International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Jun-24).)

https://dx.doi.org/10.12785/ijcds/1501120

Mitigating Data Variability and Overfitting in Deep Learning Models for Atrial Fibrillation Detection Using Single-Lead ECGs

Khadidja Benchaira[1], Salim Bitam[1] and Zineb Djihane Agli[1]

[1]*Department of Computer science, University Mohamed Khider of Biskra, Algeria.*
Received 18 Sep. 2023, Revised 2 Apr. 2024, Accepted 6 Apr. 2024, Published 1 Jun. 2024.

Abstract: Despite the growing potential of deep learning in diagnosing Atrial Fibrillation (Afib), challenges such as overfitting and limited generalizability continue to persist. These limitations are accentuated in single-lead ECGs generated from wearable devices, which frequently suffer from inadequate annotation and substantial data variability. This study seeks to address these challenges by enhancing both the accuracy and generalizability of Afib detection algorithms. We introduce Afib-CNN, a specialized Convolutional Neural Network engineered for 9-second, single-lead ECGs. The architecture comprises ten convolutional blocks and three fully connected layers, focusing on computational efficiency. To mitigate data variability, we apply advanced pre-processing techniques like Moving Average by Convolution Filter (MAConv) and Minimum-Maximum Normalization. Further dataset refinement is achieved using z-score normalization and a shifted-length overlapping technique. The effectiveness of our model is rigorously validated across three distinct ECG databases, demonstrating robust intra- and inter-patient generalizability. Employing 10-fold stratified cross-validation, Afib-CNN exhibits exemplary performance, achieving mean F1 scores of 98%, 97%, and 99% on the CinC2017, CPSC2018, and MIT-AFIB datasets, respectively. The model also attains an F1 score of 98% on the CinC2017 test set. Comparative analyses demonstrate that Afib-CNN successfully balances high performance, computational efficiency, and robust generalization. These characteristics render it well-suited for practical clinical deployment.

Keywords: Convolutional neural network (CNN) , Arrhythmia classification , Short single-lead ECG recordings , ECG Data Variability , Overfitting , Wearable ECG.

1. INTRODUCTION

Cardiac arrhythmias, notably atrial fibrillation (Afib), represent a significant cause of heart-related morbidity and mortality, challenging the conventional diagnostic paradigms [1]. The traditional reliance on the 12-lead Electrocardiogram (ECG) for assessing cardiac electrical activities, while effective, falls short in continuous, real-time monitoring [2], [3]. The emergence of wearable technologies has been a game-changer, offering continuous single-lead ECG monitoring capabilities, thus opening new avenues for real-time arrhythmia detection.

Recent advancements in wearable ECG monitoring devices have shown promising results in improving the detection and management of atrial fibrillation (Afib). These devices enable prolonged monitoring, increasing the likelihood of detecting transient arrhythmic events that might be missed during a standard ECG exam [4]. Despite their potential, the accuracy of wearable devices in diagnosing complex arrhythmias and their integration into clinical practice remain areas of ongoing research [5], [6].

The technological evolution in the field has been significantly supported by the application of deep learning techniques, such as Convolutional Neural Networks (CNNs) [7], [8], [9], [10], [11], Recurrent Neural Networks (RNNs) [12], [13], and hybrid models like Convolutional Recurrent Neural Networks (CRNNs) [14], [15], [16]. These advancements have propelled Afib classification from single-lead ECG recordings to new heights. The ability of these models to extract meaningful features from ECG signals has led to significant improvements in classification accuracy [17].

In a notable study, Rahul and Sharma [18] introduced a comprehensive approach for Afib detection utilizing both 1-D electrocardiogram signals and their time-frequency representation, achieving remarkable accuracy through sophisticated preprocessing and normalization techniques, along-

E-mail address: khadidja.benchaira@univ-biskra.dz, s.bitam@univ-biskra.dz, zinebdjihane@gmail.com

side classification with a Bi-directional Long Short-Term Memory (Bi-LSTM) network . Similarly, Zhang et al. [19] explored innovative training strategies to address overfitting in deep learning models for Afib detection, focusing on binary classification between normal and Afib ECG signals, and utilizing wearable ECG data for model training.

However, despite these advancements, the field confronts significant challenges, particularly data scarcity and ECG variability, which complicate feature extraction and model training, often leading to overfitting [20]. The variability inherent in single ECG recordings—due to factors such as respiratory sinus arrhythmia, patient movement, and autonomic tone changes—poses substantial hurdles in achieving reliable arrhythmia detection. This variability, both within and between recordings, challenges the generalization of machine learning models to unseen data, impeding the algorithms' ability to generalize effectively [21], [19].

Addressing these challenges, this paper introduces the "AFIB-CNN" model, a novel approach combining advanced preprocessing techniques with a custom-developed CNN architecture to enhance arrhythmia classification from single-lead ECG signals. Our contributions are as follows:

- **Advanced Pre-processing Techniques:** Employing sophisticated preprocessing methods like the Moving Average by Convolution Filter (MAConv), minimum-maximum normalization, and z-score normalization to mitigate signal variability and enhance data quality.
- **Addressing Data Imbalance and Overfitting:** Utilizing a shifted-length overlapping technique to augment the dataset and balance arrhythmia class representation, thereby improving model robustness and mitigating overfitting.
- **Customized CNN Architecture:** Optimizing the AFIB-CNN architecture for short, single-lead ECG recordings, addressing the unique challenges of arrhythmia detection, and balancing computational efficiency with diagnostic accuracy.
- **Comprehensive Validation:** Extensively testing our methodology across multiple independent datasets (CINC2017, CPSC2018, and MIT-AFIB), demonstrating robustness and reliability in intra- and inter-patient scenarios.

The remainder of this paper is structured to provide a detailed exploration of our methods and findings. Section 2 elaborates on the materials and methods employed in our study, Section 3 presents and discusses the experimental results, and Section 4 concludes the paper, highlighting the significant contributions of our work and outlining avenues for future research.

2. Materials and methods

A. *Overview of the approach*

Our study employs a structured methodology to classify arrhythmias from ECG data, utilizing three major steps illustrated in Figure 1. We detail our approach by focusing on dataset utilization, data preparation techniques, and the specifics of model development and evaluation.

- **Data Preparation** : We leverage three independent datasets (CINC2017, CPSC2018, and MIT-BIH) for a diverse representation of arrhythmias. The data undergo partitioning into training/validation (90%) and test (10%) sets. Preprocessing includes noise reduction and normalization, with a shifted-length overlapping technique for augmentation and balancing in the training/validation set. The test set employs non- overlapping segmentation for ensuring realistic evaluation scenarios. Each segment is accurately relabeled to match its original classification.
- **Model Development**: The Afib-CNN model was fine-tuned through an optimization process, targeting efficient arrhythmia classification from single-lead ECGs. We conducted a grid search to adjust input lengths, kernel sizes, and layer configurations, aiming to reduce complexity and enhance performance. This effort led to an optimized model that balances classification accuracy with computational efficiency, ensuring practical applicability.
- **Training, Validation, and Testing**: We adopt a stratified cross-validation approach for training and validation, ensuring a balanced representation of classes. The model's performance is then rigorously evaluated on the test set, focusing on accuracy, and F1 score to assess its classification efficacy comprehensively.

Figure 1 presents a flowchart of our ECG rhythm classification paradigm, from data preprocessing to model evaluation, encapsulating the study's structured approach.

