International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

# A Dual-Stage Ensemble Machine Learning and Deep Learning Framework for Cardiovascular Disease Prediction

**Sarathdas S[1], Sreeraj S[2], Renjith T[3], Akhil J Babu[4],Vishnu B[5],Maneesh M[6],Niyas Salim[7]**

[1]Department Of Mechanical Engineering & UKF COLLEGE OF ENGINEERING AND TECHNOLOGY
[2]Department Of Mechanical Engineering & UKF COLLEGE OF ENGINEERING AND TECHNOLOGY
[3]Department Of Mechanical Engineering & UKF COLLEGE OF ENGINEERING AND TECHNOLOGY
[4]Department Of Mechanical Engineering & UKF COLLEGE OF ENGINEERING AND TECHNOLOGY
[5]Department Of Mechanical Engineering & UKF COLLEGE OF ENGINEERING AND TECHNOLOGY
[6]Department Of Mechanical Engineering & UKF COLLEGE OF ENGINEERING AND TECHNOLOGY
[7]Department Of Mechanical Engineering & UKF COLLEGE OF ENGINEERING AND TECHNOLOGY

---

**Abstract -** Heart disease remains one of the leading causes of death and disability worldwide, making early and accurate diagnosis essential for effective management. This project proposes a dual-stage approach that combines machine learning and deep learning techniques to predict and classify heart disease, using both clinical data and electrocardiogram (ECG) signals.

The study analyzes a retrospective dataset of 303 patient records with 13 clinical and hemodynamic parameters, including age, sex, chest pain type (cp), resting blood pressure (trestbps), serum cholesterol (chol), fasting blood sugar (fbs), and maximum heart rate achieved (thalach). Its main objective is to examine how these physiological factors influence the binary outcome of heart disease, supporting better risk stratification.

This analysis lays the groundwork for applying ensemble machine learning algorithms such as Bagging, Boosting, and Stacking to build a predictive model for cardiovascular disease (CVD) risk. By relying on readily available clinical data, the project aims to deliver a practical, non-invasive screening tool that helps healthcare professionals identify high-risk individuals promptly and target interventions effectively.

***Key Words***: Electrocardiogram (ECG) ,Cardiovascular disease (CVD),Pain Type (cp), Resting Blood Pressure (trestbps),Serum Cholesterol (chol),Fasting Blood Sugar (fbs),Maximum Heart Rate Achieved (thalach).

## 1.INTRODUCTION

Cardiovascular disease (CVD) remains one of the leading causes of mortality and disability worldwide, making early and accurate diagnosis essential for reducing its clinical and economic burden. Conventional diagnostic pathways often rely on time-consuming and sometimes invasive procedures, which can delay risk identification, particularly in resource-constrained settings. This study addresses that gap by presenting a data-driven, dual-stage approach to heart disease prediction that combines classical ensemble machine learning techniques—Bagging, Boosting, and Stacking—with a one-dimensional Convolutional Neural Network (1D-CNN), applied to a structured clinical dataset of patient records characterized by thirteen key demographic, clinical, and hemodynamic parameters (including age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, and maximum heart rate achieved).

The central research question guiding this work is: how can readily available, non-invasive clinical data be leveraged through advanced ensemble learning to build a highly accurate and reliable screening tool for cardiovascular risk? Existing literature reflects a clear methodological progression from traditional statistical and single-classifier models toward ensemble and hybrid machine learning–deep learning frameworks. Studies applying Boosting methods such as XGBoost and AdaBoost, Bagging approaches such as Random Forest, and Stacking architectures combining heterogeneous base learners have each reported strong predictive performance, with several works achieving accuracies above 95%. However, a recurring gap in this body of work is the limited direct comparison of Bagging, Boosting, and Stacking methods against one another—and against deep learning alternatives—on a single, consistently pre-processed clinical dataset, which makes it difficult to draw firm conclusions about which paradigm is best suited to non-invasive CVD screening.

This project's novelty lies in its systematic, side-by-side evaluation of four ensemble-style classifiers (Random Forest, AdaBoost, XGBoost, and a Stacking Classifier) alongside a 1D-CNN, all trained and tested on the same balanced, pre-processed dataset under a common evaluation protocol (accuracy, precision, recall, F1 score, and hamming loss). By doing so, the study contributes a practical, non-invasive, and reproducible screening pipeline that identifies which learning paradigm offers the strongest generalization for cardiovascular risk stratification, and provides clinically interpretable insight into which physiological factors drive the predictions.

