Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14793
e-ISSN : 2541-2019
p-ISSN : 2541-044X

ECG-Based Arrhythmia Classification in Students Using Random Forest: A Case Study with Class Imbalance Analysis

Adya Zizwan Putra[1)*], **Ariyanto Sitorus**[2)], **Paulus Anggiat Ruben Simanjuntak**[3)],
Mega Cristin Angelina. H.[4)], **Kevin Agustino Situmorang**[5)]
[1,2,3,4,5)]Universitas Prima Indonesia
[1)]adyazizwanputra@unprimdn.ac.id, [2)]ariyantositorus1707@gmail.com, [3)]simanjuntakpaulus0@gmail.com,
[4)]megacristin15@gmail.com, [5)]kevinagustino1108@gmail.com

Submitted :May 20, 2025 | **Accepted** : May 31, 2025 | **Published** : Jun 3, 2025

Abstract: Arrhythmia is a heart rhythm disorder that can indicate a student's heart health status. This research aims to develop a Random Forest model to classify arrhythmia in students based on ECG signals. ECG data was collected from 100 students at SMK Swasta Teladan Sumatera Utara 2 after learning activities. The extracted signal features include RR interval, PR interval, QRS duration, QT interval, ST segment, beats per minute (BPM) and R/S ratio. Data labeling was carried out manually by the researchers based on the range of ECG feature values that had been determined by the doctor for each class: Normal, Abnormal, Potential Arrhythmia and Very Potential Arrhythmia. The dataset is divided into 70% for training and 30% for testing. SMOTE is applied to address class imbalance. The model achieved 80% accuracy with the best performance in normal class with precision, recall and f1-score of 94%. However, no samples were identified for Potential Arrhythmia class, as there were no extracted feature values that met the criteria set by the doctor, so model could neither learn nor make predictions for this category, even after applying balancing methods such as SMOTE. For further research, based on these findings, it highlights the need for balanced class representation and expert-guided labeling to improve the performance of ECG - based arrhythmia classification.

Keywords: Arrhythmia; Electrocardiogram; Data Labeling; *Random Forest*; *SMOTE*

INTRODUCTION

SMKS Teladan Sumatera Utara 2 is one of the schools that is the focus of research for the classification of arrhythmia in students after they complete learning activities (Data Pokok Pendidikan (DAPODIK) Direktorat Jenderal Pendidikan Anak Usia Dini, Pendidikan Dasar dan Pendidikan Menengah Kementerian Pendidikan Dasar dan Menengah 2025). The intense learning process, especially over a long period of time, can cause fatigue both physically and mentally, which can have health risks, including disturbances in heart rhythm. Therefore, it is very important to analyze the heart condition of students after they complete learning activities. This aims to detect possible arrhythmias early. The heart is an important organ that functions to pump blood throughout the body (Dwi Muthohhar and Prihanto 2023). The heart has a high-risk disease and is a major factor in the highest cause of death in the world. Based on a report from the World Health Organization (WHO), more than 2 million deaths occurred in 2000 and increased to 9 million deaths in 2019, accounting for 16% of the total deaths caused by heart disease (Switzerland 2020). In Indonesia, according to the results of the Basic Health Research (Riskesdas) that was conducted, the number of heart disease patients increased from 0.5% in 2013 to 1.5% in 2018. Currently, the number of heart disease patients has reached more than 2.78 million (Kementerian Kesehatan Republik Indonesia 2019). One type of heart disease is arrhythmia. It is a disorder of the heart rhythm characterized by an irregular heartbeat, either too fast or too slow compared to the normal (Islam et al. 2024). Arrhythmia can be identified by analyzing heart recordings or electrocardiograms (ECGs) efficiently at an early stage. An electrocardiogram is a diagnostic tool that records the electrical activity of the heart, where changes in heart rate can be detected (Anbalagan et al. 2023). An ECG records and detects changes in heart activity signals using electrodes placed on the surface of the skin (Given Hamonangan et al. 2024). The rhythm of the heartbeat can be represented in the

*name of corresponding author

form of a sinusoidal curve. Generally, the ECG curve consists of several parts, namely the P wave, Q wave, R wave, S wave, and T wave. The P wave represents atrial depolarization, while the 'QRS' complex indicates ventricular depolarization. On the other hand, the T wave illustrates the process of ventricular repolarization (Yunarti Butarbutar et al. 2022). While numerous studies (Ansari et al. 2023) focus on arrhythmia detection using benchmark datasets (e.g., MIT-BIH), limited work explores real-world ECG signals collected in non-clinical environments such as schools.

From the problems that have been outlined, the research team decided to conduct ECG-Based Arrhythmia Classification on Students Using Random Forest: A Case Study with Class Imbalance Analysis. The selection of the random forest algorithm is due to it being one of the best machine learning algorithms for classification and detection (Tamba 2022). By utilizing the Random Forest algorithm, this research is expected to enhance the accuracy and speed of classifying arrhythmias in students, so that more people, including educators and parents, can easily understand the students' heart health conditions, allowing for immediate management of arrhythmia risks. This study uniquely contributes by utilizing real ECG signals collected from students in a non-clinical setting (school), making the classification results more relevant to real-world conditions.

