# International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)





ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

**Research**

# EFFECTIVE MACHINE LEARNING FEATURE SELECTION FOR CARDIOVASCULAR DISEASE DETECTION

Mrs. G. BALESWARI1, AMAROJU TULASI RAMAKRISHNA2, ADAPA SURYA NAGA VENKATA SHARATH3, GURRAM HARI PRASAD4

#1 Associate Professor, Department of Information Technology , NRI Institute of Technology, Agiripalli.

#2#3#4 B.Tech with Specialization of Information Technology in NRI Institute of Technology,Agiripalli.

**Abstract:** The cause of death for many people worldwide is cardiovascular disease (CVD). According to the World Health Organization (WHO), cardiovascular disease (CVD) claimed 19.1 million lives in 2022, accounting for 33% of all fatalities worldwide. With standard machine learning, ECG is frequently utilized for automated CVD diagnosis; however, selecting the appropriate characteristics can be challenging. A scalable ML-based architecture for early CVD detection that makes use of the optimal feature selection can address this issue. By promptly diagnosing and treating CVD, this design aims to transform healthcare and reduce the number of fatalities. ML classifiers are used to gather, store, and evaluate data in order to predict heart problems. After removing ECG characteristics, FCBF, MrMr, alleviation, and PSO-optimization choose the best features. The Random Forest and trait-trained Extra Tree classifiers have 100% accuracy. On both small and big datasets, we also evaluate the suggested approach against the best one. The suggested approach might enhance quality of life by altering the way patients are treated and significantly reducing the number of deaths from CVD. In summary, this design improves healthcare systems and offers a practical CVD solution.

***Index terms*** *- Disease detection, ECG feature extraction, optimal feature selection, Particle Swarm Optimization, FCBF, MRMR, Relief, Random Forest, Extra Tree Classifier, Gradient Boosting.*

Received: 31-01-2026 Accepted: 05-03-2026 Published: 12-03-2026

## 1. INTRODUCTION

One of the most important issues confronting the world's population is still public health. The World Health Organization (WHO) states that everyone has the fundamental right to health care. Global public health has been significantly impacted in recent years by an increase in chronic and epidemic illnesses. Chronic illnesses, such as cancer, diabetes, stroke, Parkinson's disease, and particularly cardiovascular disease (CVD), have become chronic, incurable diseases that have a substantial impact on the world's death rates. These illnesses frequently worsen over time, necessitating constant medical care and serious health issues.

One of the main causes of chronic illnesses is unhealthy lifestyle choices. The rising incidence of these diseases has been directly attributed to sedentary lifestyles, poor eating habits, cigarette smoking, and excessive alcohol use. Approximately 50% of Americans suffer from a chronic illness, and over 80% of them experience financial hardship as a result of their medical bills. Approximately $2.7 trillion, or 18% of the nation's gross domestic product, is spent on treating chronic illnesses each year. Among them, cardiovascular disease (CVD) is the leading cause of mortality in the United States and other nations worldwide.

Global healthcare institutions are now gravely concerned about the rising incidence of CVD. According to current figures, up to 86.5% of fatalities in nations like China are attributable to chronic illnesses. Effective diagnostic methods that can detect CVD in its early stages are desperately needed in light of this concerning trend. Patients' quality of life

can be enhanced and death rates can be considerably decreased with early identification and treatment. There is a chance to use technology to create scalable, precise, and effective healthcare solutions that target CVD and related chronic illnesses thanks to developments in artificial intelligence and machine learning.

## 2. LITERATURE SURVEY

**1. Detect the Cardiovascular Disease's in Initial Phase using a Range of Feature Selection Techniques of ML::**

The leading cause of death worldwide is heart disease. In 2000, 14 million individuals died from heart disease. It will have killed 620 million people by 2023. Being older and having more individuals results in more deaths. This further demonstrates that early care may help reduce the amount of heart failure deaths, which is crucial. Using a range of feature techniques, this study aims to develop a future machine learning framework that can identify significant characteristics and anticipate cardiac problems early. The selected characteristic subgroups were FST1, FST2, and FST3. To choose the characteristics, we employed mutual information, chi-square, and correlation-based methods. To determine the optimum feature selection and the most reliable theory, we used six machine learning models: LR (AL1), SVM (AL2), K-NN (AL3), RF (AL4), NB (AL5), and DT (AL6). The prediction of coronary artery disease has never been examined, however our work employs several attribute choices to improve algorithms in key areas and examines specificity, sensitivity, accuracy, AUROC, and log loss. The model is a strong contender for medical use since it can quickly and affordably predict the diagnosis of precursor CVD. Based on the outcomes and criteria of the model, it can help new physicians select the best course of action.