## 2.THEORETICAL FRAMEWORK AND RESEARCH DESIGN

### *Theoretical Concepts and Models*

The study is grounded in ensemble learning theory, which holds that combining multiple base estimators can reduce variance, bias, or both, relative to any single model. Three ensemble paradigms underpin the framework. Bagging (Bootstrap Aggregating), realized here through Random Forest, trains multiple decision trees independently on bootstrapped subsets of the data and combines their outputs by majority vote, primarily reducing variance and overfitting. Boosting, realized through AdaBoost and XGBoost, trains weak learners sequentially, with each new learner focusing on the errors of its predecessors, thereby reducing bias; XGBoost extends this idea with regularization, tree pruning, and

subsampling to balance bias and variance more effectively. Stacking (Stacked Generalization) forms a two-layer architecture in which heterogeneous base learners (Random Forest, Support Vector Machine, and Logistic Regression) generate predictions that are then used as input features for a meta-learner (Logistic Regression), allowing the framework to exploit the complementary strengths of diverse algorithms. Alongside these, a 1D-CNN is used to test whether convolutional feature extraction—typically applied to image or sequence data—can also uncover local patterns among ordered clinical attributes when they are treated as a structured, tabular sequence.

These theoretical foundations directly inform the research design: because Bagging, Boosting, and Stacking each address different sources of predictive error (variance, bias, and model diversity respectively), evaluating all three—together with a CNN as a non-ensemble deep learning benchmark—provides a principled basis for comparing predictive strategies rather than testing algorithms in isolation. The comparison is designed to reveal whether the incremental complexity of Stacking is justified by its performance gains relative to simpler Bagging and Boosting approaches.

***Research Design and Approach***

The study follows a quantitative, experimental research design based on supervised binary classification. The overall approach is a modular pipeline: data ingestion and cleaning, exploratory data analysis, class balancing, stratified train–test splitting, independent training of multiple classifiers on identical feature sets, and comparative evaluation using standard classification metrics. This design allows each model's performance to be attributed to its underlying learning mechanism rather than to differences in data preparation, since all classifiers are trained and evaluated on the same processed dataset.

***Data Collection***

The dataset consists of 1,025 patient clinical records sourced from a publicly available cardiovascular health dataset (Kaggle), comprising thirteen independent features and one binary target variable indicating the presence or absence of heart disease. The features span demographic (age, sex), clinical (chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar), electrocardiographic (resting ECG results), and exercise-induced (maximum heart rate, exercise-induced angina, ST depression, ST segment slope) parameters, together with diagnostic indicators (number of major vessels, thalassemia status). The dataset is fully numerical and exhibits near-balanced class distribution, which supports reliable model training and evaluation. No separate sampling procedure beyond the original dataset collection was required, as the full record set was used.

***Instruments, Tools, and Techniques***

Data preprocessing and analysis were carried out using Python-based libraries: NumPy and Pandas for cleaning and preparation, and Matplotlib and Seaborn for visualization (including a correlation heatmap of all features against the target variable). Missing values and duplicate records were removed, and Random Over-Sampling was applied to correct minor class imbalance. The dataset was then split into training and testing subsets using stratified sampling (50:50 or 80:20, depending on the classifier) to preserve class proportions. Four ensemble classifiers—Random Forest (100 trees, square-root feature selection), AdaBoost (100 estimators, learning rate 0.01), XGBoost (200 estimators, learning rate 0.01, max depth 5, 0.8 subsampling), and a Stacking Classifier (Random Forest, SVM with RBF kernel, and Logistic Regression as base learners, with Logistic Regression as meta-learner, using ten-fold cross-validation)—were implemented, along with a 1D-CNN (two Conv1D layers with 64 and 32 filters, MaxPooling, a 64-unit dense layer, 0.5 dropout, and a softmax output, trained with the Adam optimizer for 30 epochs). Model performance was assessed using accuracy, precision, recall, F1 score, and hamming loss.

