the most common causes of death worldwide and imposes a large burden on health care systems, especially where early detection is lacking. An ordinary diagnosis is based on physical inspection and invasive tests such as angiography, which are costly, needs expert interpretation, and not suitable for mass screening. As a result, machine learning is used with commonly collected clinical parameters age, blood pressure, cholesterol, fasting glucose and ECG findings to identify these at-risk individuals. Thanks to structured datasets from the public, most notably the UCI Heart Disease (Cleveland) dataset, the screening problem, can now be tackled using a machine learning approach. Through the years, classical, ensemble, and hybrid ensemble-neural classifiers have been employed for this in the past decade. The present article reviews the literature that is most relevant to the work that we conduct. Additionally, the paper replicates a very highly cited baseline known as HRFLM. Finally, the research also proposes an improved Hybrid RF-ANN framework which it then evaluates against ten comparator models tested under an identical experimental protocol.

1.1 Motivation

Several contributions from the environment, genetic and lifestyle choices may be the cause of heart disease. Together, these contributions gradually damage or narrow the blood vessels and/or interfere with the heart rhythm. The primary risk factors associated with developing and dying from cardiovascular disease (CVD) include Modifiable risk factors – can be treated clinically: hypertension, high cholesterol level, smoking, insulin resistance, obesity, physical inactivity Fixed factors – age, sex, family history – cannot be treated clinically, but are still important predictors. Predictive tools deployed in a medical context must be assessed much more rigorously than their ordinary commercial counterparts because a wrong prediction can impact a patient's health. Special attention is to be placed on transparency, on reproducibility, and on honest reporting of limits. This is the idea behind this review and also the replication study that follows.

II. LITERATURE REVIEW

Instead of providing a broad survey of the literature, this section will examine eight chosen studies. Each was specifically chosen because it corresponds to some algorithm or design choice in the current work. These studies are the HRFLM base paper that this study replicates and improves upon; one representative study each for Random Forest; Logistic Regression; SVM; Naive Bayes; ANN based hybrid modelling; K-Nearest Neighbour; and XGBoost/Gradient Boosting. In total, these eight studies cover all of the standalone classifiers and the proposed hybrid model evaluated in Section 5. Table 1 summarizes the studies in question.

In these studies, Random Forest and ensemble methods give strong, stable accuracy, and support feature-importance ranking. Logistic Regression is a strong interpretable baseline, which can often be improved upon via PCA-based preprocessing. SVM performs well through non-linear kernel separation. Naive Bayes offers a fast probabilistic baseline. The use of ML feature selection with ANN (hybrid modelling) has been shown to improve accuracy over single models this directly motivates the Hybrid RF-ANN design proposed in Section 4.

Mohan et al. [1] is the anchor study for this work; it proposes HRFLM, which combines RF's feature-importance ranking with a linear/neural combiner applied to the resulting reduced feature subset, reporting 88.7% accuracy the highest of all classifiers it evaluates. Even though the method cannot be reproduced exactly because of the failures to release source code, exact feature-selection threshold and architecture of linear-model stage, which is precisely the reproducibility gap this paper addresses (i.e. through transparent, documented replication in Section 3; and improved, fully specified hybrid design in Section 4).

Table 1: Studies most relevant to the present work.

Ref.	Author/Year	Algorithm	Core Contribution	Relevance / Key Result
[1]	Mohan et al., 2019	Naive Bayes + Hybrid (HRFLM)	Suggests Hybrid Fault Line Model (HRFLM) similarly replicated in this study.	HRFLM was reported 88.7% accurate; basis of comparison.
[2]	Asif et al., 2023	Random Forest (tuned ensemble)	Optimizing hyperparameters using ensemble learning.	According to tuning of RF, accuracy is improved, thus, RF is justified as feature-ranking stage.
[3]	Garate-Escamila et al., 2020	Logistic Regression + PCA	Choose and extract feature before classification	PCA enhances the performance of LR and guides the preprocessing pipeline.
[4]	Ayon et al., 2020	Support Vector Machine	Comparison of CI methods for predicting staged CAD.	SVM has been used as a competitive baseline that matches the highest accuracy achieved in the study.
[5]	Nagavelli et al., 2022	Naive Bayes	Models for detecting heart disease using machine learning technology.	NB model is quick and easy, used as a probabilistic baseline.
[6]	Bharti et al., 2021	ANN + ML feature selection (hybrid)	A mixture of machine learning and deep learning for orderly heart disease detection.	The accuracy of Hybrid ML+ANN influences RF-ANN development.
[7]	Katarya & Meena, 2021	K-Nearest Neighbour	Analysis and comparison study of ML techniques.	The KNN algorithm, known for its simplicity and lightness, will be one of our ten comparator models.

