Research Article

Int J. Curr Inn Adv Res, Vol: 7, Issue: 1, 2024; 14-22

International Journal of Current Innovations in Advanced Research

(International Multidisciplinary Tri-annual Research Journal)

Content Available at www.ijciar.com ISSN (O) 2636-6282 ISSN (P) 2659-1553

Optimization Heart Disease Prediction Using Independent Component Analysis and Support Vector Machine

Abbas Khalifa Nawar

Department of Computer Techniques Engineering, Imam Al-Kadhum College (IKC), Baghdad, Iraq

abbas.altimimy@iku.edu.iq

DOI: https://doi.org/10.47957/ijciar.v7i1.168

Received: 15 Jan 2023 Revised: 28 Feb 2023 Accepted: 20 Mar 2024

Abstract

Prediction models play a crucial role in early detection and intervention for cardiac diseases. However, their effectiveness is often hindered by limitations. inherent in current methodologies. This paper proposes a novel approach. to address these challenges by integrating Independent .Component Analysis (ICA) with the Support Vector Machine (SVM) technique. Utilizing a comprehensive Cleveland dataset, our model achieves notable performance metrics, including an accuracy of 90.16%, an Area Under the Curve (AUC) of 96.66%, precision of 90.02%, recall of 90.00%, F1-score of 90.00%, and a minimal log loss of 3.54. Our methodology not only. surpasses previous. methodologies through extensive. comparative analysis but also. addresses common constraints identified in existing literature. These limitations encompass insufficient feature representation, overfitting, and a lack of proactive intervention strategies. By amalgamating ICA with SVM, our model. enhances feature extraction, mitigates overfitting, and facilitates proactive. diagnosis and intervention in. individuals suspected of having heart. disease. This study underscores the importance. of mitigating current literature limitations and underscores the potential of integrating contemporary machine learning. techniques to advance prediction models for heart disease.

Keyword: Confusion matrix, Heart disease, ICA, SVM, Feature selection, Heart disease, ICA, SVM, Feature selection

1. Introduction

Recent advances in artificial intelligence. (AI) and machine learning (ML) in healthcare have sparked debate. regarding these technologies' potential. to replace human physicians. However, the prevalent view. is that AI and machine learning are complimentary to human. physicians, complementing rather than replacing their expertise. The widespread availability of healthcare data. is driving the incorporation. of these technologies into medicine [1]. There is a rising interest in harnessing AI and machine learning for heart attack prediction, because these technologies have the capacity to automate signal interpretation, they produce accurate. and reliable forecasts. Predicting heart attacks. is critical .in healthcare, requiring the development. of novel solutions targeted at prevention. and general improvement in individual health and well-being. [2]. Machine Learning (ML), a subset of Artificial. Intelligence (AI), focuses on building .algorithms. and statistical models that. allow computers to gain learning skills.And. improve their performance. in certain tasks via experience, without the need for explicit .programming.Rather than being explicitly .coded for each job, machine learning .(ML) systems use data to recognize patterns, learn from instances, and make educated judgments .based on data insights [3].ML is critical in the analysis. and prediction of heart attacks, providing enhanced. approaches for early identification and. preventative measures [6]. ML algorithms excel at managing large datasets. and deciphering subtlee patterns, making themi ideal for investigating numerous. aspects of heart health. Furthermore, ML models may examine. several risk variables, including. demographic data, medical histories, lifestyle. decisions, and genetic predispositions, to forecast.ann individual's chance .of having a heart attack. This capacity allows. for tailored risk assessments and targeted. treatments. Feature extraction, a crucial stage in. data mining, is choosing or. manipulating significant features from raw data to improve. the performance. of machine. learning algorithms [4].
When. Integrated. With. Machine. learning techniques, ICA significantly. Improves. Result. accuracy [5]. It is a precise. And. crucial component of the data mining process, particularly for feature extraction [6]. ICA may be used to find significant patterns or sources.. within. a dataset, which are known. as independent. components. These. components capture. Underlying. structures that may not be seen. in the raw. data. When these.components .are employed. as features for.future .analysis or classification tasks, machine learning algorithms may leverage their representational skills, resulting in enhanced accuracy and performance across a wide range of applications. This study.stresses .the major. Benefits. of implementing. ICA into data

.mining technologies, as it is .especially .effective in. achieving .optimal .result accuracya when teaming withML algorithms, revealing criteical and important .patterns within the data [7].

This paper presents a unique hybrid technique for forecasting heart attacks that combines Independent Component Analysis (ICA) and Gradient Boosting Classifier (GBC) to leverage their complementing capabilities.