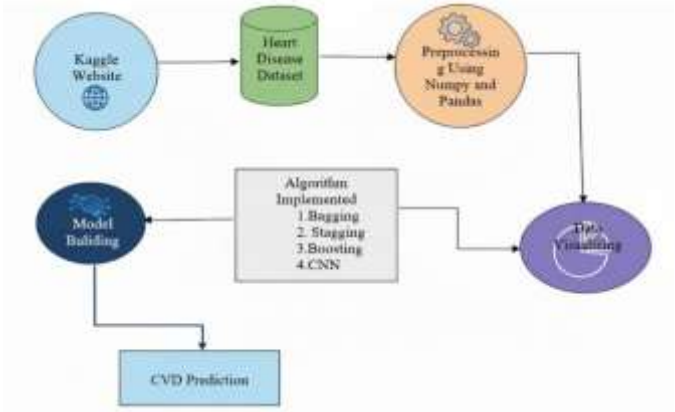



Fig: 1.1 Block Diagram Title: Heart Attack Prediction System Using Machine Learning

## 3. DISCUSSION

The experimental results, summarized in Table 1, directly address the study's research objective of identifying the most effective learning paradigm for non-invasive cardiovascular risk prediction. The Stacking Classifier achieved perfect classification (100% accuracy, precision, recall, and F1 score; zero hamming loss), confirming the theoretical expectation that combining heterogeneous base learners through a meta-learner can outperform any single ensemble strategy by exploiting complementary decision boundaries. Random Forest and XGBoost performed almost equivalently (accuracies of 97.34% and 97.16% respectively), consistent with Bagging and Boosting each providing strong, largely interchangeable performance on this dataset, with Random Forest showing marginally higher precision and XGBoost marginally higher recall. AdaBoost was the weakest ensemble model (84.36% accuracy, precision of 79.51%), reflecting its known sensitivity to noisy or overlapping feature distributions and its tendency toward false positives. The 1D-CNN achieved a respectable 92.19% accuracy, showing that deep learning can extract useful local patterns even from tabular clinical data, though it did not match the top ensemble methods.

