ISSN: 2093-5374 / E-ISSN: 2093-5382

Exploring Hybrid Optimization Techniques for High-Performance Heart Disease Prediction Models: Challenges and Solutions

B.V. Swathi[1*], and Dr.D.G. Jyothi[2]

[1*]Research Scholar, Bangalore Institute of Technology, Visvesvaraya Technological University, Belagavi, Karnataka, India; Assistant Professor, Department of CSD, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India. bvswathi92@gmail.com, https://orcid.org/0000-0003-4478-8736

[2]Professor, Department of AI & ML, Bangalore Institute of Technology, Belagavi, Karnataka, India. dgjyothi@bit-bangalore.edu.in, https://orcid.org/0000-0002-1859-6733

Received: October 06, 2025; Revised: November 29, 2025; Accepted: January 02, 2026; Published: March 31, 2026

Abstract

Cardiovascular disease (CVD) is currently among the leading causes of morbidity and mortality across the world, which is why there is an excellent demand of reliable, effective, and exact predictive mechanisms. In a typical machine learning/deep learning, the class imbalance, high-dimensional clinical data, noisy or redundant features, and computational inefficiencies are some of the key weaknesses that tend to deteriorate predictive reliability. To address these problems, scientists have resorted more and more to hybrid optimization methods, which combine machine learning models with metaheuristic methods in order to advance the process of feature selection, parameter optimization, and model generalization. In this survey paper, a detailed analysis of the hybrid optimization methods used to predict heart diseases is given, especially the way in which they improve the predictive capabilities. The paper examines popular metaheuristic algorithms such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) and explains their efficiency in the search of the most informative features of complicated medical data. Special focus is put on the classifier of the Random Forest that is optimized with the help of the following metaheuristic strategies in order to enhance precision and decrease overfitting. The survey identifies the practical challenges of hybrid optimization, including the cost of computation, convergence problems, and the heterogeneous healthcare data, as well. These limitations are mitigated by suggesting possible solutions and research directions. Lastly, the paper provides a comparison of the performance of different machine learning algorithms that have been optimized using different metaheuristic methods and it is established that hybrid models will always have a high predictability of cardiovascular disease. The results emphasize the relevance of hybrid optimization as a strong direction in the formation of high-performance and clinically reliable models of heart disease prediction.

Keywords: Hybrid Optimization, Metaheuristic Algorithms, Machine Learning, Feature Selection, Genetic Algorithm, Particle Swarm Optimization.

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 17, number: 1 (March-2026), pp. 411-424. DOI: *10.58346/JOWUA.2026.I1.023*

*Corresponding author: Research Scholar, Bangalore Institute of Technology, Visvesvaraya Technological University, Belagavi, Karnataka, India; Assistant Professor, Department of CSD, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India.

1 Introduction

Recent research highlights the importance of timely and targeted diagnostic systems due to the increasing rates of cardiovascular disease and associated complications. Traditional ML systems designed for cardiovascular disease detection lack interpretability and are subject to overfitting and dataset imbalance. Recently proposed hybrid optimization frameworks designed for the dual purpose of predictive healthcare system improvement have gained attention in the literature (Suvarna & Bharadwaj, 2024). Various hybrid optimization frameworks for feature selection, parameter tuning, and classification accuracy improvement utilize the global exploratory capacity of genetic algorithms (GA) and the local exploitative capacity of particle swarm optimization (PSO) in one of the boosts. Compared to standalone algorithms, the heart disease detection models GA+SVM, ACO+KNN, XGBoost + GA, PSO+ANN, and CNN+WOA are high-performing hybrids (Hammadi et al., 2025; Kadhim et al., 2024).

Connecting the gap between practical ML applications and metaheuristic optimization techniques, this research focuses on how hybrid methods manage computational and clinical difficulties in heart disease prediction.

This survey seeks to study hybrid optimization-based techniques for guessing heart disease methodologically, algorithmically, based on evaluation metrics, and considering possible limitations. The review integrates new works to establish contributions made towards identifying trends, gaps, and opportunities in the development of predictive models for diagnosing conditions in healthcare (Rahi & Kang, 2025). Machine learning further advances automated diagnostics and large-scale pattern recognition, transforming data-driven decision making in medicine (Flores-Fernandez et al., 2024). Given that heart disease constitutes 32% of all deaths worldwide, as reported by the World Health Organization, the need for sophisticated computational techniques to facilitate early detection is essential (Ghosh et al., 2021). As a result of ongoing growth in deep learning, hybrid optimization, and machine learning, several studies have contributed to enhancing the accuracy and reliability of cardiac diagnostic predictions (Doppala et al., 2023; Bashir et al., 2021).

