Classification of Arrhythmia Potential using the K-Nearest Neighbor Algorithm

Darren Farell, Aliecia, Elvira Valencia Lee, Sojamo Lala Hulu, Abdi Dharma, Arjon Turnip, and Mardi Turnip

***Abstract*—According to the World Health Organization in 2021, cardiovascular diseases caused around 17.9 million deaths worldwide, making them the leading cause of global mortality. This data indicates that deaths due to heart disease remain very high, largely due to a lack of tools and technology for early detection of heart conditions. In this study, the K-Nearest Neighbor (KNN) machine learning algorithm is used with Electrocardiogram (ECG) signal data. The study was conducted on 30 subjects, where the heart activity data of each subject was recorded using an ECG device. The collected data is classified into four categories: high potential for arrhythmia, potential for arrhythmia, normal, and abnormal. The implementation of the KNN algorithm resulted in an accuracy rate of 93%. The high accuracy of the KNN algorithm is expected to make a significant contribution to the early detection of cardiovascular diseases.**

***Index Terms*— Arrhythmia, Electrocardiography, Nearest Neighbor method.**

I. Introduction

ARRHYTHMIA is a heart rhythm disorder that refers to abnormalities in the frequency, regulation, and origin of electrical impulses that control the heartbeat. Common symptoms of arrhythmia include an irregular heartbeat, which may be too fast (tachycardia) or too slow (bradycardia), as well as additional symptoms such as fatigue, chest pain, and fainting [1]. Heart disease was the leading cause of death worldwide in 2021. According to the World Health Organization (WHO), approximately 17.9 million people globally died due to heart and cardiovascular diseases [2].

Statistics from the Indonesian Ministry of Health show that as many as 650,000 people died from heart disease in 2023 [3]. This high mortality rate highlights the urgency to understand and manage arrhythmias appropriately. Increasing knowledge about risk factors, early detection, and management of arrhythmias is a priority to understand this. Comprehensive measures in prevention and treatment are needed to reduce undesirable outcomes.

This high mortality rate can be caused by various factors. The main risk factors include unhealthy diet, such as consumption of foods high in saturated fat and salt, and low intake of fruits and vegetables. Lack of physical activity and poor lifestyle are other factors that increase the risk of heart disease. Smoking habits and excessive alcohol consumption are also significant contributors. Genetic factors and a family history of heart disease can increase a person's risk of developing this disease. Medical conditions such as hypertension, diabetes, and obesity also increase the likelihood of fatal heart complications. Lack of access to adequate health services for heart disease prevention worsens this situation. In addition, limited counseling and education about heart disease prevention methods, especially in areas with limited health infrastructure, result in an increased risk and prevalence of heart disease in the community.

Heart disease examination is carried out by analyzing the morphology or dynamics of the ECG [4-5]. The recorded signal, called an ECG signal, plays an important role in the diagnosis of heart disease and long-term monitoring. An electrocardiogram is a test that records the electrical signals produced when the heart pumps blood and is used to monitor the heart [6]. Most arrhythmia detectors use ECG signals consisting of P waves, QRS complexes and T waves [7]. Online ECG signal monitoring can be done in real time and using previous test participant data [8]. Manual ECG analysis requires experienced medical personnel, more time and higher effort [4]. Innovations in medical technology have made ECG monitoring more efficient and accessible. Remote monitoring can allow expert analysis of ECG signals with a high degree of accuracy.

Recent advances in electronics and data transmission infrastructure have resulted in several devices capable of monitoring human health with the help of wireless sensors. These devices offer the potential for continuous cardiac monitoring and real-time detection of Arrhythmias [9].

Scan Watch (SW) is one of the tools used to detect Arrhythmia. SW allows recording of only one ECG, which is sufficient to screen or diagnose rhythm and conduction disorders. However, SW is unable to detect Ischemic and Cardiomyopathy diseases [10]. Smart Watch can be used to take measurements physiological or such as heart rate, blood pressure, andoxygen saturation, up to the number of steps, minutes of activity, heart rate variability, and intrathoracic impedance. The data obtained will be sent digitally to the health management platform by Narrow Band Internet of Things. By combining big data obtained with artificial

D. Farell, Aliecia, E. V. Lee, A. Dharma, and M. Turnip are with Faculty of Technology and Computer Science, University of Prima Indonesia, Medan, Indonesia (e-mails: darrenfarrell09@gmail.com, alieciachai@gmail.com, elviraalee16@gmail.com, sojahulu08921@gmail.com, abdidharma@unprimdn.ac.id, marditurnip@unprimdn.ac.id*).

A. Turnip is with Department of Electrical Engineering, Universitas Padjadjaran, Indonesia (e-mail: turnip@unpad.ac.id).

intelligence (AI) techniques then the health level of the individual concerned can be predicted [11-12]. Another portable ECG is the Detector of Fetal Abnormalities Technology (DOFAT). DOFAT captures heart signal recordings with electrodes, then the results of the data processing and filtering will be sent to a smartphone via Bluetooth signal as a fetal health monitor in pregnant women [13].

