MULTI-LEVEL STACKING ENSEMBLE LEARNING FOR ENHANCED ECG-BASED ARRHYTHMIA DETECTION

Huu-Cam Nguyen, Thu-Trang T. Nguyen, Hai-Long Nguyen, Ngoc-Anh T. Phung, and Anh-Thu Pham

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Abstract: Cardiovascular diseases (CVDs), characterized by abnormal heart function, are a leading cause of mortality worldwide, emphasizing the critical need for early detection and intervention. Electrocardiogram (ECG) monitoring has emerged as a vital tool for heart rhythm assessment in CVD prevention, with recent Machine Learning applications significantly improving diagnostic accuracy. This paper presents an innovative multi-level stacking ensemble learning approach that integrates predictions from diverse base models, including Random Forest, XGBoost, Support Vector Machine, K-Nearest Neighbor, AdaBoost, and Decision Tree algorithms, applied to 12-lead ECG data preprocessed using Neurokits. The base model outputs are synthesized through a Logistic Regression meta-model to enhance overall performance. Our methodology demonstrates exceptional effectiveness across multiple evaluation metrics, including accuracy, F1-score, precision, recall, and specificity, achieving over 97% accuracy in classifying various arrhythmia types. This research underscores the significant potential of ensemble learning methods in cardiac diagnostics, offering robust and comprehensive predictions for complex clinical scenarios.

Keywords: AdaBoost, Arrhythmia classification, CVDs, Decision Tree, Machine Learning, K-Nearest Neighbor, Random Forest, XGBoost, Electrocardiogram.

I. INTRODUCTION

Cardiovascular Diseases (CVDs), encompassing disorders of the heart and blood vessels, represent the leading cause of global mortality and disability. These conditions can progress silently, with even a single cardiac irregularity potentially causing significant health complications. The World Health Organization reported that CVDs claimed 4.2 million lives in Europe during 2019, accounting for 42.5% of all deaths in the region [1]. This concerning trend extends to developing nations, where CVDs impose an increasingly heavy burden, with low and middle-income countries bearing 80% of the global CVD impact [2]. Consequently, early detection and diagnosis have become crucial components of effective treatment strategies.

Moreover, electrocardiogram (ECG) analysis has been widely adopted by both cardiac specialists and general practitioners in detecting abnormal heartbeat [3]. Its popularity stems from its ability to provide rapid, cost-effective results, making it particularly valuable in resource-limited settings. Recently, the advent of Machine Learning (ML) and Artificial Intelligence (AI) has revolutionized ECG signal processing for arrhythmia detection. An application of AI called AI-ECG studies demonstrates enhanced diagnostic accuracy and efficiency while maintaining high standards of patient care. Furthermore, AI/ML integration with smart wearable devices has created accessible, economical monitoring solutions, extending cardiac care to remote areas [4].

In electrocardiography, electrode usage is a factor in examining cardiac activity from multiple perspectives, ensuring comprehensive and reliable measurements. While numerous arrhythmia classification studies have employed single or dual-lead ECG systems due to their simplicity, cost-effectiveness, and minimal patient interference, these approaches have inherent limitations. Single-lead systems notably fail to provide a complete cardiac assessment, as certain conditions are only detectable through twelve-lead monitoring [5]. Chen's research corroborates this, demonstrating that twelve-lead systems yield more comprehensive diagnostic results [6]. Given these significant advantages, numerous researchers have focused on twelve-lead systems for cardiac arrhythmia classification. Notable contributions include Yildirim et al.'s implementation of deep neural networks (DNN) for feature extraction and Long Short-Term Memory (LSTM) for sequence learning, analyzing 10,000 patient records [7]. Additionally, Aziz et al. [8] developed an innovative approach using Discrete Wavelet Transform (DWT) for artifact removal, followed by a hybrid Support

Contact author: Huu-Cam Nguyen
Email: camnh@ptit.edu.vn
Manuscript received: 10/2024, revised: 11/2024, accepted: 12/2024.

