Available online at
https://mijuitm.com.my

e-ISSN: 2735-0703

Mathematical Sciences and Informatics Journal 6(2) 2025, 247 – 260.

Mathematical Sciences and Informatics Journal

Interpretable Arrhythmia Classification Using a Convolutional Neural Network and the LIME Technique

Adam Mohd Khairuddin[1*], Siti Armiza Mohd Aris[2], Azizul Azizan[1], Noor Jannah Zakaria[1]

[1]*Department of Intelligence Informatics,Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia*
[2]*Department of Smart Engineering and Advanced Technology, Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia*

ARTICLE INFO

Article history:
Received 10 August 2025
Revised 14 September 2025
Accepted 25 September 2025
Online first
Published 31 October 2025

Keywords:
Interpretable
Arrhythmia
Classification
Electrocardiography
Convolutional Neural Network

DOI:
10.24191/mij.v6i2.9317

ABSTRACT

Deep learning models have demonstrated strong performance in electrocardiogram (ECG) arrhythmia classification. However, their lack of interpretability limits clinical trust and adoption. By adopting an explainable artificial intelligence (XAI) technique, this study aims to enhance the interpretability of a convolutional neural network (CNN) model. More specifically, the Local Interpretable Model-Agnostic Explanations (LIME) technique is utilized to interpret the CNN model used to classify 17 classes of ECG arrhythmias. The CNN model was developed using a five-stage framework. The study uses the MIT-BIH Arrhythmia database to evaluate the performance of the CNN model. Results indicate that the model was able to accomplish precision of 97.00%, recall of 97.00%, F1-score of 97.00%, and overall accuracy of 99.00%. In addition, the LIME technique provides local explanations that help in the understanding of the decision-making process of the CNN model in classifying the 17 classes of ECG arrhythmias.

1. INTRODUCTION

For decades, cardiovascular diseases (CVDs) have been the leading cause of death worldwide. According to the World Heart Report 2023, an estimated 20.5 million people died from CVDs in 2021 alone (Di Cesare et al., 2023). More specifically, the ischemic heart disease was the primary cause of premature death in 146 countries for men and in 98 countries for women. The modifiable risk factors that contributed to CVD deaths in 2021 included (1) elevated low-density lipoprotein (LDL) cholesterol; (2) high fasting plasma

[1*] Corresponding author. *E-mail address*: adam.mk@utm.my
https://doi.org/10.24191/mij.v6i2.9317

glucose; (3) air pollution; (4) high body-mass index; (5) tobacco use; (6) low physical activity; and (7) raised blood pressure.

The prevalence of cardiovascular diseases highlights the need for early detection of heart abnormalities through the electrocardiogram. The electrocardiogram (ECG or EKG) records the electrical signals of the heart by placing electrodes on specific areas of the chest, arms, and legs. Conventionally, ECG signals are recorded using an analog ECG machine. The ECG test results can assist not only in the diagnosis of irregular heartbeats known as arrhythmias but also help to reduce the risk of severe heart complications and save lives as well. However, ECG interpretations have limitations: it is visually inspected, time-consuming, requires special expertise, and is prone to human error (Ojha et al., 2024).

More recently, with the introduction of artificial intelligence (AI) technology, modern ECG devices have started to utilize AI algorithms, including machine learning (ML) and deep learning (DL) to improve the detection and classification of arrhythmia. The ML and DL algorithms offer numerous advantages. For instance, ML algorithms can continuously monitor patients in real-time by using wearable devices and can alert clinicians when significant changes occur in ECG signals (Katal et al., 2023). In addition, by helping to eliminate noise and artefacts such as baseline wander, powerline interference, and electrode contact from the ECG signals, DL algorithms can improve feature extraction and reduce false alarms in clinical settings (Hou et al., 2023). DL algorithms also effectively identify the intricate patterns and distinguishing characteristics related to various types of arrhythmias through large-scale and diverse ECG datasets, which can lead to improve accuracy of arrhythmia classification (Narotamo et al., 2024).

The study uses a CNN model to classify 17 types of ECG arrhythmia. In addition, the study adopted the LIME technique to enhance the interpretability of the CNN model. This paper is arranged in the following manner. Section 2 presents related works on ECG arrhythmia classification approaches and XAI techniques. Section 3 provides the methodology adopted in this study. Section 4 discusses and evaluates the experimental outcomes of this work. Section 5 concludes the paper.

2. RELATED WORKS

Researchers have developed and adopted multiple ML and DL approaches to enhance the classification of ECG arrhythmias (Khairuddin et al., 2024; Khairuddin & Ku Azir, 2021). However, these approaches often lack clinical interpretability which poses challenges for decision-making in healthcare. More recent studies have emphasized on the applications of explainable artificial intelligence (XAI) techniques to enhance the interpretability of the ML and DL techniques. These studies utilized different DL techniques to classify ECG arrhythmias as well as adopted several XAI techniques to interpret the DL models.

2.1 Deep Learning Models

Several studies have explored CNN model for ECG arrhythmia classification. For example, Azzem and Harrag (2023) proposed a CNN to classify five arrhythmias, attaining an accuracy of 0.903 with corresponding precision, recall, F1-score, and area under the curve (AUC) of 0.74, 0.76, 0.75, and 0.93, respectively. Another study by (Anand et al., 2022) developed an ST-CNN-GAP-5 model with a global average pooling layer. The model was able to accomplish 95.80% accuracy and an AUC of 99.46%.

