Hybrid Deep Learning Models for Multi-Class ECG Classification: Leveraging Time-Domain and Frequency-Domain Information

Muhammad Atlas Malik
Department of Computer Science
FAST NUCES Islamabad, Pakistan
i248020@isb.nu.edu.pk

Aqdas Bibi
Department of Computer Science
FAST NUCES Islamabad, Pakistan
i248048@isb.nu.edu.pk

Aieza Noor
Department of Computer Science
FAST NUCES Islamabad, Pakistan
i248021@isb.nu.edu.pk

***Abstract*—Electrocardiogram analysis represents a critical modality for early diagnosis of cardiovascular disorders. However, traditional machine learning and many deep learning models have been unable to generalize well across diverse patient populations or to integrate temporal and spectral characteristics of signals. This study designs several progressively advanced hybrid deep-learning architectures for multi-class ECG classification from the PTB-XL dataset. A total of four different architectures were developed, ranging from the simple FFT-based CNN to the multi-scale InceptionTime–Attention framework, which extracts both time-domain morphology and frequency-domain features. All the models were subjected to comprehensive preprocessing steps like resampling, noise filtering, normalization, windowing, and class balancing. Ablation studies showed significant improvements after every architectural enhancement, with Model-4 yielding the best performance: 92% accuracy and a F1 score of 0.92. It proves that the incorporation of multi-scale temporal features, together with spectral information and an attention mechanism, significantly enhances model interpretability and diagnostic reliability. This work highlights the importance of hybrid time–frequency deep learning approaches in delivering robust clinically applicable ECG classification systems.**

***Index Terms*—ECG, Deep Learning, Signals Classification**

I. Introduction

There is significant increase in rate of deaths worldwide due to Cardiovascular diseases. According to a survey by World Health Organization (WHO), about 17.9 million individuals die annually due to cardiovascular disease. Cardiovascular disease (CVD) is a general term for conditions affecting the heart or blood vessels. It's usually associated with a build-up of fatty deposits inside the arteries (atherosclerosis) and an increased risk of blood clots. Identifying CVD in its early stage can be significant in preventing and treating the disease effectively. Approximately half of all patients diagnosed with Heart Disease die within just 1-2 years, while merely 3% of the total budget for health care is deployed on treating heart disease. To predict heart disease multiple tests are required. Lack of expertise of medical staff may results in false predictions [1].

The examination of electrocardiographic signals (ECG) is one of the most important steps in identifying cardiac disorders. Extensive research on ECG signal identification has been carried out over the past several decades. An electrocardiogram is a commonly employed non-invasive physiological signal used for screening and diagnosing cardiovascular disease. In addition, the signal is used to search for pathological patterns corresponding to diseases such as arrhythmia, atrial fibrillation, ventricular fibrillation, and others. ECG analysis tools require knowledge of the location and structural pattern of the various segments (P-QRS-T) in the ECG recordings [2].

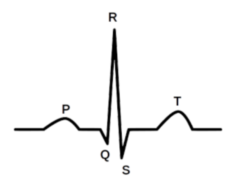


Fig. 1. The illustrative waveform of the ECG signal.

A. Problem Statement

The current study investigates ways of improving both the accuracy and interpretability of ECG classification by incorporating time-domain and frequency-domain features, including a diverse dataset, and applying multihead self-attention for improved model focus and generalization.

B. Gaps Identified

- **Limited Dataset Diversity for ECG Classification:** Many of the currently available ECG classification models, including those trained on publicly available datasets such as PTB-XL, suffer from a lack of demographic diversity; this will have implications for the generalization across different populations. Most datasets have specific patient conditions that dominate, which may not be fully representative of the myriad ECG signals one may encounter in clinical life. In this paper, the dataset used is more diverse, containing a wider spectrum of patient conditions and also more variations in ECG morphology; hence, improving the model's generalization capabilities to a wider set of real-world ECG data.