Vector Machine (SVM) and Multi-Layer Perceptron (MLP) model for anomaly detection in processed cardiac signals. Another work from Chopannejad et al. [9] emphasized using a combination of two neural networks in arrhythmia classification. He extends CNN models with Recurrent Neural Network (RNN), Bi-directional LSTM (BiLSTM), Bi-directional GRU (BiGRU), and attention mechanism to enhance learning. At first, the author suggests a new method to balance 12-lead ECG data called SMOTE-Tomek. Then, CNN extracts spatial features even when the input signal has noise, while BILSTM and BiGRU are combined to capture sequential patterns in forward and backward directions to enhance prediction accuracy with faster convergence. Finally, data is put into a multi-head self-attention mechanism to capture a broader range of features. The result achieves high accuracy in various leads in classifying multi-classes of abnormal heart rhythms.

Among advanced machine learning techniques such as bagging and boosting, which use one model with average or majority voting, stacking ensembles utilize different classification algorithms and leverage aggregate methods to combine the strength of various base model results [10]. The stacking method may be used in various machine learning issues such as classification, regression, and time-series forecasting. The primary goal of stacking is to produce a better outcome performance than a single model alone.

This paper proposes a multi-level stacking ensemble method for advanced ECG arrhythmia classification. The hierarchical integration of diverse machine learning algorithms enables comprehensive pattern recognition in ECG data, yielding exceptional performance across multiple metrics, including accuracy, precision, and specificity. The achieved accuracy exceeding 97% positions this ensemble approach as a robust diagnostic tool for early cardiac abnormality detection, potentially transforming clinical decision-making processes.

II. DATA PROCESSING

In this study, the Shaoxing database [11] is selected as the data source for training and testing. This ECG dataset contains recordings from over 10,000 patients across 12-lead. Each recording lasts 10 seconds and is sampled at a frequency of 500 Hz. The data has been meticulously annotated by clinical experts, including 11 major heart rhythms and 67 additional cardiovascular conditions, providing a comprehensive foundation for the development and optimization of machine learning models. This is one of the largest ECG datasets available, making it ideal for studies on arrhythmia classification and cardiac disease analysis.

Researchers worldwide developed methods to extract ECG features, such as Discrete Wavelet Transform (DWT) utilization by Aziz et al [8] or SMOTE-Tomek by Chopannejad et al. [9]. However, the algorithms are not packaged which is hard for other researchers who lack time and experience to re-implement. Moreover, the proposed methods are limited to a single field (for example, ECG) and incapable of handling other cases. Therefore, many neurosciences need to install different packaged software to process multimodal data inconveniently. Distributed scripts also satisfy the following criteria: Code must be well-documented and well-organized, sharable, and easy-to-use programming methods that can be reproducible instantly in several circumstances [12].

NeuroKit2 solves the above problems by providing a free and robust Neurophysiological Signal Processing toolkit for biosignal processing routines, which is easy to approach even for novices and inexperienced users. Neurokit2 adapts to various cases, including ECG, PPG (Photoplethysmography), RSP (Respiration Measures), PPG/BVP (Photoplethysmography), EMG (ElectromyoGraphy), etc [13]. It aims at clear, concise, and well-documented functionalities to create an automated pipeline for data preprocessing and analysis, such as signal simulation, important features detection, such as detecting P, Q, R, S, and T peaks in ECG diagrams and plotting, such as ECG visualization. In addition, the library allows researchers to select not only a whole pipeline but a part of a whole package, such as finding individual peaks to adapt to specific requirements. Neurokit2 is aimed at a reliable package that has been mentioned in notable publications, including QT intervals by deep learning [14], HRV estimation pipeline [15], and generating muscle function parameters from images [16]. With these huge advantages, the NeuroKit2 library was employed to process the ECG signals from this dataset. NeuroKit2 is famous in journal publications for its well-known open-source tool in the field of physiological signal analysis [17]. It aids in noise removal and key extraction features such as P, R, and T peaks, as well as PR, QT, and ST intervals, which is crucial for heart rhythm analysis and abnormalities detection. Key features extracted include the number of R-peaks, the mean and variance of the RR intervals, and the ratio of P to R peaks. Features with insufficient data or a high proportion of missing values were discarded to ensure accuracy.