The recent study by (Salvi, 2024) introduced the CNN-Bidirectional Gated Recurrent Unit (CNN-BiGRU) model to classify five arrhythmia types. The CNN-BiGRU model was able to achieve accuracy of 94.44%, precision of 94.33%, recall of 94.36%, and F1-score of 94.35%. The previous work by (Ojha et al., 2024) used the ST-CNN-5 model to classify 5 different types of arrhythmias. The performance of the model reached an accuracy of 0.891, along with precision of 0.798, recall of 0.693 and specificity of 0.934.

https://doi.org/10.24191/mij.v6i2.9317

2.2 XAI Techniques for DL Models

Different XAI techniques such as SHapley Additive exPlanations (SHAP), LIME, and Gradient-weighted Class Activation Mapping (GradCAM) have been employed to interpret these deep learning models. For instance, the work by (Sathi et al., 2024) used both SHAP and LIME with a histogram gradient boosting (Hist-GB) classifier to identify key ECG fiducial features relevant for diagnosing arrhythmia, ischemia and healthy classes. The study by (Ojha et al., 2024) compared SHAP, LIME, and GradCAM for model interpretation and found SHAP to be the most effective in emphasizing relevant ECG features. Meanwhile, the research by (Azzem & Harrag, 2023) used GradCAM to visualize influential ECG regions. Another study by (Salvi, 2024) applied LIME to highlight the important regions of the ECG waves.

Overall, prior works have demonstrated that integrating XAI with DL models can enhance interpretability in ECG arrhythmia classification. However, most studies focus on a limited number of arrhythmia classes. This study addresses this gap by integrating LIME technique with a CNN to classify 17 types of arrhythmias, providing both high predictive accuracy and interpretable local explanations. The LIME is chosen due to its model-agnostic property, which enables the interpretation of CNN predictions.

3. METHODOLOGY

In this study, the interpretable ECG classification model is developed using a five-stage framework (Fig. 1).

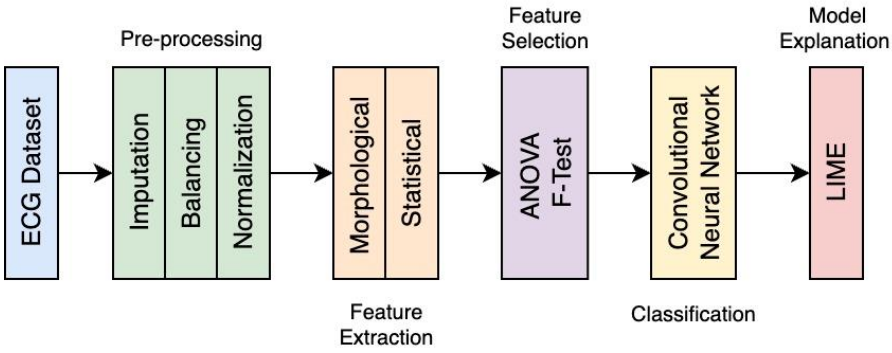

Fig. 1. Five stages of the proposed framework

3.1 ECG Dataset

The MIT-BIH Arrhythmia Database is utilized in this study to develop and evaluate the CNN model for classifying the 17 classes of ECG arrhythmia. The 48 ECG signals were obtained from the PhysioNet repository. Each ECG signal is 30 minutes long, sampled at 360 Hz. However, random selection was applied to extract 10-second fragments (3600 samples) from the ECG signals for the purpose of analysis. This study utilized only the ECG signals derived from the MLII lead.

3.2 Feature Extraction

ECG feature extraction included both morphological and statistical features. Seven morphological features were calculated from the positions of the R-peaks. An additional nine wavelet-based statistical features were extracted by using the wavelet decomposition process with the Daubechies 1 (Db1) wavelet. The following Table 1 lists the extracted features and their descriptions.

https://doi.org/10.24191/mij.v6i2.9317

Table 1. Morphological and statistical features extracted

	Name	Description
Morphological features	AvgHR	The average heart rate of the RR intervals.
	Mean_RR	Mean of the RR intervals.
	rmssd	Root-mean-square distance between successive RR intervals.
	NumR	The total count of R-peaks with differences exceeding a 30-millisecond threshold.
	sdRR	Standard deviation calculated from the RR intervals.
	sdHR	Standard deviation calculated from the heart rate.
	pse	The power spectral entropy of the R-peaks.
Statistical features	Mean	The mean of the 2nd detail coefficient.
	STD	The standard deviation of the 2nd detail coefficient.
	SK	The skewness of the 2nd detail coefficient.
	Kurt	The kurtosis of the 2nd detail coefficient.
	RMS	The root mean square of the 2nd detail coefficient.
	MR	Average ratio computed between the 1st and 2nd detail coefficient.
	Max	The highest value of the 2nd detail coefficient.
	Min	The lowest value of the 2nd detail coefficient.
	En	The energy of the 2nd detail coefficient.

3.3 Feature Selection

A univariate statistical test (SelectKBest with ANOVA F-test, $f\_classif$) identified the most relevant features. A total of 10 top-ranked features were chosen based on F-scores. The feature selection stage helped to reduce the dimensionality of the feature set to improve the model training. Fig. 2 presents the top 10 selected features based on the ANOVA F-test. Based on Fig. 2, Kurt has the highest F-score, whereas Mean has the lowest.

