Journal of Technology and Informatics (JoTI)
Vol. 6, No. 1, October 2024
P-ISSN 2721-4842
E-ISSN 2686-6102
DOI: 10.37802/joti.v6i1.792

Arrhythmia Classification with ECG Signal using Extreme Gradient Boosting (XGBoost) Algorithm

Diah Asmawati[1], Lukman Arif Sanjani[2], Christiant Dimas Renggana[3], Chastine Fatichah[4], Tanzilal Mustaqim[5]
[1,2,3,4,5]Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
e-mail: 6025232014@student.its.ac.id[1], 6025232027@student.its.ac.id[2], 6025232016@student.its.ac.id[3], chastine@if.its.ac.id[4], 7025222005@student.its.ac.id
*Corresponding author: E-mail: 6025232027@student.its.ac.id

Abstract: *Heart disease is one of the most dangerous illnesses because it has the potential to take people's lives. One of the causes of heart disease is arrhythmia, an abnormal condition of the heartbeat. To diagnose arrhythmia, analysis of electrocardiographic (ECG) signals can be performed. However, this analysis is very difficult to do conventionally and has the potential for errors, so there is a need for automatic ECG classification to detect arrhythmia. This study aims to fill the research gap by creating an ECG classification model to detect arrhythmia using the XGBoost algorithm. The results are quite good for each class, with accuracies for class N at 98.87%, class SVEB at 99.37%, class VEB at 99.4%, class F at 99.75%, and class Q at 99.99%. However, compared to existing methods in previous research, these results are still considered not better than those models.*

Keywords*: Arrhythmia; Classification; ECG; XGBoost*

INTRODUCTION

Cardiovascular disease is one of the most common diseases affecting individuals of all ages, from infants to the elderly [1]. It is also the leading cause of death worldwide, accounting for 31% of global deaths [2]. Heart disease has significant economic impacts globally. It is recorded that heart disease causes an annual economic burden of €210 billion in Europe and $555 billion in the United States [3]. One of the causes of heart disease is arrhythmia, a condition where the heart's rhythm is abnormal [4]. A common type of arrhythmia is atrial fibrillation, characterized by a rapid and irregular heartbeat, with a global prevalence of 46.3 million cases [5]. Arrhythmia is responsible for 80% of deaths caused by heart disease [6]. This highlights the significant impact of arrhythmia in addressing heart disease issues.

One method to diagnose arrhythmia is through the use of electrocardiographic (ECG) signals [7]. However, this method presents a significant challenge due to the difficulty in detecting and categorizing different waveforms and morphologies within a signal, making the analysis only possible by specialists in the field [8]. This results in ECG heartbeat classification analysis being very time-consuming and prone to errors due to fatigue [9]. Therefore, there is a need to replace conventional methods involving human labor with an automated heartbeat classification system.

This need can be addressed by leveraging the advancements in technology, which open opportunities to create a heartbeat classification model. Such a model can be developed thanks to the presence of artificial intelligence (AI) today. AI can create a classification model (supervised learning) that requires labeled data as a reference for predictions [10]. This is also supported by AI's ability to develop and improve its performance over time with minimal or no human intervention [11]. The utilization of AI in the form of machine learning for early diagnosis of various diseases has already had a substantial impact previously [12].

Several previous studies have conducted various experiments to create models for classifying types of arrhythmias based on ECG. One study utilized variable length heartbeats for classification with a combination of CNN and LSTM techniques, achieving a final accuracy of 98.10% [7]. Another study using a similar classification approach, CNN-LSTM, achieved a higher accuracy of 99.27% [13]. The use of CNN for feature extraction was also applied to another classifier, XGBoost, which achieved an accuracy of 99.69% [14]. Ensemble learning was also implemented using Random Forest and Support Vector Machine algorithms, resulting in an overall accuracy of 98.21% [15]. Another ensemble learning approach that included the kNN algorithm resulted in a lower accuracy of 97.8% [16].

This study aims to create an arrhythmia classification model based on ECG from a dataset using the XGBoost algorithm. The limitation of this study is that the model will only be tested on one dataset, and the proposed methodology will be implemented using only one algorithm. The expectation is that this research can contribute by filling the research gap where there is still no study that process the MIT-BIH Arrhythmia dataset using the XGBoost classifier with the same approach as this study.

