Low-Complexity Deep Learning Models for Accurate Atrial Fibrillation Diagnosis

Carlos A. Fajardo[a;*], Andrés S. Parra[a], Tania Castellanos-Parada[a] and Kaushik Roy[b]

[a]Universidad Industrial de Santander
[b]Purdue University
ORCiD ID:

Abstract. Timely detection and diagnosis of Atrial Fibrillation (AF) are crucial for prompt treatment and medical intervention. However, the computational complexity of AF-diagnosing algorithms, such as Deep Learning Models, presents challenges for implementation on portable devices. This study focuses on AF diagnosis using low-complexity algorithms, which introduces various challenges, including real-time processing, low power consumption, memory constraints, and handling noisy and low-quality ECG data. The study proposes multiple low-complexity deep learning models that achieve state-of-the-art accuracy. To attain this goal, several CNN models were developed, striking a balance between high accuracy and a minimal number of parameters. As a result, the study successfully achieved state-of-the-art accuracy using these low-complexity deep learning models. Our results suggest that AF detection is feasible using low-complexity models, which can be implemented on portable devices without compromising prediction accuracy.

1 Introduction

Atrial fibrillation (AF) is the most common arrhythmia diagnosed in clinical practice, and its widespread prevalence is considered alarming for some researchers. Even an epidemic is forecasted within the next 10 to 20 years [6].

Early detection of AF can reduce the risk of morbidity and mortality. Nevertheless, AF detection can be challenging because the duration of AF episodes varies over time for each individual. These episodes, also known as Paroxysmal AF (PAF), may be very sporadic and commonly asymptomatic, especially at the early stage of the disease. Consequently, specialists recommend long-term monitoring for patients who experience occasional events and symptoms associated with AF.

Systematic diagnosis of PAF is a major public health concern since early diagnosis is essential to identify candidates for oral anticoagulation and catheter ablation, which is usually curative when used at this time [9].

In this sense, machine learning advancements have enabled the detection of AF by computer-aided diagnosis (CAD) systems with high accuracy [7,8]. Computer-aided interpretation has become highly important in the field of healthcare and has proven to offer an important alternative in the clinical ECG work-flow [2].

However, the high computational cost of these machine learning algorithms has not allowed the implementation of AF system-aided diagnosis on wearable devices (at the edge).

This study addresses the challenges related to ECG data processing at the edge. Therefore, it aims at finding low-complexity algorithms that facilitate real-time processing capabilities and low power consumption. Furthermore, it deals with low-quality and noisy ECG signals, which are very likely to appear in wearable devices.

We propose different low-complexity deep learning models achieving state-of-the-art accuracy. Results regarding F1 score, and computational complexity in terms of the number of parameters were determined using the Icentia11k dataset [12].

The remaining sections of this paper are organized as follows: Section 2 introduces the dataset utilized in this study. Section 3 provides a detailed description of the proposed models. Sections 4 and 5 summarize and discuss the obtained results, respectively, and also compare our research with related works. Section 6 concludes the study.

2 Data description

In this work, we used the Icentia11k dataset [12], which contains normal sinus rhythms, noises, AF, and atrial flutter signals. As suggested by Hannun et. al. [2], we merged the AF and atrial flutter signals in one single class.

It is important to highlight that the Icentia11k stands as the largest publicly available ECG dataset, comprising data from 11 thousand patients and 2 billion labelled beats. Besides, the Icentia11k dataset contains a substantial number of noise signals, accounting for around 40% of the data. As previously mentioned, the ability to effectively handle noise signals is a critical objective in the development of wearable devices.

We split the dataset into training, validation, and test set using 64%, 16%, and 20% of the data, respectively. It is important to note that each set has separated patients; therefore, there are no shared patients between sets.

3 Models

We introduce a type of deep *elastic* model capable of adjusting its depth, width, and kernel-size through trainable hyperparameters. This model is inspired in [2, 10, 13]. Our objective is to identify the minimum values for these hyperparameters that still enable the

* Corresponding Author. Email: cafajar@uis.edu.co

model to achieve state-of-the-art accuracy. By determining these optimal values, we can create the most compact version of the model while maintaining its performance.

