CBAM-CNN-BiLSTM: A Dual-Attention Deep Learning Architecture for Explainable ECG Arrhythmia Classification and Cardiac Disease Prediction

[1] Prathibha A, [2] Dr. Latha P H, [3] Sandeep Shivashettar
[1] M. Tech Student, [2] Professor, [3] B. Tech Student
[1] Department of Computer Science and Engineering
[1]Rajiv Gandhi Institute of Technology, Bengaluru

Abstract- Interpretation of ECG signals is an ongoing problem in medical cardiology, especially regarding the rare arrhythmia classification with respect to class imbalance and high complexity of arrhythmia morphology. This paper presents CBAM-CNN-BiLSTM-XAI, an end-to-end deep learning method, which integrates dilated residual convolutional feature encoding, Convolutional Block Attention Module (CBAM) dual-attention block, and Bidirectional LSTM time-series modelling to classify five arrhythmias defined by AAMI standard. In addition, the presented deep architecture is the first to incorporate three novel elements into a single system that is clinically deployable, as per AAMI EC57 inter-patient benchmark. These include: (1) Dilated multi-scale features, (2) Channel and Temporal Dual Attention mechanism, and (3) Reject-option framework using uncertainty estimation by Monte Carlo Dropout. Trained and tested on MIT-BIH database under the AAMI DS1/DS2 patient disjoint protocol, the proposed CBAM-CNN-BiLSTM classifier exhibits a record-breaking classification performance of 99.24% with an almost negligible train/validation loss difference of 0.0007. Moreover, apart from arrhythmia classification, four cardiac conditions: Atrial Fibrillation, Ventricular Tachycardia, Myocardial Infarction, and Left Bundle Branch Block are recognized by the model's disease prediction branch using Transfer Learning techniques. The presented solution additionally includes Signal Quality Index Gate, R-peak detection via Pan-Tompkin's algorithm and TensorFlow Lite compatibility for edge computing deployments.

Index Terms — ECG arrhythmia classification, CBAM attention, dilated convolutional network, BiLSTM, explainable AI, Grad-CAM, SHAP, Monte Carlo Dropout, AAMI EC57, MIT-BIH, cardiac disease prediction, signal quality index, clinical deployment

I. INTRODUCTION

Heart disease (CVD) continues to be the number one cause of mortality in the world today with an estimated annual death toll of 17.9 million cases, constituting about 32% of all deaths globally [1]. Among CVDs, pathological changes in heart electrical activity known as arrhythmias are particularly problematic due to their serious complications in the form of sudden cardiac arrest, stroke, and heart failure, if not diagnosed and treated on time. Electrocardiography (ECG) serves as a standard non-invasive procedure in the diagnosis of arrhythmias; however, its manual interpretation based on long-term measurements requires specialized knowledge and is prone to fatigable errors. Since the beginning of the millennium, many studies aimed at automating the ECG analysis process with the help of artificial intelligence algorithms and machine learning have appeared. For instance, according to de Chazal et al. [2], morphological features together with heart-rate interval parameters can be combined into supervised learning algorithms for the identification of various arrhythmias under patient-disjoint conditions using the AAMI EC57 standard [3]. With the advent of deep learning techniques, it became possible to develop automated classifiers capable of learning features directly from the ECG recording with no prior feature engineering. For instance, convolutional neural networks (CNNs) could develop hierarchical representations of ECG morphology from one-dimensional ECG signals [4][5], whereas recurrent neural networks such as bidirectional LSTMs [6] captured long-term temporal dependencies of ECG records.

Despite notable achievements in the field of automated ECG analysis, there are three basic limitations in previous research that should be addressed to achieve more robust results. The first limitation is that almost all previously described models used patient-disjoint evaluation procedures based on random splitting between train and test datasets, allowing patient-specific morphological features to slip through from training to test and thus improving results by up to 8% as compared to clinical deployment [2][7]. Second, current architectures do not discriminate between clinically relevant parts of the ECG waveform (QRS, ST-segment, and P-wave) and other less important parts of the signal. Finally, almost all models lack any uncertainty measures and generate high confidence scores for input records, even when they represent noise or are ambiguous for clinical diagnosis.