LITERATURE REVIEW

Previous studies have explored various machine learning algorithms for classifying arrhythmias based on ECG signals. The research conducted by (Al-Shammary et al. 2024) evaluated four algorithms: K-Nearest Neighbor (KNN), Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM). This study proposes a Chi-square distance-based classification method to improve performance. The approach successfully achieved a precision level of 98% on the MIT-BIH Arrhythmia dataset. However, because it only used benchmark data, this study has not tested the model's resilience to real-world data variations, such as non-clinical populations like those in schools or with different signal characteristics. Research conducted by (Matematika and Adi n.d.) compared SVM, KNN, and RF algorithms to predict heart failure. The results showed that SVM and RF had the highest accuracy of 97%, while KNN achieved 93% accuracy on the Heart Failure Clinical Records Dataset from the UCI machine learning repository, which does not specifically represent ECG signals for arrhythmia, thus limiting its direct relevance to arrhythmia classification. Meanwhile, the research conducted by (Pal and Parija 2021) utilized the Random Forest method for predicting heart disease and achieved an accuracy of 86.9%, sensitivity of 90.6%, and specificity of 82.7%. However, the scope of the diseases studied is still general and does not focus on arrhythmia as a specific condition. The dataset was accessed from the Kaggle site. The research conducted by (Sari et al. 2023) developed a smart healthcare web application for ECG signal classification using ensemble techniques on the Random Forest algorithm and achieved an accuracy of 96% with an F1 score of 0.97 on the MIT-BIH Arrhythmia dataset. The advantage of this research is its integration with real applications, but it still uses the MIT-BIH benchmark data, which does not yet represent local data variations. On the other hand, a study conducted by (Farell and Al n.d.) classified ECG signals from 30 subjects in the age range of 20 - 25 years into four arrhythmia risk categories using the KNN algorithm with 93% accuracy. However, the limited number of subjects reduces the generalization of the results. The research conducted by (Li et al. 2021) combined the AdaBoost method with Random Forest on the MIT-BIH dataset and achieved an accuracy of 99.11%. Although the accuracy is very high, these results are still limited to standard datasets that do not yet represent the variability of signals in everyday life contexts. To clarify the comparison between studies, here is a summary presented in table 1.

Table 1. summary of previous studies

No	researchers	Algorithms	datasets	Accuracy	Information
1	(Al-Shammary et al. 2024)	KNN, RF, DT + Chi-square	MIT-BIH	98 %	The Chi-square method improves the precision of arrhythmia detection.
2	(Matematika and Adi n.d.)	SVM, KNN, RF	Heart Failure Clinical Records Dataset (UCI Machine Learning)	97% (SVM, RF)	Heart failure prediction; RF and SVM outperform KNN.
3	(Pal and Parija 2021)	Random Forest	Kaggle	86,9%	Sensitivity 90.6%; specificity 82.7%.
4	(Sari et al. 2023)	RF + Ensemble	MIT-BIH	96%	F1 Score: 0.97; web-based application.
5	(Farell and Al n.d.)	K-Nearest Neighbor	Primary data from 30 subjects (aged 20–25 years)	93%	Classification of 4 arrhythmia risk classes

6	(Li et al. 2021)	AdaBoost + Random Forest	MIT-BIH	99,11%	A combination of boosting and feature selection.

From various previous studies mentioned above, there is still an important research gap: the majority of studies only use standard benchmark datasets such as MIT-BIH, which do not adequately represent the variability of ECG signals from the local population. Not many studies have tested the effectiveness of arrhythmia classification models in real-world contexts such as educational environments, particularly on students who have just finished studying and may be experiencing physical, stress, or mental fatigue. This study offers a more practical approach because it uses actual ECG data from a student population, which allows for model testing in real - world conditions that are relevant for early screening. This approach supports the development of a contextual arrhythmia classification model that can be applied in the field of education. Furthermore, the research conducted by (Islam et al. 2024) emphasizes that most arrhythmia classification models still rely on benchmark data such as MIT-BIH, and only a few consider real conditions to support the adoption of models in real applications such as single-lead ECG-based wearable devices. Their study proposes CAT-NET, a hybrid model based on CNN, attention, and transformer that achieves an accuracy of 99.14% and a macro F1 score of 94.69%. Despite their high performance, they themselves highlight the need for validation in a broader and more realistic population context, including beyond clinical settings. This reinforces the urgency of this research as a response to these limitations.

