Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 4, August 2025, Page. 2281-2296
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

Comparison of LightGBM With XGBoost Algorithms in Determining Arrhythmia Classification in Students

Delima Sitanggang*[1], Eddrick Wilbert Solo[2], Ferdy Immanuel Sinaga[3], Stefanus Jorgi L.Tobing[4], Feliks Daniel Hutasoit[5], Agung Prabowo[6]

[1,2,3,4,5,6]Faculty of Science and Technology, Universitas Prima Indonesia, Medan, Indonesia

Email: [1]delimasitanggang@unprimdn.ac.id

Received : Jul 2, 2025; Revised : Jul 25, 2025; Accepted : Jul 28, 2025; Published : Aug 19, 2025

Abstract

Arrhythmia is a heart rhythm disorder that may occur unpredictably with life-threatening risk if it were not treated immediately. This heart disorder generally affects the elderly, but symptoms of this disorder can also arise in children and adolescents, especially for those with heart problems or are often under stress. The implementation of this research is aimed at analyzing the symptoms of early arrhythmia in adolescent children using electrocardiogram signals. In order to obtain the best possible results in determining the higher performing algorithm, two machine learning methods were used to predict the classification of arrhythmia which will be compared for their accuracy. The subjects of this study included 106 students from SMK Swasta Teladan Sumatera Utara 2 located in the city of Medan, of which 72 final subject data were used to train the capability of both models used to predict arrhythmia classification categorized into four categories, namely normal, abnormal, potential of arrhythmia, and high potential of arrhythmia. The LightGBM model outperformed the XGBoost model, with 95.11% accuracy and 95.03% F1 Score, and although the loss value of the LightGBM model is higher than the loss value of the XGBoost model, the difference between these two values is negligible and the loss value of LightGBM can be considered as excellent with a value of 0.1503. This research contributes to the advancement of digital health by demonstrating the potential of machine learning-based ECG analysis for highly accurate early arrhythmia detection in adolescent, non-clinical populations.

Keywords : *Arrhythmia, Electrocardiogram, LightGBM, Machine Learning, XGBoost*

1. INTRODUCTION

Arrhythmia is a heart rhythm disorder that can occur unexpectedly to anyone and poses a fatal risk if left untreated swiftly and appropriately [1]. The rate of incidence of arrhythmia tends to rise as people age [2]. Recent cohort studies have shown that the prevalence of atrial fibrillation (AF), a common cardiac arrhythmia, increases with age: among individuals aged 65-69 years, the prevalence of AF is about 6.4%, increasing to 10.3% at 70-74 years, 15.1% at 75-79 years, 22.4% at 80-84 years, and reaching 28.5% in those aged 85 years or older. [3]. Although many cases of arrhythmia are commonly found in adults and the elderly, children and adolescents are also susceptible to this risk, especially for those with a certain cardiac history or frequent physical and mental stress [4], [5].

One of the main ways to detect arrhythmias is by recording and analyzing ECG (electrocardiogram) signals, which is a non-invasive measurement of the heart's electrical activity that serves as the standard in the diagnosis of heart rhythm abnormalities [6]. ECG is a standard diagnostic tool for detecting heart rhythm abnormalities and may provide important information such as time intervals between waves and wave patterns [7]. However, manual analysis of ECG signals is

 Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

complicated, time-consuming, and requires the involvement of trained medical personnel [8]. Therefore, the use of Machine Learning (ML) technology has become a promising approach to automate the ECG signal classification, with recent systematic reviews reporting classification accuracies exceeding 90% across various tasks, including arrhythmia detection [9].

ML is a branch of artificial intelligence that enables computers to learn from existing data and recognize patterns without being explicitly programmed [10]. Two ML algorithms, XGBoost and LightGBM, are examples of the most widely used and high-performing gradient boosting methods [11][12]. Gradient boosting is an ensemble technique that combines predictions from many decision trees to improve model accuracy [13]. In general, XGBoost excels in terms of accuracy and overfitting controllability through strong regularization techniques, and has full-featured support for customization, meanwhile LightGBM was developed to improve the efficiency of the training process by using a leaf-wise approach in tree formation, making it faster and memory-efficient on large datasets with many features [15]. However, the leaf-wise approach of LightGBM can lead to overfitting on small or highly imbalanced datasets [16].

The performance of these two algorithms has been observed to vary depending on specific data characteristics, such as the number of features, imbalance of classes, and the size of the dataset [17], [18]. In the context of complex and sometimes noisy student ECG data, it is unclear as to which algorithm performs better for the classification of arrhythmias [19], [20]. Therefore, this study examines the performance comparison of XGBoost and LightGBM based on evaluation metrics such as accuracy, F1 score, and log loss on student ECG data. Previous research has shown that XGBoost has good generalization ability on various medical classification problems, while LightGBM excels in training time especially on large datasets [21][22]. However, studies regarding the application of these two algorithms specifically for arrhythmia classification in student populations remain scarce [23].

This research offers a novel approach in addressing a gap in arrhythmia detection studies, which typically rely on clinical datasets or data from older populations such as adults and geriatrics, by focusing on ECG data collected from a younger population rarely researched as the subject of studies in arrhythmia detection. In terms of machine learning model development, in addition to using standard evaluation metrics such as accuracy, precision, recall, and F1 score, this study also features the use of log loss to better assess the reliability of model predictions.

This research aims to provide a contribution to the development of an efficient and precise early detection system of arrhythmia using the optimal ML method for student ECG data by comparing the performance of XGBoost and LightGBM on multi-class and imbalanced ECG data from a non-clinical adolescent population, an area that remains underexplored in existing research. With the latest advances in signal processing technology and ML algorithms, it is hoped that the result of this research may assist medical personnel in quicker and more accurate diagnosis, while improving access to preventive health services among adolescents.

2. METHOD

This research aims to compare two methods with different algorithms, namely XGBoost (Extreme Gradient Boosting) and LightGBM (Light Gradient Boosting Machine). XGBoost and LightGBM are decision tree-based Machine Learning algorithms. Both of these methods are part of the Gradient Boosting Decision Tree (GBDT) method which is an ensemble technique that forms a set of weak models to gradually build strong models [24], [25].

Gradient Boosting, which is an iterative process, works in several stages. First, the initial model will be initialized as a constant model. Next, a weak model will be formed and trained to predict the difference (residual) between the actual value of the training data label and the predicted value will be calculated. This model will be combined with the previous model with certain weights controlled by the

Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

learning rate parameter. This process is repeated incrementally until certain criteria are met, such as the final number of iterations is reached or the loss function value is sufficiently low.

XGBoost and LightGBM implement similar approaches albeit with different strategies for optimization. XGBoost makes use of pruning techniques, L1/L2 regularization, and full parallelization support to improve accuracy and reduce overfitting [24], [25]. Meanwhile, LightGBM implements Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) techniques to improve efficiency on large and high-dimensional datasets [25].