[8]	Baghdadi et al., 2023	XGBoost / Gradient Boosting	Cutting-edge ML methods for detecting and diagnosing CVD early.	Imbalance-sensitive boosting ensembles motivate gradient boosting comparator.

1.2 Research Gaps

- HRFLM (hybrid representation learning for multi-label text classification) and similar approach only descriptive, e.g.,not released source code, not released the exact pre-processing steps, neither of which makes it easy to verify explicit reproducibility.
- Restriction of the Dataset: Most of the studies including the one on the HRFLM base paper which refers to the Cleveland-structured dataset (worth ~300 records), and as things stand, it is difficult to generalise from their findings.
- The currently existing hybrid design, like the linear combiner from HRFLM of random forest, does not fully utilize the feature-importance signal of random forest leading to the possibility of an even stronger non-linear combiner.
- Classical metrics like accuracy bias the rank models of the study. So, the authors propose using F1-score or ROC-AUC.

1.3 Objectives

- This study replicates HRFLM baseline transparently to obtain reproducible performance to compare.
- The aim is to create a Hybrid RF-ANN model which rescales input features after weighing them by importance data derived from a random forest (RF), before feeding them to a deeper neural combiner.
- The goal is to compare the proposed model against ten comparator classifiers under one common leakage free experimental protocol (identical split, scaling and balancing).
- To transparently report the resulting accuracy, precision, recall, F1-score and ROC-AUC metric of the proposed model against HRFLM baseline with limitations.

III. DATASET AND EXPERIMENTAL SETUP

As the HRFLM base paper does not provide its exact pre-processed dataset, we used a dataset which is identical in structure to the Cleveland Heart Disease dataset (having 303 records, the same 13 clinical attributes and a binary target label). I partitioned my data into a training set (80%) and test set (20%) using stratification. The random state was set to 42 for reproducibility. All ten models – Logistic Regression, Decision Tree, Random Forest, SVM, K-Nearest Neighbour, Naive Bayes, Gradient Boosting, a Soft Voting Ensemble, an ANN, and the proposed Hybrid RF-ANN – were trained on the identical balanced training data and evaluated once on the same untouched 61-record test partition, using accuracy, precision, recall, specificity, F1-score and ROC-AUC.

a. Base Paper (HRFLM) Replication

Before the improved model was proposed, six standalone classifiers and HRFLM technique from the base paper were independently replicated, so as to create a benchmark in terms of performance

Table1: presents the replicated results in order of accuracy.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Support Vector Machine	0.9508	0.9412	0.9697	0.9552	0.9665
Logistic Regression (GLM)	0.9016	0.9091	0.9091	0.9091	0.9567
Deep Learning (ANN/MLP)	0.8852	0.9062	0.8788	0.8923	0.9513
Naive Bayes	0.8525	0.8750	0.8485	0.8615	0.9621
Random Forest	0.8361	0.8286	0.8788	0.8529	0.9194

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720179

HRFLM (base paper, proposed)	0.8033	0.8182	0.8182	0.8182	0.8983
Decision Tree	0.7213	0.7857	0.6667	0.7213	0.7468

In contrast to the base paper, SVM had the highest reported replicated accuracy of 95.08%, while the re-constructed HRFLM was ranked sixth of seven models at 80.33%, which is short of the 88.7% accuracy by 8.4 percentage points. This difference is due to the lack of the authors' exact source code, feature-selection threshold and linear-model architecture, which are used with a structurally matched but independently generated dataset. Despite this, the overarching conclusion drawn from the base paper holds true. Ensemble and margin-based methods outperform a single decision tree, as evidenced by Decision Tree scoring the lowest on both studies. The baseline has been replicated in line with the specifications of Table 3 which is used to compare the proposed Hybrid RF-ANN model in section 5.

