IUST
Iran University of Science and Technology

Iranian Journal of Electrical and Electronic Engineering

Journal Homepage: ijeee.iust.ac.ir

Comparative Analysis of Machine Learning Techniques for Arrhythmia Detection

Biswapriyo Sen**, Maharishi Kashyap**, Jitendra Singh Tamang**, Sital Sharma***, Rijhi Dey*(C.A.)

Abstract: Cardiovascular arrhythmia is indeed one of the most prevalent cardiac issues globally. In this paper, the primary objective was to develop and evaluate an automated classification system. This system utilizes a comprehensive database of electro-cardiogram (ECG) data, with a particular focus on improving the detection of minority arrhythmia classes. In this study, the focus was on investigating the performance of three different supervised machine learning models in the context of arrhythmia detection. These models included Support Vector Machine (SVM), Logistic Regression (LR) and Random Forest (RF). An analysis was conducted using real inter-patient electrocardiogram (ECG) records, which is a more realistic scenario in a clinical environment where ECG data comes from various patients. The study evaluated the models' performances based on four important metrics: accuracy, precision, recall, and f1-score. After thorough experimentation, the results highlighted that the Random Forest (RF) classifier outperformed the other methods in all of the metrics used in the experiments. This classifier achieved an impressive accuracy of 0.94, indicating its effectiveness in accurately detecting arrhythmia in diverse ECG signals collected from different patients.

Keywords: Arrhythmia, Electrocardiography (ECG), Machine Learning, Support Vector Machine (SVM), Logistic Regression (LG), Random Forest (RF).

1 Introduction

1.1 History

THE heart, often considered one of the body's most vital organs, is responsible for sustaining life through its rhythmic contractions [1]. These contractions generate electrical activities that can be detected and recorded on the body's surface using an electrocardiogram (ECG) [2]. The ECG provides a valuable tool for monitoring and assessing the heart's electrical behavior, offering critical insights into cardiac health and functioning. Skin electrodes play a critical role in recording the electrical activity of the heart [3].

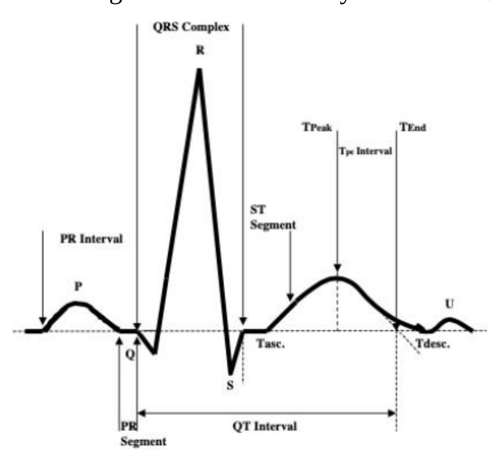

Fig. 1 Various points and intervals of a heartbeat [5]

Iranian Journal of Electrical & Electronic Engineering, 2024.
Paper first received 11 March 2024 and accepted 10 June 2024.
* The author is with the Department of Electronics and Communication Engineering, IEM New Town Campus, University of Engineering and Management, Kolkata, India.
E-mail: rijhi.dey88@gmail.com.
** The authors are with the Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim, India.
*** The author is with the Department of Artificial Intelligence and Data Science, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim, India.
Corresponding Author: Rijhi Dey.

Sl.No.	Algorithm	Findings	Accuracy	Reference
1	SVM + RF	A typical hybrid the use of machine learning model with low computational efficiency and mediocre accuracy.	77.4%	[11]
2	SVM+GB	Average accuracy on a redundant dataset with just 500 entries and 16 classes were shown.	84.82%	[13]
3	Random Forest BFS	Having just 500 entries and 16 classes, a redundant dataset was employed.	85.58%	[14]
4	OPF	An effective classifier that works well when combined with other machine learning algorithms, but which needs a significant amount of data preparation to achieve the best results.	90.6%	[12]
5	DNN	A genetic algorithm with significant computing cost paired with a DNN	94%	[15]
6	RF	Presents a comparative analysis of five popular supervised machine learning algorithms based on RR interval based feature set.	89.9%	[27]
7	Proposed Deep Learning Approach	An algorithm that uses the auto-encoder and clustering approach Which helps the training to improve faster and with less effort	92%	[28]
8	**RF**	**For detecting Arrhythmia for data set comprises of 10,646 patients and a comparison of LR and SVM method has been carried out.**	**94%**	**Present Work**

They reveal the functioning of each cardiac chamber by capturing the characteristic PQRST waves, as depicted in Figure 1. These waves represent the electrical events associated with the heart's various phases, providing valuable insights into its performance and any potential arrhythmias or irregularities [4]. Arrhythmia, derived from the Greek words "a-" meaning without and "rhythmos" meaning rhythm, refers to abnormal heart rhythms that deviate from the normal sinus rhythm [6]. It is a prevalent cardiovascular disorder affecting millions of individuals worldwide. Over the course of history, the study of arrhythmia has witnessed significant developments, leading to advancements in diagnostic techniques and treatment options. Arrhythmia, a common cardiovascular disorder, refers to ab- normal heart rhythms that deviate from the normal sinus rhythm [7]. Detecting arrhythmia accurately and promptly is vital for effective diagnosis and management of this condition [8]. Over the years, there have been remarkable advancements in arrhythmia detection techniques, resulting in improved patient outcomes and the development of personalized treatment strategies [9].