METHOD

This study uses the Random Forest algorithm implementation approach to classify arrhythmia based on electrocardiogram (ECG) signals collected from students at SMK Swasta Teladan Sumatera Utara 2. The Random Forest algorithm was chosen for its high accuracy, resistance to outliers, and ability to work well with datasets that do not always have a normal distribution. In addition, Random Forest is non-parametric and capable of handling large amounts of data without requiring specific distribution assumptions. The initial stages of the research began with the collection of primary data in the form of ECG signals taken directly from students. The ECG signals obtained will undergo pre-processing, where noise in the data will be removed, such as body movements, muscle contractions, and breathing. This step is important to ensure that the data used in the subsequent analysis process is of good quality. The pre-processing stage includes an extraction stage to obtain features with numerical values for the signal interval features rr, pr, qrs, qt, st, bpm, and r/s ratio. After the feature extraction stage, manual labelling is performed to determine the output for each signal interval. The four classes: normal, abnormal, potentially arrhythmic, and highly potentially arrhythmic are based on annotations by medical experts for each ECG wave segment. After manual labeling is complete and a dataset containing ECG features and output features is formed, the next stage is to develop a classification model using the random forest algorithm. Random Forest is a classification algorithm in machine learning that works by building a number of decision trees to group new data. Random Forest is used to classify arrhythmias into four classes by combining the prediction results from multiple decision trees with random samples of data between training data and test data based on the majority of the specified number of neighbours. The calculation of information value can be formulated as follows:

$$info\ (D) = -\sum_{i=1}^{n} pilog2(pi) \qquad (1)$$

n is the number of target classes and p_i is the proportion of class i to the data partition D. Once the initial information values are calculated, the next step is to calculate the gain value of each attribute to determine which attribute best separates the data. The gain value is calculated using the formula:

$$Gain\ (A) = Info\ (D) - \sum_{j=1}^{v} \frac{|Dj|}{|D|}\ X\ info\ (Dj) \qquad (2)$$

v is the number of partitions based on attributes A_1 $|D_j|$ is the amount of data in the j-th partition, and $|D\ |$ is the total amount of data. The attribute with the highest gain value will be selected as the root of the decision tree (Khatib Sulaiman et al. n.d.).

This research is supported by various tools and materials that play an important role in the data collection and analysis process. The main tool used is the electrocardiogram (ECG) device, which functions to record the heart's rhythm. The ECG signal data obtained from this device is the main basis for the arrhythmia classification process. A laptop with a Windows operating system is used to support the operation of the ECG device and data processing. This laptop functions as a medium for connecting to the ECG device via Bluetooth, as well as a device for running electrocardiogram software. The software allows researchers to record and store the electrical activity of the respondents' hearts. Wi-Fi connectivity is also an important component in this study. Its function is to transfer data wirelessly between the ECG machine and other devices, especially when data analysis is performed using a cloud-based platform. One of the platforms used is Google Colaboratory. Google Colaboratory is utilized as a computational experiment that supports Python-based programming. Through this platform, researchers build and train classification models using the random forest algorithm.

The dataset obtained as the main material in this study was ECG data obtained directly from experiments on students at SMK Swasta Teladan Sumatera Utara 2. The data collection procedure was carried out by asking students to participate as respondents in heart signal testing. Data collection was carried out by attaching measuring devices to the respondents' chests. The data collection process was carried out using two scenarios, namely sitting and walking, each recorded for a duration of one minute. The flow in this study can be seen in Fig. 1.

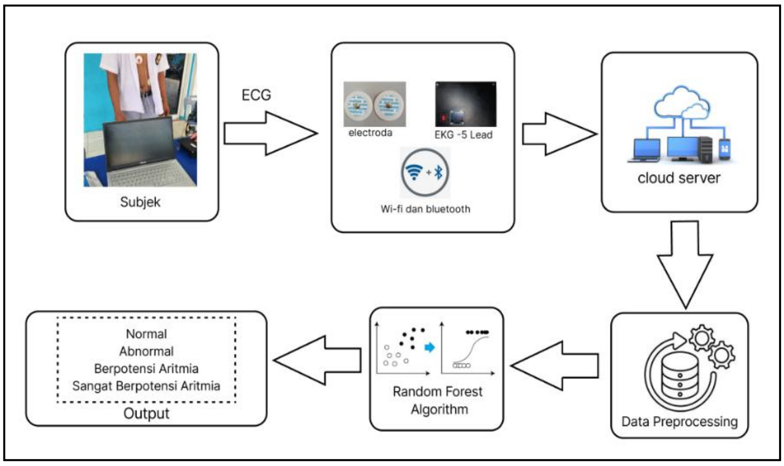


Fig. 1 Block Diagram

RESULT

This study encompasses several stages in the classification of arrhythmia using the Random Forest algorithm, including data preparation, model testing, and result discussion. The data preparation stage consists of two main processes: feature extraction from ECG signals and manual labeling to determine the output classes based on medical expert evaluation. Feature extraction was conducted in Google Colaboratory, starting from raw electrocardiogram signals recorded from Lead II in millivolts (mV). These signals represent the heart's electrical activity as amplitude over time. The data used was in its raw form, without any preprocessing such as noise filtering.