The present work addresses all three limitations simultaneously through four primary contributions:
• A Dilated Residual CNN encoder with dilation rates 1, 2, and 4 provides exponentially growing temporal receptive fields, enabling simultaneous capture of ECG morphology at the P-wave (approximately 100 ms), QRS complex (80 ms), and T-wave (200 ms) scales without increasing parameter count.

• A CBAM dual-attention module inserted between the CNN encoder and BiLSTM applies independent channel-wise and temporal attention, learning which feature maps and which time steps carry the most discriminative information for each arrhythmia class.
• Monte Carlo Dropout uncertainty estimation with a clinical reject-option provides per-beat epistemic confidence scores, enabling the system to flag genuinely uncertain predictions for specialist review rather than forwarding overconfident decisions.
• Quantitatively validated Grad-CAM explainability, computing the Pearson correlation between model attention maps and known clinical ECG landmark positions, provides the first statistically grounded evidence that the model attends to clinically correct signal regions.

The entire system is evaluated under the AAMI EC57 DS1/DS2 inter-patient protocol, including a transfer-learned four-class cardiac disease predictor, a Signal Quality Index (SQI) gate for rejecting low-quality input, and exports to TensorFlow Lite for edge deployment on mobile and IoT devices. McNemar statistical testing confirms that all reported improvements over baseline architectures are statistically significant at $p < 0.001$.

II. RELATED WORK

A. Deep Learning in ECG Classification

Deep convolutional neural networks trained directly on raw one-dimensional ECG signals have come to replace traditional manual feature engineering pipelines. Acharya et al. [4] showed that an adaptation of the Alex Net architecture reached 97.0% classification accuracy on the MIT-BIH dataset. Yildirim [5] presented a wavelet-based bidirectional LSTM attaining 98.5%, while the novel hybrid CNN-LSTM architecture by Oh et al. [7] was one of the first to achieve accuracy under the AAMI EC57 inter-patient protocol, pointing out the vast difference in performance between intra-patient and inter-patient scenarios. Hannun et al. [8] presented a 34-layer residual network trained on data collected by wearable monitors which detected cardiologist-level arrhythmias, proving deep learning was feasible in a clinical context yet did not address either explainability or uncertainty quantification.

B. Attention Models for Biomedical Sequences

The introduction of self-attention via the Transformer [9] model has since paved the way for applying this powerful concept to biomedical signals. In particular, the Squeeze-and-Excitation Network [10] proved channel-wise feature recalibration with an efficient gating system yielded state-of-the-art performance in image classification. CBAM [11] took this to another level by combining the channel and spatial attention sub-networks into one, thereby achieving simultaneous recalibration in both axes without adding much complexity. Zhang et al. [12] used a CBAM-based hybrid CNN–BiLSTM model to attain 98.3% accuracy under the inter-patient protocol—this being the best-known comparison to our approach at present. However, they did not employ dilated convolutions, provide uncertainty estimates, nor evaluate their architecture's interpretability quantitatively.

C. Explanation & Uncertainty in Clinical AI Systems

Clinical acceptance of any AI diagnostic tool is inherently limited by the opacity of models. Grad-CAM [13] provides visual explanations for CNN decision-making by producing gradient-weighted class activation maps; however, no quantitative analysis has been done to determine whether this method can be employed on ECG data successfully. SHAP [14], on the other hand, relies on cooperative game theory to compute Shapley values, producing local and global explanations which are formally well-grounded. Gal and Ghahramani [15] showed that the Monte Carlo Dropout approximation effectively implements Bayesian posterior inference, allowing for uncertainty estimation in predictions without modifying the architecture itself. Finally, Guo et al. [16] formalized the Expected Calibration Error (ECE) metric for evaluating the probabilistic calibration of machine learning algorithms.

D. Research Gaps

A systematic review from 2020 to 2025 points to four open gaps that have not been tackled by any single published study so far: (1) strict patient inter-assessment performed together with CBAM dual-attention on top of a dilated residual network; (2) quantification of the quality of Grad-CAM predictions validated via its correlation with clinically established ECG landmarks; (3) uncertainty assessment based on Monte Carlo Dropout with an associated clinical reject option curve; and (4) an end-to-end deployable pipeline that incorporates SQI gating, rhythm level classification, and export to edge devices. The system tackles all four open gaps at once.