In this section, various methods and technologies employed in the detection of arrhythmia will be delved deeper in order to provide a comprehensive understanding of their strengths, limitations and potential for further advancement. Early attempts at detecting arrhythmia relied heavily on clinical examination and patient symptoms. However, the advent of electrocardiography (ECG) in the early 20th century, discovered by Willem Einthoven, revolutionized the field by providing a non-invasive method for recording the electrical activity of the heart [10]. ECG allowed researchers to observe and analyze the intricate patterns and irregularities associated with arrhythmia, thus enabling a more objective assessment of cardiac health.

As technology progressed, researchers explored various methods to interpret ECG signals accurately. This involved the development of algorithms and mathematical models to analyze the complex waveforms present in ECG recordings. The identification arrhythmia types became possible, facilitating targeted treatment strategies and improving patient outcomes.

1.2 Approach

Electrocardiography (ECG) has played a pivotal role in revolutionizing arrhythmia detection [16]. By recording the electrical activity of the heart, ECG enables researchers and healthcare professionals to observe and analyze intricate pat- terns and irregularities associated with arrhythmia. Traditional methods involved visually analyzing ECG waveforms by skilled clinicians. While these methods provided valuable insights, they were subjective, time- consuming, and prone to human error. Consequently, to improve the precision and effectiveness of arrhythmia identification, researchers resorted to computer-aided techniques.

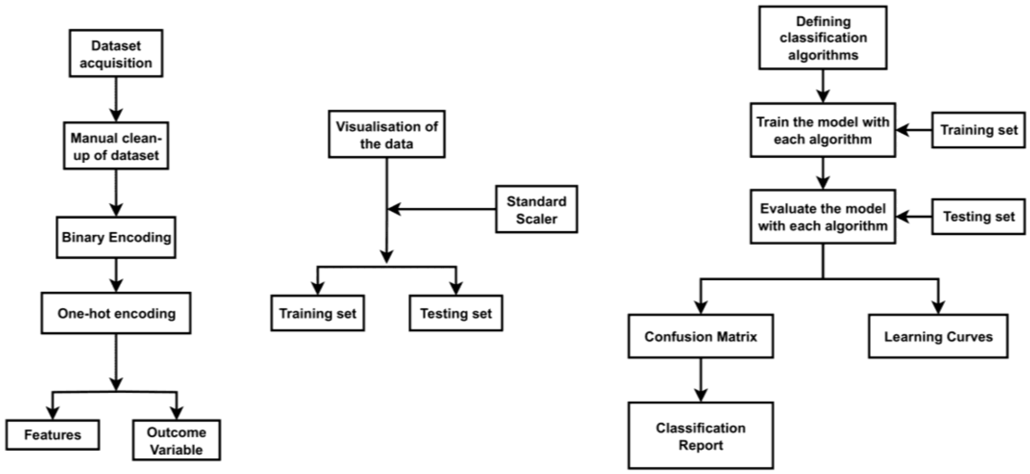

Fig. 2 Work flow diagram for the complete process

Computer-aided arrhythmia detection systems have been developed to automate the detection process and improve diagnostic accuracy [17-18]. These systems employ algorithms and mathematical models to analyze ECG signals, automatically identifying and classifying different types of arrhythmias. Signal processing techniques, such as filtering, feature extraction, and rhythm analysis, are incorporated into these systems to enhance their performance. By reducing reliance on manual analysis, computer- aided approaches enable faster and more accurate detection of arrhythmia.

Despite significant advancements, the field of arrhythmia detection continues to grapple with several challenges. Primarily, the availability of comprehensive, diverse datasets for the purpose of training and evaluating detection algorithms remain a crucial aspect for improving their performance and ensuring their generalizability across various patient populations and scenarios. The development of robust feature extraction techniques is imperative to effectively capture pertinent information from electrocardiogram (ECG) signals, thereby facilitating accurate and reliable arrhythmia detection [19]. Furthermore, the implementation of real-time monitoring solutions is imperative to enable continuous and timely detection of arrhythmias, thus allowing for prompt medical interventions and preventive measures [20].

Continued research and development efforts are vital to addressing these persisting challenges and further advancing the field of arrhythmia detection. Nonetheless, the accurate and timely identification of arrhythmias holds paramount importance in the effective management of this cardiovascular disorder. Notably, the intersection of technological advancements and cutting-edge medical research has significantly contributed to refining arrhythmia detection techniques over time. The evolution from traditional detection methods to sophisticated computer-aided approaches and the integration of machine learning and artificial intelligence (AI) have collectively enhanced the precision and efficiency of arrhythmia detection procedures. Nevertheless, ongoing research endeavors are imperative to surmount the remaining obstacles and propel the field forward, ultimately leading to improved patient outcomes and the development of tailored, personalized treatment strategies. A comparison table has been depicted in Table I to showcase the different techniques used by the researchers to detect arrhythmia and their accuracy of work.

1.3 Contribution

In this comprehensive research endeavor, a meticulous comparative analysis of three prominent machine learning algorithms—Support Vector Machine (SVM), Logistic Regression (LR) and Random Forest (RF) - assessing their capabilities in accurately identifying various forms of arrhythmia. The investigation was grounded in a rich and diverse dataset, meticulously curated to encapsulate a wide spectrum of cardiac rhythm data points, ensuring a robust and nuanced evaluation process.

