2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A New Approach to Classify Cardiac Arrythmias Using 2D Convolutional Neural Networks *

J. R. G de Santana, M. G. F. Costa, *Member*, *IEEE* and C. F. F. Costa Filho*, Member, IEEE*

Abstract— **Cardiovascular diseases are the number one cause of death worldwide. Detecting cardiovascular diseases in its early stages could effectively reduce the mortality rate by providing timely treatment. In this study, we propose a new methodology to detect arrythmias, using 2D Convolutional Neural Networks. The main characteristic of the proposed methodology is the use of 15 x15 pixels gray-level images, containing the values of a heartbeat of the ECG signal. This work aims to detect 17 arrythmias. To validate and test the proposed methodology, MIT-BIH database, the main benchmark database available in literature, was used. When compared to other results previously published, the obtained precision, 92.31%, is in the state-of-the-art.**

Clinical Relevance— **The presented work provides an automatic method to detect arrythmias in ECG signals by a new methodology.**

I. Introduction

According to the World Health Organization [1], cardiovascular diseases (CVDs) are the number one cause of death worldwide, claiming about 17.9 million lives each year. CVDs are a group of diseases of the heart and blood vessels that include rheumatic heart disease, cerebrovascular disease, coronary heart disease and other conditions. Four out of 5 deaths from CVDs are due to heart attacks and strokes, and one third of these deaths occur prematurely in people under 70 years of age.

According to the Cardiovascular Statistics Brazil 2020 report [2], of the Brazilian Society of Cardiology, CVDs were the main cause of death in Brazil in the year 2017. Among CVDs, ischemic heart diseases were the cause number 1.

Detecting CVDs in its early stages could effectively reduce the mortality rate by providing timely treatment [3]. One of the common sources of CVDs is cardiac arrhythmia, which is characterized by the fact that heartbeats deviate from your normal pattern. A normal heartbeat varies with age, body size, activity and emotions. In cases where the heartbeat seems fast or slow, the condition is known as a palpitation. An arrhythmia does not necessarily mean that the heart is beating too fast or too slowly, it indicates that the heart is following an irregular beating pattern.

Conventionally, for the diagnosis of arrhythmias, cardiologists visually inspect 12-lead ECG waveforms in a digital image format [4]. It is common that ECG signals lasting many hours and even days, like Holter monitoring, need to be analysed. This is a very time-consuming and exhausting procedure, which significantly limits the impartiality of the diagnosis. This limitation can be eliminated with the use of computational techniques for the automatic detection of arrhythmia and ECG classification.

In the literature, we have identified two groups of work developed for the automatic detection of arrhythmias. The first group comprises older works that used classic machine learning tools, while the second group comprises more recent works that use deep learning techniques, such as convolutional and recurrent neural networks.

Among others, in the first group of works we identified the use of the following classic machine learning techniques: wavelet transform [5], support vector machine [6], multilinear single value decomposition [7], hidden Markov models [8], etc.

In the second group of works, we identified two trends. The first one comprises the works that use only Convolutional Neural Networks (CNNs), the most frequent being one-dimensional networks [9][10]. Other studies use two-dimensional CNNs [11][12], converting the one-dimensional ECG tracings into a 2D array with dimensions of 128x128 or 64x64, for example. The second trend includes works that use hybrid architectures, a serial composition of CNN and Recurrent Neural Networks (RNNs) [13][14][15]. In this hybrid architecture, CNNs are used to feature extraction. The features extracted, after a dimensionality reduction step, through maxpooling layers, feed the input of a RNN, responsible for the ECG classification, according to normal or altered states (arrythmias).

In this work, we also used 2D CNNs for the classification of ECG, however, 2D ECG data was composed in a different way from that had been used in [11][12]. To obtain the input signal of the 2D CNN, 225 samples of a heartbeat of one-dimensional ECG signal are converted into a 15 x15 pixels gray-level image. From a computational point of view, working with smaller images puts less stress on low memory devices when, for example, the implementation is embedded in mobile devices.