After extraction, the features were normalized to a [0,1] range using the MinMaxScaler algorithm, facilitating the training of machine learning models. The data was divided into 80% for training and 20% for testing while maintaining the label distribution of arrhythmias across both subsets to improve model performance and generalization.

III. METHODOLOGY

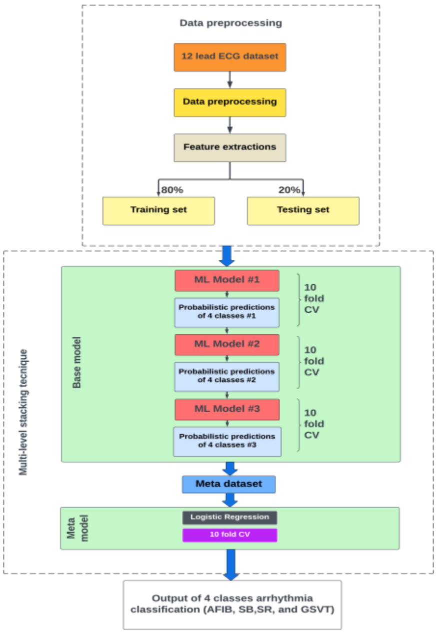

Figure 1. Workflow of multi-level stacking ensemble model over 12-lead ECG data

Figure 1 demonstrates our proposed multi-level stacking ensembled learning workflow against a 12-lead ECG arrhythmia classification. This technique combines the base-model predictions as meta-features and uses them as inputs for meta-model training. The meta-model then decides the final ECG arrhythmia prediction to improve the results generated in the base model [18]. The dataset is split into training and test sets; each set is trained independently, and predictions made at the previous level serve as input at the next level in the base- model. Our base-model has multiple levels that utilize diverse Machine Learning algorithms including AdaBoost (Ada), Random Forest (RF), XGBoost (XGB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Decision Tree (DT). Each base level in the base model applies 10-fold cross-validation. The predictions made at the final base level are used to create a meta-dataset. The meta-model combines predictions from the base-model and trains in the meta-dataset to make the final decision and compare it against the test set to assess correctness. Finally, our model performance is evaluated using various metrics to show how effective the model is regarding ECG arrhythmia categories to prove a more accurate and reliable diagnostic tool.

A. Fundamental of Machine Learning algorithms

To achieve optimal performance, the following algorithms are employed for multi-level stacking. Each base algorithm offers distinct predictive perspectives, enabling the meta-model to generate more comprehensive and accurate predictions through aggregation. A brief explanation of the selected algorithms is provided below.

- Random Forest is an ensemble algorithm based on multiple decision trees, improving accuracy by aggregating predictions from individual trees. With randomness in both sampling and feature selection, RF reduces the risk of overfitting and performs well on complex datasets. This model is employed across multiple levels to ensure stability and high predictive performance.
- XGBoost is a powerful boosting algorithm that focuses on difficult samples to incrementally improve predictions with each iteration. With parallel processing and optimized memory usage, XGB is highly effective on large and complex datasets. Using XGB at multiple levels enhances overall accuracy and system performance.
- Support Vector Machine classifies data by finding the optimal hyperplane that separates classes with the maximum margin. Through the use of kernels, SVM efficiently handles non-linear data. This model is deployed at various levels to maintain accuracy in complex classification tasks.
- K-Nearest Neighbor bases its predictions on the closest k points in the feature space. It is simple and easy to implement but can be slow on large datasets. In this system, KNN is applied at intermediate and meta levels to provide additional insights into the relationships between data points.
- Decision Tree divides data based on specific features to form decision branches, with results derived from predefined conditions. Although easy to interpret, DT can be prone to overfitting without proper tuning. In this system, DT is utilized at the meta-level to aggregate predictions from the base models.
- AdaBoost is a method that belongs to ensemble learning which transform from multiple weak learners to more robust, strong learners. AdaBoost's principle is model errors trained in training dataset is rectified in the next model. The procedure continues until errors are minimized and model predicts correctly to reach final output.

B. Base Model selection and Training

The multi-level stacking system is implemented across three levels (level-0, level-1, and level-2). Each level contains a set of machine learning models aimed at optimizing prediction performance. The selected base models include ensemble, linear, and non-linear algorithms, leveraging diverse approaches to enhance classification accuracy.