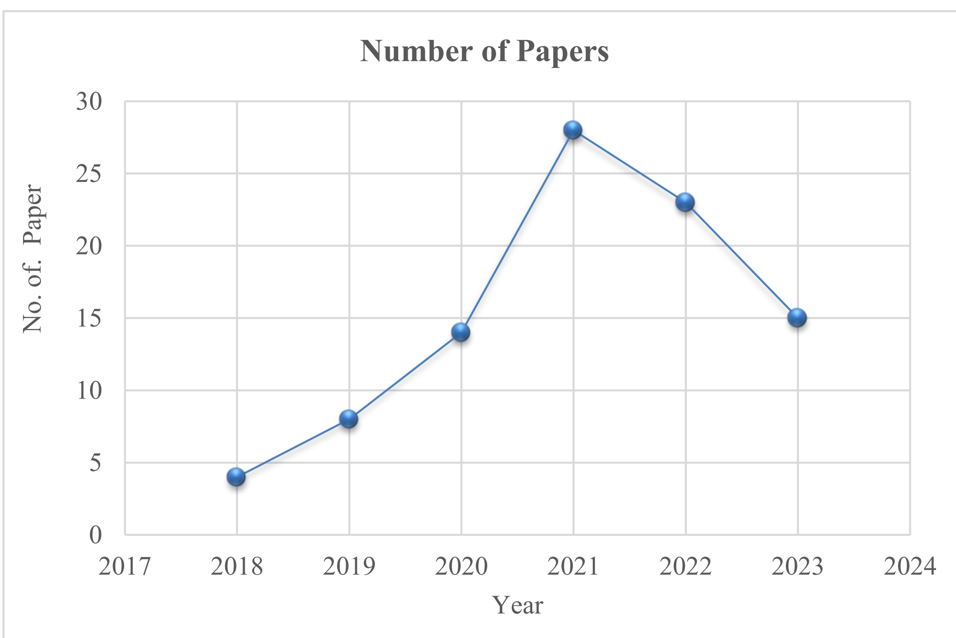

Figure 1: Number of papers distributed on machine learning

Figure 1 shows the number of papers that are connected to the ML methods in heart disease predictions. In 2018, 4 papers were distributed; in 2019, 8 papers; and in 2020, 2021, 2022, and 2023, 14, 18, 24, and 9 papers were distributed (Shinde et al., 2025).

The main contribution of this paper includes an assessment of the current research and the identification of the challenges of hybrid optimization techniques in heart disease prediction. Section 2 will provide the background information, which contains challenges, solutions, and datasets. Section 3 will discuss the hybrid optimization techniques with machine learning and metaheuristic optimization. Sections 4 and 5 examine the research gap and open issues.

2 Background Information

2.1 Challenges and Solutions

Several practical and technical obstacles exist for hybrid optimization methods used to forecast heart diseases. The computational complexity, memory needs, and runtime that result from integrating several optimization and learning algorithms can make them unsuitable for deployment in clinical situations with limited resources. When high-dimensional datasets outnumber sample variety, overfitting becomes even more severe due to feature redundancy and poor selection. Lack of interpretability in hybrid deep-learning frameworks is another issue that prevents their widespread use in clinical settings and undermines confidence in decision-making.

Regarding the identified shortcomings, hybrid metaheuristic frameworks integrate global and local search methods, along with heuristics for feature selection, parameter tuning, and stabilizing the classification model, making them more effective. Gene search, correlation filtering, and mutual-information analysis are a few ways to reduce feature redundancy. Additional benefits of data balancing techniques include the creation of generalized models (SMOTE, normalization, etc.). When compared to adaptive hypercubes, which are diverse and highly spread across clinical data, dynamic group weighting and hyperparameter adjustment provide further steps towards convergence and predictive flexibility. Tools in reasonable AI, such as SHAP and LIME, let practitioners rationalize model decisions and give transparency. In general, the methods used for hybrid optimization in CVD models highlight the significance of models that are durable, efficient in terms of computing, and optimized for interpretability. These models are both more interpretable and less computationally intensive than previous versions.

The implications of these challenges are the incorporation of hybrid metaheuristic frameworks that intelligently integrate surveying and focusing to enhance feature selection and parameter tuning while ensuring predictive stability.

2.2 Dataset and Evaluation Metrics

The Cleveland Heart Disease dataset, which is stored in the UCI Machine Learning Repository, is a popular resource for hybrid optimization frameworks that aim to forecast the prevalence of CVD (Asadi et al., 2021; Jevin et al., 2023). This dataset has 333 patient records and includes the following clinical factors: age, sex, resting electrocardiogram readings, cholesterol, fasting blood sugar, kind of chest pain, resting blood pressure, and fourteen more (Talukder et al., 2025). The target variable is used to classify individuals into five severity categories; it is typically binarized as "no heart disease" (0) or "heart disease present" (1-4).

The Framingham dataset (Ahmed & Husien, 2024; Feifei, 2024), Statlog dataset (Fitriyani et al., 2020), Hungarian dataset (Ghosh et al., 2021; Kanagarathinam et al., 2022), Switzerland dataset (Kanagarathinam et al., 2022), Long Beach VA dataset (Talukder et al., 2025), and Z-Alizadeh Sani dataset (Tyagi et al., 2021) are among many others utilized for cross-domain validation. As an example, the Framingham dataset has 4,240 records that include 16 clinical variables such as smoking status, blood pressure, glucose, cholesterol, and diabetes status. Additionally, there are binary outcomes that indicate the development of cardiovascular disease within ten years: (Ahmed & Husien, 2024; Feifei, 2024) There were 3,596 healthy patients and 644 afflicted cases. Overfitting is more common in smaller datasets, whereas larger datasets offer more consistent prediction performance. This effect is proportional to the size of the dataset.

Typical metrics for evaluating the model's performance are employed. Accuracy is the ratio of the number of instances that are correctly classified (Equation 1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where TP and TN represent false positives and false negatives, respectively, and FP and FN represent false positives and false negatives, respectively. Precision, the fraction of correctly identified positive cases over all predicted positive cases, is calculated as in equation 2,

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall, or sensitivity, indicating the fraction of actual positives correctly detected, is given by equation 3,

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The formulation of the F1 score involves the harmonic mean of both precision and recall. It can be represented as shown in equation 4 below:

$$F1 = 2.\frac{Precision.Recall}{Precision + Recall} \tag{4}$$

The correlation coefficient (MCC) of Matthew is defined in equation 5, and a good classifier of the quality of classification is given, particularly when dealing with unbalanced datasets.

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{5}$$

To assess the hybrid optimization performance in a variety of clinical tasks, a cross-domain assessment on different benchmark datasets is required. This is because of the effect that the dataset selection has on the resiliency of the model.