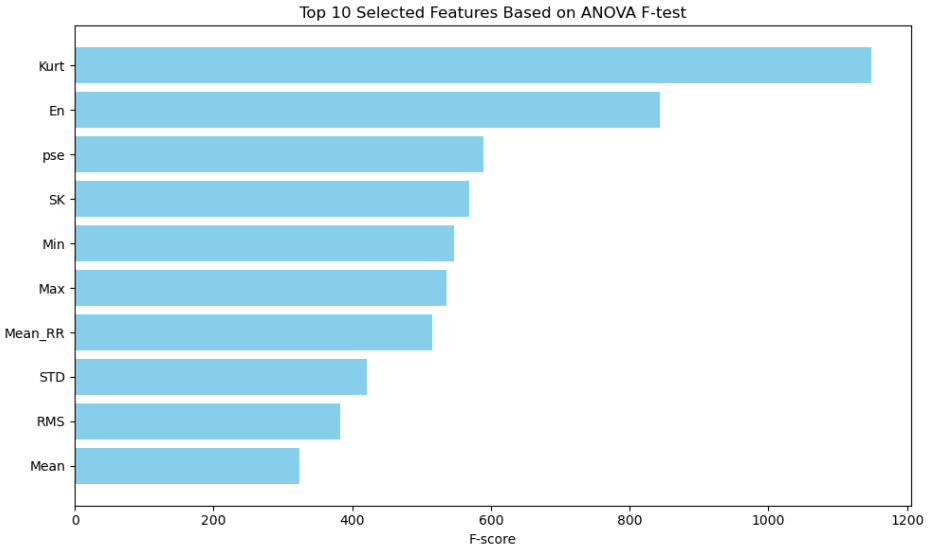

Fig. 2. The top 10 selected features based on the F-score

https://doi.org/10.24191/mij.v6i2.9317

3.4 Convolutional Neural Network Model

The proposed arrhythmia classification model is a one-dimensional CNN (1D-CNN) for sequential ECG input (Fig. 3). The following section provides a brief overview of the layers used in the CNN model.

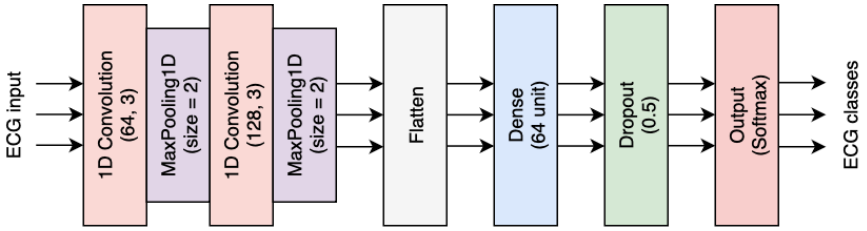

Fig. 3. The architecture framework

3.4.1 Input Layer

This input layer accepted the ECG input with a shape corresponding to the 10 selected features from the SelectKBest method. The ECG input was reshaped as a 1-dimensional sequence.

3.4.2 First Convolutional Block

A Conv1D layer with 64 filters and a kernel size of 3 were employed through the rectified linear unit (ReLU) activation function to extract local patterns from the input sequence. This was followed by a MaxPooling1D layer with a pool size of 2. This helped to reduce the spatial dimension as well as mitigate overfitting.

3.4.3 Second Convolutional Block

An additional Conv1D layer with 128 filters and a kernel size of 3 was incorporated to capture more complex and abstract features. This was followed again by another MaxPooling1D layer with a pool size of 2 for further dimensionality reduction.

3.4.4 Flattening Layer

The output of the convolutional layers was flattened into a one-dimensional vector to prepare it for the dense layers.

3.4.5 Fully Connected Layer

A dense layer with 64 neurons and ReLU activation function were used to learn high level representations. A dropout rate of 0.5 was incorporated to mitigate overfitting by randomly deactivating neurons during training.

3.4.6 Output Layer

The output layer was included with a softmax activation function to enable multiclass classification of 17 classes of ECG arrhythmia.

The CNN model was set up for training with the following hyperparameter: (1) Adam optimizer; (2) batch size of 64; and (3) epochs of 200. The loss function was the categorical cross entropy. Early stopping was employed with a patience of 5 epochs to stop the model training when the validation loss no longer improved.

https://doi.org/10.24191/mij.v6i2.9317

3.5 Performance Metrics

Four performance metrics were employed to evaluate the proposed ECG arrhythmia classification model. These performance metrics included: (1) accuracy; (2) precision; (3) recall; and (4) F1-score. TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. The section below briefly explains the four-performance metrics.

3.5.1 Accuracy

As shown in equation 1, accuracy is the ratio between the number of correct predictions over the total predictions.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FN + FP) \quad (1)$$

3.5.2 Precision

Precision indicates the ratio of correctly predicted positive instances to the total predicted positives by the CNN model.

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

3.5.3 Recall

Recall referred to the ratio of true positives to the total number of actual positive instances.

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

3.5.4 F1-score

F1-score is a metric that combines precision and recall using their harmonic mean.

$$\text{F1-score} = 2 \text{ x } ((\text{Precision x Recall}) / \text{Precision} + \text{Recall}) \quad (4)$$

3.6 Model Explanation

In this study, LIME was used for local explanation of the CNN model. This technique focused on explaining the prediction made by the CNN model for a specific data point (instance), instead of explaining the model's behavior globally. This method weighed each of the features that represented their contribution to the model prediction for the specific single instance.