Machine learning techniques such as Artificial Neural Networks have been integrated for cardiac data analysis. Artificial Neural Networks allow more complex modeling of cardiac data to detect patterns that are invisible to traditional algorithms. Deep Learning algorithms can be used to detect Arrhythmia and implemented on wearable devices. The data commonly used is data sourced from the MIT-BIH dataset, but the use of data is considered to still require quite complicated preprocessing [14]. The MIT-BIH arrhythmia database yields a multi-layer perceptron model that achieves the highest performance, robust performance while providing critical intuition [15]. The algorithm that can be used is the Deep Neural Network (DNN) algorithm with the best accuracy validation [16]. Another similar algorithm is the Support Vector Machine (SVM) algorithm, with a high level of accuracy [17]. The application of another algorithm is the Naive Bayes classification algorithm using 24 training data and 12 test data with an average computing time of 3.35 milliseconds [18]. However, the accuracy value of the research with the Naive Bayes classification algorithm is less valid because the amount of data used is too small.

Our deep learning-based algorithm based on ECG data only accurately predicts the occurrence of MVA in PLN operators [19]. Dynamic ML model and variational autoencoder provide personalized prediction of malignant ventricular arrhythmias in patients with ICD [20]. The combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) Transformer helps to capture various attributes of ECG signals [21]. Training rhythm models and sequential heart models using 61,852 ECG recordings with a macro average F1 score of 0.807 [22]. Kernelized SVM model in LM algorithm with 80.21% accuracy in predicting cardiovascular [23]. The utility of CAT-Net algorithm with high accuracy and excellent performance [24]. Replacing the cardiac electromechanical FOM with an ANN-based ROM causes the numerical simulation to fail in computational time, due to the very large computational cost, if performed with the FOM [25]. AI and machine learning can accurately detect simple arrhythmias such as sinus bradycardia, also providing valuable data sets and free software applications [26-27].

Another classification algorithm that can be used is the KNN algorithm. The KNN algorithm is one of the Supervised Machine Learning algorithms that is useful for classification and regression using several datasets [28]. The KNN algorithm is widely used for pattern recognition in various disciplines, such as face recognition, text classification, graph analysis, disease prediction, anomaly detection, and time series forecasting [29]. The classification process in this algorithm is based on the reference of all stored data. The KNN algorithm evaluates the similarity between new data points and existing data [30]. Testing the accuracy of the KNN algorithm with the confusion matrix algorithm produces an accuracy rate of 90.5% [31]. This high accuracy indicates the effectiveness of KNN in performing classification tasks [32]. Therefore, this algorithm is adopted in this study as one of the relevant methods.

The application of the KNN algorithm is expected to increase efficiency in the diagnosis of heart disease and enable faster and more precise treatment. In this study, the KNN algorithm is used to detect heart disorders. The validation process will be carried out to assess the reliability and accuracy of the KNN algorithm in identifying heart disorders in the available dataset. This is expected to lead to improved diagnosis and better management of heart disease. Another reason for using KNN in the diagnosis of heart disease is because the algorithm can handle large data with high efficiency. KNN is also effective in identifying complex patterns in medical data. In addition, KNN can work well on unstructured data. These advantages make KNN an important tool in the medical field for the diagnosis of heart disease.

II. Method

This research was conducted at Universitas Prima Indonesia, Medan in collaboration with Universitas Padjdjaran, Bandung, Indonesia. This research used a portable 3 Lead ECG machine which is useful as a compact mobile device to record the electrical activity of the heart over a period of time [33]. Portable EKG devices can be used by anyone, including those who do not have heart problems. In addition to being practical, having this device can also save on examination costs [34]. 3 Lead ECG uses 3 electrodes so it is superior because the machine is lighter. This type of ECG can also help in monitoring heart activity in places that require mobility such as outside the hospital [35]. The AD8232 sensor and ESP32 sensor are on the 3 Lead ECG to capture signals and process them. These sensors are connected using cables which will then produce signals in the form of 3 main heart rhythm waves, namely the P wave, the QRS complex, and the T wave. The data obtained is stored in an SD card before being processed. The data will be processed via Raspberry Pi so that the data can be read on the Web server.

This study involved 30 people aged 20-25 years in different conditions. The ECG components were attached to the ECG machine before being turned on. The chest, ankles, and hands of the research subjects were cleaned with alcohol cotton before the study was carried out. The three ECG electrodes were then given jelly and attached to the research subjects. The red electrode was attached to the right upper body, the yellow electrode was attached to the left upper body and the green electrode was attached to the left lower body. The research subjects were then asked to do 3 activities for 9 minutes, namely sitting for 3 minutes, walking for 3 minutes, and running for 3 minutes. During the activity, the ECG recorded and recorded their heart activity. The electrical activity of the heart will be captured through the AD8232 sensor and the Nodemcu ESP32 sensor. The AD8232 sensor is useful for increasing the accuracy level of PR and QT interval signals because ECG signals can have a lot of noise [36]. The Nodemcu ESP32 sensor reads the ECG signal which will then be converted into BPM (Beats Per Minute) [37]. Each electrode attached to the subject's chest will have direct contact. In the activity carried out by the subject, the subject is asked to breathe normally and then the data will be recorded using an ECG device. The data used include 5 distances between waves and 3 main waves, namely the P wave, QRS complex, and T wave.