By combining Random Oversampling with Transfer Learning and Extra Trees Classifier, a hybrid model (RO+TL+ETC) achieves outstanding performance improvements, according to studies conducted on the Framingham and Cleveland datasets (Talukder et al., 2025). The model's performance on the Cleveland dataset was impressive, with the following metrics: F1-score of 98.78%, accuracy of 98.82%, Kappa of 98.47, and MCC of 98.49. The validation results on the Framingham dataset demonstrated strong generalization, with an F1-score of 96.87%, precision of 97.12%, and accuracy of 96.78%. Confirmed by enhanced ROC, MSE, RMSE, and MAE measures, these findings suggest that a combination of data balancing and transfer learning strengthens the model, improves feature representation, and decreases class imbalance.

3 Hybrid Optimization Technique for Heart Disease Prediction Models

3.1 Metaheuristic Optimization Approaches for Heart Disease Prediction based on Random Forest

Machine learning predictive models applied to the study of heart disease benefit from the efficiency gained from the use of genetic algorithms (GAs). This applies primarily to feature selection and neural networks' weight optimization (Tan et al., 2024). GA imitates natural selection; for example, blood pressure, age, and cholesterol and eliminates irrelevant feature candidate solutions in blood pressure, age, and cholesterol feature candidate solutions, and blood pressure, age, and cholesterol feature candidate solutions. This technique increases the complexity of the model as well as the accuracy of the forecast. GA is able to improve the performance of neural networks by further refining the weights between neurons, thus improving classification (Jobson & Subhashini, 2016). The most common metaheuristic used in the feature selection task is Particle Swarm Optimization (PSO), which is an effective algorithm to identify the most useful patient features that also enhance the accuracy of the model. To explain why the Ant Colony Optimization (ACO) could improve the reliability of prediction by cutting redundancy and narrowing down the number of features, it equally enhances the general functionality of the machine learning classifiers in detecting heart diseases.

To improve the performance and accuracy of the classification model, the practical definition of the model and feature subgroups needs to be utilized when using the metaheuristics in the prediction of cardiac conditions (Tomar & Vyas, 2022; Bhattacharya & Deshmukh, 2021). Metaheuristic methods surpass the rest in terms of their extensive hyperparameter searching capabilities and obtaining solutions that are near optimal in under reasonable time frames. As far as intricate clinical datasets are concerned, the scalability of heuristics far surpasses that of traditional optimization techniques.

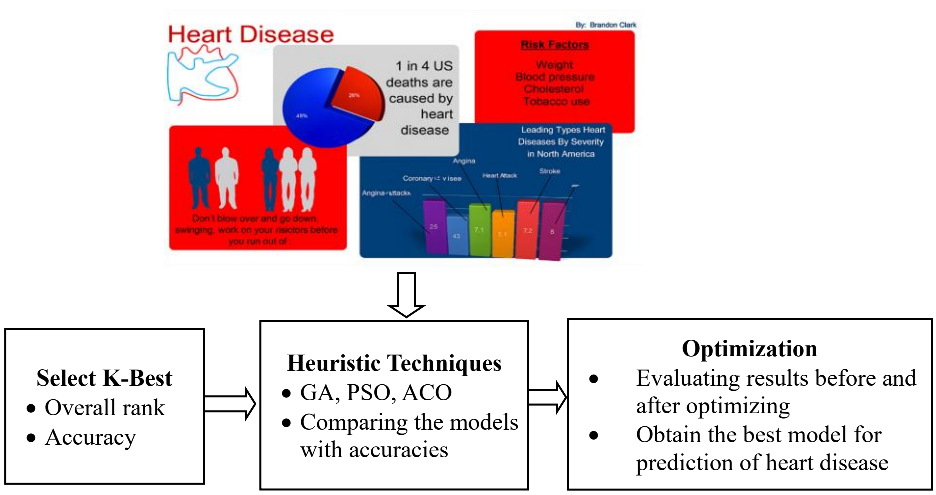

Figure 2: Optimization model for heart disease diagnosis

One binary outcome and thirteen independent factors made up the Cleveland heart disease dataset that was used as a standard in the reviewed study. The system involved preprocessing the dataset, feature selection with the help of metaheuristic techniques, and classification with the help of Random Forest (RF), as illustrated in figure 2. Once we have decided on the most essential features by employing the Select Best method, by combining GA with PSO and ACO, we were able to optimize the compact dataset. In order to increase forecast accuracy, each technique eliminated features that weren't important while keeping the ones that were. The most effective optimization technique was determined by retraining Random Forest with each optimized feature subset and comparing performance metrics. Equation 1 is used to determine the accuracy of the optimization techniques.

