Focal-based Deep Learning Model for Automatic Arrhythmia Diagnosis

Abir Boulif [1, *], Bouchra Ananou [1], Mustapha Ouladsine [1], Stéphane Delliaux [2]

[1]Aix-Marseille University, University of Toulon, CNRS, LIS, Marseille 13397, France
[2]Aix-Marseille University, INSERM, INRAE, C2VN, Marseille 13005, France
* Corresponding author.

abir.boulif@univ-amu.fr, bouchra.ananou@univ-amu.fr, mustapha.ouladsine@univ-amu.fr, stephane.delliaux@univ-amu.fr

Abstract. This paper approaches a new model for arrhythmia diagnosis based on short-duration electrocardiogram (ECG) heartbeats. To detect 8 arrhythmia classes efficiently, we design a Deep Learning model based on the Focal modulation layer. Moreover, we develop a distance variation of the SMOTE technique to address the problem of data imbalance. The classification algorithm includes a block of Residual Network for feature extraction and an LSTM network with a Focal block for the final class prediction. The approach is based on the analysis of variable-length heartbeats from leads MLII and V5, extracted from 48 records of the MIT-BIH Arrhythmia Database. The methodology's novelty consists of using the Focal layer for ECG classification and data augmentation with DTW distance (Dynamic Time Warping) using the SMOTE technique.

The approach offers real-time classification and is simple since it combines feature extraction, selection, and classification in one stage. Using data augmentation with SMOTE variant and Focal-based Deep learning architecture to identify 8 types of heartbeats, the method achieved an impressive overall accuracy, F1-score, precision, and recall of 98.61%, 94.08%, 94.53%, and 93.68% respectively. Additionally, the classification time per sample was only 0.002 seconds. Therefore, the suggested approach can serve as an additional tool to aid clinicians in ensuring rapid and real-time diagnosis for all patients with no exclusivities.

Keywords: Signal processing, Arrhythmia, Heartbeat, Electrocardiogram, Diagnosis, Classification, Deep Learning, Healthcare sustainability.

1 Introduction

According to the 2023 report of the World Heart Federation (WHF) [1], cardiovascular diseases (CVD) represent a global threat to the population. Deaths due to CVD have increased by 60% worldwide in the past 30 years. Therefore, it is essential to prioritize implementing tools to prevent premature heart attacks and strokes. To this end, arrhythmia detection proves important since it is the cause of most sudden cardiac arrests. Arrhythmia is a medical condition that results either in fast, slow, or irregular heartbeats

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

[2]. Some of the most common types of heartbeats associated with arrhythmia are Premature Atrial and Ventricular contractions. To diagnose these heartbeats, we rely on the analysis of the electrocardiogram (ECG), a non-invasive tool that records the electrical signals in the heart, to investigate symptoms of arrhythmia.

Various techniques in literature have been applied to categorize ECG signals automatically into heart rhythms or heartbeat classes. Heart rhythm describes the overall pattern of electrical activity in the heart, while heartbeat classes refer to specific types of individual heartbeats within a given rhythm. This work focuses on the classification of heartbeats. The classification is preceded by pre-processing, feature extraction, and feature selection. The pre-processing stage may comprise noise removal, data segmentation, data normalization, data reduction, and signal compression.

Noise includes power line interference, muscle noise, motion artifact, baseline wander, and high-frequency artifacts. Discrete Wavelet Transform (DWT) with its various wavelet distributions, is often used for noise removal [3-6]. More improved versions of wavelets were developed in [7, 8]. The Pan-Tompkins algorithm proposed by Pan and Tompkins in 1985 [9] was used for segmentation and QRS detection in [5, 10, 11] and for R-peak detection in [3, 6, 12]. Principal Component Analysis (PCA) is widely used for dimensionality reduction. It is also used for feature extraction [13]. Other methods can also be employed such as Discrete Wavelet Transform [14, 15], Higher Order Statistic (HOS) [5, 10, 16], Independent Component Analysis (ICA) [17], and Fast Fourier Transform (FFT) [14]. These are known to be hand-crafted methods for extraction. Deep learning with CNN layers, is also used for end-to-end extraction as in [18-20].

When dealing with imbalanced data, augmentation techniques can be used to address this issue. SMOTE [21, 22] and GANs [23] are effective methods in reducing overfitting during training, while other techniques only increase data volume by adding noise, without measurable improvement in dataset performance and variance [24].

To detect cardiac rhythms, several approaches have been employed, ranging from traditional machine learning algorithms to complex architectures. Support Vector Machines (SVM) achieved an accuracy of 98.91% [5], 94.30% [10], and 98.8% when combined with Genetic Algorithm [8]. Feed-Forward Neural Network reached respective accuracy of 98.90%, 94.52%, and 99.80% in [5, 10, 25]. Long-Short-Term-Memory (LSTM) finds numerous applications in time series, including the classification of ECG signals [4, 18]. Moreover, Convolutional Neural Networks (CNN) hold a significant position in deep learning and provide accurate results when detecting arrhythmias [3, 23, 24]. When combining CNN and LSTM, both temporal and spatial information are captured [21, 23]. More complex architectures were developed to address the problem of vanishing gradients [26], by combining a Residual Network and LSTMs to detect five heartbeats and achieve 99.4% accuracy. [27] yielded an accuracy rate of 97.9% by leveraging the proficiency of three networks - CNN, LSTM, and bi-directional GRU.