1- This study. introduces a pioneering hybrid .technique for predicting heart attacks. This study leverages. the synergies between Independent .Component Analysis (ICA) and Gradient Boosting. Classifier (GBC) to improve the. accuracy and efficacy of forecasting cardiac attacks.
2- The integration of ICA as a. preprocessing step aid in extracting relevant. features from raw data, reducing noise, and improving interpretability. Meanwhile, the Gradient .Boosting Classifier, known for its .powerful predictive. capabilities, serves as the engine for accurate risk assessment.
3- This unique hybrid technique .has the potential to significantly. enhance the present landscape. of heart attack prediction approaches. The model tries to de.liver more accurate and precise predictions. by merging sigbnal processing techniques. with machine learning algorithms.

These contributions are especially important. in the context of cardiovascular health, where early and precise detection of heart attacks can. considerably enhance preventative. measures and patient outcomes. The methodology used in .the study is likely to pave the way for. additional advances in cardiovascular .disease prediction, demonstrating the promise of hybrid approaches in healthcare research.

2. Literature Study

In this exposition, we look at the gaps .found in previous investigations, culminating in the. development of this study report. The discussion focuses on the approach. and dataset used, with a special emphasis .on notable research projects and. techniques that use machine learning .to improve the accuracy. and efficiency of such systems.

[8]The backpropagation technique in a multilayer perceptron (MLP) of Artificial Neural Networks (ANN) is employed for heart disease prediction. The outcomes are assessed against those of other models in the field, revealing enhancements in performance.[9]Data from heart disease patients gathered at the UCI laboratory is analyzed using neural networks (NN), decision trees (DT), Support Vector Machines (SVM), and Naive Bayes to identify patterns. Performance and accuracy are compared among these algorithms. The introduced hybrid approach achieves an F-measure of 86.8%, demonstrating competitiveness with established methods.[10]analysed the idea of ensemble classification, a technique used to increase the potency of weaker algorithms by merging several classifiers. The study used comparative analysis to see whether applying ensemble techniques may increase the predicted accuracy of heart disease diagnosis. The investigation's findings show that ensemble methods like bagging and boosting had an accuracy rate of 85.48%.[9]They suggested a novel method employing machine learning techniques to enhance the precision of cardiovascular disease prognosis. This predictive model integrates various feature combinations and classification methods, leading to improved performance. By utilizing a hybrid Random Forest and linear model (HRFLM), it achieves an accuracy rate of 88.7%.[11]implemented RF classifier algorithm, achieving approximately 83% accuracy in identifying heart diseases. The application complied with HIPAA regulations, ensuring medical information safety. The study discusses various ML algorithms, with the RF algorithm selected for heart disease detection, yielding significant accuracy results of about 83%.[12]They presented a pioneering training method that merges a multilayer perceptron (MLP) with particle swarm optimization (PSO) to detect heart disease. This technique underwent evaluation alongside ten distinct machine learning algorithms for heart disease prediction, utilizing diverse metrics. Results demonstrated that the MLP-PSO fusion outperformed all other methods, achieving an accuracy rate of 84.61%.[13]They presented a supervised machine learning technique for predicting the occurrence of cardiovascular disease (CVD), highlighting the efficacy of utilizing the SMOTE technique and investigating the significance of risk factors in CVD prognosis. Results demonstrated that the Stacking ensemble model, when integrated with SMOTE and 10-fold cross-validation, achieved the highest accuracy, reaching 87.8%.

3. Methods of Research

3.1 Datasets

The Cleveland dataset stands as a repository of 303 patient records, presenting a diverse demographic landscape that includes 207 males and 94 females across varying age groups. Each patient record has .14 characteristics, which serve as .the foundation for deeper analysis. Within this group, 138 people are .classified as having normal. cardiovascular conditions, w.hereas 165 are diagnosed with heart disease. The age range. of the patients is 29 to 77 years, providing a .thorough sample of adult age groups[14]. The dataset uses a binary. method to signify the presence or absence of cardiac illness, with 0 representing the absence of the condition and 1 representing the existence of a fixed abnormality. For experimentation and model assessment, the dataset has been divided into. training and testing subsets. Specifically, 80% of the data .is allotted to the training set, which allows machine learning models. to be developed and refined. The remaining 20% makes up the testing set, which serves as an impartial baseline for .evaluating the models' performance.

Table 1 in this context gives a comprehensive summary of all characteristics, encompassing information about each attribute for both datasets. This extensive collection. intends to promote a deeper understanding. of the dataset's complexities. and lay the groundwork foor thorough investigations into the domain .of cardiovascular health assessment.