4. RESULTS AND DISCUSSION

Experiments were conducted in Jupyter Notebook with Python version 3.12.7 based on the following PC specifications: (1) AMD Ryzen 9 9900X (2) 32 GB of DDR5 RAM (3) RTX 4070 GPU and (4) Windows 11. The effectiveness of the CNN model in categorizing the 17 classes of arrhythmias as well as the interpretation of the CNN model by using the LIME technique are presented and discussed in the section below.

Before the training of the CNN model, the ECG dataset was allocated into training (80%) and testing (20%) sets. More specifically, the 4,743 samples were divided into 3,794 training sets (80%) and 949 testing sets (20%). Before the dataset was split, the preprocessing techniques were applied to the ECG dataset.

https://doi.org/10.24191/mij.v6i2.9317

4.1 Data Pre-processing

This subsection explains the data preprocessing processes adopted in the study. The data preprocessing helped to improve the data quality as well as the performance of the CNN model. The data preprocessing process included the following: (1) data imputation; (2) data balancing; and (3) data normalization.

4.1.1 Data imputation

The data imputation involved replacing the missing values of numeric columns in the ECG dataset. The missing values in this dataset were imputed with the mean of the respective column values.

4.1.2 Data balancing

The ECG dataset exhibits class imbalance where certain classes contain significantly more samples than others. Specifically, the NSR class has the highest number of samples, whereas the VFL class has the lowest number of samples. Handling the class imbalance in dataset is important to avoid biased and inaccurate model predictions.

The class imbalance in the ECG dataset was addressed by adopting the random over sampling (ROS) technique. This technique selected the existing samples from the minority class randomly and duplicated them to balance out the dataset.

4.1.3 Data Normalization

After balancing the dataset, the dataset also needed to be normalized. The robust scaler function was used to normalize the 16 features in the dataset. The robust scaler function was selected due to its robustness to outliers presented in the dataset. Equation 5 below shows the expression for the robust scaler function, where Q_1 is the first quartile and Q_3 is the third quartile.

$$X = (x_i - Q_1(x)) / (Q_3(x) - Q_1(x)) \tag{5}$$

4.2 Classification of Arrhythmia

The performance of the CNN model was assessed by using performance metrics that consisted of precision, recall, and F1-score. The results of each arrhythmia classes are summarized in Table 2. The support in Table 2 refers to the total number of instances in each class.

Table 2. Performance metrics for each arrhythmia class.

Class	Precision	Recall	F1-score	Support
NSR	0.94	0.82	0.87	56
APB	0.95	0.98	0.96	56
AFL	1.00	1.00	1.00	56
AFIB	0.93	0.95	0.94	56
SVTA	1.00	1.00	1.00	56
WPW	1.00	1.00	1.00	56
PVC	0.89	0.86	0.87	56
BG	0.93	0.96	0.95	56
TG	0.98	1.00	0.99	56
VT	1.00	1.00	1.00	56
IVR	1.00	1.00	1.00	56
VFL	1.00	1.00	1.00	56

https://doi.org/10.24191/mij.v6i2.9317

Fusion	0.97	1.00	0.98	56
LBBB	0.97	1.00	0.98	56
RBBB	0.98	0.96	0.97	56
SDHB	1.00	1.00	1.00	56
PR	1.00	1.00	1.00	56

Based on the results in Table 2, the CNN model was able to predict accurately classes AFL, SVTA, WPW, VT, IVR, VFL, SDHB, and PR, with precision, recall, F1-score of 1.00, 1.00, and 1.00, respectively. In the case of the accuracy, the CNN model was able to achieve accuracy of 97.00%. Table 3 shows the macro and weighted average of precision, recall, and F1-score for the CNN model.

Table 3. The macro and weighted average of precision, recall, and F1-score.

	Precision	Recall	F1-score	Support
Macro average	0.97	0.97	0.97	949
Weighted average	0.97	0.97	0.97	949

The results in Table 3 indicate that both the macro and weighted average of precision, recall and F1-score was 0.97. Fig. 4 shows the training and validation loss plot graph. The x-axis represented the number of epochs (iterations through the entire training dataset), whereas the y-axis represented the loss, which was a measure of error.

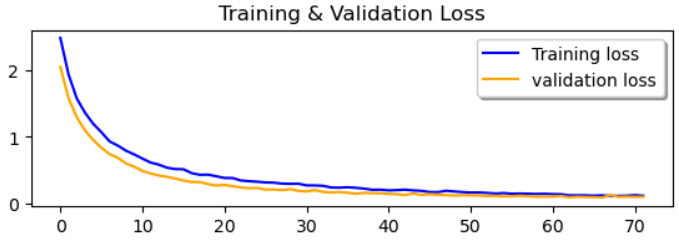

Fig. 4. The loss performance of the model

Both training and validation losses decreased consistently and converged around the same low values. The validation loss did not diverge while training loss decreased, indicating there was no significant overfitting and generalized well to unseen data. Following this, Fig. 5 presents the training and validation accuracy plot graph during the training of the CNN model. The x-axis represented the number epochs, whereas the y-axis represented the accuracy.