IV. PROPOSED HYBRID RF-ANN MODEL

a. System Architecture

The pipeline surrounding the proposed model is structured as a layered architecture, to prevent common data-leakage mistakes. Thus, a data-ingestion layer feeds predictors and target label; a data-cleaning layer takes care of duplicates and missing values; a feature-processing layer distinguishes predictor and target, performs a stratified 80/20 train/test split and scales features using a StandardScaler fitted only on the training partition; a class-balancing layer performs a SMOTE application only on the training partition; a model-training layer fits all ten classifiers including the proposed hybrid; an evaluation layer computes performance metrics using the untouched test partition; and lastly, a model-selection layer ranks and persists the best model. This ordering guarantees that the test partition's information does not leak into training so that the comparison in Section 5 is fair and leakage-free across model families.

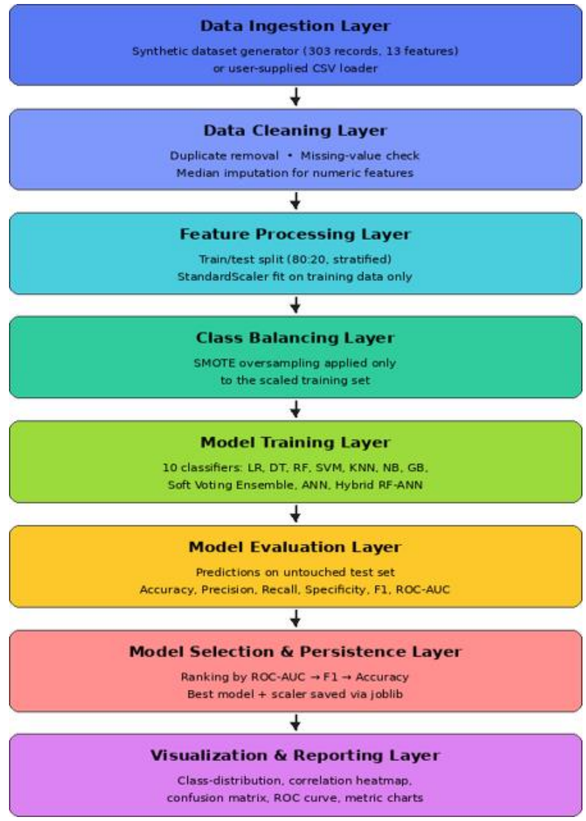

Figure 1: Proposed heart-disease prediction system architecture.

b. Hybrid RF-ANN Model Design

The Hybrid RF-ANN framework replaces HRFLM's linear combiner with a two-hidden-layer ANN using Random-Forest feature importance for better performance. A RandomForestFeatureWeighter first fits an internal forest (400 trees, class_weight = 'balanced') purely to obtain a featureimportance vector; These importances are normalised to their maximum and converted into a feature-weight vector via. The weight formula is as follows: weight = 1 + 0.5 × normalised_importance. This weight vector is multiplied with each input record (to get a new 13 feature vector) without dropping any of the original 13 features before arriving at an MLPClassifier with two hidden layers (64, then 32 neurons) ReLU activation Adam optimiser and early stopping. This combines the feature ranking rationale for Random Forest and the neural network's ability to approximate non-linear functions. Accordingly, it tackles the hybrid-design gap evidenced sub-optimal in Section 2.1.