Leveraging the powerful insights provided by confusion matrices, along with conducting a granular examination of the algorithms' detection prowess, gaining valuable insights into their abilities to correctly classify true positive and true negative cases, as well as to identify false positives and false negatives. Notably, the findings unequivocally highlighted the exceptional performance of the Random Forest algorithm, showcasing a remarkable balance between the rate of accurately identifying true positive cases and the minimization of false positive cases, thus surpassing the other algorithms in our analysis. The Random Forest algorithm's inherent capacity to adeptly handle complex, multi-dimensional data, along with its remarkable capability to discern intricate relationships within the dataset, unequivocally underscored its effectiveness in precisely identifying intricate patterns of arrhythmia. This pivotal study not only reiterates the indispensable role of machine learning in honing the accuracy of arrhythmia detection but also accentuates the immense potential of the Random Forest algorithm in advancing the diagnostic frontiers within the domain of cardiology.

1.4 Organization

This paper is organized as materials and methodology as Section 2. The output graphs and simulation comparison studies are shown in Section 3. Finally, the concluding remarks and future scope are being made in Section 4.

2 Materials and Methodology

2.1 Overview

This study is based on using machine learning techniques for predicting whether an individual will have an arrhythmia or not. The first step in the project is data pre-processing, which entails cleaning the information and getting it ready for machine learning by transforming categorical variables like gender and beat type into a format that is appropriate for algorithms. Arrhythmia is represented by the target variable, which is defined as 1 for positive cases and 0 for negative cases. Notably, the dataset is expected to have useful attributes for the prediction task, therefore the method does not entail explicit feature extraction. To further get insight into the dataset and comprehend feature interactions and data distributions, the algorithm makes use of data visualization techniques as correlation heat-maps, pair plots, and histograms. These visualizations are essential for assisting with feature selection, comprehending data trends, and ultimately aiding with classifier evaluation.

The evaluation of various machine learning classifiers forms the project's core and these include Random Forest, Logistic Regression and Support Vector Machine. Each classifier is evaluated for its propensity to correctly identify arrhythmia after being trained on a subset of the data. For each model, classification reports and confusion matrices are generated, offering thorough insights into how well they performed.

In conclusion, the project's technique incorporates data pre- processing, a comparison of machine learning classifiers, and data visualization to create, assess, and fine-tune prediction models for arrhythmia identification. The project's objectives are to determine the best classifier for this particular medical application as well as to offer insights into the features of the dataset and their role in the prediction process. The complete steps for the execution of the workflow has been depicted in Figure 2.

2.2 Dataset Used

The data set comprises of 10,646 patient ECGs, including 5,956 males and 4,690 females [21]. 17% of these individuals had a regular sinus rhythm, whereas 83% had one or more abnormalities. The largest incidence was seen in the age categories 51–60, 61–70, and 71–80, which represented 19.82%, 24.38%, and 16.9%, respectively. Table 2 provides a thorough breakdown of the baseline characteristics and rhythm frequency distribution of the recruited subjects. 4.88 volts are required for each A/D bit, and the resolution of the A/D converter was 32 bits. The microvolt was the amplitude unit. The top and lower limits were 32,767 and 32,768 respectively. Data Gathering: Data was gathered in four steps. Each individual first had a 12-lead resting ECG test performed over the course of 10 seconds. The GE MUSE ECG system was used to store the data. Each subject's rhythm, as well as additional circumstances like PVC, a right bundle branch block (RBBB), a left bundle branch block (LBBB), and an atrial premature beat (APB), are labeled. Instead of only being applied to certain beats in the 10-second readout, these extra criteria were applied to the full sample. The MUSE ECG system also contained the final diagnosis. Third, ECG information and diagnostic data were exported from the GE MUSE system to XML files that were encoded with a specific name conversion specified by General Electric (GE). Finally, we created a conversion program to extract diagnostic data and ECG data from the XML file and convert them to CSV format. It is cited from Maarten J.B. van Ettinger's efforts in doing so.
(https://sourceforge.net/projects/ecgtoolkit-cs/).

2.3 Pre-processing

Data pre-processing is an important step in getting the dataset ready for machine learning modelling to determine if an individual will have an arrhythmia or not. Feature transformation and data cleansing are the main pre-processing stages. In order to make the "Gender" column appropriate for machine learning algorithms, the gender labels are first converted into binary values, where "MALE" is represented as 1