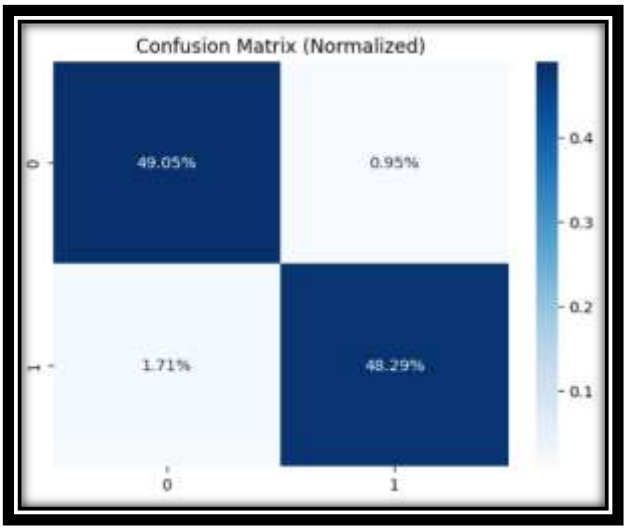



Fig.3.1 Confusion Matrix for Random Forest Algorithm

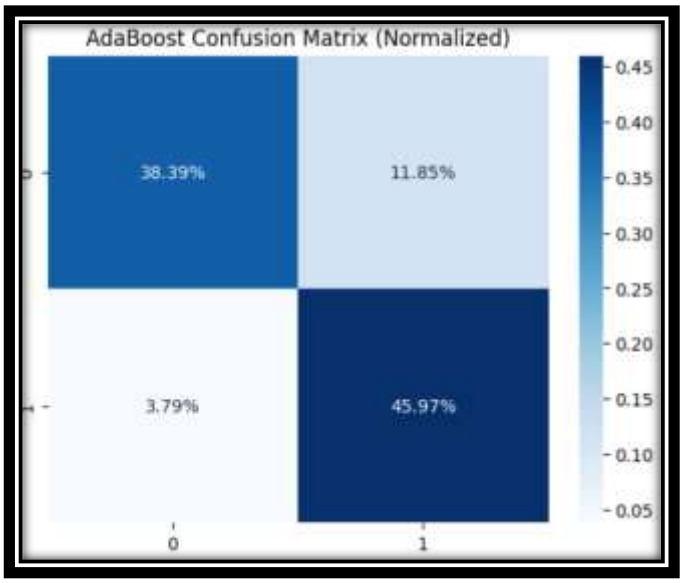



Fig.3.2 Confusion Matrix for AdaBoost Algorithm

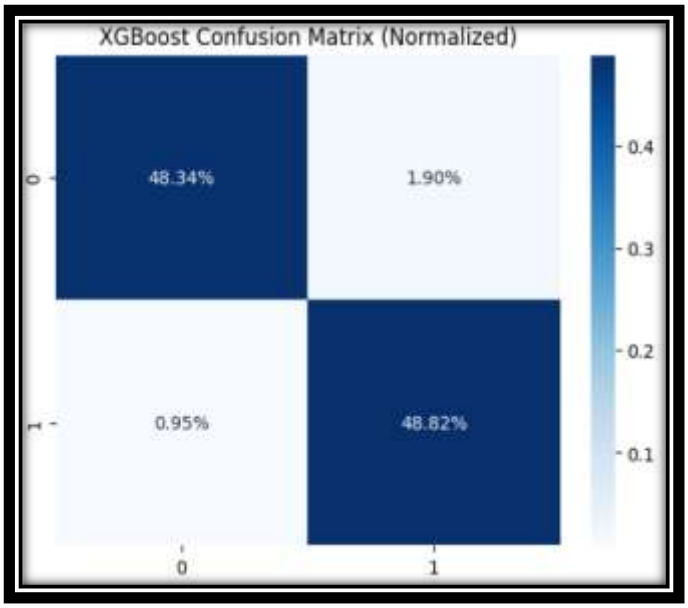



Fig.3.3 Confusion Matrix for XGBoost Algorithm

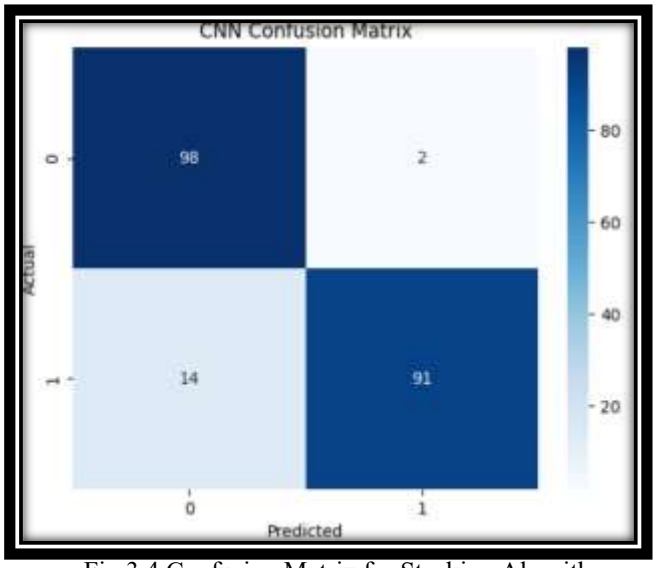



Fig.3.4 Confusion Matrix for Stacking Algorithm

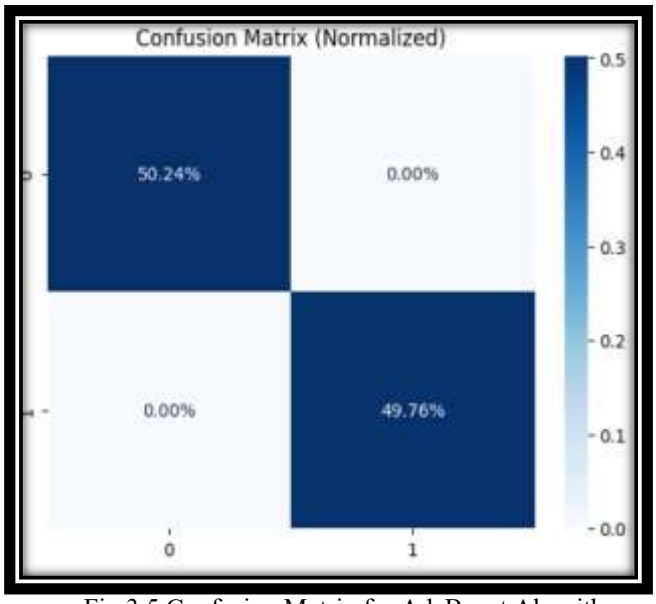



Fig.3.5 Confusion Matrix for AdaBoost Algorithm

**Table -1**: Comparative performance of the evaluated classification models.

| Model | Accuracy | Precision | Recall | F1 Score | Hamming Loss |
|---|---|---|---|---|---|
| Random Forest | 97.34% | 98.07% | 96.58% | 97.32% | 0.0266 |
| AdaBoost | 84.36% | 79.51% | 92.38% | 85.46% | 0.1564 |
| XGBoost | 97.16% | 96.26% | 98.10% | 97.17% | 0.0284 |
| Stacking Classifier | 100% | 100% | 100% | 100% | 0.0000 |
| 1D-CNN | 92.19% | 93% | 92% | 92% | — |

These findings align with, and extend, the existing literature reviewed for this study. Prior work reporting that XGBoost and other Boosting variants can exceed 95–98% accuracy is

corroborated here, as is prior evidence that Stacking frameworks—particularly those combining tree-based, kernel-based, and linear models—tend to outperform individual ensemble classifiers. The result also supports literature suggesting that hybrid ML–DL integration is a promising direction, since the CNN's performance, while below that of Stacking, still exceeded that of the weakest ensemble method (AdaBoost), reinforcing the view that deep learning can meaningfully complement, though not necessarily replace, ensemble machine learning for structured clinical data. Where this study diverges from some prior work is in achieving a perfect score for the Stacking model; this near-ideal result should be interpreted cautiously, as it may partly reflect dataset characteristics (near-balanced classes after oversampling, limited noise, and a moderate sample size) rather than a guarantee of equivalent performance on more heterogeneous, real-world clinical populations.