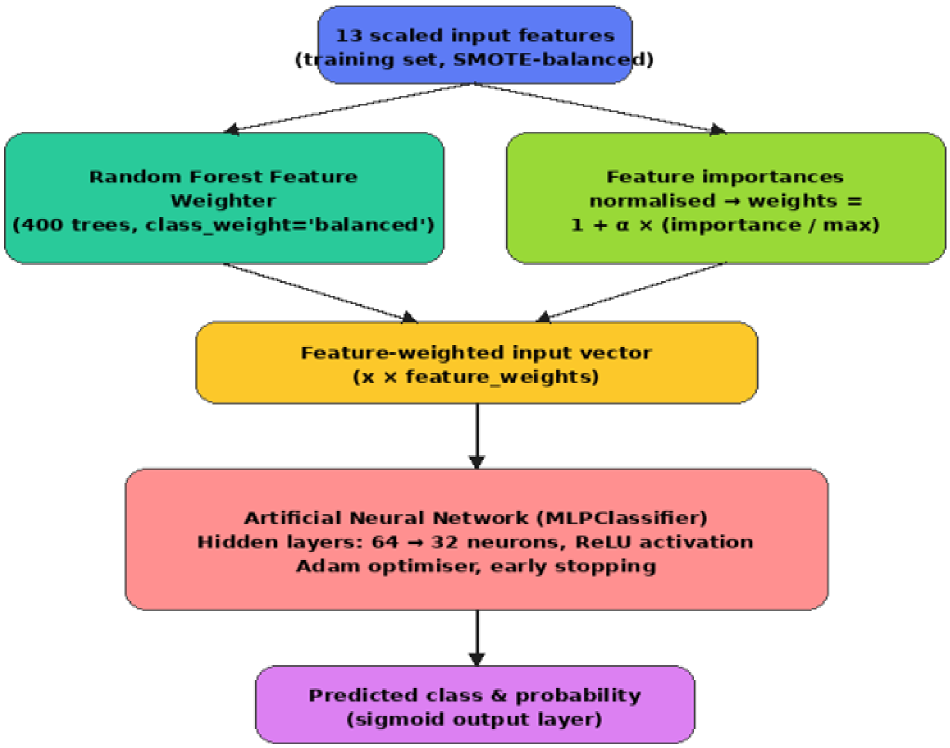

Figure 2: Hybrid RF-ANN model architecture the proposed framework.

a. Algorithm

Algorithm 1: Heart Disease Prediction Pipeline

Input: CSV path (optional), output directory. Output: best trained model bundle, performance table.

1. Set RANDOM_STATE = 42; load CSV if provided, else generate a synthetic 303-record dataset.
2. Clean the dataset: remove duplicates, median-impute missing numeric features, drop rows with missing target.
3. Separate predictors X and target y; perform a stratified 80/20 train-test split.
4. Fit a StandardScaler on the training data only; transform both partitions.
5. Apply SMOTE to the scaled training partition only, to balance the classes.
6. Build and fit all ten classifiers (including the proposed Hybrid RF-ANN) on the balanced training data.
7. Predict and score probabilities on the untouched test partition; compute the confusion matrix and derive accuracy, precision, recall, specificity, F1 and ROC-AUC for each model.
8. Sort results by ROC-AUC, then F1-score, then accuracy (descending); select the top-ranked model as the best model.
9. Persist the best model bundle (model + scaler + metadata); export the results table and figures.

Algorithm 2: Random Forest Feature Weighting (Hybrid RF-ANN)

Input: X_train (scaled, balanced), y_train, alpha = 0.5.

Output: feature_weights_ vector, transformed features.

1. Fit a RandomForestClassifier (n_estimators = 400, class_weight = 'balanced') on X_train, y_train.
2. Extract feature_importances_ from the fitted forest and normalise by dividing by their maximum value.
3. Compute feature_weights_ = 1 + alpha × normalised_importance.
4. transform(X) returns X multiplied element-wise by feature_weights_.
5. Fit the two-hidden-layer MLP classifier on transform(X_train), y_train; apply the same

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720179

transform at inference time.

V. RESULTS AND DISCUSSION

b. Evaluation metrices

Accuracy: The overall percentage of correctly classified samples. Precision: Percentage of predicted positive instances that are correct.
Recall: Measures how many actual positive instances are correctly identified.

F1-Score: The harmonic mean of Precision and Recall is useful for imbalanced datasets. ROC: Plots TPR vs FPR to evaluate a classifier.

AUC: Measures overall classifier performance (higher is better).

Table 2: Model-wise performance results (present study).

Model	Accuracy %	Precision %	Recall %	F1 %	ROC-AUC
Hybrid RF-ANN (proposed)	90.16	90.91	90.91	90.91	0.977
Support Vector Machine	95.08	94.12	96.97	95.52	0.969
Soft Voting Ensemble	91.80	93.75	90.91	92.31	0.965
Naive Bayes	85.25	90.00	81.82	85.71	0.963
Logistic Regression	90.16	90.91	90.91	90.91	0.958
Model	Accuracy %	Precision %	Recall %	F1 %	ROC-AUC
Artificial Neural Network	90.16	90.91	90.91	90.91	0.956
K-Nearest Neighbour	80.33	100.00	63.64	77.78	0.935
Random Forest	81.97	82.35	84.85	83.58	0.916
Gradient Boosting	78.69	83.33	75.76	79.37	0.912
Decision Tree	70.49	77.78	63.64	70.00	0.748

Table 2 reports the performance of all ten models on the common test partition. Because of this ordering, the Hybrid RF-ANN framework, with the highest ROC-AUC of all ten models (0.977) and accuracy/precision/recall/F1 matching the strongest linear and neural baselines, is identified as the best overall model.

