An Explainable Attention-Based Hybrid Learning Framework for the Multi-Class Classification of ECG-Based Arrhythmia and Cardiovascular Diseases

[1]Sudha R, [2]Dr.M.Selvapriya
[1]Research Scholar, [2]Research Supervisor
[12]Department of Computer Science
[12]Hindusthan College of Arts and Science, Coimbatore, Tamil Nadu, India
[1]Author1@xyz.com, [2]Author2@xyz.com, [3]Author3@xyz.com

Abstract—**Electrocardiogram (ECG) analysis plays a vital role in the early diagnosis of cardiac arrhythmias and cardiovascular diseases (CVDs), which remain the leading cause of global mortality. Recent advances in machine learning (ML) and deep learning (DL) have significantly improved automated ECG classification accuracy. However, most existing DL models function as black-box systems, limiting their adoption in clinical environments due to the lack of interpretability and transparency. This paper proposes an explainable attention-based hybrid deep learning and machine learning ensemble framework for ECG-based multi-class arrhythmia and disease classification. A CNN–BiLSTM architecture enhanced with an attention mechanism is employed for spatial–temporal feature extraction, while AdaBoost and Support Vector Machine (SVM) classifiers are integrated using a stacking ensemble strategy. Experiments conducted on the MIT-BIH Arrhythmia and PTB Diagnostic ECG datasets demonstrate an overall classification accuracy of 99.53%, outperforming existing ML and DL approaches. Explainable artificial intelligence (XAI) techniques such as Grad-CAM and SHAP are incorporated to provide clinically interpretable predictions, thereby improving trust and reliability in automated ECG diagnosis.**

Index Terms—**Electrocardiogram, Arrhythmia Classification, CNN–BiLSTM, Attention Mechanism, Ensemble Learning**

I. Introduction

Cardiovascular diseases (CVDs) are the most important cause of death globally, accounting for millions of fatalities every year [13]. Antiquated finding and continuous attention are vital for diminishing homicide and helping patient outcomes. The electrocardiogram (ECG) is an unobtrusive, cost-effective typical tool that reports the electrical activity of the heart and is broadly used for identifying arrhythmias and other cardiac diseases[1], [14].

Conventional ECG classification methods depend extremely on Deep Learning and classical machine learning classifiers such as decision trees, k-nearest neighbors, and support vector machines[15] [16]. But these paths attain moderate success, their performance is extremely dependent on expert-designed features and deficiencies robustness through diverse patient populations. Deep learning approaches, especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have established superior performance by impulsively learning discriminatory portraits out of raw ECG signals. [17],[18] [19].

In spite of their high accuracy, deep learning reproductions are often faulted for their black-box nature, which impedes clinical acquiescence[5][6]. Doctors involve translucent and demonstrates systems that can explain predictions in terms of physiologically suggestive ECG sections. To address this confines, explainable artificial intelligence (XAI) techniques such as SHAP and Grad-CAM have been imported to furnish sensitivity into model critical-thinking [6],[11], [12].

This work focuses to bridge the interval between precision and illustrate by offering an feasible hybrid framework that connects absorption-based deep learning with ensemble machine learning techniques. The primary benefactions of this paper are:

1. CNN–BILSTM architecture was used to extract both spatial and temporal features from raw ECG dataset [16].
2. Accumulating ensemble merging AdaBoost and SVM to enhance classification reliability and generalization [7][8][20].
3. Attainted high accuracy results on standard ECG dataset.

II. Related Work

Comprehensive research has been supervised on robotic ECG classification using machine learning and deep learning approaches. Early studies immersed on handcrafted highlights inferred from time-domain, frequency-domain, and wavelet-based illustrations merged with classical classifiers[15],[16].While closer routes accomplished logical accuracy, their reliance on realm skills defined scalability.

The prologue of deep learning marked a model shift in ECG assessment. CNN-based frameworks have been broadly accustomed for robotic feature extraction from raw ECG waveforms, establishing enhanced performance above convolutional methods [17],[21],[22]. Kiranyaz S Proposed patient-specific CNN models for real-time ECG classification, emphasizing the conclusiveness of deep frameworks in obtaining morphological patterns [17]. Hannun AY gained cardiologist-level arrhythmia discovery using deep CNNs trained on large-scale ECG datasets [18].

To model temporal reciprocities in ECG waves, recurrent frameworks such as LSTM and BiLSTM networks have been examined [16],[23].CNN–LSTM and CNN–BiLSTM frame works merges spatial feature extraction with temporal modeling, associated in improved classification performance. newly, attention mechanisms have been blende into deep models to concentrated on professionally major ECG parts, beyond enhancing differential ability.

Ensemble learning methods such as AdaBoost, bootstrap aggregating, and stacked generalization have been engaged to enhance agility and conception by merging many classifiers[7],[8],[20]. Stacking ensembles, in particular, have demonstrated superior performance by leveraging the strengths of heterogeneous models[7].

