Journal of Al-Qadisiyah for Computer Science and Mathematics VOL.17.(3) 2025,pp.Comp 292–308

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

A Hybrid Deep Learning Model Framework for Multi-Label ECG Classification

Mustafa Mortada Khalaf Al-Shaheen*

Department of Artificial Intelligence and Robotics College of Mechanics, Electricity, and Computers, Islamic Azad University ,Teharan , Iran.

Email: mustafamortada81@gmail.com

ARTICLE INFO

Article history:
Received: 17/05/2025
Rrevised form: 22/06/2025
Accepted : 25/06/2025
Available online: 30/09/2025

Keywords:

Heart disease, ECG, Deep learning, CNN, Bi- LSTM, Attention, Multi-label classification.

ABSTRACT

Automated interpretation of multi-label electrocardiograms (ECGs) is an essential tool for diagnosing cardiovascular diseases, as multiple concurrent cardiac events can occur in a single record. Modern deep learning techniques enable rapid and accurate automated analysis of these multi-label records. In this paper, the current study proposes a hybrid model that combines a deep convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) layers with an attention mechanism, with demographic features (age, sex) added. This study applied advanced data augmentation techniques (time warping, gaussian noise, and stochastic perturbations), during training and conducted evaluation on the PTB-XL dataset using five-fold patient-level cross-validation (5-fold CV) to ensure that patient samples do not leak between groups. The proposed model demonstrated superior performance across key evaluation metrics achieving an accuracy of 98.8%, precision of 89.4%, recall of 86.63%, F1-score of 88% and an AUC-ROC (Area Under the Receiver Operating Characteristic Curve) of 94.4%. These results confirm the effectiveness of the system in rapidly detecting cardiac abnormalities during clinical ECG screening.

MSC.

https://doi.org/10.29304/jqcsm.2025.17.32506

1. Introduction:

Cardiovascular disease (CVD) is the leading cause of mortality worldwide. Many studies indicate that a large percentage of the world's population suffers from atherosclerosis, heart disease, and other disorders. The rapid diagnosis is essential for efficient treatment and prevention of cardiovascular diseases and problems. Traditional approaches to diagnosing heart disease, clinical evaluations, lab tests, and imaging techniques can be costly and labor-intensive, therefore causing mistakes or delays in diagnosis decisions and, finally, patient loss. Deep learning (DL) technologies have revolutionized the healthcare industry by providing faster and more accurate methods for identifying heart disease, both machine learning, and deep learning approaches have shown promising results [1, 2].

Although advanced clinical imaging techniques such as magnetic resonance imaging (MRI), coronary computed tomography (CT) scans, blood tests, and ultrasound are widely available, the ECG remains a straightforward and dependable way to diagnose cardiac issues by monitoring of the electrical signal of the heart. Easy to use and non-invasive, the ECG is a representation of the heart's electrical system. Doctors and cardiologists have relied on the ECG to diagnose and predict cardiovascular outcomes for over a century [3].

*Corresponding author: *Mustafa Mortada Khalaf Al-Shaheen*

Email addresses:

Communicated by ‘sub etitor’

A significant issue mostly for middle-aged and older people is heart disease. These diseases harm the heart, arteries, and veins, therefore causing a notable global death count [4]. The world health organization (WHO) estimates that about 17.9 million people die from CVDs each year, or nearly 32% of all deaths worldwide. Low- and middle-income countries suffer the most from the disease, which comprises several types, including coronary artery disease (which by itself caused about 6.7 million of the 17.7 million deaths), ventricular arrhythmias, and congestive heart failure [5]. Often, diagnosing these diseases calls for protracted manual inspections, so fast and precise diagnostic tools are absolutely vital [6]. Deep neural networks, especially (CNNs), have shown encouraging results in ECG interpretation over the last ten years; prior research has shown CNNs can identify several diseases from ECGs with performance similar to that of cardiologists, in academia and business, machine learning (ML) especially deep Learning (DL) techniques has attracted much interest, which has resulted in dramatic changes in classification and automated detection methods [7]. Recently, in several engineering disciplines, including electrical, civil, and petroleum engineering, the efficacy of artificial intelligence and ML has been demonstrated. In medicine, DL diagnoses diseases depending on physiological criteria, classifies cardiac arrhythmias using ECGs, and identifies human activity. Several studies have used models like long short-term memory (LSTM) networks and one-dimensional convolutional networks (1D-CNNs) to sort ECGs and detect heart problems [8]. This paper presents a new model for classifying heart diseases; it first uses a 1-dimensional CNN (with Residual-style blocks) to pull out important details from raw ECG signals, then employs a bidirectional LSTM to understand the timing of the signals and a transformer-based attention module (with dropout) to emphasize the most important heartbeats, while a separate branch handles clinical information like age and sex. (MixUp/CutMix) augmentation, focus loss, K-fold cross-validation are used to fully train the model end-to-end, therefore producing total assessment metrics and visualizations. The main contributions of this study:

1-This study proposes a novel hybrid deep learning framework that integrates multi-branch residual CNNs, transformer-based attention, and BiLSTM with skip connections for robust multi-label ECG classification.

2-The model uses specific embedding layers to add demographic information (age and sex), improving prediction customization and diagnosis accuracy. The current study uses sophisticated data augmentation methods (MixUp, CutMix, gaussian noise, and random scaling), as well as adaptive focus loss and label smoothing, to increase generalization and successfully address class imbalance.

3- A comprehensive end-to-end pipeline is created and assessed on the PTB-XL dataset using 5-fold patient-level cross-validation, resulting in state-of-the-art performance with an AUC-ROC of 0.94 and an accuracy of 0.98. The suggested model beats current models in terms of accuracy, precision, recall, and F1-score, while retaining low variance across validation folds, showing good generalizability.

2. Related Works:

This study by Gour A *et al.* (2024) provides a comprehensive review of modern techniques for heart disease classification using ECG (CNN, XGBoost, LSTM) and reports that CNN achieved an average accuracy of 98.83% on MIT-BIH and PTB-XL [9]. This research conducted by E. Prabhakararao and S. Dandapat (2023) presents a set of attention-based temporal convolutional neural network (ATCNN) binary classifiers, one for each cardiac disorder designed for multi-label classification of 12-lead ECG records from the PTB-XL 2020 dataset, attaining an average F1-score of 76.51% [10]. Wei Huang *et al.* (2024) introduces the Multi-Resolution Mutual Learning Network (MRM-Net), which utilizes parallel high- and low-resolution attention branches for mutual feature learning to integrate multi-scale ECG representations. When assessed on PTB-XL and CPSC2018, MRM-Net attains an AUC-ROC of 93.59%, an accuracy of 97.96%, and an F1-score of 73.74% on the PTB-XL "all" task [11]. This study conducted by Sahil Sethi *et al.* (2025) presents ProtoECGNet, a multi-branch prototype-based DL model that emulates clinical interpretation by integrating 1D rhythm, 2D morphology, and 2D global prototype branches; it attains a macro-averaged AUC-ROC of 0.9248 on the PTB-XL test set [12]. This research was undertaken by Lei Kang *et al.* (2025) process xLSTM-ECG, which transforms 12-lead ECG signals via Short-Time Fourier Transform and processes them through extended LSTM modules capturing both temporal and inter-lead dynamics; on the PTB-XL dataset xLSTM-ECG achieves an AUC-ROC of 91.33 % and accuracy of 87.59 % [13]. This study was conducted by Qiao Xiao *et al.* (2025) incorporates a lead-wise prior knowledge framework (LPKF) into an Inception-based network to embed expert-defined correlations between ECG leads and abnormalities, facilitating feature separation for multi-label classification; on the PTB-XL superclasses, the LPKF-Inception model achieves a precision of 0.754, a recall of 0.778, and an F1-score of 0.764 [14]. This article was performed by M. Gitau *et al.* (2023), they replicated six benchmark tasks on PTB-XL using seven state-of-the-art architectures, confirming full reproducibility of the original code and results [15]. The study by Oueldji *et al.* (2024) conducted a comprehensive in comparative study of CNN, DNN, LSTM, and U-Net algorithms on the PTB-XL set, stressing the need of unified databases in facilitating fair and comparative assessment of models [16]. This article was conducted by Razin *et al.* (2023), they focused on the use of DL as a whole and tried out several deep model ensembles to see which one worked best for predicting the occurrence of various cardiac problems (such as myocardial infarction, hypertrophy, conduction abnormalities, and ST segment alterations) in ECG data [17]. This study by Śmigiel *et al.* (2021) suggests three main types of systems for processing ECG signals to identify features and classify them: a traditional convolutional network (CNN), SincNet, and a convolutional network that uses entropy-based features [18]. In this research by Pałczyński, *et al.* (2022), they investigated the applicability of few-shot learning to ECG classification from PTB-XL and compared it to traditional Softmax classification. The few-shot-capable network achieved better accuracy on 2-class and 5-class tasks [19]. This study by Strodthoff *et al.* (2021) demonstrated that ResNet- and Inception-based models achieve the best performance across various tasks on PTB-XL, encouraging the use of transfer learning as an "ImageNet" for the ECG domain [20].

The hybrid architectures combining the spatial feature extraction capabilities of CNNs with the temporal pattern capture capabilities of LSTMs are currently the best option to increase the accuracy and stability of early detection systems for heart disease. These designs should be supported by augmentation, ensemble methods, attention, or XAI (Explainable AI) mechanisms. The following Table 1 shows the summary of related work.