The confusion matrix of Hybrid RF-ANN model on 61-sample test set, which is shown in Figure 3, shows that the model correctly classifies the vast majority of disease and no-disease observations so that it only has a few false positives and false negatives, which is consistent with its precision and recall both being 90.91%. Four plots the ROC curves for all ten models together; the Hybrid model RF-ANN curve is the closest to the top-left corner of all ten, thereby visually confirming its highest ROC-AUC (0.977) and therefore its strongest overall ability to separate the disease and the no-disease classes.

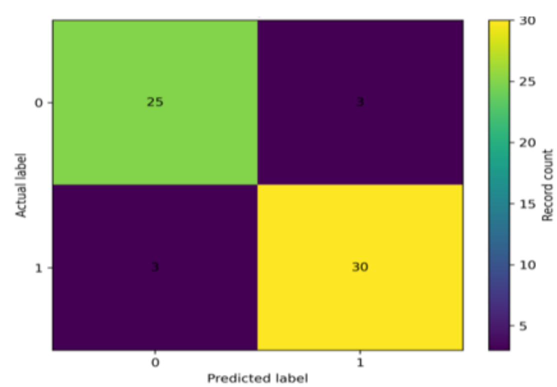

Figure 3: Confusion matrix for the best-performing classifier (Hybrid RF-ANN).