B. *Data description*

The variability inherent in ECG recordings—such as differences in quality, amplitude, duration, sampling rates, arrhythmia types, and the number of leads—poses significant challenges in ECG classification. To address these challenges, we adopted a consistent framework for ECG data evaluation, focusing on single-lead recordings with a standardized sampling rate of 300 Hz and segment lengths of 9 seconds (equivalent to 2700 samples).

Our selection of datasets—CinC2017, CPSC2018, and MIT-AFIB—was guided by their comprehensive representation of arrhythmias, including atrial fibrillation, which is of particular interest to our study. These datasets are recognized for their reliability and diversity, making them ideal for developing and validating our model. Below, we

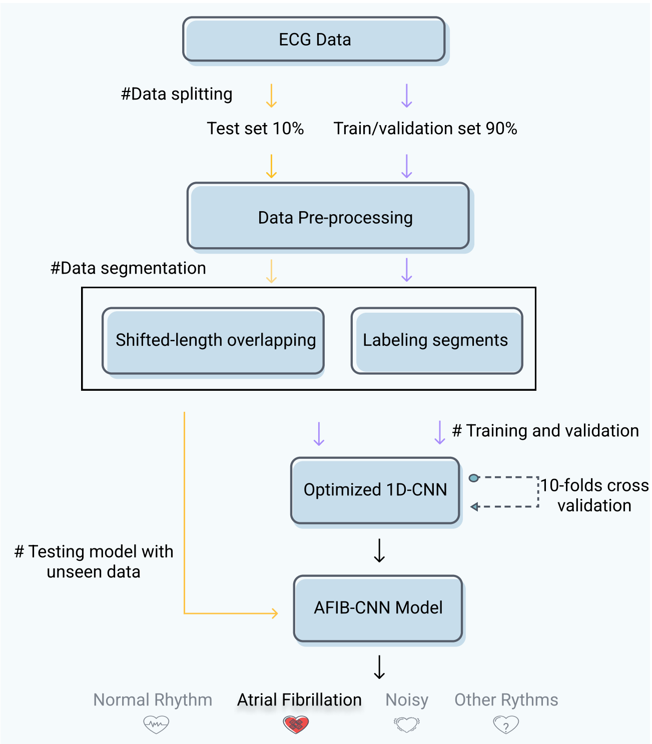

Figure 1. Flowchart of the proposed ECG rhythms classification paradigm. ECG: Electrocardiogram; CNN: Convolutional neural network.

outline the datasets used, emphasizing their relevance to our objectives.

1) PhysioNet Computing in Cardiology Challenge 2017 (CinC2017) [22] :
   - Source: AliveCor healthcare device
   - Sampling rate: 300Hz
   - Leads: 1 (single-lead or lead I)
   - Original number records: 8528
   - Duration: 9-60 seconds
   - Classes: include 4 categories (N: Normal; AF: Atrial fibrillation; P: Noisy; O: Other rhythms).
2) 2018 China Physiological Signal Challenge (CPSC2018)[23]:
   - Source: china Hospitals
   - Sampling rate: 500Hz
   - Leads: 12 (all leads)
   - Original number records: 6877
   - Duration: 6-144 seconds
   - Classes: include 9 categories(N: Normal; AF: Atrial fibrillation; I-AVB : First-degree atrioventricular block ; LBBB: Left bundle brunch block; RBBB: Right bundle brunch block; PAC: Premature atrial contraction; PVC: Premature ventricular contraction; STD: ST-segment depression; STE: ST segment elevated).
3) MIT-BIH Atrial Fibrillation Database (MIT-AFIB)[22]:
   - Source: Boston's Hospital
   - Sampling rate: 250Hz
   - Leads: 2 (lead I & lead II)
   - Original number records: 25
   - Duration: 10 hours
   - Classes: include 2 categories(N: Normal; AF: Atrial fibrillation).

Our methodological approach, including data preparation and model development, was designed to accommodate the specific characteristics of these datasets, ensuring a robust and adaptable solution to ECG classification across diverse arrhythmia presentations.

C. *Data pre-processing*

Electrocardiogram (ECG) data, especially when collected via wearable devices, frequently present with significant noise and variability. Addressing this challenge necessitates sophisticated preprocessing techniques aimed at enhancing signal quality and ensuring consistency across recordings.

The preprocessing journey begins with the application of a bandpass filter across each ECG recording. This filter, set with a pass-band of 3-45 Hz, plays a pivotal role in addressing baseline drift issues and minimizing high-frequency noise— factors that can significantly mask the actual ECG signal.

After the initial filtering, the variability within recordings—stemming from factors like respiratory sinus arrhythmia, patient movements, and autonomic tone fluctuations— becomes our next frontier. To combat this, we deploy the Convolution-Moving Average (Conv-MA) technique. This approach is engineered to smooth out ECG signals by dampening short-term fluctuations, all while preserving the integrity of the underlying physiological data.

Noise in ECG signal processing comes in various forms: baseline wander, high-frequency disturbances, muscle artifacts, and noise from electrode movements are but a few [24], [25]. Our preprocessing strategy is carefully designed to tackle each of these noise types head-on, ensuring a cleaner signal for analysis.

a) Eliminating Baseline Wander:
Through the strategic setting of the bandpass filter's pass-band, we effectively remove low-frequency shifts caused by patient motions or breathing—commonly known as baseline wander. This removal is facilitated by filtering out frequencies below 3 Hz, which, being largely non-physiological, tend to distort the ECG waveform.

b) Reducing High-frequency Noise:
Electrical interference— whether from external sources or equipment—falls under high-frequency noise, which the bandpass filter curtails by setting its upper cutoff at 45

Hz. This adjustment ensures that only frequencies pertinent to the ECG's physiological signals are maintained, thereby improving signal clarity and interpretability.

c) Mitigating Muscle Artifacts and Electrode Movements:

The Conv-MA method is instrumental in smoothing the ECG signal, significantly mitigating the effects of muscle artifacts and electrode movements. By averaging the signal across a window of 24 samples (corresponding to 0.08 seconds at a 300 Hz sampling rate), this method reduces transient noise without obscuring the ECG's diagnostic features.

The foundation of the MAConv filter's operation is captured in the following mathematical expression:

$$y(t) = \frac{1}{M} \sum_{n=0}^{M-1} x(t-n) * h(n) \tag{1}$$

Here, $y(t)$ denotes the output or the smoothed signal, M the number of points within the moving average window, $x(t-n)$ the input ECG signal at time $t-n$, and $h(n)$ the impulse response of the Hamming window employed in the convolution process [26], [27], [28]. This method ensures a smoother signal, reducing noise and variability while maintaining essential waveform characteristics crucial for accurate arrhythmia detection.

To further standardize the ECG signals and tackle inter-recording variability, we apply a minimum-maximum normalization technique. This process scales the signal amplitude to a standard range between 0 and 1. Moreover, each ECG recording is subjected to z-score normalization before segmentation, aligning morphological differences between segmented snippets and their original recordings.

Figure 2 visually summarizes the extensive preprocessing steps undertaken, illustrating the transformation of ECG recordings through each noise reduction and signal standardization stage.

D. Data balancing and augmentation

Following the initial stages of data preprocessing and the individual standardization of each electrocardiogram (ECG) recording, our methodology incorporates a shift-length overlapping technique to fulfill three critical objectives, thereby enhancing the performance and robustness of our one-dimensional convolutional neural network (1D-CNN) model. This technique is pivotal in standardizing the length of ECG recordings, augmenting the dataset, and balancing the class distribution to avert model bias towards the majority class.