Table 1: Comparison between related work in terms of method, datasets, and result:

Rf.	Method	Dataset	Result	Gaps / Challenges	Advantages / Contributions
Gour A *et al.* (2024) [9] {single-label classification}	SVM, k-NN, CNN, LSTM, XGBoost, GMM, K-Means	PTB-XL & MIT-BIH	LSTM Sensitivity = 88.01%, SVM, CNN Accuracy = 98.83%, k-NN Accuracy = 97.50%, Accuracy = 96.20%.	Single-label focus, no attention, no augmentation	Compared multiple ML/DL classifiers on PTB-XL
Prabhakararao *et al.* (2023) [10]	ATCNN	PTB-XL	F1-score = 76.51%	No multi-branch fusion, limited interpretability	Explored standardization in model comparison
Wei Huang *et al.* (2024) [11]	MRM-Ne	PTB-XL	AUC-ROC = 93.59 %, Accuracy = 97.96%, F1-score = 73.74%	No detailed metrics, weak documentation	Explored ensemble deep learning models
Sahil Sethi *et al.* (2025) [12]	ProtoECGNet	PTB-XL	AUC-ROC = 0.9248	Early design, also no attention or transformer	Used entropy-based and SincNet features
Lei Kang *et al.* (2025) [13]	xLSTM-ECG	PTB-XL	AUC-ROC = 91.33 %	Limited scalability to multi-label cases	Adapted few-shot learning for ECG
Qiao Xiao *et al.* (2025) [14]	LPKF-Inception (Lead-wise Prior Knowledge Framework)	PTB-XL	Precision = 0.754 Recall = 0.778 F1-Score = 0.764	No attention, also no clinical feature use	Suggested transfer learning in ECG domain
Gitau *et al.* (2023) [15]	CNN, LSTM, FCN, Wavelet+NN, Ensemble	PTB-XL	xresnet1d101= 0.925, Inception1d= 0.926, AUROC, Accuracy, Precision= 0.906–0.937	No demographics, also no ensemble, limited attention scope	Used per-class binary ATCNNs
Oueldji *et al.* (2024) [16]	CNN, DNN, LSTM, U-Net, (RL-LSTM)	PTB-XL	Recall = 0.75, Precision = 0.67, RL-LSTM: Accuracy = 0.98, F1-Score = 0.65.	No demographics, also don't skip-connections	Mutual attention fusion at multiple resolutions
Razin *et al.* (2023) [17]	CNNs, RNNs, Ensemble	PTB-XL	No explicit numerical results were reported.	Don't learned attention weights or demographic fusion	Prototype-based interpretability
Śmigiel *et al.* (2021) [18]	CNN, SincNet, CNN+Entropy feature extraction	PTB-XL	No explicit numerical results were reported.	No spatial CNN, also no residual connections	Combined STFT with LSTM extensions
Pałczyński, *et al.* (2022) [19]	Few-Shot Learning	PTB-XL	Accuracy = 90.4%±0.5%, Precision/F1 =	No attention, also no feature fusion	Focused on reproducibility

	(XGBoost, RF, DT, KNN, SVM)		90.6%±0.6%, Sensitivity/AUC = 93.7%±0.8%,		benchmarking
Strodthoff *et al.* (2021) [20]	CNN (ResNet- & Inception-based), LSTM, FCN, Wavelet+NN, Ensemble	PTB-XL	AUROC= (All: 0.927, Diagnostic: 0.937)	Rule-based structure, less generalizable	Used prior-knowledge in Inception model

3. Detection of Heart Diseases in the Role of Artificial Intelligence:

Artificial intelligence (AI) is the simulation of human cognitive processes via technical methods, particularly computer systems. The applications of AI include natural language processing, expert systems, speech recognition, pattern identification, and machine vision. AI systems work by absorbing massive volumes of labeled training data, analyzing it for correlations and patterns, and then using these patterns to predict future events. Radiology, electrocardiography, medical image classification and diagnosis have greatly improved as a result of DL technology, which uses electrocardiography, X-rays, computed tomography (CT), magnetic resonance imaging (MRI), and other types of imaging. [2, 8].

4. Learning Techniques:

Artificial intelligence (AI), including learning methodologies, has transformed computer technology, it mostly involves machine learning (ML) and deep learning (DL), so the AI often uses ML. Machine learning (ML) is an algorithmic intelligent artificial system that has self-learning capabilities, (ML) denotes systems that exhibit ever higher degrees of creativity and intelligence without direct human involvement, the second advancement pertains to the field of (DL). (ML) methodologies are used for the analysis of extensive data sets; the progression of ML is driven by the ideals of precision and speed. (DL) distinguishes itself from traditional (ML) methods by using numerous layers that improve abstraction and strengthen generalization, employing neural networks necessitates the extraction of features to mitigate the computational load of the network, for neural networks to assimilate features, essential to develop a comprehensive network capable of conveying exact commands, and Consequently, the probability of successful training was negligible or absent, owing to improvements in computational power and memory capacity, training extensive networks successfully is now achievable, within the same DL network, feature extraction, classification, and generalization may be performed, resulting in substantial benefits. CNN layers extract local spatial features from the heart signal, then BiLSTM captures temporal relationships in both forward and backward directions to understand the entire sequence. Attention assigns higher weights to the most relevant time frames, and the weighted representations are then combined and passed to the final layer for accurate classification [2,8].

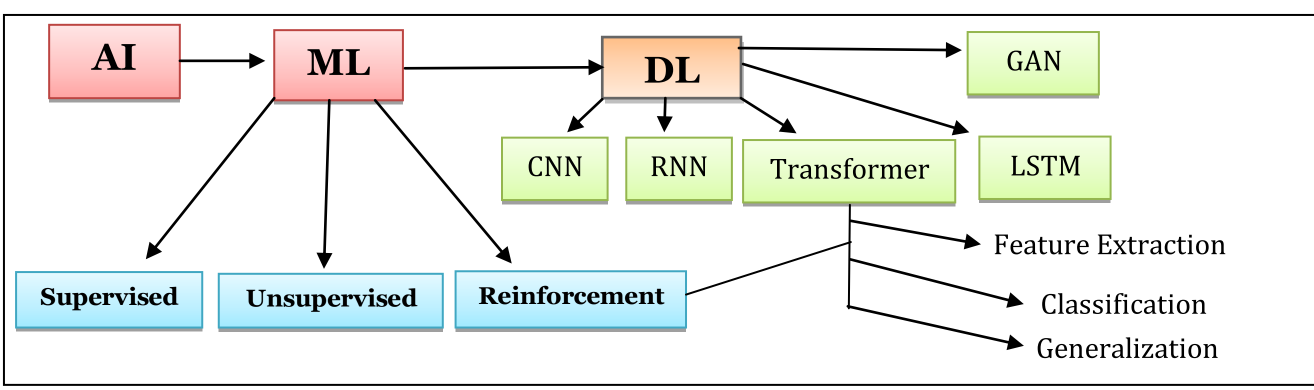

Fig 1: Learning Techniques

The figure (1) shows the architecture and components for the learning techniques.

5. Methodology and Hybrid Model Design

The suggested method is an integrated hybrid model for cardiac diagnostics, using ECG signals with clinical data. The model begins with a preprocessing step that includes (signal refinement, normalization, filtering, Z-score, converting the signal to clustered matrices), when the signal is refined and transformed into clustered matrices. The create_advanced_encoder function builds Conv1D layers with residual connections (similar to ResNet) starting from random weights and is trained directly supervised; a transformer block and then a residual BiLSTM are then integrated without any loading of pre-trained weights. The subsequent step involves the fine-tuning of a hybrid model that integrates CNN-BiLSTM architecture with attention mechanisms, in conjunction with data transformation techniques including (MixUp, CutMix), and advanced augmentation methods. A collection of models is subsequently generated utilizing a stratified (K-fold) dataset, to enhance accuracy and stability. A thorough assessment is conducted utilizing a confusion matrix, ROC curves, precision, recall metrics, and an interpretability framework that incorporates XAI tools (Explainable Artificial Intelligence) like SHAP (SHapley Additive exPlanations) or Grad-CAM (Gradient-weighted Class Activation Mapping). As show in the figure blew.