The network architecture consists of N residual blocks (Figure 1), where N, I_{ch} and S_j are hyperparameters that will allow us to vary the model size.

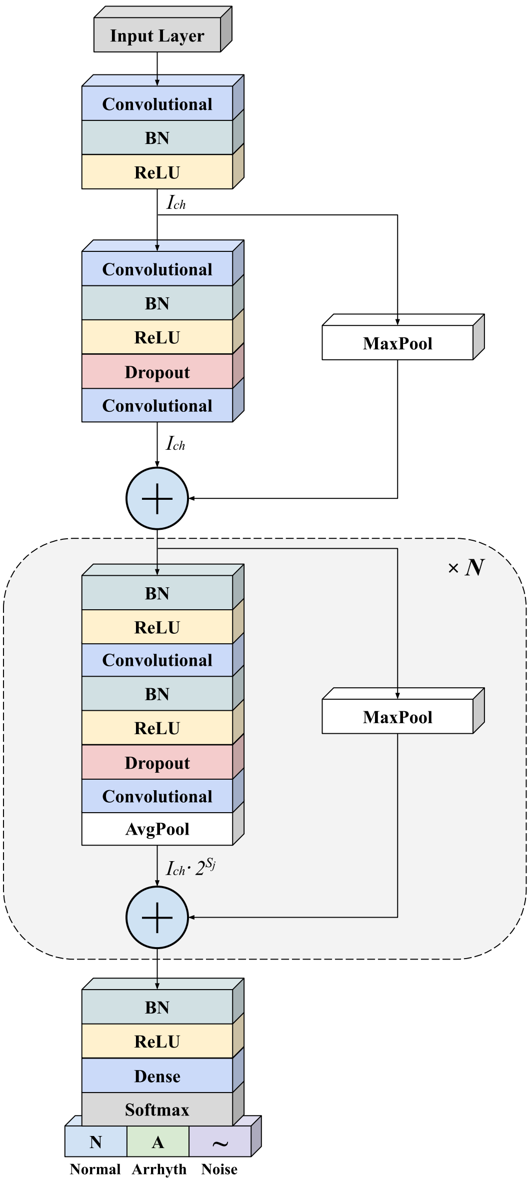

Figure 1: The CNN consist of ***N*** residual blocks and accepts raw ECG data as input, and outputs a prediction of Normal (N), cardiac Arrhythmia (A) or Noise signal ($\sim$). The number output channels of the first layers is I_{ch}. The number of output channels of the layers in the residual blocks is $I_{ch} \cdot 2^{s_j}$. The *depth* and *width* of the network are changed by the hyperparameters N, I_{ch} and s_j

The first and last layers of the model are special cases due to this pre-activation block structure [10]. The model uses residual blocks as a means to deal with the vanishing or exploding gradient problem [3]. Moreover, the model includes skip-connections which favor the propagation of the information in deep neural networks. The model uses Batch Normalization to keep values in-bounds and avoid saturation. It applies Dropout to prevent overfitting during training, followed by a Rectified Linear Activation Unit (ReLU).

Each residual block culminated in average pooling. Besides, Maxpooling layers were also used on the skip connections to maintain dimensional consistency when the two separate paths joined back together at each block. The model finishes with a fully connected layer followed by the Softmax function, which has 3 outputs corresponding to the probability for each class (normal, arrhythmia, or noise).

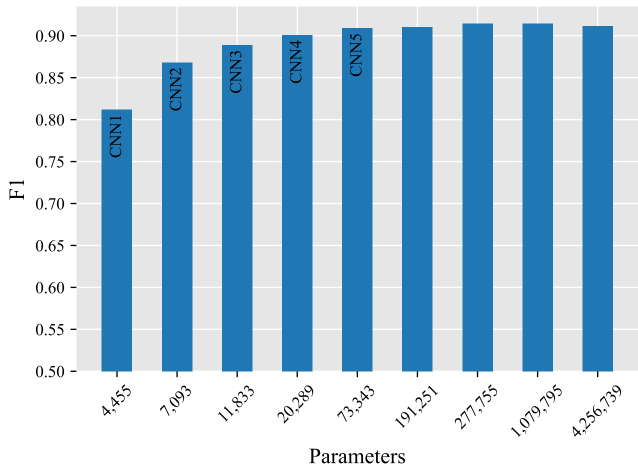

Figure 2: F1-scores of several models derived from our baseline model (Fig 1).