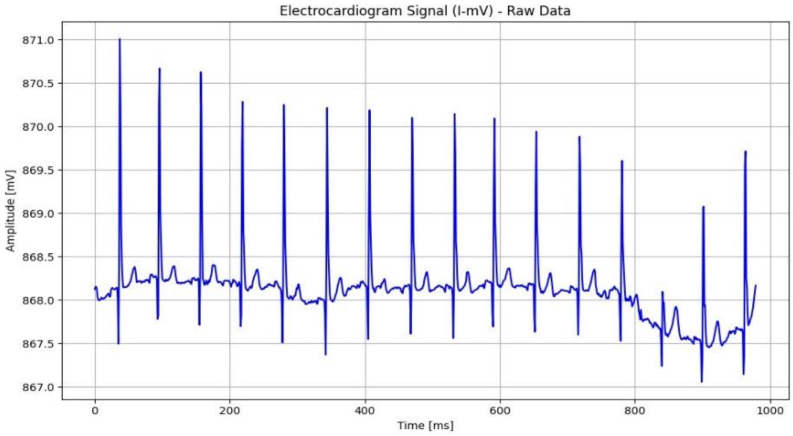


Fig. 2 Raw Signal

Before proceeding to the feature extraction stage, the data is first cleaned, as the signals are still in a raw state. The technique used involves filtering the signals using Baseline Correction, Butterworth, and FIR Filtered. The Butterworth Low Pass Filter includes detrending. The function of detrending is to maintain the stability of the signal so that it remains around the zero line.

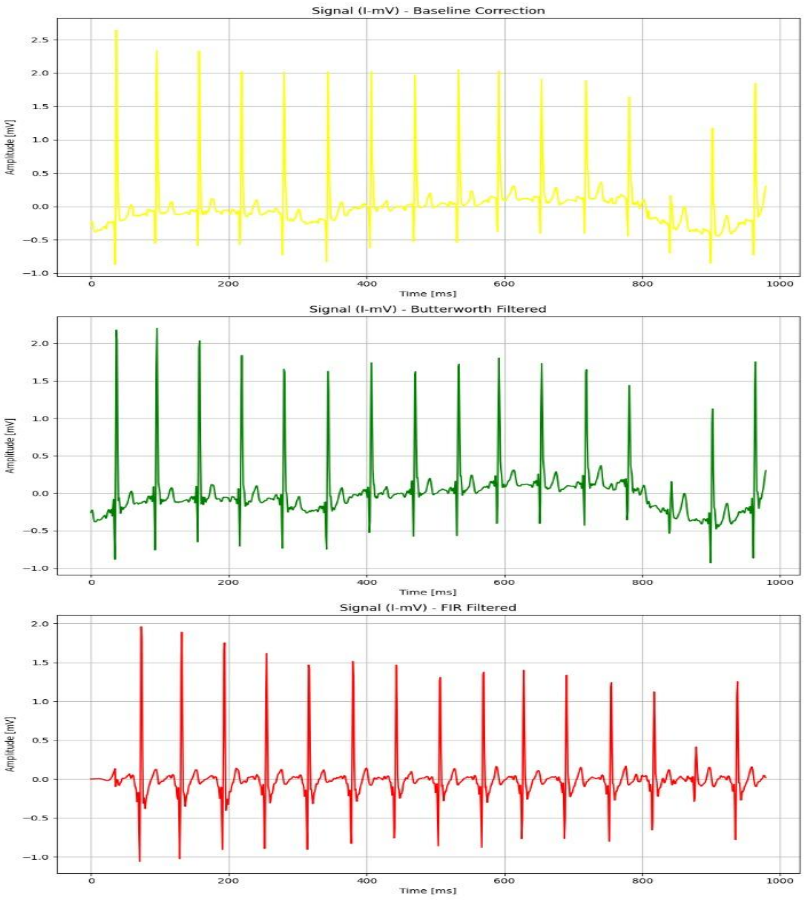


Fig. 3 Filtering

Peak detection or PQRST peak identification is used to visualize the PQRST waveform, from which combined features such as rr, pr, qt, st, bpm, and r/s ratio are extracted. These features are then used as input variables for classification using the Random Forest algorithm.

