EVOLUTION IN ELECTRICAL AND ELECTRONIC ENGINEERING
e-ISSN: 2756-8458

EEEE

Vol. 5 No. 1 (2024) 102-106
https://publisher.uthm.edu.my/periodicals/index.php/eeee

Classification of Cardiac Arrhythmia Using Supervised Learning

Nur Tasni Atifah Norizu[1] , Nur Ilyani Ramli[2]*

[1] *Faculty of Electrical and Electronic Engineering,*
Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, MALAYSIA

*Corresponding Author: ilyani@uthm.edu.my
DOI: https://doi.org/10.30880/eeee.2024.05.01.013

Article Info

Received: 10 January 2024
Accepted: 22 February 2024
Available online: 30 April 2024

Keywords

Cardiac Arrhythmia, ECG, Supervised Learning, Feature Extraction

Abstract

Cardiovascular disease (CVD) is a leading cause of mortality, often involving cardiac arrhythmias characterized by abnormal electrical activity. The sinoatrial (SA) node serves as the heart's primary electrical impulse source. This work focuses on developing an algorithm to detect cardiac arrhythmia using morphological features extracted from Electrocardiogram (ECG) data. While ECG is a widely used non-invasive tool for cardiovascular diagnosis, it has limitations, particularly in detecting infrequent arrhythmias. The proposed supervised machine learning model aims to classify various cardiac arrhythmias, including paroxysmal atrial fibrillation and congestive heart failure, providing healthcare professionals with a valuable tool for accurate categorization and monitoring of cardiac conditions. The research involves data preparation with WEKA-based processing and feature extraction, utilizing K-Fold cross-validation for a dataset of 1200 instances and 47 attributes. Classification of Arrhythmia, Atrial Fibrillation (mostly Paroxysmal), Congestive Heart Failure, and Normal Sinus Rhythm is performed in WEKA using Naïve Bayes, Decision Tree, and k-Nearest Neighbors. All three classifiers exhibit high overall accuracy from 94% to 95.58%. Naïve Bayes slightly outperformed the others with 95.58% accuracy followed closely by J48 Decision Tree at 95.08%.

1. Introduction

Cardiac arrhythmia, a medical disorder characterized by irregular heartbeats or rhythm disturbances, occurs when the intricate electrical signals orchestrating the heartbeat are disrupted, leading to the heart beating too slowly, too quickly, or irregularly [1]. This condition transcends demographic boundaries, affecting individuals irrespective of age, gender, or overall health status. The symptoms associated with arrhythmia, including palpitations, chest discomfort, shortness of breath, dizziness, and fainting, underscore the diverse ways it can manifest.

Arrhythmias, ranging from benign to potentially fatal, may arise from various factors, including underlying heart disease, genetic predispositions, lifestyle choices, or adverse effects of certain medications. The complexity of causative factors underscores the importance of individualized and comprehensive approaches to diagnosis and treatment.

In terms of management, treatment strategies vary depending on the severity of the arrhythmia. Interventions may include pharmacological approaches, lifestyle modifications, and, in more severe cases, medical procedures

Evolution in Electrical and Electronic Engineering Vol. 5 No. 1 (2024) p. 102-106

such as the installation of pacemakers or cardioverter-defibrillators. These interventions aim not only to alleviate symptoms but also to mitigate the potential risks associated with untreated or poorly managed arrhythmias.

Recognizing the significance of early identification, precise diagnosis, and effective care, healthcare professionals play a pivotal role in enhancing an individual's quality of life while concurrently reducing the risk of complications and mortality associated with cardiac arrhythmias. A holistic and patient-centered approach, considering both medical and lifestyle factors, is integral to providing optimal care for those affected by this multifaceted cardiovascular condition.