Table 1. Attributes and details of two datasets related to heart disease

No.	Attributes	Details
1	Age	Age in years
2	Gender	Gender: 1 for male, 0 for female
3	*cp*	Chest pain type (4 values)
4	*trestbps*	Resting blood pressure in mm Hg on admission to the hospital
5	*chol*	Serum cholesterol in mg/dL
6	*fbs*	Fasting blood sugar: 1 for > 120 mg/dL (true), 0 for ≤ 120 mg/dL (false)
7	restecg	Resting electrocardiographic results
8	*thalach*	Maximum heart rate achieved
9	*exang*	Exercise-induced angina: 1 for yes, 0 for no
10	*oldpeak*	ST depression induced by exercise relative to rest
11	*slp*	Slope of the peak exercise ST segment
12	*ca*	Number of major vessels colored by fluoroscopy (ranging from 0 to 3)
13	*thal*	Thal: 1 for normal, 2 for fixed defect, 3 for reversible defect
14	*output*	Heart condition: 0 for no disease, 1 for disease

3.2 Data preprocessing

Significant efforts were made to optimize the dataset for use by ML algorithms in this work. The preprocessing stages are listed below:

A. Data Exploration and Visualization: Data exploration and visualization are critical in identifying the fundamental patterns and qualities in a collection. The first stage is to compute descriptive statistics, which include measurements such as mean, median, and standard deviation. Univariate analysis allows. for the examination of individual. variables using tools like histograms and box plots, which help in the discovery .of outliers and central trends. Bivariate analysis expands on. this research by examining relationships. between pairs of variables using scatter plots and. correlation matrices. Multivariate analysis increases insights by taking into account. numerous variables at once, frequently using dimensionality reduction techniques such as PCA. Furthermore, the exploration includes .analysing feature relevance, discovering trends, locating outliers, and resolving class imbalances. The use of interactive visualization. tools and dashboards improves the exploration. experience by enabling dynamic manipulation and a more complete .comprehension of the dataset. The combination of statistical and .graphical tools during data exploration creates. the foundation for informed decision-making, directing later. preprocessing and modelling efforts.[15].

B. Normalization of the target variableis an important preprocessing step, especially in .regression problems, that ensures a steady and successful learning process. Normalization is the process of changing target variable values to a. standardized scale in order to promote consistency and convergencre during model training. Min-Max scaling and Z-score. normalization [16] are two popular approaches for normalizing the target variable.

C. Feature extractionis a key machine learning method that entails .converting raw data into a set .of useful and informative features. The objective is to minimize the. dimensionality of the data while retaining critical. information for developing efficient models. Here's the Independent Component Analysis approach. utilized in feature extraction [17].

2.3 Independent Component Analysis (ICA)

Because of its capacity to reveal significant .patterns and sources inside intricate datasets, ICA stands out as a reliable statistical. approach that has garnered significant. interest from a variety of sectors. ICA has several advantages, such as feature extraction, signal separation, and noise reduction, by dissecting mixed signals into. statistically independent components. Using ICA becomes very useful and .beneficial when predicting cardiac disease. with a dataset. Through the examination of a dataset that includes a variety of physiological .signals and patient variables, this study seeks to determine the impact. of ICA on the prediction of heart disease. The goal is to apply ICA to improve the characteristics. of the dataset, uncover hidden correlations, and raise the precision of heart disease. prediction models. In the end, this results in better patient .outcomes and more informed healthcare decisions. Three machine learning classifiers—GB, SVM, and RF—were used in this work to train the models. and forecast the incidence of heart attack illness[18].

A. Gradient Boosting (GB) Trees are a type of machine learning approach that may .be used for both regression and classification problems. It creates a predictive model by combining several .weaker prediction models into one ensemble. The idea that boosting might be seen. as an optimization method for a reasonable cost function. led to the creation of GB[19]. It builds the model repeatedly, using a step-by-step methodology like to other boosting .techniques, and expands its range of applications by permitting the optimization of several .differentiable loss functions. Through this iterative process, decision trees are created one after the other, with each tree fixing mistakes made by the .ones before it, improving the model's. overall predictive performance. GB is a versatile tool in the field of machine learning since. it can handle many data formats and manage complicated interactions .within the dataset in an efficient manner[20].

B. Random Forestis an extensively used machine learning method .that is used to build prediction models. An ensemble of regression and classification. trees makes up RF. These trees provide relatively. simple models for result prediction since they exhibit binary splits on input variables. Because of its ease of use, RF is suitable for both non-statisticians and .computer specialists. Notably, RF successfully solves overfitting and does away with the requirement for cross-validation. The stages involved in building decision trees are .outlined by the RF algorithm. The method takes into account a dataset. consisting of N training cases and M characteristics during this procedure. To introduce a random element at each node. in a tree, a random subset of m attributes is separately. chosen from the total of M attributes. Then, training examples are sampled with replacement for. each tree in the forest, adding variety by letting some. instances be selected more .than once and others not at all. To put it briefly, the RF algorithm provides a strong machine. learning strategy that minimizes the drawbacks of individual decision .trees while utilizing their benefits. The model's performance is improved by adding. randomization to attribute selection and sampling in addition to .combining the predictions, making RF a preferred option for a variety of predictive tasks[21].