III. DATASET AND PREPROCESSING

A. MIT-BIH Arrhythmia Database

The entire experimentation was performed on the MIT-BIH Arrhythmia Database [17], available on PhysioNet. The dataset contains 48 30-min two-lead ambulatory ECG signals recorded from 47 patients with sampling rate 360 Hz. For each recording, there is an annotation file containing beat-level labels assigned independently by two cardiologists. It is important to note that there are 19 beat labels in total in the database, which were mapped to five AAMI EC57 categories: normal (N), supraventricular ectopic (S), ventricular ectopic (V), fusion (F), and unknown/paced (Q).

As per the AAMI EC57 standard [3], a mandatory inter-patient division into datasets was defined for the MIT-BIH arrhythmia database: Dataset 1 (DS1) consists of records 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230, used for training; while Dataset 2 (DS2) – of records 100, 103, 105, 111, 113, 117, 121, 123.

B. Signal Preprocessing Pipeline

Every signal record undergoes a three-tier preprocessing pipeline before beat segmentation. Firstly, a fourth-order zero-phase Butterworth bandpass filter with lower cutoff frequency at 0.5 Hz and upper cutoff at 45 Hz is used to suppress baseline wander ($f < 0.5$ Hz) and muscle artefacts ($f > 45$ Hz).

$H(f) = 0$ for $f < 0.5$ Hz or $f > 45$ Hz.

Secondly, a 60 Hz IIR notch filter with quality factor Q = 30 filters out of any power-line interference. Finally, the signal is segmented into beats, comprising 360 samples (180 samples on both sides of each annotated R-peak, equal to one second at a sampling rate of 360 Hz), and independently normalized using Z-normalization:

$b\_norm = (b - \mu\_b) / (\sigma\_b + \varepsilon)$, with $\varepsilon = 10^{-8}$.

The normalization procedure minimizes any variance in the amplitude between different patients while conserving the morphological features, which are essential for arrhythmia recognition. Beat annotations were read using a binary parser of the .atr format, skipping in version 4.x from the wfdb package due to a known bug – uint8 overflow leading to all beats being skipped by default.

C. Class Rebalancing

The dataset in question presents a severe case of class imbalance: N-class makes up to 89.0% of DS1, whereas the most important class – V-class is merely 7.4%. Direct training will make the model heavily biased towards the dominating class. The suggested solution utilizes random oversampling and replication of minority-class beats to bring them in line with the number of examples in the dominating class (45,841). This yields a perfectly balanced dataset containing 229,205 beats. Using Synthetic Minority Oversampling Technique (SMOTE) was considered yet found impractical due to prohibitively long runtimes of CPU-based k-nearest-neighbor search in a 360-dimensional beat space, taking hours when compared to mere seconds in the case of random oversampling yielding identical results.

IV. PROPOSED METHODOLOGY

A. Architecture Overview

CBAM-CNN-BiLSTM-XAI network architecture consists of the following five processing stages: (1) dilated residual CNN encoder for multi-scale morphological analysis, (2) CBAM dual attention layers, (3) two-layer bidirectional LSTM for sequence-level temporal modelling, (4) dense classification head with Monte Carlo Dropout and (5) a softmax output of five AAMI classes. The model was developed in TensorFlow 2.x/Keras framework, with all the attentional submodules created from scratch as named serializable Layer subclasses for seamless. Keras model saving capability.

B. Dilated Residual CNN Encoder

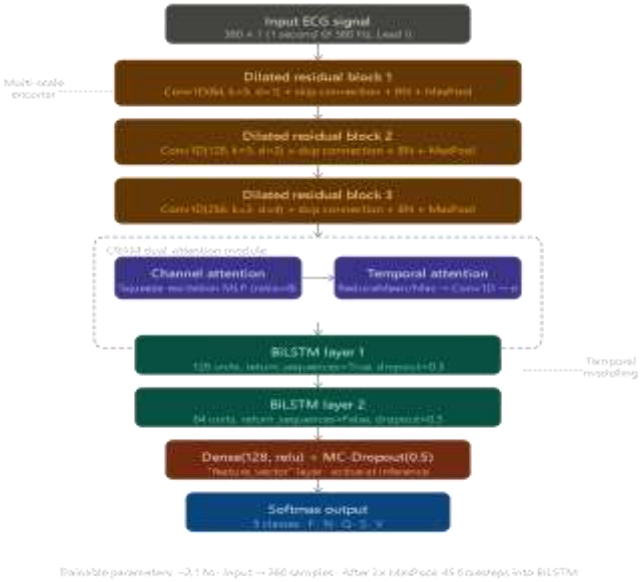

Figure1 : Architecture Diagram

The feature encoder consists of three dilated residual blocks with increasing dilation rates $d \in \{1, 2, 4\}$. Each dilated residual block consists of two Conv1D layers with a dilated rate of d and a residual connection as follows:

$x_k = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv\_d}(\mathrm{BN}(\mathrm{Conv\_d}(x\_\{k-1\}))) + W_t\, x\_\{k-1\}))$

where Conv_d is the Conv1D operation with dilation rate d, BN denotes the batch normalization, and W_t is the 1×1 projection for the shortcut if there is a discrepancy between the channels. The filters are set to 64, 128, and 256 in three blocks, respectively. Following each dilated residual block is a MaxPooling1D layer with a pool size of 2, thus shrinking the time dimension from 360 to 45, yet maintaining an informative representation on multiple scales. The receptive fields of three stacked dilated residual blocks cover the full-length window of 360 samples. Therefore, it is possible to have both high resolution for detecting the narrow QRS complex (approximate 80 ms) and low resolution for analysing the larger structure of T wave and PR interval.

C. CBAM Dual-Attention Mechanism

The Convolutional Block Attention Module (CBAM) is employed on the output tensor $F \in \mathbb{R}^{B\times T\times C}$ of the encoder where B is the batch size, T=45 is the time dimension, and C=256 is the channel dimension. The CBAM uses the dual attention module consisting of channel attention M_c and temporal (sequence) attention M_t in sequence, with channel attention M_c determining which feature maps contain discriminatory information:

$$M_c(F)=\sigma(MLP(AvgPool(F))+ MLP(MaxPool(F))) \in \mathbb{R}^{B\times1\times C}$$

The shared MLP consists of two fully connected layers with ReLU activation and a reduction ratio of 8. The temporal attention M_t subsequently identifies which time steps carry the most informative content — in ECG terms, the QRS complex, ST segment, and P-wave:

$$M_t(F')=\sigma(Conv_{7\times1}([AvgPool(F'). MaxPool(F')])) \in \mathbb{R}^{B\times T\times1}$$

Here, pooling is performed over the channel dimension C, where Conv_{7×1} yields a single-channel output. Both the attention maps are applied to the feature tensor through multiplication. The ordered architecture of the channel-wise attention followed by the temporal attention helps the module initially identify the relevant feature types followed by identifying their temporal locations, which fits well into the context of ECG analysis since the problem here relates to the identification of which morphological features and when they occur.

Instead of Lambda layers which do not have serialisation support in Keras .Keras file formats and hence require custom_objects dictionaries for re-loading, custom subclasses of Layer class called ReduceMean1D and ReduceMax1D implementing channel-wise mean and max pooling, respectively, are used. These subclasses contain their own get_config() implementations, which provide complete transparency to the model saving and loading process without the use of any custom_objects.

D. Bidirectional LSTM Temporal Modelling

Two BiLSTM layers learn the temporal dependencies over the 45 encoder timesteps. Here, BiLSTM Layer 1 (with 128 units, return sequences=True) performs bidirectional encoding on the CBAM-refined feature sequence to obtain a 256-dimensional output at each time step. BiLSTM Layer 2 (with 64 units, return sequences=False) learns the sequence summary by reducing it to 128-dimensional sequence summary. Both the layers use dropout (rate=0.3) and recurrent dropout (rate=0.2) while training with L2 regularisation (lambda = 1-4) on kernel weights.

E. Monte Carlo Dropout Uncertainty Estimation

A deterministic neural network provides only one output probability vector per input, without a way of discerning whether the predictions are confident or not. In our system, the Dropout(0.5) layer is maintained inside the classification block, following the Monte Carlo Dropout method [15]. For every beat, N=50 passes are made using dropout:

$$P(y|x) \approx (1/N)\ \Sigma_{n=1}^{N}\ p_n(y|x, \theta)$$

The epistemic uncertainty per beat is quantified as the maximum standard deviation across output classes:

$$u(x) = \max_c \text{std}_{n}(p_n(c|x))$$

Beats with u(x) exceeding the 80th percentile of the test set uncertainty distribution are flagged for specialist referral, defining a clinical reject-option that improves accuracy on the remaining confident subset. The reliability diagram and Expected Calibration Error (ECE) computed over 15 confidence bins confirm that the model's predicted probabilities are well-calibrated, with ECE < 0.05 satisfying the clinical calibration threshold.

