MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 2340-2345

Data-driven cardiovascular disease prediction using feature selection and evolutionary SVM with GA optimization

Vibhav Krashan Chaurasiya[1], Dr. Sakshi Rai[2]
[1]PhD Scholar, Computer Science Department,
LNCT University, Bhopal, India
[2]Associate Professor, Computer Science Department,
LNCT University, Bhopal, India

Article Info

Article history:
Received May 27, 2026
Revised May 29, 2026
Accepted May 31, 2026

Keywords:
Support Vector Machine (SVM)
Genetic Algorithm (GA)
Cardiovascular Disease Prediction
ANOVA Feature Selection Hyperparameter Optimization

Corresponding Author:
Vibhav Krashan Chaurasiya
PhD Scholar,
Computer Science Department,
LNCT University, Bhopal, India
Email: joyvib@gmail.com

ABSTRACT

Cardiovascular diseases $CVDs$ remain one of the leading causes of mortality worldwide, creating a need for accurate and early prediction systems. This study presents a hybrid machine learning framework that combines Support Vector Machine SVM classification with Genetic Algorithm GA-based hyperparameter optimization for cardiovascular disease prediction. The proposed model applies preprocessing, normalization, and ANOVA-based feature selection to improve classification efficiency and reduce redundant features. The optimized GA-tuned SVM model was evaluated using a publicly available cardiovascular disease dataset and compared with baseline classifiers including Random Forest, K-Nearest Neighbors, XGBoost, and conventional SVM. Experimental results demonstrated that the proposed approach achieved superior predictive performance with an accuracy of 91.11% and ROC-AUC of 93.20%. The findings indicate that the proposed hybrid framework can support intelligent clinical decision-making and early cardiovascular disease detection in healthcare applications.

1. INTRODUCTION

One of the leading causes of death globally is thought to be cardiovascular diseases (CVDs) [1]. The World Health Organization (WHO) estimates that cardiovascular illnesses cause 17.95 million deaths annually, or over 32% of all fatalities worldwide [2]. The rising number of deaths from cardiovascular diseases underscores the need for effective prevention strategies and early diagnostic systems, thus making cardiovascular diseases a major health concern globally [7,9]. Cardiovascular illnesses have multiple origins, most of which are associated with different risk factors. Numerous modifiable risk factors, such as dietary practices, inactivity, smoking, and excessive alcohol use, have a significant impact on cardiovascular illnesses [4,8]. Cardiovascular disorders are also influenced by non-modifiable risk variables such as age, gender, and genetics [3]. To identify people who are at a higher risk of cardiovascular diseases, sophisticated analytical tools that analyse multiple variables to find hidden patterns in medical data are necessary due to the complex relationship between the various risk factors contributing to cardiovascular diseases [10]. The main motivation behind the proposed research work is the identification of the early symptoms of cardiovascular diseases. Although the conventional approach is reliable in this regard, it is not always sufficient in cases where the symptoms of the diseases are not clearly visible [21-24]. This has created a significant interest in the development of computational approaches that could assist conventional approaches with the aid of data-driven insights regarding the health of the patient [11].

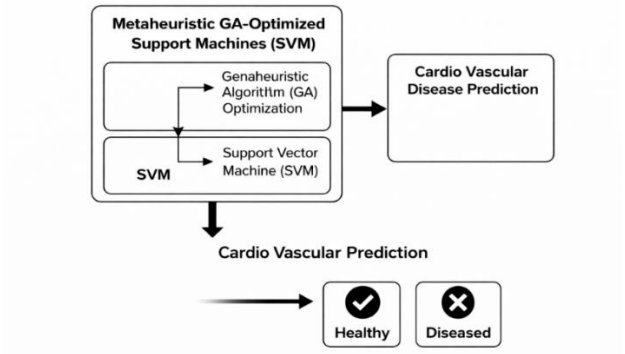

Figure 1. Architecture of improved GA-tuned SVM model [23]

1.1 Machine Learning Limitation in CVD Detection Approach: Medical clinical data has been subjected to a variety of machine learning techniques, such as ensemble methods, Random Forest, multiclass Support Vector Machine, and Neural Networks [11-16]. Fixed loss functions and assessment criteria are typically the foundation of traditional machine learning techniques. Regrettably, assessment criteria are frequently insufficient to address learning issues related to class imbalance [15]. Due to its excellent performance in addressing binary class learning difficulties, Support Vector Machine has garnered a lot of attention in machine learning among the methods created for handling class imbalance learning problems [17-20]. The selection of suitable values for the Support Vector Machine's parameters, such as regularisation parameters, kernel types, and gamma values, typically determines the machine's performance [3,4]. Inappropriate parameter selection frequently leads to either overfitting or underfitting. Grid search or manual trial-and-error methods are typically used to choose the parameters. The manual trial-and-error method typically takes a long time and does not provide the best possible outcome.