The advantage of working with a 2D signal instead of a 1D signal, is that the convolution operation with 2D kernel explores new neighborhood relations (neighborhoods above and below) in addition to the neighborhood relation already explored with 1D convolution operation (neighborhoods on the left and on the right).

To validate and test the proposed methodology we use the most used database to benchmark algorithms developed for detecting arrhythmias, known as the MIT-BIH database [16].

* This research, according for in Article 48 of Decree nº 6.008/2006, was funded by Samsung Electronics of Amazonia Ltda, under the terms of Federal Law nº 8.387/1991, through agreement nº 004, signed with the Center for R&D in Electronics and Information from the Federal University of Amazonas - CETELI/UFAM.

J. R. G de Santana, M. G. F. Costa and C. F. F. Costa Filho are with Federal University of Amazonas, Amazonas, Brazil (e-mails: joaorobertojrgs@yahoo.com.br, mcosta@ufam.edu.br, ccosta@ufam.edu.br).

Most published academic papers, that use this database, classified 5 or 8 15 cardiac disfunctions. Few studies have been carried out aiming to classify all 17 classes present in MIT-BIH database (normal ECG, pacemaker signal, and 15 cardiac disfunctions) [10][6].

We also intend to present a performance comparison between 2D CNN, 1D CNN and RNNs (Long-Short-Time-Memory -LSTM and Gated-Recurrent-Units - GRU), in the classification of the 17 classes present in the MIT-BIH database, a gap little explored in the literature.

II. METHODOLOGY

A. *Materials*

In this work, the ECG arrhythmia database of MIT-BIH was used. The ECG signals, obtained from 47 patients, were sampled at a rate of 360 Hz. The base consists of 48 annotated records, with an approximate duration of 30 minutes of heartbeat [9]. 17 classes are present in the database: Normal Sinus Rhythm (NSR), pacemaker rhythm (PR) and the following 15 cardiac disfunctions: Atrial Premature Beat (APB), Atrial Flutter (AFL), Atrial Fibrillation (AFIB), Supraventricular Tachyarrhythmia (SVTA), Pre-excitation (WPW), Premature Ventricular Contraction (PVC), Ventricular Bigeminy (B), Ventricular Trigeminy (T), Ventricular Tachycardia (VT), Idioventricular Rhythm (IVR), Ventricular Flutter (VFL), Fusion of Ventricular and Normal Beat (F), Left Bundle Branch Block Beat (LBBBB), Right Bundle Branch Block Beat (RBBBB), Second-Degree Heart Block (SDHB).

For each heartbeat, a one-dimensional record with 225 samples around an R peak, not necessarily with it centralized, was extracted. 150 records of each class were randomly extracted, totaling 2550 records.

B. *Pre-Processing*

From each one-dimensional record was generated one intensity image. Figure 1 illustrates the steps employed in this generation.

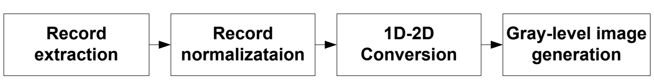


Figure 1. Steps to obtain a 15 x 15 pixels gray-level image of a record with 225 samples extracted around an R-wave peak: record extraction, record normalization, 1D-2D conversion, gray-level image generation.

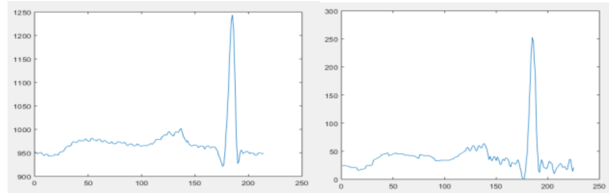
The normalization step, shown in equation (1), aims obtaining sample values in the range [0 255].

$$s_n = \frac{s - s_{min}}{s_{max} - s_{min}} . 255 \qquad (1)$$

where:

s_n: *normalized sample value*
s: *non − normalized sample value*
s_{min}: *minimum sample value*
s_{max}: *maximum sample value*

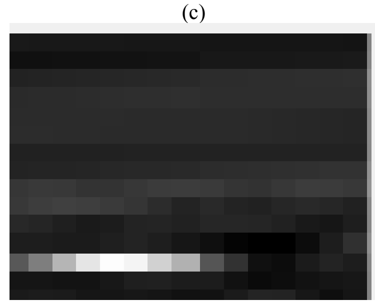
In the gray-level image, the pixels intensities correspond to the values of the normalized heartbeat ECG signal. Figure 2 shows corresponding examples of the process steps presented in Figure 1.