The authors' prior work focused on using machine learning to diagnose Atrial Fibrillation (AF). In [28], a multi-dynamics analysis of the QRS complex using SVM and MKL models yielded sensitivity of 96.54% and 95.47%, respectively. Other previous research consisted of extracting features from R-wave derivatives to aid in medical decision-making, particularly for detecting AF [29-31].

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

This work focuses on classifying eight types of heartbeats. These include Normal beats (N), Atrial Premature beats (A), which can lead to sustained arrhythmias if frequent or persistent, and Premature Ventricular beats (V), which are common and may be associated with serious ventricular arrhythmias if left untreated. Left Bundle Branch Block beats (L), which may indicate an increased risk of cardiovascular events. Right Bundle Branch Block beats (R) are considered serious as they are often associated with structural heart diseases. Paced beats (p) are generated by a pacemaker, to help the heart muscle contract when the natural heart rate is too slow or when there is a heart block. Fusion of Ventricular and Normal beats (F), occur when the electrical signals of premature ventricular beat and normal beat coincide in time. Fusion of Paced and Normal beats (f), reflect the combination of the artificially paced beat and the natural beat initiated by the sinus node.

These heartbeats were chosen because they may cause major health problems if not addressed, and their morphology may be difficult to distinguish from other heartbeats.

A novel method based on Deep Learning classification has been developed to identify eight types of heartbeats. The short-duration signal is typically used when dealing with heartbeat diagnosis, and its use is advantageous as the algorithm focuses on dynamic sequences and feature extraction within a narrowed time frame. The use of a Residual network, which is a 12-layer CNN, for feature extraction and LSTM-Focal block for prediction can be generally employed to classify other time-series data. Therefore, this approach could be widely applied. The aspects of our study encompass:

- The processing of variable-length ECG heartbeats.
- The use of MLII (modified limb lead II) and V5 leads from the ECG.
- Data augmentation using new **DTW-SMOTE** variation.
- Feature extraction using a Residual Network.
- Heartbeat classification model using a **Focal Modulation layer.**

To our knowledge, no other papers in the literature used focal layers for ECG diagnosis.

2 Material and methods

In this paper, a Deep Learning model based on Residual Network and LSTM-Focal architecture is designed to classify eight types of heartbeats. It is an end-to-end model, excluding any hand-crafted methods for feature extraction or selection. The Residual block, inspired by the Residual Network [32], which comprises 12 convolutional layers, is used for deep feature extraction. The LSTM and focal layers are mainly used for classification. The network inputs variable-length ECG heartbeats, ranging from 81 to 439 samples, and returns the heartbeat class. In terms of preprocessing, we applied noise removal, normalization, heartbeat segmentation, and data augmentation.

2.1 Assumptions

Our approach is based on the following assumptions:

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

- Heartbeat segmentation is realized based on R-peaks annotations.
- The ECG signal was de-noised, to ensure the exclusion of any significant interferences that could affect the classification accuracy.
- The majority of cardiac heartbeats, possess distinct patterns within ECG data.
- Each ECG sample contains one class heartbeat.
- The use of an End-to-end structure that combines feature extraction, selection, and heartbeat classification in one stage.

2.2 ECG database

An open-access database, hosted on Physionet [33], is used in this paper since it is regularly utilized for arrhythmia research and contains annotated heartbeats. The MIT-BIH Arrhythmia database [34] is a collection of 48 half-hour excerpts of two-channel ambulatory ECG recordings that were acquired from 47 subjects. The recordings were digitized at a rate of 360 samples per second per channel, with an 11-bit resolution and covering a range of 10 millivolts. The database contains seventeen types of heartbeats.

Eight types of heartbeats are extracted from leads MLII and V5. The signal is first cleaned then normalized and segmented. Finally, it is passed through a data augmentation process to ensure the dataset is no longer imbalanced.

The test data is extracted before data augmentation, forming 30% of the original dataset, which contained 111,471 heartbeats in total, to assess efficiently the performance of the classification model with imbalanced unseen data.

We obtain a dataset of 475,248 heartbeats in total after augmentation. The dataset is then stratified equally between classes and split into the training set (80%) which contains 380,198 heartbeats and the validation set (20%) which contains 95,050 heartbeats.

2.3 Methods: Preprocessing

Noise reduction.

The main goal of noise reduction is to eliminate or minimize the random variations in the signal. This technique aims to enhance the clarity of the underlying information whilst improving the signal-to-noise ratio. The ECG signal contains different types of noise, each should be removed with a special filter.

Baseline wander is a low-frequency artifact that arises from charged electrodes or patient movement and breathing. A high-pass Butterworth five-order filter [35] with a cutoff frequency of 0.5 Hz is used to remove baseline wander. The Gain of the n-order filter is given by (1), where ω is the filter frequency and ω_c is the cutoff frequency.

$$G_n(\omega) = \frac{1}{\sqrt{1+\left(\frac{\omega}{\omega_c}\right)^{2n}}} \quad (1)$$

Power-line interference is a high-frequency artifact caused by improper grounding of the ECG equipment. We smooth the signal with a moving average kernel, which is a Finite Impulse Response Filter [36] privileged when dealing with time series, with a width of one period of 50 Hz. The output of the FIR filter of n-order is given by (2):

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

$$y(k) = \sum_{i=0}^{n} b_i * x(k-i) \quad (2)$$

$x(k)$ is the input signal and b_i is the value of the impulse response at the i^{th} instant.