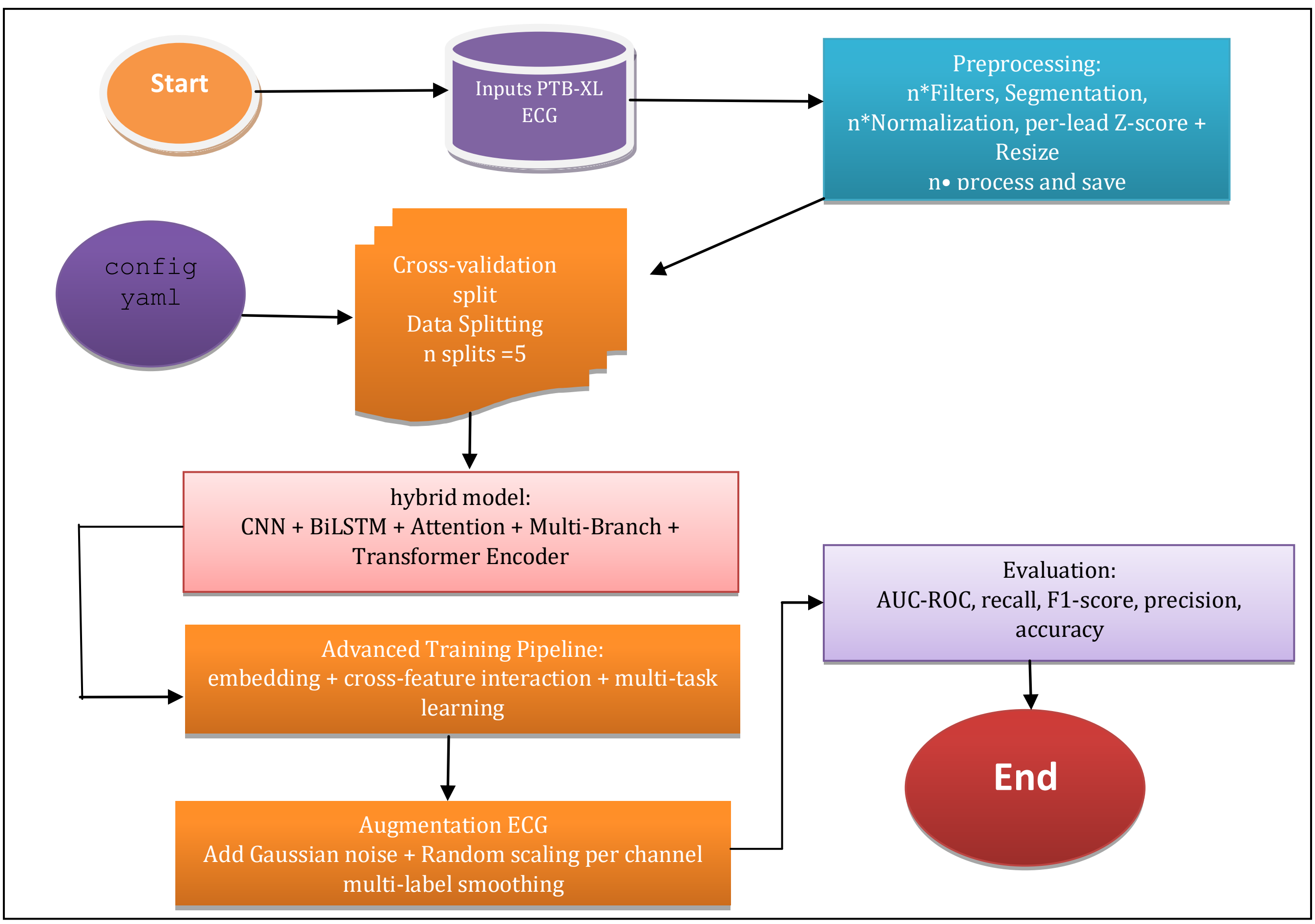

Fig2: The Proposed System Flowchart.

The figure (2) shows the flowchart of a hybrid model methodology for processing ECG signals in a way that allows for dual or multitasking may be seen in the flow diagram. An electrocardiogram (ECG) signal (5000×12) and demographic data (sex and age) are inputted in parallel via light embedding layers at the beginning of the flow. Afterwards, the ECG signal are processed through a "forward encoder" that incorporates residual connections, context-guiding automated attention layers, and multi-scale convolutional branches (Conv1D with varying kernel sizes). A BiLSTM is then used to record long-term relationships, and dilated convolutions are employed to expand the receptive space. After that, the demographic outputs and all of the characteristics that were created are merged, and then the combined vector is sent to various classification heads after passing through a dense concatenation layer (Dense 512 units with swish activation). "Loss1" and "Loss2" at the end of the figure reflect the independently estimated probabilities for each class, which are produced by each vertex's sequential Dense layers (e.g., 256 units with swish, then dropout, then 128 units with sigmoid).

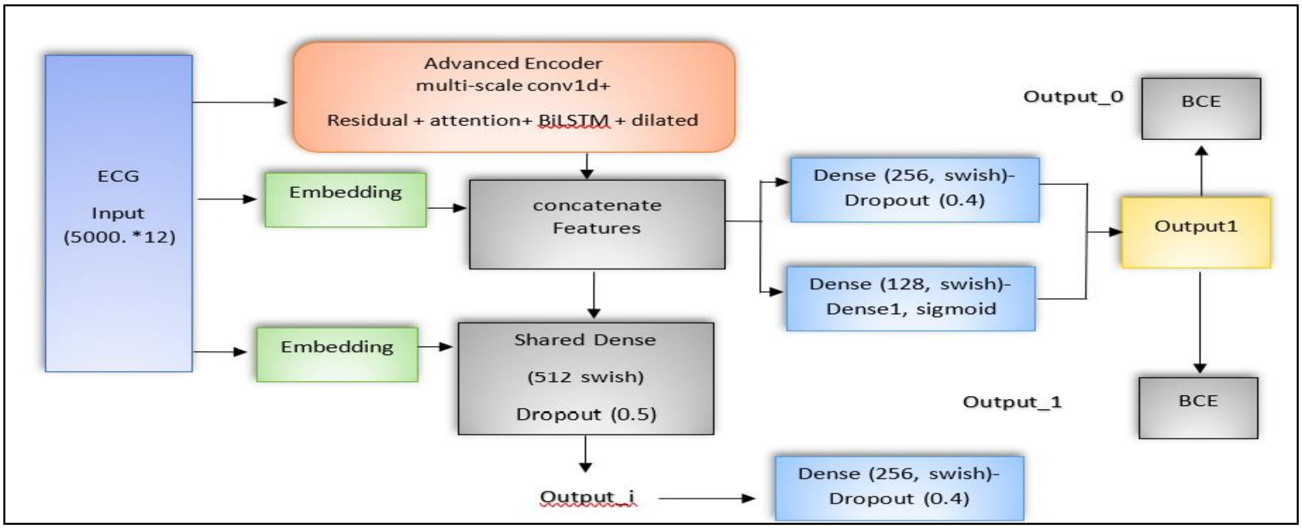

Fig. 3: A hybrid methodology.

The figure (3) shows the flowchart a hybrid methodology for processing ECG signals proposed in our model.

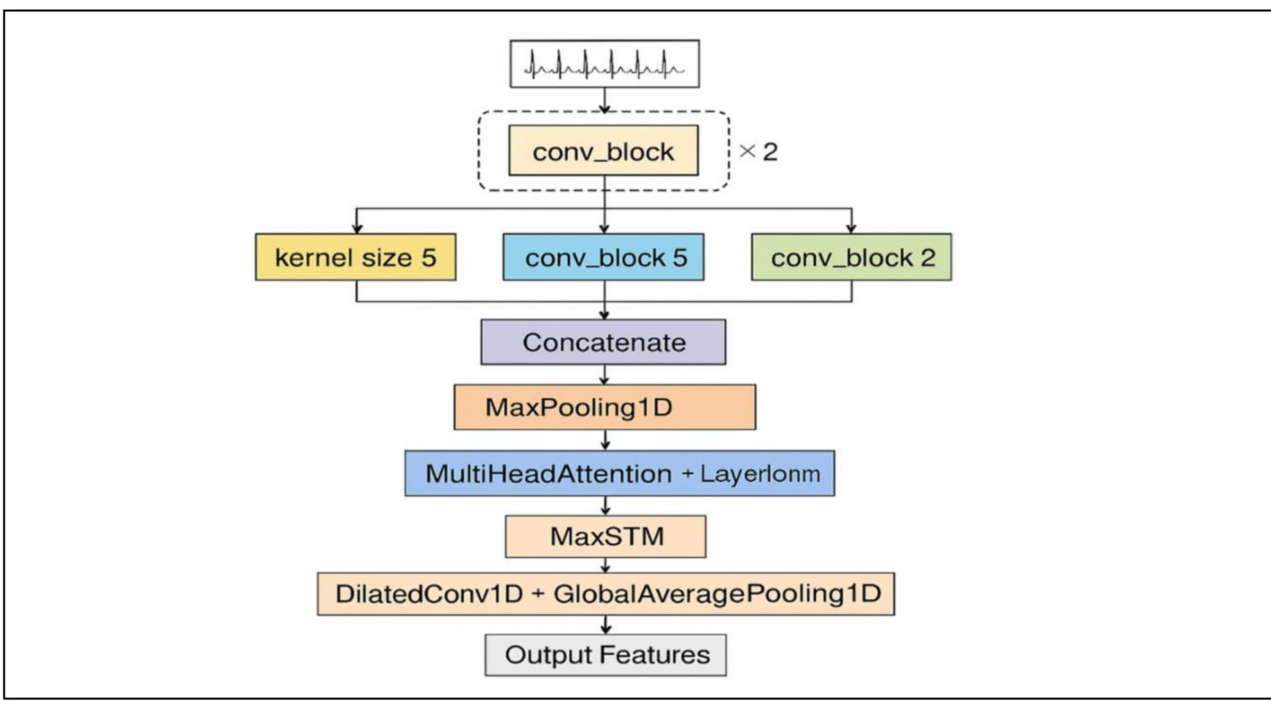

Fig 4: Visual diagram of the advanced encoder structure.

The figure (4) shows the framework architecture starts with a multi-channel ECG signal input of dimensions 5000×12, and this input is processed through two sequential convolutional blocks, which consist of Conv1D, batch normalization, swish activation, and dropout, incorporating residual connections, after that the signal subsequently diverges into three distinct paths, each employing varying kernel sizes of (5, 11, and 21). Within each path, a conv_block is applied, followed by temporal pooling using (MaxPooling1D), and concluding with an additional conv_block, after that the characteristics of the three paths are integrated using a concatenate layer, followed by (MaxPooling1D) to decrease the temporal dimension, the layer is subsequently input into a (MultiHeadAttention) mechanism that incorporates (LayerNormalization), which is then followed by a (BiLSTM) layer featuring residual connections. A (DilatedConv1D) layer is implemented, succeeded by global pooling through (GlobalAveragePooling1D), which generates the final feature vector for each ECG sample.