Generally, both algorithms optimize an objective function consisting of two main components, namely loss function and regularization function. The general formula of the objective function in the gradient boosting method can be written as shown in (1).

$$L(t) = \sum_{i=1}^{n} l(yi,\ \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (1)$$

Where $l(y_i, \hat{y}_i)$ is the loss function (e.g. log-loss), $\hat{y}_i^{(t-1)}$ is the model prediction at iteration (t-1), $f_t(x_i)$ is the new tree model at iteration t, and $\Omega(f_t)$ is the regularization function for model complexity.

The flow of this research is presented as a block diagram in Figure 1, starting with the collection of ECG data from multiple students. The collected data then were preprocessed with multiple filters and its features extracted. After going through the process of normalization, the data is split into training and testing dataset with Stratified Cross Validation method, which were then used to train and test the XGBoost and LightGBM model. Confusion matrix was used to visualize the prediction results, from which the evaluation parameters such as Accuracy and F-1 Scores can be determined. These parameters could further be used to compare the performance of the models using visual tools such as bar graphs.

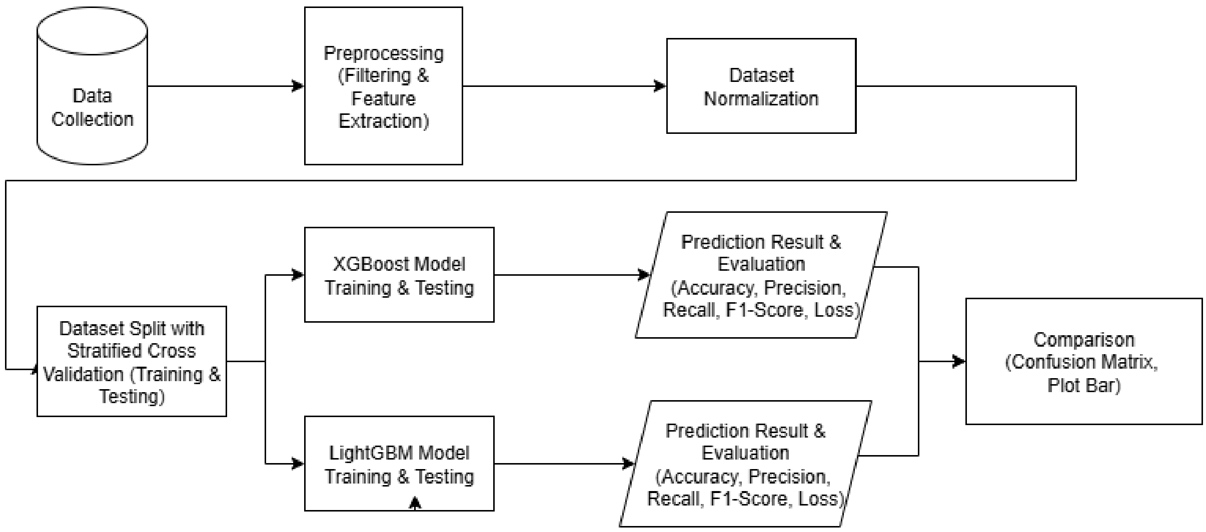

Figure 1. Block Diagram Representing the Flow of Research

2.1. Materials

The research was conducted at SMK Swasta Teladan Sumatera Utara 2, Jl. Pendidikan No.62, Cinta Damai, Kec. Medan Helvetia, Medan City, North Sumatera in partnership with UNPRI. Collaborating parties involved in this research involved the research participants, as well as the field supervisors. The tools and materials used in this study were in the form of a 5-lead ECG Smart Holter tool totaling 3 units based on Raspberry Pi 4, electrode cables, and laptops used in data recording and the use of data recording programs. Respondents in this study consisted of 106 student subjects aged between 15 and 20 years. A total of 101 male students and 5 female students were involved.

Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

This experiment was conducted under the condition of the subject sitting for approximately two minutes. Subjects were then requested to remove their upper clothing to facilitate the application of the electrodes on the necessary areas of the subject's body, as shown in Figure 2. During data collection, the researcher conducted interviews with the subjects to obtain additional data regarding the subject's name, age, height, weight, and medical history.

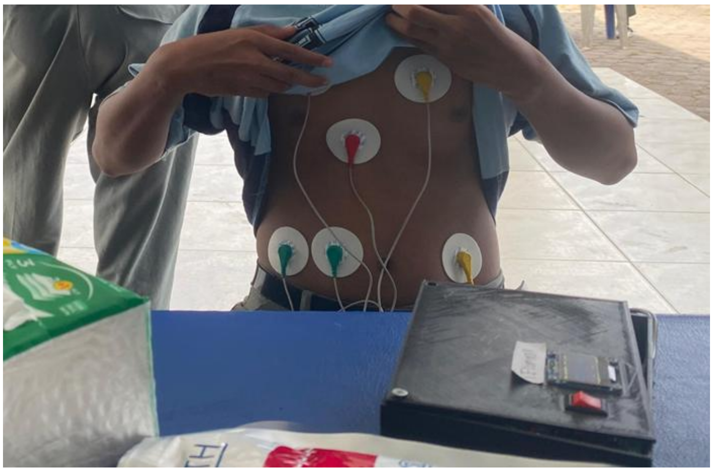

Figure 2. Recording using Smart Holter ECG 5 lead device

2.2. Data Collection

Data collection was performed by running a Serial Data Acquisition program that reads and parses the raw ECG signals from the Smart Holter 5-lead ECG system with 6 electrodes, which are then transmitted via the serial port of the ECG hardware as depicted in Figure 3. At Jack 1, the red electrode is attached to the upper right chest (RA), the yellow electrode to the upper left chest (LA), and the green electrode to the lower right body (RL). Whereas on Jack 2, the red electrode is placed at the sternal area leaning to the right (V1), the yellow electrode at the bottom of the left (LL), and the green electrode remains at the bottom of the right (RL). The Smart Holter ECG machine works by recording electrical signal data from the subject’s torso, which is then converted into integer values, in this case with the assumption that the signal has a range of 2.4 Volts and 24-bit resolution.

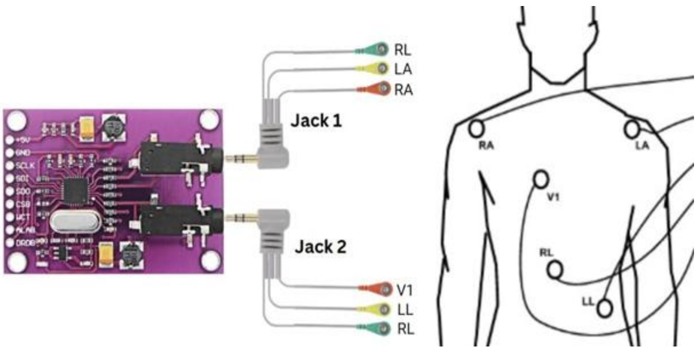

Figure 3. Smart Hoster ECG 5 Lead applied at specific points on the torso.

2.3. Data Preprocessing

Before model building and training can begin, the initial dataset will be preprocessed and its features extracted to produce ECG data suitable for preprocessing by utilizing a custom made Python-based pipeline implemented in Google Colab designed for automatic processing of large-scale ECG data, implementing libraries such as NeuroKit2, pandas, NumPy, and SciPy for signal processing and feature extraction.

Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

Raw signals collected from the subjects were first organized and converted in the form of a compressed ZIP file containing the raw ECG data for each subject. The data then went through several filtering processes to remove noise and signal drift. Once the ECG signal is cleaned and filtered, the P, Q, R, S, and T peaks will be identified which will then be used to extract key features such as heart rate and duration of various intervals. The features that have been obtained will be arranged in the form of numerical data with a fixed dimension of 7, namely RR Interval, PR Segment, QRS Duration, QT Interval, ST Segment, R/S Ratio, heart rate. This ensures that the ECG data is consistent and ready to be used to analyze heart function and rhythm patterns.

The training of machine learning models was conducted using a supervised learning approach, which refers to a machine learning technique that requires external assistance in the form of output labels which serve as objective values [10]. These output labels are used during model training to learn the relationship between output targets and features. In the testing data, these labels provide a reference for the model in evaluating the accuracy of the prediction.

Feature normalization was then applied using the StandardScaler utility from scikit-learn library. This method standardizes the input features by subtracting the mean and scaling them to unit variance. Normalization is performed to ensure that all features are on a similar scale, allowing each one to contribute equally during the training process. This step also helps improve the learning efficiency and performance of algorithms that are sensitive to feature scaling, such as XGBoost and LightGBM.

In order to address the class and gender imbalance, the Stratified Cross-Validation method is implemented using the StratifiedKFold utility from scikit-learn library. K-Fold Cross Validation works by dividing the data into several equally sized folds, where each fold takes turns being the validation set while the rest are used for training. However, in classification tasks with imbalanced data, this approach may result in unequal class distributions across the folds. To keep the class proportion balanced in each fold, stratification was applied to ensure that each fold possessed approximately the same class distribution as the original dataset.

2.4. Model Building

The model used in this research was built in the Visual Studio Code software which supports the Python programming language, the language used by the LightGBM and XGBoost models. The main libraries used to support the design of this model are libraries from LightGBM and XGBoost, Seaborn for Confusion Matrix visualization, Matplotlib.pyplot for bar and line graph comparison visualization, Pandas for reading dataset files so that they can be used by the model, Numpy for mathematical operations, and Sklearn for cross-validation using StratifiedKFold, normalizing features, and model evaluation such as accuracy, precision, recall, F1-Score and log loss.

The two models XGBoost and LightGBM were built using the XGBClassifier and LGBMClassifier classes respectively, with hyperparameters for each model customized for the general boosting settings which were fine-tuned with multiple testing in order to achieve the highest accuracy and efficiency. Both models use similar parameters to maintain training consistency, such as objective to specify the type of multi-class classification problem, num_class for the number of output classes, learning_rate of 0.01, n_estimators of 3000, max_depth 8, subsample 0.8, colsample_bytree 0.8, and regularization of L1 (reg_alpha) and L2 (reg_lambda) both set to 1.0. The main difference lies in the parameters specific to each model that uses different approaches, where XGBoost uses eval_metric='mlogloss' while LightGBM uses boosting_type='gbdt' as well as num_leaves=31 and min_child_samples=20 which are specific settings to control tree complexity in LightGBM. Unlike XGBoost, LightGBM does not include eval_metric explicitly in the model initialization. Although both models share similar basic settings, the different parameters reflect the different architectures and internal optimizations of the two algorithms.

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

2.5. Evaluation

In this study, the confusion matrix is first used to visualize and quantify the performance of the classification models by comparing predicted class labels against the actual class labels. From this matrix, the fundamental values of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) can be obtained to serve as the basis for calculating various evaluation parameters such as accuracy, precision, recall, F1-score. Log loss however, is gained from calculating predicted probabilities of each class. These metrics provide a comprehensive understanding of the model's effectiveness, particularly in the presence of imbalanced classes.

Accuracy reflects the overall correctness of the model's predictions by measuring the proportion of correctly classified instances among all data, as shown in (2). However, accuracy alone may not fully represent performance when class imbalance is present.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

Precision measures the proportion of the correct positive predictions out of all positive predictions by the model, indicating how reliable the model is when it predicts a certain class to reduce false alarms, calculated with (3).

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

Recall measures the ability of the model to correctly identify all actual positive instances with the (4) formula. This metric is crucial in medical diagnosis where missing a positive case carries serious consequences.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

The F1-score combines precision and recall into a single metric by calculating their average as shown in (5). It provides a balanced evaluation of the model's performance, especially when the dataset is imbalanced.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (5)$$

Log loss evaluates the uncertainty of the predicted probabilities by penalizing confident but wrong predictions more. Lower log loss values indicate better calibrated probability estimates, with the formula defined in (6).

$$Log\ Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} log(p_{ij}) \quad (6)$$

3. RESULT

3.1. Data Selection

In the process of evaluating ECG data recorded and collected using the Smart Holter device, a number of incomplete subject data were identified. These incompletenesses were characterized by missing iterations or segments of data, indicating interference or imperfections in the process of recording the data. To address this issue, the data of these 106 subjects was pruned through a selection process to remove the data of subjects that did not meet the criteria.

After this pruning process, the number of subjects with usable data for this study decreased. From the initial total number of 106 subjects, 72 subjects remained with complete and valid data for the purposes of this study. Although this number is fewer than planned, the remaining data is expected to have sufficient signal quality to be analyzed accurately, allowing for the analysis of the ECG parameters which were the primary focus in this study.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 4, August 2025, Page. 2281-2296
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

3.2. Preprocessing Result

The preprocessing and feature extraction pipeline started by converting raw ADC values into millivolts (mV). However, the results of this ADC conversion is too high as shown with an example of a raw ECG signal in Figure 4, where ECG signals generally range between -5 and +5 mV, while the value of the illustrated raw ECG signal data was observed to reach up to 871 mV. To obtain the correct and accurate values, filtration will be applied with three different methods gradually. These methods consist of baseline correction, Butterworth Filter, and FIR (Finite Impulse Response) Filter.

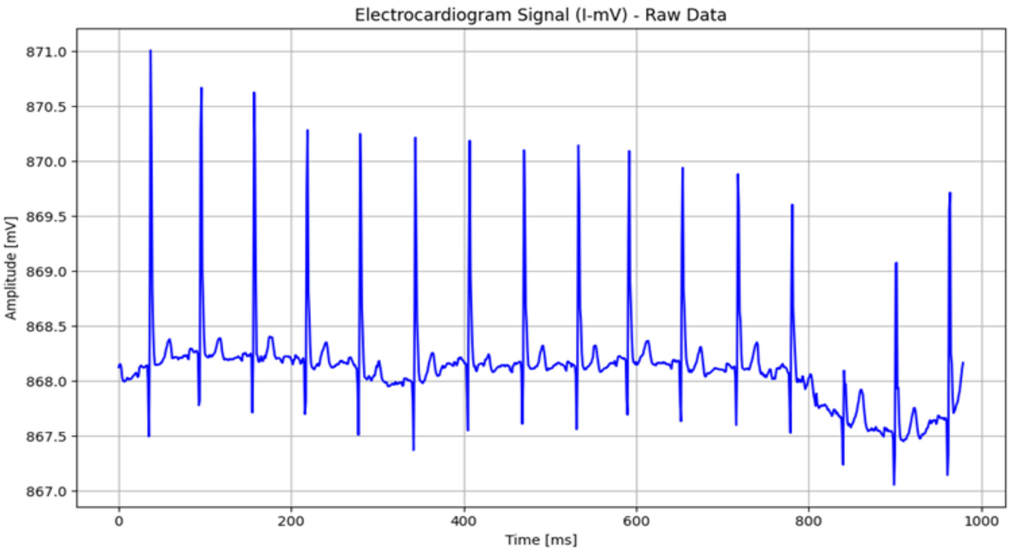

Figure 4. Plot of raw ECG signal.