Data Normalization.

Applying Min-Max normalization results in smaller standard deviations, which helps to eliminate the impact of outliers. In addition, rescaling improves the backpropagation process during Deep Learning by speeding up the convergence rate. Fig. 1 highlights the effect of data normalization and noise removal on the ECG signal.

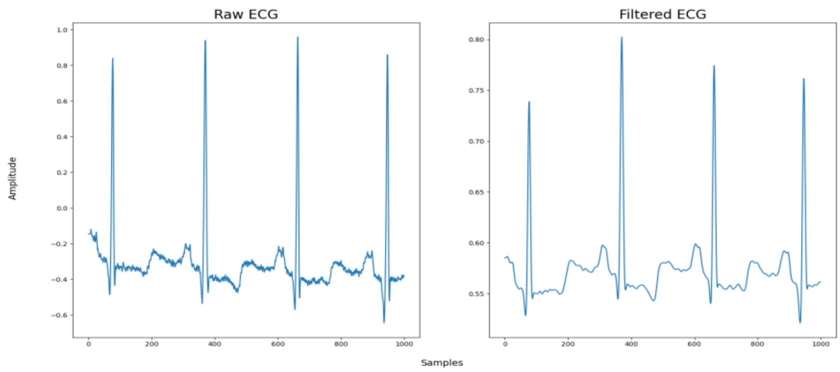

Fig. 1. Noisy Vs de-noised and normalized ECG signal.

Heartbeat segmentation.

The Discrete Wavelet Transform (DWT) technique is used for dynamic heartbeat segmentation by detecting P-wave onsets and T-wave offsets based on the annotations of R-peaks provided by the MIT-BIH Arrhythmia database.

DWT enables signal decomposition by passing it through a series of low-pass and high-pass filters to extract the required information. It facilitates the capture of both spatial and temporal information in a signal. Fig. 2 shows the delineation of cardiac heartbeat where the morphological and temporal characteristics can be depicted.

The segmented heartbeats vary in size, but to meet the requirements of the Deep Learning model, we apply zero padding to ensure uniform input size.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

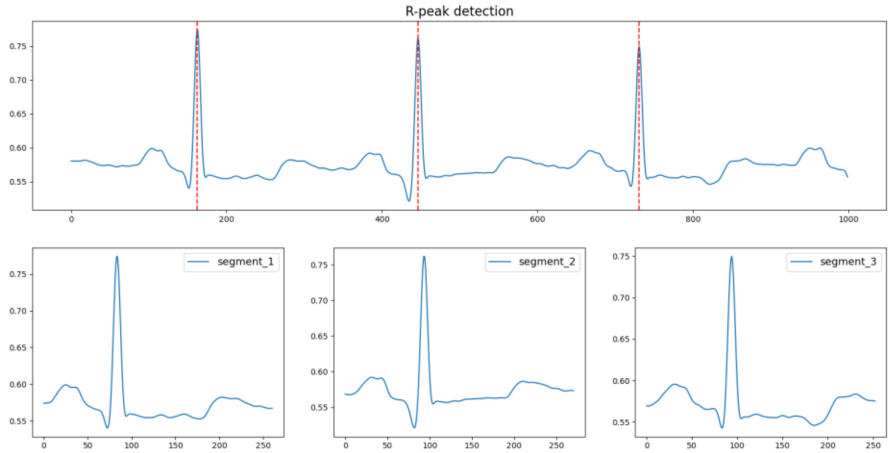

Fig. 2. R-peaks annotations from Physionet (Top), heartbeat segmentation with DWT (Bottom).

Data augmentation.

In machine learning, models trained on unbalanced datasets may have biased behavior in favor of the majority class. As a result, the model may perform poorly on the minority class, leading to limited generalization ability to unseen data.

To address this issue, SMOTE [37] is widely used for dealing with class imbalance, by oversampling the minority class. The technique generates synthetic examples by interpolating between existing neighbor instances found with the KNN algorithm. The classic version is based on the Euclidean distance to detect the KNN neighbors.

This paper presents a new variant of SMOTE that uses the DTW (Dynamic Time Warping) [38] similarity distance to detect the K nearest neighbors. The aim is to find the best mapping with minimal Euclidean distance matching. As the data consists of variable-size heartbeats, their distance cannot be measured with the classic Euclidean formula. However, DTW can handle series that have different lengths and may be warped in the time domain, to find the optimal alignment where the patterns like troughs and peaks can be correctly matched due to the one-to-many mapping.

The DTW algorithm computes the optimal alignment path dynamically by considering various alignments and selecting the one that minimizes the cumulative distance.