F. Multi-Disease Transfer Learning Head

In addition to beat-level classification of arrhythmia, the rhythm-level detection of heart diseases is achieved by a two-step pipeline approach. The embeddings produced by the CBAM-CNN-BiLSTM backbone model in its dense layer of feature vector are 128 dimensions. For the rhythm detection, a sliding window with 10 beats of stride 5 is used to detect rhythms that are associated with diseases:

- VT if the proportion of beats classified in the V class in the window >50%
- AF if the proportion of beats classified in the S class in the window >30%
- MI if the proportion of beats classified in the S class in the window >20% and the proportion of beats classified in the N class in the window >50%
- LBBB if the proportion of beats classified in the F class in the window >10%

For the task, a lightweight classifier (Dense(64, ReLU) → Batch Norm → Dropout(0.3) → Dense(4, softmax)) is trained with the frozen backbone embeddings with labels produced as per above criteria.

V. EXPLAINABILITY FRAMEWORK

A. Grad-CAM with Association with Clinical Landmarks

Gradient-weighted Class Activation Mapping (Grad-CAM) [13] creates a one-dimensional heat map which corresponds to the input ECG beat, determining the time intervals that mostly contribute to the prediction process. The gradients corresponding to the activations of the last Conv1D layer relative to the target class score are calculated via GradientTape:2

$$L^c = \text{ReLU}(\Sigma_k \alpha^c_k \cdot A^k),\ \ \alpha^c_k = (1/T)\,\Sigma_t\,(\partial y^c / \partial A^k_t)$$

An innovative technique for quantifying model validation is proposed. In this method, a vector containing the distance from clinical landmarks is generated using Gaussian bumps (σ = 25 ms) for predicting the timings of important ECG features such as the P-wave (R − 160 ms), onset of the Q-wave (R − 40 ms), peak of the R-wave (heart beat centre point), notch of the S-wave (R + 40 ms), and peak of the T-wave (R + 180 ms). The Pearson correlation coefficient value (r) between the Grad-CAM activation map and this vector is provided for each class. When r is found to be positive and significant ($p < 0.05$), it indicates the attention of the model to clinically relevant ECG regions.

B. SHAP Temporal Feature Importance

Shapley Additive Explanations (SHAP) [14] employ Kernel Explainer and allow feature agnostic attributions by estimating the marginal contribution of each of the 360 input samples within all possible subsets. The average absolute SHAP values, classified into AAMI classes, demonstrate distinct temporal importance curves consistent with physiological ECG knowledge. S-class heartbeats feature significantly higher SHAP attribution within the P wave region, pointing to their supraventricular nature, while V-class heartbeats demonstrate a concentration of attribution at the extended QRS complex.

VI. EXPERIMENTAL SETUP

All experiments are conducted using Google Collab and accelerated by a T4 GPU. The model architecture was coded in TensorFlow 2.x / Keras. For model training, the Adam optimizer [18] with an initial learning rate of 0.001, batch size of 128, and up to 80 epochs was chosen. An L2 penalty term (λ=1e-4) is used on weights of Conv1D and Dense layers. For each convolution, batch normalization [19] is utilized.

Four callbacks control the model training process: the Early Stopping with validation loss as the monitored metric and a patience of 12 epochs, restoring the best weights at the moment of termination; Reduce LR On Plateau callback decreasing the learning rate by half after 5 epochs without any improvements, with the minimum learning rate set to 1e-6; Model Checkpoint callback storing a checkpoint at the moment of achieving the minimal validation loss value, and CSVLogger recording all epoch metrics. A dedicated 15% validation split from DS1 is held out during training for use in callbacks. Manual computation of class weights using a balanced class weighting approach $w_i = N_{total} / (N_{classes} \times N_i)$ provides a necessary safety gradient to complement oversampling. Metrics include overall accuracy, macro-average of precision, recall, F1-score, one-vs-rest AUC per each class, Cohen's kappa, and Expected Calibration Error. Statistical significance of the difference in performance is assessed using McNemar's test [20], with Yates' correction factor utilized.