First, baseline correction is performed to remove baseline drift, a slow shift of the baseline of the ECG signal caused by factors such as breathing, body movement, or noise from the electrodes. This was achieved using linear detrending, allowing the basic shape of the signal to become more visible, including the PQRST wave potential, which can be observed in Figure 5.

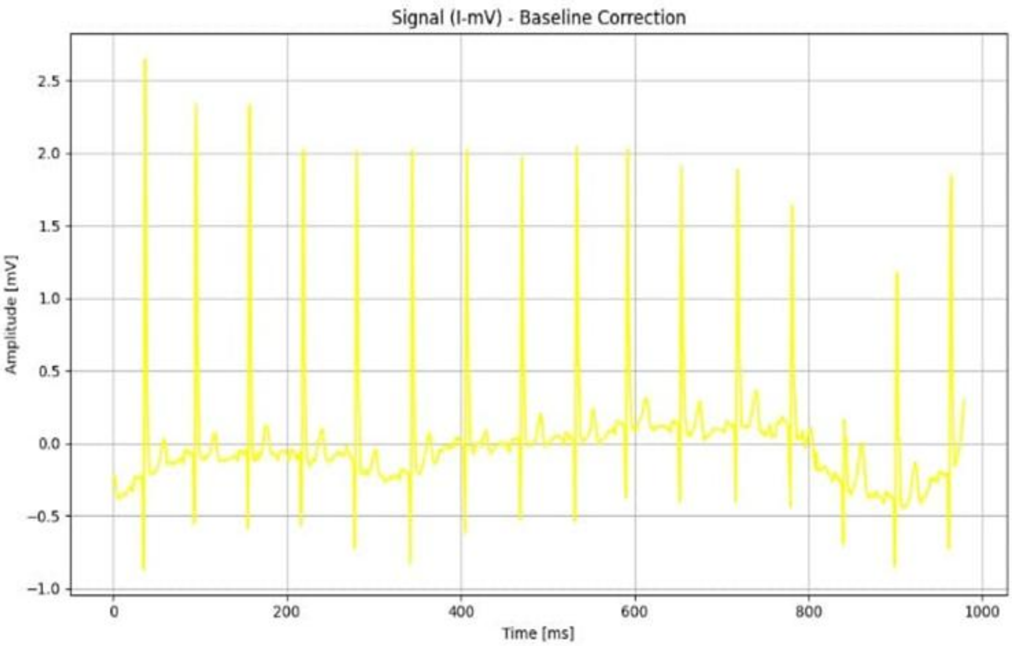

Figure 5. Plot of ECG signal after baseline correction process

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 4, August 2025, Page. 2281-2296
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

Following the baseline correction, Butterworth low-pass Filter is used to filter out noise or high frequencies such as electrical interference or muscle activity without causing distortion to the features, implemented as a 4th-order filter with a cutoff frequency of 0.6 Hz. As a result, the PQRST signal components became cleaner and more recognizable, as shown in Figure 6.

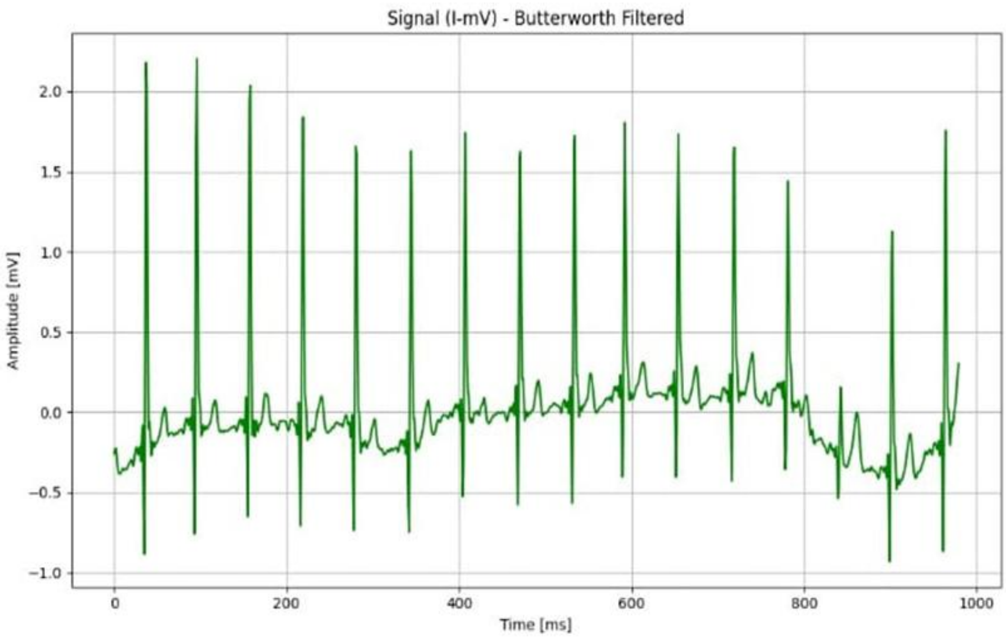

Figure 6. Plot of ECG signals filtered using the Butterworth Filter method.

Finally, FIR filter is applied using a Kaiser window with a cutoff frequency of 4 Hz to smoothen the overall signal while maintaining an accurate waveform, as shown in Figure 7.