Heart disease is a broad term that encompasses a variety of coronary conditions also referred to as cardiovascular disease. Cardiovascular disease is the primary cause of death in Malaysia, although it may not be the first thing that comes to mind when contemplating mortality causes. According to the Department of Statistics Malaysia, ischemic heart disease remained the leading cause of death in Malaysia, accounting for 17.0% of the 109,155 fatalities that were medically certified in 2020. The primary cause of cardiac arrhythmia is a disruption in the heart's normal rhythm, which leads to malfunctioning of the cardiovascular system. Due to cardiac dysfunction, electrical conduction across the heart is disrupted [1]. Numerous cardiac arrhythmias, including atrial fibrillation, atrial flutter, ventricular tachycardia, ventricular fibrillation, and supraventricular [2], are common among cardiovascular disease patients. Numerous arrhythmias are asymptomatic. When symptoms are present, palpitations or a cessation of heartbeats may occur. Serious symptoms may consist of vertigo, syncope, difficulty of breath, or chest pain. While most arrhythmias are not hazardous, some can place a person at risk for cardiac failure or a stroke. Others may result in cardiac arrest [3]. This research endeavors to achieve several objectives: firstly, to formulate a machine learning model aimed at classifying cardiac arrhythmias; secondly, to utilize supervised learning techniques to discern beats and rhythms characteristic of cardiac arrhythmias, including paroxysmal atrial fibrillation, congestive heart failure, and normal sinus rhythm; and finally, to assess the efficacy and accuracy of the developed machine learning model in classification tasks.

2. Materials and Method

Data preparation will involve WEKA-based pre-processing and feature extraction. K-Fold cross validation was implemented for model training. Datasets type is nominal with 1200 instances and 47 attributes. This work employs supervised learning in WEKA for the classification of Atrial Fibrillation, as outlined in Fig. 1. The methodology involves pre-processing and feature extraction of raw data, followed by dataset modeling using Training and Testing Split and K-Fold Cross Validation. Atrial Fibrillation classification employs Naïve Bayes, Decision Tree, and k-Nearest Neighbors algorithms. Evaluation includes assessing the performance of testing options and classification algorithms (Naive Bayes, Decision Tree, k-Nearest Neighbors) to optimize the model's accuracy and precision [4-10].

2.1 Feature Extraction and Data Processing

The process of feature extraction plays a crucial role in the effectiveness of machine learning algorithms, and in the context of this particular work, a meticulous extraction of 54 characteristics was undertaken. These features, a combination of identified and computed attributes, provide a comprehensive insight into the analyzed signals.

Among the extracted characteristics are durations, encompassing P, QRS, and T durations, as well as the duration of one cycle of the ECG signal. Additionally, features related to the areas and perimeters, such as the QRS area and perimeter, contribute to the comprehensive dataset. The inclusion of angles, capturing PQR, QRS, RST, PonPQ, and STToff, adds a nuanced understanding of the signal morphology.

Slope features, including PQ, QR, RS, and ST slopes, offer insights into the signal's inclinations. Interval features, spanning PQ, PT, QR, QT, RS, and ST segments, provide further dimensions for analysis. Statistical features, such as RR mean, PP mean, and ratios of QR to QS and RS to QS intervals, contribute to the dataset's statistical richness.

Heart rate variability features, including IBIM, SDRR, IBISD, NN50, Pnn50, SDSD, RMSSD, RRTot, NNTot, along with a feature representing Heartbeat per minute, add a temporal and frequency domain perspective to the dataset. These extracted features collectively contribute to the formation of a 1200x54 feature vector. In this vector, each of the 1200 signals is characterized by 54 distinct features, with 300 signals attributed to each of the four diseases. This feature vector serves as a robust foundation for subsequent machine learning models, facilitating a comprehensive analysis and classification of cardiac conditions based on the rich set of characteristics derived from the ECG signals. After extracting features from 1200 signals, we obtain a dataset with dimensions 1200x54. It contains diverse feature values, but some may be missing or NaN. To prepare for classification, scaling is crucial for consistent feature ranges. Handling missing values is also vital for dataset integrity. This meticulous preprocessing ensures a solid foundation for accurate and reliable classification models on the cardiac arrhythmia dataset.