❖ 5-1 Study Design:

The study aims to develop and evaluate a hybrid DL model that combines Conv1D residual networks, transformer units, and BiLSTM layers with an attention mechanism to diagnose cardiac disease using PTB-XL ECG data.

❖ 5-2 Materials & Data Sources:

The current study uses the PTB-XL dataset Electrocardiogram (ECG) Database, Version 1.0.3. We download the dataset from the official website of medical data, (https://physionet.org/content/ptb-xl/1.0.3/).

The PTB-XL ECG dataset comprises (21,799) 12-lead ECG records, each with a duration of 10 seconds. The data was gathered from (18,869) patients aged (0 to 95) years, with a mean age of 62 years and a fairly balanced sex distribution (52% male and 48% female). This dataset exhibits significant diagnostic diversity. It encompasses a broad spectrum of co-morbid conditions and a substantial number of healthy samples, rendering it an invaluable resource for the advancement and assessment of automated ECG interpretation algorithms. Each record was annotated by two seasoned cardiologists in accordance with the SCP-ECG standard, and each record may possess multiple labels from a set of 71 diagnostic statements encompassing (1) cardiac diagnoses, (2) waveforms, and (3) rhythm disturbances. This dataset is enhanced by extensive metadata, encompassing demographic details, infarct attributes, diagnostic probabilities, and signal features, thereby offering a thorough resource for research in ECG analysis. The variation in age, sex, and diagnostic conditions within this cohort illustrates the complexities of reality, thereby augmenting the generalizability of the models developed from it. The PTB-XL ECG dataset have many attributes, the 28 features in the PTB-XL include: Patient information (age, sex), signal characteristics (quality, noise), medical interpretations (SCP-ECG codes), technical data (e.g., infarction phase) [23, 24, 25].

Table 2: Distribute and partition the original database:

Dataset	ECG record	No. of Patients	No. of patients/have disease	No. of patients/don't have disease	No of diagnostic phrases	No. of Attributes	Classes
PTB=XL	21,799 ECG recordings	18,869	18,459	410	diagnostic statements=71	28	Super classes = 5 subclasses = 23

❖ 5-3 Procedures & Experimental Protocol:

These waveforms were gathered using Schiller AG devices over the course of almost seven years, from October 1989 to June 1996. They make up the PTB-XL ECG collection. After the PTB bought the original information from Schiller AG, they were given full permission to use it. As part of a long-term project at the Physikalisch-Technische Bundesanstalt (PTB), the records were organized and turned into a standardized database. When the current database was made public in 2019, it was simplified with the ML community's ease of use and accessibility in mind. Waveforms and information were changed into open data types that common tools can easily work with [23, 24, 25].

1/ Waveform Formatting: All signals were recorded in a proprietary compressed format and then converted to WFDB, with a standard set of 12 leads (I, II, III, aVL, aVR, aVF, V1–V6) with a duration of 10 seconds per recording [23, 24, 25].

2/ Demographic metadata: The age and sex data for each patient, as well as height in 31.98% of samples and weight in 43.18%, were stored in CSV tables attached to the primary data [23, 24, 25].

3/ Annotation produces an automated preliminary interpretation for each ECG recording. Each interpretation undergoes review by two independent cardiologists and is subsequently converted into final SCP-ECG codes.

Additionally, heart_axis and infarction_stadium1 & infarction_stadium2 values are extracted when available [23, 24, 25].

4/ Technical verification of signal properties Important signal properties (such as baseline drift and noise levels) are manually tagged as fields in the technical metadata, to ensure the quality of the recordings before in-depth analysis [23, 24, 25].

❖ 5-4 Data Preprocessing:

Multiple preprocessing techniques are utilized to ready the PTB-XL ECG dataset for effective multi-label classification. The PTB-XL metadata is initially loaded from ptbxl_database.csv, and the scp_codes field of each record is parsed into (Python) dictionaries using `ast.literal_eval` to enable subsequent processing. A sorted vocabulary of diagnostic codes is constructed, and each recording's diagnoses are encoded as a fixed-length multi-hot vector, allowing the model to learn co-occurring conditions concurrently. The metadata table includes only essential columns: patient_id, filename_hr, age, sex, and labels. Age is normalised across individuals using Z-score scaling, while sex is binarised to 0/1. Wfdb.rdsamp analyzes the raw signal for each 12-lead ECG, then normalizes the Z-score by removing the mean and dividing by the standard deviation using a modest constant to reduce amplitude fluctuation. The processed signals are saved as NumPy (`.npy` files), with metadata enhanced by including file paths and multi-hot label vectors. This comprehensive dataset is exported in both {CSV and pickle} formats to facilitate reproducibility. To enable thorough patient-level evaluation, the script utilizes GroupKFold (n_splits=5) to divide the data into five non-overlapping folds, producing separate train_meta_fold{i}.csv and val_meta_fold{i}.csv files for each fold. Figures (5 & 6) shows examples of clean and malformed signals, with the latter being excluded during preprocessing to ensure the utilization of high-quality data.

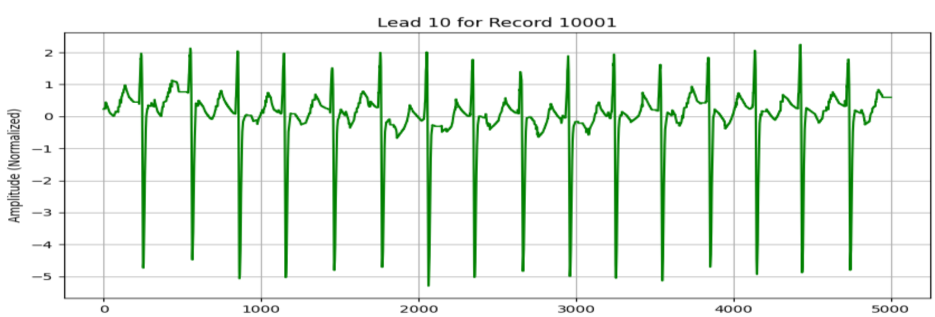

Fig. 5: Log 10001/ clean signal valid for analysis.

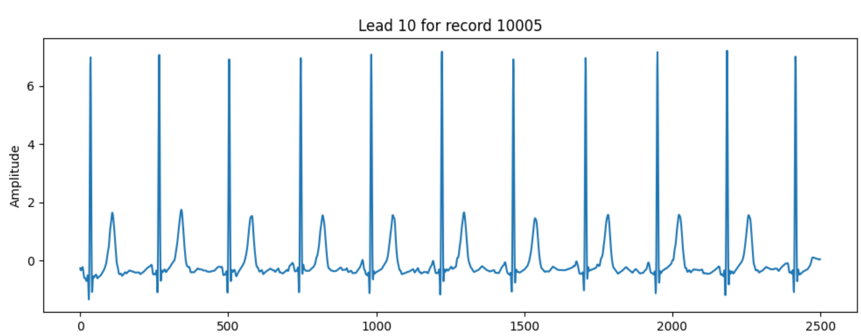

Fig. 6: Log 10005/ malformed signal excluded.

The figures (5 & 6) shows examples of clean and malformed signals, with the latter being excluded during preprocessing to ensure the utilization of high-quality data.

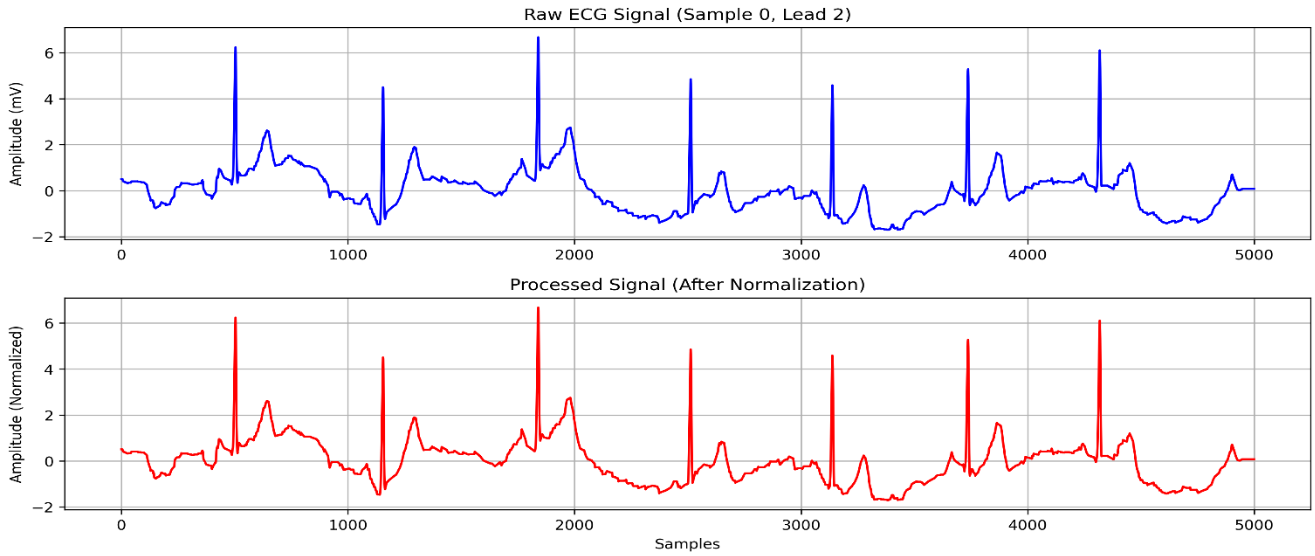

Fig 7: process signal after Normalization.