- Level-0 Layer: At the initial layer, models are trained on the original data to generate the first set of predictions. The models employed include RF, XGB, KNN, and SVM, providing different perspectives on the data and generating diverse predictions for the following layers.
- Level-1 Layer: The predictions from level-0 models are passed to level-1, where additional models are trained to

aggregate information from the previous layer. In this layer, RF, SVM, XGB, and K-Nearest Neighbors (KNN) are reused. DT is introduced. The repetition of models across multiple layers ensures system stability and improves generalization.

- Level-2 Layer: In the final layer (level-2), predictions from level-1 are used to train another set of base models, which generate the final output. This layer includes DT, and KNN, alongside RF, SVM, and XGB, which are carried over from the previous layers. AdaBoost is added to this layer to boost performance. Although level-2 builds upon the predictions from the previous layer, it remains focused on refining the base models to enhance the system's overall classification performance. The integration of multiple models across different levels ensures the system achieves optimal performance in classification tasks.

C. Meta model training and validation

The meta-model uses an algorithm different from the other six algorithms from the meta-dataset obtained from the base-model. This meta-model is used to learn how to optimally choose the perfect configuration to improve efficiency and accuracy. The training data for the meta-model contains a base-model prediction, along with true class labels.

The training data, along with classes, are supplied for the Machine Learning algorithm to predict results in the test set. Here, Logistic Regression is selected as a meta-model classifier. Ten-fold cross-validation is applied at this level to ensure the algorithm learns a generalized way of combining data and preventing overfitting. This results in higher performance capabilities rather than utilizing a single model in terms of heart arrhythmia classification.

D. Classification Model Evaluation

Six machine learning algorithms in the base model and one algorithm in the meta-model are studied to classify heart arrhythmia problems. The performance of a multi-level stacking ensemble method is evaluated using various metrics such as accuracy, F1-score, Precision, Recall, and Specificity criteria. The arrhythmia class is chosen based on the performance of the proper stacking model ensemble settings.

IV. SIMULATION RESULTS AND DISCUSSION

A. Performance Evaluation Metrics

Five key metrics are utilized to measure the effectiveness of our algorithms, namely Accuracy, F1-score, Precision, Recall, and Specificity. Accuracy refers to the ratio of correctness in both true positive and true negative over total predictions. F1-score represents harmony between precision and recall, which is useful when facing imbalanced dataset input. Precision is the number of actual positive instances in which to view how relevant data are. Recall models all relevant positive instances to answer the question about how many were correctly predicted. Specificity measurement is calculated to measure how effectively the model is capable of avoiding false positives, where mistaking the opposite result leads to severe consequences.

Denote TP, TN, FP, and FN are True Positives, True Negatives, False Positives, and False Negatives respectively. In the context of ECG-based arrhythmia classification, True Positives are cases where the model correctly identifies an arrhythmia when it is present, True Negatives are cases where the model correctly identifies a normal heartbeat when there is no arrhythmia, while False Positives are cases where the model incorrectly identifies an arrhythmia when the heartbeat is normal, and False Negatives are cases where the model fails to detect an arrhythmia when it is present (This is a critical error in medical diagnostics as it can lead to undetected health risks). The key performance metrics are calculated as follows.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{4}$$

$$Specificity = \frac{TN}{TN+FP} \tag{5}$$

These metrics collectively help in understanding the performance of an algorithm, especially in cases where class imbalance might affect results.

B. Base Model

Data are fed into several levels of the base model, which generally uses RF, XGB, SVC, KNN, DT, and AdaBoost algorithms in 12-lead ECG classifications of 4 arrhythmia types containing AFIB, SB, SR, and GSVT. Prediction of each level is served as input of another layer. Table I summarizes the algorithms along with their hyperparameters tuning used.