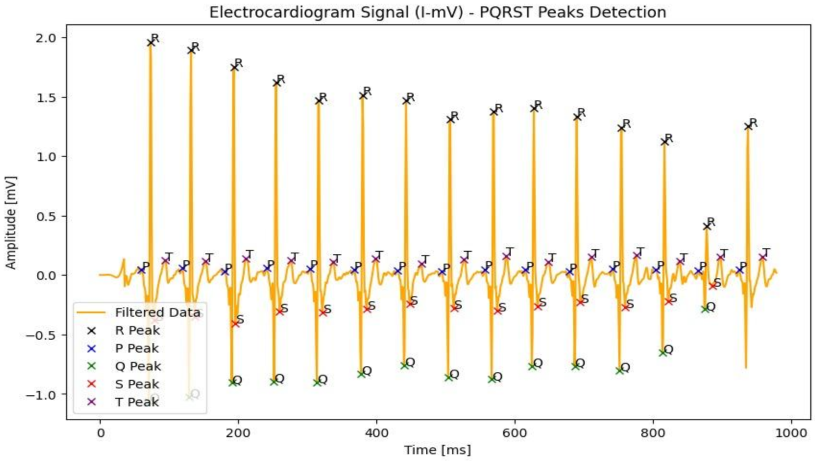


Fig. 4 PQRST Peaks

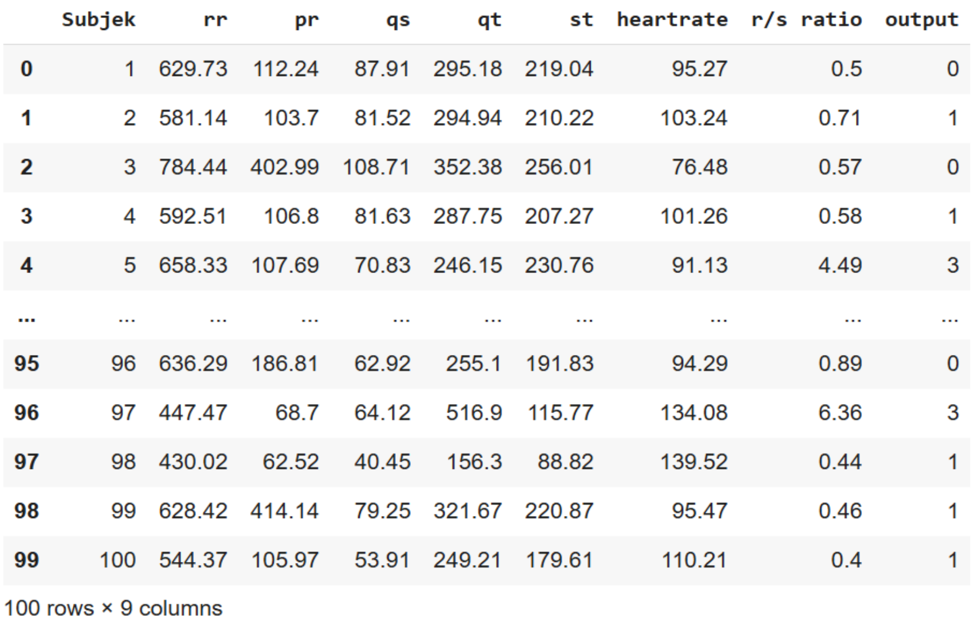
After the feature extraction process from the ECG signal has been successfully completed and the combined features rr, pr, qt, st, bpm, and r/s ratio have been obtained, manual labeling is carried out to determine the output feature of the dataset in this study. The classification into four output classes is based on medical criteria established by expert physicians. The Excel file containing the labeling results, discussed and validated by medical experts, can be accessed via the following link: https://bit.ly/medical_labeling_ecg. A summary of the labeling criteria is presented in table 2.

*name of corresponding author

Table 2. Class labeling criteria based on ECG parameters (based on medical criteria)

RR (ms)	PR (ms)	QRS (ms)	QT	ST	BPM	Rasio R/S	Kelas
600 - 100	120-200	60-100	350-440	80-120	60-100	<1	Normal
<600	<120	<60	<350	<80	<60	≥ 1	Abnormal
600-1000	190-200	101-120	441-460	120-130	101-110	≥ 1	Potential Arrhythmia
600-1000	>200	>120	>460	>130	>111	≥ 1	Very Potentially Arrhythmia

After the entire feature extraction process from the ECG signal has been successfully completed and the data has been manually labeled based on medical criteria defined by expert doctors, the data preparation stage can be considered complete.



	Subjek	rr	pr	qs	qt	st	heartrate	r/s ratio	output
0	1	629.73	112.24	87.91	295.18	219.04	95.27	0.5	0
1	2	581.14	103.7	81.52	294.94	210.22	103.24	0.71	1
2	3	784.44	402.99	108.71	352.38	256.01	76.48	0.57	0
3	4	592.51	106.8	81.63	287.75	207.27	101.26	0.58	1
4	5	658.33	107.69	70.83	246.15	230.76	91.13	4.49	3
...	...	...	...	...	...	...	...	...	...
95	96	636.29	186.81	62.92	255.1	191.83	94.29	0.89	0
96	97	447.47	68.7	64.12	516.9	115.77	134.08	6.36	3
97	98	430.02	62.52	40.45	156.3	88.82	139.52	0.44	1
98	99	628.42	414.14	79.25	321.67	220.87	95.47	0.46	1
99	100	544.37	105.97	53.91	249.21	179.61	110.21	0.4	1

100 rows × 9 columns

Fig. 5 Research Dataset

The result of this process is a structured dataset containing feature values such as RR, PR, QS, QT, ST, BPM, and R/S ratio, as well as class labels representing the heart condition of each subject (Normal, Abnormal, Potential Arrhythmia, and High Potential Arrhythmia). These class labels are then converted into numerical form to facilitate the model training process. The numerical coding used is as follows: 0 for Normal, 1 for Abnormal, 2 for Potential Arrhythmia, and 3 for High Potential Arrhythmia. This resulting dataset is then used as input in the training and testing phases of the classification model for arrhythmia detection based on ECG signals.