- **Lack of Integration Between Temporal and Frequency Domain Features:** Although many ECG classification methods have been focused on either time-domain or frequency-domain signals, there is limited research that explores the integration of both domains to achieve better model performance. Temporal features capture the shape and structure of the ECG waveform, while frequency-domain features, such as those extracted using FFT, provide insight into the frequency characteristics of the signal. This research bridges this gap by implementing a hybrid model that simultaneously integrates both time-domain and frequency-domain features to allow the model to better capture the complex patterns inherent in the ECG signal, which are crucial for the proper identification of conditions such as Myocardial Infarction (MI) and ST-T Changes (STTC).
- **Absence of Attention Mechanisms in ECG Classification Models:** The majority of ECG classification systems lack sophisticated components which enable models to concentrate on essential parts of ECG signals. The model's ability to focus on essential areas of interest including abnormal heartbeats and morphological changes remains restricted which affects its interpretability and accuracy. The research fills this knowledge gap through multi-head self-attention mechanisms which enable the model to concentrate on the most important sections of ECG signals. The model becomes more effective at detecting essential features which indicate particular medical conditions through this improvement which leads to better performance and clearer model prediction interpretation.

II. Literature Review

Ba Mahel et al. (2025) propose an explainable hybrid CNN-GRU model for the classification of ECG arrhythmias, extending prior work in which more traditional ML approaches (like SVM, KNN, and Random Forest) had reached only moderate accuracy due to manual feature engineering and poor scalability. Deep learning models represented by 1D-CNNs, LSTMs, BiLSTMs, and hybrid CNN-RNN architectures further improved performance through the extraction of morphological and temporal ECG patterns, often at the cost of interpretability and generalization across datasets. In order to bridge these gaps, the authors integrate an attention mechanism along with Grad-CAM into a lightweight CNN-GRU architecture to achieve improved accuracy, specificity, and interpretability on both the MIT-BIH and PTB datasets. The findings in this work are that attention-enhanced hybrid models outperform classical, deep, and hybrid baselines, thereby providing a clinically more reliable approach toward automated arrhythmia detection [3]. Safdar et al. (2023) propose a novel data-augmentation technique for ECG classification that addresses two major challenges in PTB-XL—class imbalance and lack of generalizable training data. Unlike traditional augmentation methods such as flipping, rotation, noise injection, and GAN-based synthesis, which often distort ECG morphology or generate unrealistic waveforms, the authors introduce a simple yet effective segmentation-and-rearrangement algorithm that reorders signal segments to produce new samples with low structural but high feature similarity to the originals. They evaluate the augmented data using both transfer-learning models (VGG-16, DenseNet) and a lightweight custom 4-layer CNN, showing significant improvements in accuracy—from 64% on the raw dataset to nearly 90% after augmentation—demonstrating that structural diversification without feature loss can enhance ECG classification performance, especially in imbalanced diagnostic classes [4]. Wang et al. (2024) propose an ECG classification framework that combines a lightweight 1D-CNN with multi-resolution feature extraction and an improved attention mechanism to enhance arrhythmia detection accuracy on PTB-XL and MIT-BIH datasets. Prior studies relied on either traditional ML with handcrafted features or deep CNN/LSTM models, but these approaches often suffered from high computational cost, weak generalization, and limited focus on subtle morphological variations across different ECG leads. Wang et al. address these issues by designing a multi-scale convolution block to capture fine-grained and global waveform patterns, while the attention module enhances feature weighting for clinically relevant