The P wave in the ECG represents atrial depolarization, the electrical activity that causes the atria of the heart to contract. The P wave is usually the first indication of the cardiac cycle and appears as a small wave that precedes the QRS complex. Normally, the P wave has a duration of less than 0.12 seconds, reflecting the time required for atrial depolarization. The amplitude of the P wave is usually no more than 2.5 mm, reflecting the electrical voltage generated during the atrial depolarization process. The QRS complex in the ECG represents ventricular depolarization, the electrical activity that causes the ventricles of the heart to contract to pump blood out into the pulmonary artery and aorta. Normally, the duration of the QRS complex is between 0.06 and 0.10 seconds. A longer duration may indicate conduction blockade or ventricular hypertrophy. Therefore, analysis of the QRS complex is very important in evaluating the electrical condition of the heart and can help in the diagnosis of various heart disorders such as arrhythmias, myocardial infarction, and cardiomyopathy conditions. The T wave in the ECG represents ventricular repolarization, the process of restoring the ventricles of the heart after contraction so that they are ready for the next cardiac cycle. Normally, the T wave has a variable amplitude depending on the ECG lead used, but is usually no more than 5 mm in the limb leads and 10 mm in the precordial leads.

The data obtained from the sensor will then be processed using the Matlab application. Matlab plays a role in validating research results with various methods whose results can be clearly visualized. With the Matlab application, data will be processed and classified using the KNN algorithm. The use of KNN is implemented to evaluate the possibility of arrhythmia based on experimental data from participants. The data processing stage involves normalizing numeric attributes, removing missing values, and dividing the dataset into training and testing subsets. Normalization is done to ensure that attributes with different value ranges have a balanced impact on the distance calculation in the KNN algorithm. After the data is prepared, the KNN algorithm is applied for classification. Testing uses various K parameter values and measures model performance using evaluation metrics such as accuracy, precision, recall, and F1-score. Evaluation is carried out using cross-validation techniques to avoid overfitting and ensure model generalization.

KNN is a Supervised Machine Learning algorithm used in classification and regression problems. KNN classifies unlabeled data by calculating the distance between each unlabeled data point and all other points in the data set. Then, assigning each unlabeled data point to the most identically labeled data class by finding patterns in the data set. [38]. The straight-line distance formula (Euclidean distance) is the most common algorithm for finding distance in KNN. The distance between each test sample x and training data point X where x_i, = (x_1, x_2, x_3,…,x_n) and X_i = (X_1, X_2, X_3,…,X_n) can be calculated using the Euclidean Distance Equation.

$$d(x_i - X_i) = \sqrt{\sum_{i=1}^{n}(x_i - X_i)}\,. \tag{1}$$

In addition to Euclidean Distance, other distance metrics can also be used, such as Manhattan Distance which calculates the total absolute difference between the coordinates of data points:

$$d(x_i, X_i) = \sum_{i=1}^{n} \mid x_i - X_i \mid . \tag{2}$$

Or using the Minkowski Distance, which is a generalization of the Euclidean and Manhattan Distance:

$$d(x_i, X_i) = \left(\sum_{i=1}^{n} \mid x_i - X_i \mid^{x}\right)^{\frac{1}{x}}. \tag{3}$$

Classification is done based on KNN with the smallest distance, where K is the number of nearest neighbors involved in the most processes. Classification is done based on KNN with the smallest distance, where K is the number of nearest neighbors involved in the most processes. The class label assigned to the test sample is classified based on the majority vote (the class with maximum members in K) from KNN. The KNN algorithm has many features that make it a strong competitor to other classification algorithms. These include simplicity, understandability, and robustness to noisy data x_i=2 x_i=1 [39]. The main wave signals are obtained from 3 electrodes attached to the patient's chest during various activities such as sitting, walking, and running.

The ECG cable is designed for cardiac monitoring so that it can help in diagnosing heart problems. The AD8232 sensor is used to collect and measure changes in heart rate when sitting, walking, standing. The collected signals will then be input to the ESP32 for processing. The ESP32 device has demonstrated its ability to consistently sample sensors at high speeds with excellent stability [40]. In this study, the ESP32 sensor can improve the accuracy of the noise signal received by the ECG with the help of a micro SD card as additional storage. Raspberry Pi is a small and affordable computer that is increasingly popular in the Internet of Things (IoT) and other domains [41]. The system on the raspberry pi will read the ECG signal accurately through the AD8232 signal sensor that has received and recorded. The collected signal will be processed through the AD8232 sensor and the ESP32 sensor which can increase the accuracy of the noise signal received by the ECG. The data is then processed using the KNN algorithm and will be classified into 4 categories, namely Abnormal, normal, potential Arrhythmia and very potential Arrhythmia. The classified data will be sent to a database or cloud system via the internet so that it can be accessed and read by users as in Fig. 1.