Handling NaN values is crucial for dataset integrity. In this work, feature reduction is employed, removing columns with NaN values, resulting in a refined set of 47 features from the original 54. This streamlined dataset,

with dimensions 1200x47, is optimized for classification, enhancing the resilience and accuracy of machine learning models in identifying cardiac arrhythmias.

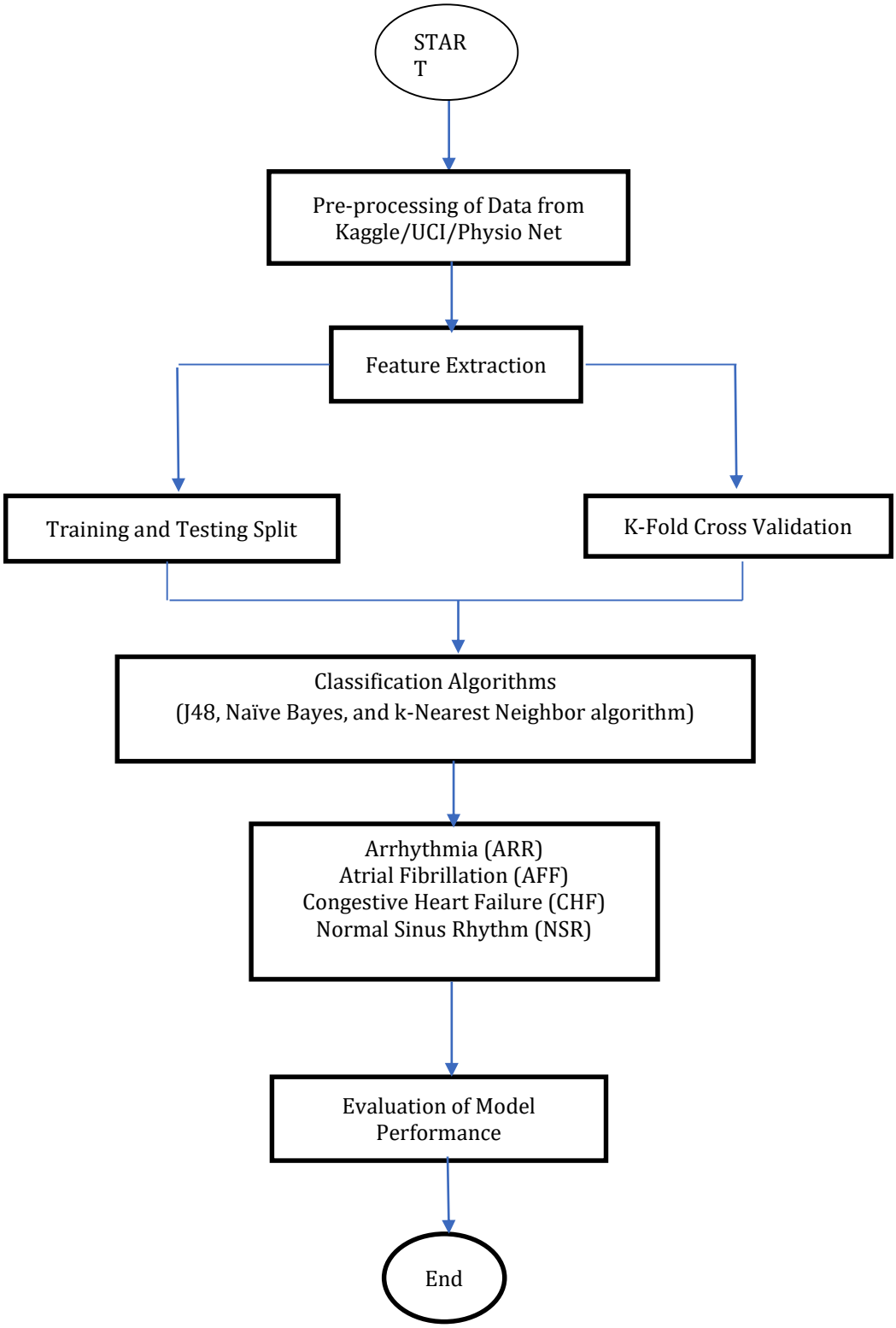

Fig.1 *Flowchart of the Methodology*

2.2 K-Fold Cross Validation

The dataset is divided into 10 subsets for k-fold cross-validation. Each fold involves distinct training and testing phases, ensuring comprehensive coverage. A decision tree model is constructed in the training phase using 9 subsets and evaluated in the testing phase. Metrics include class-specific accuracy and overall accuracy, providing