segments. Their results outperform several CNN and RNN baselines in both accuracy and F1-score, showing that multi-resolution representation combined with attention can significantly improve ECG signal classification efficiency and robustness [5]. Nitta et al. (2024) present a lightweight deep learning framework for ECG arrhythmia classification that integrates a 1D-CNN feature extractor with a bidirectional GRU and an improved attention mechanism to enhance interpretability and feature weighting. Prior ECG research using classical ML approaches—such as SVM, KNN, and Decision Trees—relied on handcrafted morphological features and delivered inconsistent performance across datasets, while deep learning models like CNNs, LSTMs, and hybrid CNN-RNN architectures achieved higher accuracy but remained computationally heavy and lacked clear explainability. Addressing these limitations, the authors design a compact CNN-BiGRU model capable of capturing both local waveform morphology and long-range temporal dependencies, while an attention layer strengthens discriminative focus on clinically relevant ECG segments. Experiments on MIT-BIH and PTB-XL demonstrate improved accuracy and F1-scores over existing CNN and LSTM baselines, indicating that combining gated recurrent units with attention can provide efficient and interpretable arrhythmia classification suitable for real-time screening applications [6]. Acharya et al. (2022) present an automated ECG arrhythmia detection system using a deep convolutional neural network designed to learn morphological and temporal features directly from raw ECG signals without manual feature extraction. Earlier approaches—such as SVM, Random Forest, and wavelet-based feature engineering—were limited by dependence on handcrafted descriptors and inconsistent performance across different ECG datasets. Deep learning solutions like CNNs and LSTMs improved accuracy but were often computationally heavy and prone to overfitting on small, imbalanced datasets. Acharya et al. address these issues by designing an optimized CNN architecture capable of extracting hierarchical features while minimizing preprocessing requirements, demonstrating strong performance on MIT-BIH and outperforming several classical and deep-learning baselines. Their work shows that end-to-end CNN models can effectively capture discriminative waveform characteristics, offering a robust foundation for automated arrhythmia screening [7]. Śmigiel et al. propose a modular CNN framework for PTB-XL ECG classification that integrates raw 12-lead signals, extracted QRS complexes, and entropy-based descriptors. The authors introduce a multi-lead R-peak detection scheme that fuses outputs from six classical single-lead detectors and refines peak positions using k-means clustering, yielding more accurate segmentation. They compute a comprehensive set of entropy measures for both full signals and isolated QRS complexes, demonstrating that QRS-level entropy enhances discriminative power. Their architecture combines separate encoders for raw signals, QRS segments, and entropy vectors, with fused representations used for multi-class classification. Experiments on 2-, 5-, and 20-class PTB-XL tasks show that QRS-based and entropy-augmented models outperform raw-signal baselines, highlighting the utility of entropy and precise R-peak localization in improving ECG classification performance [9]. Pałczyński, Śmigiel, Ledziński, and Bujnowski review how deep learning has advanced automated ECG analysis but still faces challenges such as data scarcity, class imbalance, and the high cost of retraining large models. While prior ECG research has relied mainly on CNNs, LSTMs, and QRS-detection networks that require extensive labeled data, emerging Few-Shot Learning (FSL) methods in biomedical imaging and biosignals show promise for learning from limited samples. Studies using Siamese networks, meta-learning, and FSL-based segmentation in areas such as leukocyte classification, COVID-19 imaging, dermatology, EEG motor imagery, and arrhythmia detection demonstrate improved adaptability with minimal data. Building on this trend, the authors investigate FSL for ECG classification using PTB-XL, positioning it as a more flexible and data-efficient alternative to traditional softmax-based deep learning models [10].