METHOD

The proposed method consists of four major parts: pre-processing, feature extraction using TSFEL, model training, and model testing. The workflow of this method can be seen in Fig 1.

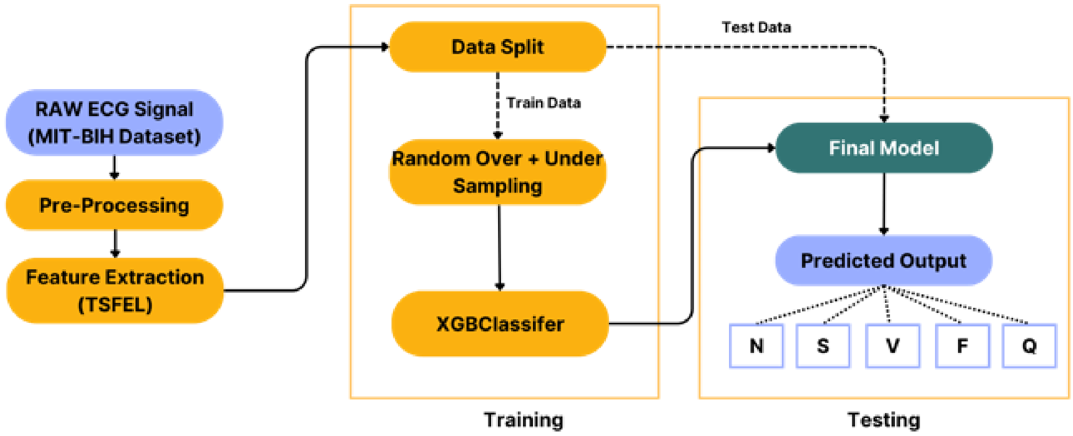


Figure 1. Proposed Methodology Flow

Dataset Description

The dataset used in this study is the public MIT-BIH Arrhythmia Database, which consists of 48 ECG recordings, each spanning 30 minutes. These samples were taken at a frequency of 360Hz per channel with an ADC resolution of 11 bits at a 10mV interval [17], [18]. This dataset began distribution in 1980 and has undergone various updates, with the latest revision edited in 2018. Heartbeat labels are provided in the database based on annotations independently made by one or more cardiologists. Classification labels are then categorized based on the AAMI (Association for the Advancement of Medical Instrumentation) standards, where classes are divided into normal (N), supraventricular (SVEB or S), ventricular (VEB or V), fusion (F), and unclassified (Q) [19]. Each annotation code given in the MIT-BIH dataset is grouped according to the AAMI standards as shown in Table 1.

Table 1. Mapping Of Heartbeat Class Categories in the MIT-BIH Dataset Based on AAMI Standards

ID	Heartbeat Types Based on AAMI Standards	Heartbeat Types Based on MIT-BIH
0	N (Normal or other than S, V, F, Q)	*normal beat (N), left bundle branch block beat (L), right bundle branch block beat (R), atrial escape beat(e), nodal (junctional) escape beat (j)*
1	SVEB (*Supraventricular Ectopic Beat*)	*atrial premature beat (A), aberrated atrial premature beat (a), nodal (junctional) premature beat (J), supraventricular premature beat (S)*
2	VEB (*Ventricular Ectopic Beat*)	*premature ventricular contraction (V), ventricular escape beat (E)*
3	F (*Fusion Beat*)	*fusion of ventricular and normal beat (F)*
4	Q (*Unknown Beat*)	*paced beat (P), fusion of paced and normal beat (f), unclassified beat (U)*

In each recording for each patient, there are 2 readable ECG channels. However, in this case, only one channel, the MDII channel, is used for the classification process.

Pre-process and Segmentation

Pre-processing of the ECG recording data aims to filter and reduce noise in the signal, making it optimal for recognizing classes in the learning approach in the subsequent process. The filtering is performed by applying a median filter with signal widths of 200ms and 600ms, then subtracting the original signal and baseline values to produce a corrected signal. Next, the denoising process is carried out on this signal using the db4 Discrete Wavelet Transform (DWT), applying a high-pass filter to the DWT coefficients, and then inverting it to obtain a cleaner signal.

