A Lightweight Unidimensional Deep Learning Model for Atrial Fibrillation Detection

Quenaz B Soares[1,2], Rosangela Monteiro[1], Fábio B Jatene[1], Marco A Gutierrez[1]

[1] Heart Institute Clinics Hospital, University of Sao Paulo Medical School, Brazil
[2] Polytechnic School, University of Sao Paulo, Brazil

Abstract

Continuous rhythm monitoring using wearable devices is a potential tool for early identification of atrial fibrillation (AF), the most frequent cardiac arrhythmia (with 0,51% worldwide prevalence, increasing with time), and is also a tool for remote monitoring patients after cardiac surgery. However, AF detection directly through wearable devices is limited by the computational complexity of the classifier model.

In this work we propose a lightweight AF classifier model based on the VGG-11 architecture (LiteVGG-11), focusing on reducing the number of parameters and numerical operations. Using a low number of filters, depthwise separable convolution, and global pooling, this model has only 20,454 parameters and needs 6.9 MFLOP to make an inference for an input of 10 seconds of the ECG leads I and II, sampled at 200 Hz.

To test its effectiveness for AF detection we used the PhysioNet/CinC Challenge 2021 public dataset, stratifying the classes into sinus rhythm, AF, and other rhythms. After 10 Monte Carlo cross-validation splits, with 24,260 unbalanced samples for training and 1,536 balanced samples for validation and testing, the observed metrics (mean±standard deviation) were: Se 94.1±0.1%; Sp 91.9±0.8%; F1-Score 89.50.7±%; and AUC 96.1±0.6%.

1. Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia, with a global prevalence of 0.51% [1]. According to the Global Burden of Disease study from 2019, the age-standardized prevalence rate in Brazil is 0.537% [2], but data from the Telehealth Network of Minas Gerais shows a prevalence of 1.33% across all ages and up to 7.0% in octogenarians [3]. The prevalence of AF increases as the population ages [4], leading to estimations of a 1.7% prevalence in 2025 [5]. Furthermore, the incidence of AF after cardiac surgery varies according to the type of surgery, ranging from 16% to 63% [6].

Continuous rhythm monitoring via wearable devices powered by artificial intelligence algorithms may improve early AF diagnosis, allowing for prompt treatment and potentially preventing malign outcomes (like stroke and heart failure) [3, 7]. According to a 2004 study [8], 32.3% of the patients after coronary artery bypass graft (CABG) surgery developed AF; from these 76.8% were first diagnosed by using continuous monitoring and only 17.5% by traditional 12-lead ECG. Although high, this incidence may be even higher as continuous ECG monitoring is needed to detect most of the paroxysmal AF events [9].

Classical solutions for AF screening include hand-held ECG devices, used periodically over multiple weeks [10, 11]. As wearable technologies become increasingly affordable and utilized by the general public, AF detection algorithms included in these devices can be an attractive alternative to existing ECG-based solutions [12].

However, the computational complexity of most classification models, particularly those based on deep learning algorithms, is one of the challenges for using wearable devices in continuous rhythm monitoring. To improve accuracy, these models frequently use deeper and more complex models [13, 14] or ensembles of models [15], usually without taking into account the energy consumption or computational limitations of portable devices [16].

In this work, we propose a lightweight deep convolutional neural network based on the VGG-11 architecture [17] and apply it as a rhythm classifier, focusing on the AF but also detecting the presence of other abnormalities.

2. Materials and method

2.1. Dataset

The dataset ensemble made available by the George B. Moody PhysioNet Challenge 2021 - "Will Two Do? Varying Dimensions in Electrocardiography" [18] was used in this study to validate the proposed model as an AF classifier. This ensemble consists of seven publicly available datasets which included 88,253 annotated ECG recordings, with AF as one of the diagnoses in 5,284 of them.

As this dataset comes from multiple sources, its signals have different characteristics. The sampling rate varies from 257 Hz to 1000 Hz, and the duration from 6 seconds up to 30 minutes. The set of possible diagnoses also changes from source to source, leading to a total of 111 diagnostic classes that include heart rhythms and disorders, as well as other ECG findings.

The ECG signals were preprocessed as follows:

1. Selection of the leads I and II;
2. Downsampling to 200 Hz using polyphase filtering;
3. Z-score normalization for individual channels.
4. Cropping/zero-padding of the first 10 seconds of signal;
5. Exclusion of signals with constant voltage;

The 200 Hz sampling rate and leads I and II were chosen to simulate a wearable ECG condition. No further preprocessing, such as filtering or data augmentation, was applied in this study.