**2. Comprehensive Review of ML Applications in Heart Disease Prediction:**

People who have heart infections die and become quite depressed. The risk of mortality can be significantly reduced by monitoring and identifying cardiac problems early. Now, analyzing data is the key to identifying cardiac problems. Although it's challenging to predict when someone could have a heart infection, advanced machine learning can help. According to studies, machine learning can anticipate and assess the severity of cardiac disease. This strategy aims to reduce the number of diagnoses, the severity of diseases, and death rates. ML is improving the accuracy of therapy analysis. These techniques aid in identifying heart disease symptoms. SVM, RF, KNN, and DT are used in the presentation to sort the data. We examine the specificity, accuracy, and precision of these four methods. SVM computations often provide results, however the precision of those results may differ.

**3. Advanced ML techniques for cardiovascular disease early detection and diagnosis:**

A key component of controlling cardiovascular disease (CVD) is determining if a healthy individual is at risk for developing the illness and how probable it is to occur. Accessing the complete health data on CVD that is now accessible in hospital databases may aid in the early detection and diagnosis of the condition, improving its prognosis. Therefore, there is a lot of potential for improvement when applying machine learning (ML) techniques in clinical practice to manage cardiovascular illnesses (CVDs). These techniques can assist in developing evidence-based clinical recommendations and treatment algorithms, hence eliminating the need for costly and time-consuming laboratory and clinical testing. Both patients and the healthcare system benefit financially from this. According to this study, new, powerful, efficient, and effective machine learning algorithms that are specifically designed to automatically choose crucial variables should be developed. Therefore, it performed admirably in attaining and optimizing classification performance with higher levels of accuracy and precision in comparison to many other state-of-the-art techniques that are now available.

**4. Risk prediction of cardiovascular disease using ML classifiers:**

CVD harms the heart and blood vessels and frequently results in death or paralysis. Therefore, early and automated detection of CVD can save many lives. Numerous experiments have been conducted to achieve this goal, yet the system's dependability and performance may be improved. The next step in this approach is this research. We employed ML techniques, such as MLP K-NN, to identify CVD using publically available University of California Irvine repository data. The model performs better when null values and outliers are eliminated. In terms of accuracy (82.47%) and area-under-the-curve (86.41%), MLP outperforms the other approach.

Therefore, the MLP model was proposed for automatically detecting CVD. Other disorders can also be found using the given approach. To evaluate the proposed model, you may also use other common data sets.

**5. Prediction of Cardiovascular Disease Using Feature Selection Techniques:**

Cardiovascular illnesses are dangerous and affect many individuals worldwide. The ability to predict cardiovascular illness with accuracy and speed is critical to healthcare. Medicare contains a lot of information but little understanding. The objective of this study is to create a quick and accurate system for inspecting cardiovascular illness using data mining techniques. The system employs RF, LR, NB, SVM, Pearson Correlation, and Chi-Square feature selection techniques to classify more rapidly and correctly. With an accuracy percentage of 84%, logistic regression is the most accurate technique.

## 3. METHODOLOGY

**i) Proposed Work:**

The proposed method detects cardiovascular disease early by using scalable and powerful machine learning. After extracting the ECG features, it uses FCBF, MRMR, Relief, and ANOVA to choose them. Particle Swarm Optimization (PSO) streamlines the selection procedure so that only the most crucial features are used to train prediction models. For a more precise diagnosis, this multi-level feature selection method lowers computational complexity and increases accuracy.