1) Rationale and impact

The decision to employ the shift-length overlapping technique is grounded in its ability to address several challenges inherent in the processing of ECG data for machine learning applications. ECG recordings, by their nature, vary significantly in length and exhibit imbalances in class distribution, representing different cardiac conditions. These variances can lead to substantial challenges in training machine learning models, such as overfitting and bias towards more frequently represented classes. The shift-length overlapping technique directly addresses these challenges through the following mechanisms:

- **Standardization of ECG Lengths:** By dividing each ECG recording into segments of a uniform length, this technique ensures that the input to the 1D-CNN is consistent, facilitating more effective learning and generalization. This standardization is crucial for capturing the temporal dynamics of cardiac signals within a fixed-dimensional input space.

- **Augmentation of the Dataset:** Data augmentation is a widely recognized approach to enhance the robustness of machine learning models by artificially increasing the diversity of training data. The shift-length overlapping method augments the dataset by creating multiple, slightly shifted segments from each ECG recording. This process not only expands the dataset but also introduces variations that help in reducing overfitting, as the model learns to recognize cardiac patterns across slightly different segments.

- **Balancing Class Distribution:** Imbalanced datasets can lead to models that are biased towards the majority class, impairing their ability to accurately identify less represented classes. By adjusting the overlap between segments, this technique allows for a controlled increase in the representation of minority classes, thereby balancing the class distribution. This balance is crucial for developing a model that performs well across all classes of cardiac conditions.

Additionally, the overlapping technique ensures the preservation of continuity between adjacent ECG segments, maintaining morphological consistency within individual recordings. This continuity is essential for retaining the contextual integrity of cardiac signals, allowing the model to better learn and generalize from the temporal patterns present in ECG data.

The impact of employing the shift-length overlapping technique is quantitatively illustrated in Table I, which details the significant increase in data size and the improvement in class balance for the CinC2017 dataset. This balanced distribution is further supported by Shannon entropy calculations, providing a comprehensive measure of the dataset's uniformity across different classes.

2) Selecting the fit parameters for ECG data segmentation

To estimate the class distribution, we initially base our calculations on fixed-length segments. This is necessary as the original recordings have variable lengths, which could affect the real estimate of the distribution. The formula used

Figure 2. Illustration of the preprocessing steps applied to ECG signals from the CinC2017 dataset, demonstrating the transformation of recordings from Normal (N), Atrial Fibrillation (AF), Other rhythms, and Noisy categories. Panels (A) show the original ECG recordings. Panels (B) illustrate the effect of filtering and reducing intra-recording variations, showcasing the signal smoothing and noise reduction. Panels (C) depict the process of reducing inter-recording variations, standardizing the signal amplitude across different recordings. Finally, panels (D) demonstrate the discrimination of the morphology of each ECG recording, highlighting the morphological consistency achieved before segmentation. These visualizations collectively underscore the comprehensive preprocessing strategy employed to enhance signal quality and consistency for subsequent analysis.

TABLE I. The size and distribution of the data before and after employing the shifted-length overlapping technique for the CinC2017 dataset.

Recordings class	Original number recordings	Not-overlapping segments	ShiftLen overlapping segments
N	5050	21077	27794
Afib	783	3078	25185
O	2456	10832	26191
P	284	936	14760
Total	**8528**	**35923**	**93930**
Balance	**0.71**	**0.70**	**0.98**

to calculate the number of fixed-length segments for each class $c_k \in \{c_1, c_2, c_3, \ldots, c_m\}$ is given by:

$$Number\_Segments_{c_k} = \sum_{i=1}^{n} \frac{RecordLen_i}{WindowLen} \tag{2}$$

Here, RecordLen_i is the original length of the i-th recording, and WindowLen = 2700 samples corresponds to a fixed length of 9 seconds.

After calculating the number of segments for each class without overlap, we estimate the optimal overlap length ShiftLen_{c_k} for each category to maintain class balance.

$$ShiftLen_{c_k} = rnd(WindowLen - (1 - \frac{NumberOfSegments_{c_k}}{\sum_{j=1}^{m} NumberOfSegments_{c_j}})) \tag{3}$$

To calculate the number of overlapping segments for each class, we use:

$$Shift\_Number\_Segments_{c_k} = \sum_{i=1}^{n} \frac{RecordLen_i - shiftlen_{c_k}}{WindowLen - shiftlen_{c_k}} \tag{4}$$

Finally, the balance of the data is evaluated using Shannon entropy (H).

$$Balance = \frac{H}{log_2(k)} = \frac{-\sum_{j=1}^{k}(\frac{Number\_Segments_{c_k}}{\sum_{j=1}^{m} Number\_Segments_{c_j}})log_2(\frac{Number\_Segments_{c_k}}{\sum_{j=1}^{m} Number\_Segments_{c_j}})}{log_2(k))} \tag{5}$$

As indicated in Table I, using the overlapping technique has enlarged the dataset and balanced the different categories with a balance ratio of 0.98.

3) ECG data segmentation

As illustrated in Figure 3, the optimized parameters for the shift length ($ShiftLen$) are applied to each category in the dataset. The continuous ECG signals are then segmented into fixed overlapping windows of 9 seconds in duration, which equates to 2700 samples at a sampling rate of 300Hz. Each segmented window is subsequently relabeled in accordance with its root recording.

For segments that are shorter than 9 seconds, zero-padding is employed to reach the fixed window size. On the contrary, segments that exactly match the fixed window size undergo no modifications. The segmentation algorithm is elaborated further in (appendixA).

E. Deep convolutional network architecture

The primary objective of this study is to design a highly efficient and low-complexity deep neural network for the classification of rhythm categories in short, single-lead ECG records.

Given the experimental nature of deep learning, various hyperparameters were empirically tested to identify a configuration that maximizes accuracy while minimizing complexity.

The final architecture, referred to as Afib-CNN, is outlined in Table II. The model accepts as input a 1D timeseries vector consisting of 2700 ECG samples (corresponding to 9 seconds of data). The architecture comprises ten convolutional blocks each followed by a max-pooling layer, and then three fully connected layers for classification. The model is trained to classify ECG recordings into one of several categories from three independent datasets: Cinc2017 (4 classes), CPSC2018 (9 classes), and MIT-AFIB (2 classes).

To mitigate the risk of overfitting and reduce computational complexity, three dropout layers with a dropout rate of 50% are incorporated into the architecture. These layers effectively prune a random subset of nodes during training, thereby simulating a variety of network structures and making the model more robust.

Additionally, the architecture utilizes a total of 1,984 feature maps with an overall capacity of 242,592 and an optimized number of 1,641,796 parameters. After each convolutional layer, a max-pooling layer condenses the feature maps, preserving only the most significant features essential for arrhythmia classification.

During the training phase, the Rectified Linear Unit (ReLU) function serves as the activation function for each convolutional block to prevent the vanishing gradient problem. For the fully connected layers, two dense layers activated by ReLU are employed, while the output layer uses the Softmax function for classification.

F. Deep model training and evaluation

As previously stated, this study incorporates three widely recognized arrhythmia datasets: CinC2017, CPSC2018, and MIT-AFIB. To rigorously evaluate the proposed model, the data was partitioned into two subsets before any preprocessing steps: 90% of the data was allocated to the training/validation set, and the remaining 10% was designated as the test set.

Subsequently, each subset underwent independent preprocessing to remove noise and standardize ECG signals, employing techniques like moving average filtering, min-max normalization, and z-score normalization for each recording.