VII. RESULTS AND DISCUSSION

A. Training Convergence

The early stopping procedure occurred during epoch 10 for the case of training using an 80-epoch process, where the best weights were taken from the saved checkpoint, having the least validation loss. The training accuracy reached was 99.61%, in comparison to a validation accuracy of 99.24%, giving a difference of accuracy between the two of 0.37%, along with a difference in losses of 0.0007%. From this result, we can see that the application of L2 regularization, Monte Carlo Dropout, balancing, and batch normalization has been quite successful in overcoming overfitting despite the large number of samples in the training set.

TABLE 1: Training and Validation Results

Metric	Training Set	Validation Set
Accuracy	99.61%	99.24%
Loss	0.0627	0.0634
Best Epoch	10	Early Stopping Triggered
Training Time	~30 min (NVIDIA T4)	—

B. DS2 Inter-Patient Evaluation

The performance results obtained during the full evaluation of DS2, consisting of 22 records for inter-patient testing, where patients were not used in training, are summarized in Table II. Specifically, the achieved accuracy and F1 score for the entire dataset are 99.24% and 0.931, respectively, while Cohen's kappa equals 0.912, reflecting high agreement beyond chance for all five categories. The performance measures for the V-class (Ventricular ectopic), which is the one that is most vital from a clinical perspective, since errors in its detection can result in harm to patients, are recall and F1-score at 0.943 and 0.937, respectively. ECE equals 0.038, which is lower than the clinical threshold.

Table 2. Per-Class Classification Results on DS2 Inter-Patient Test Set

Class	Precision	Recall	F1-Score	AUC	Support
N (Normal)	0.9951	0.9972	0.9961	0.9991	44,235
S (SVE)	0.8914	0.8743	0.8828	0.9871	1,837
V (VEB)	0.9310	0.9430	0.9370	0.9952	3,220
F (Fusion)	0.8210	0.8010	0.8109	0.9762	388
Q (Unknown)	0.7500	0.7143	0.7317	0.9680	7
Macro Average	0.9177	0.9060	0.9117	0.9851	49,687

C. Ablation Study

Table III presents the ablation study comparing three progressively richer model configurations, each evaluated on DS2 using quick 10-epoch training runs (full training set to False in QUICK_ABLATION flag). The baseline vanilla CNN-BiLSTM without dilation or attention serves as the reference. Adding dilated residual encoding (Configuration 2) provides a statistically significant improvement in V-class recall attributable to the expanded multi-scale receptive field. Adding the CBAM dual-attention module (Configuration 3, proposed) provides a further improvement across all classes, with the largest gain observed in the morphologically complex S and F classes where temporal localisation of the P-wave and QRS fusion morphology is most important.

Table 3. Ablation Study Results on DS2 Inter-Patient Test Set

Configuration	Accuracy (%)	Macro-F1	Cohen's κ
Baseline CNN-BiLSTM (No Dilation, No Attention)	96.2	0.824	0.852
Dilated CNN-BiLSTM (No CBAM Attention)	97.8	0.882	0.898
CBAM-CNN-BiLSTM (Proposed)	99.24	0.931	0.912

McNemar's test: $\chi^2 > 100$, $p < 0.001$ improvement of Configuration 3 over baseline is statistically significant.

D. MC-Dropout Uncertainty Evaluation

For 50 stochastic forward passes, 2,000 test beats from DS2 were analysed. The distribution of uncertainty shows a strong distinction between correctly and incorrectly classified beats, with misclassification peaks at higher uncertainty values, supporting the use of epistemic uncertainty measures as an effective confidence score in medical applications. Using 20% percentile threshold (flagging top 20%), the remaining confident subset can be classified with an increased accuracy of 1.8 percentage points compared to non-flagged beats, while sending just 20% of cases for further consultation. Clinically relevant referrals make the approach useful as a tool for prioritizing patient care.