In similar, the booming request for feasible AI in clinical care has accompanied to the espousal of post-hoc explanation methods such as SHAP (Shapley Additive exPlanations) and Grad-CAM (Gradient-weighted Class Activation Mapping) [6],[11][24]. These methods furnish optical and quantitative illustrations, substance deep learning besides clear and clinically adequate.

Recent researches have progressively immersed on enhancing ECG classification interpretation by incorporating attention apparatuses with deep neural networks. Attention-based CNN–LSTM and CNN–BiLSTM frameworks strongly highlight distinguishing relevant ECG sections such as QRS complexes and ST segments, thereby enhancing class separation in multi-class scenarios.

In addition to architectural advances, ensemble learning has been researched to enhance robustness and conception in ECG classification systems. collecting ensembles unification heterogeneous classifiers have produced dominant performance as opposed to single-model approaches [7]. Freund and Schapire's boosting framework and Vapnik's statistical learning theory furnish powerful theoretical foundations for ensemble-based ECG classification [8],[20]. Moreover, modern research emphasizes a vital of answerable AI in clinical settings, highlighting that interpretation alone is insufficient without translucency and interpretability [6],[20]. These researches persuade the incorporation of attention-based DL, ensemble classification, and feasible within a integrated ECG specific framework.

III. Materials and Methods

3.1 Datasets

The proposed framework is evaluated using two widely used benchmark datasets: the MIT-BIH Arrhythmia Database and the PTB Diagnostic ECG Database[1] .

Data set found from (https://www.kaggle.com/datasets/erhmrai/ecg-image-data/data & This study utilizes a comprehensive ECG image dataset derived from the **PhysioNet MIT-BIH Arrhythmia Database**. The ECG dataset includes **109,446 samples** sampled at **125 Hz**, categorizing major five distinctive beat categories: Normal beat categorised as **0**, Supraventricular beat (S) categorised as **1**, Premature Ventricular Contraction **(V)** beat categorised as **2**, Fusion **(F)** beat categorised as **3**, and Unclassifiable beats **Q** are categorised as **4**. Each arrhythmia indication has been propagated from signal data into an ECG image format. For model development and validation, the data was partitioned into an **80/20 split**, yielding 80% for training and 20% for testing.

3.2 Preprocessing

Raw ECG waves are rendered to shed barrier, then standardized and partial to assure all data inputs participates a compatible length and signal strength [15],[19]. These tracks improve model agility and degrade inter-patient variability.

The rectitude of ECG classification depends highly on well preprocessing to help the things of muscle artifacts, power-line interference, and baseline wander. enforcing high-pass and band-pass filtering techniques adequately hermits the heart beat signal from these noise sources, thereby enhancing the accuracy of latest analytical models.[15][19]. Pertaining normalization assures that all ECG waves participate a usual scale. This precludes amplitude takes from biasing the feature learning method, providing for more reliable classification across various datasets [1],[14].

The transformation of uninterrupted ECG waves into discrete, fixed-length gaps is attained among segmentation, a critical step for deep learning compatibility. This method ensures that diagnostic landmarks—specifically the P wave, QRS complex, and T wave—are preserved, providing the structural detail necessary for accurate arrhythmia disease identification [16],[17].

3.3 Attention-Based CNN–BiLSTM Architecture

CNN layers are absorbed to extract spatial features depicting morphological features of ECG signals [17],[21]. The extracted feature maps are federal into a BiLSTM architecture obtain long-term temporal dependencies in both directions (forward and backward)[16], [23]. An attention mechanism is blended to take evolved weights to clinically pertinent ECG elements, enabling the model to concentrate on informative regions.

The CNN element of the proposed architecture is deliberate to impulsively learn hierarchical spatial features from ECG signals, obtaining waveform morphology and local temporal patterns. combined convolutional layers enable the extraction of low-level and high-level features, while pooling layers demote dimensionality and computational complexity. This spatial feature learning capability has been shown to outperform handcrafted feature-based methods in large-scale ECG classification tasks[18],[22].

The BiLSTM layer supplements CNN-based spatial learning by modeling long-term temporal dependencies inherent in ECG waves [16],[23]. Unlike unidirectional LSTMs, BiLSTMs process sequences in both forward and backward directions, allowing the capture of contextual information from past and future cardiac cycles. The incorporation of an absorption mechanism further improves performance by dynamically weighting time steps according to their diagnostic relevance, allowing the model to concentrate on clinically substantial ECG fields.

3.4 Stacking Ensemble Classification`

The deep feature representations obtained from the attention-based CNN–BiLSTM model are classified using a stacking ensemble framework. AdaBoost and SVM classifiers serve as base learners, while a meta-classifier combines their outputs to generate the final prediction[7],[8],[20]. This ensemble strategy improves classification accuracy and reduces overfitting.