The figure (7) shows a comparison between the raw ECG signal (top, Sample 0, Lead 2) and the post-processing signal (bottom). The raw signal exhibits variability and baseline drift, while the post-Z-score normalization signal exhibits a centering around zero and reduced amplitude variation over time. This demonstrates how effective the normalization step is in making the signal more consistent and ready for automated processing.

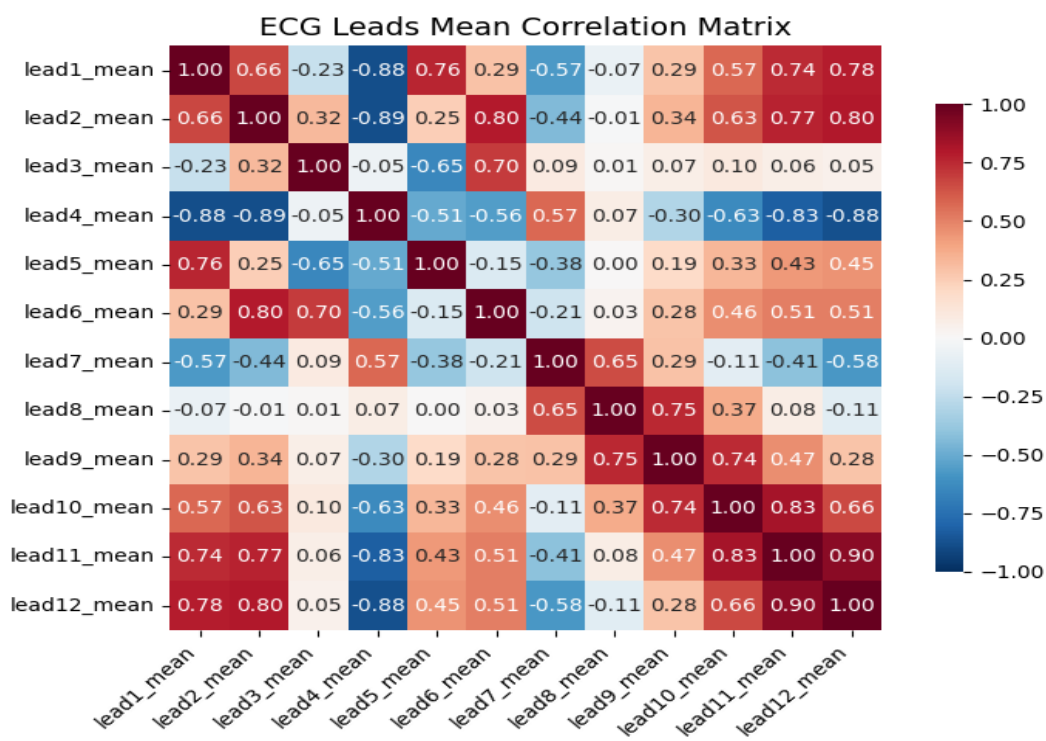

Fig. 8: Correlation Matrix Preprocessing.

The figure (8) shows the pearson correlation matrix computed between the mean ECG signals across all 12 ECG leads, each cell in the matrix represents the correlation value between the mean signals of two different channels, with values ranging from −1 (perfect inverse correlation) to +1 (perfect direct correlation), dark red shades indicate a strong positive correlation (e.g., lead11 and lead12 with values ≈0.90), while blue shades indicate a strong negative correlation (e.g., lead1 and lead4 with values ≈−0.88), this correlation matrix helps in identifying leads that provide similar or redundant information (high correlation), and those that may offer complementary or contrasting signals, which is useful for tasks like dimensionality reduction or optimal channel selection in model development.

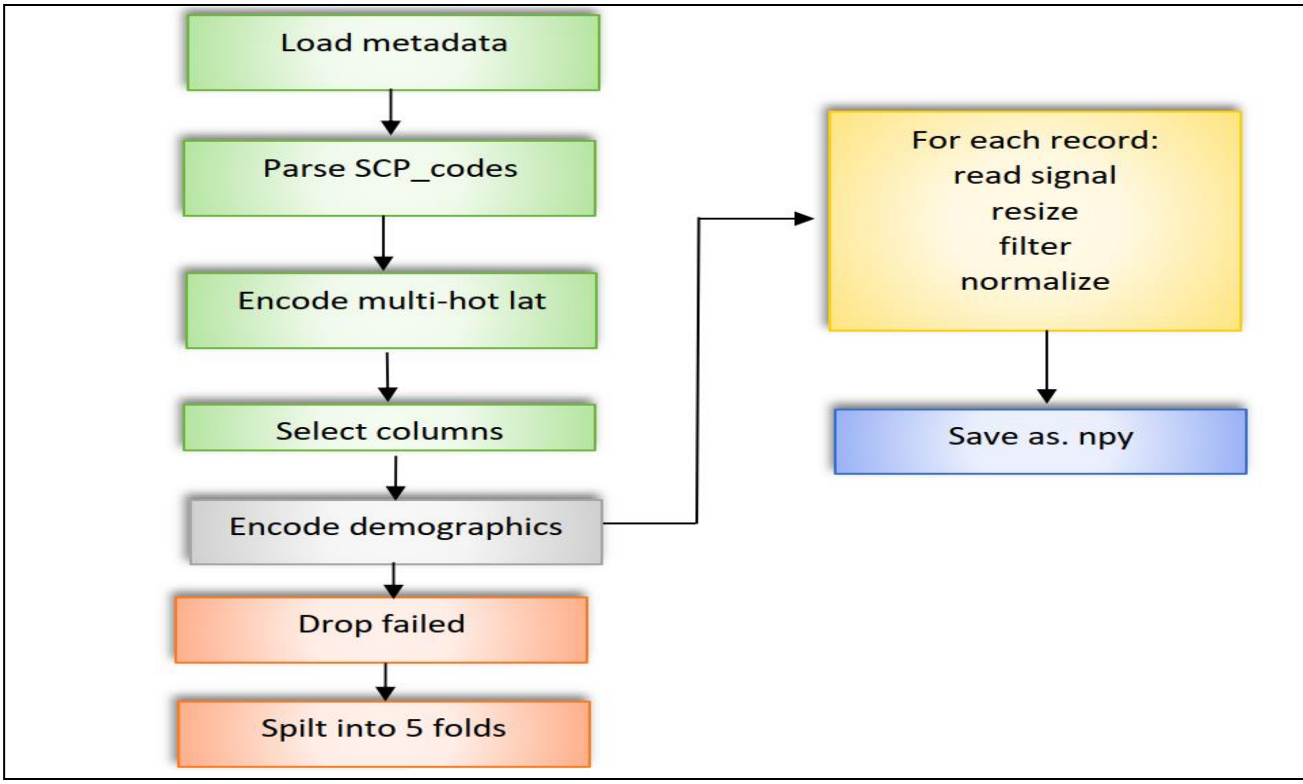

Fig. 9: The flowchart of preprocessing model

The figure (9) illustrates the flowchart detailing the ECG data preprocessing steps within the proposed model. The process initiates with the loading of metadata, subsequently converting the SCP_Codes field into Python dictionaries. This involves constructing a diagnosis code dictionary and encoding each record as a multi-hot vector. Key columns, including patient_id, filename_hr, age, sex, and labels, are selected. Demographic attributes are encoded, the sex converted to binary values (0/1) and age standardized using Z-score. The diagram outlines the signal processing steps for each record, which include reading the signal, adjusting its length, filtering, applying Z-score normalization, and saving the processed signal to a (`. npy` file). Ultimately, unsuccessful records are removed, and the metadata table is exported in both {CSV and pickle} formats. The data is partitioned at the ID level into five independent folds utilizing (Group K Fold) for the preparation of training and testing datasets.

Algorithm: Proposed ECG + Clinical Diagnosis System:

Begin

Step one: Create output dataset folders

Input:

– CSV_FILE

1. Load configuration

 1.1 read config.yaml → config

 1.2 set random seed, create results directory

2. Load & preprocess data

 2.1 (metadata_df, code_list) ← load_and_process_metadata(CSV_FILE)

 2.2 processed_df ← process_and_save(metadata_df)

 2.3 config['n_labels'] ← len(code_list)

  config['label_names'] ← code_list

Step two: Prepare cross-validation

 1.1 gkf ← GroupKFold(n_splits=config['n_splits'])

 1.2 results ← []

Step three: For each fold ← enumerate (gkf.split(processed_df, groups=processed_df.patient_id)):

 a. Split data:

  train_df ← processed_df.iloc[train_idx]

  val_df ← processed_df.iloc[val_idx]

```
  b. Oversample minorities:
    train_df ← oversample_multilabel(train_df)
  c. Instantiate model:
    classifier ← ECGClassifier(config)
    classifier.compile_model()
  d. Create data generators:
    train_gen ← ECGDataGenerator(train_df, batch_size, augment=True)
    val_gen   ← ECGDataGenerator(val_df,   batch_size, augment=False)
  e. Trainning:
    classifier.train(train_gen, val_gen)  # with early stopping, LR schedule, checkpointing
  f. Save best weights:
    classifier.model.save("best_model_fold{fold}.h5")
  g. Evaluation:
    y_true ← stack(val_df.labels)
    y_pred ← concat(classifier.model.predict(val_gen))
    metrics, report ← compute_metrics(y_true, y_pred)
    results.append(metrics)
    save report CSV
  h. Plotting:
    plot_confusion_matrix(y_true, y_pred, config['label_names'], fold)
    plot_roc_curves(y_true, y_pred, config['label_names'], fold)
    plot_pr_curves(y_true, y_pred, config['label_names'], fold)
  i. Cleanup:
    tf.keras.backend.clear_session()
Step four:  Aggregate & save results
  pd.DataFrame(results).to_csv("cv_results.csv")
  plot_metrics_comparison("cv_results.csv", config['label_names'])
```