The next step is to test the classification model using the random forest algorithm. The available dataset is divided into two parts, namely training data and test data, with a ratio of 70:30. After the model is trained, it is tested on the test data to determine the performance of the random forest model. The evaluation is carried out using a confusion matrix that describes the number of correct and incorrect predictions for each class.

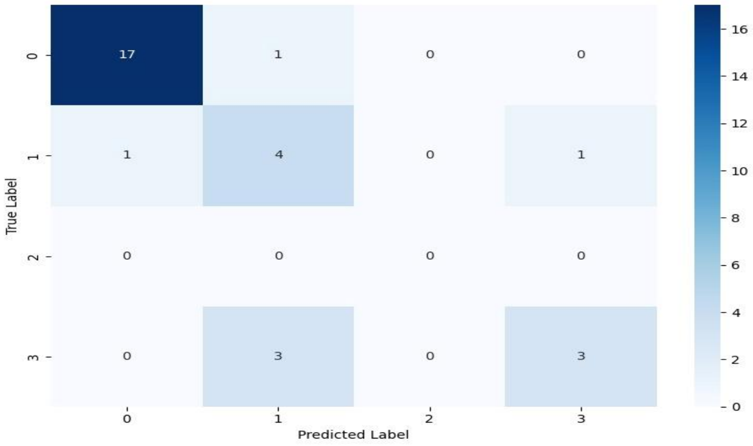


Fig. 6 Confusion Matrix

Confusion matrix results of model testing for four classes, namely 0 for Normal, 1 for Abnormal, 2 for Potential Arrhythmia, and 3 for Highly Potential Arrhythmia. The results show that the model performs reasonably well in classifying the Normal class, with 17 correct predictions out of 18 data points, and the Abnormal class, with 4 correct predictions out of 6 data points. However, the model still makes prediction errors in the Highly Potential Arrhythmia class, where only 3 out of 6 data points were correctly classified, while the remaining 3 were incorrectly classified as Abnormal. The Potential Arrhythmia class was not predicted at all. This is due to the manual labeling process, which referenced expert physician data, not including any data that met the criteria for that class, leaving the model without sufficient samples to learn from. Consequently, this class was underrepresented during model training and testing.

Table 3. Model Evaluation Results

Kelas	Accuracy	Precision	Recall	F1-Score	Support
0	0.80	0.94	0.94	0.94	18
1		0.50	0.67	0.57	6
2		0.00	0.00	0.00	0
3		0.75	0.50	0.60	6

Table 3 presents the evaluation results of the model based on the confusion matrix using the Random Forest algorithm. It reports the performance metrics for each class, including precision, recall, F1-score, and support, along with the model's overall accuracy of 0.80 (80%). For class 0, the model demonstrates excellent performance, with precision, recall, and F1-score all at 0.94, supported by a sample count (support) of 18. Meanwhile, for class 1, the model achieved a precision of 0.50, recall of 0.67, and F1-score of 0.57, with a support of 6. This suggests that the model is reasonably capable of recognizing this class, although its precision remains moderate. In class 2, all evaluation metrics are recorded as zero due to the absence of data (support = 0) in the evaluation set; thus, the model's performance for this class could not be assessed. For class 3, the model obtained a precision of 0.75, recall of 0.50, and F1-score of 0.60, with a support of 6, indicating that the model can partially recognize this class, though the recall still requires improvement. Overall, the model demonstrates reasonably good accuracy, but its performance is inconsistent across classes—particularly impacted by the lack of data in class 2, which affects the overall evaluation.

For comparison, the performance of the proposed model is evaluated against the results reported by (Sari et al. 2023), who also employed the Random Forest algorithm. The comparison of evaluation metrics is presented in table 4 below:

Table 4. Comparison of Model Evaluation Results with the Study by (Sari et al. 2023)

Model	Accuracy	Precision	Recall	F1-Score
(Sari et al. 2023)	0.96	1.00	0.94	0.97
This Research	0.80	0.79	0.78	0.78

From table 4, it can be seen that the model developed by (Sari et al. 2023) shows higher performance compared to the model in this study, especially in the recall metric (100%) and F1-score (0.97). The accuracy of their model is also higher, at 96%, while the model in this study achieved an accuracy of 80%. Nevertheless, this study has a more complex approach, namely performing multi-class classification of four categories of arrhythmia risk, while most previous studies, including (Sari et al. 2023), tend to use a two-class classification approach. This complexity adds value in the context of early detection with more levels of risk categories. The main weakness of this study is the underrepresentation of the "Potential Arrhythmia" class (class 2) in the dataset. Although efforts have been made to balance the data using the SMOTE method, this class remains unformed because, medically, no data meeting the numerical feature criteria for inclusion in this class has been found based on expert doctor annotations. As a result, the model cannot learn to recognize patterns from this class, so the evaluation metrics for that class are zero. This affects the overall performance of the model, which appears uneven.