1) Model training and validation

The model training and validation process consists of the following five stages:

1) Application of the shifted-length overlapping technique to standardize the length of all ECG recordings to 9 seconds while also balancing and augmenting the dataset.
2) Utilization of the stratified k-fold cross-validation strategy for data partitioning into ten mini train-test subsets (k-folds). The model iteratively trains on $(k-1)$ folds, validating against the remaining fold.
3) Implementation of the Adam optimizer over 300 epochs with a batch size of 30 and a constant learning rate of 0.001. The model's loss and accuracy metrics, derived from cross-entropy, are assessed after each epoch on both the training and validation datasets, as illustrated in Figures 5 and 6.
4) Performance evaluation of the model using metrics such as the confusion matrix, accuracy, and F1 score.

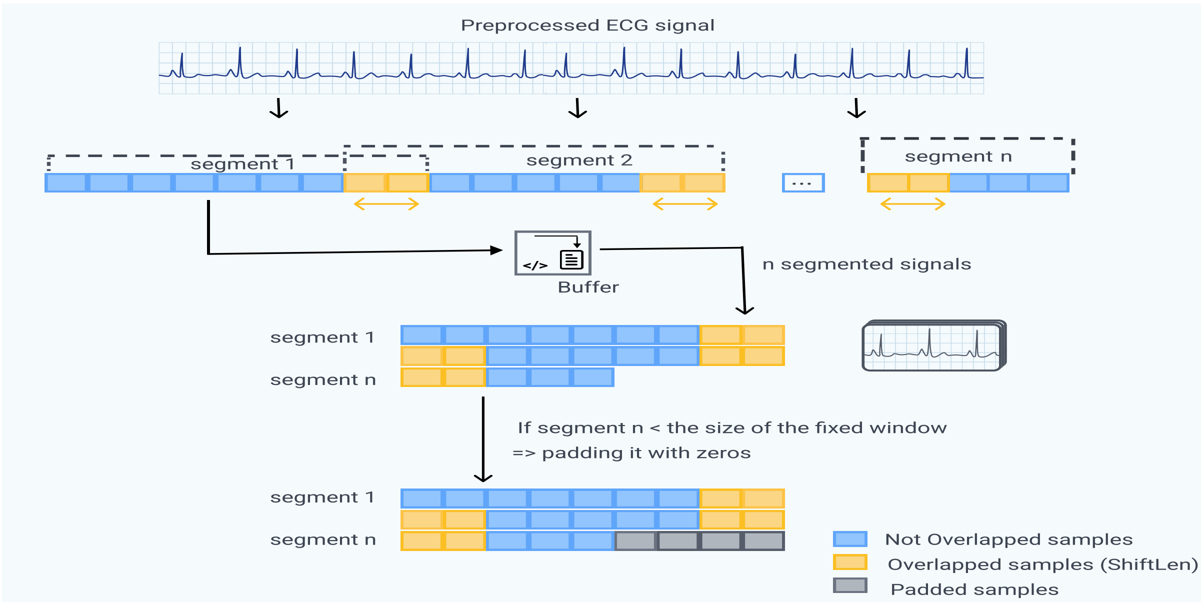

Figure 3. Illustration of ECG signal segmentation.

TABLE II. Optimized 1D-CNN architecture for cardiac arrhythmias classification based on short single lead ECG recordings. Convolution(padding="valid", bais = True, activation function = ReLu) ; MaxPooling1D padding="valid";Dropout rate = 0.5.

Layer	Input nodes	Filter number	Kernel size/pool size	Output nodes	Parameters	Feature interpretation
Input	2700,1	_	_	_	_	ECG amplitude for one segment
Convolution 1	2700,1	32	3×1,stride 1	2698×32	128	32 feature map
Maxpooling 1	2698×32	_	2×1,stride 0	1349×32	_	Feature reduction (1349 nodes for one episode)
Convolution 2	1349×32	32	3×1,stride 1	1347×32	3104	32 feature map
Maxpooling 2	1347×32	_	2×1,stride 0	673×32	_	Feature reduction (673 nodes for one episode)
Convolution 3	673×32	64	3×1,stride 1	671×64	6208	64 feature map
Maxpooling 3	671×64	_	2×1,stride 0	335×64	_	Feature reduction (335 nodes for one episode)
Convolution 4	335×64	64	3×1,stride 1	333×64	12352	64 feature map
Maxpooling 4	333×64	_	2×1,stride 0	166×64	_	Feature reduction (166 nodes for one episode)
Convolution 5	166×64	128	3×1,stride 1	164×128	24704	128 feature map
Maxpooling 5	164×128	_	2×1,stride 0	82×128	_	Feature reduction (82 nodes for one episode)
Convolution 6	82×128	128	3×1,stride 1	80×128	49280	128 feature map
Maxpooling 6	80×128	_	2×1,stride 0	40×128	_	Feature reduction (40 nodes for one episode)
Dropout 1	_	_	_	_	_	complexity reduction (eliminating 50 % of nodes)
Convolution 7	40×128	256	3×1,stride 1	38×256	98560	256 feature map
Maxpooling 7	38×256	_	2×1,stride 0	19×256	_	Feature reduction (19 nodes for one episode)
Convolution 8	19×256	256	3×1,stride 1	17×256	196864	256 feature map
Maxpooling 8	17×256	_	2×1,stride 0	8×256	_	Feature reduction (8 nodes for one episode)
Dropout 2	_	_	_	_	_	complexity reduction (eliminating 50 % of nodes)
Convolution 9	8×256	512	3×1,stride 1	6×512	393728	512 feature map
Maxpooling 9	6×512	_	2×1,stride 0	3×512	_	Feature reduction (3 nodes for one episode)
Dropout 3	_	_	_	_	_	complexity reduction (eliminating 50 % of nodes)
Convolution 10	3×512	512	3×1,stride 1	1×512	786944	512 feature map
Flatten	512	_	_	_	_	Dot product between 1 nodes and 512 feature map
Dense 1	512	_	_	128	65664	Weight params
Dropout 4	_	_	_	_	_	complexity reduction (eliminating 50 % of nodes)
Dense 2	128	_	_	32	4128	Weight params
Output	32	_	_	2 / 4 / 9	132	Class
						Total capacity of features map : 242.592
						Total number of parameters : 1,641,796

5) Repetition of steps 1-4 for each of the three ECG datasets after adjusting parameters like sampling rate to 300Hz and the number of output classes in the convolutional network to match each dataset.

2) Model testing

Model testing involves the following four steps:

1) Segmentation of the continuous ECG recording into N fixed-length segments, each lasting 9 seconds.
2) Classification of each ECG segment using the final selected Afib-CNN model. The predicted classes belong to the set $\{c_{\mathrm{seg1}}, c_{\mathrm{seg2}}, c_{\mathrm{seg3}}, \ldots, c_{\mathrm{segN}}\}$.
3) Final class determination for each recording based on the predicted classes of its constituent segments. The rule is as follows: if the segments contain both normal and abnormal classes, the recording is labeled as abnormal; if multiple types of arrhythmias appear, the most frequent one is selected; and if multiple arrhythmias occur with equal frequency, the one appearing first is chosen.
4) Construction of a confusion matrix using the final selected classes and evaluation of the model's performance through the F1 score metric.

3) Model evaluation

To assess the efficacy of our model in classifying arrhythmias from short, single-lead ECG recordings, we rely on the confusion matrix to summarize the model's predictions. We also benchmark our results against those in the existing literature by utilizing two key metrics: accuracy and mean F1 score.