❖ 5-5 Model Architecture Training & Evaluation Algorithms for Multi-Label Classification:

Upon completion of data preprocessing, load the PTB-XL metadata parse the SCP diagnostic codes into multi-hot label vectors, encode sex, normalize age, and retain only the essential columns: patient_id, filename_hr, age, sex, and labels. Each raw ECG record is read, resized to a fixed length of 5000 samples, normalized using per-lead Z-scores, and subsequently saved as a ('.npy') file. The processed dataset is divided at the patient level utilizing a (5-fold) (GroupKFold) cross-validation approach. An (ECGDataGenerator) implements real-time augmentations, including time warping, gaussian noise, and random scaling, for each training batch. The core architecture comprises an advanced encoder (It consists of: Multi-branch, Branch-level Attention Fusion, Positional Encoding, Transformer Self-Attention,BiLSTM, Dilated Convolution) utilizing multi-scale 1D convolutional blocks, which incorporate residual connections, transformer-based multi-head attention with positional encoding, bidirectional LSTM layers featuring skip connections, and dilated convolutions succeeded by global average pooling to extract deep ECG features. Demographic inputs, specifically age and sex, are processed through individual dense layers prior to concatenation with ECG features. The integrated representation is then transmitted through both shared and task-specific dense layers, incorporating dropout, resulting in a single sigmoid output for each label in the context of multi-label classification. The model is compiled with the (AdamW) optimizer and an adaptive focal loss for binary classification (with dynamic α and γ to balance positive/negative classes and focus on hard examples), combined with label smoothing to prevent overconfidence. Class weights, computed from tag frequencies are also applied to further mitigate class imbalance. Metrics tracked include [AUC, precision, recall, and accuracy]. Training is conducted with callbacks such as (EarlyStopping, ReduceLROnPlateau, ModelCheckpoint, and CSVLogger). Following each fold, performance is assessed using ROC AUC, average precision, and F1-score, and generate confusion matrices, ROC/PR curves, and cross-fold metric comparison plots are generated for visualization. The classification accuracy is calculated for each fold, Finally, the final average accuracy is calculated.

❖ 5-6 Evaluation Metrics:

The performance of the suggested model is assessed through the many evaluation criteria available are accuracy, recall, precision, and F1-score, all of which are used to gauge the performance of ML algorithms depending on a confusion matrix. Often referred to as a confusion matrix, an error matrix shows the count of correct and false predictions for every class, additionally, ROC and precision-recall curves are generated to compute AUC-ROC and average precision (AUPR), 5-fold patient-level cross-validation is performed: classification accuracy (and the other metrics) is calculated for each fold and then averaged to obtain the final performance estimate. Confusion matrices and ROC/PR plots are saved for further analysis. See equations (1–6).

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

$$2 - \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$3 - \text{Recall}\ (TPR) = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

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

$$5 - \text{FPR} = \frac{\text{FP}}{\text{FP} + TN}$$

$$6 - \text{AUC} - ROC = \int_{n}^{1} TPR(FPR)d(FPR)$$

6. Experimental Results & Discussion:

The proposed work is implemented using the processor Intel (R) Core (TM) i7-7700HQ CPU @ 2.80 GHz, RAM 16GB, an x64-based processor, windows 10, and 64-bit, and (PyCharm) 2025.1, which is an open-source platform used to manage python and use the python 3.11 interpreter, of the (18,869) PTB-XL patients (21,799 total ECG recordings), (18,459) had at least one cardiac diagnosis and (410) had no diagnostic labels. After preprocessing, the dataset was partitioned at the patient level using 5-fold GroupKFold cross-validation, yielding approximately (80 %) of patients per fold for training and (20%) for validation, during training a ML model, hyperparameters are important variables that control the model's learning process based on data. Fine-tuning a ML model involves adjusting its hyperparameter values after initial selection is referred to as hyperparameter tweaking. Adjusting hyperparameters improves model performance by determining optimal values for certain data sets, choosing appropriate hyperparameters improves model accuracy, speed of learning, adaptability to new data, and prevents overfitting/underfitting. This study uses hyperparameter tuning via with scikit-learn's "GridSearchCV" (wrapped with "KerasClassifier") to optimize key parameters, including learning rate, batch size, and dropout rate. The search grids were defined prior to cross-validation, and the combination yielding the highest average accuracy across folds was selected. Final model performance was evaluated using AUC-ROC, accuracy, precision, recall, and F1-score.

Table 3: The proposed model evaluation in terms of AUC-ROC, accuracy, recall, precision, and f1-score with the other models that use the same dataset (PTB-XL):

Method	AUC-ROC	Precision	F1-Score	Recall	Accuracy
ATCNN [11].	92.00	71.00	77.00	84.00	88.05
MRM-Ne [9].	93.59	—	73.74	—	97.96
ProtoECGNet [14].	92.48	—	—	—	—
xLSTM-ECG [15].	91.33	—	—	—	87.59
LPKF-Inception [12].	—	75.40	76.40	77.80	—
The proposed model.	**94.40**	**89.40**	**88.00**	**86.63**	**98.80**

The table 3 shows how the accuracy achieved by each of the proposed architectures. The (CNN + LSTM) baseline attained an accuracy of 90.12 %, which was substantially improved to 96.45 % by the (CNN + BiLSTM + Attention) model. Further incorporating data augmentation and assembling in the (CNN-LSTM + Attention + Augmentation + Ensemble) architecture raised the accuracy to 97.96 %, while the proposed model (CNN + BiLSTM + Attention + Augmentation+ Transformer) when collecting the stratified 5-fold results model achieved the highest accuracy of 98.8 %. The model was evaluated on the test set, achieving an accuracy of 98.8%, with a precision of 89.4%, recall of 86.63%, AUC-ROC of 94.4%, and an F1-score about of 88%. The performance of the proposed model was compared with traditional ML algorithms. The hybrid model outperformed these traditional models in all metrics, demonstrating its superior ability to handle complex heart disease data as well as the ability to distinguish between the less represented categories despite the difference in weights. The proposed model takes several essential steps that distinguish it from previous research [10, 12, 34]: instead of a single-scale TCN layer as in ATCNN [10], the current study adopt three multi-scale Conv1D branches (kernel sizes 5, 11, and 21 with expansion rates 1, 2, and 4) to capture fine and coarse patterns in the signal; then, the current study apply a learning attention mechanism to calculate differential weights for each branch using a Softmax layer and merge its outputs, unlike ProtoECGNet [12] which relied only on variance without learned subweights; third, the proposed model include a positional encoder and a multi-head Transformer block to enhance the capture of long-range relationships, which ATCNN, ProtoECGNet, and M2MASC-enabled CNN–BiLSTM; fourth, we integrate residual skip connections with BiLSTM to facilitate information flow and mitigate gradient vanishing, unlike the single-scale BiLSTM in [34]; Fifth, the proposed model add a branch to process demographic features (age and sex) via Embedding and Dense layers, while previous studies were limited to signal alone; Sixth, the proposed model adopts Adaptive Focal Loss with dynamic adjustment of alpha and gamma parameters in addition to (Label Smoothing), to address class distribution imbalance, this approach is used instead of relying on (Weighted) BCE in [10] and [34]; Seventh, the proposed

model use an advanced data generator that combines (Time Warping), (Gaussian Noise), (Random Scaling), (MixUp and CutMix techniques), unlike the simple additions in [10] and [12]. While the use of DL techniques such as CNNs, LSTMs, and attention mechanisms has been explored in previous studies for ECG classification, the proposed model presents a novel and integrated framework that surpasses existing approaches in multiple aspects. Unlike prior works that often rely on single-scale convolution or isolated components (e.g., CNN-only or LSTM-only architectures), this study introduces a multi-branch residual CNN structure with varying kernel sizes, fused with a transformer-based attention mechanism and BiLSTM layers equipped with skip connections. Furthermore, the inclusion of demographic features (age and sex) through dedicated embedding layers, as well as the use of advanced data augmentation techniques (e.g., MixUp, CutMix), and an adaptive focal loss function with label smoothing, are rarely combined in earlier works. These design enhancements contribute significantly to the model's ability to generalize across patient cohorts and handle imbalanced multi-label outputs. Therefore, although similar algorithms have been individually employed in the literature, this study distinguishes itself through its comprehensive hybrid design, achieving superior performance across all key metrics and offering a highly generalizable ECG classification pipeline.
1-Multi-branch CNN with residual connections (kernels=5,11,21),
2-Transformer attention fused with BiLSTM skip-connections,
3-Demographic embeddings (age/ sex),
4-Advanced augmentation (MixUp/CutMix),
5-Adaptive focal loss collectively outperform prior works (e.g., +0.8% accuracy over MRM-Net).