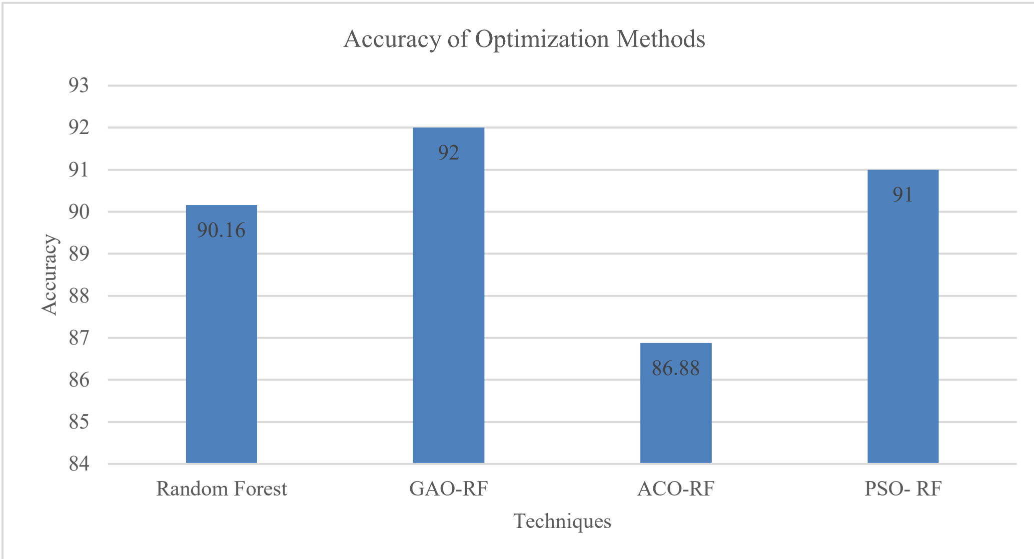

Figure 3: Accuracy of optimization techniques

Figure 3 displays the results of the comparison of the accuracy of alignment of the optimization strategies. In the Cleveland dataset, the Random Forest model optimized by the GA (GAORF) model recorded an impressive 92 percent accuracy, far better than its rivals. With applied adaptive operation, such as mutation, selection, and convergence to select the features effectively, GA produced small and highly selective sets of inputs. PSO and ACO also made better accuracies, but both of them had slightly lower generalization scores by making features more relevant. The combination of hybrid metaheuristics and machine learning, especially the GA–RF models, reveals the prospects of engineering and the capacity to search globally and adjust models on a local scale**.** Genetic algorithms streamline the optimization of neural weights and feature spaces, and random forests classify via ensemble averaging, branch ensemble averaging, and prediction fusion. This partnership enhances precision, decreases dimensionality, and improves interpretability. Hence, for advanced integration automation to become clinically deployable, the GAORF hybrid model for early heart disease detection was focused on. In addressing the primary concern discussed in Section 2, this approach demonstrates that careful feature selection directly improves accuracy and interpretability.

3.2 Hybrid Approach Using GA and PSO for Heart Disease Prediction

A GAPSO-RF system, which combines Genetic Algorithms with Particle Swarm Optimization and Random Forest, was presented in the study (El-Shafiey et al., 2022) for high-performance prediction of cardiac illness. To optimize the feature subsets utilized in Random Forest classification, the approach integrates GA's global search capabilities with PSO's local convergence behavior. The first GA population is formed by identifying the most influential clinical variables using multivariate statistical analysis. Next, a discriminative mutation technique is used to increase genetic diversity. PSO uses velocity and position updates to refine prospective solutions, re-evaluating previously rejected individuals to see if they are more suited. Successful balancing between exploration and exploitation makes this cooperative hybridization better with regard to feature selection and more stable in classification. As figure 4, which is in (El-Shafiey et al., 2022), illustrates, the model was better than the

GA-RF and the PSO-RF models when employed singly in predictive performance. As an expansion of the GA-RF model, hybridization with PSO enhances the merging velocity and the data appropriateness because it improves the global and local search capacities.

