Automated atrial fibrillation recognition in 12-lead electrocardiographic records: a signal to image and transfer learning approach

A case-control accuracy study

Elena Caires Silveira

Medical Faculty, Federal University of Bahia, Multidisciplinary Institute of Health, Vitória da Conquista, Brazil

Received: May 11, 2021

Revised: July 6, 2021

Accepted: November 4, 2021

Corresponding author:

Elena Caires Silveira
Medical Faculty, Federal
University of Bahia,
Multidisciplinary Institute of
Health, Hormindo Barros Street,
58 Candeias, Vitória da Conquista
45029-09, Brazil
Tel: +55-77-999258012
E-mail: elenacaires@gmail.com

ABSTRACT

Purpose: Atrial fibrillation (AF), the most common among cardiac arrhythmias, is associated with significant morbidity and mortality. For its diagnosis, documentation of the electrocardiographic tracing is required. The use of electrocardiogram has been established as a valuable noninvasive diagnostic tool, and the interpretation of electrocardiographic records using deep learning models has attracted significant attention in recent years. Relying on signal-to-image and transfer learning approaches, this study is aimed at the development of a deep neural network for classifying binary electrocardiographic records according to their rhythm, i.e., normal or AF.

Methods: Electrocardiographic records labeled as normal (n=917) or AF (n=1,097) from the China Physiological Signal Challenge 2018 were collected and used to generate images, which were split into training and test sets and used as inputs to a dense convolutional neural network (DCNN). For the training, transfer learning with a fine tuning of all layers was applied. For a performance evaluation of the test set, the accuracy, sensitivity, specificity, F1-score, and area under the curve (AUC) were used as metrics.

Results: For the test set, the proposed model achieved an accuracy of 99.34%, sensitivity of 98.85%, specificity of 100.00%, F1-score, of 99.42%, and AUC of 0.99.

Conclusion: To validate the methodology, as well as apply it to the multilabel classification of arrhythmia, it is important that further studies adopting this approach be conducted for the detection of AF in larger volumes of data.

Keywords: Arrhythmias, cardia; Artificial intelligence; Atrial fibrillation; Deep learning

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>).

INTRODUCTION

Atrial fibrillation (AF), which affects more than 33 million individuals worldwide and whose incidence and prevalence has grown considerably in recent decades, is the most common among cardiac arrhythmias [1,2]. This condition is associated with significant morbidity and

mortality—being a well-established risk factor for cardiovascular events such as ischemic stroke and heart failure—and imposes a significant burden on health systems globally [3]. Moreover, it is a condition whose diagnosis and management can be challenging, imposing difficulties in terms of its quantification and the measurement of its impact [4].

AF is defined as atrial tachyarrhythmia with uncoordinated atrial electrical activation and consequently ineffective atrial contraction. For the diagnosis of clinical AF, documentation of the electrocardiographic tracing for at least 30 seconds is required. The characteristics of this rhythm disorder shown through an electrocardiogram (ECG) include irregular atrial activations, the absence of repeating P waves, and irregularly irregular intervals between R waves (R-R intervals) [5].

Electrocardiographic signals contain information about the morphology, heart rate, regularity, wave segments, relative amplitudes, wave intervals, and normalized energy of a given heart rhythm [6]. In view of this, an ECG has been established as a valuable noninvasive tool for the identification and classification of cardiac rhythm abnormalities. Considering its clinical relevance and popularity, the interpretation of electrocardiographic records using artificial intelligence techniques, more specifically, deep learning models (which have shown significant potential in the medical field), has attracted attention in recent years [7,8].

For most of the classification models used in electrocardiographic data developed thus far, the use of one-dimensional (1D) ECG signals (that is, 1D time series) has been adopted as inputs to dense convolutional neural networks (DCNNs). Less commonly, such data are converted into images, which are used as two-dimensional (2D) inputs [7,9]. Although not as usual, one should consider the potential of this second approach, which allows as an example the use of pretrained weights on large image sets, whose number of instances usually significantly exceeds the time series datasets.