C.Support Vector Machineshows greater accuracy in comparisont to other machine learning techniques. In order to properly separate data pooints into two classes—those showing the existence of heart disease and those. showing its absence—SVM functions by creating a hyperplane. To ensure that the model is resilient to noise and outliers, the hyperplane is chosen with the goal of maximizing the distance. between data points from these two groups. Finding the hyperplane that best separates .the two classes depends on the support vectors, which show the data points closest to the hyperplane on either side[4]. A key component of SVM .is the dot product, a mathematical .procedure that gauges. vector similarity. The dot product in thiss. case measures. how close a data point. is to the hyperplane. The data points that are furthest from. the hyperplane are known as support vectors, and they are crucial in establishing a useful divisionbetween tthe classes. This is a flexible approach that may be used .for both regression and. classification applications. But one significant disadvantage of SVM is that it is .highly dependent on the kernel function and related parameters that are selected. Projecting data points .into a higher-dimensional space, the kernel function functions as a mathematical .transformation, and its properties are determined by. the kernel parameters. The SVM model's performance can be greatly impacted. by selecting the incorrect kernel function or by. incorrectly specifying its parameters, which could result. in less-than-.ideal outcomes like higher variance or decreased accuracy[22].

4. The Proposed Research

The suggestion highlights how .crucial early identification is to enhancing .heart attack patient outcomes. In order to improve the precision and effectiveness. of heart attack diagnosis, the suggested method. presents a unique strategy that combines machine learning .with Independent Component Analysis (ICA). Through the utilization of ICA's capacity to separate independent components that capture crucial sources of variation, the objective is to augment the machine learning model's discriminative capabilities, therefore elevating the overall detection accuracy.

A typical machine learning pipeline for tasks like anomaly detection or classification is shown in Figure 1. There are two main stages to the process: model training/evaluation and preprocessing. The input data from the database is subjected to data exploration, normalization, and feature extraction (using ICA) techniques during the preparation stage. Following pre-processing, the generated datasets are split into training (80%) and testing (20%) categories. Using the 80% training data, a variety of classifiers are learned during the model training phase, including Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB). The remaining 20% of the testing data is then used to evaluate the trained models. Data instances are categorized into two classes at this review stage: Normative or Anomaly, fulfilling functions such as anomaly identification or binary categorization.

The figure 1, which is particularly designed for supervised learning tasks, summarizes a typical machine learning pipeline and highlights the critical phases of data preparation, model training, and assessment.

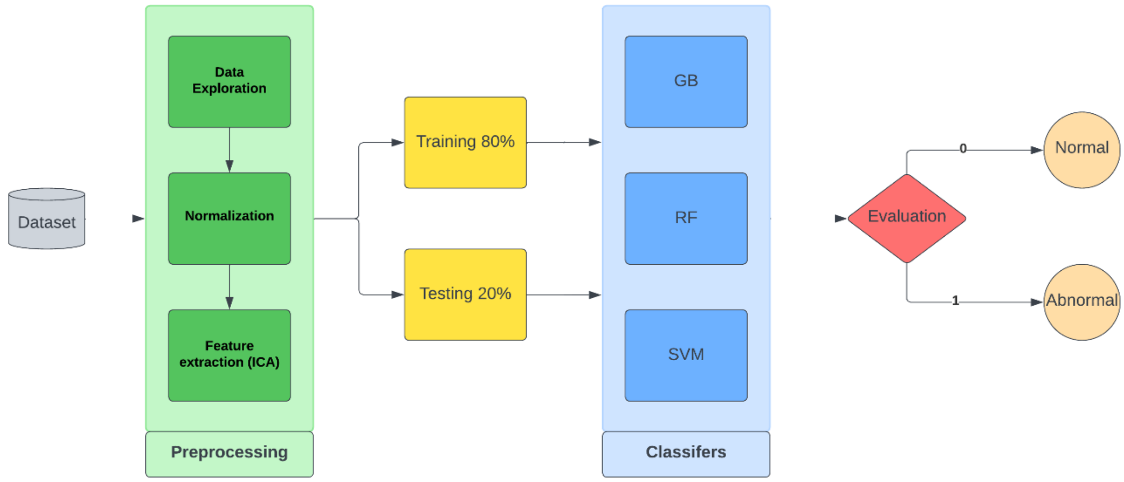

Figure 1. The block diagram of the proposed model

5. RESULTS AND DISCUSSION

5.1 Performance Evaluation Metrics

The following measures are included in the evaluation in order to determine how effective the suggested system is: accuracy, precision, recall, F1-score, and ROC curve. These metrics serve as numerical standards by which to compare the system's performance when it comes to anomaly detection.