1.2 Contribution of Proposed Research : The main contribution this research will make is that the hybrid method is proposed as a successful combination of Support Vector Machine and Genetic Algorithms methods. Genetic Algorithm optimizes the hyperparameters of SVM model which enhances the accuracy, reliability and generalizability of the results of the classification. Since the cardiovascular disease prediction issue demands a high level of accuracy and reliability, suggested SVM-GA model is rather relevant. The hybrid model that has been proposed can be associated with the principles of precision medicine that focuses on making medical interventions specific to the individual characteristics of a patient. The proposed SVM-GA model would help medical professionals to exercise informed decision-making that may improve the quality of healthcare services and reduce their prices. The proposed hybrid model of SVM and GA is an innovation in the sphere of computational diagnostics since it covers the weakness of the current methods of diagnosis and is highly efficient in early diagnosis of cardiovascular disorders.

2. LITERATURE REVIEW

The issue of cardiovascular disease prediction with the help of advanced computational algorithms has been discussed by a great number of research studies in the last few years. Fast, precise, and efficient diagnostic technologies are required because cardiovascular disease remains a significant issue in the world population. This section of the study addresses the current research in the field of machine learning, optimization, and hybrid methodologies that are used in prediction of cardiovascular disease.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 2340-2345

2.1 AI Approaches for CVD Prediction: Medical diagnosis has successfully employed a number of machine learning classifiers, such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor's (KNN) and deep learning algorithms, due to their ability to scan large datasets and extract underlying patterns [20]. Recent research claims that machine learning types of classifiers may enhance early detection of heart diseases. It demonstrated superior results in their systematic examination of machine learning as well as deep learning algorithms in predicting heart disease as compared to old fashioned methods. Nonetheless, the issues of deploying machine learning classifiers to real-time systems and the interpretability of machine learning classifiers were not discussed in the study in detail [1]. The effectiveness of individual and hybrid machine learning classifiers in the detection of cardiovascular diseases in the initial stages of their development was validated. Nevertheless, the challenges of optimization of machine learning classifier and imbalance in class in machine learning classifier were not addressed in the research [2].

2.2 Hyperparameter Tuning in Classification Models : Hyperparameter tuning is also a very important part of the implementation of machine learning model performance. Despite being used, the classic methods of hyperparameter optimisation such as grid search and manual tuning are impractical to optimize models, particularly in high-dimensional space. Although optimization methods such as random search and Bayesian optimization can be used to optimize models, they are limited in their ability to perform a global search. In a recent study, an ensemble-based model to operate an image classification based on the Genetic Algorithm is introduced. Their proposed EGACNN model has a 99.91% accuracy and CNN-spiking neural network model had an accuracy of 99.68%. These results made it clear that hyperparameter optimization techniques can boost the performance of image categorization accuracy [3]. To optimize a Random Forests model with a FOX algorithm and predict cardiovascular disease in another research. In their proposed model, they used the UCI cardiovascular disease dataset with 13 variables and 303 samples. One of the preparation tasks included them using SMOTE to balance the data, treating missing values, and normalising the data. Their findings showed that they were more reliable than the conventional methods and that the precision was 97.83, recall 97.88, and the F1-score was 97.89 [4].

2.3 Genetic Algorithm in Machine Learning Optimization : In this respect, we are motivated by natural selection in the population-based metaheuristic optimization strategy called Genetic Algorithm (GA). It is used to determine solutions of optimization problems with a fitness function. Genetic algorithms are a widely used approach to optimization and have been successfully used by various scholars in improving the performance of machine learning models. The suggested a learning-assisted evolutionary algorithm that improves the performance of small-scale multi-objective optimization tasks. They demonstrated the effectiveness of their strategy in the context of other optimization strategies that are currently being used in their study [5]. The study demonstrated the effectiveness of their approach to enhance the efficiency of optimization of different optimization problems [6]. As an optimization method, a genetic algorithm can be useful in boosting the accuracy of a machine learning model. However, the issues of reliability in health information are beyond the reach of the different methods currently in existence of optimization. To make the models of predicting cardiovascular diseases more accurate, a better hybrid approach to machine learning is needed, combining feature selection in optimizations.

3. MATERIALS AND METHODS

This study presents a proposed hybrid machine learning model, which is founded on predicting the cardiovascular disease (CVD). The method is a combination of Support Vector Machine (SVM) and Genetic technique (GA) optimisation, which makes it classified with accuracy and possesses better predictive robustness. The proposed framework also avoids the weaknesses that are inherent in the traditional machine learning algorithms, which automatically identify the optimal hyperparameters through global search methods. In general, the strategy includes mathematical modelling, kernel-based classification and evolutionary optimization algorithms, which are intended to provide a good analysis of medical data. The given proposal applies to the publicly available UCI Heart Disease data containing 303 cases of patient experiences and several clinical attributes that can be used to predict cardiovascular disease. The summary of data attributes in the present study is presented in Table 1 which shows the sample size, features and classes distribution.