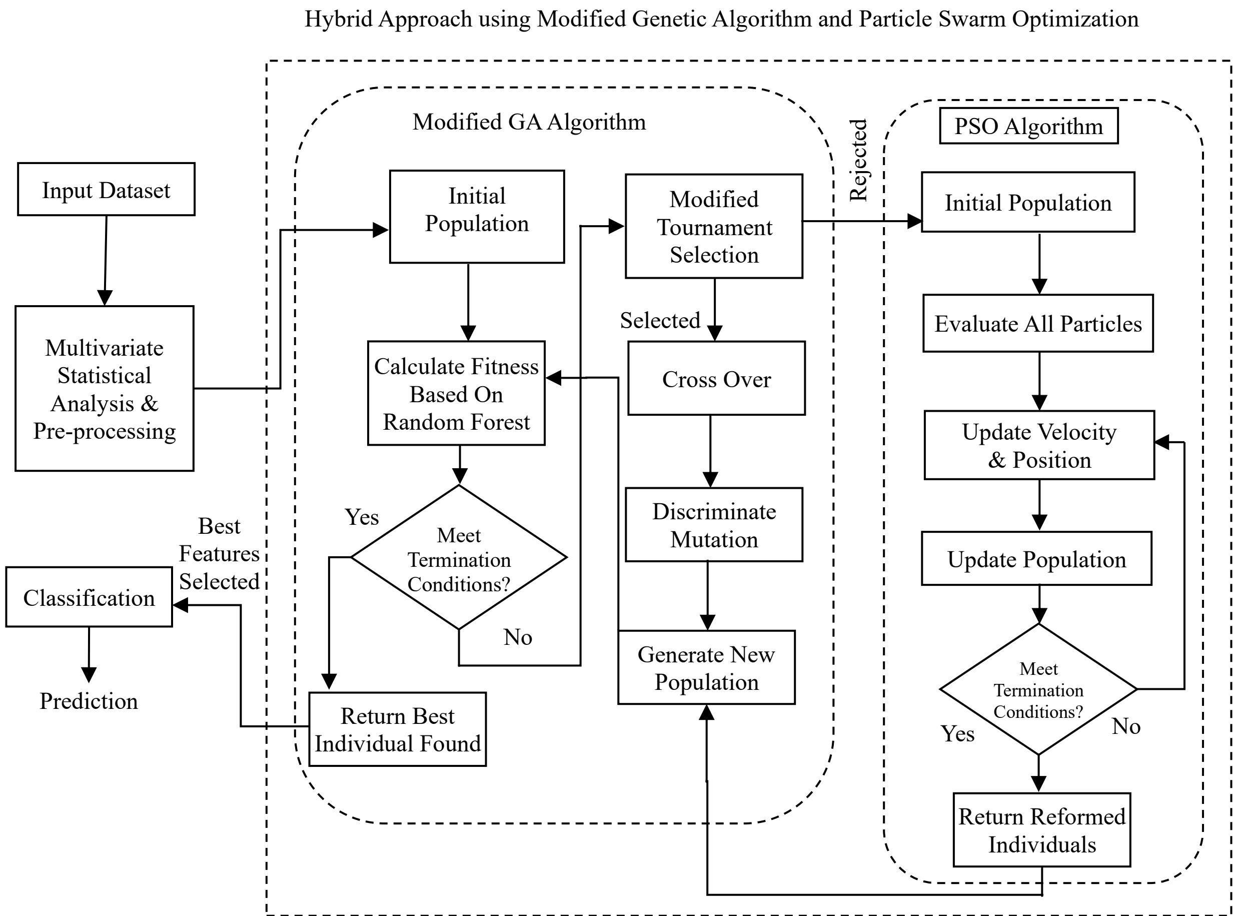

Figure 4: Workflow of GAPSO- RF approach

Figure 4 also depicts the GAPSO-RF hybrid methodology of prediction of cardiac diseases using a Genetic Algorithm to search global features and Particle Swarm Optimization to refine the features before applying Random Forest classification. Such a cooperative hybridization demonstrates that there exists a trade-off between predictive performance and computational complexity in that hybrid methods are capable of improved performance without causing models to become significantly less understandable. Two datasets that are publicly accessible and stored on the UCI Machine Learning Repository were used as controls to measure the effectiveness of the proposed GAPSO-RF approach: the Statorlog Heart Disease dataset and the Cleveland Heart Disease dataset. Some of the customer characteristics that are reported in the records presented in the databases include age, BP, cholesterol, fasting glucose, and exercise angina. Every patient record also contains a binary variable that shows if the patient has a cardiac disease or not. For evaluating the classifiers, the metrics used include accuracy, sensitivity, specificity, and the area under the ROC curve (AUC). Results showed that for the experimental models, GPASO-RF was the most dependable and the most generalizable optimization-based model, achieving an accuracy of 95.6% on the Cleveland dataset and 91.4% on the Statlog dataset. Cross-dataset comparative analysis has established the generalizability of GAPSO-RF, proving its applicability across diverse patient profiles and real-world clinical situations.

3.3 Comparison of Machine Learning Techniques for Predicting Heart Disease

In table 1, the authors are able to provide a thorough representation of the latest papers that use machine learning and hybrid optimization strategies to predict heart diseases. The works reviewed present a variety of algorithms, such as Random Forest, Support Vector Machines, Logistic Regression, Neural Networks, Gradient Boosting, and ensemble techniques that are used on popular datasets, such as Cleveland, Framingham, Hungarian, Statlog, UCI, and ECG specialist datasets. Numerous works combine metaheuristic or feature-selection algorithms, including Genetic Algorithms, Particle Swarm Optimization, Random Search Approach, or hybrid reduction approaches, to improve the performance of the models. Accuracies are reported to be significantly variable, with an average accuracy of between 75 and more than 99 percent, with the best outcomes delivered mainly by hybrid or ensemble systems like RF-based combinations, GA-RBF, CatBoost, and mixed classifier systems. The effects of preprocessing, feature selection, and trait reduction on enhancing the classification results are also mentioned in several works. In general, the table shows how the traditional ML models have evolved to more advanced hybrid and ensemble models with an increasing emphasis on the optimization method to obtain more confident and precise cardiovascular disease prediction.

Table 1: The comparative analysis of the state-of-the-art heart disease prediction models

Reference	Proposed Work	ML Methods' Comparison	Proposed Method	Accuracy	Dataset
(Asadi et al., 2021)	Developed a framework by using a prediction algorithm and feature selection for heart failure disease diagnosis	RF, RSA	RSA + RF	93.34	Cleveland
(Anjaneyulu et al., 2023)	Established a hybrid random model	LR, DT, DL, RF, SVM, NB	Hybrid RF with LM	88.7	Cleveland
(Jevin et al., 2023)	Fast mutual detail feature selection algorithm	LR, ANN, NB, DT, SVM	Feature selection + SVM	92.35	Cleveland
(Fitriyani et al., 2020)	Proposed clinical decision support system	Xgboost, DBScan, Smote NN, KNN	Xgboost	95.90	Statlog and Cleveland
(Katarya & Meena, 2021)	Studied machine learning methods for the CHD prediction	KNN, LR, KNN, NB, ANN, MLP	SVM	92.29	Cleveland
(Rahim et al., 2021)	ML framework for detecting CVD	Ensemble of LR and KNN	Ensemble for LR and KNN	95.50	Framingham
(Swathy & Saruladha, 2022)	Comparison of ML, DM, and DL models for heart disease prediction	ANN, NN	-	-	-
(Amarbayasgalan et al., 2021)	Heart disease prediction with two datasets designed	DNN + FS	DNN	-	Own
(Doppala et al., 2023)	A hybrid approach for heart disease prediction using trait reduction	SVM, RBF, DT, LR	GA-RBF	94.20	Cleveland

(Ghosh et al., 2021)	Established new hybrid classifiers	RF, KNN, Xgboost, DT	RF	99.05	Combination of Hungary, Cleveland
(Bashir et al., 2021)	A four-ensemble related voting scheme is planned	SVM + NB + AUTO NLP	SVM + NB + AUTO NLP	83	Cleveland, Hungary
(Pathan et al., 2022)	Significant risk factor for disease occurrence and disease classification	DT, LR, MLP, ET, RF	SVM	75	CVD, Framingham
(Chang et al., 2022)	An AI model using ML methods for CHD detection	DT, SVM, KNN, RF, LR	RF	83	Cleveland
(Kanagarathinam et al., 2022)	A hybrid dataset for the identification of heart disease	KNN, MLP, SVM, NB, XGBoost	Cat Boost	94.34	Hungarian, Switzerland
(Tyagi et al., 2021)	Feature selection method Through ML methods	RF, SVM, KNN, DT, Ensemble	Ensemble	96.7	Z- Alizadeh Sani
(Aleem et al., 2021)	Feature selection for developing a detection	PSO, NB, SVM, DT, LR, RF, GA	-	-	Hungarian
(Tyagi et al., 2021)	Clustering model for CHD prediction	K-means clustering	K-means clustering	-	ECG signal data
(Lutimath et al., 2021)	Genetic algorithm used for heart disease detection	Xgboost, NN, RF, bagged trees	-	-	UCI
(Gupta et al., 2022)	New popular noting ensemble technique	LR, KNN, SVM, DT, NB, LGBM	Ensemble of NB, SVM, LGBM, Rf	88.33	Cleveland
(Ahmad et al., 2022)	A model that matches the actual world	DT, SVM, KNN, LDA	DT, RF	99.40	Hungary, Cleveland
(Sarah et al., 2022)	Compared different ML techniques and provided a better one	LR, SVM, DT, RF	LR	85.25	Cleveland
(Baghdadi et al., 2023)	DNN using CNN	CNN, BiLSTM	CNN, BiLSTM	94.50	UCI
(Absar et al., 2022)	A system that detects CHD combinations with datasets	DT, RF, ADB, KNN	KNN, RF	93.443	Cleveland
(Rani et al., 2021)	A hybrid system of machine learning methods	NB, LR, RF, SVM, DT	RF	86.60	Cleveland