TABLE I. MULTI-LEVEL ALGORITHMS HYPERPARAMETERS

Level	Algorithms	Hyper parameters
0	RF	criterion= 'log_loss', max_depth= 15, max_features= 'sqrt', n_estimators= 40
	XGB	gamma= 0,learning_rate= 0.05,max_depth= 5,min_child_weight= 1,n_estimators= 1000
	SVM	C= 15, gamma= 'scale', kernel= 'rbf', probability= True

	KNN	algorithm= 'auto', n_neighbors= 5, p= 1, weights= 'uniform'
1	SVM	C= 15, gamma= 'scale', kernel= 'rbf', probability= True
	RF	criterion= 'entropy', max_depth= 18, max_features= 'sqrt', n_estimators= 40
	XGB	gamma= 0, learning_rate= 0.05, max_depth= 5,min_child_weight= 1,n_estimators= 1000
	KNN	algorithm= 'auto', n_neighbors= 5, p= 1, weights= 'uniform'
	DT	criterion= 'gini',max_depth= 6,max_features= 'sqrt',splitter= 'best'
2	RF	criterion= 'log_loss', max_depth= 15, max_features= 'sqrt', n_estimators= 40
	XGB	gamma= 0, learning_rate= 0.05, max_depth= 5, min_child_weight= 1,n_estimators= 1000
	SVC	C= 15, gamma= 'scale', kernel= 'rbf', probability= True
	DT	criterion= 'gini', max_depth= 6, max_features= 'sqrt', splitter= 'best'
	KNN	algorithm= 'auto', n_neighbors= 5, p= 1, weights= 'uniform'
	AD	algorithm= 'SAMME.R', learning_rate=0.1, n_estimators=61

In Table I, various classification models such as RF, XGB, SVM, KNN, and DT are demonstrated. Across all models, the choice of the parameter focuses on regularization, limit overfitting, and balancing bias-variance.

Multi-level stacking creates a highly flexible and powerful model. The diversity of algorithms with fine-tuned hyperparameters enables the model to capture broader patterns, enhance generalization, and prevent overfitting. This approach leverages the strength of each algorithm to produce more accurate predictions to serve to meta-model.

Meta-model utilizes GridSearchCV to find the best combination of hyperparameter tuning and configurations for machine learning. In this circumstance, GridSearchCV will search through possible hyperparameter combinations to find the best configuration. The primary job of the base-model is to capture patterns and define relationships to bring diversity, while the meta-model combines them to maximize performance.

C. Performance of Multi-level Stacking Ensembles

This experiment uses a multi-level stacking ensemble learning technique with one level of meta-model where each base level trains data independently to produce result probability. The first uses a combination of RF, XGB, SVM, and KNN, while the next level appends SVM, RF, XGB, KNN, and DT. The last level applies five algorithms: RF, XGB, SVM, DT, KNN, and AdaBoost. Characteristics of each level are fed into the next as input features. Meta-model aggregates predictions from base-model to produce a final prediction using LogisticRegression with optimized parameters tuning. To measure performance, a Confusion matrix with metrics is applied in the meta-model to ensure the model's accuracy is comprehensively and deeply analyzed.

The model performance of the multi-level stacking ensemble is measured against four heart rhythm types of arrhythmia classes: AFIB, SB, SR, and GSVT, which are shown in Table II accordingly. It's clear from Table II that the model demonstrates high accuracy and reliability in predicting different rhythm types above 0.97. SB detection achieves the highest performance with an accuracy of 0.994 and 0.995 in the F1-score, reflecting minimal wrong classification. The high specificity and precision indicate high successful avoidance of false positives, which is important in healthcare treatments. Overall, the model is well-suited to classifying diverse ECG arrhythmia classes, ensuring reliable predictions under various heart conditions.

TABLE II. MULTI-LEVEL STACKING MODEL PERFORMANCE

Classification		**Parameters**				
		Accuracy	***F1***	***Precision***	***Recall***	***Specificity***
Individual class	*AFIB*	0.971	0.939	0.923	0.931	0.979
	GSVT	0.975	0.926	0.955	0.941	0.988
	SB	0.994	0.995	0.989	0.992	0.993
	SR	0.987	0.971	0.969	0.970	0.992
Average	*macro*	-	0.958	0.959	0.958	-
	micro	-	0.963	0.963	0.963	-
	weighted	-	0.963	0.963	0.963	-

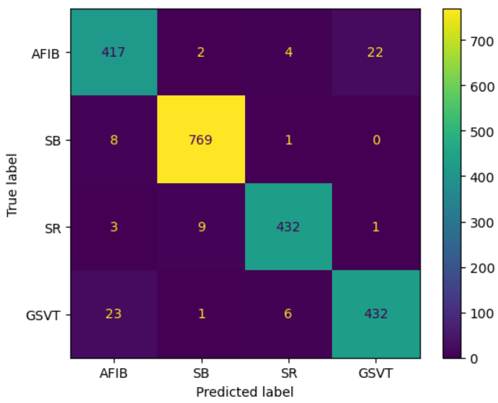

Figure 2. Confusion matrix of Multi-level stacking algorithm.