Table 1. Dataset description

Parameter	Description
Dataset Name	UCI Heart Disease Dataset
Source	UCI Machine Learning Repository (24)
Total Samples	303
Total Features	14
Selected Features	Top 10 (ANOVA-based)
Target Classes	Disease / No Disease
Data Type	Clinical patient records

3.1 Support Vector Machine Framework: Support Vector Machine (SVM) is a supervised machine learning technique that can handle high-dimensional data and has a high degree of generalisation, making it widely useful to classification. Building an ideal hyperplane that maximises the separation margin between different classes is the primary goal of SVM. The training dataset will be displayed as follows the labels have been applied:

$$D = \{(x_i, y_i)\}_{i=1}^{N} \quad (1)$$

where: $x_i \in R^d$ represents the feature vector of the i^{th}patient, $y_i \in \{-1, +1\}$ is the label of corresponding class
The classification function of SVM is defined as:

$$f(x) = \omega^T \phi(x) + b \quad (2)$$

where: ω denotes the weight vector, $\phi(x)$represents nonlinear feature transformation, b represents bias
SVM is aimed at maximizing the distance between two classes and reducing errors in classification. This is optimized by the use of the following optimization:

$$\min_{\omega,b,\varepsilon} \frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{N} \varepsilon_i \quad (3)$$

subject to:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (4)$$

where: ε_i are the slack variables, C refers to the regularization parameter of the classification tolerance. This is an expression that allows SVM to address both the linearly separable and non-separable data sets.

3.2 Kernel Function for Nonlinear Classification: When it comes to variables, medical data is typically nonlinear. Kernel functions are used to convert input data in a higher-dimensional feature space to solve the nonlinear separability. This study adopts the Radial Basis Function (RBF) kernel due of its versatility and ability to handle nonlinear classification problems. The definition of the RBF kernel is:

$$K(x_i, x_j) = \exp\left(-\gamma \| x_i - x_j \|^2\right) \quad (5)$$

where: γ is the kernel width parameter, A higher value of γproduces narrower decision boundaries. The RBF kernel is the most common kernel used in medical predictions because it can represent the complicated decision surfaces.

3.3 Genetic Algorithm-Based Hyperparameter Optimization: Genetic Algorithm (GA) is a metaheuristic optimization algorithm which is a population-based method based on biological evolution. It works by enhancing candidate solutions with genetic operation like selection, crossover and mutation. In the suggested framework, GA is used to optimize such variables of critical SVM hyperparameters as regularization parameter, kernel coefficient, and kernel type. Candidate's solutions are indicated by the use of chromosomes. The structure of the chromosomes is determined as:

$$\theta = (C, \gamma, k) \quad (6)$$

The initial population is generated as:

$$P^{(0)} = \{\theta_1, \theta_2, \ldots, \theta_M\} \quad (7)$$

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 2340-2345

where: θ_j represents the hyperparameter vector, M denotes population size. The population is changing over many generations to enhance performance of models.

3.4 Fitness Evaluation Strategy: Each chromosome's performance within the population is evaluated using the fitness function. In the present study, the Receiver Operating Characteristic-Area Under Curve (ROC-AUC) and classification metrics are combined to quantify the prediction capacities. The following function:

$$Fitness(\theta_j) = \alpha \cdot Accuracy + \beta \cdot ROC\text{-}AUC \quad (8)$$

where: $\alpha + \beta = 1, \alpha, \beta \geq 0$. The weighting parameters α and β are selected to balance classification accuracy and discrimination capability. This combined evaluation ensures reliable model selection and prevents bias toward a single performance metric.

3.5 Genetic Operations: Genetic operations are taken in a repetitive manner to transform the population of candidate solutions to optimal hyperparameter values. These operations are the simulation of biological evolution mechanisms and enhance the performance of models over generations. Selection, crossover, and mutation are the primary genetic processes used in this work.

3.5.1 The Process of Selection : Based on their fitness value, the selection procedure will identify the most suitable parent chromosomes. The likelihood that a chromosome will be selected for use in reproduction increases with its fitness value. The selection of parents is described as:

$$\theta_{parent} = \arg\max_{\theta \in P} Fitness(\theta) \quad (9)$$

This operation helps to make sure that the best solutions are taken for the next generation to enhance the efficiency of convergence.

3.5.2 Crossover Operation: New offspring chromosomes are produced using crossovers so that it mixes features of two parent chromosomes. The operation increases diversity of the population and increases the likelihood of finding the best solutions. The crossover operation will be given as:

$$\theta_{Child} = \lambda\theta_{p1} + (1-\lambda)\theta_{p2} \quad (10)$$

where: θ_{p1}, θ_{p2} represent parent chromosomes, λ denotes crossover coefficient. The step facilitates exchange of genetic information among parent solutions.

3.5.3 Mutation Operation: Mutation leads to the introduction of random changes in the chromosomes to ensure there is a variation in the chromosomes to avoid converging early. It is defined that the mutation will be performed as follows:

$$\theta^* = \theta + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2) \quad (11)$$

where: ε represents a random mutation variable, σ^2represents mutation variance

Mutation guarantees the search of new solution spaces and enhances the ability to search globally.