insights into the model's proficiency across all classes. This iterative process enhances the reliability of assessing the decision tree model's generalization capabilities, contributing to robust model development and validation.

2.3 WEKA

Weka serves as a comprehensive library of machine learning algorithms designed for various data mining tasks. This versatile toolkit encompasses a range of functionalities, including data preparation, classification, regression, clustering, association rules mining, and visualization tools. The name "Weka" is inspired by a flightless bird characterized by its inquisitive nature, native to the islands of New Zealand. Pronounced as "wek-uh," the software draws parallels with the distinctive sound of the bird. It's important to note that Weka is freely available as open-source software, distributed under the GNU General Public License, making it accessible for researchers, developers, and practitioners engaged in diverse machine learning and data mining endeavors [9].

3. Result and Discussion

The comparison of performance metrics across three classification algorithms, namely J48 Decision Tree, Naïve Bayes, and k-Nearest Neighbors (k-NN) Classifier as shown in Table 1, provides valuable insights into the strengths and weaknesses of each model in the context of cardiac arrhythmia classification.

All three classifiers exhibit high overall accuracy, ranging from 94% to 95.58%. This implies that the models are proficient in making correct predictions across all classes of cardiac arrhythmias. Naïve Bayes slightly outperforms the others with 95.58% accuracy, followed closely by J48 Decision Tree at 95.08%. The MCC, a measure of the quality of binary classifications, is consistently high across all models, indicating their effectiveness in capturing true and false positives and negatives. Naïve Bayes demonstrates the highest mean MCC at 94.2%, followed by J48 Decision Tree at 93.7%, and k-NN at 92.1%. Sensitivity, reflecting the ability to correctly identify positive instances, and precision, indicating the accuracy of positive predictions, are crucial metrics. Naïve Bayes and J48 Decision Tree show comparable overall sensitivity and precision, while k-NN lags slightly behind in precision. The weighted F1 score, which balances precision and recall, is consistently high for all models, ranging from 94.1% to 95.6%. This emphasizes the models' robustness in achieving a balance between precision and recall across multiple classes. The classification error, representing the proportion of misclassifications, is the lowest for Naïve Bayes at 4.42%, followed by J48 Decision Tree at 4.92%. k-NN exhibits a slightly higher classification error at 6%.

The results indicate that all three models perform remarkably well in classifying cardiac arrhythmias, with minimal variations in overall accuracy. Naïve Bayes excels in terms of mean MCC, demonstrating its strength in capturing both positive and negative instances effectively. J48 Decision Tree showcases high precision, especially in differentiating normal sinus rhythm signals. The choice between these models may depend on specific requirements and considerations, such as interpretability, computational efficiency, and the importance of false positives and false negatives in the clinical context. The slightly higher classification error for k-NN suggests a marginally higher rate of misclassifications compared to the other models.

Table 1 *Comparison of Performance Metrics*

Overall Performance Metrics	J48 Decision Tree	Naïve Bayes	k-Nearest Neighbors
Accuracy of Model	95.08%	95.58%	94%
Mean of MCC	93.7%	94.2%	92.1%
Sensitivity	95.1%	95.6%	94%
Precision	98.3%	95.8%	94.2%
F1 Score	95.1%	95.6%	94.1%
Classification Error	4.92%	4.42%	6%

4. Conclusion

The development and evaluation of machine learning models for the classification of cardiac arrhythmias, including beats and rhythms-based arrhythmias, paroxysmal atrial fibrillation, congestive heart failure, and normal sinus rhythm, have been implemented. The comparison of three classification algorithms—J48 Decision Tree, Naïve Bayes, and k-Nearest Neighbors (k-NN) Classifier—reveals their high overall accuracy, ranging from 94% to 95.58%. This indicates the proficiency of these models in making accurate predictions across various classes of cardiac arrhythmias.