- Accuracy: Taking into account both regular and anomalous occurrences throughout the dataset, this statistic assesses the general accuracy of the system's categorization. Equation 1 illustrates this point particularly well for datasets where the distribution of instances is balanced.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad Eq.(1)$$

- Precision: By comparing the total number of cases categorized as anomalies to the number of correctly diagnosed anomalies, precision evaluates the accuracy of real positive results. It provides information on how well the system can detect affirmative cases, as shown in Equation 2.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad Eq.(2)$$

- Recall: Also known as sensitivity, this metric calculates the proportion of accurately identified anomalies relative to the total count of actual anomalies. It proves particularly useful when the goal is to detect the maximum number of anomalies while minimizing false negatives, as presented in Equation 3.

$$\text{Recall} = \frac{TP}{TP + FN} \qquad Eq.(3)$$

- F1-score: This composite measure considers both precision and recall, calculating the harmonic mean of these two metrics. It provides a balanced evaluation of the system's performance, especially in scenarios where the class distribution is imbalanced, as outlined in Equation 4

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \qquad Eq.(4)$$

5.2 Experimental Results on Cleveland Database

Two sets of experiments were performed using the Cleveland database. In the initial experiment, three machine learning classifiers underwent training and testing to detect heart attacks. The subsequent experiment incorporated the Independent Component Analysis (ICA) technique with the machine learning classifiers. The objective of this integration was to isolate separate components that might identify hidden patterns and sources of variation linked to heart disease. Then, the independent components that were retrieved were used as input characteristics for the classifiers that used machine learning.
Table 2 displays When heart attack detection. in the Cleveland database is examined using .performance evaluation metrics in Table 2 without the use of Independent Component. Analysis (ICA), important information about the. efficacy of three distinct algorithms is revealed: Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB).
Different trends were found while .assessing the effectiveness of the Support .Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB) algorithms for heart attack identification in the Cleveland database. The SVM displayed the following results: 74.82% accuracy, 70.84% area. under the curve (AUC), 80.00% precision, 79.00% recall, 79.00% F1-score, and 9.45 log

loss. Proceeding to the RF, its performance. demonstrated a 75.40% accuracy, an 84.55% AUC, a 75.00% precision, a 74.00% recall, a 75.00% F1-score, and a log loss of 8.86. The results show that the GB algorithm .performed better than the others: 85.24% accuracy, 91.66% AUC, 85.00% precision, 86.00% recall, 85.00% F1-score, and 5.31 log loss. These results demonstrate the Gradient Boosting. algorithm's higher performance and show how useful. it is for detecting heart attacks based .on the given metrics when Independent Component .Analysis is not available.

Table 2. Performance Evaluation Metrics on Cleveland database without ICA

Algorithm	Accuracy %	AUC %	Precision %	Recall %	F1-score %	log loss %
SVM	74.82	70.84	80.00	79.00	79.00	9.45
RF	75.40	84.55	75.00	74.00	75.00	8.86
GB	85.24	91.66	85.00	86.00	85.00	5.31

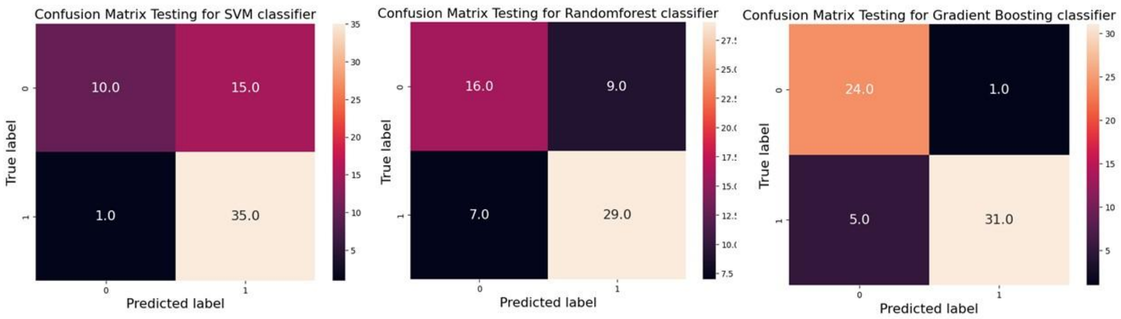

Figure 2. Confusion matrix on the Cleveland database without ICA

On the other hand, Figure 2 compares the performance of three distinct machine learning classifiers: Support Vector Machine (SVM), Random Forest, and Gradient Boosting. Each confusion matrix illustrates the count of instances correctly and incorrectly classified. To interpret a confusion matrix, consider the following breakdown

- True Positives (TP): Instances correctly predicted as the positive class.
- True Negatives (TN): Instances correctly predicted as the negative class.
- False Positives (FP): Instances incorrectly predicted as the positive class (Type I error).
- False Negatives (FN): Instances incorrectly predicted as the negative class (Type II error).