3.6 Optimized Hyperparameter Selection: The most successful chromosome is chosen after several generations of genetic development through the highest fitness value. The optimal set of hyperparameters for training the final model is this chromosome. The chosen optimised parameters are expressed as follows:

$$\theta^* = \arg\max_{\theta \in P} Fitness(\theta) \quad (12)$$

This step determines the hyperparameters that are optimized globally to give the classification performance maximum.

3.7 Final Optimized SVM Classification Model: The final Support Vector Machine model is trained to do classification using the optimised hyperparameters found by the Genetic Algorithm. The definition of an optimised decision function is:

$$f^*(x) = \sum_{i=1}^{N} \alpha_i\, y_i K(x_i, x) + b \quad (13)$$

where: $\alpha_i$represents support vector coefficients, y_i represents class labels, $K(x_i, x)$is the kernel function, b *is the* bias. Combination of Genetic Algorithm and Support Vector machine will make it possible to optimize and globalize the hyperparameters as the classification accuracy and predictive reliability is enhanced. This hybrid structure is especially applicant when dealing with nonlinear and imbalanced cardiovascular data, which enhances the ability of the early detection in the clinical decision-making system.

3.8 Summary of Proposed Hybrid Framework : The methodology that would be recommended in the study would involve the use of the efficiency of the Support Vector Machine classification and the Genetic Algorithm optimisation to develop an effective prediction model. The SVM component gives a high performance of learning correct classification according to the margin and the GA component will optimize the performance automatically by automatically optimizing the hyperparameters. These combination techniques result in stronger generalization, reduction of classification errors and prediction of cardiovascular diseases.

3.9 Data Processing and Experimental Setup: The proposed hybrid GA-tuned SVM model would be implemented on Jupyter Lab on Python language. The implementation process comprised the data set acquisition, its preprocessing, feature selection, model training and performance assessment. The objective of the workflow was to ensure reproducibility and reliable execution of the workflow on the task of cardiovascular disease prediction.

3.9.1 Dataset Description: The proposed system will be based on a publicly available cardiovascular dataset that was acquired through the UCI Machine Learning Repository. The data set is a combination of clinical variables, which are associated with the diagnosis of cardiovascular disease, and demographic and physiological variables. It was trained and tested on 303 records of patients who had numerous clinical features (24).

3.9.2 Data Preprocessing: Data preprocessing is required to enhance model performance and come up with a sound prediction output. First, the encoding of the category variables sex, type of chest pain (cp), thal, and slope into numerical numbers were performed through the use of label encoding methods. The change allowed even categorical data to be processed by machine learning algorithms. Two methods of statistics that were used to determine the outlier include Z-score analysis and box plot visualisation. There were numerical variables that were scrutinized under scrutiny, which are cholesterol level (chol), the maximal heart rate (thalach), and the ST depression (oldpeak). Extreme values that fell out of the three standard deviations were verified and further processing was done to keep the extreme values that were clinically important. This was done through feature scaling in which numerical attributes were standardized with the use of normalization methods. Standardization also ensured that the contribution of all the features in the classification process was equal as well as it increased convergence during model training.

3.9.3 Feature Selection Using ANOVA: The Analysis of Variance (ANOVA) F-test was employed in determining the most relevant variables to make a prediction of cardiovascular illness. The SelectKBest method was used to identify the ten most significant features in relation to the target variable in terms of their statistical value. The characteristics that were selected to be used in the proposed model are indicated in Table 2. The characteristics selected made the classification more effective and reduced the dimensions and eliminated the redundant variables. The computation performance and the model complexity were also enhanced by the process.

Table 2. Features of Data Set

age	sex	smoke	years	ldl	chp	height	weight	fh	active	ihd	hr	dm	bpsys	bpdias	htn	ivsd
65	0	0	0	69	4	168	111	1	0	1	98	1	120	80	1	0
54	1	0	0	117	2	145	81	0	0	0	85	0	130	80	0	0
61	0	1	45	86.2	2	160	72	0	0	0	63	1	150	70	1	0
57	0	0	0	76	2	176	78	1	0	1	74	1	120	70	0	0
62	1	0	0	160	3	154	61	0	0	0	89	1	110	70	0	0
52	1	0	0	125	2	146	95	1	0	1	85	1	110	60	1	0
53	1	0	0	260	4	158	74	0	0	0	118	1	170	90	1	0
50	1	0	0	121	1	160	79	1	0	0	69	0	120	80	1	0

3.9.4 Dataset Splitting Strategy: The dataset was divided into training and testing sets using stratified sampling in order to assess the model's performance. To ensure that the distribution of classes in both subsets was uniform, stratified splitting was employed. The distribution of the dataset consisted of:

- Approximately 80% of the training data
- Approximately 20% of the testing data

The dataset of 297 valid records after preprocessing included an equal number of the two classes of non-disease and disease cases (approximately 54% and 46% respectively).