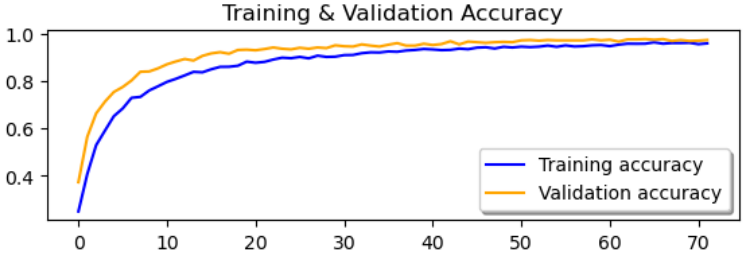

Fig. 5. The accuracy performance of the model

As shown in Fig. 5, there was a rapid increase in both training and validation accuracy during the early epochs (0-10). In the middle epochs (10-30), the validation accuracy was slightly higher. In the later epochs (30-70), both training and validation accuracy plateau close to 1.0. There was very little difference between

https://doi.org/10.24191/mij.v6i2.9317

the training and validation accuracy, which suggested no significant overfitting as well as the model had good generalization on unseen data.

Accordingly, the following Fig. 6 presents the 17x17 confusion matrix representing the classification outcomes derived from the 17 classes of arrhythmias. The analysis of the confusion matrix suggested that the CNN model predicted accurately the AFL, SVTA, WPW, BG, TG, VT, IVR, VFL, Fusion, LBBB, RBBB, SDHB, and PR classes. However, the CNN model predicted NSR, APB, AFIB, and PVC classes were less accurate.

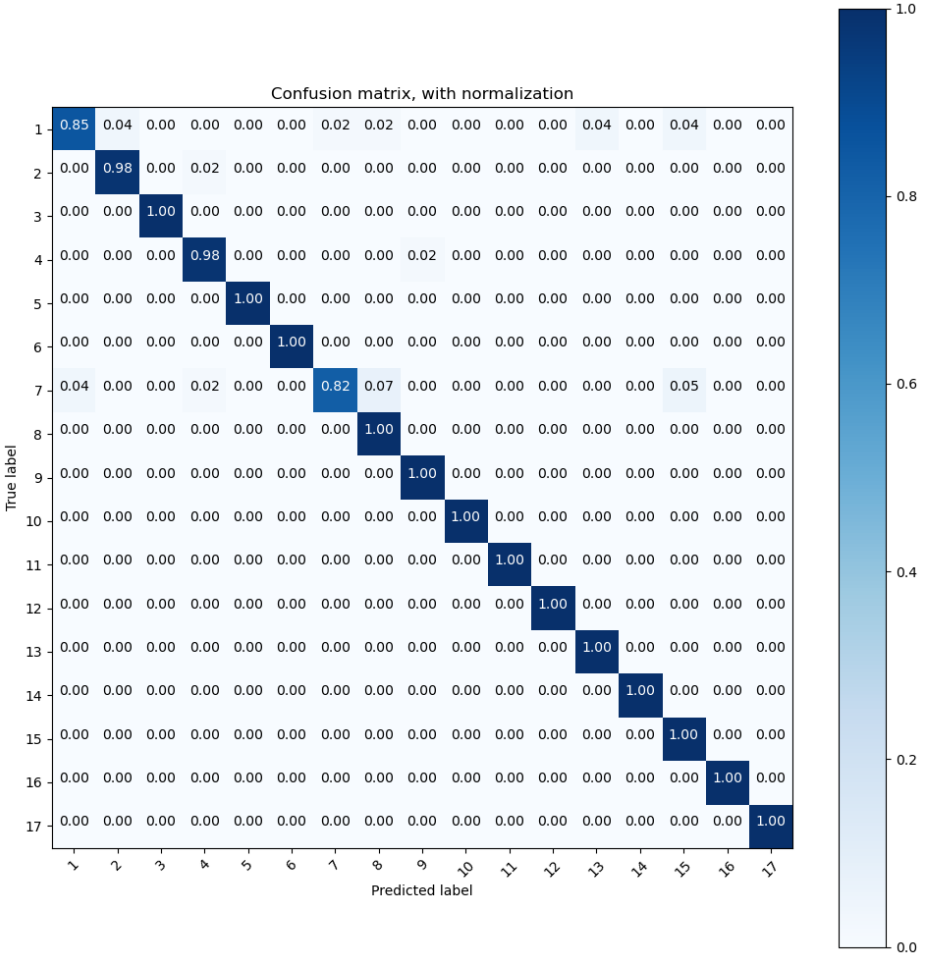

Fig. 6. The confusion matrix for the CNN model

4.3 Model Explanation

This section provides the results associated with the CNN local interpretations by using the LIME technique. The instance 180 of class 8 and instance 400 of class 9 were chosen randomly for the local explanations. The LIME technique was able to present features that contributed the most to the target outcome of class 8 and 9, respectively.