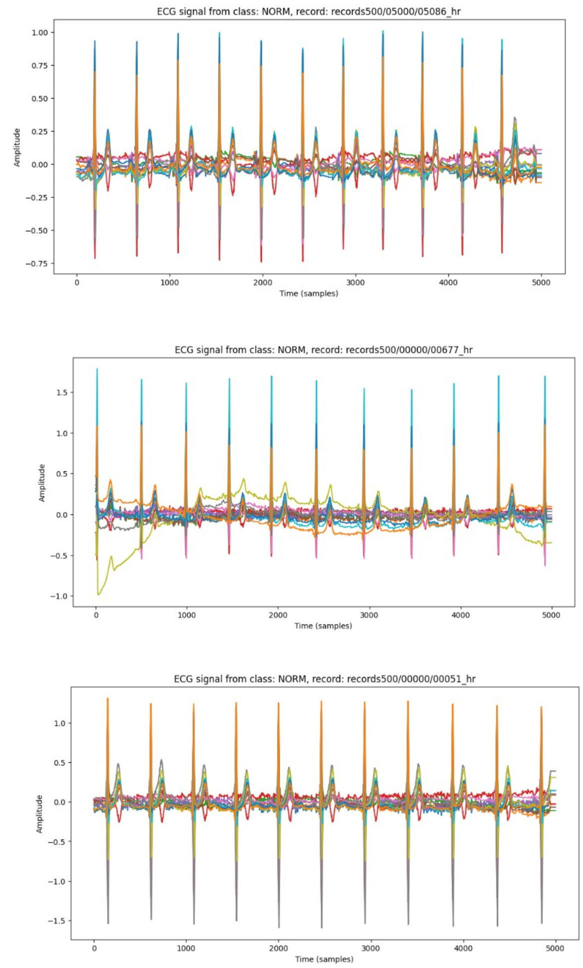
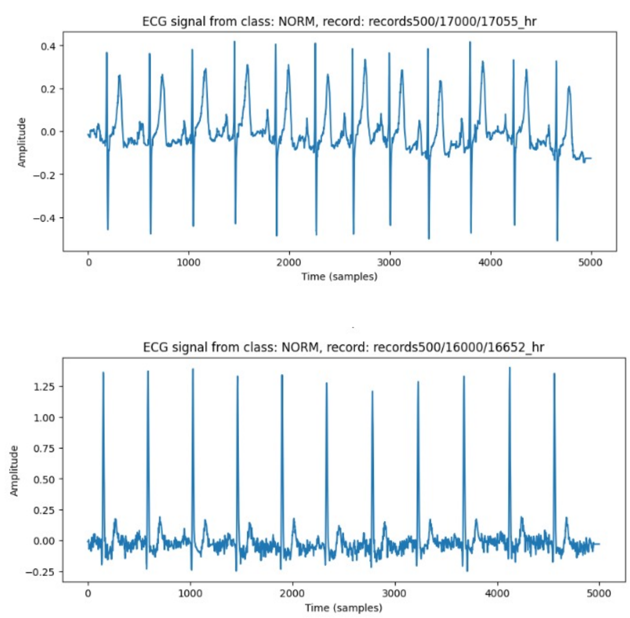


Fig. 2. Figures showing ECG Signals of 12 leads of NORM Class



III. Proposed Methodology

A. Dataset

In this study, we used a PTB-XL ECG dataset sourced from Physionet. The PTB-XL ECG dataset is a large dataset of 21799 clinical 12-lead ECGs from 18869 patients of 10 second length, where 52% are male and 48% are female with ages covering the whole range from 0 to 95 years (median 62 and interquantile range of 22) [8]. It is publicly available dataset which is annotated by expert cardiologists and is also complemented by extensive metadata.

B. Preprocessing

In this study, we applied a series of consistent preprocessing steps across all models to ensure the quality and consistency of the input data. The preprocessing pipeline for the ECG signals involved the following steps:

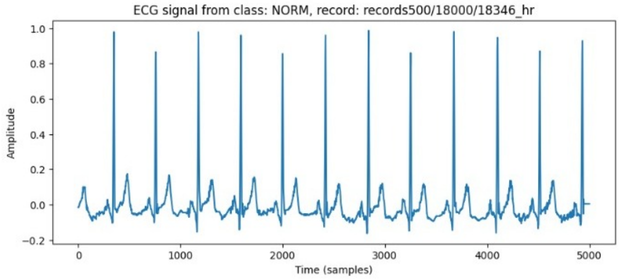


Fig. 3. Figures showing ECG Signals of only 1 lead of NORM Class

- **Resampling:** The ECG signals received uniform processing through resampling at 250 Hz because their original sampling rates differed. The preprocessing step became essential because it standardized the input data before model processing.
- **Filtering:** The ECG signals underwent multiple filtering operations to eliminate both noise and artifacts. The filtering process included three stages which operated as follows:
  - **High-pass filter:** The high-pass filter operated at 0.5 Hz to eliminate low-frequency noise that produced baseline wander in ECG signals.
  - **Bandpass filter:** The bandpass filter operated between 0.5 Hz and 40 Hz to maintain the essential frequency range for ECG signal evaluation.
  - **Notch filter:** The notch filter operated at 50 Hz to remove powerline interference which represents a typical ECG recording noise source.
- **Normalization:** The ECG signals underwent normalization through a process that involved subtracting the mean value and then dividing by the standard deviation of each signal. The normalization process brought all signals to a uniform state with zero mean and unit variance which supported stable model training.
- **Windowing:** The ECG signals received windowing treatment through 5-second segments (1250 samples per segment) with 50% overlapping sections between adjacent segments. The windowing process created workable signal lengths which helped researchers detect patterns that emerged from ECG data across time.
- **Stratified Train-Validation Split:** The research team divided the data into training and validation sets through stratified sampling to maintain equal class representation between the two sets. The method protected the dataset from class distribution biases which could affect the results.
- **Class Balancing:** The different models used two methods to handle class imbalances during training: Model-1 used oversampling while Models 2, 3 and 4 applied class weight adjustments to prevent minority class underrepresentation.

C. Models

To systematically evaluate how architectural components influence multi-class ECG classification, we developed four progressively more advanced models ranging from a simple spectral CNN baseline to a state-of-the-art hybrid InceptionTime–Attention architecture.

Model 1 is a lightweight FFT-based CNN that operates solely in the frequency domain using short 5-second windows. Its shallow structure and lack of temporal modeling limit its ability to capture important ECG morphology such as QRS shape, ST-segment trends, or T-wave abnormalities. The preprocessing sequence for model 1 consisted of standard resampling followed by filtering and windowing according to the previous steps. The ECH signals entered the time domain analysis without undergoing any further feature extraction process. This model did not incorporate frequency-domain features, and no additional preprocessing steps (such as FFT) were applied. The focus was purely on time-domain analysis using 12-lead ECG windows.

Model 2 extends this baseline to a dual-branch hybrid framework by combining a time-domain CNN with an FFT domain CNN. This early fusion of temporal and spectral information leads to moderate performance gains, confirm ing that frequency-domain cues complement raw waveform morphology, although the shallow design still restricts its capacity to learn long-range dependencies. Model 2 followed the same preprocessing methods as model 1 but included an additional step, which applied FFT (Fast Fourier Transform) to ECG windows. The ECG windows underwent two processing steps, which included time-domain analysis and frequency-domain analysis to extract vital cardiovascular features. The model combined time-domain features with frequency-domain features to create a hybrid system that processed both temporal and spectral data.

Model 3 implements advanced representation learning techniques together with explicit time-based analysis through residual CNN blocks and BiLSTM layers, and Multi-Head attention. The new architecture design enables better detection of both short-term and extended patterns in ECG signals, which leads to enhanced performance in both accuracy and macro-F1 metrics. Model 3 extended the model 2 preprocessing system through residual CNN layers for enhancing feature extraction and BiLSTM layers for detecting extended temporal patterns. The preprocessing operations from Model-2 carried over to Model-3, but residual connections improved training gradient propagation for deep neural network architectures. The model architecture received an attention mechanism, but the preprocessing operations maintained their original configuration.

Finally, **Model 4** achieved the highest performance through its combination of multi-scale Inception ime blocks with residual connection and FFT branch and multi-head attention with in a single fusion framework. the model extrects information from different time scales while using spectral data and attention mechanisms to achieve excellent performance on all five diagnostic tasks. The sequence of models shows that ECG classification requires the combination of temporal and spectral and multi-scale and attention mechanisms to reach high performance levels. The preprocessing system of Model-4 used advanced techniques which matched those of Model-3. The main distinction between Model-4 and previous models stemmed from its implementation of InceptionTime blocks which performed multi-scale feature extraction. The model extracted ECG patterns at different scales through its 3, 9, 19 and 39 scale extraction capability which helps detect various ECG signal patterns. The preprocessing sequence included resampling and filtering and FFT and windowing but Model-4 introduced innovative time-frequency feature combination through its dual-branch architecture and InceptionTime feature extraction methods.