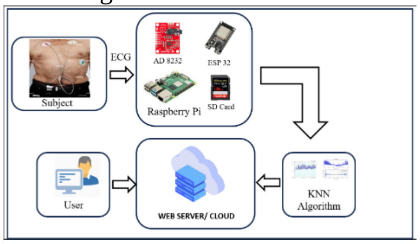

Fig. 1. ECG System Process

III. Results and Discussions

Early identification of potential heart abnormalities using ECG data and the KNN algorithm has great potential in the medical field. This approach can greatly assist in detecting heart disease earlier, allowing for faster and more effective medical intervention. ECG data, which records the electrical activity of the heart, is a widely used diagnostic tool. By using the KNN algorithm to analyze ECG data, decisions can be made more quickly and objectively compared to traditional manual algorithms that often rely on subjective judgment. This approach can help reduce errors associated with visual observation by medical professionals, resulting in more consistent and accurate diagnoses. To achieve optimal results, the KNN model needs to be optimized through proper feature selection, parameter tuning, and thorough model validation. The development of this model requires a deep understanding of ECG theory and signal processing. In this study, 30 normal subjects participated in three experimental scenarios: sitting, walking, and running. The ECG data from each subject produced waveform structures with different characteristics. The amplitude (wave height) in the ECG data reflects the depolarization and repolarization of the myocardium, which is used by healthcare professionals to analyze heart rhythms, identify potential Arrhythmias, and evaluate overall heart health.

The use of KNN was implemented to evaluate the likelihood of arrhythmia based on experimental data from participants. The data processing stage involved normalizing numeric attributes, removing missing values, and dividing the dataset into training and testing subsets. Normalization was performed to ensure that attributes with different ranges of values had a balanced impact on the distance calculation in the KNN algorithm. After the data was prepared, the KNN algorithm was applied for classification. Testing was performed using various values of the K parameter. Model performance was measured using evaluation metrics such as accuracy, precision, recall, and F1-score. Evaluation was performed using cross-validation techniques to avoid overfitting. This ensures that the model has good generalization capabilities.

The average data taken are organized and visually displayed in groups based on the experimental scenario, divided by sitting (in blue), walking (in orange), and running (in gray) in Fig. 2. In Fig. 2, the RR graph shows the interval between two consecutive R waves, called the RR interval, typically ranging from 500 to 1000 ms. Three lines depict the RR values corresponding to each activity: the blue line represents sitting, the orange line represents walking, and the gray line represents running. Among the three, running displays the most significant fluctuations in RR values, with peaks exceeding 1,000 at various points, indicating that running leads to greater variability in the RR intervals or respiratory rate. Walking shows moderate variations, with its RR values generally remaining lower than those seen during running. In contrast, the data for sitting is more stable, with consistently lower RR values, reflecting less variation during this activity.

The graph suggests that higher-intensity activities, such as running, lead to more variability in RR values, whereas sitting results in the most stable values. Such information can be valuable for studies that analyses heart rate variability (HRV) or changes in respiratory rate across different levels of physical activity.

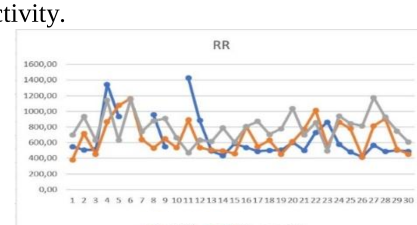

Fig. 2. Recorded RR wave graph with sitting (blue), walking (orange), and running (gray) colors.

Other graph such as the PR period graph shows the time between the beginning of the P wave (marking the beginning of atrial depolarization) and the beginning of the QRS complex (marking the beginning of ventricular depolarization), typically ranging from 120 to 200 ms. The QS graph depicts the simultaneous activation of the right and left ventricles and contraction of the large ventricular muscle, with the duration of the QS typically ranging from 0.06 to 0.10 seconds. The QT graph shows the time from the beginning of the Q wave to the end of the T wave, reflecting the duration from the beginning of ventricular contraction to the end of relaxation. The normal range for the QT interval is usually less than 0.44 seconds or 440 milliseconds. The ST graph depicts the segment shift between the previous T wave and the current P wave (the previous TP segment), which is used as a reference where the TP segment is not isoelectric. Finally, the HR (Heart Rate) graph shows the number of heartbeat per minute with a normal heart rate ranging from 60 - 100 BPM. HR calculations use the RR interval formula.

The results of the graph above are obtained by entering the values of each wave into the KNN algorithm. Classification in the KNN algorithm uses the nearest neighbor of the k value to predict the resulting value. The benchmark limit values are divided into 4 groups, namely abnormal (<1618), normal (1618 - 2579), potential for Arrhythmia (2579 - 3541), and very potential for Arrhythmia (>3541). The division of value limits is based on the range of test data used. The classification process is carried out based on KNN with the smallest distance, where K is the number of nearest neighbors involved in the process as much as possible. The class label assigned to the test sample is classified based on the majority vote (the class with the most members in K) from KNN. To achieve a high level of accuracy, it is important to perform class labels with adequate division of the data set for training and testing.

The scatter plot provides a visual representation of the relationship between two key variables, RR (the R-R interval) and ST, along with their associated values and classifications related to heart health. The horizontal axis, representing RR values, ranges from approximately 400 to 2200 milliseconds, indicating the variability in heart rate intervals; lower values signify a faster heart rate, while higher values indicate a slower heart rate. For instance, an RR interval of around 400 ms suggests a very high heart rate, whereas an RR of approximately 2200 ms indicates a much slower heart rate. The vertical axis shows ST values ranging from about 50 to 450, likely representing the electrical activity of the heart as recorded in an electrocardiogram (ECG), with higher ST values suggesting greater cardiac workload or strain.