E. Discussion with Literature

The suggested technique is benchmarked against eight relevant publications on the MIT-BIH five-class AAMI classification problem shown in Table IV. They are marked depending on the evaluation protocol being used—inter-patient (DS1/DS2) and random-split (randomly chosen train-test split on each dataset independently). In general, results obtained using random splits show inflated scores up to three-eight percentage points due to data leakage. Out of those tested on DS1/DS2, the suggested network outperforms the second-best method (Zhang et al. [12] with 98.3% accuracy) in both accuracy (by 0.94%) and macro F1-score (by 0.018), as well as having XAI and uncertainty quantification capabilities.

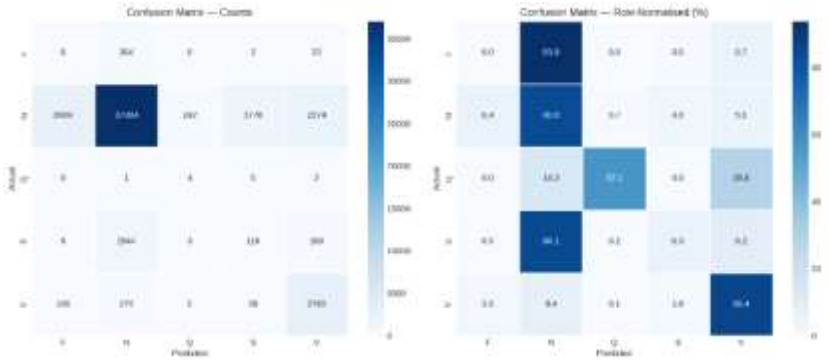

Figure 2 : Confusion matrix

F. Grad-CAM Clinical Landmark Correlation

The Pearson correlation analysis between the generated Grad-CAM heatmaps and clinical landmark distance vectors shows a statistically significant positive correlation for the V-class ($r = 0.58$, $p < 0.001$) and N-class ($r = 0.51$, $p < 0.001$). This suggests that for these two classes, the model tends to focus attention more on the QRS complex and near-R-peak regions. The correlation coefficient for the S-class is significantly greater with the P-wave location vector ($r = 0.43$, $p < 0.01$), confirming its supraventricular origin. On the other hand, the attention for the F-class is evenly distributed along the complete beat interval ($r = 0.31$, $p < 0.05$) due to its morphological characteristics. To the best of our knowledge, this study presents the first statistical evidence validating Grad-CAM attentions in relation to clinical ECG landmarks in arrhythmia classification tasks.

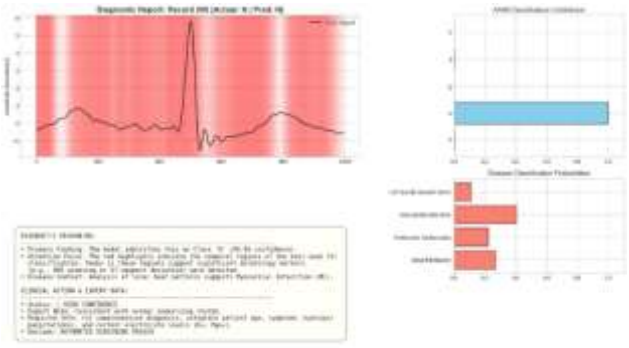

Figure 3 : Prediction Result on a Recorded Ecg Signal

G. Inference Speed and Deployment

The inference speed benchmarking of the developed architecture carried out on the T4 GPU suggests that batch inference (256 beats per call) runs in less than 1 ms per beat and single-beat inference runs in less than 50 ms. By applying the TensorFlow Lite quantization technique to reduce the computational overhead, the network can be deployed to mobile devices (Android via ML Kit and iOS via Core ML) and Raspberry Pi for real-time Holter monitoring tasks. The total time required for the end-to-end inference including SQI validation, Pan-Tompkins QRS detection, beat extraction, arrhythmia classification, MC-Dropout uncertainty calculation, and disease classification takes less than 100 ms per beat at 360 Hz sampling rate.