After pre-processing, the signal is segmented based on the R-Peak annotations provided in the MIT-BIH dataset. As a result, the data generated totals 101,147 samples with the number of each class as

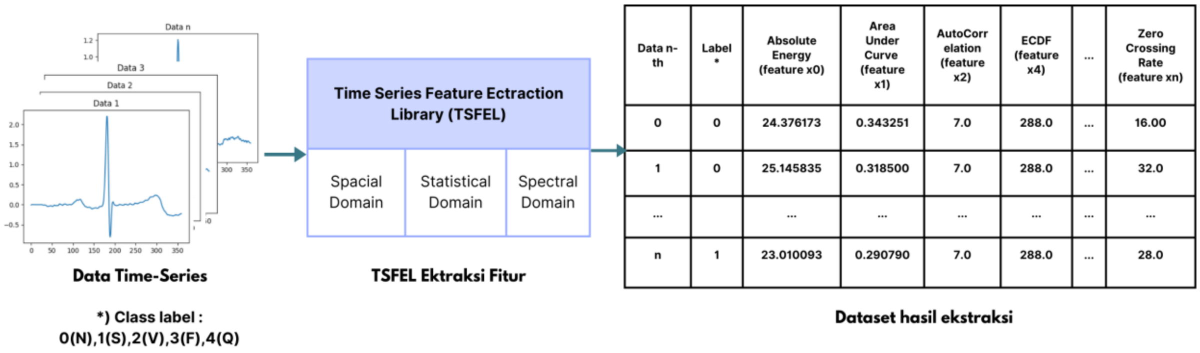


Figure 3. Feature Extraction Using TSFEL Process

follows: 90,320 (N), 7,229 (VEB), 2,781 (SVEB), 802 (F), and 15 (Q). Sample data of each class signal can be illustrated in Fig. 2.

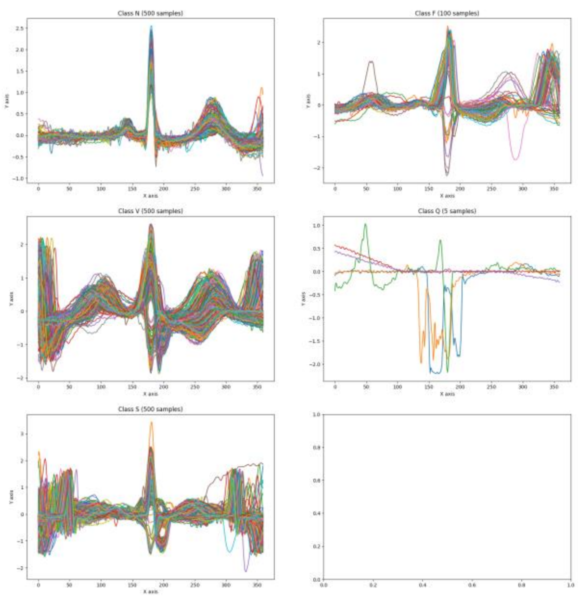


Figure 2. Example of Signal for Each Class

Feature Extraction

The segmented signal is a single data point that represents a specific class. This data inherently contains information that can be extracted and turned into features to be trained in a learning model approach. In this study, the Time Series Feature Extraction Library (TSFEL) is used to extract features from the signal data. TSFEL can handle multidimensional time series data, with available features divided into three domains: statistical, temporal, and spectral [22]. The total categories cover more than 60 features and based on its default parameters used for ECG signals in this study, it can extract 314 features from each data point. The illustration of the feature extraction process using TSFEL can be seen in Fig. 3. The list of features obtained after extraction using TSFEL can be seen in Table 2.