Let's consider two Times Series of the Normal heartbeat class, $X = (x_1, \dots, x_N)$ and $Y = (y_1, \dots, y_M)$ of lengths $N \in \mathbb{N}$ and $M \in \mathbb{N}$ respectively. The DTW distance is expressed by (3) where D is an N by M matrix defining the accumulated distance between x_i and y_j. It is computed dynamically as shown in (4).

$$DTW(X,Y) = \sqrt{D[x_N, y_M]} \qquad (3)$$

$$D[x_i, y_j] = d(x_i, y_j) + \min\left(D[x_{i-1}, y_{j-1}], D[x_{i-1}, y_j], D[x_i, y_{j-1}]\right) \qquad (4)$$

$d(x_i, y_j)$ is the Euclidean distance between elements x_i and y_j.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

The optimal alignment path is then found by backtracking from $D[x_N, y_M]$ to $D[x_1, y_1]$, while following the minimum cumulative distance.
We apply DTW-SMOTE by resampling all the classes but the majority class. Table 1 shows the distribution of heartbeats after applying the DTW-SMOTE variant.

Table 1. Original data Vs DTW-SMOTE data

Type of heartbeats	Original distribution	Smote Distribution
Normal beat (N)	59,406	59,406
Right bundle branch block beat (R)	5032	59,406
Left bundle branch block beat (L)	4338	59,406
Paced beat (p)	4027	59,406
Premature Ventricular beat (V)	2657	59,406
Atrial Premature beat (A)	1700	59,406
Fusion of paced and normal beats (f)	451	59,406
Fusion of ventricular and normal beats (F)	418	59,406

To evaluate the performance of the DTW-SMOTE variant, we applied two Machine-Learning techniques for the classification of the augmented data using both classic SMOTE and DTW-SMOTE. The Decision Tree model applied to the SMOTE variant reached a better classification performance with an overall accuracy of 94.40% while the SVM model achieved almost similar results on both classic SMOTE data and DTW-SMOTE data. The evaluation metrics (in %) are detailed in Table 2.

Table 2. Evaluation metrics of Decision Tree and SVM on SMOTE and DTW-SMOTE data.

	Decision Tree		SVM	
	SMOTE	DTW-SMOTE	SMOTE	DTW-SMOTE
F1-score	79.94	81.78	70.64	70.59
Precision	75.47	78.40	63.80	63.75
Recall	86.20	85.86	87.77	86.83
Accuracy	93.95	94.40	88.44	88.53

2.4 Methods: LSTM-Focal classification model

A deep neural network is designed to classify eight types of heartbeats. It consists of a Residual Network [32] for feature extraction and LSTM-Focal for classification.

Feature extraction with Residual Network (ResNet).
The usage of residual blocks, which include skip connections, allows the network to directly learn residuals (the difference between input and output), making the model easier to train without encountering the vanishing gradient problem. The Residual Network used in this approach is a 12-layer convolutional neural network.

Convolution layers with different kernel sizes are performed on cardiac heartbeats to obtain feature maps, each one selecting various features from the ECG signal, i.e.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

peaks, troughs, waves, and local patterns... An adaptive average pooling layer is applied at the end of the model, to condense all of the feature maps into a single one, capturing all relevant information. We obtain as an output a 1-D vector of 512 features.

Heartbeat classification with LSTM-Focal based model.

The last block of the heartbeat classification model is the prediction block. The architecture of the global model is shown in Fig.3.

The prediction block takes as an input the ResNet features and outputs the ECG heartbeat class. The architecture consists of two LSTM layers [39] that are very adept at capturing patterns and variations in ECG sequences with temporal dependencies.

The LSTM layers are then followed by a focal modulation layer which is a part of the FocalNets [40], designed by Microsoft in 2022, where the self-attention mechanism [41] is completely replaced by a focal modulation module for modeling token interactions in vision. This module allows better generalization ability through dynamic focal layers instead of static convolution kernels.

In the present approach, we implement only the focal modulation layer given that it represents the core of the FocalNet. This layer allows the model to selectively focus on specific parts of its input with more lightweight operations. The process is based on aggregation followed by interaction between the aggregated parts of the input. Unlike self-attention which gives priority to the interaction over the aggregation.

First, the output of the LSTM layers H is projected linearly and then split into query q, context Y^0, and gates G^l. These three components are passed through three major operations of the focal modulation layer:

- Hierarchical Contextualization: In this stage, the initial context $Y^0 \in \mathbb{R}^{H*W*C}$ is passed through a series of Depth-Wise Convolution (DWConv) and GeLU layers. These blocks are termed focal levels l and the output of each level is a context Y^l. The final feature map Y^{L+1} goes through a Global Average Pooling layer. The corresponding equations of the aforementioned steps are defined in (5) to (7).

$$q, Y^0, G = f(H) \quad (5)$$

$$Y^l = g^l(Y^{l-1}) \in \mathbb{R}^{H*W*C}, l \in \{1, \dots, L\} \quad (6)$$

$$Y^{L+1} = AvgPooling(Y^l) \in \mathbb{R}^C \quad (7)$$

  f is the linear projection layer and g^l is the focal layer where l denotes the focal level. (H, W, C) is the output size corresponding to Height, Weight, and Channels. l denotes the focal level where $l \in \{1, \dots, L\}$.
- Gated Aggregation: Gates $G^l \in \mathbb{R}^{H*W*(L+1)}$ are employed to perform a weighted aggregation over the context Y^l. Next, the output Z is passed through a convolutional layer h to obtain the modulator M. The equations are shown in (8) and (9):

$$Z = \sum_{l=1}^{L+1} G^l \odot Y^l \in \mathbb{R}^{H*W*C} \quad (8)$$

$$M = h(Z) \in \mathbb{R}^{H*W*C} \quad (9)$$

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

- Interaction: where the initial query q interacts with the context aggregation. Another post- linear projection function is applied with Dropout to provide the final focal output according to (10).