(a) (b)

26	24	23	22	23	23	24	22	20	19	21	21	21	18	19
18	18	19	19	21	20	20	19	23	24	26	25	27	29	32
34	34	35	34	36	40	42	42	42	41	42	44	45	43	44
43	44	46	47	44	44	42	44	48	48	46	45	45	43	46
46	46	44	42	42	44	44	42	41	40	39	40	42	40	38
39	38	41	42	42	42	40	40	40	42	39	37	35	37	38
38	38	36	35	35	36	38	38	35	35	34	35	35	36	35
35	37	37	38	39	38	40	41	45	46	46	46	46	46	50
51	53	54	53	55	58	59	56	54	52	52	57	59	58	57
60	63	64	58	52	48	46	42	44	39	37	36	34	36	36
37	36	35	32	34	35	34	34	34	30	33	34	35	34	32
31	34	34	33	33	29	27	19	16	11	4	0	5	26	57
85	121	169	215	246	255	232	173	94	31	5	9	23	28	26
22	23	25	26	26	24	22	24	23	23	23	19	19	19	23
23	23	23	22	23	24	25	24	23	21	21	23	25	24	24

(c)



(d)

Figure 2. Corresponding examples of process steps shown in Figure 1: (a) Extracted record; (b) Normalized record; (c) 1D-2D conversion and (d) gray-level image obtained.

C. *Deep Neural Networks and Training Parameters*

Figure 3 shows the architectures of neural networks implemented. The 1D and 2D CNN architectures are shown in Figure 3(a), while the LSTM and GRU architectures are shown in Figure 3(b).

(a) (b)

Figure 3: (a) 1D or 2D CNN architecture and (b) LSTM or GRU architecture

The differences between the 1D CNN and 2D CNN architectures are the kernel sizes and the input layer. In 1D CNN, the kernel size is [1 n], while in 2D CNN the kernel size is [n n]. Both convolutional kernels have $n = 2$. The size of input layer is [1 225] in 1D CNN, corresponding to a vector with 255 values of a heartbeat. The size of input layer is 15x15 in 2D CNN, and corresponds to a gray-level image, as shown in Figure 2(d). In the CNN architecture there are seven features extraction blocks, consisting of the following layers: Conv→Batch→Relu. The number of filters of the convolutional layers is equal to 96. A stride of 1 is used. In the sequence, there are the following layers: fully connected layer (*200 neurons*)→ dropout layer → fully connected layer (*17 neurons*) →softmax layer → classification layer.

The recurrent neural networks, LSTM or GRU, are formed by the following layers: a sequence input layer (input size = 225) → a LSTM or GRU layer (*4096 neurons*) → a fully connected layer (*512 neurons),* a dropout layer → a fully connected layer *(17 neurons)* → a softmax layer and a classification layer.

The methodology used for training and testing was the 5-fold-cross-validation. In each fold 80% of data is used for training and 20%, for validation. In the training, after an exhaustive search, the following parameters were chosen *bath size = 512*, *L_2 regularization factor = 0.1*, *dropout factor = 0.5*, *maximum number of epochs* = 2000. For learning rate, a piecewise schedule was employed, with *initial learning rate = 0.01*, a *drop period = 200 epochs* and a *drop factor = 0.75*. Figure 4 shows the precision during a training section of a 2D CNN. At the beginning of the training, due to the high values of the learning rate, several oscillations occurred. However, as the learning rate decreased with the increase in the number of epochs, these oscillations also decreased. Three optimization methods were evaluated: Stochastic Gradient Descent with Momentum - SGDM, Adaptive Moment Estimation Optimizer - ADAM and Root Mean Square Propagation - RMSProp. The best performances were obtained with the SGDM optimizer. We notice that the best performance of the training, evaluated in the validation set, does not necessarily occur at the end of the training. That is why we use the MATLAB checkpoint feature, recording the network with the best performance during the training.