2.2. Proposed model

The VGG models were proposed for use in computational vision tasks taking images as input [19]. When used with unidimensional inputs, they usually are adapted by only swapping the 2D convolutional layers for the 1D version.

Our approach uses depth-wise separable 1D convolution layers (DWConv), a reduced number of filters, global average pooling as flattening, and a reduced number of units in the dense layers. Figure 1 shows the proposed architecture, named LiteVGG, alongside the traditional VGG. As with the VGG, the total number of convolutional/dense layers can be customized. In this study we chose the 11-layer configuration.

The usage of DWConv layers reduces the number of parameters and computational costs by a factor of $1/N + 1/K$ [20], being N the number of output filters and K being the kernel size. Further reduction on the model complexity is achieved by using only 16 filters in the first block and multiplying this number by a factor of $\sqrt{2}$ after each pooling operation. This reduced multiplying factor results in an effective reduction in the feature maps size, unlike if we doubled the number of filters as in the original VGG. However, most parameters of a VGG model are in the fully connected block. The global pooling and reduced number of units by dense layer resulted in a reduction from 148.9 million parameters to only 6339 in this block.

2.3. Training and evaluation

Among all the available diagnoses in the dataset, we chose to classify the ECG exams into three classes: 1) **NSR**: Exams that contain only normal sinus rhythm. Sinus bradycardia and tachycardia were also included as normal because a patient on continuous rhythm monitoring may enter into these states by doing everyday activities; 2) **AF**: Exams that contain atrial fibrillation diagnostic; 3) **Others**: Exams that do not fit into the previous classes.

The main goal is to classify the AF, which could be used to trigger notifications. But **Other** events may indicate a condition that needs posterior cardiologist analysis.

Our class stratification reduced the data unbalance, but the prevalence of the *AF* class remains low (6%). A Monte Carlo cross-validation with 10 sampling were used to reduce the unbalance during training and tests. Validation and test sets were both balanced by sampling

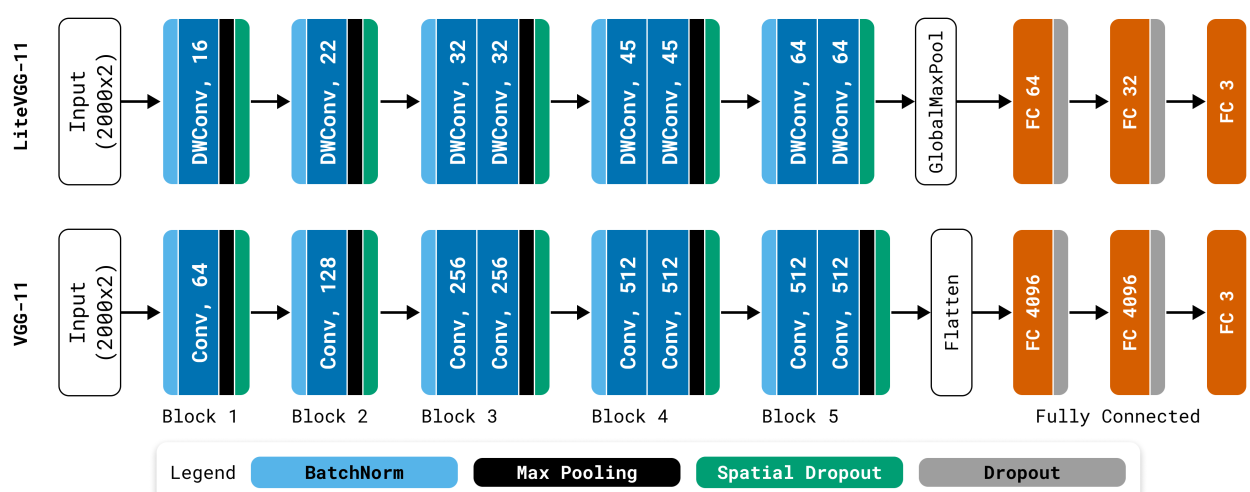

Figure 1. Comparison between the implemented LiteVGG-11 and VGG-11 neural network architectures.

Table 1. Model complexity and computational costs.