$$F^{out} = linear(q \odot M) \tag{10}$$

The output of the focal modulation is then normalized with the batch-normalization technique, which helps accelerating the learning process. Finally, a Fully-connected layer is added to the model to output the eight heartbeat classes.

Furthermore, this paper includes a comparison experiment using the LSTM-attention model, which employs an attention mechanism instead of the Focal module. This comparison serves to highlight the impact of the Focal layer on the classification result.

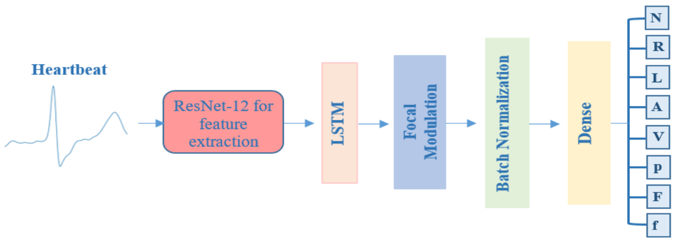

Fig. 3. The prediction block based on LSTM-Focal layers.

Environment setup.

For the hardware characteristics, all the experiments were run on a cluster equipped with GPU node. The model of the GPU graphic card is NVIDIA GA102 [GeForce RTX 3090] with 192 G of RAM.

3 Results of classification

For heartbeat recognition, a heartbeat fragment of 439 sample-long is fed to the input layer of the Residual Network in batches. Each time series accumulates 512 features. The output is then processed in the LSTM-Focal block to produce the heartbeat class.

Fig. 4 shows the training and validation accuracy. The LSTM-Focal model is trained during 50 epochs, during which the learning is stabilized and reaches a maximum training accuracy of 99.94% and validation accuracy of 99.87%.

The model is then tested on unbalanced heartbeats. It succeeded in recognizing correctly 32,977 fragments out of 33,442, yielding an overall accuracy of 98.61%. The evaluation criteria for the test data are shown in Table 3. As can be observed, the model reached high F1-score performances for classes N, L, R, and p since they form the majority of the test data. The model yielded an overall precision, recall, and F1 score of 94.53%, 93.68%, and 94.08% respectively. The lowest recognition performance was observed for the F class which contained only 179 fragments during the test.

The LSTM-attention model is trained to assess the effectiveness of the Focal layer on the classification. Fig. 4 shows that the model takes 15 epochs to stabilize, it reaches

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

an overall training and validation accuracy of 99.31% and 99.5% respectively. According to Table 4, we conclude that the LSTM-Focal model reaches a better classification performance with an overall accuracy of 98.61% Vs 97.65% for the LSTM-Attention model. The use of the Focal layer enhanced the ability of the model to correctly identify the positive heartbeats during the classification with an overall precision of 94.53% Vs 89.94% for the model with the attention mechanism.

The time required for testing one heartbeat is measured by averaging the prediction time over all samples. It is shown in equation (11) where e represents the execution time at step t.

$$T = \sum_{t=0}^{n} \frac{e_t}{n} \tag{11}$$

We measure the variation of the classification time by calculating the mean and standard deviation over all iterations. As a result, in each time, we find a constant average classification time of 0.002 seconds with zero deviation for a single heartbeat.

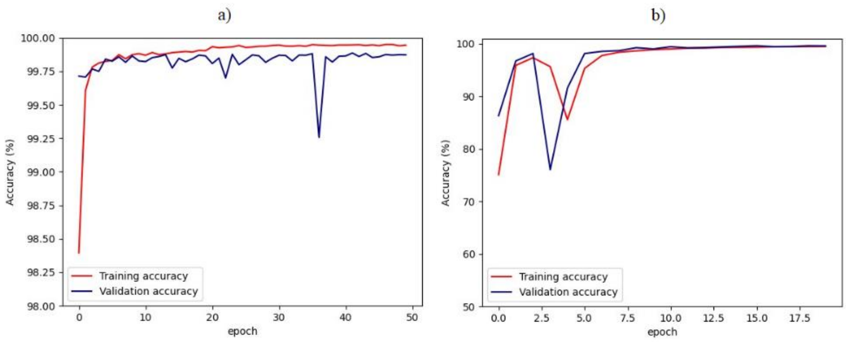

Fig. 4. Training and validation accuracy of LSTM-Focal (a) and LSTM-attention (b) models.

Table 3. Evaluation metrics (in %) of the LSTM-Focal model using test data.