The confusion matrix, as shown in Figure 2, indicates model performance against four types of cardiac diseases, namely AFIB, SB, SR, and GSVT. The model accurately identifies 417 instances of AFIB, though it misclassifies a few with 2 SB, 4 SR, and 22 GSVT. This indicates high accuracy for AFIB. For SB, the model demonstrates strong performance, classifying 769 instances of SB. This suggests that SB is well differentiated by the model. The classification of SR is also successful with minor misclassification. In the case of GSVT, 432 instances are identified correctly; however, some notable numbers are identified, such as 23 for AFIB, 1 for SB, and 6 for SR. The misclassification between AFIB and GSVT shows these classes share some similar features that the model finds challenging to distinguish.

D. Discussion

The multi-level stacking ensemble model demonstrates exceptional performance in classifying four distinct cardiac rhythm types: Atrial Fibrillation (AFIB), Supraventricular Tachycardia (GSVT), Sinus Bradycardia (SB), and Sinus Rhythm (SR). The model achieves remarkable accuracy across all classes, with SB showing the highest accuracy at 0.994, followed by SR at 0.987, GSVT at 0.975, and AFIB at 0.971. Notably, SB classification exhibits outstanding performance across all metrics, with F1-score, precision, and recall all exceeding 0.989, suggesting highly reliable detection of SB cases. While AFIB shows slightly lower precision (0.923) compared to other classes, it maintains high specificity (0.979), indicating a strong capability to rule out false positives. The model's robust performance is further evidenced by the consistent macro, micro, and weighted averages of 0.958-0.963 across F1-score, precision, and recall metrics. GSVT classification demonstrates excellent specificity (0.988) with strong precision (0.955), indicating reliable identification of this challenging arrhythmia type. These results suggest that the ensemble approach effectively captures the distinctive characteristics of each arrhythmia class while maintaining high discrimination capability, making it a reliable tool for clinical applications. This suggests that using multiple levels in an ensemble contributes to more nuanced decision boundaries, thus contributing to overall accuracy. Previous studies in which stacking ensemble only comprises one level in base-model and one level for meta-learner demonstrates high accuracy, our strategy utilizes more than one level is less frequently explored to reproduce due to its high complexity. This recent finding is consistent with other research to prove that deeper ensemble structures can offer performance gain. The reason behind our success is our meta-model enhances the performance of the base-model prediction probabilities and leverages the use of cross-validation in balancing results. The result of this study is very advantageous for applications that require high reliability such as medical detections or weather forecasts. By improving accuracy, this solution provides a more reliable framework that assists in applications in which reliability is a key point. However, our method also faces some limitations while processing predictions. One limitation of this technique is that it requires high computational cost, which is not efficient when the dataset volume is large. This is obvious in calculating to select the best hyperparameter combination both in base-model and meta-model using GridSearchCV, and interactions between different models at different levels are very complex, which shows in multiple outputs for multiple levels. However, this issue could be addressed in the future with the help of parallel computing to minimize performance duration and maximize algorithm prediction generation. In addition, the number of algorithms used in multi-level stacking could be re-considered to yield a performance boost. Further research on the applicability of multi-level ensemble learning in a variety of different domains could also be investigated to solve complex predictive tasks.

V. CONCLUSION

In conclusion, this study demonstrates the potential of multi-level stacking ensemble method to enhance ECG classification results. By combining multiple machine learning algorithms in different layers, this approach captures patterns in ECG data and achieves high accuracy, precision, and specificity. With model accuracy beyond 97%, this ensemble method offers a reliable diagnostic tool for early detection, which impacts clinical decision-making in healthcare. This solution suggests promising applications in healthcare systems, particularly for early detection in lack-of-resource settings.