Model	# Params	Model size	FLOP	Inference Time[1]	Memory footprint[1]
ResNet-34	16.61M	66.4 MB	3.22G	198.31 ms	128.37 MB
VGG-11	152.00M	608.0 MB	2.25G	259.84 ms	1160.77 MB
LiteVGG-11	20.45k	95.2 kB	6.87M	3.24 ms	1.17 MB
Katsaouni[2]	290	6.7 kB	118.31k	386 µs	87.1 kB

[1] ADFParameters obtained with the TensorFlow Lite benchmark tool on a Samsung Galaxy A01 Core using the XNNPACK delegate with 4 threads. [2] Best-performing model in our evaluation (kernel with size 9).

512 signals for each class. Training sets contained 10,000 ECGs for *NSR* and *Others* classes, but only 4260 for *AF*.

For comparison, we also used ResNet-34 [21] and VGG-11 [19] models, both adapted to unidimensional input. Residual networks (and variants) were very common in the ”Will two do?” challenge [14, 15, 22] and other ECG classification tasks [23, 24]. And as our proposed model is an adaptation of the VGG-11, the original model is used to compare the impact of the reduction in complexity. These models, however, are computationally expensive. Energy-efficient models proposed by Katsaouni [16] were used as an additional comparison.

All models were implemented using TensorFlow. Training was performed over up to 200 epochs with an initial learning rate of 10^{-3}, using the ADAM optimizer and categorical cross entropy loss. The learning rate was reduced by a factor of 0.2 after 30 epochs without a reduction in the validation loss, and the training was stopped early after 50 epochs without a validation loss reduction.

3. Results and discussion

Table 1 compares the computational costs of our models against all the other models. Compared to the original VGG-11, our model requires 99.997% fewer parameters, 99.58% fewer floating point operations (FLOP), and 98.75% less time for inference. Katsaouni's model is even less computationally expensive. It has only 290 parameters and requires 88.09% less time for inference than our model.

Even though our model is not the lightest, it achieved lower computational costs while preserving a classification performance comparable to that of the ResNet-34, as shown in table 2. Katsaouni's models were designed in a way to minimize their requirements but balance the low number of parameters with a recurrent scenario that was not reproduced in this work. LiteVGG has the advantage of high accuracy in an instantaneous rhythm classification using lower computing resources.

This work did not evaluate the models using the challenge's metric because we used different class definitions. The unique class that was not modified in comparison to the original challenge's dataset is the *AF*, enabling some degree of comparison. The challenge winner, ISIBrno-AIMT team, used an ensemble of residual networks approach that achieved 97.1% AUROC and an 83% F1-score when using only the leads I and II [15]. Considering only the AF class (and two lead approaches), the snu_adsl team achieved the best AUROC (98.0%) and F1-score (88.3%) by using representational learning and

Table 2. Cross-validation metrics. The best metrics for each class are highlighted in bold.

Model	Target class	Sensitivity	Specificity	F1-score	AUROC	Accuracy
ResNet-34	AF	87.1 ± 3.5	$\mathbf{93.1 \pm 1.3}$	86.7 ± 1.8	96.0 ± 0.6	80.3 ± 1.0
	NSR	89.2 ± 3.3	$\mathbf{88.8 \pm 1.8}$	84.3 ± 0.7	$\mathbf{95.1 \pm 0.3}$	
	Others	$\mathbf{64.8 \pm 3.8}$	88.6 ± 2.5	$\mathbf{69.0 \pm 1.5}$	$\mathbf{84.1 \pm 1.2}$	
VGG-11	AF	91.6 ± 2.6	90.3 ± 1.5	86.8 ± 1.0	95.2 ± 0.4	78.2 ± 1.2
	NSR	92.7 ± 2.6	84.1 ± 1.9	82.6 ± 0.8	93.6 ± 0.7	
	Others	50.4 ± 3.8	93.0 ± 1.7	61.2 ± 2.8	80.0 ± 1.4	
LiteVGG-11	AF	$\mathbf{93.9 \pm 0.6}$	91.6 ± 1.1	$\mathbf{89.1 \pm 1.1}$	$\mathbf{96.2 \pm 0.5}$	$\mathbf{81.2 \pm 1.1}$
	NSR	$\mathbf{93.2 \pm 1.8}$	86.3 ± 1.1	$\mathbf{84.5 \pm 0.9}$	94.9 ± 0.6	
	Others	56.7 ± 3.3	$\mathbf{94.0 \pm 0.8}$	67.2 ± 2.4	$\mathbf{84.1 \pm 1.4}$	
Katsaouni	AF	83.2 ± 4.7	78.2 ± 5.9	73.5 ± 5.3	86.2 ± 4.3	62.8 ± 4.2
	NSR	75.0 ± 4.7	81.3 ± 2.6	70.6 ± 2.5	86.1 ± 1.8	
	Others	30.2 ± 6.7	84.7 ± 2.3	37.3 ± 7.0	61.1 ± 4.6	

an EfficientNet-B3 model [13]. LiteVGG-11 has a lower AUROC than both models but a higher F1-score.