4 Results

The deep learning models were implemented in Python using the Keras library with a TensorFlow backend. The networks were trained by Adam optimizer with the default parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$, using a mini-batch size of 128. The deployed learning rate was 0.001 and was reduced by a factor of 10 when the loss plateaus. We chose the model that achieved the lowest loss on the validation dataset.

We tested several configurations by varying the *kernel-size* (K), *depth* (N), and *width* (I_{ch} and s_j). The *kernel-size* is a highly influential hyperparameter that underwent extensive training. Our findings indicate that the best value for the size of the kernel is $K = 16$. We also tested different ways to apply down-sampling in the CNN by setting a stride greater than 1. The downsampled late strategy [4] was used and has shown to improve the accuracy on a limited budget of parameters. This strategy aims to have large activation maps at the beginning of the network, which can lead to higher classification accuracy.

We trained several models with the aim of finding their F1-score. Figure 2 illustrates the top-performing models discovered in our study. Notably, with just 20,289 parameters, we achieved an F1-score of 0.902, and larger models did not present a significant F1-score improvement. For this reason, it was decided to focus on the four smallest models (CNN1-CNN4). Table 1 presents the key characteristics of these four models.

Table 1: Main characteristics of the CNN 1 - CNN 4 models.

Models	Kernel Size	N	I_{ch}	$C_{out}^{j} = I_{ch} \cdot 2^{s_j}$	Parameters
CNN1	16	8	2	2, 2, 2, 2, 2, 2, 2, 2	4,455
CNN2	16	13	2	2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 4, 4, 4	7,093
CNN3	16	11	2	2, 2, 2, 2, 4, 4, 4, 4, 8, 8, 8	11,833
CNN4	16	13	2	2, 2, 2, 2, 4, 4, 4, 4, 8, 8, 8, 8, 16	20,289

Table 2: Recent studies, on larger data sets, classifying ECG signals using Deep Learning techniques.

Study	Database	Total Data & Classes	Number of patients	Technique	Number of parameters	Results
Hannun et al., 2019 [2]	Zio Monitor	91,232 records 12 classes	53,549 patients	CNN	10,473,635	F1 = 0.837
Rubin et al., 2018 [11]	PhysioNet Challenge 2017	8,528 records 4 classes	8,528 patients	CNN	262,344	F1 = 0.820
Yao et al., 2020 [14]	1st China Physiological Signal Challenge	9,831 records 8 classes	9,831 patients	ATI-CNN	4,984,640	F1 = 0.812
Fonseca et al., 2022 [1]	Icentia11k	550,000 records 4 classes	11,000 patients	CNN Constrastive *Few-shot	250,000	F1 = 0.801*
CNN1 (**Ours**)	Icentia11k	550,000 records 4 classes	11,000 patients	CNN	4,455	F1 = 0.813
CNN4 (**Ours**)	Icentia11k	550,000 records 4 classes	11,000 patients	CNN	20,289	F1 = 0.902

5 Discussion

Deep learning has had a profound impact on enhancing state-of-the-art accuracy across various classification and detection tasks. As a result, there is a growing trend in deep learning research concerning arrhythmia detection [8]. However, these studies have primarily focused on improving accuracy by increasing model size, leading to oversized models [5].

The substantial size of large deep learning models poses significant challenges, including high memory consumption and demanding computational resources, rendering their implementation on edge devices impractical. Additionally, the incorporation of recurrent layers, such as short-term memory (LSTM) cells, or the adoption of transformer-based approaches can further increase computational complexity and hinder real-time inference.