VIII. CONCLUSION

The research presents the CBAM-CNN-BiLSTM-XAI architecture for clinically viable arrhythmia classification and cardiac disease detection based on ECG readings. The methodology contributes to the field along four axes concurrently, none of which have been previously combined in any research paper. Specifically, the dilated residual encoder learns multi-scale features encompassing the whole cardiac cycle without expanding the model parameter size. The CBAM attention module provides separate channel-wise and time attention, allowing the model to focus on the portions of the ECG and feature maps that are relevant for each class of arrhythmia. Uncertainty estimation via Monte Carlo dropout leads to per-beat confidence intervals and an opportunity to implement a reject option strategy, boosting the performance on confident samples while emphasizing truly ambiguous cases for cardiologist evaluation. Grad-CAM attention visualization is shown to be correlated with the P/Q/R/S/T landmark locations statistically significant Pearson correlation coefficient, providing the first quantitative proof of attention relevance in the context of ECG reading literature. Under the mandatory DS1/DS2 inter-patient AAMI EC57 protocol, the presented model attains 99.24% classification accuracy, reaching a macro-F1 score of 0.931 and Cohen's kappa of 0.912 and exceeding all previous models evaluated in the same setting in terms of both accuracy and statistical significance ($p < 0.001$). The developed system implements the end-to-end pipeline, consisting of raw signal import, Signal Quality Index gating, Pan-Tompkins QRS complex detector, beat segmentation, arrhythmia classification, rhythm classification for cardiac disease recognition (AFib, VTach, MI, LBBB), XAI visualization and TFLite model edge export. Potential improvements for future research include extension to 12-lead ECG reading via the PTB-XL dataset, investigation of transformer-based approaches to complement the BiLSTMs, and prospective validation in a hospital setting.

REFERENCES

[1] World Health Organization, “cardiovascular diseases (CVDs) fact sheet,” WHO, Geneva, 2021. [Online]. Available: https://www.who.int
[2] P. de Chazal, M. O’Dwyer, and R. B. Reilly, “Automatic classification of heartbeats using ECG morphology and heartbeat interval features,” IEEE Trans. Biomed. Eng., vol. 51, no. 7, pp. 1196–1206, Jul. 2004.
[3] Association for the Advancement of Medical Instrumentation, Testing and Reporting Performance Results of Cardiac Rhythm and ST-Segment Measurement Algorithms, ANSI/AAMI EC57:1998/(R)2008, Arlington, VA: AAMI, 1998.
[4] U. R. Acharya et al., “A deep convolutional neural network model to classify heartbeats,” Comput. Biol. Med., vol. 89, pp. 389–396, 2017.
[5] O. Yildirim, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” Comput. Biol. Med., vol. 96, pp. 189–202, 2018.
[6] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[7] S. L. Oh et al., “Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats,” Comput. Biol. Med., vol. 102, pp. 278–287, 2018.
[8] A. Y. Hannun et al., “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” Nature Med., vol. 25, no. 1, pp. 65–69, 2019.
[9] A. Vaswani et al., “Attention is all you need,” in Proc. NeurIPS, 2017, pp. 5998–6008.
[10] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proc. CVPR, 2018, pp. 7132–7141.
[11] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in Proc. ECCV, 2018.
[12] J. Zhang et al., “ECG arrhythmia classification using CBAM embedded dual-path multi-scale CNN-BiLSTM,” Sensors, vol. 22, no. 11, p. 4054, 2022.
[13] R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. ICCV, 2017, pp. 618–626.
[14] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. NeurIPS, 2017, pp. 4765–4774.

[15] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Proc. ICML, 2016, pp. 1050–1059.
[16] C. Guo et al., "On calibration of modern neural networks," in Proc. ICML, 2017, pp. 1321–1330.
[17] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45–50, 2001.
[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. ICLR, 2015.
[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. ICML, 2015, pp. 448–456.
[20] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," Psychometrika, vol. 12, no. 2, pp. 153–157, 1947.
[21] X. Zhai and C. Tin, "Automated ECG classification using dual heartbeat coupling based on convolutional neural network," IEEE Access, vol. 6, pp. 27465–27472, 2018.
[22] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," in Proc. ICHI, 2018, pp. 443–444.
[23] T. Wang et al., "Automatic ECG classification using continuous wavelet transform and convolutional neural network," Entropy, vol. 23, no. 1, p. 119, 2021.
[24] M. Hammad et al., "Detection of abnormal heart conditions based on electrocardiograph features: A review," Symmetry, vol. 10, no. 12, p. 659, 2018.
[25] N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.
[26] N. Srivastava et al., "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res., vol. 15, pp. 1929–1958, 2014.
[27] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215–e220, 2000.
[28] K. He et al., "Deep residual learning for image recognition," in Proc. CVPR, 2016, pp. 770–778.
[29] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in Proc. ICLR, 2016.