According to the STM32Cube.AI analysis tool, LiteVGG-11 models can run directly into an Arm Cortex-M4 microcontroller AI compatible (or superior), requiring 80 KiB of flash and 101 KiB of RAM, enabling even an AF detection direct by a wearable ECG.

4. Conclusion

LiteVGG-11 model achieved an AF classification metrics comparable to more robust models while retaining low computing requirements. Our model could be embedded into mobile applications or even directly into a wearable ECG, enabling continuous AF monitoring for patients after cardiac surgeries or for early AF identification.

Acknowledgments

The authors acknowledge the support of Lenovo, under Brazilian Informatics Law, and Zerbini Foundation as part of the research project “Telemonitoramento Remoto Assistido de Arritmias” (TRAdA).

References

[1] Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. International Journal of Stroke feb 2021; 16(2):217–221.

[2] Roth GA, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019. Journal of the American College of Cardiology December 2020;76(25):2982–3021.

[3] Moraes de Oliveira GM, et al. Cardiovascular statistics - brazil 2021. Arquivos Brasileiros De Cardiologia 2022; 118(1):115–372.

[4] Kornej J, Börschel CS, Benjamin EJ, Schnabel RB. Epidemiology of atrial fibrillation in the 21st century: Novel methods and new insights. Circ Res June 2020;127(1):4–20.

[5] de Moraes ERFL, et al. Prevalence of atrial fibrillation and stroke risk assessment based on telemedicine screening tools in a primary healthcare setting. European Journal of Internal Medicine 2019;67:36–41.

[6] Passannante AN. Prevention of atrial fibrillation after cardiac surgery. Current Opinion in Anaesthesiology February 2011;24(1):58–63.

[7] Joung S, et al. Early detection of atrial fibrillation-atrial flutter using remote patient monitoring. Journal of Medical Cases 2019;10(2):31–36.

[8] Mathew JP, et al. A Multicenter Risk Index for Atrial Fibrillation After Cardiac Surgery. JAMA 04 2004; 291(14):1720–1729.

[9] Kaireviciute D, Aidietis A, Lip GY. Atrial fibrillation following cardiac surgery: clinical features and preventative strategies. European Heart Journal 01 2009;30(4):410–425.

[10] Svennberg E, et al. Mass screening for untreated atrial fibrillation. Circulation 2015;131(25):2176–2184.

[11] Aronsson M, et al. Cost-effectiveness of mass screening for untreated atrial fibrillation using intermittent ECG recording. EP Europace 04 2015;17(7):1023–1029. ISSN 1099-5129.

[12] Carpenter A, Frontera A. Smart-watches: a potential challenger to the implantable loop recorder? Europace February 2016;18(6):791–793.

[13] Suh J, et al. Learning ecg representations for multi-label classification of cardiac abnormalities. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[14] Magni S, Sansonetti A, Salvi C, Tabiadon T, García-Isla G. Combining resnet model with handcrafted temporal features for ecg classification with varying number of leads. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[15] Nejedly P, et al. Classification of ecg using ensemble of residual cnns with attention mechanism. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[16] Katsaouni N, et al. Energy efficient convolutional neural networks for arrhythmia detection. Array March 2022; 13:100127.

[17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition, 2014.

[18] Reyna MA, et al. Will two do? varying dimensions in electrocardiography: The physionet/computing in cardiology challenge 2021. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition, 2014.

[20] Howard AG, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.

[21] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition, 2015.

[22] Li X, et al. Towards generalization of cardiac abnormality classification using ecg signal. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[23] Jin Y, et al. A novel domain adaptive residual network for automatic atrial fibrillation detection. Knowledge Based Systems SEP 5 2020;203.

[24] Zhang D, Yang S, Yuan X, Zhang P. Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram. ISCIENCE APR 23 2021;24(4).

Address for correspondence:

Quenaz Bezerra Soares
Instituto do Coração (Incor) - Av.Dr Enéas Carvalho de Aguiar, 44 - Cerqueira César, São Paulo -SP, Brazil
quenaz.soares@incor.usp.br