In terms of data distribution, the blue dots, representing "Normal" readings, are concentrated in the lower left region of the graph, typically showing RR values between 400 and 800 and ST values between 50 and 200. This indicates that individuals with normal heart function tend to have lower RR intervals and lower ST values. The orange dots, which signify "Potential Arrhythmia," are more spread out and overlap with both the normal and abnormal categories, suggesting that these individuals may have some risk factors for arrhythmias, with RR values ranging from about 600 to 1200 and ST values typically between 100 and 300. The red dots, indicating "Abnormal" readings, appear more in the mid-range of RR (around 800 to 1600) and ST values (150 to 350), highlighting significant abnormalities in heart function. Meanwhile, the sparse purple dots (Highly Potential Arrhythmia) are found at higher ST values (around 300 to 450) with RR values ranging from 1000 to 1600, signifying a higher likelihood of serious arrhythmias or heart dysfunction.

The numerical values on both axes enable an understanding of how RR intervals correlate with ST levels. As RR values increase, indicating slower heart rates, ST values also tend to rise, suggesting a relationship between slower heart rates and higher electrical activity levels. The spread and distribution of the points reflect variability among individuals, where those with RR values below 800 typically exhibit lower ST levels, while those with RR values above 1200 show higher ST levels, indicating potential cardiac issues.

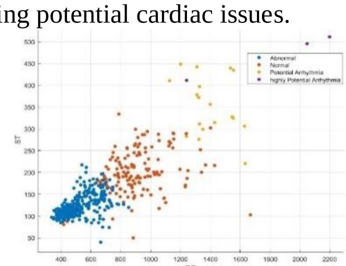

Fig. 3. Scatter Plot RR-ST

The confusion matrix is a powerful tool used to evaluate the performance of a classification model by summarizing the model's predictions on test data in a table format. It consists of four key components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positives refer to cases where the model correctly predicts the positive class, while True Negatives refer to correct predictions of the negative class. False Positives occur when the model incorrectly predicts the positive class, and False Negatives occur when the model incorrectly predicts the negative class. These metrics provide the foundation for calculating various performance measures, such as accuracy, precision, recall, and F1-score, offering deep insights into the model's reliability and effectiveness.

The confusion matrix is use to pinpoint areas where a model may need improvement, such as bias toward specific classes or a tendency to make particular types of errors. For example, in the matrix, different colors like dark blue (indicating good performance) and pink (representing errors) enhance the visualization and make it easier to understand the model's performance. A well-performing model, such as the Fine KNN model depicted in the confusion matrix, may show a strong classification of 281 observations as Abnormal (dark blue) while misclassifying only 4 as Normal (pink box). The confusion matrix can also display results in terms of percentages, providing values such as the Positive Predictive Value (PPV) and False Discovery Rate (FDR). For instance, the Abnormal class might show a PPV of 97.6%, indicating that 97.6% of Abnormal predictions are correct, while the FDR is 2.4%, meaning 2.4% of the Abnormal predictions are wrong. Additionally, metrics like True Positive Rate (TPR) and False Negative Rate (FNR) are represented, offering insights into how well the model detects specific classes. For example, the Abnormal class might have a TPR of 98.6% and an FNR of 1.4%, indicating that the model is excellent at detecting Abnormal cases.

However, the model struggles with the Potential Arrhythmia and Highly Potential Arrhythmia classes. For Potential Arrhythmia, 84.2% of cases are accurately identified, but 10.5% of cases are wrongly predicted, which indicates some overlap in features between this class and others. The model's performance deteriorates significantly for Highly Potential Arrhythmia, where only 33.3% of cases are correctly classified, and a staggering 66.7% are misclassified. This suggests that the model has difficulty distinguishing between this class and others, likely due to data complexity or insufficient differentiation between features. The confusion matrix also uses color-coding—dark blue representing high accuracy and pink representing errors—to help visualize the model's performance.

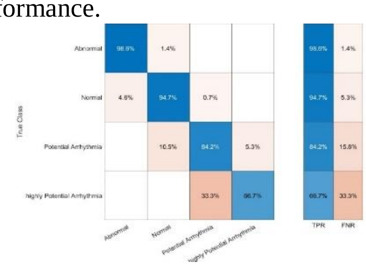

Fig. 4. TPR (True Positive Rate) and FNR (False Negative Rate)

ROC (Receiver Operating Characteristic) is a graph used to evaluate the performance of a classification model by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various classification thresholds. True Positive Rate, also known as sensitivity or recall, measures the proportion of positive cases that are actually detected by the model, while False Positive Rate measures the proportion of negative cases that are incorrectly classified as positive. The ROC curve allows for analysis of the trade-off between sensitivity and specificity (1 - FPR) at various thresholds, providing an idea of the model's ability to distinguish between positive and negative classes. A specific point on the curve, with coordinates X = 0.004545 and Y = 0.8421, shows that the model has a very low FPR (approximately 0.45%), meaning it rarely misclassifies negative cases as positive. Simultaneously, it achieves a TPR of 84.21%, correctly identifying 84.21% of positive cases (potential arrhythmia). The curve also includes a point (0.00, 0.84), demonstrating that even with a highly conservative approach (no false positives), the model still maintains a TPR of 84%.