The experiments were performed using MATLAB® version 2018b, a 3.2 GHz Intel i7-8700 processor computer with 16 GB RAM and 8 GB GeForce GTX 1070 GPU.

D. Evaluation Metrics

The following metrics were used for evaluation: accuracy (Acc), sensitivity (Sens) or recall, specificity (Spec) and precision (P) and F1-score. For each fold, these metrics are calculated according to equations (2), (3), (4), (5) and (6), respectively [10]. Accuracy can be a misleading metric for imbalanced data sets. To make up for this deficiency, we also calculated specificity and sensitivity. Precision talks about how precise/accurate the model is out of those predicted positive, how many of them are actual positive. F1 Score might be a better measure to use if we need to seek a balance between Precision and Recall and there is an uneven class distribution (large number of Actual Negatives).

$$Acc = \frac{1}{17}\sum_{c=1}^{17}\frac{T_p^c+T_n^c}{T_p^c+T_n^c+F_p^c+F_n^c} \quad (2)$$

$$Sens = \frac{1}{17}\sum_{c=1}^{17}\frac{T_p^c}{T_p^c+F_n^c} \quad (3)$$

$$Spec = \frac{1}{17}\sum_{c=1}^{17}\frac{T_n^c}{T_n^c+F_p^c} \quad (4)$$

$$P = \frac{1}{17}\sum_{c=1}^{17}\frac{T_p^c}{T_p^c+F_p^c} \quad (5)$$

$$\text{F1-score} = 2 * \frac{(P*Sens)}{(P+Sens)} \quad (6)$$

where T_p^c denotes the true positives: all *c* instances that are classified as *c; T_n^c* denotes true negatives: all non-*c* instances that are not classified as *c; F_p^c* denotes the false positives: all non-*c* instances that are classified as *c*; F_n^c denotes false negatives: all *c* instances that are not classified as *c*.

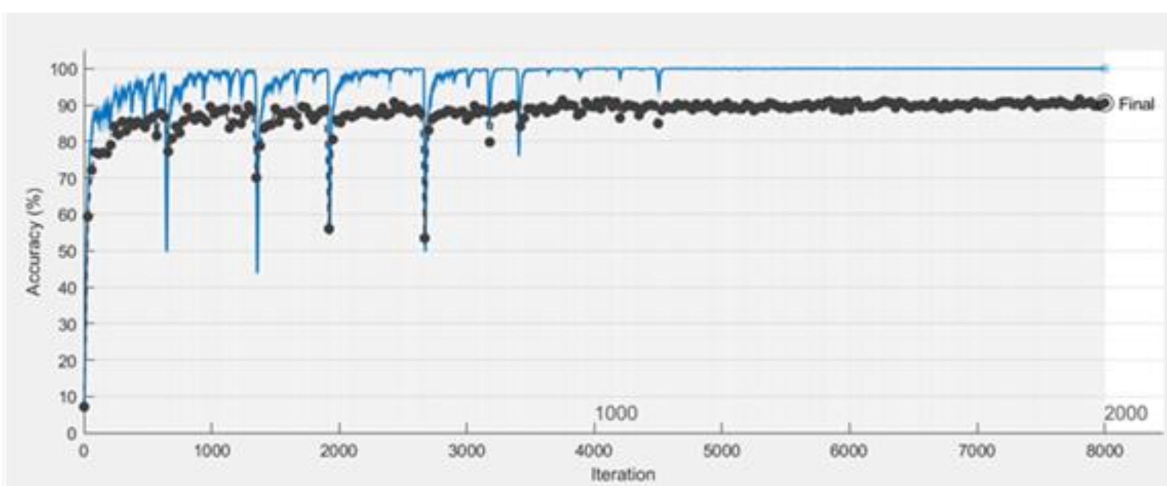


Figure 4. Precision during the training section of a 2D CNN. blue curve: training precision. black-curve – validation precision.