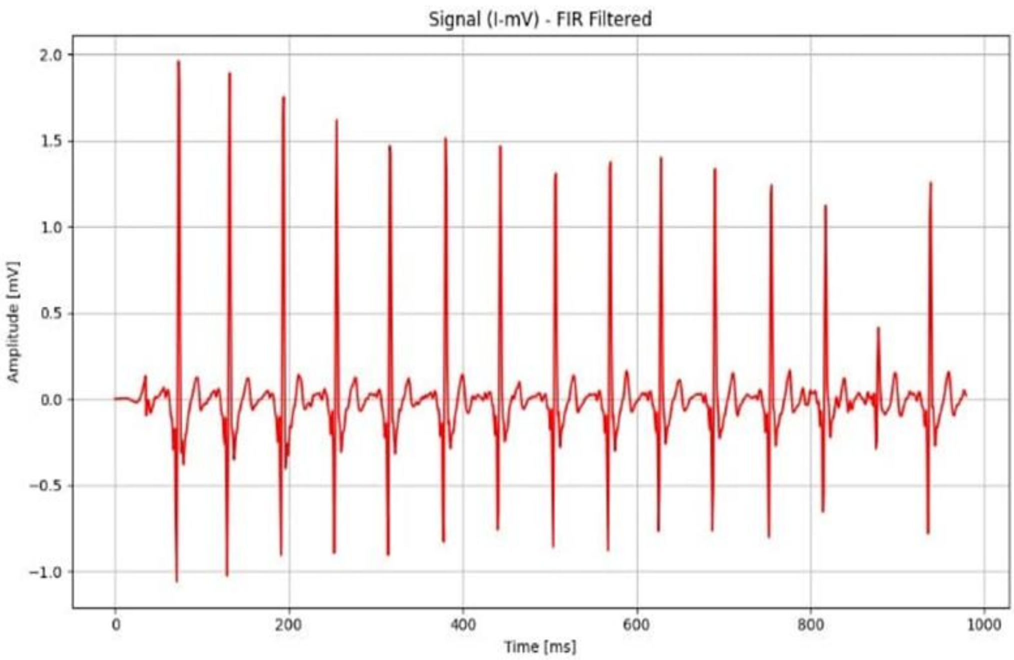

Figure 7. Plot of ECG signal further filtered using FIR filter.

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

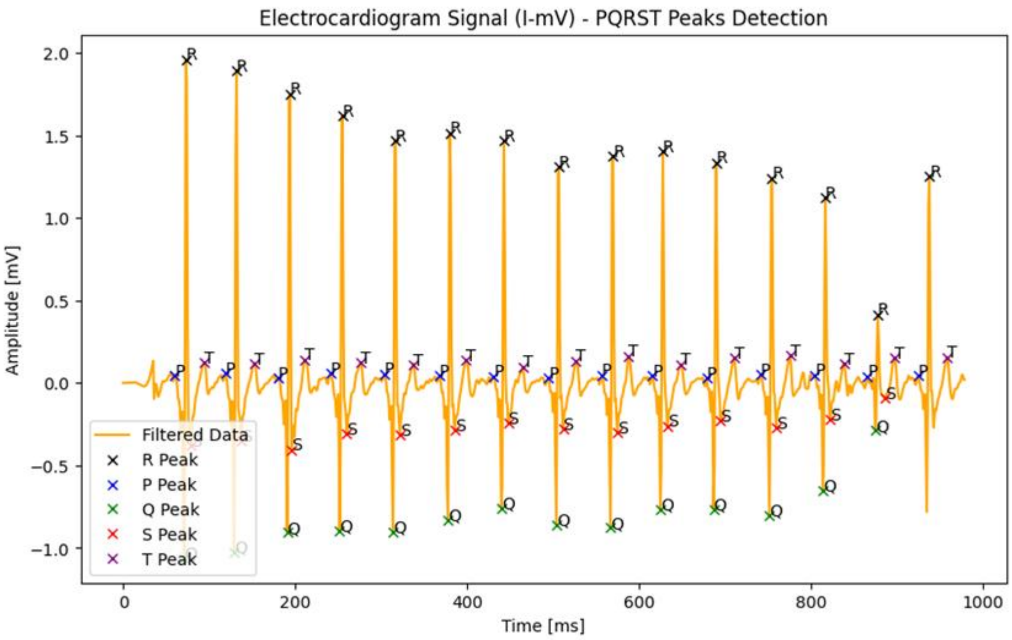

Figure 8. Plot of the filtered and preprocessed PQRST waveform.

The result of this filtration is used to detect the PQRST peaks of the ECG signal, performed with NeuroKit2's Discrete Wavelet Transform (DWT) method, allowing for accurate identification of the P, Q, R, S, and T peaks as well as their corresponding onsets and offsets. The processed ECG waveform, now within the approximate range of -1.0 mV to 2.0 mV, is displayed with identifiable marking for each peak as shown in Figure 8.

With the results of PQRST peak detection, extraction was performed to obtain features in the form of interval and duration values. The average ST interval was obtained from Lead I, the average intervals of RR, PR, QT, QRS duration, and heart rate (bpm) were obtained from Lead II, and the R/S ratio was obtained from Lead V1. This process was then carried out on all data from 72 subjects, with a total of 655 pieces of feature extraction data, which were packaged in a CSV (Comma Separated Value) file. Data labeling is done manually based on specific physiological features, with the criteria shown in Table 1.

Table 1. Range of ECG values based on classification category

Output Label	RR Interval (ms)	PR Interval (ms)	QRS Complex (ms)	QT Interval (ms)	ST Segment (ms)	R/S Ratio	Heart Rate (bpm)
Normal	600-1000	120-200	60-100	350-440	80-120	<1	60-100
Abnormal	<600	<120	<60	<350	<80	>1	<60
Potential of Arrhythmia	600-1000	190-200	101-120	441-460	120-130	>1	101-110
High Potential of Arrhythmia	600-1000	>200	>120	>460	>130	>1	>111

Each row corresponds to one of four classification categories, namely Normal, Abnormal, Potential of Arrhythmia, and High Potential of Arrhythmia. Meanwhile, each column indicates the name of the variable used in determining the classification such as RR interval, PR interval, QT interval, QRS

Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

complex duration, ST segment, R/S ratio, and heart rate to a range of physiological values that have been determined according to medical standards. These value ranges reflect both normal and pathological conditions and are used as a basis in determining the classification category of the ECG signal.

The preprocessing, feature extraction, and labeling results in a dataset consisting of 665 instances of ECG recordings, each represented by 7 numerical features as well as the output label. The class spread consists of 236 counts for class 0 representing Normal, 312 counts for class 1 representing Abnormal, and 107 counts for class 3 representing High Potential of Arrhythmia, while no instance of class 2, representing Potential of Arrhythmia, is observed in the process. The result is then compiled into one CSV file for subsequent training and evaluation, with the sample of the complete data shown in table 2.

Each row represents a data instance corresponding to a specific subject's ECG reading, while each column reflects one of the extracted features: RR interval, PR interval, QRS complex duration, QT interval, ST segment, R/S ratio, and heart rate. The final column, labeled output, indicates the class label assigned to the instance based on predefined medical thresholds, with values representing different arrhythmia risk levels. This structured dataset, compiled after preprocessing, feature extraction, and labeling, serves as the input for training and evaluating the classification models.

In the initial stage, the dataset will be read using the pandas library, which will then be stored as a dataframe. The variables x and y will then be taken from this dataframe, where x is all the columns excluding the column containing the subject's identification number and output label as the required features, while y is the column containing the output label as the actual value.