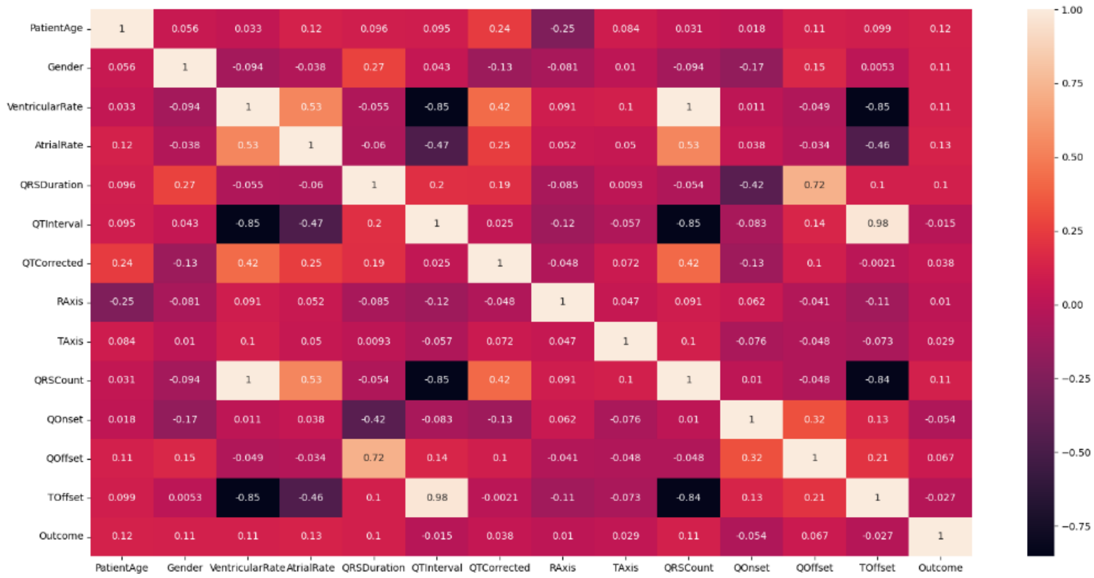

Fig. 3 Correlation Heat Map of the dataset

and “FEMALE” as 0. Next, the “Beat” column, which was initially categorical, is subjected to one-hot encoding. With the help of this transformation, the model can successfully handle categorical data by creating binary columns for each distinct category. This step is crucial to ensuring that the classifier can understand these categories as it learns.

The target feature, which denotes whether arrhythmia exists or not, is set to 1 in the instance of arrhythmia and 0 in its absence. The goal variable for the prediction model is this binary result. Overall, these pre-processing processes get the data ready for the classification task and make sure it’s in a way that allows for comparisons of the three machine learning algorithms’ accuracy in predicting arrhythmia.

2.4 Feature Used

There was a feature file provided along with the dataset [21] having total 13 features. The features were Patient’s Age, Gender, Ventricular Rate, Atrial Rate, QRS Duration, QT Interval, QT Corrected, R Axis, T Axis, QRS Count, Q Onset, Q Offset and T Offset. All the 13 features had been used further for data visualization and model prediction.

2.5 Data Visualization

Instead of explicitly using feature extraction, the project concentrates on data pre-processing, including data cleaning, transformation, and encoding, as well as modeling and evaluating various classifiers. For this classification assignment, it is expected that the dataset has previously undergone preprocessing with the necessary features. However, the code uses a number of data visualization techniques, including the correlation heat-map, pair-plot, and histograms to obtain insight into the information and aid in the model evaluation process.

a. **Correlation Heat-map:** Understanding the connections between the various elements in the dataset can be done with the help of the correlation heat-map [22]. It aids in determining how characteristics connect to one another and to the project’s goal variable (the presence or absence of arrhythmia). Strong correlations can show which features may be more useful for making predictions. The feature selection or additional analysis can be guided by this display. The correlation heat-map for the dataset used is illustrated in Figure 3. It illustrates the relationships between different features in the dataset, aiding in understanding how these features are interconnected and their relationship to the presence or absence of arrhythmia. Strong correlations highlighted in the heat map indicate which features are more influential in making predictions, guiding feature selection and further analysis. The figure is not an exact copy from the references but is generated based on the dataset used in the study. Prominent points in the heat map could include high correlation values (close to 1) indicating strong positive relationships, while values close to -1 suggest strong negative correlations. Understanding these correlations can help in feature selection and model evaluation for arrhythmia detection.