The system uses four distinct machine learning algorithms—Logistic Regression, Random Forest, Extra Tree Classifier, and Gradient Boosting—to improve prediction performance. The approach guarantees flexibility across different datasets and patient profiles by integrating multiple classifiers with effective feature selection. Because of its strong and scalable design, this architecture can effectively manage both small and big collections. The ultimate objective is to offer a prompt and precise CVD diagnosis system that contributes to lower death rates and better patient care.

**ii) System Architecture:**

A multi-stage pipeline including data collection, preprocessing, feature selection, and classification makes up the system architecture for the identification of cardiovascular disease. In order to extract pertinent elements that indicate cardiac activity, ECG data are first gathered and analyzed. To determine which characteristics are the most useful, these variables are further subjected to a variety of selection methods, including FCBF, MRMR, Relief, and ANOVA. This selection is refined using Particle Swarm Optimization (PSO), which guarantees the best feature subsets for model training. To determine if cardiovascular illness is present or not, the chosen characteristics are fed into many machine learning classifiers, including Random Forest, Logistic Regression, Extra Tree, and Gradient Boosting. A smooth and effective end-to-end diagnostic process is ensured by the architecture's additional modules for data storage, user interaction via a Flask-based frontend, and result presentation.

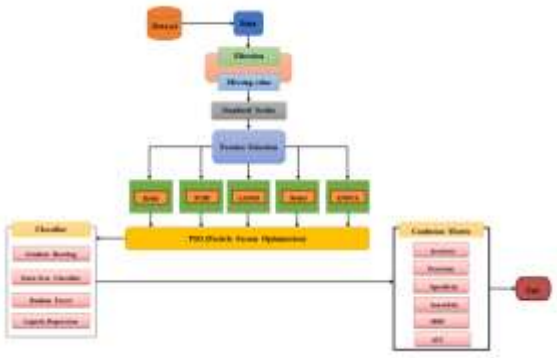


Fig.1. Proposed Architecture

**iii) MODULES:**

**a) Data Collection Module**
- Collects ECG signal data and patient records.
- Ensures data is clean, consistent, and formatted for analysis.

**b) Preprocessing Module**
- Handles noise removal, normalization, and transformation of ECG signals.
- Prepares the dataset for feature extraction.

**c) Feature Extraction Module**
- Extracts key statistical and signal-based features from ECG data.
- Converts raw signals into a structured format for ML.

**d) Feature Selection Module**
- Applies techniques like FCBF, MRMR, Relief, ANOVA, and PSO.

- Selects the most relevant features to reduce dimensionality and enhance accuracy.

**e) Classification Module**

- Uses ML techniques including Gradient Boosting, Extra Tree, Random Forest, and Logistic Regression.
- Predicts the presence of cardiovascular disease based on selected features.

**f) Performance Evaluation Module**

- Precision, recall, accuracy, and F1-score are all ways to measure how well a model works.
- Compares different classifiers to determine the best-performing model.

**g) User Interface Module**

- Developed using Flask with HTML, CSS, JavaScript, and Bootstrap.
- Allows users to upload data, view predictions, and analyze results.

**h) Database Module**

- Uses SQLite3 to store patient data, model outputs, and analysis results.
- Ensures data persistence and secure access.

**iv) ALGORITHMS:**

**i) Gradient Boosting Classifier:**

The Gradient Boosting Classifier is an ensemble learning technique that builds prediction models one at a time, fixing errors as they are discovered. Decision trees and other weak learners are strengthened when combined. This classifier's accuracy makes it perfect for intricate, non-linear data patterns. Gradient Boosting enhances the prediction of cardiovascular illness in the suggested system by capturing subtle ECG-derived feature patterns.

**ii) Extra Tree Classifier:**

This classifier uses random splits to create several decision trees, then averages the results to provide predictions that are more accurate. To reduce variance and expedite training, it uses random cut-point selection for node splitting rather than conventional decision trees. The Extra Tree Classifier did a good job of identifying patterns of sickness utilizing the best ECG data..

