Proceedings of the Second International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI-2026)
IEEE Xplore Part Number: CFP26VH0-ART; ISBN: 979-8-3315-6898-6

Scopus
J. Anu 18/4/26

A Cost-Aware Stacking Ensemble Framework for Early Heart Disease Detection

R. Devi 18/4/26
[Dr. R. Devi Professor & Head DACE]

Jhansi B
Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India
jhansisathish.bcn@gmail.com
ORCID: 0009-0004-5243-4422

Dr. T. Kamalakannan
Department of DACE, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India
kkannan.scs@velsuniv.ac.in
ORCID: 0000-0002-5982-8983

Yet to Index in Scopus

***Abstract*—Heart disease remains a leading cause of mortality worldwide, emphasizing the need for reliable and early screening mechanisms. Although machine learning techniques have been widely explored for heart disease prediction, many existing approaches primarily optimize overall accuracy and provide limited attention to misclassification cost asymmetry and interpretability, which are critical for clinical screening applications. This paper proposes a cost-aware stacking ensemble framework for early heart disease detection that integrates heterogeneous base classifiers, optimized decision thresholding, and explainable artificial intelligence techniques within a unified pipeline. Logistic regression, support vector machine, and random forest models are employed as base learners, and their outputs are combined using a logistic regression–based meta-learner. Cost-sensitive learning is incorporated to prioritize reducing false negatives, while the final classification threshold is optimized using receiver operating characteristic analysis. Model transparency is enhanced through Shapley value–based local explanations and permutation-based global feature importance analysis. Experimental evaluation on the UCI Heart Disease dataset using stratified cross-validation demonstrates that the proposed framework achieves an accuracy of 87.17%, a sensitivity of 0.92, a specificity of 0.84, and an area under the ROC curve of 0.87. The results indicate that the proposed approach provides a robust and interpretable decision-support system suitable for early heart disease screening.**

Keywords*—*clinical decision support, cost-sensitive learning, explainable artificial intelligence, Heart disease prediction, stacking ensemble.

I. Introduction

Cardiovascular diseases remain one of the leading causes of mortality worldwide, accounting for a significant proportion of premature deaths and long-term disability. Among these, heart disease poses a major public health challenge due to its multifactorial nature and often asymptomatic progression during early stages. Timely identification of individuals at risk is therefore critical for initiating preventive interventions and reducing adverse clinical outcomes. Conventional diagnostic approaches rely heavily on clinical expertise, laboratory investigations, and imaging procedures, which can be costly, invasive, and time-consuming, particularly in large-scale screening scenarios [1].

The increasing availability of electronic health records and structured clinical datasets has enabled the application of machine learning techniques to assist in heart disease prediction. By learning patterns from historical patient data, machine learning models can help clinicians more efficiently and consistently identify high-risk individuals. Early studies employing traditional classifiers such as logistic regression, decision trees, and support vector machines demonstrated the feasibility of data-driven heart disease prediction [2][3].

However, the predictive capability of single models is often constrained by limited generalizability, sensitivity to data characteristics, and an inability to capture the complex nonlinear relationships inherent in clinical data. To address these limitations, ensemble learning methods have been increasingly adopted in cardiovascular disease prediction. Techniques such as random forests and boosting algorithms combine multiple learners to improve robustness and classification performance [4]. While ensemble approaches generally outperform individual classifiers, many existing studies focus predominantly on maximizing overall accuracy. In medical screening contexts, such an objective is insufficient, as false negatives—cases where diseased patients are incorrectly classified as healthy—can have serious clinical consequences. Consequently, models optimized solely for accuracy may fail to meet the practical requirements of early disease detection.

Another important limitation of many existing machine learning–based diagnostic systems is the lack of interpretability. Complex ensemble and nonlinear models are often treated as black boxes, providing limited insight into the factors driving their predictions. This lack of transparency poses a significant barrier to clinical adoption, where trust, accountability, and alignment with medical knowledge are essential [5]. Recent advances in explainable artificial intelligence have sought to address this issue by introducing methods such as Shapley value–based explanations, which quantify feature contributions to individual predictions [6]. However, explainability is often used as a post hoc analysis rather than integrated into the model development process.

More recently, stacking ensemble learning has emerged as a powerful technique that combines the strengths of multiple base classifiers through a meta-learning framework. By learning how to optimally fuse model outputs, stacking ensembles have demonstrated improved generalization performance in various healthcare applications [7][8]. Despite their potential, existing stacking-based heart disease prediction models often neglect clinically relevant considerations such as misclassification cost asymmetry and decision threshold optimization. Furthermore, the integration of stacking ensembles with explainability mechanisms remains limited in current literature.

Motivated by these observations, this work proposes a cost-aware and explainable stacking ensemble framework for early heart disease screening. The proposed approach

2026 Second International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI) | 979-8-3315-6898-6/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICMSCI67830.2026.11469762

VELS

INSTITUTE OF SCIENCE, TECHNOLOGY & ADVANCED STUDIES (VISTAS)
(Deemed to be University Estd. u/s 3 of the UGC Act, 1956)
PALLAVARAM, THALAMBUR, PERIYAPALAYAM-CHENNAI
ACCREDITED BY NAAC WITH 'A++' GRADE

DETAILS OF PH.D PROGRAMME OF ....Comp. Sc....

Name: ....Mrs B. Jhansi.... Year: ....

Payment of Research Fees	I	II	III	IV	V	VI
	✓	✓	✓			

Date of Registration	2/9/2023		FT	PT
M.Phil qualified	Yes ✓		No	
Minimum period of Research				
Date of First RAC / Minutes	10/2/2024			
Six Monthly Report 1	Sep'23 - Feb 24			
Research Methodology: Paper I	I -	II -	III -	IV -
Confirmation sent date				
Date of Second RAC / Minutes	12/11/2024			
Six Monthly Report 2	Mar 24 - Aug 24			
Date of Third RAC / Minutes	20/5/2025			
Six Monthly Report 3	Sep'24 - Feb 25			
Date of Fourth RAC / Minutes				
Six Monthly Report 4	Mar'25 - Aug 25			
Date of Fifth RAC / Minutes				
Six Monthly Report 5	Sep'25 - April '26			
Date of Sixth RAC /Minutes				
Six Monthly Report 6				
Six Monthly Report 7				
Six Monthly Report 8				
Six Monthly Report 9				
Six Monthly Report 10				
Six Monthly Report 11				
Six Monthly Report 12				
No. of Publications	2 publications (Scopus indexed).			
Synopsis Meeting				
Date of Submission of Synopsis				
Date of Submission of Thesis				