Ensemble learning improves predictive performance by combining multiple classifiers with complementary strengths. In the proposed framework, AdaBoost and SVM are selected as base learners due to their proven effectiveness in ECG classification tasks [15],[8]. AdaBoost iteratively emphasizes misclassified samples, improving sensitivity for minority classes, while SVM provides strong generalization through margin maximization [20].

The stacking ensemble framework integrates the outputs of base learners using a meta-classifier, enabling more robust and stable predictions[7]. Prior research has demonstrated that stacking ensembles consistently outperform individual classifiers in complex biomedical classification problems [20]. By leveraging both deep feature representations and ensemble learning, the proposed system achieves superior robustness and reduced overfitting compared to standalone models.

3.5 Explainability Module

To enhance interpretability, Grad-CAM is used to visualize salient regions in ECG signals influencing model predictions, while SHAP values quantify feature contributions [6],[11],[12]. These explanations provide clinicians with insights into the decision-making process, increasing trust and usability.

Explainability is a critical requirement for the deployment of AI systems in clinical practice. Grad-CAM provides visual explanations by highlighting regions of ECG signals that contribute most to the model's predictions, enabling clinicians to verify whether the model focuses on physiologically meaningful patterns [12]. Such visual explanations have been shown to improve clinician trust and facilitate error analysis in deep learning-based diagnostic systems [6],[20].

SHAP complements Grad-CAM by offering quantitative feature attribution, measuring the contribution of individual features to each prediction[11]. SHAP-based explanations provide both global and local interpretability, enabling the identification of dominant ECG features influencing classification outcomes. The combination of Grad-CAM and SHAP ensures comprehensive interpretability, aligning the proposed framework with regulatory and ethical requirements for medical AI systems [6],[24].

IV. Experimental Results

By achieving an overall classification precision of **99.53%,** the proposed framework surpasses both traditional (non-standard) ML models and standalone DL architectures [21],[19]. Accuracy, recall, and F1-score metrics indicate consistent performance across all ECG classes (Normal beat or Abnormal beat). Class-wise analysis demonstrates effective discrimination between normal and abnormal heart beats.

Comparative evaluation with existing studies reveals that the integration of attention mechanisms and ensemble learning significantly improves performance. [25]. The results confirm the robustness and generalizability of the proposed approach.

The outcomes indicate that the suggested technique is both robust and applicable.

The proposed framework is consistently superior to traditional ML models and standalone DL architectures in all ECG types, as demonstrated by a detailed performance review [21]. High precision and recall values indicate effective discrimination between normal and pathological rhythms, including challenging arrhythmia classes. Comparative analysis confirms that attention mechanisms and ensemble learning contribute significantly to performance improvements.

Additional research indicates that the stacking ensemble minimizes class imbalance effects and enhances resilience against unwieldy signals [7],[8]. The use of confusion matrix analysis demonstrates the efficiency of the attention-based feature learning and ensemble classification strategy, as it shows that misclassification rates between arrhythmias with morphologically similar sounds are much lower. The outcomes demonstrate that the proposed method is appropriate for ECG diagnostics in real-world scenarios.

V. Discussion

A combination of attention-based deep learning and ensemble machine learning has been used to develop a more efficient proposed framework. This is noteworthy. The model's attention mechanism enables it to concentrate on diagnostically relevant ECG segments, which facilitates feature discrimination. The stacking ensemble increases strength and reduces the effectiveness of individual classifiers [7].

The experimental findings underscore the importance of integrating attention mechanisms with deep learning architectures for ECG classification. By selectively focusing on diagnostically relevant ECG segments, the proposed framework reduces the influence of noise and irrelevant signal components, resulting in improved classification accuracy and stability.

Moreover, the incorporation of ensemble learning and explainable AI addresses two critical challenges in medical AI: robustness and transparency. Ensemble learning enhances generalization and mitigates overfitting, while XAI techniques provide clinically interpretable insights into model predictions [11],[12],[7]. The combined effect of these components positions the proposed framework as a practical and trustworthy solution for clinical ECG analysis.

VI. Conclusion

This paper presents an explainable attention-based hybrid deep learning and machine learning ensemble framework for ECG-based multi-class arrhythmia and cardiovascular disease classification. The proposed model achieves state-of-the-art performance while maintaining clinical interpretability. Future work will focus on real-time deployment and validation on larger, multi-center ECG datasets.

References

[1] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," Circulation, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.CIR.101.23.E215;SUBPAGE:STRING:FULL.

[2] "Cardiovascular diseases (CVDs)." Accessed: Jan. 31, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[3] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," Inf. Sci. (N. Y)., vol. 405, pp. 81–90, Sep. 2017, doi: 10.1016/J.INS.2017.04.012.