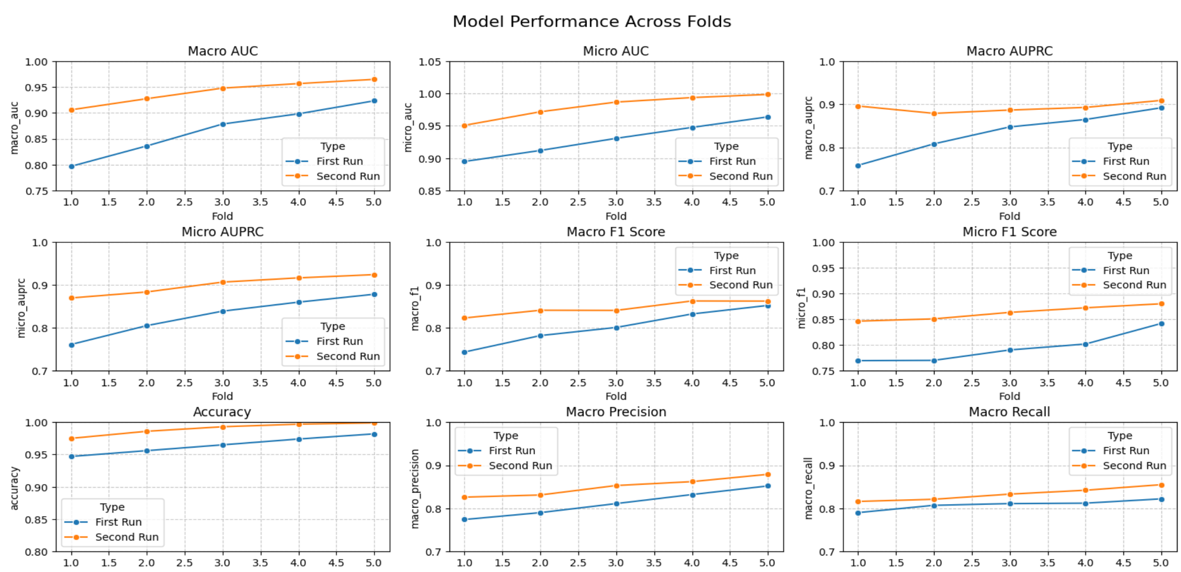

Fig. 10: The performance of the proposed hybrid model.

The figure (10) shows the result and model Performance across folds of proposed model.

The Area Under the Curve (AUC) scores and Receiver Operating Characteristic (ROC) curves were calculated for each class in addition to the previously mentioned evaluation metrics. With an average AUC-ROC of 0.944, the discrimination performance was outstanding overall. The ROC curves for the most common diagnostic classes (shown in figure 11), where positive and negative examples are clearly separable. In this study we developed a hybrid deep-learning pipeline that combines multi-scale convolutional blocks, transformer-based attention, bidirectional LSTM, and dilated convolutions to extract rich features from 12-lead ECG signals, while including demographic inputs (age and sex) via parallel embedding branches. The proposed model demonstrated consistently high discrimination (average accuracy= 0.98, AUC-ROC= 0.94) and precision-recall balance (precision= 0.90, recall= 0.89) across five folds of GroupKFold cross-validation, along with robust F1-scores (F1-scores= 0.88). The low inter-fold variability in these measures shows that the architecture generalizes well to unseen patient cohorts and preserves stability across data splits. Age and sex were included in the model as demographic covariates due to clinical relevance in the diagnosis of cardiovascular disease.

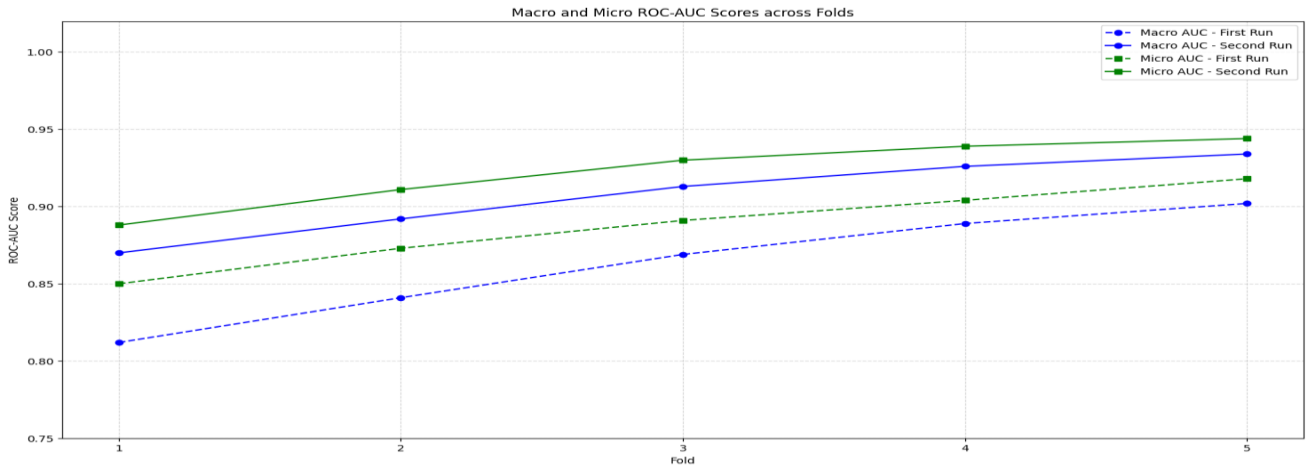

Fig. 11: Corresponding ROC-AUC curve and values in the results.

The figure (11) shows the ROC curve and AUC values corresponding in the results

Age is a key predictor of arrhythmias and ischemic patterns, while sex differences influence baseline ECG parameters and disease symptoms. Furthermore, these characteristics were available for the entire dataset, ensuring consistency in training and validation. Other demographic parameters (such as height, weight, and BMI) were not included due to significant missing values, which would require complex imputation strategies. However, future studies may explore the contribution of these additional factors using multimodal data fusion and interpretable AI tools to assess their incremental value. The current study attributes these results to three key design choices:

1-multi-branch receptive fields, which capture both fine-grained and coarse temporal patterns;

2-Attention-based fusion, which dynamically weights branch outputs;

3-Dilated convolutions with BiLSTM skip-connections, which enhance temporal context without excessive parameter growth.

Still, some restrictions have to be handled, First, while cross-validation indicates internal generalizability, the current study have not yet tested the model on completely independent cohorts or data from other recording devices; external validation is thus vital, Although the training loop already applies the AdaptiveFocalLoss class to address class imbalance, is currently use a fixed decision threshold of 0.2 for all labels. The future work should therefore explore per-class threshold optimization guided by precision-recall curve analysis to further boost F1-score for underrepresented diagnoses. Though this paper mentioned an ensemble step, it is not yet in the script, where assembling several fold-trained models or architectures could increase final accuracy over 98%. Looking forward, we intend to (1)/ conduct external validation on distinct ECG datasets (e.g., CPSC & UK Biobank), (2)/ apply and contrast focal loss against weighted BCE schemes, (3)/ optimize decision thresholds per diagnosis using Youden's index or F1 maximization, and (4)/ investigate model calibration methods (e.g., Platt scaling) to enhance probability estimates for clinical decision-making. Addressing these issues will help us to provide a thoroughly validated ECG classification system ready for use in actual diagnostic environments [21, 22, 23, 24, 25, 26, 27 ,28, 29, 30, 31, 32, 33, 35, 36].

7. Conclusion:

This paper presents a sophisticated hybrid framework for processing binary/multi-label ECG signals that combines multi-scale convolutional branches, transformer-based attention mechanisms, BiLSTM layers with skip connections, and dilated convolutions with efficient integration of demographic inputs (age and sex). Experiments over five folds of cross-validation showed great stability and better performance (accuracy = 0.98, AUC-ROC = 0.94, precision = 0.90, F1-score = 0.88), allowing precise identification of even rare classes. The multi-level design of feature extraction, loss adjustment to handle class imbalance, and careful performance threshold monitoring during training all contribute to this superiority. External assessment on independent datasets and real integration of adaptive focal loss, as well as threshold calibration for every diagnosis, still urgently calls, though. These results confirm the effectiveness of the system in rapidly detecting cardiac abnormalities during clinical ECG screening, the system could be further improved through ensemble analysis, calibration optimization, and threshold adjustment to achieve a high level of reliability and support real clinical decision-making, for future work, this study aims to:

1- Evaluating the proposal model with external datasets, such as CPSC & UK Biobank.
2- Implementing ensemble inference using multiple model variants to improve robustness.
3- Applying threshold optimization approaches (e.g., *Youden's Index*, *F1 maximization*).
4- Investigate calibrating approaches (e.g., *temperature scaling*, *Platt scaling*) to increase probability reliability.
5- Extend the model for real-time and mobile health applications with emphasis on deployment feasibility.

References

[1] H. Khraishah, B. Alahmad, R. L. Ostergard Jr., A. AlAshqar, M. Albaghdadi, N. Vellanki, M. M. Chowdhury, S. G. Al-Kindi, A. Zanobetti, A. Gasparrini, and S. Rajagopalan, "Climate change and cardiovascular disease: implications for global health," Nat. Rev. Cardiol., vol. 19, no. 12, pp. 798–812, Dec. 2022. https://doi.org/10.1038/s41569-022-00720-x

[2] Z. Wu and C. Guo, "Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management," Biomed. Eng. Online, vol. 24, no. 1, p. 23, Feb. 2025. https://doi.org/10.1186/s12938-025-01349-w

[3] S. S. Barold, "Willem Einthoven and the Birth of Clinical Electrocardiography a Hundred Years Ago," Cardiac Electrophysiology Review, vol. 7, no. 1, pp. 99–104, 2003. https://doi.org/10.1023/a:1023667812925

[4] G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour, N. C. Barengo, A. Z. Beaton, E. J. Benjamin, C. P. Benziger, and A. Bonny, "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study," J. Am. Coll. Cardiol., vol. 76, no. 25, pp. 2982–3021, Dec. 2020. https://www.jacc.org/doi/10.1016/j.jacc.2020.11.010

[5] World Health Organization, "cardiovascular diseases (CVDs) fact sheet," World Health Organization, Sep. 15, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). [Accessed: May 07, 2025].