Particularly, Naïve Bayes emerges as a top performer with a 95.58% accuracy and the highest mean Matthews Correlation Coefficient (MCC) at 94.2%. The MCC consistently reflects the effectiveness of all models in capturing

true and false positives and negatives. J48 Decision Tree demonstrates high precision, especially in differentiating normal sinus rhythm signals, while k-NN lags slightly behind in precision.

The weighted F1 score, balancing precision and recall, remains consistently high for all models, emphasizing their robustness in achieving a balance between these two metrics across multiple classes. Additionally, the classification error, representing the proportion of misclassifications, is impressively low for Naïve Bayes and J48 Decision Tree, with k-NN exhibiting a slightly higher rate. In summary, each model, whether Naïve Bayes, J48 Decision Tree, or k-NN, presents a strong candidate for cardiac arrhythmia classification. The choice among these models should consider specific requirements and considerations such as interpretability, computational efficiency, and the importance of false positives and false negatives in the clinical context. The overall high performance of all three models suggests their potential applicability in real-world healthcare scenarios, contributing to more accurate and efficient identification of cardiac arrhythmias for improved patient care.

Acknowledgement

The authors would like to thank the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

References

[1] Rahul, J., & Sharma, L. D. (2022). Automatic cardiac arrhythmia classification based on hybrid 1-D CNN and Bi-LSTM model. *Biocybernetics and Biomedical Engineering*, *42*(1), 312–324. https://doi.org/10.1016/j.bbe.2022.02.006

[2] Sinha, N., & Das, A. (2020). Automatic diagnosis of cardiac arrhythmias based on three stage feature fusion and classification model using DWT. *Biomedical Signal Processing and Control*, *62*, 102066. https://doi.org/10.1016/j.bspc.2020.102066

[3] Sowmya, S., & Jose, D. (2022). Contemplate on ECG signals and classification of arrhythmia signals using CNN-LSTM deep learning model. *Measurement: Sensors*, *24*, 100558. https://doi.org/10.1016/j.measen.2022.100558

[4] Mohonta, S. C., Motin, M. A., & Kumar, D. K. (2022). Electrocardiogram based arrhythmia classification using wavelet transform with deep learning model. *Sensing and Bio-Sensing Research*, *37*, 100502. https://doi.org/10.1016/j.sbsr.2022.100502

[5] Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, *51*(5), 675–687. https://doi.org/10.1016/j.beth.2020.05.002

[6] Miric, M., Jia, N., & Huang, K. G. (2022). Using Supervised Machine Learning to Create Categorical Variables for Use in Management Research: The Case for Identifying Artificial Intelligence Patents. *Strategic Management Journal*. https://doi.org/10.1002/smj.3441

[7] Bansal, S. (2019, April 17). *Supervised and Unsupervised learning - GeeksforGeeks*. GeeksforGeeks. https://www.geeksforgeeks.org/supervised-unsupervised-learning/

[8] Merlini, D., & Rossini, M. (2021). Text categorization with WEKA: A survey. *Machine Learning with Applications*, *4*, 100033. https://doi.org/10.1016/j.mlwa.2021.100033

[9] *Weka – Graphical User Interference Way to Learn Machine Learning*. (n.d.). Analytics Vidhya. Retrieved January 18, 2021, from https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/weka-gui-learn-machine-learning/#:~:text=Weka%20is%20a%20collection%20of

[10] Rahul, J., Sora, M., Sharma, L. D., & Bohat, V. K. (2021). An improved cardiac arrhythmia classification using an RR interval-based approach. *Biocybernetics and Biomedical Engineering*, *41*(2), 656–666. https://doi.org/10.1016/j.bbe.2021.04.004