III. RESULTS AND DISCUSSIONS

Table I shows the performance of all deep neural networks evaluated in this work: 1D CNN, 2D CNN, LSTM and GRU, with 5-fold-cross validation. As shown, the best performance was obtained with the 2D CNN network, with a precision of 92.31%. Figure 5 shows a confusion table obtained with a 2D CNN.

Table II shows a comparison of the performance of the 2D CNN with other results previously published in the literature for classification of 17 classes of MIT-BIH database. As shown, the results obtained in this paper are in the state-of-the-art.

As mentioned, some previous arrhythmias detection works use gray-level images [11] [14] [12]. In these works, the authors used, as input signal, a gray-level image of the ECG tracing (like an ECG signal photo) with different sizes: 256 x 256, 192 x 128, and 64 x 64 pixels, respectively. In this work, we also use a gray-level image. However, the intensity values of the pixels correspond to the values of the normalized heartbeat ECG signal. Therefore, all pixels of the gray-level image carry relevant information about ECG signal.

Also considering that, this image is significantly smaller than those used in aforementioned papers, the proposed method is more suitable to be implemented on platforms with reduced memory capacity, such as mobile devices.

Although the most recent studies for arrhythmias detection use hybrid architectures, CNN-LSTM, in this work we managed to obtain cutting-edge performance, using only a CNN architecture.

We credit this better performance of the 2D-CNN to the fact that new neighborhood relationships are explored through a proposed gray-level image that all pixels carry information about a heartbeat of the ECG signal.

In future works, we intend to 1) explore data augmentation, creating new images through small gray level changes in the pixels of the gray-level image; 2) evaluate other arrythmia classifications in 2, 5 and 8 classes; 3) implement the proposed method in mobile devices.

TABLE I RESULTS OF THE 4 ARCHITECTURES EVALUATED IN THIS STUDY (MEAN±SD)

Architecture	Accuracy	Sensitivity	Specificity	Precision	F1-Score
1D CNN	94.29±2.14	47.64±18.05	97.00±1.92	52.27±27.28	49.85±22.37
2D CNN	99.13± 0.65	93.07±6.10	99.52±0.42	92.31±6.87	92.68±5.64
LSTM	96.21±1.65	64.65±15.15	98.09±1.48	68.13±24.72	66.34±18.53
GRU	95.71±1.82	62.99±16.12	97.70±1.55	63.51±24.19	63.24±18.35