Table 2. Available Features in TSFEL

Domain	Feature Types
Spectral	FFT Mean coefficient Fundamental Frequency Human range energy LPCC (Linear Prediction Cepstral Coefficient) MFCC (Mel-Frequency Cepstral Coefficient) Max power spectrum Maximum frequency Median frequency Power bandwith Spectral centroid Spectral decrease Spectral distance Spectral entropy Spectral kurtosis Spectral positive turning points Spectral roll-off Spectral roll-on Spectral skewness
Statistical	Absolute energy Average power ECDF (Empirical Cumulative Distribution Function) ECDF Percentile ECDF Percentile Count Entropy Histogram Interquartile range Kurtosis Max Mean Mean absolute deviation Median Median absolute deviation Min Peak to peak distance Root mean square Skewness Standard deviation Variance

Domain	Feature Types
Temporal	Area under the curve Autocorrelation Centroid Mean absolute diff Mean diff Median absolute diff Median diff Negative turning points Neighbourhood peaks Positive turning points Signal distance Slope Sum absolute diff Zero crossing rate

Data Balancing

In the previously mentioned dataset distribution, it is evident that the data is imbalanced. Most of the data belongs to the normal class, whereas a good classification requires balanced data across the minority classes. Therefore, data balancing performed to make the dataset balanced [13]. The data balance technique used is a combined Random Over Sampling and Random Under Sampling processes. The result of this phase can be seen in Table 3.

Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is an algorithm used for classification and regression. XGBoost is an enhancement of the boosting algorithm, which combines several weak models to form a strong model.

Table 3. Data Distribution Before Balanced, After Random Over Sampling, and After Random Under Sampling

Class	Before Data Balancing	After Random Over Sampling	After Random Under Sampling
N	72255	144510	36127
SVEB	5783	11566	5783
VEB	2225	4450	2225
F	642	1284	642
Q	12	24	12

This algorithm develops an ensemble sequential number of trees [14].

In this case, the dateaset will be divided into training and testing data with a ratio of 3:1. The training data will be trained using XGBoost Classifier with the parameter gamma set to 0.1 as a way to control the complexity threshold for decision tree splits. The process is shown in Fig. 4.

Evaluation Matrix

Evaluation of the performance of a classifier is necessary to ensure the accuracy of its predictions. For this evaluation, a recommendation matrix from AAMI is used, which includes TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative) [19]. Such a matrix is needed because the classification performed is not binary but multiclass, so the confusion matrix cannot only encompass positive and negative classes. TP (True Positive) represents the condition when a positive instance is correctly predicted, while FP (False Positive) is the incorrect prediction of a negative instance as positive. On the other hand, a result is considered TN (True Negative) when a negative instance is correctly

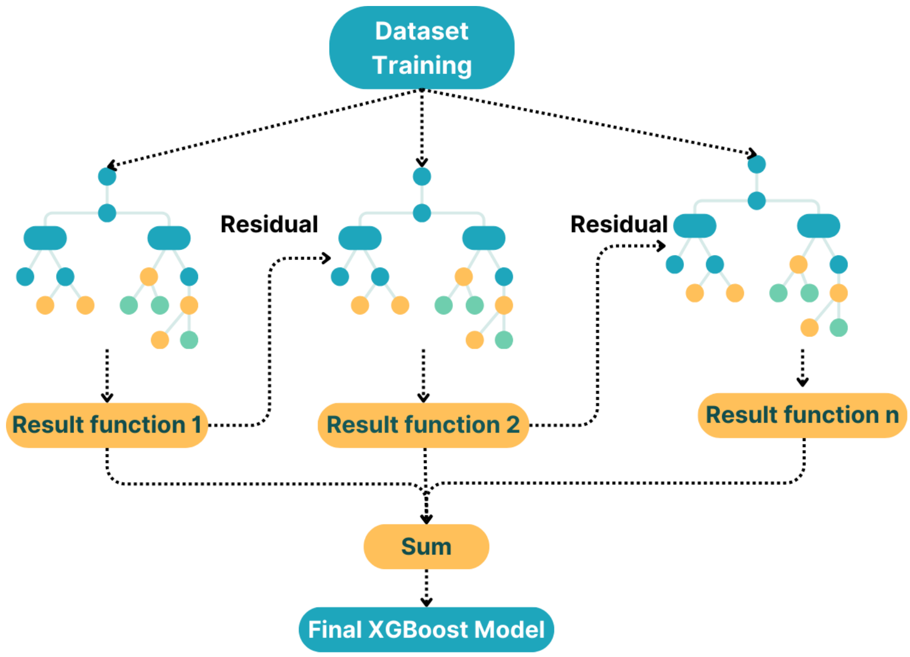


Figure 4. XGBoost Classifier Flow

predicted, whereas FN (False Negative) is the result when a positive instance is incorrectly predicted as negative.