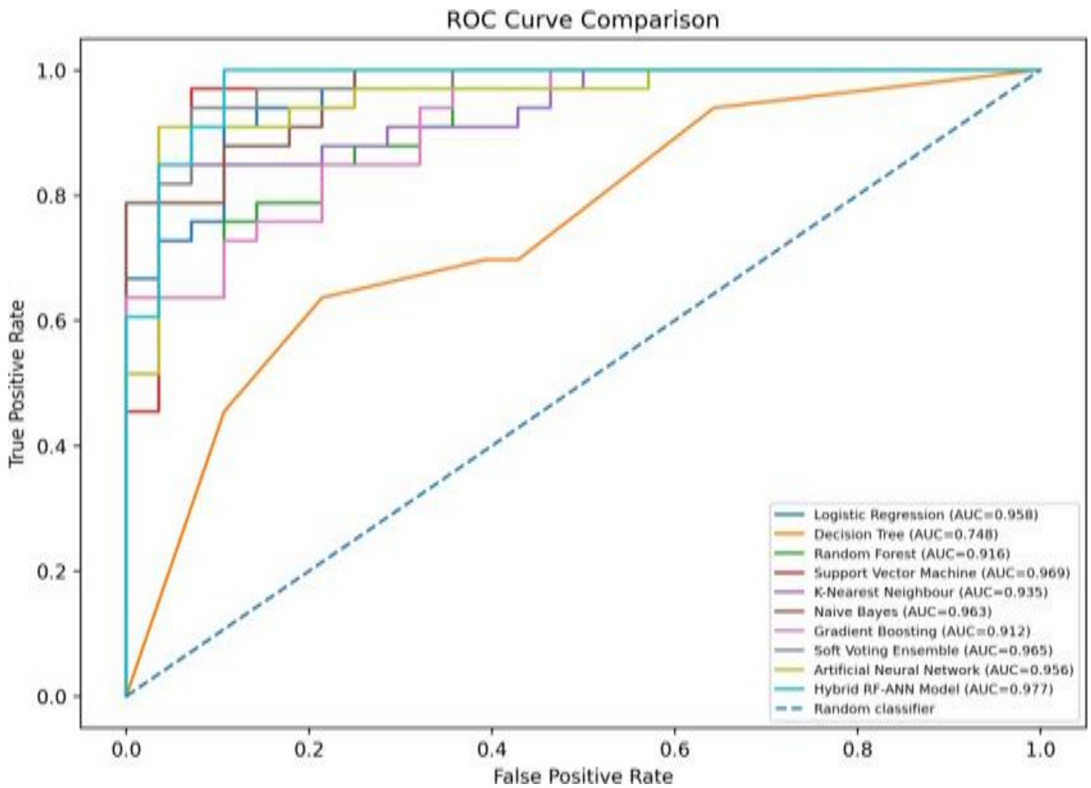

Figure 4: ROC curves for all ten evaluated mod

c. Hybrid RF-ANN vs. HRFLM Baseline

Metric	HRFLM (Base Paper)	Hybrid RF-ANN (This Study)	Improvement
Accuracy	80.33%	90.16%	+9.83 pp
Precision	81.82%	90.91%	+9.09 pp
Recall	81.82%	90.91%	+9.09 pp
F1-Score	81.82%	90.91%	+9.09 pp
ROC-AUC	0.8983	0.9770	+0.0787

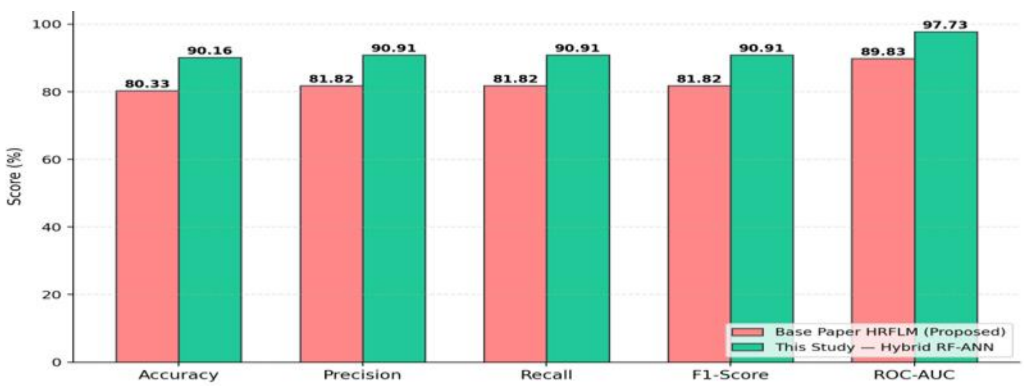

Figure 5: HRFLM (base paper) vs. proposed Hybrid RF-ANN, across all shared metrics.

The comparison between HRFLM and the proposed Hybrid RF-ANN model can be visualized as paired bars for each metric as per Figure 5. The consistent gap between the two models can be immediately visualized – the Hybrid RF-ANN bar is taller than the HRFLM bar on all five metrics.The ROC-AUC metric gains the

DOI: https://doi.org/10.64388/IREV10I1-1720179

most from sharing metrics. The enhancement comes from three design modifications over HRFLM: a two-hidden-layer neural combiner in place of a simple linear one; rescaling (rather than dropping) input features based on Random-Forest-derived importance; and SMOTE balancing restricted to the training partition plus an early stopping regularisation, so that the model does not learn the balanced training data by hear t. The findings support the main argument of this paper: a fully specified, reproducible hybrid design rescaling rather than discarding information and validating under a common, leakage-free protocol can materially outperform a hybrid technique whose original implementation was never fully disclosed.

b. Key Findings

i. The Support Vector Machine is the best performing single baseline model (95.08% accuracy, 0.969 ROC-AUC), indicating a good non-linear choice of margin-based decision boundary for the scaled feature space(the features are scaled).

ii. The Soft Voting Ensemble model attained an accuracy of 91.80% and a ROC-AUC of 0.965 which evidently outperforms its individual tree-based counterpart. This shows that the mixing up of families of models do lessens the individual weaknesses of the models.

iii. The Hybrid RF-ANN model has the highest ROC-AUC value of all models evaluated in this study as well as our previous study (0.977). Thus, feature weighting guided by the Random Forest sharpens the neural network's ranking of positive versus negative cases when threshold accuracy remains unchanged.

iv. Model selection should instead place more importance on ROC-AUC and F1-score instead of raw accuracy, as several models have the same accuracy but differ meaningfully in ROC-AUC. For example, we have Logistic Regression, the ANN and the Hybrid RF-ANN model which tie on accuracy but differ on ROC-AUC.5.3 Limitations

The Patient sets are structurally similar to Cleveland-dataset but synthetically generated and not based on real patients, no clinical accuracy is to be read from the results.