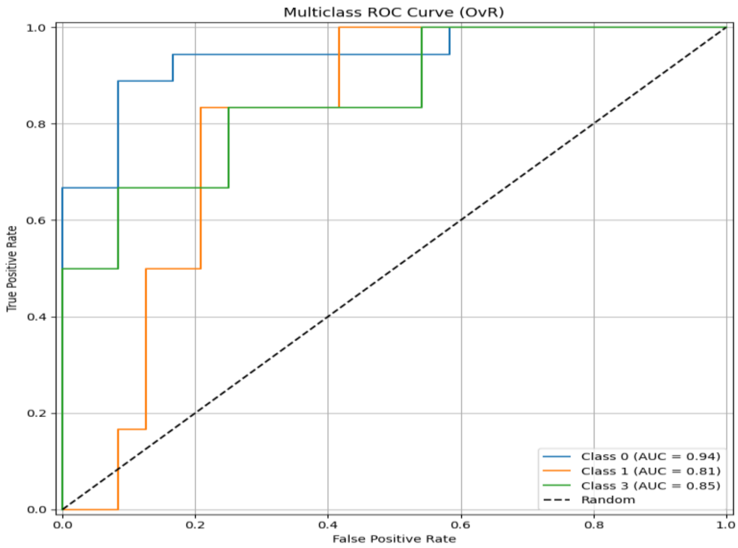


Fig. 7 ROC Curve

As an additional evaluation, the measurement of the AUC (Area Under Curve) for each class was also carried out to understand the model's ability to distinguish between positive and negative classes. The AUC (Area Under the Curve) value per class provides a more detailed picture of the model's ability to differentiate each class. In the image, it can be seen that class 0 (Normal) performs best with an AUC of 0.94, indicating the model's very good capability in classifying normal conditions. Meanwhile, class 1 (Abnormal) has an AUC value of 0.81, and class 3 (Very Potential Arrhythmia) received an AUC of 0.85. Although this value still shows quite good performance, there remains a chance of misclassification between classes, especially in classes with less data representation or overlapping feature characteristics. This AUC evaluation complements the analysis of basic classification metrics such as accuracy and F1-score, and provides additional insights into the sensitivity and specificity of the model for each class. In the future, this approach can be used as a basis for further model optimization, particularly in addressing class imbalance and strengthening classification in the minor class.

CONCLUSIONS

This study makes an important contribution to the development of an arrhythmia classification system based on electrocardiogram (ECG) signals in students at SMK Swasta Teladan Sumatera Utara 2, particularly through the provision of an ECG dataset taken from students with labels referenced from medical criteria. This dataset was generated through feature extraction from raw signals and manual labeling performed by researchers based on expert medical criteria, then divided into 70% training data and 30% test data. The results of the experiment show that the Random Forest model is capable of achieving an accuracy of 80%, with good performance in majority classes such as “Normal” and “Abnormal.” However, the model failed to classify the “Potential Arrhythmia” class due to the absence of samples that met the criteria for that class during labeling. Although data balancing techniques such as SMOTE were applied, the limited number of minority samples caused the model to remain biased toward the majority class. The main limitations of this study include the relatively small sample size and the uneven distribution of data across classes, especially in the minority class. This indicates that data distribution and labeling quality factors significantly influence the overall model performance. Therefore, for future research, it is recommended to compare the performance of random forest with other algorithms such as XGBoost, Gradient

Boosting, or deep learning-based methods that may be sensitive to minority classes. Additionally, redefining the classification problem into a binary classification between 'Normal' and 'Arrhythmia' can be a more stable alternative, especially when minority data is very limited. Data labeling should also involve more than one doctor to improve the quality and objectivity of the labels. Further testing can be conducted on datasets that include other medical conditions, such as patients with a history of other cardiovascular disorders, to evaluate the model's generalization in classifying more diverse ECG signals.

REFERENCES

Al-Shammary, Dhiah, Mustafa Noaman Kadhim, Ahmed M. Mahdi, Ayman Ibaida, and Khandakar Ahmed. 2024. “Efficient ECG Classification Based on Chi-Square Distance for Arrhythmia Detection.” *Journal of Electronic Science and Technology* 22(2). doi:10.1016/j.jnlest.2024.100249.