D. Evaluation Metric

Models were evaluated using the metrics described below:

- **Accuracy:** Accuracy is widely known performance metric for machine learning algorithms. It measures how often a machine learning model predicts the correct output.
- **Precision:** Precision is a metric that measures how frequently a machine learning model predicts the positive class.
- **Recall:** Recall is a performance metric that how often a machine learning model correctly identifies positive instances (true positives) from all the actual positive samples in the dataset.
- **F1 Score:** The F1 score ranges from 0 to 1. It is known as balancing of precision and recall.

E. Implementation Environment

The research implementation took place in a Python-based system which utilized multiple strong machine learning frameworks and libraries. The main library for model development and training used TensorFlow version 2.x while Keras functioned as the high-level API for building and training deep learning models. The signal processing operations for preprocessing involved SciPy and NumPy to execute numerical computations and signal processing tasks. The dataset management and preprocessing tasks used Pandas while Matplotlib and Seaborn handled data visualization for model performance assessment and training curve analysis and confusion matrix generation. The deep learning model received speed boosts from NVIDIA Tesla T4 GPUs which operated in a GPU-enabled training environment. The PTB XL ECG dataset and other datasets were stored in local Kaggle notebook storage for easy cloud-based experiment execution. The scikit-learn library performed three essential functions which included class balancing and model evaluation and performance metric computation. The research environment supported quick model development and training operations which produced reliable results for ECG classification applications.

IV. Experimental Results

Model 4 stands as the top-performing model in our research because it combines InceptionTime with Attention architecture to process data at multiple time scales while using residual learning and FFT spectral analysis and multi-head self-attention. The model reaches 92% accuracy while achieving a macro-F1 score of 0.92 which surpasses all previous models. The model detects both small-scale morphological details and large-scale rhythm patterns through its multiple-scale convolutional filters and maintains stable deep learning through residual connections. The model gains improved discriminatory power for classes with specific spectral characteristics through the addition of frequency-domain information. The attention mechanism in the model identifies crucial waveform areas which include ST-segment changes and T-wave irregularities to increase the model's focus on essential clinical features. The model achieves state-of-the-art PTB-XL benchmarks through its combined features which produce excellent results for all five diagnostic categories.

The better performance of this model proves that deep learning systems for ECG interpretation need to combine different feature types during their design process. The combination of temporal and spectral models in Model 4 produces better results than models that focus on either time or frequency data because it uses attention-based weighting to combine temporal morphology with frequency structure. The research shows that ECG classification requires deep learning models to use multiple scales and multiple domains for achieving reliable clinical-grade results.

A. Ablation Study

To validate the contribution of each architectural component in our ECG classification framework, we conducted a comprehensive ablation study across four progressively enhanced models, with results summarized in Table I. The evaluation process demonstrates how each model improvement leads to better performance starting from the basic spectral CNN up to the final multi-scale InceptionTime–Attention hybrid model. The evaluation shows that time-domain processing addition followed by deeper residual and temporal modeling and multi-head attention and spectral–temporal fusion with multi-scale convolutions leads to performance improvements at each stage. The experiments show that the best model requires all three components which include multi-scale temporal features and frequency domain context and attention-based feature weighting to achieve its high diagnostic accuracy.