Analysing the confusion matrix allows for an assessment of how well each classifier executed on the dataset. Ideally, a desirable confusion matrix exhibits a concentration of values along the diagonal, indicating accurate predictions by the model.

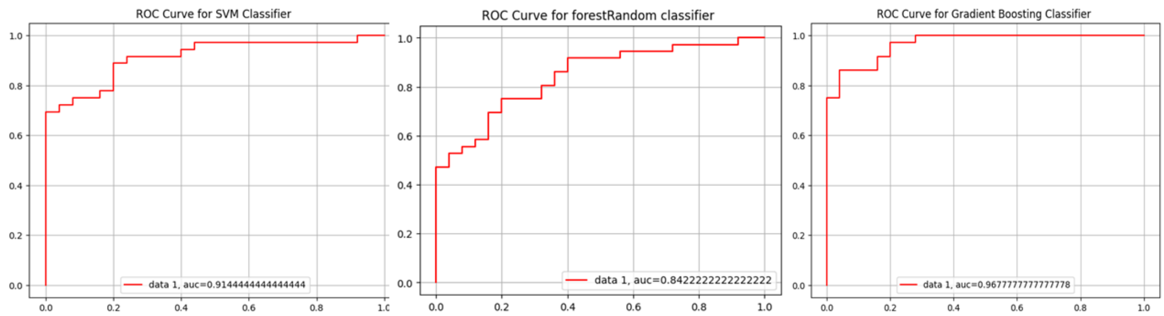

Figure 3. Evaluating ROC Curves on the Cleveland dataset without utilizing ICA

Table 3. Performance Evaluation Metrics on the Cleveland database with ICA

Algorithm	Accuracy %	AUC %	Precision %	Recall %	F1-score %	log loss %
RF	85.24	94.00	86.00	92.00	89.00	5.31
GB	88.52	94.55	89.00	87.00	88.00	4.13
SVM	90.16	96.66	90.02	90.00	90.00	3.54

In Table 3. The assessment metrics for the Cleveland dataset, employing the ICA algorithm, unveil distinct performance attributes across three machine learning classifiers. For Random Forest (RF), a precision of 85.24% is attained, coupled with an AUC of 94.00%, precision reaching 86.00%, recall at 92.00%, an F1-score of 89.00%, and a log loss of 5.31. Gradient

Boosting (GB) showcases an accuracy of 88.52%, AUC of 94.55%, precision of 89.00%, recall of 87.00%, an F1-score of 88.00%, and a log loss of 4.13. Notably, Support Vector Machine (SVM) distinguishes itself with an accuracy of 90.16%, AUC of 96.66%, precision at 90.02%, recall at 90.00%, an F1-score of 90.00%, and a log loss of 3.54. All of these measures provide a thorough understanding of how well the classifiers perform on the chosen dataset. They include accuracy, precision, recall, area under the curve, F1-score, and log loss.

There are significant differences in the algorithms' performance metrics when comparing the first results—obtained without using the ICA technique—with the latter results that use the ICA approach. When Support Vector Machine (SVM) was taken into account, using ICA produced notable improvements in every statistic. AUC increased from 70.84% to 96.66%, while accuracy increased significantly from 74.82% to 90.16%. Recall increased from 79.00% to 90.00%, precision increased from 80.00% to 90.02%, and improvements were seen in F1-score and log loss as well. In a similar vein, using ICA enhanced Random Forest's (RF) performance measures. AUC increased from 84.55% to 94.00%, while accuracy increased from 75.40% to 85.24%. Recall significantly improved from 74.00% to 92.00%, while both F1-score and log loss showed improvements. Precision increased from 75.00% to 86.00%. Moreover, the use of ICA showed significant improvements in Gradient Boosting (GB). AUC went from 91.66% to 94.55%, but accuracy improved from 85.24% to 88.52%. The precision increased to 89.00% from 85.00%, the recall increased to 87.00% from 86.00%, and the F1-score and log loss also exhibited gains.