The accuracy metric calculates the ratio of correctly predicted observations to the total number of observations. On the other hand, the F1 score is a balanced metric that computes the weighted average of precision and recall. Specifically, precision is the ratio of true positive observations to the total number of predicted positives, while recall measures the ratio of true positives to all observations that actually belong to the positive class [29].

The overall performance of our classifier is quantified by averaging the F1 scores for each class within each dataset. For example, for the CinC2017 dataset, which has four classes, the averaged F1 score is computed as shown in Equation 6. Likewise, the process is repeated for the CPSC dataset, which comprises nine classes, and for the MIT-AFIB dataset, which includes two classes.

$$F1_{\mathrm{CinC2017}} = \frac{F1_N + F1_{Afib} + F1_P + F1_O}{4} \tag{6}$$

3. Results and discussion

This section presents our key experimental findings. We begin by discussing the steps taken to mitigate overfitting and to select the optimal deep classifier for detecting Afib and other arrhythmias using the CinC2017 dataset. Next, we delve into the analysis and generalization phase, employing confusion matrices to examine the performance of our proposed method across three independent ECG datasets (CinC2017, CPSC2018, and MIT-AFIB). Finally, we compare our results with existing works in the literature, focusing on the accuracy and F1 score metrics.

A. Process of selecting the optimized Afib-CNN model

1) Reducing model parameters

To minimize the model's complexity, we conducted a grid search based on three key hyperparameters: input length, kernel size, and number of layers.

Selection of Input Length: We evaluated three different input lengths:

- 60 seconds: The maximum record length in the CinC2017 dataset.
- 30 seconds: Clinically recognized as the gold standard for detecting atrial fibrillation (AF) [30].
- 9 seconds: The shortest record length in the CinC2017 dataset that still provides a meaningful context for rhythm classification.

Selection of Kernel Size: We executed two experiments to find the optimal kernel size, as summarized in Tables III and IV. In these experiments, we set the kernel sizes to 3 and 5 across all convolutional layers.

Selection of Number of Layers and Filters: We explored models with varying layers, where the number of filters for each layer configuration was as follows:

- 10-layer model: [32, 32, 64, 64, 128, 128, dropout, 256, dropout, 256, 512, dropout, 512].
- 9-layer model: [32, 32, 64, 64, 128, 128, dropout, 256, dropout, 256, 512, dropout].
- 8-layer model: [32, 32, 64, 64, 128, 128, dropout, 256, dropout, 256].
- 7-layer model: [32, 32, 64, 64, 128, 128, dropout, 256, dropout].

2) Improving model accuracy

In our quest to identify the most effective yet parsimonious model for arrhythmia classification, we scrutinized all lower-complexity models identified in the preceding section. The outcome of this exercise is summarized in Table V.

Overfitting and Input Length: Our findings reveal that utilizing a longer ECG input length (60 seconds) resulted in overfitting. Interestingly, we found a direct correlation between ECG input length and overfitting; the

TABLE III. Grid search for reducing deep model parameters with fixed the kernel size of (3) for all convolutional layers.

Input length	number layers			
	7	8	9	10
60s	4,720,708	**2,623,812**	2,952,004	3,607,876
30s	2,426,948	**1,476,932**	1,772,356	2,428,228
9s	**821,316**	**657,732**	**985,924**	**1,641,796**

TABLE IV. Grid search for reducing deep model parameters with fixed the kernel size of (5) for all convolutional layers.

Input length	number layers			
	7	8	9	10
60s	4,784,260	**2,818,436**	3,343,236	4,392,324
30s	2,490,500	**1,671,556**	2,163,588	3,212,676
9s	**884,868**	**852,356**	**1,377,156**	unworkable

shorter the input, the better the model was at avoiding overfitting.

Trade-Off between Complexity and Performance: We observed that certain models with a lower number of parameters did not offer satisfactory performance. For instance, the model labeled *model9_s 3_k 8_l* demonstrated an accuracy of 96% on both the training and validation sets while having a total of 657,732 parameters.

Optimal Model: The highest performance was attained by the model designated as *model9_s 3_k 10_l*, which had a total of 1,641,796 parameters. This model achieved an impressive accuracy of 99% on the training set and 98% on the validation set.

Test Set Performance: Moreover, the Afib-CNN model (*model9_s 3_k 10_l*) testing on the CinC2017 dataset showcases a remarkable average F1 score of 98%, as detailed in Figure 4. This performance underscores the model's precision in atrial fibrillation detection, achieving a perfect F1 score of 100%. The confusion matrix highlights the model's effectiveness across all categories, including a 99% for normal rhythms, 93% F1 score for other rhythms and a 100% score for noisy signals, demonstrating its overall classification strength.

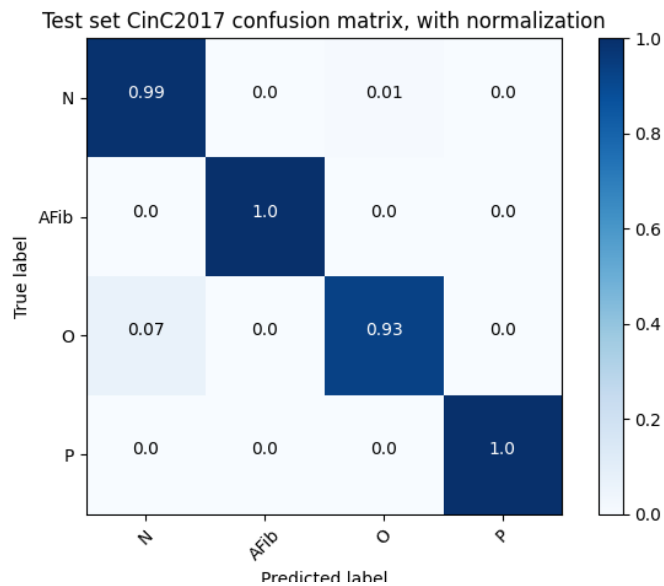

Figure 4. Confusion matrix for the Afib-CNN model predictions on the test set of Cinc2017; $F1_N$ = 99%; $F1_{Afib}$ = 100%; $F1_O$ = 93%; $F1_{Noisy}$ = 100%.

3) Improving generalization and tackling overfitting

In this study, we successfully mitigated the problem of overfitting and enhanced the generalization capabilities of our ECG-based classifier by adopting two novel training strategies: data preprocessing and data segmentation.

As illustrated in Table VI, we evaluated the effectiveness of each strategy separately using the optimal architecture previously identified (*model9_s3_k10_l*). Initially, **Model 0** was trained without any of these improvement strategies.

During the data preprocessing phase, we minimized the wide-ranging variations within and between ECG recordings by employing a Moving Average by Convolution (MAConv) filter and the min-max normalization method. Additionally, we utilized z-score normalization to further differentiate the morphology of each ECG recording. These preprocessing techniques significantly reduced overfitting and enhanced generalization, as evidenced by **Model 1**, which showed a decrease in the overfitting ratio to 7% and 8% for the validation and test sets, respectively.

TABLE V. Grid search for selecting the optimal deep model for arrhythmias classification. Signature Meaning of Model Name (Model: Input length + kernel size + number of convolutional layers).