https://doi.org/10.24191/mij.v6i2.9317

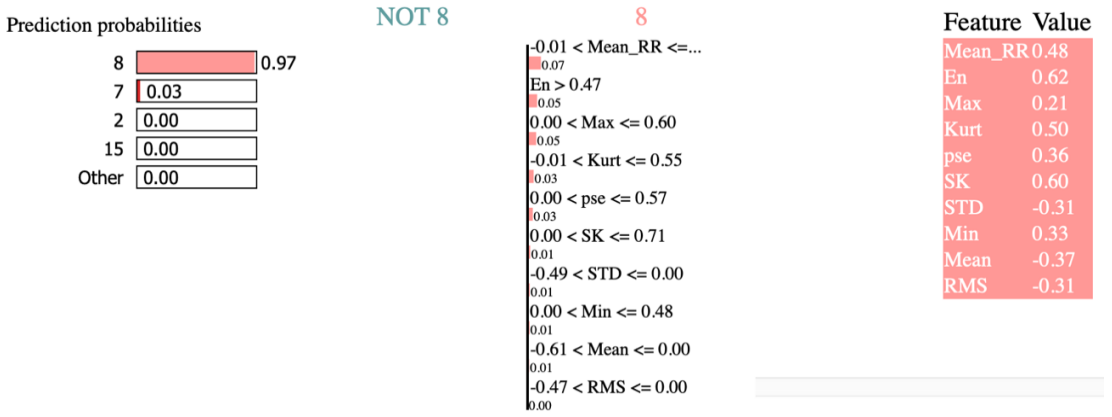

Fig. 7. Local interpretability prediction for instance 180 of class 8

The three graphs that showed each of the essential information about the ECG features and classification classes are presented in Fig. 7. The left graph shows that the instance 180 in the dataset showed the confidence interval indicating that this instance was 97% belonged to class 8, whereas only 3% stated that this instance belonged to class 7. The center graph showed the feature importance scores on the instance 180 with $Mean\_RR$ had 7% feature importance score, followed by En with 5%, and Max with 5%. The RMS had the lowest feature importance score. The right graph showed the 10 features and their respective values. The features highlighted in peach color contributed toward class 8.

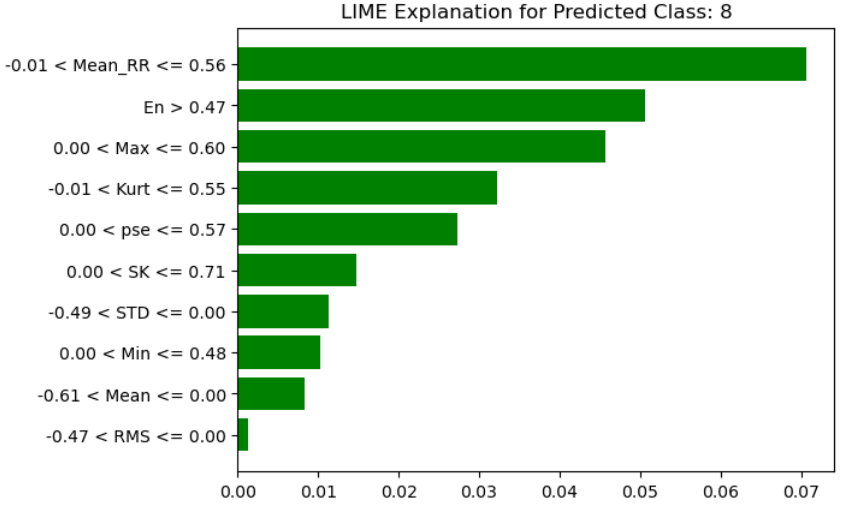

Fig. 8. The range of local interpretability prediction for instance 180 of class 8

Fig. 8 shows the bar plot that indicated the range of local interpretability predictions, on instance 180 for class 8. More specifically, the $Mean\_RR$ feature lied within the range $-0.01 < Mean\_RR <= 0.56$, the En feature was greater than 0.47, and Max feature lied within the range $0.00 < Max <= 0.60$. The RMS feature, which had the lowest feature importance score lied within the range $-0.47 < RMS <= 0.00$.

https://doi.org/10.24191/mij.v6i2.9317

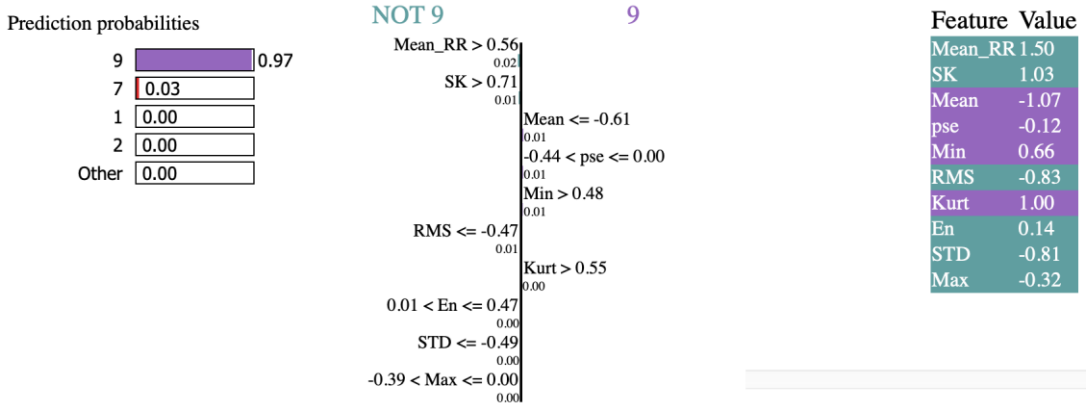

Fig. 9. Local interpretability prediction for instance 400 of class 9

Fig. 9 presents the local interpretability prediction, for instance 400. The left graph indicated that the instance 400 in the dataset showed the confidence interval stating that this instance was 97% belonged to class 9, whereas only 3% indicated this instance belonged to class 7. The center graph showed the feature importance scores on the instance 400. More specifically, the Mean had 1% feature importance score, followed by pse with 1%, and Min with 1%. The Kurt had the lowest feature importance score. In the right graph, the features highlighted in purple contributed toward class 9, whereas features highlighted in teal color contributed toward class not 9.

