

CARDIOVASCULAR DISEASES

Artificial intelligence for the electrocardiogram

Deep-learning algorithms can be applied to large datasets of electrocardiograms, are capable of identifying abnormal heart rhythms and mechanical dysfunction, and could aid healthcare decisions.

Ana Mincholé and Blanca Rodriguez

Artificial intelligence (AI) is expected to revolutionize many aspects of society through its ability to optimize processes and decisions using computer-based algorithms. Healthcare innovation is an area of great promise for AI, with enormous potential socioeconomic impact. In the last 15 years, two main advances have laid the foundations for a boom in AI in healthcare innovation. Firstly, databases and electronic health records have started to systematically collect integrated digitalized medical data for large numbers of patients and asymptomatic individuals¹. Secondly, powerful computing platforms based on high performance computing and cloud computing have become available through hardware improvements, particularly in graphics processing units.

Machine learning is a field of AI and is based on computational statistical algorithms that allow computers to learn directly from data, without being explicitly programmed. Thus, for example, machine-learning techniques have the potential to automatically identify the most important features related to key differences in patient data, that is, disease versus healthy. The potential applications of machine learning in healthcare are vast, including screening, disease detection and classification, patient risk stratification, and optimal therapy selection (Fig. 1). In this issue of *Nature Medicine* two studies^{2,3} demonstrate the power of machine learning applied to cardiology. Both studies describe deep-learning algorithms applied to large datasets of electrocardiograms (ECGs), the most widely used and perhaps simplest recordings in the clinic, and indicate that machine learning can be applied to identify heart rhythm abnormalities and mechanical dysfunction.

Deep learning, a type of machine learning inspired by how the brain is connected and works, is based on algorithms, also known as models, consisting of interconnected nodes called neurons arranged in a network layout. Neurons in the deep-learning network are activated according to the data inputs, which were ECGs in these studies^{2,3}. Their

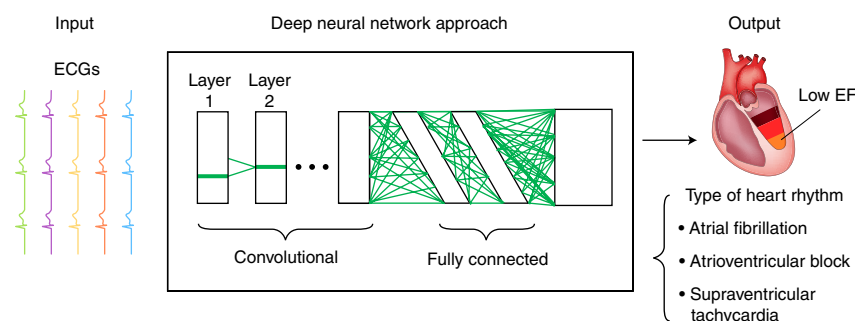


Fig. 1 | A deep neural network approach for analyzing electrocardiograms. Hannun et al.² and Attia et al.³ employ deep neural network approaches to classify heart rhythms and impaired left ventricular systolic function from electrocardiogram.

activation propagates through the network, providing the ability to learn the relationship between input and the output data (such as patient or rhythm classification) through a process called empirical learning. Once the neural network is trained, its performance is evaluated on an independent dataset that was not used for training, and statistical metrics such as sensitivity, specificity, and accuracy are calculated.

The focus of the study by Hannun et al.² is cardiac rhythm classification using a deep neural network built on a large clinical dataset of 91,232 single-lead ambulatory ECG recordings from 53,877 patients. The authors' goal was the identification of 12 classes of cardiac rhythm from the ECG data. Validation was performed against an independent test dataset annotated by a certified board of cardiologists. The algorithm showed excellent performance in its ability to identify individual rhythm classes. The authors assessed the accuracy of the results using the average area under the receiver operating characteristic curve (AUC), which would be 1 for perfect rhythm classification and 0.5 for a random classification of rhythms, and achieved an impressive AUC of over 0.91 for each of the individual rhythm classes. The algorithm outperformed all individual cardiologists who conducted the classification by eye before reaching consensus with additional colleagues.

Automatic detection of cardiac rhythms (such as that shown in ref. ²) is becoming increasingly important with the emergence of wearable devices, which allow continuous individual monitoring and require robust, rapid, real-time analysis of ECG signals to avoid storage of large amounts of data. An example is KardiaBand from AliveCor⁴, a smartphone application based on machine learning for the identification of atrial fibrillation episodes from ECG. These technologies come with substantial technical, ethical, and clinical challenges.

Rhythm classification has also been the focus of previous machine-learning studies using a range of techniques such as linear and quadratic discriminants, support vector machines, and random forests and neural networks, as explained in a recent review⁵. Most of these studies relied on the derivation of ECG-based features able to classify different beat types as well as on techniques such as feature selection for better generalization and performance properties. A clear advantage of the deep-learning approaches presented here is that they do not require extraction of ECG features, as these are learned from the data as part of the algorithm training. Furthermore, even if accuracies in heart-rhythm classification presented in previous studies are comparable to that from Hannun et al.², these previous studies included smaller datasets and considered a smaller variety of rhythms⁵.