These results are then used to calculate several performance metrics, namely:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}\ x\ 100\% \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN+FP}\ x\ 100\% \quad (3)$$

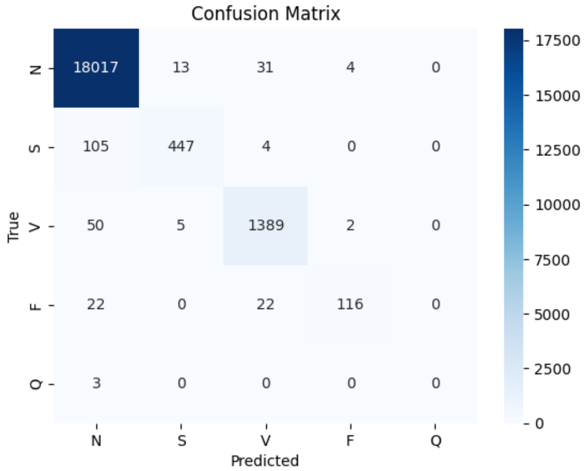
$$\text{F1-Score} = 2\ x\ \frac{Precision*Recall}{Precision+Recall}\ x\ 100\% \quad (4)$$

Formula 1 shows the calculation method for accuracy, which is the ratio of correct predictions to the total number of predictions. Accuracy itself gives a straightforward overview of the prediction results. However, it becomes problematic when dealing with an imbalanced dataset, as accuracy can yield misleading results in such cases. To address this issue, sensitivity is calculated to determine the true positive rate through Formula 2, and specificity is measured to gauge the true negative rate through Formula 3. Finally, to get an understanding of the balance between the true positive rate and the true negative rate, the F1-Score is calculated as shown in Formula 4. The results of these calculations will be used as a reference for comparison with previous research studies.

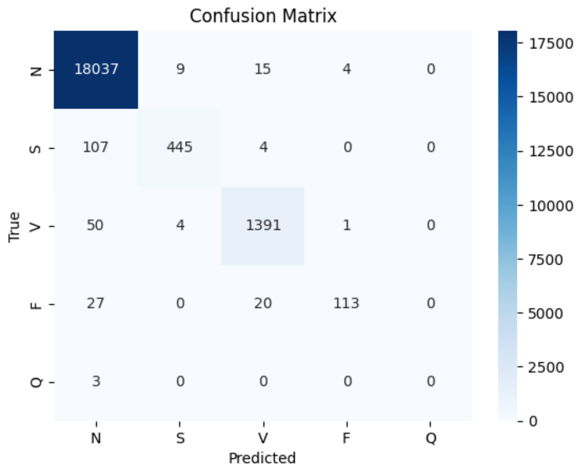
RESULT AND DISCUSSION

The model training was carried out using a dataset that had already undergone pre-processing, segmentation, feature extraction, and data balancing. The data used for training contains 314 features, and the distribution of each class can be seen in Table 2. In this study, experiments were also conducted to predict using a model trained on data that did not undergo the resampling process first.

After successfully training the model and testing it on the test dataset, the prediction results were summarized in a confusion matrix. The confusion matrix helps map the results that can be categorized as TP, TN, FP, or FN. The confusion matrix results for the prediction of 5 classes according to the AAMI recommendations can be seen in Fig. 5. Based on the obtained results, there were a total of 261 prediction errors out of 20,230 test data points in the prediction model using resampled data. In contrast, the prediction model using unresampled data achieved fewer prediction errors, with only 244. However, this cannot be the main reference in determining the performance of each model. Therefore, it is necessary to pay attention to the specific prediction results of each class as shown in Table 4 and Table 5.