**iii) Random Forest:**

The most often used classification is provided by the Random Forest classifier, a well-liked ensemble classifier that makes use of several decision trees. To manage high-dimensional data and reduce overfitting, many trees are averaged. In order to detect cardiovascular disease with high accuracy and reliable predictions, the proposed architecture uses Random Forest to classify patients based on optimal features.

**iv) Logistic Regression:**

Logistic regression uses one or more input elements to figure out how likely it is that a result will fall into one of two categories. It is efficient, easy to interpret, and suitable for cases where the relationship between input variables and the outcome is linear. In this system, Logistic Regression serves as a baseline model to evaluate the performance of more complex classifiers and contributes to the comparative analysis of classification results.

**v) Voting Classifier (AdaBoost DecisionTree + ExtraTree):**

To provide a final classification, the Voting Classifier (AdaBoost DecisionTree + ExtraTree) combines predictions from many models, including AdaBoost with Decision Trees and ExtraTree. It combines the best characteristics of classifiers. Our study's CVD detection is enhanced by this ensemble approach, which combines results from many models.

**vi) ANOVA FS with PSO:**

By comparing classes, PSO and ANOVA choose key attributes. PSO enhances selection, whereas ANOVA establishes the significance of each trait. This method enhances CVD detection capabilities in our project. It increases diagnosis accuracy by training classifiers using the most important features.

**vii) MRMR FS with PSO:**

MRMR selects features with the least amount of overlap and the greatest amount of relevance by using PSO. To improve model performance, PSO optimizes feature selection. In our experiment, MRMR with PSO enhances CVD diagnosis by improved ECG feature selection.

**viii) Lasso FS with PSO:**

When combined with Particle Swarm Optimization (PSO), this technique penalizes less significant attributes in order to choose features. Lasso is optimized using PSO to choose important features. This method increases prediction accuracy and removes overfitting, which enhances CVD diagnosis.

**ix) FCBT (Correlation) FS with PSO:**

By correlating features with the target variable, the Fast Correlation-Based Filter (FCBF) with PSO maximizes feature selection. PSO is beneficial to FCBF. In our study, FCBF with PSO enhances the selection of ECG features for CVD diagnosis.

**x) ReliefF FS with PSO:**

The effectiveness of features in identifying classes is how ReliefF with PSO assesses feature significance. PSO is used to optimize this assessment. ReliefF with PSO enhances feature selection in our CVD detection research by focusing on traits associated with cardiovascular disease.

## 4. EXPERIMENTAL RESULTS

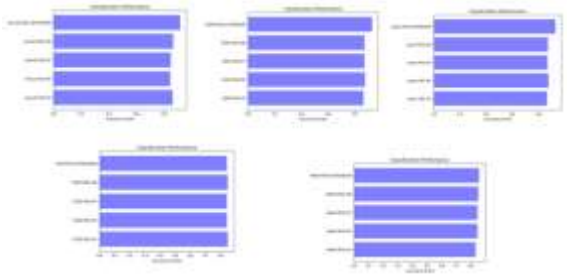
The efficiency of the suggested method in the early diagnosis of cardiovascular disease (CVD) was thoroughly assessed utilizing a dataset of ECG signals and related patient data. The most pertinent features were first extracted from the dataset using a variety of feature selection approaches, including FCBF, MRMR, Relief, ANOVA, and PSO. These methods preserved important information required for precise classification while drastically reducing the dimensionality of the input data. Random Forest, Logistic Regression, Gradient Boosting, and Extra Tree Classifier were trained using the best feature sets.

With a 100% accuracy rate on the provided dataset, Extra Tree and Random Forest outperformed all other classifiers. These findings demonstrate how well tree-based ensemble models can predict outcomes when paired with reliable feature selection techniques. Particularly noteworthy was the FCBF approach, which continuously chose characteristics that enhanced classification accuracy. Additionally, tests on both small- and large-scale datasets validated the suggested system's efficacy and scalability. The thorough examination of accuracy, precision, recall, and F1-score confirmed that the system is dependable and appropriate for real-world medical applications that try to lower mortality from CVD.