Figure 5 (Shinde et al., 2025) shows a variety of machine learning approaches, including RF, XgBoost, DNN, KNN, GA-RBF, SVM, and other ensemble methods, that try to predict cardiovascular illness. Deep Neural Networks (DNNs) are the backbone of the most accurate prediction methods, according to figure 5's findings and DNN classification.

Based on the analyses conducted, the hybrid optimization frameworks, specifically GA-RF and GAPSO-RF, can confidently be described as practical, reliable, and clinically viable options for the prediction of heart disease, offering feasibility to scalable and generalizable frameworks for improving preemptive and personalized cardiovascular care. Although the DNN-based models, as compared with the hybrid models, are shown to be reasonably accurate, the hybrid metaheuristic approaches indeed exhibit greater computational efficiency and clinical interpretability due to their simultaneous optimization of feature selection and classification.

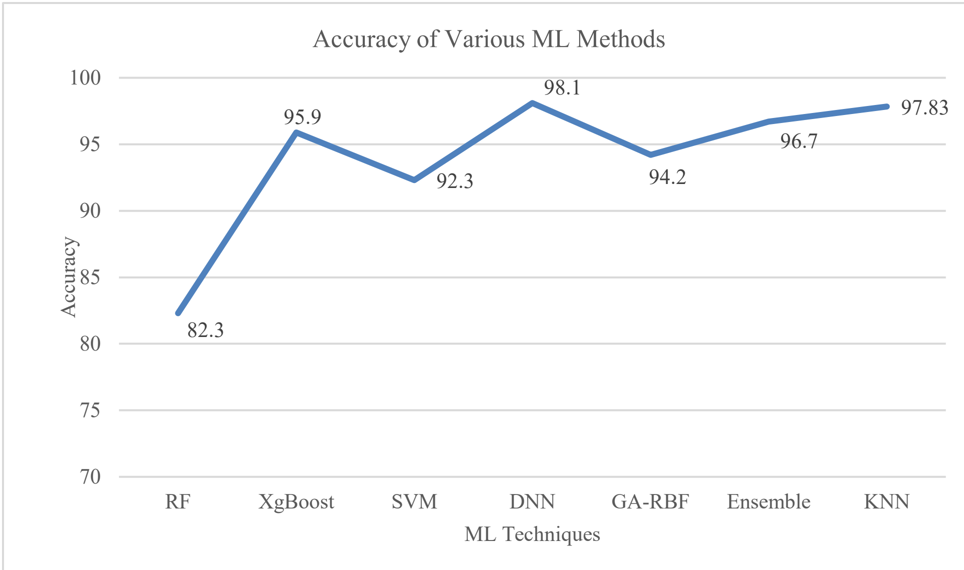

Figure 5: Comparing the accuracy of different machine learning techniques

4 Research Gap

Despite advancements, there are still several methodological and technical gaps in hybrid optimization for CVD prediction. The lack of strategy integration to enhance computing effectiveness and accuracy in predicting resources is experienced because most studies only use one metaheuristic method, such as GA, PSO, or ACO (Castillo & Al-Mansouri, 2025). The trust of doctors and patients is at stake because there has been underfunding in studies that compare the advantages and disadvantages of the complexity, interpretability, and performance of models.

Also, various models have been proven using fixed or limited datasets and, thus, their extensiveness and extrapolation to massive, dynamic electronic health records (EHRs) are constrained. Adaptive resampling and dynamic weighting are yet to be embraced in most models, and this further complicates the deal with class imbalances, noisy data, and missing values. The absence of benchmark standards complicates reproducibility and comparability across studies (Mehra & Patel, 2024). The lack of patient-level personalization, remarkably adaptive, robust, and cross-clinically pertinent hybrid optimization frameworks, underscores the need for enhanced attention to the genetic, lifestyle, and comorbidity diversity within patient populations.

5 Open Issues

Several issues prevent hybrid optimization models from being practically helpful in predicting cardiac disease, even though they have made great strides in this area.

Models are hard to generalize because of problems with imbalance in datasets, heterogeneity, and uneven annotation of regions. Another impediment to cross-validation, benchmarking, and repeatability is the absence of large, labelled, and privacy-compliant clinical datasets. The majority of the frameworks are fixed and are not able to adapt to the dynamic patient information, further slowing down the process. Additionally, there is no automatic hyperparameter adjustment. Interpretability continues to pose a formidable challenge, given the insufficient validation on actual clinical populations and the erosion of practitioner confidence from hybrid "black-box" systems. Sophisticated metaheuristic systems tend to consume excessive memory and processing capacity, which severely limits their applicability in resource-constrained healthcare environments. The solution to such matters will require the creation of new algorithms that would be flexible and applicable in hospitals in a wide variety of scenarios, as well as be practical and easy to digest in terms of their complexity.