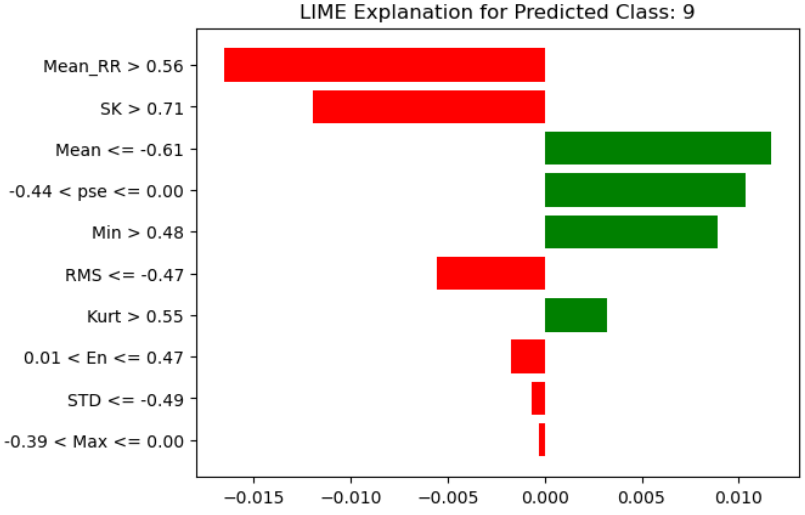

Fig. 10. The range of local interpretability prediction for instance 400 of class 9

The bar plot that indicated the range of local interpretability predictions, for instance 400 for class 9 is provided in Fig. 10. More specifically, the Mean feature was less than or equal to -0.61, the pse feature lies within the range $-0.44 < pse <= 0.00$, and Min feature was greater than 0.48. The Kurt feature, which had the lowest feature of importance, score was greater than 0.55.

Although the findings presented in Fig. 11 suggested that the LIME technique can provide local interpretability prediction, this technique did not offer consistency in terms of providing the same

https://doi.org/10.24191/mij.v6i2.9317

confidence interval and feature importance scores for similar instance. For instance, after second execution of the LIME technique on CNN model, the confidence interval indicated that the instance 400 was 99% belonged to class 9, whereas only 1% indicated this instance belonged to class 7. In contrast to the feature importance scores in Fig. 9, the RMS feature contributed to class not 9, while the Mean and Kurt features contributed to class 9. However, both findings in Fig. 9 and Fig. 11 indicated that the instance 400 belonged to class 9.

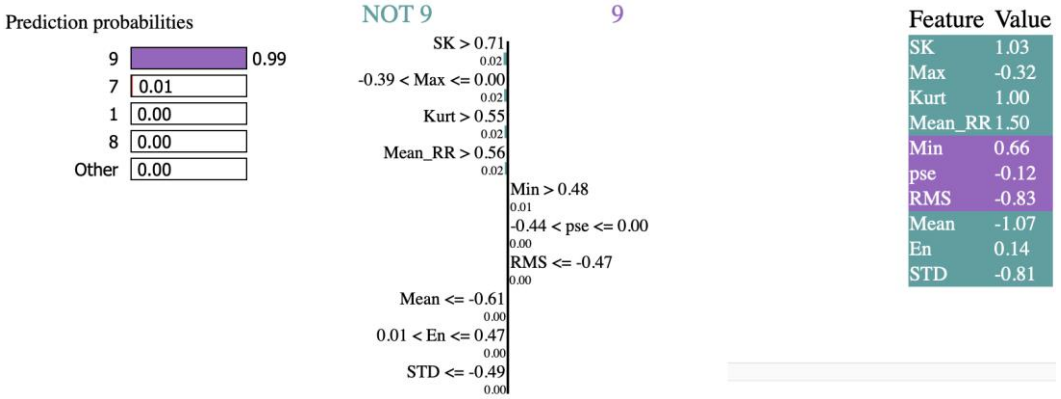

Fig. 11. Local interpretability prediction for instance 400 of class 9 (after second execution)

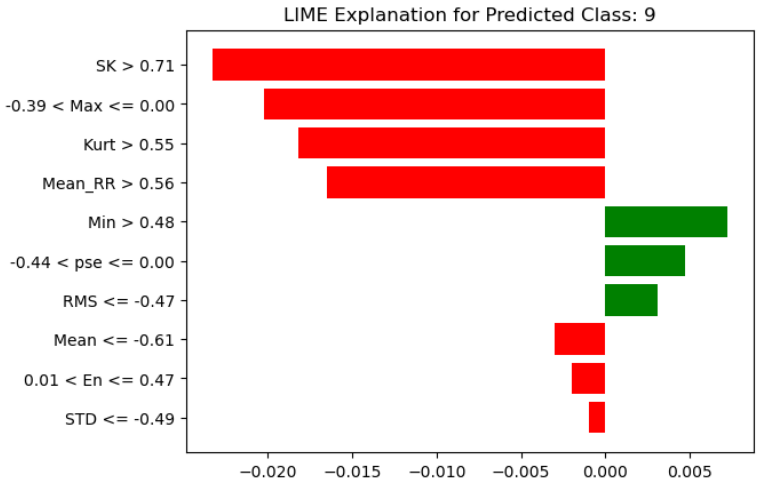

Fig. 12. The range of local interpretability prediction for instance 400 of class 9 (after second execution)