(a)



(b)

Figure 5. Confusion Matrix for Resampled Data (a) and Unresampled Data (b)

Table 4. Evaluation Result for Resampled Data

Class	Acc%	Spe%	Sen%	F1%
N	98.87	91.69	99.73	99.37
SVEB	99.37	99.91	80.40	87.56
VEB	99.44	99.70	96.06	96.06
F	99.75	99.97	72.50	82.27
Q	99.99	100	0	0

Table 5. Evaluation Result for Unresampled Data

Class	Acc%	Spe%	Sen%	F1%
N	98.94	91.36	99.85	99.41
SVEB	99.39	99.93	80.04	87.87
VEB	99.54	99.79	96.20	96.73
F	99.74	99.98	70.63	81.36
Q	99.99	100	0	0

The results indicate that almost all evaluation outcomes show that the model with unresampled data is better compared to the model with resampled data. However, it is important to note that the classes that need to be considered are the SVEB, VEB, and F classes. In the medical field, sensitivity (recall) is an evaluation metric that requires more attention because it can affect patient treatment. The model with resampled data demonstrated better sensitivity values for the VEB and F classes. There was a significant increase, especially in the sensitivity of the F class, which increased by approximately 2%. This shows that the performance of the model with resampled data is better compared to the model with unresampled data in a medical context.

Table 6. Result Comparison

Model	Metric	N	S	V	F	Q
CNN + LSTM [13]	Acc%	97.95	99.16	**99.80**	99.25	99.48
	Sen%	-	-	-	-	-
	Spe%	-	-	-	-	-
AdaBoost + Random Forest [20]	Acc%	**99.24**	**99.58**	99.67	99.79	99.94
	Sen%	**99.95**	**82.61**	**97.45**	70.88	**99.24**
	Spe%	95.86	**99.99**	99.84	**100**	99.99
Ensemble RF & SVM [15]	Acc%	98.48	98.74	99.37	**99.84**	**99.99**
	Sen%	99.50	74.20	94.22	73.21	0
	Spe%	89.82	99.69	99.72	99.88	**100**
CNN + Focal loss [21]	Acc%	98.71	99.16	99.36	99.75	99.84
	Sen%	99.49	77.88	94.54	**82.10**	98.51
	Spe%	**97.60**	99.88	**99.85**	99.92	99.97
Proposed method	Acc%	98.87	99.37	99.44	99.75	**99.99**
	Sen%	99.73	80.40	96.06	72.50	0
	Spe%	91.69	99.91	99.70	99.97	**100**

From the comparison of results, it was found that the proposed method has a lower accuracy compared to existing methods. In Table 6, the bold numbers indicate the best accuracy for the corresponding matrix and class. The method using XGBoost as the classifier only excels in class Q, which cannot be considered a reliable reference due to the severe imbalance in the data, making the results for class Q inherently unreliable. This indicates the subpar performance of XGBoost as a classifier in achieving better accuracy compared to the methods used in previous studies. However, XGBoost does outperform certain methods in specific matrices and classes, but it is not the best when compared to all four other methods simultaneously. For example, the method using the XGBoost classifier outperforms the accuracy of each class when compared to CNN + Focal loss [21]. Another example is that XGBoost can also outperform most of the specificity values produced by the ensemble RF & SVM method, except for class F [15].

This research also has certain limitations that may have prevented the XGBoost algorithm from performing at its best. These limitations include the highly imbalanced dataset, which required sampling that may have compromised data integrity, the scope for further exploration of feature extraction techniques, and the use of hyperparameters to better optimize the XGBoost algorithm. To ensure the performance of XGBoost, further exploration is needed, especially regarding pre-processing, as it also plays a crucial role in determining the final evaluation results.

CONCLUSION AND SUGGESTION

The evaluation results of arrhythmia classification using the XGBoost classifier show that its performance is not quite satisfactory. This is evidenced by the comparison of performance with other models applied in previous research. Although the comparison shows that XGBoost can outperform certain models in specific matrices for certain classes, there are still other models that perform better when compared to XGBoost. Therefore, reconsideration is needed regarding the use of XGBoost for arrhythmia classification. There are certainly several areas for future work, such as trying this classifier on a different and larger dataset, as the current dataset is quite imbalanced. Additionally, ensemble learning could also be applied to improve the accuracy of the created classifier.

REFERENCES

[1] A. K. Sangaiah, M. Arumugam, and G. Bin Bian, "An intelligent learning approach for improving ECG signal classification and arrhythmia analysis," *Artif Intell Med*, vol. 103, Mar. 2020, doi: 10.1016/j.artmed.2019.101788.

[2] Shanthi. Mendis, P. Puska, Bo. Norrving, World Health Organization., World Heart Federation., and World Stroke Organization., *Global atlas on cardiovascular disease prevention and control*. World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization, 2011.