Model name	Number parameters	Training acc	Validation acc	Optimal epoch
$model60_s3_k8_l$	2,623,812	99	82	46
$model60_s5_k8_l$	2,818,436	99	84	28
$model30_s3_k8_l$	1,476,932	99	96	71
$model30_s5_k8_l$	1,671,556	99	97	31
$model9_s3_k7_l$	821,316	96	92	205
$model9_s3_k8_l$	657,732	96	96	288
$model9_s3_k9_l$	985,924	96	96	297
$model9_s3_k10_l$	**1,641,796**	**99**	**98**	**295**
$model9_s5_k7_l$	884,868	99	97	285
$model9_s5_k8_l$	852,356	98	97	280
$model9_s5_k9_l$	1,377,156	98	97	283

The second strategy involved the application of a shifted-length overlapping technique to balance and increase the size of the input ECG data. **Model 2** , trained with this strategy, exhibited a reduction in the overfitting ratio to 1% for the validation set, although it did not generalize as well to the test set.

Finally, **Model 3**, which combined both training strategies, achieved the best performance with an average accuracy of 98% on both the validation and test sets. This combination effectively eradicated overfitting and significantly enhanced the generalization capabilities of our arrhythmia classifiers.

Figures 5 and 6 illustrate the metrics of the training and validation datasets over epochs for the baseline and final models (**Model 0** and **Model 3**), respectively, highlighting the effectiveness of our techniques for tackling overfitting. These figures visually demonstrate the impact of preprocessing and data augmentation strategies on model performance. Figure 5 shows the performance variations of the baseline model without overfitting prevention strategies, indicating potential areas of overfitting through the gap between training and validation accuracy. Conversely, Figure 6 showcases the improved alignment between training and validation metrics in the final model, emphasizing the effectiveness of the employed techniques in enhancing generalization and reducing overfitting. This direct comparison underscores the necessity and impact of preprocessing and data augmentation strategies for achieving robust model performance.

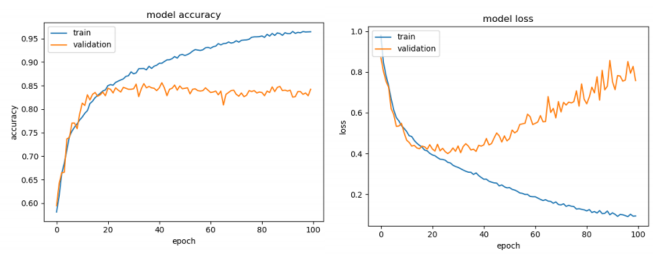

Figure 5. Training and validation dataset metrics as a function of training epoch (cross-entropy loss, accuracy) for the baseline model(without activating overfitting tackling techniques).

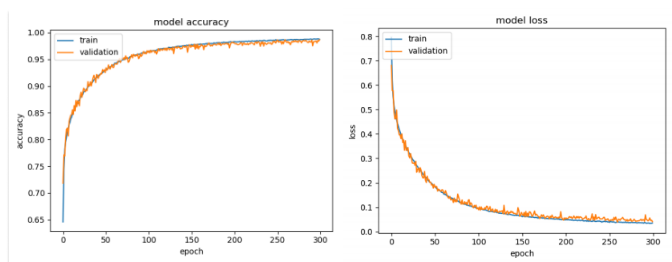

Figure 6. Training and validation dataset metrics as a function of training epoch (cross-entropy loss, accuracy) for the final model(with activating overfitting tackling techniques).

B. *Analysis and generalization of the deep model*

To assess the generalizability of our proposed Afib-CNN model for Afib and other rhythm classifications, we evaluated it on three widely-used datasets. After fixing the architecture of the deep CNN and updating the number of output classes according to each dataset (Cinc2017 has 4 classes, CPSC2018 has 9 classes, and MIT-AFIB includes

TABLE VI. the experiments steps that illustrate the importance of the overfitting tackling technique used in this study, The over-fitting ratio is the difference between the training and the validation accuracies.

Model	Pre-process data	shifted-length overlapping	Training acc	Validation acc	Testing acc
Model 0	✗	✗	98	82	72
Model 1	✓	✗	98	91	90
Model 2	✗	✓	99	98	78
Model 3	✓	✓	99	98	98

2 classes), we trained each of these three models separately without further hyperparameter tuning.

As evident from the confusion matrices in Figure 7, our model demonstrates high performance in classifying various arrhythmias across three distinct ECG datasets. This illustrates the model's robustness and broad applicability to different types of arrhythmias and ECG data distributions. Errors made by the model, as visible in the confusion matrices, are largely explainable. For instance, numerous arrhythmias are confused with normal rhythms, often due to CPSC2018 poses a particular challenge for automatic feature extraction due to the presence of both rhythm- and heartbeat-based arrhythmias. For example, atrial fibrillation (AF) is best analyzed at the rhythm level, while other arrhythmias like LBBB and RBBB are more appropriately classified based on individual heartbeats [31]. This suggests that multiple electrodes may be needed to better differentiate between arrhythmia classes.
Finally, the confusion matrices reveal that the model performs less robustly on minority classes, such as SET from CPSC2018 and P from CinC2017. This underscores the importance of large datasets for improving the performance of deep classification models.

C. Comparison with existing arrhythmia classification methods

Table VII compares the performance of our proposed method with state-of-the-art deep convolutional models across three independent datasets: CinC2017, CPSC2018, and MIT-AFIB. Our model, comprising a total of 1,641,796 parameters and utilizing 9-second single-lead ECG signals, outperforms the other listed methods in terms of classification accuracy.

For instance, the study by Jeong et al. [34] proposed a network with fewer parameters and a shorter input length, yet its classification precision proved to be suboptimal. Similarly, Mousavi et al. [36] introduced an optimized deep architecture with a 5-second input signal length but demonstrated inconsistent performance across the two datasets they validated—achieving 98.17% accuracy for MIT-AFIB and only 72.62% for CinC2017.

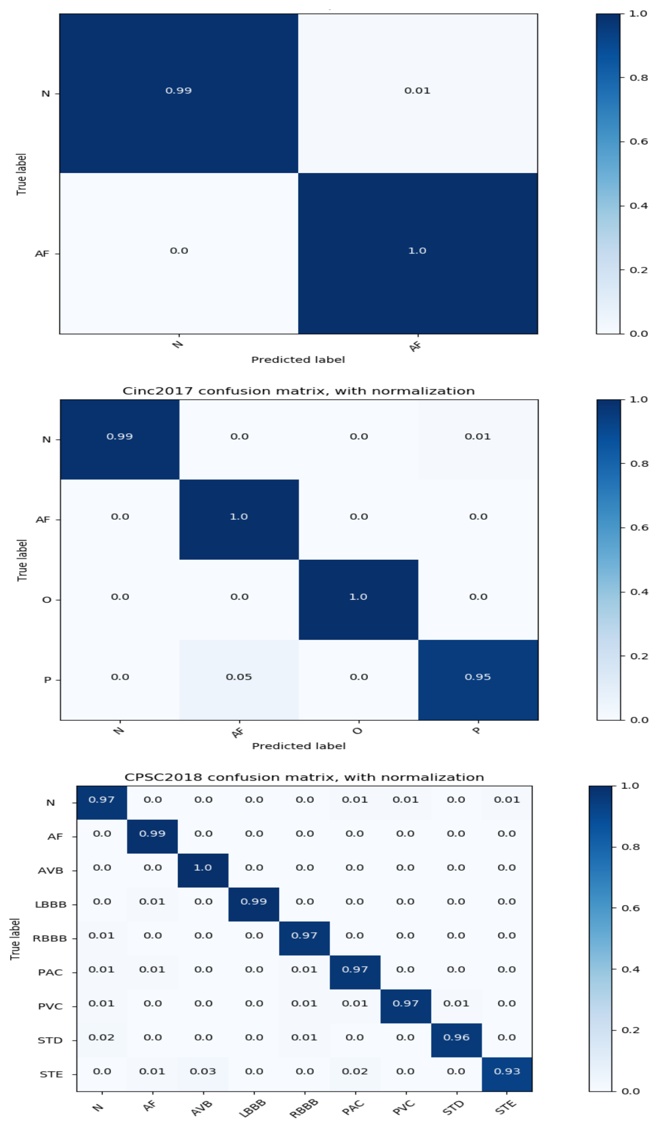

Figure 7. The normalized confusion matrices (CM) of the Afib-CNN classifier classification results in the validation set of three data sets (MIT-AFIB, Cinc2017 and CPSC2018, respectively).