[6] 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. https://doi.org/10.1038/s41598-021-97118-5

[7] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," Knowl.-Based Syst., vol. 227, p. 107187, Sep. 2021. https://doi.org/10.1016/j.knosys.2021.107187

[8] Shuvo SB, Alam SS, Ayman SU, Chakma A, Salvi M, Seoni S, Barua PD, Molinari F, Acharya UR. Application of Wavelet Transformation and Artificial Intelligence Techniques in Healthcare: A Systemic Review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2025 Jun;15(2): e70007. https://doi.org/10.1002/widm.70007

[9] A. Gour, M. Gupta, R. Wadhvani, and S. Shukla, "ECG based heart disease classification: advancement and review of techniques," Procedia Comput. Sci., vol. 235, pp. 1634–1648, Jan. 2024. https://doi.org/10.1016/j.procs.2024.04.155

[10] E. Prabhakararao and S. Dandapt, "Multi-label ECG classification using temporal convolutional neural network," arXiv:2306.03844 [Preprint], Jun. 6, 2023. Available: https://arxiv.org/abs/2306.03844

[11] W. Huang, N. Wang, P. Feng, H. Wang, Z. Wang, and B. Zhou, "A multi-resolution mutual learning network for multi-label ECG classification," in Proc. 2024 IEEE Int. Conf. Bioinformatics Biomed. (BIBM), Dec. 3–6, 2024, pp. 3303–3306. https://doi.org/10.1109/bibm62325.2024.10822741

[12] S. Sethi, D. Chen, T. Statchen, M. C. Burkhart, N. Bhandari, B. Ramadan, and B. Beaulieu-Jones, "ProtoECGNet: case-based interpretable deep learning for multi-label ECG classification with contrastive learning," arXiv:2504.08713 [Preprint], Apr. 11, 2025. Available: https://arxiv.org/abs/2504.08713

[13] L. Kang, X. Fu, J. Vazquez-Corral, E. Valveny, and D. Karatzas, "xLSTM-ECG: multi-label ECG classification via feature fusion with xLSTM," arXiv:2504.16101 [Preprint], Apr. 14, 2025. Available: https://arxiv.org/abs/2504.16101

[14] Q. Xiao, Y. Li, K. Lee, S. A. Mokhtar, I. Ismail, A. L. Pauzi, and P. Y. Lim, "Integrating learnable expert knowledge into deep learning-based multi-label ECG classification," Meas. Sci. Technol., Apr. 15, 2025. https://doi.org/10.1088/1361-6501/adcce7

[15] A. M. Gitau, A. Adeyemi, B. Tavashi, and B. J. Singstad, "[Re] Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL," medRxiv [Preprint], Jan. 29, 2025. https://doi.org/10.1101/2025.01.27.25321112

[16] K. Oueldji and N. Farah, "Machine learning models to help classification of cardiovascular disease," in Proc. Int. Conf. Emerging Intell. Syst. Sustain. Dev. (ICEIS), Aug. 31, 2024, pp. 215–230. https://doi.org/10.2991/978-94-6463-496-9_17

[17] V. Razin, A. Krasnov, D. Karchkov, V. Moskalenko, D. Rodionov, N. Zolotykh, L. Smirnov, and G. Osipov, “Solving the problem of diagnosing a disease by ECG on the PTB-XL dataset using deep learning,” in Proc. Int. Conf. Neuroinformatics, Oct. 12, 2023, pp. 13–21. https://doi.org/10.1007/978-3-031-44865-2_2

[18] S. Śmigiel, K. Pałczyński, and D. Ledziński, “ECG signal classification using deep learning techniques based on the PTB-XL dataset,” Entropy, vol. 23, no. 9, p. 1121, Aug. 28, 2021. https://doi.org/10.3390/e23091121

[19] K. Pałczyński, S. Śmigiel, D. Ledziński, and S. Bujnowski, “Study of the few-shot learning for ECG classification based on the PTB-XL dataset,” Sensors, vol. 22, no. 3, p. 904, Jan. 2022. https://doi.org/10.3390/s22030904

[20] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, “Deep learning for ECG analysis: benchmarks and insights from PTB-XL,” IEEE J. Biomed. Health Inform., vol. 25, no. 5, pp. 1519–1528, Sep. 9, 2020. https://doi.org/10.1109/jbhi.2020.3022989

[21] C. V. Nguyen, H. X. Nguyen, D. D. Minh, and C. D. Do, “Comparing deep neural network for multi-label ECG diagnosis from scanned ECG,” arXiv:2502.14909 [Preprint], Feb. 19, 2025. Available: https://arxiv.org/abs/2502.14909

[22] T. Rahman, R. Ahommed, N. Deb, U. K. Das, M. Moniruzzaman, M. A. Bhuiyan, F. Sultana, and M. K. Kausar, “ECG Signal Classification of Cardiovascular Disorder using CWT and DCNN,” J. Biomed. Phys. Eng., vol. 15, no. 1, pp. 77–92, Jan. 2025. https://doi.org/10.31661/jbpe.v0i0.2307-1636

[23] P. Wagner, N. Strodthoff, R. Bousseljot, W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset (version 1.0.1)," PhysioNet, 2020. [Online]. Available: https://doi.org/10.13026/x4td-x982

[24] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Scientific data. 2020 May 25;7(1):1-5. https://doi.org/10.1038/s41597-020-0495-6. https://doi.org/10.1038/s41597-020-0495-6

[25] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” Circulation, vol. 101, no. 23, pp. e215–e220, Jun. 13, 2000. https://doi.org/10.1161/01.cir.101.23.e215

[26] I. J. Selvam, M. Madhavan, and S. K. Kumarasamy, “Detection and classification of electrocardiography using hybrid deep learning models,” Hell. J. Cardiol., vol. 81, pp. 75–84, Jan. 1, 2025. https://doi.org/10.1016/j.hjc.2024.08.011

[27] A. S. Yahya and N. M. Edan, “ECG signal classification using hybrid and non-hybrid learning technologies,” Int. J. Percept. Cogn. Comput., vol. 11, no. 1, pp. 114–121, Jan. 30, 2025. https://doi.org/10.31436/ijpcc.v11i1.503

[28] A. A. Sahlu, S. M. Gizaw, M. H. Yimer, A. M. Mekbibu, M. W. Teshome, T. N. Yehualashet, “ConvNeXt based hybrid models with multi-modal feature fusion for ECG classification,” in: Artificial Intelligence and Human-Computer Interaction 2025, IOS Press, 2025, pp. 166–175. https://doi.org/10.3233/faia250118

[29] W. Chen, H. Wang, L. Zhang, and M. Zhang, “Temporal and spatial self-supervised learning methods for electrocardiograms,” Sci. Rep., vol. 15, no. 1, p. 6029, Feb. 19, 2025. https://doi.org/10.1038/s41598-025-90084-2

[30] M. Cai, J. Jiang, W. Huang, C. Liu, and R. Arcucci, “SuPreME: a supervised pre-training framework for multimodal ECG representation learning,” arXiv:2502.19668 [Preprint], Feb. 27, 2025. Available: https://arxiv.org/abs/2502.19668

[31] M. M. Hossain, M. S. Ali, M. M. Ahmed, M. A. Rakib, M. A. Kona, S. Afrin, M. K. Islam, M. M. Ahsan, S. M. Raj, and M. H. Rahman, “cardiovascular disease identification using a hybrid CNN-LSTM model with explainable AI,” Informatics Med. Unlocked, vol. 42, p. 101370, Jan. 1, 2023. https://doi.org/10.1016/j.imu.2023.101370

[32] P. K. Shrivastava, M. Sharma, and A. Kumar, “HCBiLSTM: a hybrid model for predicting heart disease using CNN and BiLSTM algorithms,” SN Comput. Sci., vol. 4, no. 2, p. 172, Feb. 1, 2023. https://doi.org/10.1016/j.measen.2022.100657

[33] Ö. Yıldırım, P. Pławiak, R. S. Tan, and U. R. Acharya, “Arrhythmia detection using deep convolutional neural network with long duration ECG signals,” Comput. Biol. Med., vol. 102, pp. 411–420, Nov. 1, 2018. https://doi.org/10.1016/j.compbiomed.2018.09.009

[34] V. Pandey, U. K. Lilhore, R. Walia, R. Alroobaea, M. Alsafyani, A. M. Baqasah, and S. Algarni, “Enhancing heart disease classification with M2MASC and CNN-BiLSTM integration for improved accuracy,” Sci. Rep., vol. 14, no. 1, p. 24221, Oct. 16, 2024. https://doi.org/10.1038/s41598-024-74993-2

[35] A. L. Golande and T. Pavankumar, “Optical electrocardiogram-based heart disease prediction using hybrid deep learning,” J. Big Data, vol. 10, no. 1, p. 139, Sep. 9, 2023. https://doi.org/10.1186/s40537-023-00820-6

[36] V. K. Sudha and D. Kumar, “Hybrid CNN and LSTM network for heart disease prediction,” SN Comput. Sci., vol. 4, no. 2, p. 172, Jan. 18, 2023. https://doi.org/10.1007/s42979-022-01598-9