Anbalagan, Thivya, Malaya Kumar Nath, D. Vijayalakshmi, and Archana Anbalagan. 2023. “Analysis of Various Techniques for ECG Signal in Healthcare, Past, Present, and Future.” *Biomedical Engineering Advances* 6:100089. doi:10.1016/j.bea.2023.100089.

Ansari, Yaqoob, Omar Mourad, Khalid Qaraqe, and Erchin Serpedin. 2023. “Deep Learning for ECG Arrhythmia Detection and Classification: An Overview of Progress for Period 2017–2023.” *Frontiers in Physiology* 14.

Data Pokok Pendidikan (DAPODIK) Direktorat Jenderal Pendidikan Anak Usia Dini, Pendidikan Dasar dan Pendidikan Menengah Kementerian Pendidikan Dasar dan Menengah. 2025. https://dapo.dikdasmen.go.id/sekolah/2EDAE0CCE0B6BA1B89C2.

Dwi Muthohhar, Jalis, and Agus Prihanto. 2023. “Analisis Perbandingan Algoritma Klasifikasi Untuk Penyakit Jantung.” *Journal of Informatics and Computer Science* 04.

Farell, E., & Al. (n.d.). “Classification of Arrhythmia Potential using the K-Nearest Neighbor Algorithm,” https://www.internetworkingindonesia.org/index.php/iij/article/view/49.

Given Hamonangan, Jonathan, Anita Sofia Rahmi, Sean Julius Lase, Nendi Suhendi Syafei, and Arjon Turnip. 2024. “JIIF (Jurnal Ilmu Dan Inovasi Fisika) DETEKSI DINI ARITMIA MENGGUNAKAN K-NEAREST NEIGHBOUR.” 08(01):86–95. doi:10.24198/jiif.v8i1.

Islam, Md Rabiul, Marwa Qaraqe, Khalid Qaraqe, and Erchin Serpedin. 2024. “CAT-Net: Convolution, Attention, and Transformer Based Network for Single-Lead ECG Arrhythmia Classification.” *Biomedical Signal Processing and Control* 93. doi:10.1016/j.bspc.2024.106211.

Kementerian Kesehatan Republik Indonesia. 2019.“Penyakit Jantung.” https://p2ptm.kemkes.go.id/informasi-p2ptm/penyakit-jantung?page=13.

Khatib Sulaiman, Jln, Abd A. Mizwar Rahim, Inggrid Yanuar Risca Pratiwi, Muhammad Ainul Fikri, and Universitas Amikom Yogyakarta. n.d. “Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier.” *Indonesian Journal of Computer Science Attribution* 12(5):2023–2995.

Li, Runchuan, Wenzhi Zhang, Shengya Shen, Jinliang Yao, Bicao Li, Bing Zhou, Gang Chen, and Zongmin Wang. 2021. “An Intelligent Heartbeat Classification System Based on Attributable Features with AdaBoost+Random Forest Algorithm.” *Journal of Healthcare Engineering* 2021. doi:10.1155/2021/9913127.

Matematika, Jurnal Ilmiah, and Socayo Adi. n.d. “MATHunesa Tahun 2022 KOMPARASI METODE SUPPORT VECTOR MACHINE (SVM), K-NEAREST NEIGHBORS (KNN), DAN RANDOM FOREST (RF) UNTUK PREDIKSI PENYAKIT GAGAL JANTUNG Atik Wintarti.”

Pal, Madhumita, and Smita Parija. 2021. “Prediction of Heart Diseases Using Random Forest.” in *Journal of Physics: Conference Series*. Vol. 1817. IOP Publishing Ltd.

Sari, Juni Nurma, Putri Madona, Hari Kusryanto, Muhammad Mahrus Zain, and May Valzon. 2023. “Electrocardiogram Signals Classification Using Random Forest Method for Web-Based Smart Healthcare.” *International Journal of Advances in Applied Sciences* 12(2):133–43. doi:10.11591/ijaas.v12.i2.pp133-143.

Switzerland, Geneva. 2020.“WHO Reveals Leading Causes of Death and Disability Worldwide: 2000-2019.” https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019.

Tamba, Parsaoran. 2022. “PREDIKSI PENYAKIT GAGAL JANTUNG DENGAN MENGGUNAKAN RANDOM FOREST.” *Jurnal Sistem Informasi Dan Ilmu Komputer Prima)* 5(2).

Yunarti Butarbutar, Serly, Christian Deniro Napitupulu, Nessa Sanjaya Ginting, Evta Indra, and Delima Sitanggang. 2022. “Http://Infor.Seaninstitute.Org/Index.Php/Infokum/Index INFOKUM Is Licensed under a Creative Commons Attribution-Non Commercial 4.0 International License (CC BY-NC 4.0) CLASSIFICATION OF ELECTROCARDIOGRAM (ECG) WAVES OF HEART DISEASE USING THE XGBOOST METODE METHOD.” *JURNAL INFOKUM* 10(2).