Heartbeats	**Precision**	**Recall**	**F1-score**
Atrial Premature beats (A)	91.01	91.63	91.32
Fusion of Ventricular and Normal beats (F)	85.09	76.54	80.59
Left bundle branch block beats (L)	96.85	97.47	97.16
Normal beats (N)	99.21	99.23	99.22
Right bundle branch block beats (R)	98.56	98.56	98.56
Premature Ventricular beats (V)	95.47	94.29	94.88
Fusion of Paced and Normal beats (f)	90.86	92.27	91.56
Paced beats (p)	99.19	99.48	99.33
Macro average	94.53	93.68	94.08
Weighted average	98.60	98.61	98.60
Overall accuracy	98.61		

Table 4. Overall evaluation metrics (in %) of the LSTM-Focal and the LSTM-attention models.

Model	**Precision**	**Recall**	**F1-score**	**Accuracy**

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

LSTM-Focal	94.53	93.68	94.08	98.61
LSTM-attention	89.94	92.49	91.17	97.65

4 Discussion

In order to confirm the effectiveness of the proposed model, we compare it with some ECG diagnosis approaches in literature, that have used the MIT-BIH Arrhythmia database. The comparison criteria are listed in Table 5. They include the number of heartbeat classes, the feature set, the classification algorithm, and the overall accuracy.

Our model, indicates one of the best performances recorded in arrhythmia diagnosis. It outperformed certain models that used more complex architectures such as U-Net [44] and Google-Net [45], with an accuracy of 96.30% and 97.32% respectively Vs 98.61% for our model. Other models that outperformed the current approach either used feature extraction before classification, as in [21], or used only one ECG lead.

The majority of state-of-the-art methods work on the classification of 5 types of heartbeats: N, V, L, R, and A. These are the most common in open-access databases. When compared to [3, 43, 46], the proposed method achieves a higher accuracy, particularly when using the same type of input (raw data). In comparison to techniques working with 17 classes [11, 43], the proposed method outperformed these models with an accuracy of 98.61% Vs 89.95% and 91.33% respectively. This is due to the fact that training the model with a large number of imbalanced classes reduces performance and causes the model to fail to correctly distinguish minority classes.

Among the Machine Learning methods shown in Table 5, it can be said that the SVM are efficient for heartbeat classification and can outperform DL models when combined with appropriate feature extraction methods such as Wavelet Transform and PCA.

Table 5. Comparison of the proposed methods to Literature methods.

Paper	N° of beats	Feature set	Classifier	Accuracy
Martis et al. [10]	5	HOS+PCA	Feed Forward NN	94.52
Park et al. [11]	17	RR interval, R and P waves' positions and amplitudes	Decision tree	89.95
Li et al. [8]	5	DWT, PCA, LDA, KICA	SVM and Genetic algorithm	98.80
Acharya et al. [3]	5	Raw data	CNN	94.03
Qin et al. [42]	6	Wavelet multi-resolution and PCA	One-Versus-One SVM	99.70
Yang et al. [19]	5	PCA-Net	SVM	97.77
Yildirim et al. [43]	17	Raw data	CNN	91.33
Oh et al. [44]	5	Raw data	Modified U-network	97.32
Yildirim et al. [18]	5	Raw data	LSTM	99.23
Kim et al. [45]	5	Raw data	GoogleNet with 2 inceptions	96.30
Zubair et al. [46]	5	Raw data	CNN	96.36

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

Irfan et al. [21]	5	PCA	CNN + LSTM	99.35
The authors' approach	8	Raw data	LSTM-Focal modulation	98.61

The present approach using the LSTM and Focal Modulation for classification seems to be efficient for most heartbeats due to the focus of the Focal layer on specific ECG features. Yet, we noticed that F heartbeat, which is the fusion of Normal and Ventricular beats, was misclassified into categories N or V. Also, p heartbeat was mainly misclassified into paced and Normal beats. Therefore, the classification of types f and F heartbeats led to the lowest accuracy. To encounter this problem, another feature should be extracted, to distinguish between the fused heartbeats.

5 Conclusion

In this paper, we propose a SMOTE variant based on DTW distance for data augmentation and a new Focal-based model for ECG heartbeat diagnosis. The DTW-SMOTE allows addressing the issue of data imbalance since the DTW similarity measure finds with more precision the neighbors to select for data generation. A Residual Network consisting of convolutions is employed for feature extraction. It is based on skip connections that avoid the loss of information derived from earlier layers, and take into account the spatial dimension of the ECG data. The model uses LSTM layers to capture any temporal dependencies and to keep in memory the long-term context information. Afterward, a Focal Modulation layer is introduced for more feature enhancement. Due to the use of dynamic kernels, the mechanism can effectively focus on features differently and help improve the heartbeat classification. The model reached an accuracy of 98.61% in detecting eight types of heartbeats, on the MIT-BIH Arrhythmia database.

To sum up, the essential elements of our work include:

- The use of variable-length ECG heartbeats, extracted from 2 leads (MLII and V5),
- Data augmentation using new **DTW-SMOTE** variation,
- End-to-end structure, including deep feature extraction and **Focal classification**.

As for the sustainable aspect of this study, the automatic diagnosis model can help with the early detection of heart disorders. This can alleviate the overall burden on healthcare systems.

Our model can be extended to recognize other types of heartbeats by using other databases. Additionally, the Focal layer can be applied for the classification of rhythm categories instead of heartbeat classes.