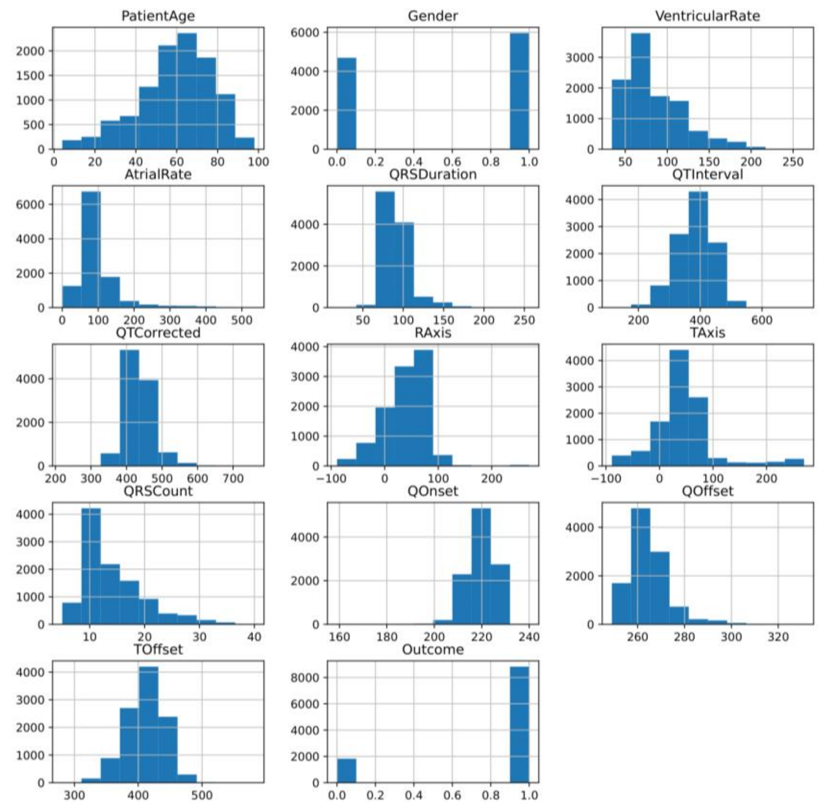

Fig. 4 Histogram of the dataset

b. Histograms: The histograms individually display how each feature is distributed in the dataset [23]. This is essential to comprehend the skewness, core tendencies and distribution of the data. Analysing the feature distributions in the context of arrhythmia prediction can offer information about the frequency of particular values or ranges in the data. Setting appropriate model parameters or comprehending the statistical features of the data can both benefit from this knowledge. Figure 4 is showing the histogram for the dataset used.

Although these visualizations don't immediately aid in feature extraction, they are essential for exploring and comprehending the data. As a result, it is simpler to understand the model results and improve the classification algorithms for arrhythmia prediction when they are used to assist uncover data patterns and linkages.

2.6 Machine Learning Models

1) Support Vector Machines (SVM): SVMs, or support vector machines, are strong models for applications involving regression and classification [24]. Support vectors are used to anchor the optimal hyperplane that maximizes the margin between distinct classes of data points. The decision function may be expressed mathematically as equation 1, for data that can be separated linearly.

$$f(x) = sign(w^T x + b) \tag{1}$$

SVM seeks to maximize the margin, expressed as an optimisation problem, min by equation 2, while minimizing the classification error for each and every data point, subject to the restrictions in equation 4.

$$r_i = \frac{|w^T x_i + b|}{|w^T|} \tag{2}$$

$$\min \frac{1}{2} |w|^2 \tag{3}$$

$$y_i(w^T x_i + b) \geq 1 \tag{4}$$

SVM employs the kernel approach to convert the input space into a higher-dimensional one for linear separability while dealing with nonlinear data. Polynomial and RBF kernels are popular options. SVMs provide a strong foundation for classification by utilizing the concept of margin maximization for efficient decision boundary determination.

2) Logistic Regression (LR): Logistic Regression is a fundamental statistical technique for binary classification, estimating the likelihood of an input belonging to a particular category [25]. Despite its name, it primarily serves categorization tasks. It establishes the relationship between one or more independent variables and the categorical dependent variable by fitting a logistic function to the observed data. The model identifies the optimal parameters that minimize the error between the predicted probability and the actual outcome using optimization techniques. It employs the maximum likelihood estimation approach to determine the best parameters.

3) Random Forest (RF): An ensemble learning method called Random Forest creates a large number of decision trees during training and determines a class by either finding the class mode (classification) or averaging the predictions (regression). It increases robustness and efficiency by assembling the outputs of a forest of decision trees [26].

Using random feature subsets at each node, the approach chooses the best splits and trains each tree on a random subset of the data. It aggregates the results from each individual tree during prediction. In terms of mathematics, it entails the creation of numerous decision trees, with the average (regression) or majority vote (classification) serving as the final forecast. Generalization is improved and overfitting is decreased when features and data sample selection are randomized. Given its ability to handle large datasets, high dimensionality, resilience against overfitting, and difficulty with classification problems, Random Forest is a popular machine learning approach.

2.7 Model Evaluation

In this study, classification reports and confusion matrices—two crucial tools—were used to assess the models. These evaluation measures offer insightful information about the effectiveness and precision of the machine learning classifiers.

a. Classification Reports: A classification report provides a thorough analysis of a model's performance in this area. Usually, it contains a variety of measures, including precision, recall, F1-score, and support. While recall (sensitivity) analyzes the proportion of true positive predictions among all real positive cases, precision measures the proportion of true positive forecasts across all positive predictions. The F1-score, which offers a fair evaluation of a model's overall performance, is the harmonic mean of precision and recall. The number of instances of each class in the true labels is counted as the support metric. When working with imbalanced datasets, classification reports are essential since they give a more thorough knowledge of a model's advantages and disadvantages. They aid in recognizing false positives and false negatives, which can be crucial in the field of medicine.

b. Confusion Matrices: A model's effectiveness in a binary classification challenge is tabulated and shown as a confusion matrix. It categorizes predictions into four groups: true positives (positive cases accurately predicted), true negatives (negative cases correctly predicted), false positives (positive cases mistakenly predicted), and false negatives (negative cases incorrectly predicted). Confusion matrices are crucial because they enable a detailed analysis of how a model behaves when differentiating between the two groups. They can assist in identifying potential error sources and if a model is more likely to produce false positives or false negatives. Knowing these faults is essential in a medical context like arrhythmia detection because it allows one to evaluate the model's dependability and safety in practical applications.

For a thorough evaluation of the model's predictive power, it is crucial to incorporate both classification reports and confusion matrices. Particularly when working with sensitive medical data where false positives or false negatives might have serious repercussions, they provide insights beyond a mere accuracy score and aid in finding areas for development and fine-tuning. These evaluation criteria are useful tools for making sure the model works and is reliable in practical applications.