3.9.5 Genetic Algorithm Parameter Configuration

The SVM hyperparameters, including the kernel coefficient (γ) and regularisation parameter (C), were optimised using a genetic algorithm. To find the best set of parameters to maximise classification performance, the algorithm was run on multiple generations.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 2340-2345

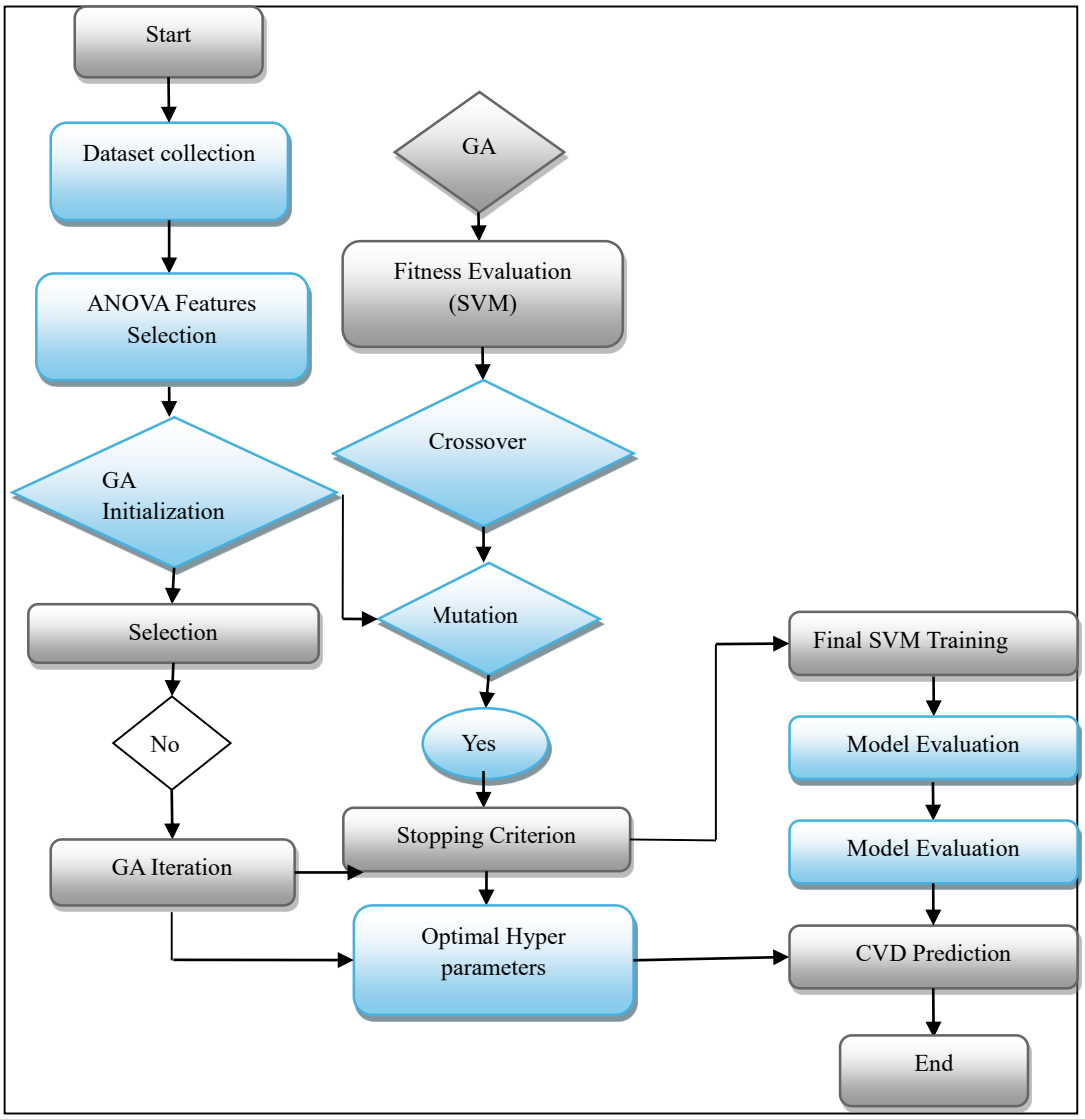

Figure 2. Flowchart of the proposed improved GA-tuned SVM model

Algorithm Steps:
1. Initialization: Create an initial chromosomal population at random.
2. Fitness Evaluation: Use cross-validation to gauge each chromosome's performance (fitness).
3. Selection: Use methods like tournament selection to pick the fit individuals for procreation.
4. Crossover: To produce kids, portions of two parental chromosomes are combined.
5. Mutation: To preserve variety, randomly modify genes.
6. Replacement: Select the best individuals from both the current population and the most recent offspring to form a new population.
7. Termination: Repeat steps 2 through 6 for a predetermined number of generations or until convergence.

The flow diagram will be expanded with all the steps, including the data preparation and model evaluation. The workflow of the proposed model is shown in Figure 2.