V. Conclusion

This work has also determined that state-of-the-art results in multi-class ECG classification definitely require integration of diverse signal representations and advanced deep-learning mechanisms. From a basic FFT-based CNN, each model progressively brought an important architectural improvement—hybrid time–frequency fusion, residual learning, BiLSTM-based temporal modeling, and multi-head self-attention—that collectively contributed significantly to better performance. The last InceptionTime–Attention hybrid model yielded the best performances, 92% accuracy and F1 score

TABLE I
ABLATION STUDY

Model ID	Model Description	Accuracy	Precision	Recall	F1-Score
Model-1	Baseline FFT-CNN using 12-lead FFT windows (simple CNN, no residuals, no attention, no temporal modeling)	0.73	0.72	0.73	0.71
Model-2	Hybrid: Time + FFT Branch (Simple CNN fusion)	0.78	0.71	0.66	0.70
Model-3	Improved Hybrid: Residual CNN + FFT + BiLSTM + Multi-Head Self Attention + Class Weights	0.85	0.80	0.84	0.81
Model-4	InceptionTime (multi-scale) + Residuals + Multi-Head Attention + Dual-branch Fusion (Time + FFT) + Class Weights + Early Stopping	0.92	0.90	0.90	0.92

of 0.92, hence underlining the importance of capturing multi-scale temporal patterns along with frequency-domain features. The attention mechanism further enhanced interpretability by enabling the model to focus on clinically relevant waveform regions such as ST-segment deviations and T-wave abnormalities.

These findings confirm that ECG classification systems have indeed benefited immensely from combining multi-domain and multi-scale features, provided that this process is supported by robust preprocessing and class-balancing strategies. The proposed framework puts forward a strong basis for the development of clinically reliable diagnostic tools and points out several directions of future work: real-time deployment, cross-dataset generalization, and integration with explainability frameworks in view of further support for clinical decision-making.

REFERENCES

[1] Pronab Ghosh, Sami Azam, Mirjam Jonkman, Asif Karim, FM Javed Mehedi Shamrat, Eva Ignatious, Shahana Shultana, Abhijith Reddy Beeravolu, and Friso De Boer. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques. IEEE Access, 9:19304–19326, 2021.

[2] Rashkovska, A.; Depolli, M.; Tomaši´c, I.; Avbelj, V.; Trobec, R. Medical-grade ECG sensor for long-term monitoring. Sensors 2020, 20, 1695.

[3] Mahel, A. S. B., Alotaibi, F. M. G., Lonseko, Z. M., Rao, N. N. (2025). XABH-CNN-GRU: Explainable attention-based hybrid CNN-GRU model for accurate identification of common arrhythmias. Journal of Electronic Science and Technology, 23(3), 100322.

[4] Safdar, M. F., Pałka, P., Nowak, R. M., Al Faresi, A. (2023). A novel data augmentation approach for enhancement of ECG signal classification. Biomedical Signal Processing and Control, 86, 105114.

[5] Eleyan, A., Bayram, F., Eleyan, G. (2024). Spectrogram-based arrhythmia classification using three-channel deep learning model with feature fusion. Applied Sciences, 14(21), 9936.

[6] Eleyan, A., AlBoghbaish, E., AlShatti, A., AlSultan, A., AlDarbi, D. (2024). RHYTHMI: A deep learning-based mobile ECG device for heart disease prediction. Applied System Innovation, 7(5), 77.

[7] Madan, P., Singh, V., Singh, D. P., Diwakar, M., Pant, B., Kishor, A. (2022). A hybrid deep learning approach for ECG-based arrhythmia classification. Bioengineering, 9(4), 152.

[8] Bousseljot, R., Kreiseler, D., SO Central Secretary. (2000). Waveform recognition with 10,000 ECGs. Computers in Cardiology, 27, 331-334. PhysioNet. https://physionet.org/content/ptb-xl/1.0.3/files-panel

[9] Śmigiel, S., Pałczyński, K., Ledziński, D. (2021). Deep learning techniques in the classification of ECG signals using R-peak detection based on the PTB-XL dataset. Sensors, 21(24), 8174.

[10] Pałczyński, K., Śmigiel, S., Ledziński, D., Bujnowski, S. (2022). Study of the few-shot learning for ECG classification based on the PTB-XL dataset. Sensors, 22(3), 904.