TABLE VII. Comparison of the performances of convolutional neural network models that have been validated on the three independent databases (Cinc 2017, CPSC 2018, MIT-AFIB); TF: Time-frequency; STFT: Short-term fourier transform; SWT: Stationary wavelet transform .

Team	*Dataset*	*InputLength*	*NumberParameters*	*Accuracy*	*Method*
Our proposal.	Cinc2017 CPSC2018 MIT-AFIB	9 seconds	1,641,796	98% 97% 99%	Optimized 1D-CNN
Goodfellow et al[32].	Cinc2017	60 seconds	5,491,392	88%	1D-CNN
Hsieh, et al [10].	Cinc2017	30 seconds	3,212,740	77.8%	1D-CNN
Wang et al [33].	CPSC2018	10 seconds	//	87.3%	Densely 1D-CNN
Jeong, et al [34].	CPSC2018	0.6 seconds	92,857	74%	TF 2D-CNN
Xia et al [35].	MIT-AFIB	5 seconds	53,400 4,864,462	98.29% 98.63%	STFT 2D-CNN SWT 2D-CNN
Mousavi et al [36].	Cinc2017 MIT-AFIB	5 seconds	91,332 173,380	72.62% 98.17%	1D-CNN

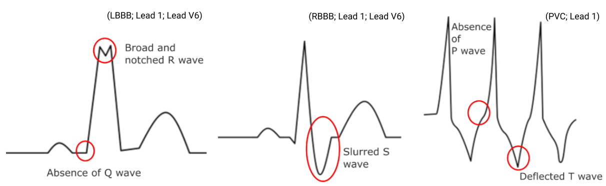

Figure 8. Arrhythmias based heartbeat examples RBBB; LBBB; PVC beats [31].

Additionally, Xia et al. [35] attained high classification accuracy on the MIT-AFIB database with a model comprising only 53,400 parameters. However, they did not evaluate their model's performance on different datasets, limiting its proven generalizability.

In summary, our results indicate that our proposed model offers a balanced combination of high performance and low complexity for classifying various arrhythmias using short, single-lead ECG records. Moreover, the model demonstrates excellent flexibility and generalizability across different data distributions.

4. Conclusions

This study introduced the Afib-CNN model, an optimized 1D-CNN designed for the robust and accurate classification of atrial fibrillation (Afib) and other arrhythmias using short, single-lead ECG records. Validated across three widely recognized datasets—CinC2017, CPSC2018, and MIT-AFIB—our model achieved high classification accuracy, exceeding 97% across all datasets, and demonstrated remarkable generalizability across different ECG data distributions.

The practical errors encountered by our model highlight the inherent challenges in arrhythmia classification, such as the episodic nature of certain arrhythmias and the occurrence of multiple arrhythmia types within single ECG records. Despite these challenges, Afib-CNN's performance compares favorably with state-of-the-art models, offering superior results with notably lower complexity—evidenced by fewer parameters and shorter input signal lengths.

This balance of high performance and computational efficiency makes Afib-CNN an attractive solution for real-world applications, where timely and accurate arrhythmia detection is paramount, yet computational resources may be limited.

Looking forward, our research will pivot towards further optimizing Afib-CNN's architecture and investigating the integration of additional data types to enhance the model's diagnostic robustness. An important direction for future work involves expanding the evaluation to include a broader range of datasets, particularly those representing diverse patient demographics and arrhythmia manifestations. This expansion aims to address the limitation noted in the current study—the evaluation on a limited set of public datasets—and to ensure the model's effectiveness across a wider spectrum of arrhythmias and patient conditions.

In conclusion, the Afib-CNN model marks a significant leap forward in arrhythmia classification technology. Achieving a synergistic balance of accuracy, efficiency, and generalizability, it holds the promise of enhancing healthcare delivery through the improved and expedited diagnosis of atrial fibrillation and other arrhythmias. Future enhancements and broader clinical validations will be instrumental in realizing this potential, contributing to better health outcomes and advancing the field of cardiac care.

Acknowledgement

This work experimented with the IBNKHALDOUN high-performance computing platform provided by the university Mohamed Khider Biskra (http://hpc.univ-biskra.dz/).

Appendix

Algorithm 1: shifted-length overlapping ECG signal algorithm.

Result: Overlapped fixed size ECG recordings.
ECGRecording: The continuous ECG recording.
WindowLen: the length of the fixed size window.
ShiftLen: the length of samples to be overlapped.

```
if ECGRecording length > WindowLen then
    - calculate the number of segments for the current
      recording
    NumberSegments ← (ECGRecording length -
      ShiftLen )/(WindowLen - ShiftLen);
    - segmented the continues ECG Recording into
      WindowLen length with ShiftLen overlaps
      between adjacent segments;
    - if the last segment ¡ WindowLen;
    padding it with zeros;
    - save segments;
else
    if ECGRecording length < WindowLen then
        - Padding the recording with zeros into the
          windowLen length Save recording;
    else
        - Save the recording;
    end
end
```

References

[1] S. Goldsworthy, *Compact clinical guide to arrhythmia and 12-lead ekg interpretation: foundations of practice for critical care nurses*. New York, NY: Springer Publishing Company, LLC, 2016, oCLC: 1030357368.

[2] A. B. M. Abdullah, *ECG in medical practice*. New Delhi: Jaypee Brothers Medical Pub., 2010, oCLC: 957382130.

[3] R. Stroobandt, S. Barold, and A. Sinnaeve, *ECG from Basics to Essentials: Step by Step*. Wiley, 2016.

[4] S. Kwon, E.-K. Choi, S.-R. Lee, S. Oh, H.-S. Song, Y.-S. Lee, S.-J. Han, and H. E. Lim, "Comparison of novel telemonitoring system using the single-lead electrocardiogram patch with conventional telemetry system," vol. 54, no. 3, p. 140. [Online]. Available: https://e-kcj.org/DOIx.php?id=10.4070/kcj.2023.0252

[5] N. Manetas-Stavrakakis, I. M. Sotiropoulou, T. Paraskevas, S. Maneta Stavrakaki, D. Bampatsias, A. Xanthopoulos, N. Papageorgiou, and A. Briasoulis, "Accuracy of artificial intelligence-based technologies for the diagnosis of atrial fibrillation: A systematic review and meta-analysis," vol. 12, no. 20, p. 6576. [Online]. Available: https://www.mdpi.com/2077-0383/12/20/6576

[6] H. Y. Gu, J. Huang, X. Liu, S. Q. Qiao, and X. Cao, "Effectiveness of single-lead ECG devices for detecting atrial fibrillation: An overview of systematic reviews," vol. 21, no. 1, pp. 79–86.