REFERENCES

[1] "Cardiovascular diseases." [Online]. Available: https://www.who.int/europe/news-room/fact-sheets/item/cardiovascular-diseases

[2] K. K. Teo and T. Rafiq, "Cardiovascular Risk Factors and Prevention: A Perspective From Developing Countries," *Canadian Journal of Cardiology*, vol. 37,

no. 5, pp. 733–743, May 2021, doi: 10.1016/j.cjca.2021.02.009.

[3] K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management," *Nat Rev Cardiol*, vol. 18, no. 7, pp. 465–478, Jul. 2021, doi: 10.1038/s41569-020-00503-2.

[4] A. H. Kashou, A. M. May, and P. A. Noseworthy, "Artificial Intelligence-Enabled ECG: a Modern Lens on an Old Technology," *Curr Cardiol Rep*, vol. 22, no. 8, p. 57, Aug. 2020, doi: 10.1007/s11886-020-01317-x.

[5] A. Abdou and S. Krishnan, "Horizons in Single-Lead ECG Analysis From Devices to Data," *Front. Signal Process.*, vol. 2, p. 866047, Apr. 2022, doi: 10.3389/frsip.2022.866047.

[6] Y.-J. Chen, C.-L. Liu, V. S. Tseng, Y.-F. Hu, and S.-A. Chen, "Large-scale Classification of 12-lead ECG with Deep Learning," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Chicago, IL, USA: IEEE, May 2019, pp. 1–4. doi: 10.1109/BHI.2019.8834468.

[7] O. Yildirim, M. Talo, E. J. Ciaccio, R. S. Tan, and U. R. Acharya, "Accurate deep neural network model to detect cardiac arrhythmia on more than 10,000 individual subject ECG records," *Comput Methods Programs Biomed*, vol. 197, p. 105740, Dec. 2020, doi: 10.1016/j.cmpb.2020.105740.

[8] S. Aziz, S. Ahmed, and M.-S. Alouini, "ECG-based machine-learning algorithms for heartbeat classification," *Sci Rep*, vol. 11, no. 1, p. 18738, Sep. 2021, doi: 10.1038/s41598-021-97118-5.

[9] S. Chopannejad, A. Roshanpoor, and F. Sadoughi, "Attention-assisted hybrid CNN-BILSTM-BiGRU model with SMOTE–Tomek method to detect cardiac arrhythmia based on 12 - lead electrocardiogram signals," *Digit Health*, vol. 10, p. 20552076241234624, Jan. 2024, doi: 10.1177/20552076241234624.

[10] G. Kunapuli, *Ensemble methods for machine learning*. Shelter Island, NY: Manning, 2023.

[11] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Sci Data*, vol. 7, no. 1, p. 48, Feb. 2020, doi: 10.1038/s41597-020-0386-x.

[12] D. Makowski *et al.*, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behav Res*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, doi: 10.3758/s13428-020-01516-y.

[13] "Overview — NeuroKit2 0.2.11 Documentation," https://neuropsychology.github.io/NeuroKit/introduction.html.

[14] R. Alam, A. D. Aguirre, and C. M. Stultz, "QTNet: Deep Learning for Estimating QT Intervals Using a Single Lead ECG," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, Jul. 2023, pp. 1–4. doi: 10.1109/EMBC40787.2023.10341204.

[15] M. G. Frasch, "Comprehensive HRV estimation pipeline in Python using Neurokit2: Application to sleep physiology," *MethodsX*, vol. 9, p. 101782, 2022, doi: 10.1016/j.mex.2022.101782.

[16] T. H. Farook, T. M. Haq, L. Ramees, and J. Dudley, "Deep learning and predictive modelling for generating normalised muscle function parameters from signal images of mandibular electromyography," *Med Biol Eng Comput*, vol. 62, no. 6, pp. 1763–1779, Jun. 2024, doi: 10.1007/s11517-024-03047-6.

[17] S. Zhang, C. Lian, B. Xu, Y. Su, and A. Alhudhaif, "12-Lead ECG signal classification for detecting ECG arrhythmia via an information bottleneck-based multi-scale network," *Inf Sci (N Y)*, vol. 662, p. 120239, Mar. 2024, doi: 10.1016/j.ins.2024.120239.