2.8 Performance Evaluation Measures

When evaluating classification results, several performance metrics are commonly used to assess the performance of a classification model. In the literature, four of the most frequently employed metrics include:

1) **Accuracy:** It calculates the ratio of correctly predicted instances to the total number of instances in the dataset. Accuracy provides a general measure of the model's effectiveness.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

2) **Precision:** Precision, also known as positive predictive value, assesses the model's ability to make accurate positive predictions. It calculates the ratio of true positive predictions to the total positive predictions made by the model.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

3) **Recall (Sensitivity):** It quantifies the model's capability to identify all positive instances correctly. It calculates the ratio of true positive

predictions to the total actual positive instances in the dataset. Recall is important when it's crucial to detect all positive cases.

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

4) **F1-Score:** The F1-Score is the harmonic mean of precision and recall. It provides a balanced measure that considers both false positives and false negatives. The F1- Score is particularly useful when there is an imbalance between the classes in the dataset.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

These metrics are essential for evaluating the performance of a classification model and provide a comprehensive view of its accuracy, precision, recall, and overall effectiveness in making predictions. The choice of which metric to emphasize often depends on the specific goals and requirements of the classification task.

3 Result

Learning curves were used for the model evaluation to examine the effectiveness of the three classifiers: Random Forest, Support Vector Machine and Logistic Regression. They offer a graphic depiction of the model's performance during training and cross-validation in relation to the size of the training set. The Random Forest learning curve shows that both training and cross-validation scores stable at a high level as the training set size increases, demonstrating the model's potent predictive power. Comparably, the learning curve for Logistic Regression shows rapid convergence, suggesting that this classifier functions effectively even with little data. The SVM learning curve, on the other hand, converges more slowly, indicating the need for additional data or model adjustment. Along with the learning curves, a table was also created that summarizes the recall, accuracy, precision, and F1-score metrics for Random Forest, SVM, and Logistic Regression to supplement our findings which is depicted in Table II. An extensive overview of each classifier's performance on the specified dataset is provided by this table. Understanding the model's predictive capacity, accuracy in categorizing arrhythmia, and capacity to distinguish between real positives and true negatives depends on these metrics. True positives and false negatives with arrhythmia are represented by "1" s, and false positives and false negatives with arrhythmia are represented by "0" s.

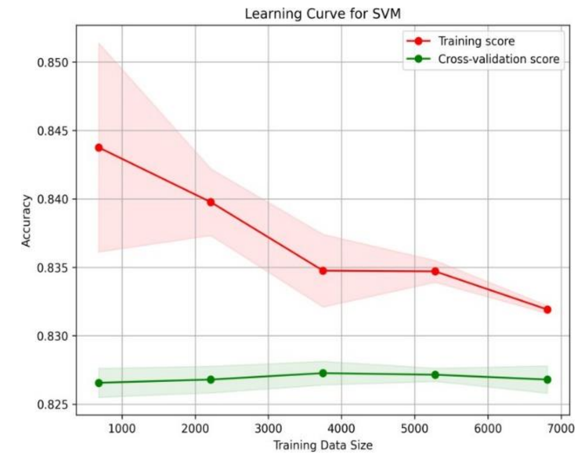

Fig. 5 Learning Curve for SVM

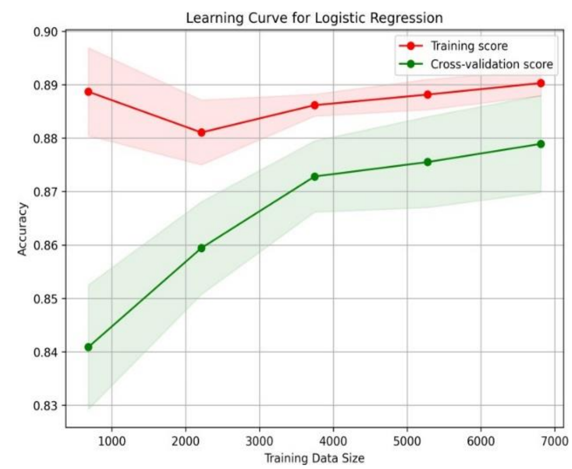

Fig. 6 Learning Curve for LR

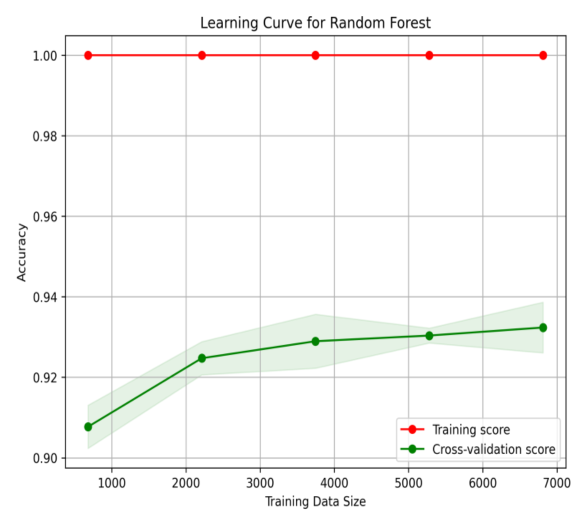

Fig. 7 Learning Curve for RF

Table. 1 Prediction's performance comparison

Sl. No.	Classifier	Accuracy	F1-Score		Precision		Recall	
			0	1	0	1	0	1
1.	SVM	0.83	0.01	0.91	0.29	0.83	0.01	1
2.	LR	0.88	0.60	0.93	0.67	0.91	0.55	0.95
3.	**RF**	**0.94**	**0.82**	**0.96**	**0.82**	**0.96**	**0.83**	**0.96**

These metrics are crucial for the comparison study of the model and hence, collectively help us to make informed decisions about the model's performance. SVM and LR gave a close accuracy of 83% and 88% respectively while RF had a better and more dependable accuracy of 94%. And for a better visualization of the performance of the models, a proper illustration of accuracy is illustrated in Figure 5, 6 and 7.