[3] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," 2020, doi: 10.1016/j.eswax.2020.10.

[4] C. Chen, Z. Hua, R. Zhang, G. Liu, and W. Wen, "Automated arrhythmia classification based on a combination network of CNN and LSTM," *Biomed Signal Process Control*, vol. 57, Mar. 2020, doi: 10.1016/j.bspc.2019.101819.

[5] M. Halomoan, "Epidemiologi Artimia," https://www.alomedika.com/penyakit/kardiologi/aritmia/epidemiologi.

[6] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, "Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network," *Information Fusion*, vol. 53, pp. 174–182, Jan. 2020, doi: 10.1016/j.inffus.2019.06.024.

[7] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Comput Biol Med*, vol. 102, pp. 278–287, Nov. 2018, doi: 10.1016/j.compbiomed.2018.06.002.

[8] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG Heartbeat Classification: A Deep Transferable Representation," Apr. 2018, doi: 10.1109/ICHI.2018.00092.

[9] D. K. Atal and M. Singh, "Arrhythmia Classification with ECG signals based on the

Optimization-Enabled Deep Convolutional Neural Network," *Comput Methods Programs Biomed*, vol. 196, Nov. 2020, doi: 10.1016/j.cmpb.2020.105607.

[10] J. Lemantara, "Penerapan Algoritma Naïve Bayes dan ID3 untuk Memprediksi Segmentasi Pelanggan pada Penjualan Mobil," *Journal of Technology and Informatics (JoTI)*, vol. 4, no. 1, pp. 31–40, Oct. 2022, doi: 10.37802/joti.v4i1.265.

[11] N. A. B. Ibrahim, "The Existence of Artificial Intelligence in the Future," *Journal of Technology and Informatics (JoTI)*, vol. 5, no. 1, pp. 25–33, Oct. 2023, doi: 10.37802/joti.v5i1.349.

[12] M. Guhdar, A. Ismail Melhum, and A. Luqman Ibrahim, "Optimizing Accuracy of Stroke Prediction Using Logistic Regression," *Journal of Technology and Informatics (JoTI)*, vol. 4, no. 2, pp. 41–47, Jan. 2023, doi: 10.37802/joti.v4i2.278.

[13] G. R. Patra, M. K. Naik, and M. N. Mohanty, "ECG Signal Classification Using a CNN-LSTM Hybrid Network," in *2023 2nd International Conference on Ambient Intelligence in Health Care, ICAIHC 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICAIHC59020.2023.10431449.

[14] R. Mogili and G. Narsimha, "Detection of Cardiac Arrhythmia from ECG Using CNN and XGBoost," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 2, pp. 414–425, Apr. 2022, doi: 10.22266/ijies2022.0430.38.

[15] S. Bhattacharyya, S. Majumder, P. Debnath, and M. Chanda, "Arrhythmic Heartbeat Classification Using Ensemble of Random Forest and Support Vector Machine Algorithm," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 3, pp. 260–268, Jun. 2021, doi: 10.1109/tai.2021.3083689.

[16] R. Kumar and Jyoti, "A Hybrid Approach Using SVM, kNN and Random Forest for ECG Classification," in *ISED 2023 - International Conference on Intelligent Systems and Embedded Design*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ISED59382.2023.10444570.

[17] "MIT-BIH Arrhythmia Database," http://circ.ahajournals.org/content/101/23/e215.full.

[18] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.

[19] "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," in *ANSI/AAMI EC57:2012/(R)2020; Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms*, AAMI, 2013. doi: 10.2345/9781570204784.ch1.

[20] R. Li *et al.*, "An Intelligent Heartbeat Classification System Based on Attributable Features with AdaBoost+Random Forest Algorithm," *J Healthc Eng*, vol. 2021, 2021, doi: 10.1155/2021/9913127.

[21] T. F. Romdhane, H. Alhichri, R. Ouni, and M. Atri, "Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss," *Comput Biol Med*, vol. 123, Aug. 2020, doi: 10.1016/j.compbiomed.2020.103866.

[22] M. Barandas et al., "TSFEL: Time series feature extraction library," SoftwareX, vol. 11, 2020, Art. no. 100456.