Relying on the conversion of signals into images and the use of transfer learning, in line with the above, the present work aims at the development of a 2D deep neural network for a binary classification of electrocardiographic records according to their rhythm, i.e., normal or AF.

METHODS

The 12-lead electrocardiographic records used to develop the model were collected from the repository of the China Physiological Signal Challenge 2018 [10], consisting of 9,831 ECG recordings sampled as 500 Hz obtained from 11 hospi-

tals, and for inclusion in the present study were labeled as normal ($n=917$) or AF ($n=1,097$). The records, originally in the format of 12 1D time series (each corresponding to an ECG channel, i.e., I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6) were converted into images, each of which encompasses the signal plots of each channel in a 4×3 arrangement.

The generated images were subdivided at an 85%/15% ratio into training and validation set, respectively, and converted into arrays with dimensions of $349 \times 231 \times 3$. The training set encompassed 789 records labeled as normal and 923 records labeled as AF, whereas the validation set encompassed 128 normal records and 174 AF records. To make them suitable inputs to the proposed architecture, arrays were preprocessed using the densenet function. In addition, 361 records (199 normal and 162 AF) randomly selected from a different dataset, i.e., a 12-lead ECG database for arrhythmia research published by Zheng et al. [11], were used as an external test set.

To build the binary classifier, the DenseNet-201 architecture, a densely connected convolutional network with four dense blocks and a total of 200 convolutional layers, was used. This architecture is based on the presence of dense convolutional blocks (which connect each layer to every other layer in a feed-forward fashion), with convolutional and pooling transition layers between them [12]. The connectivity pattern of the layers of a dense block and an overview of the DenseNet-201 architecture are illustrated in Fig. 1.

The original top layer (a dense layer with sigmoid activation and 1,000 units) was replaced with a dense layer having a sigmoid activation and 1 unit, and is thus adjusted based on the purpose of the binary classification. The output of the last convolutional block was resized from four-dimensions (4D) into 2D by average pooling. The architecture was initialized with weights pretrained on the ImageNet dataset, and proceeded the fine tuning of all layers.

The model was trained for 50 epochs on the training set, which was divided into batches of 4 data. As a regularization measure, a dropout layer was added previously to the final layer, with a rate of 0.2. Finally, the performance of the classifier was evaluated on the test set, with the determination through the following metrics: accuracy, sensitivity, specificity, F1-score, and area under the curve (AUC). All steps of the model development and evaluation were applied using Python version 3.6.9 (Python, Wilmington, DE, USA), applying the Keras library [13].

Because this is a retrospective study and all data used were retrieved from a public, open-source, and anonymized data-

PRECISION AND FUTURE MEDICINE

Automated AF recognition with transfer learning

Fig. 1. Schematic representation of dense blocks and DenseNet-201 architecture. (A) It shows the pattern of dense connections in a 4-layer block, while (B) shows the general structure of the DenseNet-201, with 4 dense blocks and 200 convolutional layers. NC, number of convolutions.

Fig. 2. Receiver operating characteristic curve for atrial fibrillation detection in electrocardiographic records of validation and test sets using the proposed 2D-dense convolutional neural network model. This curve demonstrates the great trade-off between sensitivity and specificity achieved, with an area under the curve (AUC) of 1.00 for validation set (A) and of 0.99 for test set (B).

Table 1. Metrics for model performance in validation and test set

Set	Number	Metrics (%)			
		Accuracy	Sensitivity	Specificity	F1-score
Validation	302	99.67	99.22	100.00	99.61
Test	361	94.45	100.00	89.90	94.19
Total unseen data	663	96.83	99.70	93.88	96.96

The total unseen data represents the sum of both validation and test sets.

set, a review by the institutional review board and written informed consent were waived.