Although the LIME technique can be used to interpret the ECG classification model, clinical experts are still needed to evaluate the global and local explanations generated by this technique. This is because as clinical experts, they can identify potential biases or inconsistencies that may lead to harmful decisions as well as in ensuring that the explanations are aligned with the real-world understanding of the ECG arrhythmia.

https://doi.org/10.24191/mij.v6i2.9317

5. CONCLUSION

This study attempted to integrate the LIME technique with the CNN model for classifying 17 types of ECG arrhythmia. The ECG classification model consisted of five stages that included pre-processing, feature extraction, feature selection, classification and model explanation. Findings of the study indicated that the CNN model was able to attain precision, recall, F1-score of 97.00% with an overall accuracy of 99.00%. In addition, the findings of the study revealed that LIME technique managed to provide local explanation by determining which of the 10 features contributed to the class prediction of the specific instance. The integration of LIME technique in the study has contributed to a more interpretable CNN model, which initially possessed low interpretability in its decision-making process. Nevertheless, this study was limited by the absence of expert validation. For future research, it is recommended that researchers not only adopt the other alternative XAI approaches including SHAP or Grad-CAM but also validate their results by using clinical experts.

6. ACKNOWLEDGEMENTS/FUNDING

The authors wish to express their gratitude to the Faculty of Artificial Intelligence, Universiti Teknologi Malaysia (UTM), Kuala Lumpur for their support in this research.

7. CONFLICT OF INTEREST STATEMENT

The authors declare that this research was conducted without any personal, commercial, or financial conflicts and report no conflicts of interest with the funders.

8. AUTHORS' CONTRIBUTIONS

Adam Mohd Khairuddin conducted research, prepared the manuscript, and revised its content. Siti Armiza Mohd Aris validated the results reported in the article. Azizul Azizan designed the framework used in the research methodology. Noor Jannah Zakaria reviewed and edited the article for submission.

REFERENCES

Anand, A., Kadian, T., Shetty, M. K., & Gupta, A. (2022). Explainable AI decision model for ECG data of cardiac disorders. *Biomedical Signal Processing and Control, 75*, 103584. https://doi.org/10.1016/j.bspc.2022.103584

Azzem, Y. C. H., & Harrag, F. (2023). Explainable deep learning–based system for multilabel classification of 12-lead ECG. In *Proceedings of the 6th International Conference on Networking and Advanced Systems (ICNAS 2023)*. https://doi.org/10.1109/ICNAS59892.2023.10330482

Hou, Y., Liu, R., Shu, M., & Chen, C. (2023). An ECG denoising method based on adversarial denoising convolutional neural network. *Biomedical Signal Processing and Control, 84*, 104964. https://doi.org/10.1016/j.bspc.2023.104964

Katal, N., Gupta, S., Verma, P., & Sharma, B. (2023). Deep-learning-based arrhythmia detection using ECG signals: A comparative study and performance evaluation. *Diagnostics, 13*(24). https://doi.org/10.3390/diagnostics13243605

Khairuddin, A. M., & Ku Azir, K. N. F. (2021). Using the Haar wavelet transform and k-nearest neighbour algorithm to improve ECG detection and classification of arrhythmia. In H. Fujita, A. Selamat, J.

C.-W. Lin, & M. Ali (Eds.), *Advances and trends in artificial intelligence: From theory to practice* (pp. 310–322). Springer International Publishing.

Khairuddin, A. M., Ku Azir, K. N. F., & Rashidi, C. B. M. (2024). R-peaks and wavelet-based feature extraction on k-nearest neighbor for ECG arrhythmia classification. In N. S. Ahmad, J. Mohamad-Saleh, & J. Teh (Eds.), *Proceedings of the 12th International Conference on Robotics, Vision, Signal Processing and Power Applications* (pp. 523–529). Springer Nature Singapore.

Di Cesare, M., Bixby, H., Gaziano, T., Hadeed, L., Kabudula, C., Vaca McGhie, D., Mwangi, J., Pervan, B., Perel, P., Piñeiro, D., Taylor, S., & Pinto, F. (2023). *World Heart Report 2023: Confronting the world's number one killer*. World Heart Federation.

Narotamo, H., Dias, M., Santos, R., Carreiro, A. V., Gamboa, H., & Silveira, M. (2024). Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches. *Biomedical Signal Processing and Control, 93*, 106141. https://doi.org/10.1016/j.bspc.2024.106141

Ojha, J., Haugerud, H., Yazidi, A., & Lind, P. G. (2024). Exploring interpretable AI methods for ECG data classification. *ACM International Conference Proceeding Series*, 11–18. https://doi.org/10.1145/3643488.3660294

Salvi, R. (2024). An explainable CNN-BiGRU model for arrhythmia classification from ECG signals. In *Proceedings of the 2024 Conference on AI, Science, Engineering, and Technology (AIxSET 2024)* (pp. 316–319). https://doi.org/10.1109/AIxSET62544.2024.00054

Sathi, T. A., Jany, R., Ela, R. Z., Azad, A. K. M., Alyami, S. A., Hossain, M. A., & Hussain, I. (2024). An interpretable electrocardiogram-based model for predicting arrhythmia and ischemia in cardiovascular disease. *Results in Engineering, 24*, 103381. https://doi.org/10.1016/j.rineng.2024.103381

https://doi.org/10.24191/mij.v6i2.9317