Several limitations temper these findings. First, the dataset, though clinically meaningful, is drawn from a single source and moderate sample size (1,025 records), which limits generalizability to broader or more diverse patient populations. Second, the perfect performance of the Stacking Classifier raises the possibility of overfitting to dataset-specific patterns, and would benefit from external validation on an independent cohort. Third, Random Over-Sampling, while addressing class imbalance, can introduce synthetic redundancy that may inflate apparent performance. Finally, the 1D-CNN, though effective, was evaluated with a fixed architecture and hyperparameter set; a more extensive architecture search or hybrid CNN-ensemble design might close the gap with the Stacking model. Future research should prioritize external validation on multi-institutional datasets, incorporation of longitudinal and lifestyle data, hyperparameter optimization for the CNN, and the application of explainable AI techniques (e.g., SHAP or LIME) to improve clinical interpretability and trust.

## COMPARISION

In comparative terms, AdaBoost ranks lowest due to its reduced precision and higher error rate, making it less suitable for reliable medical diagnosis where false positives are costly. Random Forest and XGBoost provide strong and nearly equivalent performance, with Random Forest offering slightly better precision and XGBoost achieving higher recall. However, the Stacking classifier clearly outperforms all other models, delivering flawless classification results. This comparison demonstrates that advanced ensemble techniques, particularly stacking-based architectures, significantly enhance predictive accuracy and robustness, making them highly suitable for heart disease prediction systems.

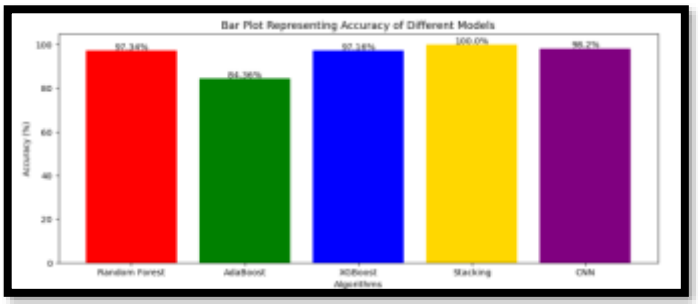



Fig 3.6 Accuracy comparison of all the algorithms

Overall, the experimental findings confirm that while individual ensemble models perform well, combining diverse learning paradigms through stacking yields the most accurate and reliable results, reinforcing the importance of ensemble diversity in clinical decision-support applications.