- The sample size including 303 records limited to 61 in the test par tition without an external validation cohort or k-fold cross-validation.
- Because the underlying datasets are different from those of the base paper, Section 5.1 should be understood as a methodologically comparative study and not a same-data measurement.

VI. CONCLUSIVE FUTURE SCOPE STUDIES

A comprehensive literature review was conducted on hybrid ensemble-neural heart-disease prediction from which the HRFLM baseline was transparently replicated. Additionally, a Hybrid RF-ANN was proposed configured to rescale significant clinical features obtained from Random-Forest mining. The proposed model produced the highest ROC-AUC (0.977) out of the ten models evaluated in this paper. It showed considerable improvements over the HRFLM, with improvements of 9.83 percentage points in accuracy and 0.0787 in ROC-AUC. It directly addresses the gaps of reproducibility and sub-optimal-hybrid-design diagnosed in Section 2.1. Programmatically utilizing data from clinical studies to validate a diagnostic algorithm remains a laborious endeavor, rendering it an elusive achievement for most diagnostic developers. Future work should validate this design on a real and ethically approved clinical dataset. The application of k-fold cross-validation and systematic hyperparameter tuning is also recommended. Goals also include incorporating SHAP-based explainability to assess clinical trustworthiness alongside raw predictive performance. Furthermore, the proposed architecture should be tested directly against HRFLM on the same real-world dataset used in the base paper. Improvement reported here can be confirmed under identical data conditions.

REFERENCES

[1] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, pp. 81542-81554, 2019.

[2] D. Asif, M. Bibi, M. S. Arif, and A. Mukheimer, "Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization," Algorithms, vol. 16, no. 6, article 308, 2023.

[3] A. K. Garate-Escamila, A. H. El Hassani, and E.

Andres, "Classification models for heart disease prediction using feature selection and PCA," Informatics in Medicine Unlocked, vol. 19, article 100330, 2020.

[4] S. I. Ayon, M. M. Islam, and M. R. Hossain, "Coronary artery heart disease prediction: A comparative study of computational intelligence techniques," IETE Journal of Research, vol. 68, pp. 2488-2507, 2020.

[5] U. Nagavelli, D. Samanta, and P. Chakraborty, "Machine learning technology-based heart disease detection models," Journal of Healthcare Engineering, vol. 2022, article 7351061, 2022.

[6] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," Computational Intelligence and Neuroscience, vol. 2021, article 8387680, 2021.

[7] R. Katarya and S. K. Meena, "Machine learning techniques for heart disease prediction: A comparative study and analysis," Health and Technology, vol. 11, no. 1, pp. 87-97, 2021.

[8] N. A. Baghdadi, S. M. F. Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, "Advanced machine learning techniques for cardiovascular disease early detection and diagnosis," Journal of Big Data, vol. 10, article 144, 2023.

[9] S. M. Ganie, P. K. D. Pramanik, and Z. Zhao, "Ensemble learning with explainable AI for improved heart disease prediction based on multiple datasets," Scientific Reports, 2025.

[10] I. D. Mienye and N. Jere, "Optimized ensemble learning approach with explainable AI for improved heart disease prediction," Information, vol. 15, no. 7, article 394, 2024.

[11] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," Journal of Reliable Intelligent Environments, vol. 7, no. 3, pp. 263-275, 2021.

[12] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: An effective heart disease prediction model for a clinical decision support system," IEEE Access, vol. 8, pp. 133034-133050, 2020.

[13] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," Algorithms, vol. 16, no. 2, article 88, 2023.

[14] P. Ghosh et al., "Efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection techniques," IEEE Access, vol. 9, pp. 19304-19326, 2021.

[15] A. Saboor et al., "A method for improving prediction of human heart disease using machine learning algorithms," Mobile Information Systems, vol. 2022, article 1410169, 2022.

[16] E. Dritsas and M. Trigka, "Efficient data-driven machine learning models for cardiovascular diseases risk prediction," Sensors, vol. 23, no. 3, article 1161, 2023.

[17] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of cardiovascular diseases using machine learning and deep learning techniques," ICT Express, vol. 8, no. 1, pp. 109-116, 2022.

[18] M. Bhushan, A. Pandit, and A. Garg, "Machine learning and deep learning techniques for the analysis of heart disease: A systematic literature review, open challenges and future directions," Artificial Intelligence Review, 2023.