6 Conclusion

Based on the evidence of the study, hybrid optimization frameworks that use GA, PSO, and Random Forest classifiers are theoretically powerful and clinically effective for predicting cardiac events. Hybrid optimization systems, where metaheuristic algorithms are used to create forecasting of heart diseases with the help of ML classifiers such as the Random Forest (RF), outperformed more traditional forecasting methods. The frameworks that are employed include GA, PSO, and the Ant Colony Optimization (ACO). GA-RF had a 92 percent success rate in the selection of discriminative features on the Cleveland dataset. The GAPSO-RF hybrid method combines the global searching capabilities of genetic algorithms (GA) with the local refinement abilities of Particle Swarm Optimization (PSO) to optimally feature subsets, achieving 95.6% accuracy on the Cleveland dataset and 91.4% on the Statlog dataset. Composite algorithms are lauded for demonstrating reliable predictive performance and strong generalization across various benchmark datasets, as measured by standardized evaluation metrics. These frameworks address the reduction of feature redundancy and high dimensionality and imbalanced datasets while improving the stability of classification and increasing clinical utility. Scalability, difficulty of interpretability, and adaptation to changing clinical data remain of concern. To establish superior predictive performance, generalization and interpretability, hybrid models such as GA-RF and GAPSO-RF systematically deal with inequality of datasets, feature selection and parameter optimization. These models provide a guideline to the further studies and practice. Integrating energy-efficient hybrid architectures with an emphasis on automated hyperparameter optimization, explainable AI, and privacy-preserving approaches would lead to therapeutically more adoptable, generalizable, and personalized predictive cardiovascular care, which is the goal of ongoing research.

References

[1] Absar, N., Das, E. K., Shoma, S. N., Khandaker, M. U., Miraz, M. H., Faruque, M. R. I., ... & Pathan, R. K. (2022, June). The efficacy of machine-learning-supported smart system for heart disease prediction. In *Healthcare* (Vol. 10, No. 6, p. 1137). MDPI. https://doi.org/10.3390/healthcare10061137

[2] Ahmad, G. N., Fatima, H., Ullah, S., & Saidi, A. S. (2022). Efficient medical diagnosis of human heart diseases using machine learning techniques with and without Grid Search CV. *IEEE Access*, *10*, 80151-80173.

[3] Ahmed, M., & Husien, I. (2024). Heart disease prediction using hybrid machine learning: A brief review. *Journal of Robotics and Control (JRC), 5*(3), 884-892.

[4] Aleem, A., Prateek, G., & Kumar, N. (2021, December). Improving heart disease prediction using feature selection through genetic algorithm. In *International Conference on Advanced Network Technologies and Intelligent Computing* (pp. 765-776). Cham: Springer International Publishing.

[5] Amarbayasgalan, T., Pham, V. H., Theera-Umpon, N., Piao, Y., & Ryu, K. H. (2021). An efficient prediction method for coronary heart disease risk based on two deep neural networks trained on well-ordered training datasets. *IEEE Access*, *9*, 135210-135223.

[6] Anjaneyulu, M., Degala, D. P., Devika, P., & Hema, V. (2023). Effective heart disease prediction using hybrid machine learning techniques. *Advancements in Aeromechanical Materials for Manufacturing: ICAAMM-2021*, *2492*(1), 030070. https://doi.org/10.1063/5.0114370

[7] Asadi, S., Roshan, S., & Kattan, M. W. (2021). Random forest swarm optimization-based for heart diseases diagnosis. *Journal of biomedical informatics*, *115*, 103690. https://doi.org/10.1016/j.jbi.2021.103690

[8] Baghdadi, N. A., Farghaly Abdelaliem, S. M., Malki, A., Gad, I., Ewis, A., & Atlam, E. (2023). Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. *Journal of Big Data*, *10*(1), 144. https://doi.org/10.1186/s40537-023-00817-1

[9] Bashir, S., Almazroi, A. A., Ashfaq, S., Almazroi, A. A., & Khan, F. H. (2021). A knowledge-based clinical decision support system utilizing an intelligent ensemble voting scheme for improved cardiovascular disease prediction. *IEEE Access*, *9*, 130805-130822.

[10] Bhattacharya, M., & Deshmukh, A. (2021). A Smart Grid-oriented Energy Optimization System Using Renewable Energy Sources and Machine Learning. *International Academic Journal of Science and Engineering, 8*(2), 26–30.

[11] Castillo, M. F., & Al-Mansouri, A. (2025). Big Data Integration with Machine Learning Towards Public Health Records and Precision Medicine. *Global Journal of Medical Terminology Research and Informatics, 3*(1), 22-29.

[12] Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics*, *2*, 100016. https://doi.org/10.1016/j.health.2022.100016

[13] Doppala, B. P., Bhattacharyya, D., Chakkravarthy, M., & Kim, T. H. (2023). A hybrid machine learning approach to identify coronary diseases using feature selection mechanism on heart disease dataset. *Distributed and Parallel Databases*, *41*(1), 1-20.

[14] El-Shafiey, M. G., Hagag, A., El-Dahshan, E. S. A., & Ismail, M. A. (2022). A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. *Multimedia Tools and Applications, 81*(13), 18155-18179.

[15] Feifei, W. (2024). Optimization Algorithm of Public Service Facilities Layout in Earthquake-stricken Areas based on the SA Algorithm. *Archives for Technical Sciences, 2*(31), 70–85. https://doi.org/10.70102/afts.2024.1631.070

[16] Fitriyani, N. L., Syafrudin, M., Alfian, G., & Rhee, J. (2020). HDPM: an effective heart disease prediction model for a clinical decision support system. *IEEE Access*, *8*, 133034-133050.

[17] Flores-Fernandez, G. A., Jimenez-Carrion, M., Gutierrez, F., & Sanchez-Ancajima, R. A. (2024). Genetic Algorithm and LSTM Artificial Neural Network for Investment Portfolio Optimization. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15*(2), 27-46. https://doi.org/10.58346/JOWUA.2024.I2.003

[18] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. J. M., Ignatious, E., ... & De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*, *9*, 19304-19326.