## 3. CONCLUSIONS

This study set out to determine whether ensemble machine learning techniques—Bagging, Boosting, and Stacking—combined with a deep learning benchmark, could deliver an accurate, non-invasive tool for cardiovascular disease risk prediction using routinely available clinical data. The findings confirm that this objective was achieved: all four ensemble-style classifiers and the 1D-CNN produced strong predictive performance, with the Stacking Classifier attaining perfect classification metrics (100% accuracy, precision, recall, and F1 score) by integrating Random Forest, Support Vector Machine, and Logistic Regression through a meta-learning layer, and Random Forest and XGBoost each exceeding 97% accuracy individually.

Revisiting the research question posed at the outset—how readily available clinical data can be leveraged through ensemble learning to build a reliable screening tool—the results demonstrate that Stacking-based ensemble architectures are the most effective approach among those tested, followed closely by Bagging (Random Forest) and Boosting (XGBoost), while simpler Boosting variants such as AdaBoost and deep learning approaches such as the 1D-CNN, though still useful, offer comparatively lower reliability for this task.

The implications of this work are twofold. Practically, it provides evidence that a carefully pre-processed, moderately sized clinical dataset combined with an appropriately chosen ensemble architecture can support a non-invasive, low-cost, and rapid cardiovascular risk-screening tool suitable for integration into clinical decision-support systems. Theoretically, it reinforces the value of combining complementary learning paradigms—variance reduction (Bagging), bias reduction (Boosting), and model diversity (Stacking)—over relying on any single algorithm. While the near-perfect results warrant cautious interpretation and further external validation, this research lays a solid foundation for future work on explainable, generalizable, and deployable machine learning systems for early cardiovascular disease detection

## REFERENCES

1. C. Mansoor, S. K. Chettri and H. Naleer, “Advancing Heart Disease Prediction through Synergistic Integration of Machine Learning and Deep Learning Techniques,” 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2024, pp. 1 - 6.
2. M. S. Gangadhar, K. V. S. Sai, S. H. S. Kumar, K. A. Kumar, M. Kavitha, and S. Aravinth, “Machine learning and deep learning techniques on accurate risk prediction of coronary heart disease,” in *Proceedings of the 2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, (2023), pp. 227–232.
3. Ouyang, Shuge. “Research of heart disease prediction based on machine learning.” 2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE). IEEE, 2022.
4. A. A. Romalt and R. M. S. Kumar, “Prediction of Cardio Vascular Disease by Deep Learning and Machine Learning-A Combined Data Science Approach,” 2022 International Conference on Computer, Power and Communications (ICCPC), 2022, pp. 83 - 85, doi: 10.1109/ICCPC55978.2022.10072141.

5. C. R. Shaker, A. Sidhartha, A. Praveena, A. Chrsity and B. Bharati, “An Analysis of Heart Disease Prediction using Machine Learning and Deep Learning Techniques,” 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 1484–1491, doi: 10.1109/ICOEI53556.2022.9776745.
6. Kumar S, Thakur B. Heart Disease Prediction Using a Stacked Ensemble Learning Approach. SN Comput Sci. 2025;6:3.
7. Sultan, S.Q.; Javaid, N.; Alrajeh, N.; Aslam, M. Machine Learning-Based Stacking Ensemble Model for Prediction of Heart Disease with Explainable AI and K-Fold Cross-Validation: A Symmetric Approach. Symmetry 2025, 17, 185.
8. Shahid Mohammad Ganie, Pijush Kanti Dutta Pramanik, Majid Bashir Malik3, Anand Nayyar, Kyung Sup Kwak, An Improved Ensemble Learning Approach for Heart Disease Prediction Using Boosting Algorithms, CSSE, Vol.46, No.3, 2023.
9. Abdellatif, A.; Mubarak, H., Abdellatef, H., Kanesan, J., Abdelltif, Y., Chow, C.-O., Chuah, J.H., Gheni, H.M., Kendall, G. Computational detection and interpretation of heart disease based on conditional variational auto-encoder and stacked ensemble-learning framework. Biomed. Signal Process. Control 2024, 88,105644.
10. Ankit Saxena, Renu Sharma, and Vinod Jain, “Advanced Heart Disease Prediction System Using Hybrid Machine Learning Techniques,” 2023 5th International Conference on Computing, Communication and Intelligent Systems (ICCCIS), 2023, pp. 638 - 643.