3.9.6 Model Training and Evaluation

The SVM classifier was optimized using the selected characteristics and the optimized hyperparameters obtained with the help of the Genetic Algorithm. The model performance was assessed based on conventional performance measurements such as accuracy, precision, recall, F1-score, and ROC-AUC. These assessment measures have given an overall understanding of diagnostic effectiveness and reliability of classification. Reproducibility and computational efficiency Python-based machine learning frameworks were used to perform preprocessing, optimizations and classification.

4. RESULTS AND DISCUSSION

The next section demonstrates the performance study of proposed improved GA-tuned Support Vector Machine (SVM -GA) model in cardiovascular disease prediction. The model's performance was evaluated using standard classification measures and compared to baseline machine learning methods like Random Forest, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), XGBoost, and CatBoost.

4.1 Metrics for Performance Evaluation

Standard evaluation metrics like accuracy, precision, recall, and F1-score have been used to assess the suggested model.

Accuracy is calculated as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Precision is defined as:

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

Recall (Sensitivity) is defined as:

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 2340-2345

The F1-score, representing the harmonic mean of Precision and Recall, is calculated as:

$$F1 = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (17)$$

These measures are an overall assessment of the accuracy of classification and prediction reliability.

4.2 Cross-Validation Outcome Analysis: The suggested enhanced GA-tuned SVM model attained an accuracy of 91.11%, that is indicative of high level of generalization in terms of validation datasets. The model achieved a precision of 91.77%, this implies that the model makes fewer false positive forecasts and this is of great value in medical diagnosis to avoid misdiagnosis of the target disease on a healthy person The model's ability to accurately identify true positive situations is demonstrated by its recall value of 91.11%. Additionally, the correctness of the suggested model is demonstrated by the F1-score of 91.02%, which shows an equal trade-off between precision and recall.

Table 3. Comparative performance analysis

Model	Precision (%)	Recall (%)	Accuracy (%)	ROC–AUC (%)
Random Forest	84.60	84.44	84.44	86.10
SVM	90.88	90.00	90.00	91.40
KNN	89.48	88.89	88.89	90.20
XGBoost	82.22	82.22	82.22	84.00
CatBoost	83.02	82.22	82.22	84.60
Improved GA-Tuned SVM	91.77	91.11	91.11	93.20

The results shown in Table 3 demonstrate that the recommended GA-tuned SVM model outperforms all baseline classifiers across all assessment metrics.

4.3 Comparison of Graphical Performance

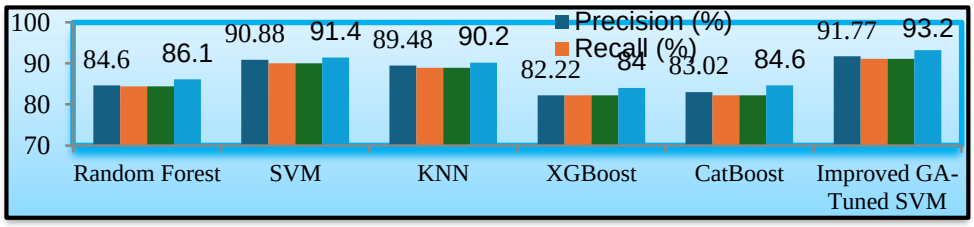

Figure 3. Graphical comparative performance analysis

The relative performance of different machine learning models is shown in Figure 3. Particularly in classifier performance and ROC-AUC rates, the recommended GA-tuned SVM model consistently outperforms the other models.

4.4 Confusion Matrix Analysis: The confusion matrix of the suggested GA-tuned SVM model and the traditional machine learning model are compared in Figure 4. When compared to other examined models, the suggested model had the lowest false positive (FP = 1) and false negative (FN = 7). Reduced false negativity lowers the frequency of cases of missed diseases, which makes false negatives particularly crucial for disease diagnosis. These results imply that the proposed approach would have higher diagnostic reliability and clinical utility.