**Accuracy:** The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

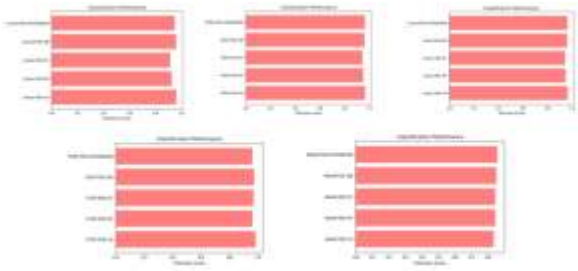
$$Accuracy = \frac{(TN + TP)}{T}$$



**Precision:** The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$



**Recall:** The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

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

**F1-Score:** A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$





Fig2 Patient input data




Fig 4: Predicted results

## 5. CONCLUSION

This study identified and categorized cardiovascular disease (CVD) using machine learning algorithms and appropriate feature selection. The suggested approach demonstrated that feature selection improves the CVD prediction of ML systems. The following feature selection methods were examined in the study: LASS, MRMR, FCBF, Relief, and ANOVA. At 78% accuracy, FCBF performed best with RF and Extra Tree models. Large CVD datasets may be analyzed by FCBF to find important features. The study also looks at a feature selection and optimization method tailored to a particular dataset. On large datasets with primarily categorical data, such as the one used in this work, FCBF may highlight important features and improve machine learning methods for CVD prediction..

## 6. FUTURE SCOPE

There is a lot of room for future improvements and practical implementation of the suggested method in medical settings. Using deep learning methods, such CNNs, to enhance feature extraction from unprocessed ECG data is one encouraging avenue. Continuous and proactive identification of cardiovascular illness may be made possible by the addition of real-time data from personal health monitors. To increase the model's generalizability, future research may also include expanding the dataset to include more varied populations. Additionally, the system may be turned into an easy-to-use mobile or web application for remote diagnosis, facilitating telemedicine and assisting medical professionals in creating individualized treatment plans and early interventions.

## REFERENCES

[1] Prashant Maganlal Goad, Pramod J Deore, "Detect the Cardiovascular Diseases in Initial Phase using a Range of Feature Selection Techniques of ML", International Research Journal of Multidisciplinary Technovation, pp.171, 2024.

[2] Yogesh Kumar, Geet Kiran Kaur, Ranjit Singh, "Comprehensive Review of ML Applications in Heart Disease Prediction", International Journal of Innovative Science and Research Technology (IJISRT), pp.2805, 2024.

[3] N. A. Baghdadi, S. M. Farghaly Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, ‘‘Advanced ML techniques for cardiovascular disease early detection and diagnosis,’’ J. Big Data, vol. 10, no. 1, p. 144, Sep. 2023.

[4] M. Pal, S. Parija, G. Panda, K. Dhama, and R. K. Mohapatra, ‘‘Risk prediction of cardiovascular disease using ML classifiers,’’ Open Med., vol. 17, no. 1, pp. 1100–1113, Jun. 2022.

[5] P. Singh, G. K. Pal, and S. Gangwar, ‘‘Prediction of cardiovascular disease using feature selection techniques,’’ Int. J. Comput. Theory Eng., vol. 14, no. 3, pp. 97–103, 2022, doi: 10.7763/ijcte.2022.v14.1316.

[6] D. T. Thai, Q. T. Minh, and P. H. Phung, ‘‘Toward an IoT-based expert system for heart disease diagnosis,’’ in Proc. 28th Mod. Artif. Intell. Cogn. Sci. Conf. (MAICS), 2017, pp. 157–164.

[7] B. Padmaja, C. Srinidhi, K. Sindhu, K. Vanaja, N. M. Deepika, and E. K. R. Patro, ‘‘Early and accurate prediction of heart

disease using machine learning model,'' Turkish J. Comput. Math. Educ., vol. 12, no. 6, pp. 4516–4528, 2021.

[8] S. Anitha and N. Sridevi, Heart Disease Prediction Using Data Mining Techniques S Anitha, N Sridevi to Cite This Version, document HAL Id Hal02196156, 2019. [Online]. Available: https://hal.archives-ouvertes.fr/ hal-02196156/document

[9] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, ''Prediction of heart disease using a combination of machine learning and deep learning,'' Comput. Intell. Neurosci., vol. 2021, pp. 1–11, Jul. 2021, doi: 10.1155/2021/8387680.