Disclosure of interests.
The authors have no competing interests to declare that are relevant to the content of this article.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

References

1. World heart report 2023. World Heart Federation. (2023). https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf
2. What is an arrhythmia?. National Heart Lung and Blood Institute. (Last updated 2022). https://www.nhlbi.nih.gov/health/arrhythmias
3. Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., & Tan, R. S. (2017). A deep convolutional neural network model to classify heartbeats. Computers in Biology and Medicine, 89, 389–396. https://doi.org/10.1016/j.compbiomed.2017.08.022
4. Gao, J., Zhang, H., Lu, P., & Wang, Z. (2019). An Effective LSTM Recurrent Network to Detect Arrhythmia on Imbalanced ECG Dataset. Journal of Healthcare Engineering, 2019, 1–10. https://doi.org/10.1155/2019/6320651
5. Elhaj, F. A., Salim, N., Harris, A. R., Swee, T. T., & Ahmed, T. (2016). Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. Computer Methods and Programs in Biomedicine, 127, 52–63. https://doi.org/10.1016/j.cmpb.2015.12.024
6. Rajagopal, R., & Ranganathan, V. (2018). Design of a hybrid model for cardiac arrhythmia classification based on Daubechies wavelet transform. Advances in Clinical and Experimental Medicine, 27(6), 727–734. https://doi.org/10.17219/acem/68982
7. Yildirim, Ö. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology and Medicine, 96, 189–202. https://doi.org/10.1016/j.compbiomed.2018.03.016
8. Li, H., Yuan, D., Wang, Y., Cui, D., & Cao, L. (2016). Arrhythmia Classification Based on Multi-Domain Feature Extraction for an ECG Recognition System. Sensors, 16(10), 1744. https://doi.org/10.3390/s16101744
9. Pan, J., & Tompkins, W. J. (1985). A Real-Time QRS Detection Algorithm. IEEE Transactions on Biomedical Engineering, BME-32(3), 230–236. https://doi.org/10.1109/TBME.1985.325532
10. Martis, R. J., Acharya, U. R., Lim, C. M., Mandana, K. M., Ray, A. K., & Chakraborty, C. (2013). Application of higher order cumulant features for cardiac health diagnosis using ECG signals. International Journal of Neural Systems, 23(04), 1350014. https://doi.org/10.1142/s0129065713500147
11. Park, J., & Kang, K. (2014). PcHD: Personalized classification of heartbeat types using a decision tree. Computers in Biology and Medicine, 54, 79–88. https://doi.org/10.1016/j.compbiomed.2014.08.013
12. Raj, S., & Ray, K. C. (2018). Automated recognition of cardiac arrhythmias using sparse decomposition over composite dictionary. Computer Methods and Programs in Biomedicine, 165, 175–186. https://doi.org/10.1016/j.cmpb.2018.08.008
13. Kim, J., Shin, HS., Shin, K., et al. (2009). Robust algorithm for arrhythmia classification in ECG using Extreme Learning Machine. BioMedical Engineering OnLine, 8(1):31. https://doi.org/10.1186/1475-925x-8-31
14. Chen, G., Hong, Z., Guo, Y., & Pang, C. (2019). A cascaded classifier for multi-lead ECG based on feature fusion. Computer Methods and Programs in Biomedicine, 178, 135–143. https://doi.org/10.1016/j.cmpb.2019.06.021
15. Sumathi, S., Beaulah, H. L., & Vanithamani, R. (2014). A Wavelet Transform Based Feature Extraction and Classification of Cardiac Disorder. Journal of Medical Systems, 38(9), 98. https://doi.org/10.1007/s10916-014-0098-x