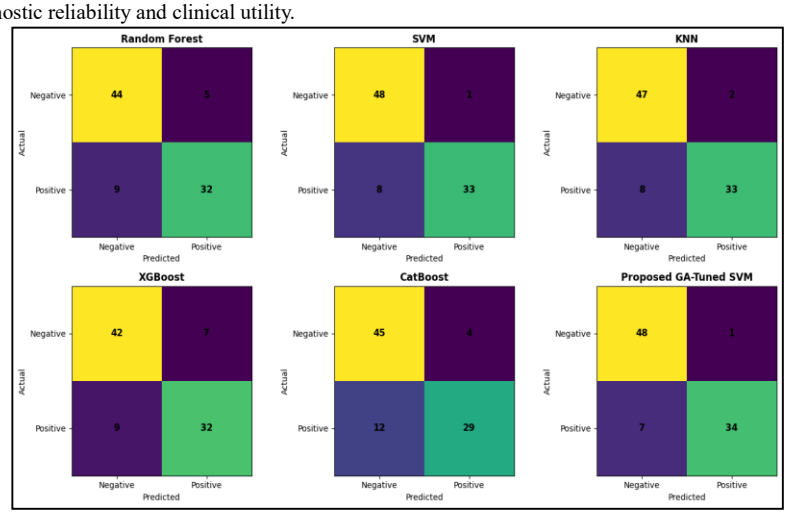

Figure 4. Confusion matrix comparison of classification models

4.5 ROC Curve Analysis: The classification performance at various threshold values were obtained by drawing Receiver Operating Characteristic (ROC) curves. The maximum ROC-AUC of 93.20%, was obtained with the proposed model of GA-tuned SVM, and this means that it has a high ability to discriminate between disease and non-disease classes. The ROC curve of the proposed model seems the closest to the top-left corner that is the ideal classifier performance.

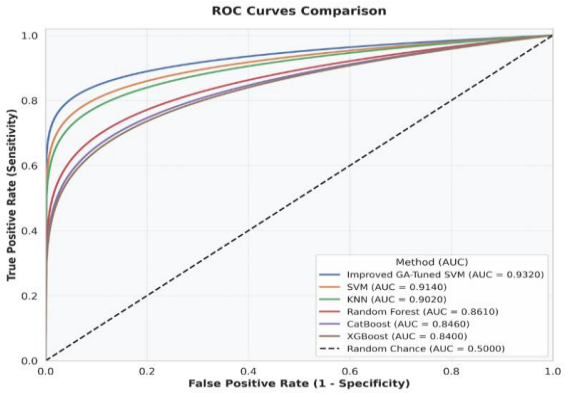

Figure 5. ROC curves for evaluated classification models

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 2340-2345

5. CONCLUSION

This study proposed a combined machine learning model that combines Support Vector Machine (SVM) classification and hyperparameter search using Genetic Algorithm (GA) to predict cardiovascular diseases. It was found that the suggested approach outperforms the traditional machine learning framework in terms of classification accuracy (91.11) and ROC -AUC (93.20). Evolutionary optimization made it possible to effectively tune the parameters, which led to a higher degree of model stability and prediction consistency. The developed framework is a scalable and reliable solution to cardiovascular disease early detection by doing feature selection with optimized classification. The results of the present research show that hybrid machine learning can be an essential part of the creation of intelligent healthcare systems and the assistance of clinical decision-making processes. The combination of data balancing strategies like Synthetic Minority Oversampling Technique (SMOTE) and ensemble learning strategies can be further employed in the future to improve on prediction performance, especially with imbalanced medical data.

ACKNOWLEDGMENT: The authors acknowledge their debt to the UCI Machine Learning Repository which has made publicly available cardiovascular disease datasets to be used in this study. The presence of standardized datasets also played a major role in the analysis of the experiment and the validation of the proposed model.

CONFLICT OF INTEREST : No conflict of interest.

ETHICAL ACCEPTANCE: This experiment has used publicly available anonymous datasets that are acquired in the open-source repositories; hence, no ethical consent was needed.

FUNDING INFORMATION: Authors state no funding involved.

DATA AVAILABILITY: The dataset used in this study is publicly available from the UCI Machine Learning Repository at: https://archive.ics.uci.edu/dataset/45/heart+disease

The processed data and experimental results supporting the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS STATEMENT *(mandatory)* (10 PT)

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Vibhav Krashan Chaurasiya	✓	✓	✓	✓	✓	✓		✓	✓		✓		✓	
Dr. Sakshi Rai	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