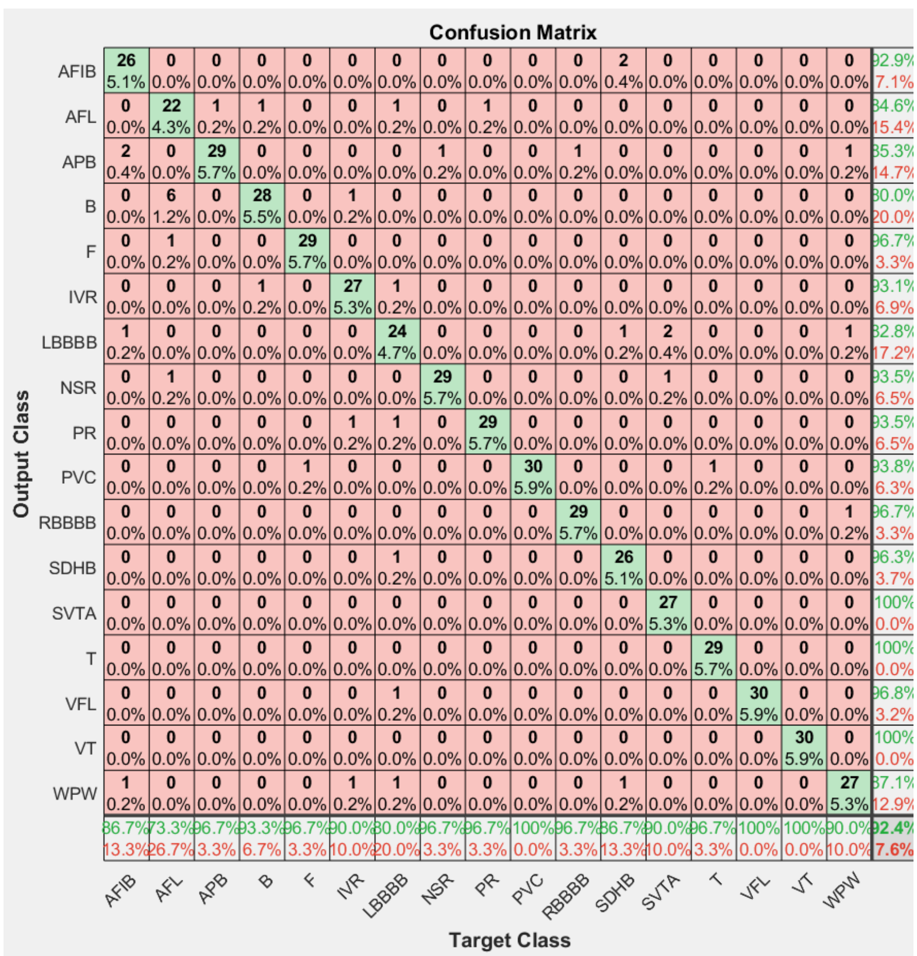


Figure 5. Confusion Matrix obtained with 2D CNN showing precision values.

TABLE II COMPARISON OF THE 2D CNN WITH OTHER RESULTS PREVIOUSLY PUBLISHED IN LITERATURE

Work	Year	Precision
Yildirim [10]	2018	91.33
Plawiak [6]	2018	91.40
Proposed Method (2D CNN)	2021	92.31

IV. CONCLUSION

In this paper we proposed a new methodology for arrythmias detection in ECG signals, using 2D CNNs. The main characteristic of the proposed method was the use of 15x15 pixels gray-level images of the ECG signal, with the pixel's intensities corresponding to the ECG values. The obtained results, with a precision of 92.31% is in the state-of-the-art. In future works we intend to explore data augmentation to improve the results. We also intend to implement the proposed method in mobile devices.

REFERENCES

[1] WHO, "WHO," 2021. [Online]. Available: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1.

[2] G. M. M. Oliveira *et al.*, "Cardiovascular Statistics – Brazil 2020," *Arq. Bras. Cardiol.*, vol. 115, no. 3, pp. 308–439, 2020.

[3] A. Mustaqeem, S. M. Anwar, and M. Majid, "A modular cluster based collaborative recommender system for cardiac patients," *Artif. Intell. Med.*, vol. 102, no. November 2019, p. 101761, 2020.

[4] G. A. Tadesse *et al.*, "Cardiovascular disease diagnosis using cross-domain transfer learning," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 4262–4265, 2019.

[5] Y. Jung and W. J. Tompkins, "Detecting and Classifying Life-threatening ECG Ventricular Arrhythmias using Wavelet Decomposition," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 3, pp. 2390–2393, 2003.

[6] P. Pławiak, "Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals," *Swarm Evol. Comput.*, vol. 39, no. April, pp. 192–208, 2018.

[7] M. Bousse, G. Goovaerts, N. Vervliet, O. Debals, S. Van Huffel, and L. De Lathauwer, "Irregular heartbeat classification using Kronecker Product Equations," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 438–441, 2017.

[8] D. Novák, D. Cuesta-Frau, T. Al Ani, M. Aboy, P. Mico, and L. Lhotská, "Speech recognition methods applied to biomedical signals processing," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 26 I, pp. 118–121, 2004.

[9] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.

[10] A. Ullah, S. Muhammad Anwar, M. Bilal, and R. Majid Mehmood, "Classification of arrhythmia by using deep learning with 2-d ecg spectral image representation," *arXiv*, vol. 2, pp. 1–14, 2020.