Table 2. Table of dataset as input to the models

Subject	rr	pr	qs	qt	st	r/s ratio	heart rate	output
1	629.7376	115.3846	87.91209	295.1895	217.93	0.50375	95.27778	0
1	494.6667	103.5294	80.47059	221.7778	160.8889	0.482003	121.2938	1
1	511.1972	114.4551	73.41541	238.1568	173.5573	0.900144	117.3715	1
	...							
106	506.6138	87.76844	71.42857	213.4039	181.2169	7.587139	118.4334	1
106	472.0648	76.11336	78.13765	241.2955	180.7692	11.03366	127.1012	1
106	451.9531	65.37829	81.82566	243.75	177.3438	9.05171	132.7571	1

The dataset of x that will then be used as an input for the model will be normalized using StandardScaler which is a method from the scikit-learn library that serves to modify feature values in order to obtain an average value (mean) of 0 and standard deviation of 1. This process is carried out to increase the effectiveness of the learning model and the stability of the results.

Data splitting is then implemented using the Stratified K-Fold Cross Validation method by dividing the data into folds using the StratifiedKFold function. This method divides the dataset as equal in size as possible to balance the class proportion in each fold. In this research, the data is split into 5 folds, chosen to ensure the balance of reliability and computational efficiency. The training and testing process is repeated 5 times, with each fold used once as the testing data while the remaining folds were used for training. The random state parameter is also defined for consistency and reproducibility in the division of data into folds.

3.3. Testing Results

In the process of assessing and comparing the XGBoost and LightGBM models, one of the tools used is the Confusion Matrix table. Confusion Matrix describes the performance of the model by comparing the predicted results with the actual labels. Confusion Matrix is useful for evaluating

Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

classification, especially in the case of multi-class classification in this study. In addition, this research also uses a number of quantitative evaluation parameters to measure accuracy and consistency in model prediction, which consist of Accuracy, Precision, Recall, F1 Score, and Log Loss. These parameters are calculated by comparing the model's predicted results against the true labels using the concepts of True Positive, False Positive, True Negative, and False Negative to provide a comprehensive overview of the model's performance. Log Loss, however, differs by utilizing predicted class probabilities to assess the confidence and calibration of the models.

3.4. Result of XGBoost Model Testing

The results of testing the XGBoost model with datasets that have gone through the Stratified Cross Validation process as the main method of dividing training data and testing data show excellent predictive performance, with the prediction accuracy value obtained from testing this model reaching 93.74%, indicating that most of the testing data could be predicted accurately. The precision and recall values also reached 93.74%, indicating that this model is able to provide accurate predictions and also recognize samples from each class evenly. F1 Score obtained from calculating the results of precision and recall produces a value of 93.66%, indicating a good balance between the two values. The log loss value of 0.1471 shows that this model is able to predict the probability with high confidence in the correct arrhythmia class.

The Confusion Matrix graph in Figure 9 shows that most of the correct predictions are on the main diagonal, where the vertical axis shows the actual value, while the horizontal axis shows the predicted value. In class 0, a total of 230 data were correctly predicted, while 2 data and 4 data were incorrectly predicted as class 1 and class 3, respectively. While in class 1, a total of 298 data were correctly predicted, but there were 8 data and 6 data that were incorrectly predicted into class 0 and class 3, respectively. Finally, in class 3, 86 data were correctly predicted, with 13 and 8 data incorrectly predicted as class 0 and class 1, respectively.

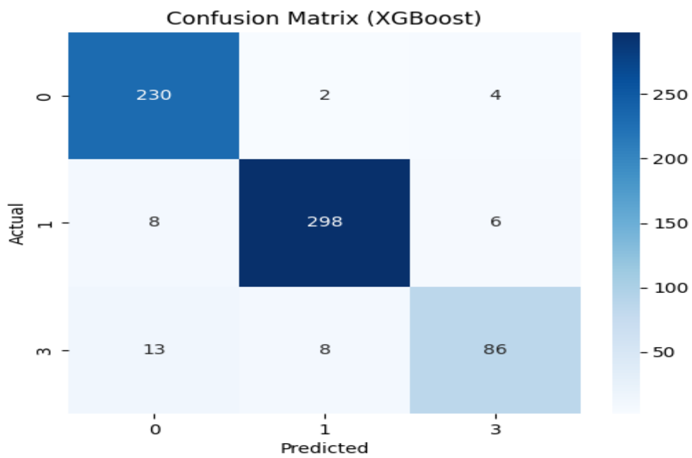

Figure 9. Confusion Matrix of XGBoost Model prediction results.

3.5. LightGBM Model Testing

As in the case with the previous model, the LightGBM Model also shows excellent prediction performance with the same dataset. The accuracy achieved by this model is 95.11%, precision and recall reached 95.10% and 95.10% respectively, F1 Score reached 95.03%, and the log loss value amounted to 0.1503.

In the Confusion Matrix graph of the LightGBM Model prediction results shown in Figure 10, the results are similar to the Confusion Matrix graph of the XGBoost prediction results, where most of

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

the correct prediction results are on the main diagonal of this graph. This Confusion Matrix graph shows that in class 0, there are 230 correctly predicted data, with errors of 2 and 4 data found in class 1 and 3 respectively. Class 1 gets 306 accurate data, while errors are found in 3 data each for class 0 and class 3. Finally, in class 3, 87 data are correctly predicted, errors of 15 data and 5 data are found in class 0 and class 1 respectively.

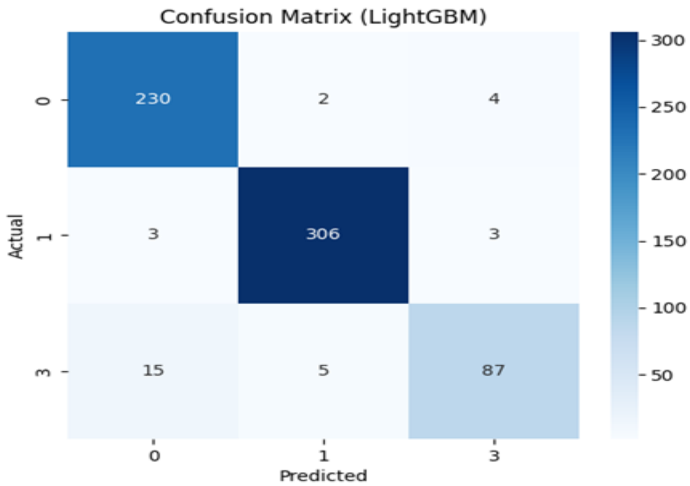

Figure 10. Confusion Matrix of LightGBM Model prediction results.

3.6. Comparison

Based on the outcome of testing both Machine Learning models above using the ECG dataset of students, it can be seen that the LightGBM Model obtained prediction results which tended to outperform the prediction results of the XGBoost Model. The LightGBM Model achieved an accuracy of 95.11% compared to the accuracy of the XGBoost Model which achieved 93.74%, showing a difference of 1.37%. Similar value differences can also be seen in the Precision, Recall, and F1 Score values, showing a difference of 1.36%, 1.37%, and 1.37% respectively. The comparison of these four test parameters can be seen in Figure 11.