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

16. Afkhami, R.G., Azarnia, G., & Tinati, M.A. (2016). Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. Pattern Recognition Letters, 70,45–51. https://doi.org/10.1016/j.patrec.2015.11.018
17. Yu, S.N., Chou, K.T. (2019). Selection of significant independent components for ECG beat classification. Expert Systems with Applications, 36(2),2088–96. https://doi.org/10.1016/j.eswa.2007.12.016
18. Yildirim, O., Baloglu, U. B., Tan, R.-S., Ciaccio, E. J., & Acharya, U. R. (2019). A new approach for arrhythmia classification using deep coded features and LSTM networks. Computer Methods and Programs in Biomedicine, 176, 121–133. https://doi.org/10.1016/j.cmpb.2019.05.004
19. Yang, W., Si, Y., Wang, D., & Guo, B. (2018). Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. Computers in Biology and Medicine, 101, 22–32. https://doi.org/10.1016/j.compbiomed.2018.08.003
20. Wang, R., Fan, J., & Li, Y. (2020). Deep Multi-Scale Fusion Neural Network for Multi-Class Arrhythmia Detection. IEEE Journal of Biomedical and Health Informatics, 24(9), 2461–2472. https://doi.org/10.1109/JBHI.2020.2981526
21. Irfan, S., Anjum, N., Althobaiti, T., Alotaibi, A. A., Siddiqui, A. B., & Ramzan, N. (2022). Heartbeat Classification and Arrhythmia Detection Using a Multi-Model Deep-Learning Technique. Sensors, 22(15), 5606. https://doi.org/10.3390/s22155606
22. Luo, X., Yang, L., Cai, H., Tang, R., Chen, Y., & Li, W. (2021). Multi-classification of arrhythmias using a HCRNet on imbalanced ECG datasets. Computer Methods and Programs in Biomedicine, 208, 106258. https://doi.org/10.1016/j.cmpb.2021.106258
23. Ullah, W., Siddique, I., Zulqarnain, R. M., Alam, M. M., Ahmad, I., & Raza, U. A. (2021). Classification of Arrhythmia in Heartbeat Detection Using Deep Learning. Computational Intelligence and Neuroscience, 2021, 1–13. https://doi.org/10.1155/2021/2195922
24. Iftene, A., Burlacu, A., & Gifu, D. (2022). Atrial Fibrillation Detection Based on Deep Learning Models. Procedia Computer Science, 207, 3752–3760. https://doi.org/10.1016/j.procs.2022.09.436
25. Anwar, S. M., Gul, M., Majid, M., & Alnowami, M. (2018). Arrhythmia Classification of ECG Signals Using Hybrid Features. Computational and Mathematical Methods in Medicine, 2018, 1–8. https://doi.org/10.1155/2018/1380348
26. Ma, S., Cui, J., Xiao, W., & Liu, L. (2022). Deep Learning-Based Data Augmentation and Model Fusion for Automatic Arrhythmia Identification and Classification Algorithms. Computational Intelligence and Neuroscience, 2022, 1–17. https://doi.org/10.1155/2022/1577778
27. Wang, J. (2021). Automated detection of premature ventricular contraction based on the improved gated recurrent unit network. Computer Methods and Programs in Biomedicine, 208, 106284. https://doi.org/10.1016/j.cmpb.2021.106284
28. Trardi, Y., Ananou, B., Haddi, Z., & Ouladsine, M. (2018). Multi-dynamics analysis of QRS complex for atrial fibrillation diagnosis. 5th International Conference on Control, Decision and Information Technologies (CoDIT). https://doi.org/10.1109/codit.2018.8394935
29. Trardi, Y., Ananou, B., Haddi, Z., & Ouladsine, M. (2018). A novel method to identify relevant features for automatic detection of atrial fibrillation. 26th Mediterranean Conference on Control and Automation (MED). https://doi.org/10.1109/med.2018.8442479
30. Trardi, Y., Ananou, B., & Ouladsine, M. (2020). An advanced arrhythmia recognition methodology based on R-waves time-series derivatives and benchmarking machine-learning algorithms. European Control Conference (ECC). https://doi.org/10.23919/ecc51009.2020.9143678

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_31

31. Trardi, Y., Ananou, B., & Ouladsine, M. (2022). Computationally efficient algorithm for atrial fibrillation detection using linear and geometric features of RR Time-series derivatives. International Conference on Control, Automation and Diagnosis (ICCAD). https://doi.org/10.1109/iccad55197.2022.9853910
32. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2016.90
33. Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. Ch., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation, 101(23). https://doi.org/10.1161/01.cir.101.23.e215
34. Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH Arrhythmia Database. IEEE Engineering in Medicine and Biology Magazine, 20(3), 45–50. https://doi.org/10.1109/51.932724
35. Butterworth, S. (1930). On the Theory of Filter Amplifiers. Experimental Wireless and the Wireless Engineer, 7, 536–541.
36. Oppenheim, A.V., Willsky, A. S., & Young, I.T. (1983). Signals and Systems.
37. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953
38. Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing, 26(1), 43-49. https://doi.org/10.1109/tassp.1978.1163055
39. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
40. Yang, J., Li, C., Dai, X., Yuan, L., & Gao, J. (2022). Focal Modulation Networks. Advances in Neural Information Processing Systems 35 (NeurIPS 2022). https://proceedings.neurips.cc/paper_files/paper/2022/hash/1b08f585b0171b74d1401a5195e986f1-Abstract-Conference.html
41. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems 30 (NeurIPS 2017). https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
42. Qin, Q., Li, J., Zhang, L., Yue, Y., & Liu, C. (2017). Combining Low-dimensional Wavelet Features and Support Vector Machine for Arrhythmia Beat Classification. Scientific Reports, 7(1), 6067. https://doi.org/10.1038/s41598-017-06596-z
43. Yıldırım, Ö., Pławiak, P., Tan, R.-S., & Acharya, U. R. (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in Biology and Medicine, 102, 411–420. https://doi.org/10.1016/j.compbiomed.2018.09.009
44. Oh, S. L., Ng, E. Y. K., Tan, R. S., & Acharya, U. R. (2019). Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types. Computers in Biology and Medicine, 105, 92–101. https://doi.org/10.1016/j.compbiomed.2018.12.012
45. Kim, J.-H., Seo, S.-Y., Song, C.-G., & Kim, K.-S. (2019). Assessment of Electrocardiogram Rhythms by GoogLeNet Deep Neural Network Architecture. Journal of Healthcare Engineering, 2019, 1–10. https://doi.org/10.1155/2019/2826901
46. Zubair, M., & Yoon, C. (2022). Cost-Sensitive Learning for Anomaly Detection in Imbalanced ECG Data Using Convolutional Neural Networks. Sensors, 22(11), 4075. https://doi.org/10.3390/s22114075