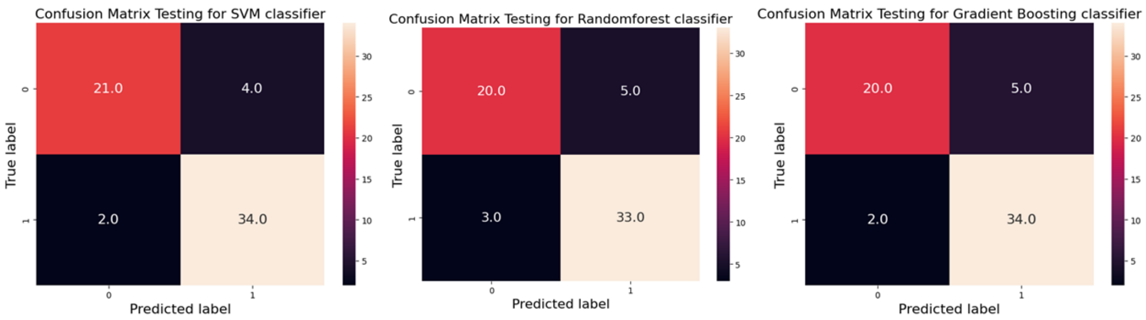

Figure 4. Confusion matrix on the Cleveland database with ICA

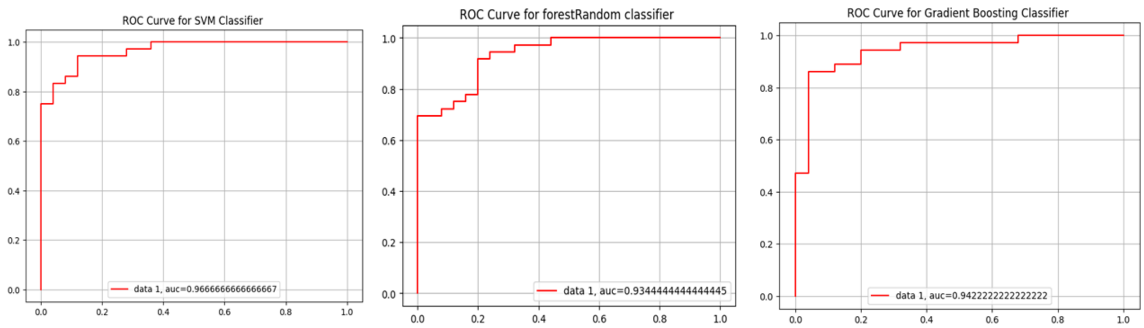

Figure 5. Evaluating ROC Curves on the Cleveland dataset with utilizing ICA

5.3 Comparative Studies with Recent Research

In this segment, we conducted a comparative analysis of the performance of our proposed model against recent research employing machine learning techniques for heart disease prediction. Table 4 illustrates a comparison of the evaluation metrics attained by our model with those of existing methods utilizing the Cleveland dataset. Based on the table provided, which showcases the final results of your research papers, the Support Vector Machine (SVM) algorithm emerges as the most favorable option among those listed. It boasts the highest accuracy, AUC (Area Under the Curve), precision, recall, F1-score, and the lowest log loss in comparison to Random Forest (RF) and Gradient Boosting (GB). Therefore, based on these performance metrics, SVM appears to be the superior algorithm.

Table 4. Evaluating the Accuracy of the Proposed Model against Other Methods

Ref.	Algorithms	Accuracy%
[23], 2019	Ensemble Classification	85.48
[24],2021	HRFLM	88.74
[25],2022	MLP-PSO hybrid	84.61
[13],2023	SC	87.30
Proposed Approach	SVM-ICA	90.16

6. CONCLUSION

In conclusion, our research has demonstrated the effectiveneess of our suggested model for machine learning-based heart disease prediction, especially when Independent Component Analysis (ICA) is incorporated into the Suppart Vector Machine (SVM) algorithm. Our model's amazing performance measures, which included excellent accuracy (90.16%), AUC (96.66%), precision (90.02%), recall (90.00%), F1-score (90.00%), and minimum log loss (3.54), were achieved by utilizing the Cleveland dataset and optimizing SVM with ICA. When ICA-SVM was combined with other techniques like Random Forest (RF) and Gradient Boosting (GB), the results were better. When our work is thoroughly compared with other current research endeavors, it demonstrates a considerable advancement in the prediction of heart disease. We have exceeded the performance parameters of previous approaches, demonstrating the efficacy and novelty of our strategy in improving the precision and reliability of heart disease prognosis. Additiionally, our model's proactive features, enabled by ICA-SVM, help to enhance public health outcomes by facilitating early diagnosis and mitigation of future heart disease cases. Subsequent research directions might encompass enhancing the ICA-SVM model, investigating supplementary datasets or characteristics, and converting our discoveries into useful clinical applications with the objective of reducing the impact of heart disease on heealthcare systems.

References

[1] S. A. Jebur, M. A. Mohammed, and A. K. Abdulhassan, "Covid-19 detection using medical images," in *AIP Conference Proceedings*, 2023, vol. 2591, no. 1. doi: https://doi.org/10.1063/5.0119758.

[2] R. Manne and S. C. Kantheti, "Application of Artificial Intelligence in Healthcare: Chances and Challenges," *Current Journal of Applied Science and Technology*, vol. 40, no. 6, pp. 78–89, 2021, doi: 10.9734/cjast/2021/v40i631320.