REFERENCES

[1] S. I. Ayon, M. M. Islam, and M. R. Hossain, "Coronary artery heart disease prediction: A comparative study of computational intelligence techniques," *IETE J. Res.*, vol. 68, no. 4, pp. 2488–2507, 2020, doi: 10.1080/03772063.2020.1713916.
[2] G. B. Berikol, O. Yildiz, and I. T. Özcan, "Diagnosis of acute coronary syndrome with a support vector machine," *J. Med. Syst.*, vol. 40, no. 4, p. 84, 2016.
[3] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, p. 88, 2023, doi: 10.3390/a16020088.
[4] A. C. Dimopoulos *et al.*, "Machine learning methodologies versus cardiovascular risk scores in predicting disease risk," *BMC Med. Res. Methodol.*, vol. 18, no. 1, p. 179, 2018, doi: 10.1186/s12874-018-0580-6.
[5] D. Elreedy and A. F. Atiya, "A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance," *Inf. Sci.*, vol. 505, pp. 32–64, 2019, doi: 10.1016/j.ins.2019.07.070.
[6] H. F. El-Sofany, "Predicting heart diseases using machine learning and different data classification techniques," *IEEE Access*, vol. 12, pp. 106146–106160, 2024, doi: 10.1109/ACCESS.2024.3437181.
[7] A. Gheorghe *et al.*, "The economic burden of cardiovascular disease and hypertension in low- and middle-income countries: A systematic review," *BMC Public Health*, vol. 18, no. 1, p. 975, 2018, doi: 10.1186/s12889-018-5806-x.
[8] H. Guo *et al.*, "ConfigX: Modular configuration for evolutionary algorithms via multitask reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 39, no. 25, pp. 26982–26990, 2025.
[9] A. Hammoud, A. Karaki, R. Tafreshi, S. Abdulla, and M. Wahid, "Coronary heart disease prediction: A comparative study of machine learning algorithms," *J. Adv. Inf. Technol.*, vol. 15, no. 1, pp. 27–32, 2024, doi: 10.12720/jait.15.1.27-32.
[10] W. Hussain *et al.*, "Ensemble genetic and CNN model-based image classification by enhancing hyperparameter tuning," *Sci. Rep.*, vol. 15, no. 1, p. 1003, 2025, doi: 10.1038/s41598-024-76178-3.
[11] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.
[12] V. V. R. Karna *et al.*, "A comprehensive review on heart disease risk prediction using machine learning and deep learning algorithms," *Arch. Comput. Methods Eng.*, vol. 32, no. 3, pp. 1763–1795, 2025, doi: 10.1007/s11831-024-10194-4.
[13] D. Krishnani *et al.*, "Prediction of coronary heart disease using supervised machine learning algorithms," in *TENCON 2019 - IEEE Reg. 10 Conf.*, 2019, pp. 367–372, doi: 10.1109/TENCON.2019.8929434.
[14] A. Masbakhah, U. Sa'adah, and M. Muslikh, "Heart disease classification using random forest and FOX algorithm as hyperparameter tuning," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 7, no. 4, pp. 964–976, 2025, doi: 10.35882/jeeemi.v7i4.932.
[15] I. D. Mienye, Y. Sun, and Z. Wang, "An improved ensemble learning approach for the prediction of heart disease risk," *Informatics Med. Unlocked*, vol. 20, p. 100402, 2020, doi: 10.1016/j.imu.2020.100402.
[16] F. Ming *et al.*, "Automated configuration of evolutionary algorithms via deep reinforcement learning for constrained multiobjective optimization," *IEEE Trans. Cybern.*, 2025, doi: 10.1109/TCYB.2025.3603251.
[17] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
[18] S. A. Nabofa-Ebiaredoh, "Improved machine learning methods for classification of imbalanced data," M.S. thesis, Univ. Johannesburg, Johannesburg, South Africa, 2022.
[19] S. Tiwari, N. Gupta, and P. Yadav, "Diabetes type-2 patient detection using Lasso based CFFNN machine learning approach," in *Proc. 8th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, 2021, pp. 602–608, doi: 10.1109/SPIN52536.2021.9565965.
[20] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," *Informatics Med. Unlocked*, vol. 26, p. 100655, 2021, doi: 10.1016/j.imu.2021.100655.
[21] S. Vernekar, S. Nair, D. Vijaysenan, and R. Ranjan, "A novel approach for classification of normal/abnormal phonocardiogram recordings using temporal signal analysis and machine learning," in *Comput. Cardiol. Conf. (CinC)*, 2016, pp. 1141–1144.
[22] S. Wan, F. Wan, and X. J. Dai, "Machine learning approaches for cardiovascular disease prediction: A review," *Arch. Cardiovasc. Dis.*, 2025, doi: 10.1016/j.acvd.2025.04.055.
[23] M. Wanyonyi *et al.*, "Enhanced machine learning and hybrid ensemble approaches for coronary heart disease prediction," *PLOS One*, vol. 20, no. 12, p. e0328338, 2025, doi: 10.1371/journal.pone.0328338.
[24] A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Heart disease dataset," *UCI Mach. Learn. Repository*, 1989, doi: 10.24432/C52P4X.

BIOGRAPHIES OF AUTHORS

Vibhav Krashan Chaurasiya is currently pursuing a PhD in Computer Science at LNCT University, Bhopal, India. His research interests include machine learning, artificial intelligence, healthcare analytics, deep learning, and intelligent medical decision support systems. His current research focuses on the development of optimized hybrid machine learning frameworks for cardiovascular disease prediction using evolutionary algorithms and feature selection techniques. He has worked on multiple research projects involving predictive analytics, healthcare data mining, and computational intelligence applications. He contributed to conceptualization, methodology development, implementation, experimentation, data analysis, and manuscript preparation for this study. He can be contacted at email: joyvib@gmail.com.

Dr. Sakshi Rai is an Associate Professor in the Department of Computer Science at LNCT University, Bhopal, India. Her research interests include artificial intelligence, machine learning, data science, cloud computing, and intelligent healthcare systems. She has guided several undergraduate, postgraduate, and doctoral research projects in emerging computational technologies. Her academic contributions include research in predictive analytics, optimization techniques, and advanced computing methodologies. In this study, she contributed to research supervision, methodology validation, manuscript review, and technical guidance. She can be contacted at email: sakshi@lnct.ac.in.