The Area Under the Curve (AUC) is 0.92, reflecting the overall effectiveness of the model. An AUC of 1 represents perfect classification, while 0.5 is equivalent to random guessing. With an AUC of 0.92, the model shows strong performance in distinguishing between positive and negative cases. Both the ROC curve and the AUC are closely linked to the confusion matrix, which breaks down the model's TP, FP, TN, and FN values. These, in turn, determine the TPR and FPR values plotted on the ROC curve. In conclusion, the combined insights from the confusion matrix and the ROC curve demonstrate that the model effectively balances sensitivity and specificity, offering high reliability in detecting arrhythmia with minimal errors.

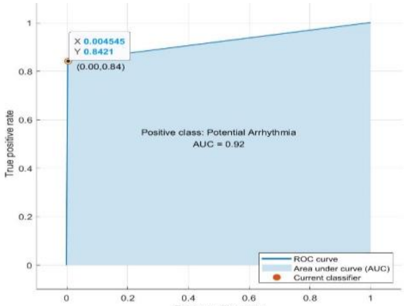

Fig. 5. Results of data processing using the ROC curve showing Average Positive Data “Potential for Arrhythmia”

IV. Conclusion

The results of the experiment demonstrate that the KNN model is highly effective in classifying potential arrhythmias based on data gathered from research subjects, particularly when utilizing a three-lead ECG system. The evaluation metrics reveal an impressive average accuracy exceeding 93% for the dataset employed, indicating that the KNN algorithm is not only robust but also reliable in detecting arrhythmias.

This high level of accuracy underscores the potential of the KNN algorithm to serve as an essential tool in clinical settings. Its ability to predict the risk of arrhythmia from clinical data is significant, offering medical practitioners an advanced method for early diagnosis and proactive treatment planning for patients. Early detection of arrhythmias is crucial, as timely intervention can significantly reduce the risk of severe cardiovascular events, including stroke or sudden cardiac arrest.

Moreover, the KNN model's performance in identifying positive cases of arrhythmia suggests its applicability in routine clinical practice, where continuous monitoring of patients' heart rhythms is essential. The algorithm's efficacy in pinpointing underlying risk factors associated with arrhythmias also holds substantial clinical value. By highlighting these risk factors, the KNN model can guide healthcare providers in implementing targeted preventive strategies tailored to individual patients, thereby enhancing patient outcomes.

Additionally, the integration of this algorithm with existing ECG monitoring technologies can lead to a more comprehensive and efficient approach to cardiovascular health. This could facilitate the development of predictive analytics platforms that not only identify arrhythmias but also provide real-time insights into patients' heart health, enabling clinicians to make informed decisions swiftly.

The application of the KNN algorithm in analyzing three-lead ECG data for arrhythmia detection represents a promising advancement in cardiovascular medicine. With its high accuracy and capability to uncover crucial clinical insights, this approach could significantly impact patient care by fostering early diagnosis, optimizing treatment strategies, and ultimately improving overall patient prognosis. Further research and development are warranted to refine this methodology and explore its integration into clinical workflows, paving the way for enhanced monitoring and management of arrhythmias in diverse patient populations.

Declaration of Competing Interest

This research is fully supported by the Faculty of Science and Technology, Universitas Prima Indonesia. Deepest appreciation to Universitas Prima Indonesia, and also to all friends (senior and junior) who participated in our research.

Acknowledgement

This research was supported by Faculty of Science and Technology Universitas prima Indonesia, Medan, Indonesia and Department of Electrical Engineering, Padjadjaran University, Bandung, Indonesia.

References

[1] Q. Ayuni, R.C. Wihandika, N. Yudistira, “Classification of Arrhythmia from Electrocardiogram Results Using Support Vector Machine Method,” *Journal of Information Technology and Computer Science Development*, vol. v, no. 6, pp. 2163-2170, 2021.

[2] Rokom. 2022. *Penyakit Jantung Penyebab Utama Kematian, Kemenkes Perkuat Layanan Primer* https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20220929/0541166/penyakit-jantung-pembebab-utama-matitian-kemenkes-perkuat- jasa-primer/. Accessed 30 October 2023.

[3] Humas FKUB. 2023. World Heart Day 2023: Use Heart Know Heart. https://prasetya.ub.ac.id/world-heart-day-2023-use-heart-know-heart/#:~:text=WHO%20menybutkan%20that%20every%20years,650.%20000 %20population%20per%20years Accessed 30 November 2023.

[4] D. N. Fauzi, Y. N. Fuadah, and I. Safitri, “Supraventricular Arrhythmia Classification Based on ECG Signals Using Convolutional Neural Network,” *e-Proceeding of Engineering*, vol. 8, no. 6, pp. 3255-3262, 2022.

[5] A. Turnip et al., “Classification of ECG signals using decision trees and linear discriminant analysis,” *in Proc. 2024 IEEE Int. Conf. Artif. Intell. Mechatron. Syst. (AIMS)*, Bandung, Indonesia, 2024, pp. 1–6, doi: 10.1109/AIMS61812.2024.10512398.