[3] M. M. Ahsan and Z. Siddique, "Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review," Dec. 2021, [Online]. Available: http://arxiv.org/abs/2112.06459

[4] R. Baxani and M. Edinburgh, "Heart Disease Prediction Using Machine Learning Algorithms Logistic Regression, Support Vector Machine and Random Forest Classification Techniques," *Support Vector Machine and Random Forest Classification Techniques (July 1, 2022)*, 2022, [Online]. Available: https:/papers.ssrn.com/sol3/papers.cfm?abstract_id=4151423s. Int J Adv Sci Technol. 2020;29(03):12273-12282.

[5] M. O. Adebiyi, A. A. Adebiyi, O. Okesola, and M. O. Arowolo, "ICA Learning Approach for Predicting RNA-Seq Data Using KNN and Decision Tree Classifiers," *International Journal of Advanced Science and Technology*, vol. 29, no. 03, pp. 12273–12282, 2020.

[6] S. M.R.Taha and A. K. Nawar, "A New Quantum Radial Wavelet Neural Network Model Applied to Analysis and Classification of EEG Signals," *International Journal of Computer Applications*, vol. 85, no. 7, pp. 7–11, 2014, doi: 10.5120/14851-3216.

[7] D. Giri *et al.*, "Automated diagnosis of Coronary Artery Disease affected patients using LDA, PCA, ICA and Discrete Wavelet Transform," *Knowledge-Based Systems*, vol. 37, pp. 274–282, 2013, doi: 10.1016/j.knosys.2012.08.011.

[8] A. I. B. Shimaa Ouf, "A proposed paradigm for intelligent heart disease prediction system using data mining techniques," *Journal of Southwest Jiaotong University*, vol. 56, no. 4, 2021.

[9] 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.

[10] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informatics in Medicine Unlocked*, vol. 16, p. 100203, Jan. 2019, doi: 10.1016/J.IMU.2019.100203.

[11] V. Chang, V. R. Bhavani, A. Q. Xu, and M. A. Hossain, "An artificial intelligence model for heart disease detection using machine learning algorithms," *Healthcare Analytics*, vol. 2, p. 100016, 2022.

[12] A. Al Bataineh and S. Manacek, "MLP-PSO hybrid algorithm for heart disease prediction," *Journal of Personalized Medicine*, vol. 12, no. 8, p. 1208, 2022.

[13] H. C. De Albuquerque, F. Palumbo, P. Barsocchi, A. K. Bhoi, E. Dritsas, and M. Trigka, "Efficient Data-Driven Machine Learning Models for Cardiovascular Diseases Risk Prediction," *Sensors 2023, Vol. 23, Page 1161*, vol. 23, no. 3, p. 1161, Jan. 2023, doi: 10.3390/S23031161.

[14] "Heart Disease - UCI Machine Learning Repository." https://archive.ics.uci.edu/dataset/45/heart+disease (accessed Sep. 25, 2023).

[15] X. Qin, Y. Luo, N. Tang, and G. Li, "Making data visualization more efficient and effective: a survey," *The VLDB Journal*, vol. 29, no. 1, pp. 93–117, 2020.

[16] W. Ma, T. Zhou, J. Qin, X. Xiang, Y. Tan, and Z. Cai, "Adaptive multi-feature fusion via cross-entropy normalization for effective image retrieval," *Information Processing & Management*, vol. 60, no. 1, p.

103119, 2023.

[17] A. N. Khalifa, H. R. Ali, S. A. Jebur, and S. A. Jahefer, "Automate facial paralysis detection using vgg architectures," *International Journal of Current Innovations in Advanced Research*, pp. 1–8, 2024.

[18] R. Shaaque, A. Mehmood, G. S. Choi, R. Shafique, and S. Ullah, "Cardiovascular Disease Prediction System Using Extra Trees Classiier Cardiovascular Disease Prediction System Using Extra Trees Classifier," 2019, [Online]. Available: https://europepmc.org/article/ppr/ppr135668%0Ahttps://doi.org/10.21203/rs.2.14454/v1

[19] M. S. Al-Batah, M. Alzyoud, R. Alazaidah, M. Toubat, H. Alzoubi, and A. Olaiyat, "Early Prediction of Cervical Cancer Using Machine Learning Techniques," *Jordanian Journal of Computers and Information Technology*, vol. 8, no. 4, 2022.

[20] R. Aggarwal, P. Podder, and A. Khamparia, "ECG Classification and Analysis for Heart Disease Prediction Using XAI-Driven Machine Learning Algorithms," in *Intelligent Systems Reference Library*, vol. 222, Springer Science and Business Media Deutschland GmbH, 2022, pp. 91–103. doi: 10.1007/978-981-19-1476-8_7.

[21] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, no. July, p. 104672, 2021, doi: 10.1016/j.compbiomed.2021.104672.

[22] S. A. Jebur, K. A. Hussein, H. K. Hoomod, and L. Alzubaidi, "Novel deep feature fusion framework for multi-scenario violence detection," *Computers*, vol. 12, no. 9, p. 175, 2023.