Directly contrasting with the study by Hannun et al.², only very few studies have attempted ECG classification using deep learning. For example, in one study⁶, the average detection accuracies of both ventricular and supraventricular ectopic beats were over 97% using the MIT-BIH arrhythmia database⁷ for both training and testing of the algorithm, but data in this set were obtained from only 47 different individuals.

The study by Attia et al.³ aimed to identify patients with impaired left ventricular (LV) systolic function (ejection fraction (EF) < 35%, a measurement of the amount of blood leaving the heart each time it contracts) using a deep-learning approach known as a convolutional neural network, which was applied to a large dataset of ECGs measured in 12 leads (which are electrodes positioned on the patient torso). The authors hoped to capitalize on the ECG as an inexpensive and widely available test to identify patients with asymptomatic LV dysfunction, a common condition that is treatable when diagnosed. Twelve-lead ECGs and echocardiograms from 44,959 patients were available to train the network, which was then tested on data from more than 52,870 patients. Results showed an AUC of 0.93 for diagnosis of mechanical dysfunction from the ECG data, representing a clear improvement with respect to current B-type natriuretic peptide (BNP) screening blood tests, which have an AUC of 0.79–0.89. However, 1,335 patients were assigned to have an abnormal cardiac pump function by the algorithm when in fact their EF was normal. The authors suggest that these false positives may represent an early detection of abnormalities, although the 5-year incidence of one episode of abnormal EF in those patients was only 9.5%.

These two studies^{2,3} demonstrate that deep-learning algorithms are powerful and accurate tools for patient screening. They show the potential to advance patient stratification by capitalizing on population-based information. A known limitation of current machine-learning methods is that it is challenging to understand the rationale behind their results. The algorithms are not able to provide explanations for the pathophysiological basis of classification outcomes, as they are unable to reveal the functional dependencies between data inputs and classes. To this end, cardiologists generally use information on well-known ECG properties, such as the QRS width, the QT interval, or the T wave morphology, as these are directly linked to the contraction and the relaxation of the heart and have been investigated using human-based computational modeling and simulation⁸. However, extraction of ECG biomarkers requires the delineation of the ECG, i.e., the identification of key points of the ECG waveforms for which expert system techniques are still the state of the art⁹, even though deep-learning attempts are showing promise¹⁰. In contrast to the established success of deep-learning approaches for image segmentation¹¹, deep learning for ECG delineation remains a challenge. Thus, one of the limitations of deep-learning analyses is that they are not able to reliably and automatically yield key physiologically meaningful ECG biomarkers from the large datasets that they analyze.

As with other deep-learning applications, the main challenge for ECG analysis is not necessarily computational but the availability of digitalized large-scale datasets that are annotated with the required information. Expert consensus,

as shown in ref. ², is also key, as individual expert annotations often differ among different cardiologists^{2,10}. In addition to collaborations between computational scientists and clinicians, the availability of freely available databases (as in Physionet¹² for the ECG) is an important catalyst for advancements in AI and data science for healthcare innovation, such as those described in the two studies presented in this issue^{2,3}. □

Ana Mincholé and Blanca Rodríguez*

Department of Computer Science, British Heart Foundation Centre of Research Excellence, University of Oxford, Oxford, UK.

*e-mail: blanca@cs.ox.ac.uk

Published online: 7 January 2019

<https://doi.org/10.1038/s41591-018-0306-1>

References

1. Bycroft, C. et al. *Nature* **562**, 203–209 (2018).
2. Hannun, A. Y. et al. *Nat. Med.* <https://doi.org/10.1038/s41591-018-0268-3> (2019).
3. Attia, Z. I. et al. *Nat. Med.* <https://doi.org/10.1038/s41591-018-0240-2> (2019).
4. Lau, J. K. et al. *Int. J. Cardiol.* **165**, 193–194 (2013).
5. Lyon, A., Mincholé, A., Martínez, J. P., Laguna, P. & Rodríguez, B. *J. R. Soc. Interface* **15**, 20170821 (2018).
6. Kiranyaz, S., Ince, T. & Gabbouj, M. *IEEE Trans. Biomed. Eng.* **63**, 664–675 (2016).
7. Moody, G. B. & Mark, R. G. *IEEE Eng. Med. Biol. Mag.* **20**, 45–50 (2001).
8. Lyon, A., Bueno-Orovio, A., Zacur, E., Ariga, R., Grau, V., Neubauer, S., Watkins, H., Rodríguez, B. & Mincholé, A. *Europace* **20**, iii102–iii112 (2018).
9. Martínez, J. P., Almeida, R., Olmos, S., Rocha, A. P. & Laguna, P. *IEEE Trans. Biomed. Eng.* **51**, 570–581 (2004).
10. Camps, J., McCarthy, A., Rodríguez, B. & Mincholé, A. Deep learning based QRS multilead delineator in electrocardiogram signals. In *Proc. 3rd International Workshop on Biomedical Informatics with Optimization and Machine Learning* (2018).
11. Bai, W. et al. *J. Cardiovasc. Magn. Reson.* **20**, 65 (2018).
12. Goldberger, A. L. et al. *Circulation* **101**, E215–E220 (2000).

Competing interests

The authors declare no competing interests.