[19] Gupta, P., Mala, S., Shankar, A., & Asirvadam, V. S. (2022). Heart disease detection scheme using a new ensemble classifier. In *Advances in Data and Information Sciences: Proceedings of ICDIS 2021* (pp. 99-110). Singapore: Springer Singapore.

[20] Hammadi, J. F., Latif, A. B. A., & Cob, Z. B. C. (2025). Artificial intelligence approaches for cardiovascular disease prediction: a systematic review. *Indonesian Journal of Electrical Engineering and Computer Science*, *38*(2), 1208-1218. https://doi.org/10.11591/ijeecs.v38.i2.pp1208-1218

[21] Jevin, M. J., Jayant, H., Sanjay, R., Hemasai, V., & Venkatasrinivas, P. V. (2023). Heart disease identification method using machine learning classification in e-healthcare. *Heart Disease*, *10*(3).

[22] Jobson, M. C, & Subhashini, R. (2016). To Assess the Social Maturity Level among Congenital Heart Defect Children. *International Academic Journal of Social Sciences, 3*(2), 1–6.

[23] Kadhim, A. A., Mohammed, S. J., & Al-Gayem, Q. (2024). DVB-T2 Energy and Spectral Efficiency Trade-off Optimization based on Genetic Algorithm. *Journal of Internet Services and Information Security, 14*(3), 213-225. https://doi.org/10.58346/JISIS.2024.I3.012

[24] Kanagarathinam, K., Sankaran, D., & Manikandan, R. (2022). Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset. *Data & Knowledge Engineering*, *140*, 102042. https://doi.org/10.1016/j.datak.2022.102042

[25] Katarya, R., & Meena, S. K. (2021). Machine learning techniques for heart disease prediction: a comparative study and analysis. *Health and Technology*, *11*(1), 87-97.

[26] Lutimath, N. M., Ramachandra, H. V., Raghav, S., & Sharma, N. (2021, September). Prediction of heart disease using genetic algorithm. In *Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021* (pp. 49-58). Singapore: Springer Singapore.

[27] Mehra, P., & Patel, K. (2024). A Metrics Driven Approach to Brand Management and Brand Health Check. In *Brand Management Metrics* (pp. 31-47). Periodic Series in Multidisciplinary Studies.

[28] Pathan, M. S., Nag, A., Pathan, M. M., & Dev, S. (2022). Analyzing the impact of feature selection on the accuracy of heart disease prediction. *Healthcare Analytics*, *2*, 100060. https://doi.org/10.1016/j.health.2022.100060

[29] Rahi, P., & Kang, S. S. (2025). A novel paradigm in cardiovascular disease risk prediction through hybrid machine learning. *Системная инженерия и информационные технологии*, *7*(3 (22)), 48-65.

[30] Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M. A., & Muzaffar, A. W. (2021). An integrated machine learning framework for effective prediction of cardiovascular diseases. *IEEE access*, *9*, 106575-106588.

[31] Rani, P., Kumar, R., Ahmed, N. M. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments*, *7*(3), 263-275.

[32] Sarah, S., Gourisaria, M. K., Khare, S., & Das, H. (2022). Heart disease prediction using core machine learning techniques—a comparative study. In *Advances in Data and Information Sciences: Proceedings of ICDIS 2021* (pp. 247-260). Singapore: Springer Singapore.

[33] Shinde, P., Sanghavi, M., & Tran, T. A. (2025). A survey on machine learning techniques for heart disease prediction. *SN Computer Science*, *6*(4), 334. https://doi.org/10.1007/s42979-025-03860-2

[34] Suvarna, N. A., & Bharadwaj, D. (2024). Optimization of System Performance through Ant Colony Optimization: A Novel Task Scheduling and Information Management Strategy for Time-Critical Applications. *Indian Journal of Information Sources and Services, 14*(2), 167–177. https://doi.org/10.51983/ijiss-2024.14.2.24

[35] Swathy, M., & Saruladha, K. (2022). A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques. *ICT express*, *8*(1), 109-116.

[36] Talukder, M. A., Talaat, A. S., & Kazi, M. (2025). Hxai-ml: a hybrid explainable artificial intelligence based machine learning model for cardiovascular heart disease detection. *Results in Engineering*, *25*, 104370. https://doi.org/10.1016/j.rineng.2025.104370

[37] Tan, W., Sarmiento, J., & Rosales, C. A. (2024). Exploring the Performance Impact of Neural Network Optimization on Energy Analysis of Biosensor. *Natural and Engineering Sciences, 9*(2), 164-183. https://doi.org/10.28978/nesciences.1569280

[38] Tomar, A., & Vyas, N. (2022). Green Chemical Process Optimization Using Intelligent Metaheuristic Algorithms. *International Academic Journal of Innovative Research, 9*(3), 1–6. https://doi.org/10.71086/IAJIR/V9I3/IAJIR0918

[39] Tyagi, A., Singh, V. P., & Gore, M. M. (2021, December). Improved detection of coronary artery disease using dt-rfe based feature selection and ensemble learning. In *International conference on advanced network technologies and intelligent computing* (pp. 425-440). Cham: Springer International Publishing.

Authors Biography

B.V. Swathi is currently pursuing her Ph.D. in Computer Science and Engineering at Bangalore Institute of Technology under Visvesvaraya Technological University and working as an Assistant Professor in Dept of Computer Science and Design at Dayananda Sagar College of Engineering. Her research focuses on Machine Learning, Deep Learning, Deep Belief Networks, and Hybrid Optimization algorithms. Her academic interests also include Machine Learning, Data Mining & Data Warehousing, Database Management and systems, and Natural Language Processing.

Dr.D.G. Jyothi is a Professor and Dean R&D in the Department of Artificial Intelligence and Machine Learning at Bangalore Institute of Technology with 28 years of Teaching and Research Experience. Her major areas of research interest are Computer Networks, Artificial Intelligence and Machine Learning, IoT, and Blockchain.