Table 2 presents a comparison between our proposed models and selected recent ECG classification studies that have specifically focused on larger datasets (comprising more than 8000 patients) and have reported the number of parameters. The table provides detailed information on various aspects, including the amount of data and classes, the number of patients involved, the deep learning techniques employed, and the corresponding accuracy results achieved by each study.

Notably, our CNN-4 model, consisting of only $20,289$ parameters, achieved an impressive F1-score of 0.902.

6 Conclusions and Future Work

This research tackled the challenges related to AF detection by proposing low-complexity deep learning models suitable for deployment on portable devices.

Future work will focus on employing compression strategies, such as pruning and quantization, to effectively reduce model size while preserving performance. Additionally, we aim to deploy compressed models on edge devices like FPGAs and System on Chip (SoC). This will involve model adaptation and performance optimization to ensure efficient execution on resource-constrained edge devices.

Ultimately, our goal is to contribute to early AF detection at the edge, enabling timely treatment and reducing the associated risks of morbidity and mortality. By pursuing these research areas and objectives, we aspire to make significant advancements in the field of AF detection, leading to improved patient outcomes and healthcare practices.

References

[1] Karen Fonseca, Sergio Osorio, Jeyson Castillo, and Carlos Fajardo, ‘Contrastive learning for atrial fibrillation detection in challenging scenarios’, in *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 1218–1222. IEEE, (2022).

[2] Awni Y. Hannun et al., ‘Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network’, *Nature Medicine*, **25**(1), 65–69, (2019).

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, ‘Deep residual learning for image recognition’, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, (2016).

[4] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer, ‘Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size’, *arXiv preprint arXiv:1602.07360*, (2016).

[5] Giosué Cataldo Marinó, Alessandro Petrini, Dario Malchiodi, and Marco Frasca, ‘Deep neural networks compression: A comparative survey and choice recommendations’, *Neurocomputing*, **520**, 152–170, (2023).

[6] Carlos A Morillo et al., ‘Atrial fibrillation: the current epidemic’, *Journal of geriatric cardiology: JGC*, **14**(3), 195, (2017).

[7] Fatma Murat, Ozal Yildirim, Muhammed Talo, Ulas Baran Baloglu, Yakup Demir, and U. Rajendra Acharya, ‘Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review’, *Computers in Biology and Medicine*, **120**(March), 103726, (2020).

[8] S. Parvaneh, Jonathan Rubin, S. Babaeizadeh, and Minnan Xu-Wilson, ‘Cardiac arrhythmia detection using deep learning: A review’, *Journal of Electrocardiology*, **57**, S70–S74, (2019).

[9] Joao Primo et al., ‘Prevalence of paroxysmal atrial fibrillation in a population assessed by continuous 24-hour monitoring’, *Revista Portuguesa de Cardiologia (English Edition)*, **36**(7-8), 535–546, (2017).

[10] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng, ‘Cardiologist-level arrhythmia detection with convolutional neural networks’, *arXiv preprint arXiv:1707.01836*, (2017).

[11] Jonathan Rubin, Saman Parvaneh, Asif Rahman, Bryan Conroy, and Saeed Babaeizadeh, ‘Densely connected convolutional networks for detection of atrial fibrillation from short single-lead ecg recordings’, *Journal of electrocardiology*, **51**(6), S18–S21, (2018).

[12] Shawn Tan, Guillaume Androz, Ahmad Chamseddine, Pierre Fecteau, Aaron Courville, Yoshua Bengio, and Joseph Paul Cohen, ‘Icentia11k: An unsupervised representation learning dataset for arrhythmia subtype discovery’, *arXiv preprint arXiv:1910.09570*, (2019).

[13] Zhaohan Xiong, Martin K. Stiles, and Jichao Zhao, ‘Robust ECG signal classification for detection of atrial fibrillation using a novel neural network’, *Computing in Cardiology*, **44**, 1–4, (2017).

[14] Qihang Yao, Ruxin Wang, Xiaomao Fan, Jikui Liu, and Ye Li, ‘Multiclass Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network’, *Information Fusion*, **53**(June 2019), 174–182, (2020).