RESULTS

The predictive model developed consisted of a 201-convolu-

tional layer DCNN (feature extractor), whose weights were pretrained on the large ImageNet dataset and fine-tuned using the training set, and connected to a final dense layer output with a sigmoid activation (binary classifier). The training set encompassed 789 records labeled as normal and 923 records labeled as AF.

An Adam optimizer was adopted as the model optimization algorithm, with a learning rate of 0.001, and the loss function used was binary cross entropy. During the training, the classes were weighted inversely proportional to their frequencies.

The accuracy, sensitivity, specificity, and F1-score for the internal validation set (128 normal records and 174 AF records), external test set (199 normal records and 162 AF records), and total unseen data (validation + test sets) are shown in Table 1. The numbers of true positives and true negatives in the validation set were 173 and 128, respectively, and in the test set were 162 and 179. The ROC curves for both sets and their respective AUC values are presented in Fig. 2.

For the records coming from the China Physiological Signal Challenge 2018 dataset (which made up the training set and the internal validation set), a mean age of 71.4 years (± 18.4) was observed for the AF records and 41.6 years (± 12.6) for the normal records. Among the AF patients, there were 476 women and 622 men, whereas among the healthy patients, there were 555 women and 363 men. For the external test set, the AF records comprised 64 women and 98 men, with a mean age of 73.3 years (± 12.2), whereas the normal records comprised 114 women and 85 men, with a mean age of 55.5 years (± 16.6).

DISCUSSION

AF is a major global health problem, associated with severe adverse outcomes, and its burden is expected to increase up to 60% by 2050 [2,14]. Thus, the automation of the diagnostic decision processes related to this arrhythmia has significant potential to contribute both clinically and economically in the coming decades. Consistent with this context, the predictive model based on a pretrained 2D dense neural network proposed in this study was able to classify 12-channel electrocardiographic records according to the presence or absence of AF with an accuracy of 96.83%, sensitivity of 99.70%, and specificity of 93.88% on unseen data.

Previous studies have documented the identification of AF in ECG data using different approaches. For example, Tutuko et al. [15] proposed a 1D-DCNN that is able to detect AF in unseen data, differentiating it from normal recordings, with an accuracy of 98.8%. A multilabel (AF vs. normal vs. other arrhythmia) 1D-DCNN reported an AF detection accuracy of 82% [16]. In another study, a fine-tuned stack sparse autoencoder achieved an accuracy of 98.3% for AF recognition [17]. Xia et al. [18] used short-term Fourier and stationary wavelet trans-

forms to generate 2D inputs for DCNNs, and through this approach they were able to detect AF with 98.6% accuracy.

Based on the results obtained through this study and a comparison with the results obtained in previous research, we demonstrate herein the potential of the signal-to-image approach when combined with deep learning, transfer learning, and a fine-tuning for the analysis of ECG data used in the detection of AF and potentially other types of arrhythmia. In this sense, the generation of 2D data, which is even used in an analysis of a typical ECG interpretation achieved by physicians, given from an analysis of graphical representations of cardiac electrical signals, allows the use of pretrained architectures in large datasets, with a significant contribution in terms of accuracy. Considering that deep learning has not yet been widely used in an ECG analysis owing to a small training collection and the specificity of ECGs [17], this signal-to-image approach may help to expand such use. Another great contribution of this methodology is to dismiss the need for manual feature extraction.

A potential limitation of this study is the fact that records of other types of arrhythmia were not included in the analysis, and only the discrimination between AF and normal ECG was made. However, the validity of this binary approach is understood in the sense of a methodological contribution concerning the investigation of a greater adequacy regarding the interpretation of a certain type of signal entity. Another limitation considered is the relatively small volume of data used, a fact that contrasts with the high quality of the data applied.

Finally, it is important to highlight the importance of testing and validating different algorithmic approaches, particularly in the context of deep learning, for the automated detection of AF, enabling the development of accurate and reliable diagnostic tools. The present study is aligned with this goal. With the development of such systems, a significant contribution to the clinical approach to dealing with this type of arrhythmia and the reduction of the costs related to such an approach, has become feasible particularly when considering that the interpretation of electrocardiographic records is a time-consuming activity, which requires high qualification and practice.