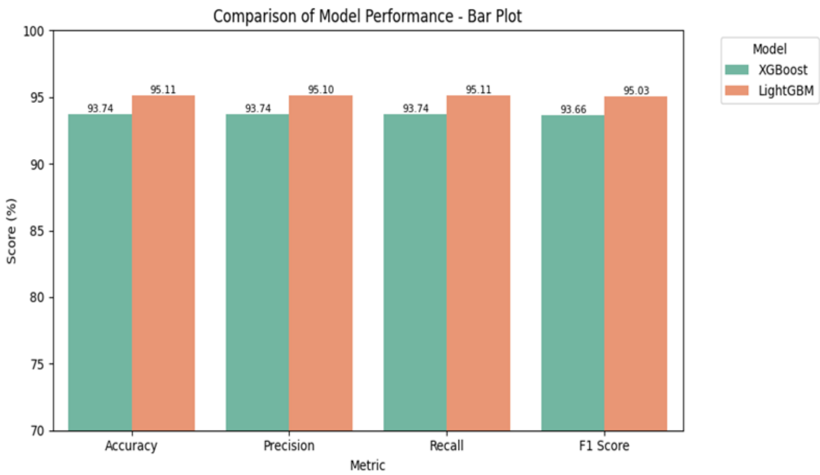

Figure 11. Comparison chart of Accuracy, Precision, Recall, and F1 Score of XGBoost and LightGBM models.

Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

In addition to using the four test parameters above, comparisons have also been made using log loss. In Figure 12, it is shown that the prediction results of the XGBoost Model have a lower log loss than the prediction results of the LightGBM Model. This difference indicates that although the LightGBM Model has a higher accuracy rate, the XGBoost Model has a slightly better confidence level in its prediction probability.

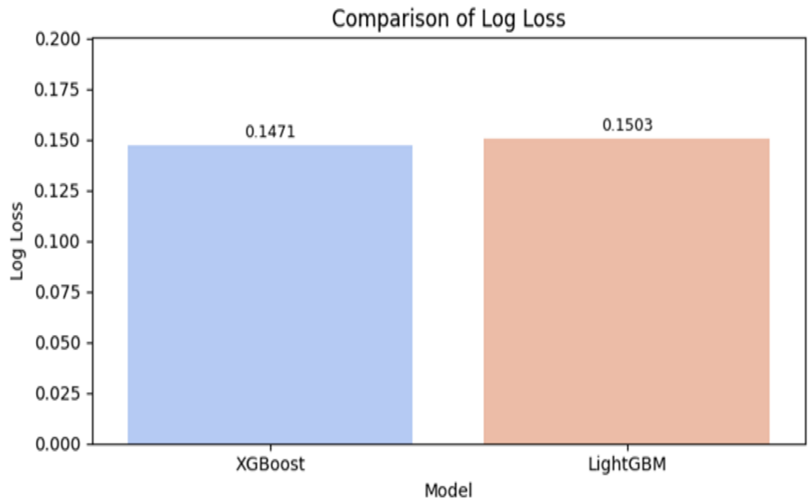

Figure 12. Log Loss comparison graph of XGBoost and LightGBM models.

4. DISCUSSIONS

This research compares the performance of two popular boosting algorithms, XGBoost and LightGBM, in multi-class classification of student ECG signals across four arrhythmia categories: normal, abnormal, potentially arrhythmic, and high potential of arrhythmic. Both models demonstrated strong predictive capabilities with high accuracy, precision, recall, and F1 scores. LightGBM showed a slight edge over XGBoost in accuracy and other classification metrics, indicating better overall prediction performance on this dataset. Analysis of the confusion matrices revealed that both models correctly classified the majority of samples, with minimal misclassification across classes. However, XGBoost displayed a marginally lower log loss value, suggesting slightly better calibration or confidence in its predicted probabilities. Despite this, the differences in log loss were minor and did not outweigh the higher accuracy metrics favoring LightGBM. One possible reason why LightGBM performs better in this study is due to its leaf-wise tree growth approach, which enables the model to capture more complex patterns in the data while still maintaining computational efficiency. In addition, LightGBM is known to handle sparse features well and includes built-in regularization mechanisms, making it less susceptible to overfitting. These characteristics may have contributed to the model's strong performance when applied to the interval-based ECG features used in this research.

Comparison with recent studies using machine learning on ECG data in arrhythmia prediction conducted by Xie et al. [9] and Li et al. [22] is favourable, showing improvements in accuracy. Xie et al. reported an average accuracy of approximately 92% across multiple models evaluated in their systematic review of atrial fibrillation detection using machine learning. Similarly, Li et al., by utilizing K-means clustering for feature extraction combined with LightGBM, reported an overall classification accuracy of 93.5%. The performance achieved by the present LightGBM model, showing 95.11% and supported by the F1 score of 95.03% and the log loss value 0.1503 highlight that the present model not only matches, but also outperforms many recent approaches in terms of both discriminative power and calibration, suggesting strong generalization on the tested arrhythmia classes.

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

5. CONCLUSION

This study compares the performance of two machine learning methods with popular boosting models, namely XGBoost and LightGBM, in multi-class ECG signal classification using four categories of arrhythmia classes. These classes are normal, abnormal, potential of arrhythmia, and high potential of arrhythmia. Based on the evaluation results using test parameters such as accuracy, precision, recall, F1 score, and log loss, both models show excellent performance, where LightGBM appears to be superior in prediction accuracy supported by additional test parameters with similar results. The confusion matrix analysis also shows the accurate prediction distribution and minimal misclassification of both models. However, XGBoost has a slightly lower log loss value, indicating a slightly more stable prediction probability compared to the LightGBM Model. However, this difference is minimal and does not detract from the advantages of the LightGBM Model in predicting potential arrhythmias in students. These findings suggest that the LightGBM model, with its strong predictive performance, has high potential to be integrated into a machine learning-based clinical decision support system, assisting medical professionals in the early detection and classification of arrhythmias from ECG data. Further development could also be done to adapt for real-time prediction by integrating it with streaming ECG data sources or wearable devices, enabling continuous arrhythmia monitoring using multi-lead ECGs.

Despite the promising results of this study, there are some limitations that should be considered. The research utilized ECG data from only three leads (I, II, and V1) combined with intervals like RR and PR, which may not fully capture the complexity of cardiac signals across diverse populations. This could affect the generalizability of the findings. Additionally, this study did not explore variations in model hyperparameters or alternative feature engineering approaches, which could potentially improve performance further. Moreover, evaluation was limited to stratified k-fold cross-validation on the same dataset without validation on completely unseen external data, restricting the assessment of the models' robustness in more practical settings.

These limitations present valuable opportunities for future work. Expanding the dataset to include additional ECG leads and a more diverse pool of subjects may improve generalizability. Systematic hyperparameter tuning and experimentation with advanced feature extraction methods may also boost the accuracy of the classification. Moreover, validating the models on external unseen data may help to confirm their practical applicability and stability for general use. Overall, despite these limitations, the current findings provide a solid foundation for ongoing research into optimizing machine learning models for arrhythmia classification using student ECG data.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest between the authors or with research objects in this paper.

ACKNOWLEDGEMENT

The authors express their sincere gratitude towards Universitas Prima Indonesia for the support and facilities provided, and to SMK Swasta Teladan Sumatera Utara 2, including all of the participating students, teachers, and staffs, for their valuable contributions and assistance in providing the materials and cooperation necessary for the success of this research.