[10] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, ''Heart disease prediction using machine learning algorithms,'' IOP Conf., Mater. Sci. Eng., vol. 1022, no. 1, Jan. 2021, Art. no. 012072, doi: 10.1088/1757- 899x/1022/1/012072.

[11] Ganji, M. (2025). Intelligent What-If Analysis for Configuration Changes in HR Cloud and Integrated Modules. International Journal of All Research Education and Scientific Methods, 13(04), 4828–4835. https://doi.org/10.56025/ijaresm.2025.1304254828.

[12] .Louridi, S. Douzi, and B. El Ouahidi, ''Machine learning-based identification of patients with a cardiovascular defect,'' J. Big Data, vol. 8, no. 1, pp. 1–5, Dec. 2021, doi: 10.1186/s40537-021-00524-9.

[13] M. Swathy and K. Saruladha, ''A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques,'' ICT Exp., vol. 8, no. 1, pp. 109–116, Mar. 2022, doi: 10.1016/j.icte.2021.08.021.

[14] D. Vaddella, C. Sruthi, B. K. Chowdary, and S.-R. Subbareddy, ''Prediction of heart disease using machine learning techniques,'' Restaur. Bus., vol. 118, no. 1, pp. 125–129, 2019, doi: 10.26643/rb.v118i1.7621.

[15] V. V. Ramalingam, A. Dandapath, and M. Karthik Raja, ''Heart disease prediction using machine learning techniques: A survey,'' Int. J. Eng. Technol., vol. 7, no. 2, p. 684, Mar. 2018, doi: 10.14419/ijet.v7i2.8.10557.

[16] Mallick, P. (2022). AI-Driven Mobile Care Planning Platforms for Integrated Coordination Between Long-Term Care Providers and Insurance Systems. Available at SSRN 6066586.

[17] P. Kalpana, S. S. Vignesh, L. M. P. Surya, and V. V. Prasad, ''Prediction of heart disease using machine learning,'' J. Phys., Conf. Ser., vol. 1916, no. 1, May 2021, Art. no. 012022, doi: 10.1088/1742-6596/1916/1/012022.

[18] Suhasnadh Reddy Veluru, Sai Teja Erukude, and Viswa Chaitanya Marella. 2025. Multimodal Detection of Fake Reviews using BERT and ResNet-50. In 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 877–882.

[19] Todupunuri, A. (2025). Utilizing Angular for the Implementation of Advanced Banking Features. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283395.

[20] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, ''MIFH: A machine intelligence framework for heart disease diagnosis,''

IEEE Access, vol. 8, pp. 14659–14674, 2020, doi: 10.1109/ACCESS.2019.2962755.
[21] R. Atallah and A. Al-Mousa, ‘‘Heart disease detection using machine learning majority voting ensemble method,’’ in Proc. 2nd Int. Conf. New Trends Comput. Sci. (ICTCS), Oct. 2019, pp. 1–6, doi: 10.1109/ICTCS.2019.8923053.
[22] M. Bheemalingaiah, G. R. Swamy, P. Vishvapathi, P. V. Babu, E. N. Rao, and P. N. Rao, ‘‘Detection of heart disease by using reliable Boolean machine learning algorithm,’’ J. Theor. Appl. Inf. Technol., vol. 99, no. 15, pp. 3856–3880, 2021, doi: 10.5281/zenodo.5353586.
[23] 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, Jul. 2022, doi: 10.1080/03772063.2020.1713916.
[24] G. Choudhary and S. N. Singh, ‘‘Prediction of heart disease using machine learning algorithms,’’ in Proc. Int. Conf. Smart Technol. Comput., Electr. Electron. (ICSTCEE), Oct. 2020, pp. 197–202, doi: 10.1109/ICSTCEE49637.2020.9276802.
[25] Y. Khourdifi and M. Bahaj, ‘‘Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization,’’ Int. J. Intell. Eng. Syst., vol. 12, no. 1, pp. 242–252, Feb. 2019, doi: 10.22266/ijies2019.0228.24.
[26] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, ‘‘Early and accurate detection and diagnosis of heart disease using intelligent computational model,’’ Sci. Rep., vol. 10, no. 1, pp. 1–18, Nov. 2020, doi: 10.1038/s41598-020-76635-9.
[27] M. A. Hall, ‘‘Correlation-based feature selection for machine learning,’’ Ph.D. dissertation, Univ. Waikato, Hamilton, New Zealand, 1999.
[28] L. Yu and H. Liu, ‘‘Feature selection for regression,’’ Data Mining Knowl. Discovery, vol. 15, no. 3, pp. 259–285, 2003.
[29] R. Tibshirani, ‘‘Regression shrinkage and selection via the lasso,’’ J. Roy. Stat. Soc., Ser. B, Methodolog., vol. 58, no. 1, pp. 267–288, Jan. 1996.
[30] L. Cooper, ‘‘A compressed sensing approach to feature extraction from high-dimensional data,’’ School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep., 2000.