4 Conclusion

In this research paper, the focus was on developing an automatic classification system that utilizes a comprehensive ECG database. The goal was to enhance the detection of arrhythmic classes without the need for feature extraction. We evaluated the performance of three supervised machine learning models: Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF).

To assess the effectiveness of these models, the key performance metrics was measured, including accuracy, precision, recall, and the F1-score. The simulation results clearly indicated that the Random Forest (RF) model outperformed the other methods, demonstrating its significance as a predictive tool for detecting ECG arrhythmias. This outcome highlights the potential of RF as a valuable tool in the field of ECG arrhythmia prediction.

For future scope, more features can be extracted and hybrid machine learning techniques can be implemented to enhance the performance indices.

Acknowledgement

The authors are thankful to Sikkim Manipal University for providing the research facilities.

Conflict of Interest

The author has no conflicts of interest to disclose.

Funding

The author has no funding sources to disclose

References

[1] S. W. Patterson, H. Piper, and E. Starling, "The regulation of the heart beat," The Journal of physiology, vol. 48, no. 6, p. 465, 1914.

[2] D. B. Geselowitz, "On the theory of the electrocardiogram," Proceedings of the IEEE, vol. 77, no. 6, pp. 857–876, 1989.

[3] H. Wu, G. Yang, K. Zhu, S. Liu, W. Guo, Z. Jiang, and Z. Li, "Materials, devices, and systems of on-skin electrodes for electrophysiological monitoring and human–machine interfaces," Advanced Science, vol. 8, no. 2, p. 2001938, 2021.

[4] S. K. Saini and R. Gupta, "Artificial intelligence methods for analysis of electrocardiogram signals for cardiac abnormalities: state-of-the-art and future challenges," Artificial Intelligence Review, vol. 55, no. 2, pp. 1519–1565, 2022.

[5] C. Pater, "Methodological considerations in the design of trials for safety assessment of new drugs and chemical entities," Current controlled trials in cardiovascular medicine, vol. 6, no. 1, pp. 1–13, 2005.

[6] R. L. Royster, "Arrhythmia or dysrhythmia: Let's standardize our nomenclature." Anesthesia & Analgesia, vol. 70, no. 1, pp. 125–126, 1990.

[7] A. C¸ ınar and S. A. Tuncer, "Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ecg signals using lstm and hybrid cnn-svm deep neural networks," Computer methods in biomechanics and biomedical engineering, vol. 24, no. 2, pp. 203–214, 2021.

[8] O¨ . Yıldırım, P. Pławiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ecg signals," Computers in biology and medicine, vol. 102, pp. 411–420, 2018.

[9] V. D. Nagarajan, S.-L. Lee, J.-L. Robertus, C. A. Nienaber, N. A. Trayanova, and S. Ernst, "Artificial intelligence in the diagnosis and mmmm pp. 3904–3916, 2021.

[10] S. S. Barold, "Willem einthoven and the birth of clinical electrocardio- graphy a hundred years ago," Cardiac electrophysiology review, vol. 7, pp. 99–104, 2003.

[11] M. Sraitih, Y. Jabrane, and A. Hajjam El Hassani, "An automated system for ecg arrhythmia detection using machine learning techniques," Journal of Clinical Medicine, vol. 10, no. 22, p. 5450, 2021.

[12] E. J. d. S. Luz, T. M. Nunes, V. H. C. De Albuquerque, J. P. Papa, and D. Menotti, "Ecg arrhythmia classification based on optimum-path forest," Expert Systems with Applications, vol. 40, no. 9, pp. 3561–3573, 2013.

[13] S. Irfan, N. Anjum, T. Althobaiti, A. A. Alotaibi, A. B. Siddiqui, and N. Ramzan, "Heartbeat classification and arrhythmia detection using a multi-model deep-learning technique," Sensors, vol. 22, no. 15, p. 5606, 2022.

[14] N. Singh and P. Singh, "Cardiac arrhythmia classification using machine learning techniques," in Engineering Vibration, Communication and Information Processing: ICoEVCI 2018, India. Springer, 2019, pp. 469–480.