[6] S. Kumar, A. Mallik, A. Kumar, J. D. Ser, and G. Yang, “Fuzz-ClustNet: Coupled fuzzy clustering and deep neural networks for arrhythmia detection from ECG signals,” *Computers in Biology and Medicine*, vol. 153, 2023.

[7] C. Wijaya, Andrian, M. Harahap, C. Natalis, M. Turnip, and A. Turnip, “Abnormalities State Detection from P-Wave, QRS Complex, and T-wave in Noisy ECG,” *Journal of Physics*, 2019.

[8] M. Turnip, R. I. E. Saragih, A. Dharma, D. E. Kusumandari, A. Turnip, D. Sitanggang, and S. Aisyah, “Extraction of ECG signal with adaptive filter for health abnormalities detection,” *Journal of Physics*, no. 1007, 2018.

[9] Y. J. Nurriski and Alamsyah, “Optimization of Deep Convolutional Neural Network (Deep CNN) for Arrhythmia Detection Through ECG Signals Using Conv1D Architecture,” *Indonesian Journal of Mathematics and Natural Sciences*, vol. 46, no. 1, pp. 10-20, 2023.

[10] S. Touiti, I. Medarhri, K. Marzouki, N. Ngote, and A. Tazi-Mezalek, “Feasibility and reliability of Whiting ScanWatch to record 4 lead Electrocardiogram: A comparative analysis with a standard ECG,” *Heliyon*, 2023.

[11] X. Y. Chen, B. X. Yang, S. A. Zhao, J. Ding, P. Sun, and L. Gan, “Intelligent health management based on analysis of big data collected by wearable smart watch,” *Cognitive Robotics*, vol. 3, pp. 1-7, 2023.

[12] G. J. Williams, A. Al-Baraikan, F. E. Rademakers, F. Cirabegna, F. N. V. D. Vosse, and A. Lawrie, “Wearable technology and the cardiovascular system: the future of patient assessment,” *The Lancet Digital Health*, vol. 5, no. 7, pp. 467-476, 2023.

[13] C. N. Ginting, I. N. E. Lister, M. Silaen, E. Girsang, Y. Laila, M. Turnip, and A. Turnip, “Fetal heart detection based on wide area network technology with wireless sensor transmission,” *Journal of Physics: Conference Series*, 2019.

[14] R. L. B. Buana and I. Hudati, “Review: Analysis of Arrhythmia Detection Features and Deep Learning Methods for Wearable Devices,” *National Journal of Electrical Engineering and Information Technology*, vol. 11, no. 1, pp. 61-69, 2022.

[15] E. P. Moghaddam, D., Muguli, A., Razavi, M., & Aazhang, B., “A graph-based cardiac arrhythmia classification methodology using one-lead ECG recordings,” Intelligent Systems with Applications, p. 22, 2024.

[16] M. Z. Nugraha, T. S. P. H. Fauzi, and R. Magdalena, “Arrhythmia Detection Using Deep Neural Network (DNN) Algorithm on Electrocardiogram Signals,” *e-Proceeding of Engineering*, vol. 10, no. 5, pp. 4155-4158, 2023.

[17] A. Turnip, M. I. Rizqwan, D. E. Kusumandari, M. Turnip, and P. Sihombing, “Classification of ECG signals with Support Vector Machine for Arrhythmia Detection,” *Journal of Physics: Conference Series*, vol. 970, 2018.

[18] R. L. Mampioper, R. Maulana, and B. H. Prasetio, “Implementation of Premature Atrial Contraction (PAC) Detection System using Naive Bayes Classifier Method,” *Journal of Information Technology and Computer Science Development*, vol. 6, no. 7, pp. 3291-3300, 2022.

[19] R. R. van de Leur, R. de Brouwer, H. Bleijendaal, T. E. Verstraelen, B. Mahmoud, A. Perez-Matos, J. Dickhoff, J. Schoonderwoerd, L. Germans, N. Houweling, A. van der Zwaag, P. Cox, P. van Tintelen, A. Te Riele, M. van den Berg, P. Wilde, A. Doevendans, T. de Boer, and E. van Es, “ECG-only explainable deep learning algorithm predicts the risk for malignant ventricular arrhythmia in phospholamban cardiomyopathy,” *Heart Rhythm*, vol. 21, no. 7, pp. 1102-1112, 2024.

[20] M. Z. H. Kolk, S. Ruipérez-Campillo, L. Alvarez-Florez, B. Deb, E. J. Bekkers, C. P. Allaart, A.-L. C. J. Van Der Lingen, P. Clopton, I. Išgum, A. A. M. Wilde, R. E. Knops, and S. M. Narayan, “Dynamic prediction of malignant ventricular arrhythmias using neural networks in patients with an implantable cardioverter-defibrillator,” *eBioMedicine*, vol. 99, p. 104937, 2024.

[21] S. Din, M. K. Qaraqe, O. Mourad, K. A. Qaraqe, and E. Serpedin, “ECG-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial-temporal and long-range dependency features,” *Artificial Intelligence in Medicine*, vol. 150, p. 102818, 2024.