In conclusion, AF is a condition of significant clinical and epidemiological relevance whose diagnosis is established from the interpretation of electrocardiographic recordings. Contributing to the efforts to automate this diagnosis using deep learning systems, in this paper, the integration of signal-to-image conversion and transfer learning with fine tuning approaches is proposed for use in a dense neural network capable of classifying ECG data

as normal or AF with high accuracy and optimal sensitivity.

Further studies adopting this approach must be conducted for the detection of AF in larger volumes of data, thereby validating the methodology, as well as for the multilabel classifications of arrhythmias and other physiological signals from other types of tests.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ORCID

Elena Caires Silveira <https://orcid.org/0000-0003-3470-9205>

AUTHOR CONTRIBUTIONS

Conception or design: ECS.

Acquisition, analysis, or interpretation of data: ECS.

Drafting the work or revising: ECS.

Final approval of the manuscript: ECS.

REFERENCES

1. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation* 2014;129:837-47.
2. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. *Int J Stroke* 2021;16:217-21.
3. Adderley NJ, Nirantharakumar K, Marshall T. Risk of stroke and transient ischaemic attack in patients with a diagnosis of resolved atrial fibrillation: retrospective cohort studies. *BMJ* 2018;361:k1717.
4. Deshpande S, Catanzaro J, Wann S. Atrial fibrillation: prevalence and scope of the problem. *Card Electrophysiol Clin* 2014;6:1-4.
5. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomstrom-Lundqvist C, et al. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J* 2021;42:373-498.
6. de Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans Biomed Eng* 2004;51:1196-206.
7. Weimann K, Conrad TO. Transfer learning for ECG classification. *Sci Rep* 2021;11:5251.
8. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: a review. *Comput Methods Programs Biomed* 2018;161:1-13.
9. Wu Y, Yang F, Liu Y, Zha X, Yuan S. A comparison of 1-D and 2-D deep convolutional neural networks in ECG classification. In: *Proceedings of 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; 2018 Jul 17-21; Honolulu, HI. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2018. p. 324-7.
10. Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *J Med Imaging Health Inform* 2018;8:1368-73.
11. Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Sci Data* 2020;7:48.
12. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017 Jul 21-26; Honolulu, HI. Piscataway (NJ):IEEE; 2017. p. 2261-9.
13. Gulli A, Pal S. *Deep learning with Keras*. Birmingham (UK): Packt Publishing; 2017.
14. Korucuk N, Polat C, Gunduz ES, Karaman O, Tosun V, Onac M, et al. Estimation of atrial fibrillation from lead-I ECGs: Comparison with cardiologists and machine learning model (CurAlive), a clinical validation study. *ArXiv* 2021 Apr 15. <https://arxiv.org/abs/2104.07427>.
15. Tutuko B, Nurmaini S, Tondas AE, Rachmatullah MN, Darmawahyuni A, Esafri R, et al. AFibNet: an implementation of atrial fibrillation detection with convolutional neural network. *BMC Med Inform Decis Mak* 2021;21:216.
16. Xiong Z, Stiles MK, Zhao J. Robust ECG signal classification for detection of atrial fibrillation using a novel neural network. In: *2017 Computing in Cardiology (CinC)*; 2017 Sep 24-27; Rennes, FR. Piscataway (NJ): IEEE; 2017. p. 1-4.
17. Yuan C, Yan Y, Zhou L, Bai J, Wang L. Automated atrial

- fibrillation detection based on deep learning network. In: 2016 IEEE International Conference on Information and Automation (ICIA); 2016 Dec 16-19; Galle, LK. Piscataway (NJ): IEEE; 2016. p. 1159-64.
18. Xia Y, Wulan N, Wang K, Zhang H. Detecting atrial fibrillation by deep convolutional neural networks. *Comput Biol Med* 2018;93:84-92.