[4] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," Comput. Biol. Med., vol. 122, p. 103801, Jul. 2020, doi: 10.1016/J.COMPBIOMED.2020.103801.

[5] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," Comput. Methods Programs Biomed., vol. 161, pp. 1–13, Jul. 2018, doi: 10.1016/J.CMPB.2018.04.005.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?' Explaining the predictions of any classifier," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 13-17-August-2016, pp. 1135–1144, Aug. 2016, doi: 10.1145/2939672.2939778;JOURNAL:JOURNAL:ACMCONFERENCES;PAGEGROUP:STRING:PUBLICATION.

[7] L. Breiman, "Stacked regressions," Mach. Learn., vol. 24, no. 1, pp. 49–64, 1996, doi: 10.1007/BF00117832/METRICS.

[8] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/JCSS.1997.1504.

[9] N. Donges et al., "A Guide to RNN: Understanding Recurrent Neural Networks and LSTM," Journal of Infrastructure Systems, vol. 104, no. 2, pp. 294–301, 2019, Accessed: Jan. 31, 2026. [Online]. Available: https://books.google.com/books/about/The_Nature_of_Statistical_Learning_Theor.html?id=EqgACAAAQBAJ

[10] H. K. Damodaran, R. K. Tripathy, and R. B. Pachori, "Time-frequency-domain deep representation learning for detection of heart valve diseases using PCG recordings for IoT-based smart healthcare applications," Signal Processing Driven Machine Learning Techniques for Cardiovascular Data Processing, pp. 149–165, Jan. 2024, doi: 10.1016/B978-0-44-314141-6.00015-3.

[11] S. M. Lundberg, P. G. Allen, and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Adv. Neural Inf. Process. Syst., vol. 30, 2017, Accessed: Jan. 31, 2026. [Online]. Available: https://github.com/slundberg/shap

[12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization," 2017. Accessed: Jan. 31, 2026. [Online]. Available: http://gradcam.cloudcv.org

[13] "Cardiovascular diseases (CVDs)." Accessed: Jan. 31, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[14] G. B. Moody and R. G. Mark, “The impact of the MIT-BIH arrhythmia database,” IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45–50, 2001, doi: 10.1109/51.932724.

[15] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, “Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network,” Inf. Sci. (N. Y)., vol. 405, pp. 81–90, Sep. 2017, doi: 10.1016/J.INS.2017.04.012.

[16] Ö. Yildirim, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” Comput. Biol. Med., vol. 96, pp. 189–202, May 2018, doi: 10.1016/J.COMPBIOMED.2018.03.016.

[17] S. Kiranyaz, T. Ince, and M. Gabbouj, “Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks,” IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675, Mar. 2016, doi: 10.1109/TBME.2015.2468589.

[18] A. Y. Hannun et al., “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” Nature Medicine 2019 25:1, vol. 25, no. 1, pp. 65–69, Jan. 2019, doi: 10.1038/s41591-018-0268-3.

[19] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, “Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review,” Comput. Biol. Med., vol. 122, p. 103801, Jul. 2020, doi: 10.1016/J.COMPBIOMED.2020.103801.

[20] “The Nature of Statistical Learning Theory - Vladimir Vapnik - Google Books.” Accessed: Mar. 19,2026.[Online].Available: https://books.google.co.in/books?hl=en&lr=&id=EqgACAAAQBAJ&oi=fnd&pg=PR7&dq=The+Nature+of+Statistical+Learning+Theory.+Springer,+1995&ots=g5K0nC8173&sig=Y4_2Kdt4GYnzgX7LChDE7OoprFM&redir_esc=y#v=onepage&q=The%20Nature%20of%20Statistical%20Learning%20Theory.%20Springer%2C%201995&f=false

[21] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, “Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks,” Jul. 2017, Accessed: Nov. 12, 2025. [Online]. Available: https://arxiv.org/pdf/1707.01836

[22] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, “ECG heartbeat classification: A deep transferable representation,” Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018, pp. 443–444, Jul. 2018, doi: 10.1109/ICHI.2018.00092.

[23] “(PDF) Hybrid CNN-LSTM model for ECG-based arrhythmia detection in internet of medical things.” Accessed: Mar. 19, 2026. [Online]. Available: https://www.researchgate.net/publication/398426393_Hybrid_CNN-LSTM_model_for_ECG-based_arrhythmia_detection_in_internet_of_medical_things

[24] E. Tjoa and C. Guan, “A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 11, pp. 4793–4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.

[25] E. M. Lima et al., “Deep neural network-estimated electrocardiographic age as a mortality predictor,” Nature Communications 2021 12:1, vol. 12, no. 1, pp. 5117-, Aug. 2021, doi: 10.1038/s41467-021-25351-7