[18] L. Rokaḥ, *Ensemble learning: pattern classification using ensemble methods*, Second edition., no. volume 85. in Series in machine perception and artificial intelligence. Singapore: World Scientific Publishing Co. Pte. Ltd, 2019. doi: 10.1142/11325.

GIẢI PHÁP HỌC TẬP TỔNG HỢP NHIỀU LỚP CẢI THIỆN PHÁT HIỆN RỐI LOẠN NHỊP TIM DỰA TRÊN TÍN HIỆU ECG

Tóm tắt: Các bệnh tim mạch (CVDs) bao gồm các triệu chứng của bất thường của nhịp tim và là tác nhân gây ra nhiều ca tử vong trên toàn thế giới. Để phát hiện sớm và điều trị hiệu quả, điện tâm đồ (ECG) được sử dụng để theo dõi nhịp tim nhằm phòng ngừa các bệnh tim mạch. Gần đây với sự tiến bộ của khoa học công nghệ, việc kết hợp học máy trong chẩn đoán nhịp tim ngày càng trở nên phổ biến. Bài báo này đề xuất một phương pháp mới sử dụng mô hình học tập tổng hợp nhiều lớp trong đó kết hợp các kết quả của mô hình dự đoán cơ sở sử dụng các phương pháp Học Máy đa dạng bao gồm Rừng Ngẫu Nhiên, XGBoost, Máy Vector Hỗ trợ, K-Hàng xóm gần nhất, AdaBoost, và Cây Quyết Định trên tập dữ liệu ECG 12-lead xử lý bởi Neurokits. Kết quả của mô hình cơ sở được đưa vào mô hình meta sử dụng Logistic Regression để cải thiện hiệu suất tổng thể. Hiệu quả của phương pháp này được đo lường qua các chỉ số như độ chuẩn xác, F1-score, độ chính xác, độ nhạy, và độ đặc hiệu cho thấy độ chính xác của phương pháp trên 97% trong việc nhận diện các bệnh rối loạn nhịp tim. Công trình này nhấn mạnh tiềm năng của các phương pháp tập hợp trong điều trị tim mạch, cung cấp các dự đoán chính xác và toàn diện trong các tình huống lâm sàng phức tạp.

Từ khoá: AdaBoost, Phân loại rối loạn nhịp tim, Cây quyết định, Học máy, Mô hình học tập tổng hợp nhiều lớp, K-Hàng xóm gần nhất, Rừng ngẫu nhiên, XGBoost, Tín hiệu điện tâm đồ.

Huu Cam Nguyen received the B.E degree from Hanoi University in 2013, and the M.E of Information System Design of University of Central Lancashire, United Kingdom in 2016. He joined the Research Institute of Posts and Telecommunications in Vietnam. He is currently a lecturer in Telecommunications Faculty at PTIT. His research interests include Machine Learning, Bioinformatics, Blockchain and Information System Design.

Thu Trang T. Nguyen is a B.E. student in the Telecommunications Department at the Posts and Telecommunications Institute of Technology (PTIT) in Vietnam. Her research interests include machine learning, IoT systems, and hardware engineering.

Hai-Long Nguyen is currently a research student in Data and Intelligent Systems Laboratory at Posts and Telecommunications Institute of Technology (PTIT), Viet Nam. His research interests include machine learning, bioinformatics, and data science.

Ngoc Anh T. Phung is a B.E. student in Information Technology Department of Posts and Telecommunications Institute of Technology (PTIT) of Vietnam. Her research interests include machine learning, bio-informatics, network security.

Anh Thu Pham received B.E degree of Telecommunication engineering from Posts and Telecommunications Institute of Technology (PTIT), Viet Nam, in 2003, and M.E degree of Telecommunication engineering from Royal Melbourne Institute of Technology, Australia, in 2008. She received the Ph.D. degree in Telecom- munication engineering from PTIT, in 2010. Now, she is a lecturer in the Telecommunications Faculty of PTIT. Her research interests include networking, radio over fiber, and broadband wireless networks.