REFERENCES

[1] B. Takase *et al.*, "JCS/JHRS 2022 Guideline on Diagnosis and Risk Assessment of Arrhythmia," *Journal of Arrhythmia*, vol. 40, no. 4, pp. 655–752, Apr. 2024, doi: 10.1002/joa3.13052.

Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

[2] A. B. Curtis, R. Karki, A. Hattoum, and U. C. Sharma, "Arrhythmias in patients ≥ 80 years of age: pathophysiology, management, and outcomes," *J. Am. Coll. Cardiol.*, vol. 71, no. 17, pp. 2041–2057, May 2018, doi: 10.1016/j.jacc.2018.03.019.

[3] S. Khurshid, J. M. Ashburner, P. T. Ellinor, D. D. McManus, S. J. Atlas, D. E. Singer, and S. A. Lubitz, "Prevalence and incidence of atrial fibrillation among older primary care patients," *JAMA Network Open*, vol. 6, no. 2, p. e2255838, 2023, doi: 10.1001/jamanetworkopen.2022.55838.

[4] A. C. Lubis, A. N. Nasution, and H. A. P. Lubis, "Prevalence of Congenital Heart Disease and Pediatric Rhythm Disorder or Arrhythmia in Children in Rantau-Prapat City, North Sumatra, Indonesia," *J. Endocrinol. Trop. Med. Infect. Dis. (JETROMI)*, vol. 5, no. 4, 2024, doi: 10.32734/jetromi.v5i4.14332.

[5] K. P. Yasa, A. A. Katritama, I. K. A. P. Harta, and I. W. Sudarma, "Prevalence and risk factors analysis of early postoperative arrhythmia after congenital heart surgery in pediatric patients," *J. Arrhythm.*, vol. 40, no. 2, pp. 356–362, Mar. 2024, doi: 10.1002/joa3.13011.

[6] Y. Ansari *et al.*, "Deep learning for ECG arrhythmia detection and classification: an overview of progress for period 2017–2023," *Frontiers in Physiology*, vol. 14, p. 1246746, Sep. 2023, doi: 10.3389/fphys.2023.1246746.

[7] P. Kligfield, L. Gettes, J. Bailey, R. Childers, B. Deal *et al.*, "Recommendations for the Standardization and Interpretation of the Electrocardiogram. Part I: The Electrocardiogram and Its Technology," *Circulation*, vol. 49, no. 10, pp. 1109–1136, 2004, doi: 10.1161/CIRCULATIONAHA.106.180200.

[8] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," *Expert Systems with Applications: X*, vol. 7, p. 100033, 2020, doi: 10.1016/j.eswax.2020.100033.

[9] C. Xie, Z. Wang, C. Yang, J. Liu, and H. Liang, "Machine Learning for Detecting Atrial Fibrillation from ECGs: Systematic Review and Meta-Analysis," *Rev Cardiovasc Med.*, vol. 25, no. 1, p. 8, Jan. 2024, doi: 10.31083/j.rcm2501008.

[10] B. Mahesh, "Machine Learning Algorithms – A Review," *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, Jan. 2019, doi: 10.21275/ART20203995.

[11] Z. A. Ali, Z. H. Abduljabbar, H. A. Tahir, A. B. Sallow, and S. M. Al Mufti, "Exploring the power of eXtreme gradient boosting algorithm in machine learning: a review," *Academic Journal of Nawroz University*, vol. 12, no. 2, pp. 320–334, May 2023, doi: 10.25007/ajnu.v12n2a1612.

[12] M. R. Machado, S. Karray, and I. T. de Sousa, "LightGBM: an Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry," *2019 14th International Conference on Computer Science & Education (ICCSE)*, Toronto, ON, Canada, 2019, pp. 1111-1116, doi: 10.1109/ICCSE.2019.8845529.

[13] D. Boldini, F. Grisoni, D. Kuhn *et al.*, "Practical guidelines for the use of gradient boosting for molecular property prediction," *J. Cheminform.*, vol. 15, p. 73, 2023, doi: 10.1186/s13321-023-00743-7.

[14] XGBoost Developers, "Parameter Tuning," *XGBoost Documentation*, [Online]. Available: https://xgboost.readthedocs.io/en/stable/tutorials/param_tuning.html. [Accessed: 2-Jun-2025].

[15] LightGBM Developers, "Features — LightGBM 4.3.0.99 documentation," *LightGBM Documentation*, [Online]. Available: https://lightgbm.readthedocs.io/en/stable/Features.html. [Accessed: 2-Jun-2025].

[16] R. Sibindi, R. W. Mwangi, and A. G. Waititu, "A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices," *Engineering Reports*, vol. 8, 2023, Art. no. e12599, doi: 10.1002/eng2.12599.

[17] P. S. Rizky, R. H. Hirzi, dan U. Hidayaturrohman, "Perbandingan metode LightGBM dan XGBoost dalam menangani data dengan kelas tidak seimbang," *J. Statistika*, vol. 15, no. 2, pp. 228–236, Des. 2022, doi: 10.36456/jstat.vol15.no2.a5548.

Vol. 6, No. 4, August 2025, Page. 2281-2296
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5015

[18] T. W. Cabral, F. B. Neto, E. R. de Lima, G. Fraidenraich, and L. G. P. Meloni, "Analysis of variance combined with optimized gradient boosting machines for enhanced load recognition in home energy management systems," *Sensors*, vol. 24, no. 15, p. 4965, 2024, doi: 10.3390/s24154965.

[19] C. T. Chung, G. Bazoukis, S. Lee *et al.*, "Machine learning techniques for arrhythmic risk stratification: a review of the literature," *Int. J. Arrhythm.*, vol. 23, p. 10, 2022, doi: 10.1186/s42444-022-00062-2.

[20] Q. Xiao, K. Lee, S. A. Mokhtar, I. Ismail, A. L. b. M. Pauzi, Q. Zhang, and P. Y. Lim, "Deep learning-based ECG arrhythmia classification: a systematic review," *Appl. Sci.*, vol. 13, p. 4964, 2023, doi: 10.3390/app13084964.

[21] A. Jafari, F. Yousefirizi, and V. Seydi, "DeepBoost-AF: A novel unsupervised feature learning and gradient boosting fusion for robust atrial fibrillation detection in raw ECG signals," *arXiv*, vol. 2505, no. 24085, May 2025, doi: 10.48550/arXiv.2505.24085.

[22] H. Li, X. Hong, and B. Hu, "Analysis of arrhythmia features based on LightGBM and K-means feature extraction," *Front. Comput. Intell. Syst.*, vol. 5, no. 2, pp. 67–71, 2023, doi: 10.54097/fcis.v5i2.12471.

[23] G. Silva, P. Silva, G. Moreira, V. Freitas, J. Gertrudes, and E. Luz, "A systematic review of ECG arrhythmia classification: adherence to standards, fair evaluation, and embedded feasibility," *arXiv*, Mar. 2025, doi: 10.48550/arXiv.2503.07276.

[24] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[25] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.