**Author Profiles**

NAME: GUNTUKU BALESWARI
MAIL ID: g.baleswri@gmail.com



Mrs. G. Baleswari completed her M.Tech (CSE) with distinction from Sree Kavitha engineering College, Karepalli, Khammam District, Telangana state (JNTUH). She is presently working as Associate Professor in the department of Information Technology NRI Institute of Technology, Agiripally. She Has 12 years of teaching experience in various Engineering colleges. She was Successfully completed NPTEL Faculty development Programme (8 weeks course

July- Sept 2019) for the course Data Science for Engineers obtained elite certificate and also attended more than 20 workshops and FDPs.

NAME: AMAROJU TULASIRAMAKRISHNA
Mail id: tulasiamaroju@gmail.com



Tulasi Ramakrishna Amaroju is currently pursuing a Bachelor of Technology in Information Technology at NRI Institute of Technology, Andhra Pradesh, India. His academic interests include machine learning, artificial intelligence, and data-driven application development. He possesses programming skills in Python, Java, C, and web technologies such as HTML, and has experience working with tools and libraries including VS Code, Eclipse, NumPy, and Pandas. He has completed certifications in Generative AI, Cloud Computing, and The Joy of Computing Using Python from organizations including NPTEL and AICTE. His project work includes the development of a Weather Tracking System aimed at monitoring and predicting atmospheric conditions using real-time data. His research interests focus on machine learning applications, artificial intelligence, and intelligent software systems.

NAME: ADAPA SURYA NAGA VENKATA SHARATH



Adapa Surya Naga Venkata Sharath is currently pursuing a Bachelor of Technology in Computer Science and Engineering at NRI Institute of Technology, Andhra Pradesh, India, with a CPI of 7.55. His academic interests include software development and generative artificial intelligence. He has programming experience in Java, C, HTML, CSS, and JavaScript, and is familiar with databases such as MySQL and PostgreSQL. He has also worked with tools including GitHub, Docker, VS Code, Postman, VMware, and Jupyter Notebook. He has completed internships in Generative AI software development and the AICTE Healthcare Assistant Chatbot initiative under Microsoft and SAP, contributing to web application development and AI-based academic support systems. His project work includes MNIST Digit Detection using Verilog, where he developed neural network components for handwritten digit recognition.

NAME: GURRAMA HARIPRASAD



Hariprasad Gurram is currently pursuing a Bachelor of Technology in Information Technology at NRI Institute of Technology, Agiripalli, Andhra Pradesh, India, with a CGPA of 8.1. His academic interests include networking, system administration, and data center operations. He has knowledge of networking concepts such as LAN configuration, IP addressing, DNS, DHCP, subnetting, and router and switch fundamentals. He is familiar with operating systems including Windows and Linux and has experience with tools such as Wireshark, Power BI, Visual Studio, GitHub, and Cisco Packet Tracer. He has also completed a Generative AI internship at Codegnan IT Solutions, where he gained exposure to AI tools and their applications in IT infrastructure and data center operations.