[15] M. Hammad, A. M. Iliyasu, A. Subasi, E. S. Ho, and A. A. Abd El- Latif, "A multitier deep learning model for arrhythmia detection," IEEE Transactions on Instrumentation and Measurement,

vol. 70, pp. 1–9, 2020.
[16] H. Li and P. Boulanger, "A survey of heart anomaly detection using ambulatory electrocardiogram (ecg)," Sensors, vol. 20, no. 5, p. 1461, 2020.
[17] N. Alamatsaz, M. Yazdchi, H. Payan, N. Alamatsaz, F. Nasimi et al., "A lightweight hybrid cnn-lstm model for ecg-based arrhythmia detection," arXiv preprint arXiv:2209.00988, 2022.
[18] H. W. Loh, C. P. Ooi, S. L. Oh, P. D. Barua, Y. R. Tan, F. Molinari, S. March, U. R. Acharya, and D. S. S. Fung, "Deep neural network technique for automated detection of adhd and cd using ecg signal," Computer methods and programs in biomedicine, vol. 241, p. 107775, 2023.
[19] L. B. Marinho, N. de MM Nascimento, J. W. M. Souza, M. V. Gurgel, P. P. Rebouc¸as Filho, and V. H. C. de Albuquerque, "A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification," Future Generation Computer Systems, vol. 97, pp. 564– 577, 2019.
[20] Y. Hagiwara, H. Fujita, S. L. Oh, J. H. Tan, R. San Tan, E. J. Ciaccio, and U. R. Acharya, "Computer-aided diagnosis of atrial fibrillation based on ecg signals: A review," Information Sciences, vol. 467, pp. 99–114, 2018.
[21] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," Scientific data, vol. 7, no. 1, p. 48, 2020.
[22] C. Reimann, P. Filzmoser, K. Hron, P. Kyncˇlova´, and R. Garrett, "A new method for correlation analysis of compositional (environmental) data–a worked example," Science of the Total Environment, vol. 607, pp. 965–971, 2017.
[23] R. M. Tayebi, Y. Mu, T. Dehkharghanian, C. Ross, M. Sur, R. Foley, R. Tizhoosh, and C. J. Campbell, "Automated bone marrow cytology using deep learning to generate a histogram of cell types," Communica- tions medicine, vol. 2, no. 1, p. 45, 2022.
[24] Y. Xu, B. Klein, G. Li, and B. Gopaluni, "Evaluation of logistic regression and support vector machine approaches for xrf based particle sorting for a copper ore," Minerals Engineering, vol. 192, p. 108003, 2023.
[25] A. Abdulhafedh, "Comparison between common statistical modeling techniques used in research, including: Discriminant analysis vs logistic regression, ridge regression vs lasso, and decision tree vs random forest," Open Access Library Journal, vol. 9, no. 2, pp. 1–19, 2022.
[26] A. Shafiei, A. Tatar, M. Rayhani, M. Kairat, and I. Askarova, "Artificial neural network, support vector machine, decision tree, random forest, and committee machine intelligent system help to improve performance prediction of low salinity water injection in carbonate oil reservoirs," Journal of Petroleum Science and Engineering, vol. 219, p. 111046, 2022.
[27] Igiri, C. G., V. I. E. Anireh, N. D. Nwiabu, and Daniel Matthias. "Comparative Analysis of Supervised Machine Learning Algorithms for ECG Arrhythmia Detection using Small Dataset.", International Journal of Computer Science and Mathematical Theory (IJCSMT), Vol 9. No.4, 2023.
[28] Mohebbian, Mohammad Reza, Hamid Reza Marateb, and Khan A. Wahid. "Semi-supervised active transfer learning for fetal ECG arrhythmia detection." Computer Methods and Programs in Biomedicine Update 3 (2023): 100096

Mr. Biswapriyo Sen is a final-year Electronics and Communication Engineering student with a keen interest in artificial intelligence and machine learning. He is driven by the potential of technology to solve real-world problems. His passion extends beyond the academic and technical aspects, as he is deeply motivated to apply his skills toward noble causes, aiming to make a positive impact on society. Through innovative AI/ML projects, he aspires to contribute meaningfully to fields such as healthcare, environmental sustainability, and social good.

Mr. Maharishi Kashyap is a B.Tech. student specializing in Electronics and Communication Engineering, graduating in 2024. His strong interest is using technology to improve not just healthcare but other fields as well. He is passionate about finding innovative solutions to enhance the livelihood of normal people. Hence, his research focuses on comparing different machine learning techniques, hoping to make a small but significant impact on medical technology and research.

Dr. Jitendra Singh Tamang is an Assistant Professor of Electronics & Communication Engineering Department at Sikkim Manipal Institute of Technology, specializing in Plasmonics and Nanosensors. He received his Ph.D. from National Institute of

Technology, Mizoram and has extensively published on the significance of Graphene and SPR phenomenon in various reputed journals and conferences.

Mr. Sital Sharma is an Assistant Professor in the Department AI and DS at SMIT. He is pursuing his PhD degree from SMIT. His primary research interest is in the field of Natural Language processing, Natural Language Understanding and Large Language Models. Presently he is working on developing techniques for preserving and digitizing low resources Himalayan Languages for NLP. Apart from teaching and research, he is actively involved in various activities like Mini and Major project coordinator, Project based learning coordinator etc. He has been actively involved in various funded projects amounting to a total sum of 39 Lakhs.

Ms. Rijhi Dey is an Assistant Professor of Electronics & Communication Engineering Department at IEM Newtown Campus, UEM, Kolkata, specializing in Biomedical Signal Processing and Control System. She is pursuing her Ph.D. from National Institute of Technology, Mizoram and has publications on various reputed journals and conferences.