[22] X. Zhang, H. Wu, W. Zhang, and L. Zhu, “Automatic diagnosis of arrhythmia with electrocardiogram using multiple instance learning: From rhythm annotation to heartbeat prediction,” *Artificial Intelligence in Medicine*, vol. 133, p. 102379, 2022.

[23] R. G. S. B. and K. V. Sangeetha, “Cat-net: Convolution, attention, and transformer based network for single-lead ECG arrhythmia classification,” *Biomedical Signal Processing and Control*, vol. 105, p. 104032, 2024.

[24] M. R. Qaraqe and S. E. Islam, “Cat-net: Convolution, attention, and transformer based network for single-lead ECG arrhythmia classification,” *Biomedical Signal Processing and Control*, vol. 105, p. 93, 2024.

[25] F. Regazzoni, M. Salvador, L. Dede, and A. Quarteroni, “A machine learning method for real-time numerical simulations of cardiac electromechanics,” *Computer Methods in Applied Mechanics and Engineering*, vol. 393, 2022, Art. no. 111083.

[26] A.A. Munoz-Macho, M.J. Dominguez-Morales, and J.L. Sevillano-Ramos, “Analyzing ECG signals in professional football players using machine learning techniques,” *Biomed. Signal Process. Control*, vol. XX, pp. 1-7, 2024.

[27] S. J. Lase, A. Trisanto, N. S. Syafei, N. L. Hoa, and A. Turnip, “Identification and classification of arrhythmias on portable ECG using ANN method,” *in Proc. 2022 IEEE Int. Conf. Sustain. Eng. Creative Comput. (ICSECC)*, Kabupaten Bekasi, Indonesia, 2022, pp. 73–77, doi: 10.1109/ICSECC56055.2022.10331533.

[28] S. Adhikary and S. Banerjee, “Introduction to Distributed Nearest Hash: On Further Optimizing Cloud Based Distributed kNN Variant,” *Procedia Computer Science*, no. 218, pp. 1571-1580, 2023.

[29] A. X. Wang, S. S. Chukova, and B. P. Nguyen, “Ensemble k-nearest neighbors based on centroid displacement,” *Information Sciences*, vol. 629, pp. 313-323, 2023.

[30] S. Saifullah, D. B. Prasetyo, Indahyani, R. Dreżewski, and F. A. Dwiyanto, “Palm Oil Maturity Classification Using K-Nearest Neighbors Based on RGB and Lab Color Extraction,” *Procedia Computer Science*, vol. 225, pp. 3011-3020, 2023.

[31] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, “Implementation of K-Nearest Neighbor (KNN) Classification Algorithm for Scholarship Recipient Selection Classification,” *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, pp. 118-127, 2021.

[32] A. Turnip, J. G. Hamonangan, G. F. Yohanes, P. Turnip, E. Sitompul, and N. Le Hoa, “Detection of drug effects on brain-based EEG signals using K-nearest neighbours,” *in Proc. 2023 IEEE 9th Inf. Technol. Int. Seminar (ITIS)*, Batu Malang, Indonesia, 2023, pp. 1–5, doi: 10.1109/ITIS59651.2023.10420084.

[33] S. C. Mohonta, M. A. Motin, and D. K. Kumar, “Electrocardiogram based arrhythmia classification using wavelet transform with deep learning model,” *Sensing and Bio-Sensing Research*, vol. 37, 2022.

[34] Fadhili Rizal Makarim. Elektrokardiogram (EKG) - Gejala, Penyebab, dan Pengobatan," *Halodoc*. https://www.halodoc.com/kesehatan/elektrokardiogram-ekg?srsltid=AfmBOorxI6UpQqLJQ3IFLKX_az7ugA7owfTpcD_il4rgRstOUgoSFBlF. Accessed: 30 October 2023.

[35] “EKG: Menganalisis Aktivitas Jantung,” *HaloSehat*. https://www.halosehat.com. Accessed: 30 October 2023.

[36] R. Hariri, L. Hakim, and R. F. Lestari, “Heart Rate Monitoring System Using Ad8232 Sensor,” *Zetroem*, vol. 2, 2020.

[37] D. F. Nugroho, R. Munardi, and I. H. Santoso, “Internet of Things Based Heart Rate Condition Monitoring System Using EKG Sensor With Android Application Media,” *e-Proceeding of Engineering*, vol. 8, p. 5530, 2021.

[38] S. Ray, “A quick review of machine learning algorithms,” in *2019 International Conference on Machine Learning, Big Data, Cloud, and Parallel Computing (COMITCon)*, pp. 35-39, 2019.

[39] A. Almomany, W. R. Ayyad, and A. Jarrah, “Optimized implementation of an improved KNN classification algorithm using Intel FPGA platform: Covid-19 case study,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, pp. 3815-3827, 2022.

[40] D.-L.-R.J. Espinosa-Gavira, J.M. Muñoz-Cruz, J.A. Sánchez-Ceja, “Characterization and performance evaluation of ESP32 for real-time synchronized sensor networks,” *Procedia Computer Science*, vol. 237, pp. 261-268, 2024.

[41] K.N.K. Mathe, S. M. Dhole, “A comprehensive review on applications of Raspberry Pi,” *Computer Science Review*, vol. 56, p. 100636, 2024.