[7] B. Król-Józaga, "Atrial fibrillation detection using convolutional neural networks on 2-dimensional representation of ECG signal," vol. 74, p. 103470. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421010673

[8] S. Nurmaini, A. E. Tondas, A. Darmawahyuni, M. N. Rachmatullah, R. Umi Partan, F. Firdaus, B. Tutuko, F. Pratiwi, A. H. Juliano, and R. Khoirani, "Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks," vol. 113, pp. 304–317. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X20305410

[9] X. Q. Chang, A. F. Chew, B. C. M. Choong, S. Wang, R. Han, W. He, L. Xiaolin, R. C. Panicker, and D. John, "Atrial fibrillation detection using weight-pruned, log-quantised convolutional neural networks," in *2022 IEEE 13th Latin America Symposium on Circuits and System (LASCAS)*, pp. 1–4, ISSN: 2473-4667.

[10] C.-H. Hsieh, Y.-S. Li, B.-J. Hwang, and C.-H. Hsiao, "Detection of Atrial Fibrillation Using 1D Convolutional Neural Network," *Sensors (Basel)*, vol. 20, no. 7, Apr. 2020.

[11] W. Cai, Y. Chen, J. Guo, B. Han, Y. Shi, L. Ji, J. Wang, G. Zhang, and J. Luo, "Accurate detection of atrial fibrillation from 12-lead ECG using deep neural network," *Computers in Biology and Medicine*, vol. 116, p. 103378, Jan. 2020.

[12] M. Wang, S. Rahardja, P. Fränti, and S. Rahardja, "Single-lead ECG recordings modeling for end-to-end recognition of atrial fibrillation with dual-path RNN," vol. 79, p. 104067. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809422005353

[13] J. Kim, S. J. Lee, B. Ko, M. Lee, Y.-S. Lee, and K. H. Lee, "Identification of atrial fibrillation with single-lead mobile ECG during normal sinus rhythm using deep learning," vol. 39, no. 5, p. e56. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10843976/

[14] A. Fatih Gündüz and M. Fatih Talu, "Atrial fibrillation classification and detection from ECG recordings," vol. 82, p. 104531. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809422009855

[15] G. Petmezas, K. Haris, L. Stefanopoulos, V. Kilintzis, A. Tzavelis, J. A. Rogers, A. K. Katsaggelos, and N. Maglaveras, "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets," vol. 63, p. 102194. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809420303323

[16] V. V. S. Tallapragada, D. V. Reddy, K. N. V. S. Varma, and G. S. Sarma, "Improved atrial fibrillation detection using CNN-LSTM," in *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1050–1055.

[17] F. Murat, F. Sadak, O. Yildirim, M. Talo, E. Murat, M. Karabatak, Y. Demir, R.-S. Tan, and U. R. Acharya, "Review of deep learning-based atrial fibrillation detection studies," vol. 18, no. 21, p. 11302, number: 21 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1660-4601/18/21/11302

[18] J. Rahul and L. D. Sharma, "Artificial intelligence-based approach for atrial fibrillation detection using normalised and short-duration

time-frequency ECG," vol. 71, p. 103270. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421008673

[19] X. Zhang, J. Li, Z. Cai, L. Zhang, Z. Chen, and C. Liu, "Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection," vol. 59, no. 1, pp. 165–173. [Online]. Available: https://doi.org/10.1007/s11517-020-02292-9

[20] G. D. Clifford, C. Liu, B. Moody, L.-w. H. Lehman, I. Silva, Q. Li, A. E. Johnson, and R. G. Mark, "AF Classification from a Short Single Lead ECG Recording: the PhysioNet/Computing in Cardiology Challenge 2017," *Computing in cardiology*, vol. 44, Sep. 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5978770/

[21] X. Ying, "An Overview of Overfitting and its Solutions," *Journal of Physics: Conference Series*, vol. 1168, p. 022022, Feb. 2019.

[22] A. L. Goldberger, L. A. N. 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, 2000 (June 13), circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[23] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, and E. N. Yin Kwee, "An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, Sep. 2018. [Online]. Available: http://www.ingentaconnect.com/content/10.1166/jmihi.2018.2442

[24] D. B. Seema Nayak, M. K. Soni, "Signal processing techniques for removing noise from ecg signals." *International Journal of Research in Engineering and Applied Sciences*, 2012. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a5c75c969814f48aff522f19226e70fe2e890673

[25] R. Kher, "Signal processing techniques for removing noise from ecg signals." *journal of Biomedical Engineering and Research*, 2019. [Online]. Available: https://www.semanticscholar.org/paper/Signal-Processing-Techniques-for-Removing-Noise-ECG-Kher/373580265b077d5d71254b505a41536c14231145

[26] S. C. HC Chen, "A moving average based filtering system with its application to real-time qrs detection," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2003.

[27] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing: Chapter 15 - Moving Average Filters*, 1997-1998, accessed: date.

[28] CodeMonk, "Moving average filter - towards signal processing," https://codemonk.in/blog/moving-average-filter/, 2021, accessed: 02-26-2024.

[29] R. Joshi, "Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures," Sep. 2016, section: Blog Homepage. [Online]. Available: https://blog.exsilio.com/?s=f1+score/

[30] L. S. Z. P. Steinberg JS, O'Connell H, "Thirty-second gold standard definition of atrial fibrillation and its relationship with subsequent arrhythmia patterns: Analysis of a large prospective device database." *Circulation: Arrhythmia and Electrophysiology*, vol. 11, no. 7, Jul. 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30002065/

[31] S. L. Oh, "Application of deep learning algorithms for automated detection of arrhythmias with ECG beats," Ph.D. dissertation, Nanyang Technological University, 2019. [Online]. Available: http://hdl.handle.net/10220/47829

[32] S. Goodfellow, A. Goodwin, R. Greer, P. Laussen, M. Mazwi, and D. Eytan, "Towards Understanding ECG Rhythm Classification Using Convolutional Neural Networks and Attention Mappings," in *MLHC*, 2018.

[33] T. X. L. B. Wang C., Yang S., "A 12-lead ecg arrhythmia classification method based on 1d densely connected cnn." *springer*, vol. 11794, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-33327-0_9

[34] L. K. Jeong DU, "Convolutional neural network for classification of eight types of arrhythmia using 2d time-frequency feature map from standard 12-lead electrocardiogram." *Scientific Reports*, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8516863/

[35] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Computers in Biology and Medicine*, vol. 93, pp. 84–92, Feb. 2018.

[36] F. A. F. K. Mousavi, Sajad and U. R. Acharya., "Ecg language processing (elp): A new technique to analyze ecg signals." *Computer Methods and Programs in Biomedicine*, 2021. [Online]. Available: https://doi.org/10.1016/j.cmpb.2021.105959

Khadidja Benchaira is an academic at the LESIA Laboratory, Department of Computer Science, University of Biskra, Algeria. With a Master's in 'Decision Support Systems and Multimedia,' she's now a Ph.D. candidate exploring Machine Learning, Deep Learning, and Biomedical Engineering to enhance healthcare outcomes.

Salim Bitam , a Full Professor at the University of Biskra, Algeria, has an extensive background in computer science with degrees from the University of Constantine and Biskra, and a Doctorate of Sciences from the Higher School of Computer Science - ESI, Algiers. His research spans vehicular networks, cloud computing, and bio-inspired optimization. Bitam has published over 35 works, received two best paper awards, and actively contributes to the academic community as an editorial board member, reviewer, and technical program committee member for prestigious journals and international conferences.

Zineb Djihane Agli holds a Master's in Computer Science from the University of Mohamed Khider Biskra, specializing in biomedical signal processing, machine learning, and bio-inspired optimization. Her research focuses on advancing healthcare technologies.