

A deep neural network trained to interpret results from electrocardiograms: better than physicians?



Heart function depends on precisely orchestrated cell membrane depolarisations of billions of heart muscle cells that elicit voltage differences on the body surface—recordable by an electrocardiogram (ECG).

Pathological cardiac conditions result in pathognomonic changes of the ECG waveforms.¹ Although single-channel ECGs, which record voltage differences between two points on the body surface, are often sufficient to diagnose rhythm disorders, the precise diagnosis of other abnormalities, such as conduction disorders or ischaemia-related abnormalities, frequently requires the consideration of different leads recorded from different points of the body surface. A standard ECG typically consists of 12 channels.

Correct interpretation of ECG recordings is a non-trivial task, requiring both knowledge and experience. First attempts to use computers to aid this interpretation date back to the 1960s.² Although analysis algorithms are implemented in many state-of-the-art ECG recorders, the quality of these algorithms needs to be improved,³ and physicians should not base therapeutic decisions solely on them.

Deep neural networks could improve the quality of automatic ECG interpretation. In 2019, Hannun and colleagues⁴ showed that a deep neural network trained to do a limited task within the wide field of ECG interpretation—arrhythmia detection in single-channel ECGs—was more accurate than individual board-certified cardiologists.

In *The Lancet Digital Health*, Hongling Zhu and colleagues⁵ report the results of a similar study in which they widened the approach to diagnose an array of different rhythm or conduction disorders from recordings of 12-channel ECGs. Zhu and colleagues' study used a large training dataset of 10 s ECGs (180 112 recordings from 70 692 patients) and 20 different arrhythmias or conduction disorders (eg, atrial fibrillation, atrioventricular block) were annotated by trained physicians; 10% of the ECGs in this dataset contained more than one abnormality.

A convolutional deep neural network was developed that accepts the raw ECG data as its input and provides an output vector of length 21, its elements representing

normal rhythm and the 20 types of arrhythmias and conduction defects. After the deep neural network was trained on the training dataset, it was validated on an independent test dataset of ECGs from 828 patients that had been annotated by a panel of three cardiologists; 24% of the ECGs in this dataset contained more than one abnormality.

The deep neural network correctly identified all abnormalities in 658 (80%) of the 828 ECGs in the test dataset. To compare the network's accuracy with that of human physicians, the test dataset ECGs were also evaluated by 53 physicians. The mean accuracy of the physicians (70% across all experience levels) was worse than that of the deep neural network. Even physicians with more than 12 years of experience of working with ECGs were less accurate (accurate in 621 [75%] of 828 cases) than the deep neural network.

Although Zhu and colleagues' study suggests that ECG interpretation by deep neural networks could become a diagnostic standard in the future, it also has some important limitations. First, all ECGs in the training and test dataset were recorded in China, which makes it difficult to predict the accuracy of the network for interpreting ECG results from patients of different ethnicities.

Second, there is no objective gold standard for ECG interpretation. Although a consensus decision of a panel of three cardiologists seems to be a sensible choice, it is difficult to judge the level of experience of the 53 physicians who rated the ECGs in the test dataset compared with the three panellists. In the work by Hannun and colleagues,⁴ nine experienced cardiologists were divided into three panels, each of which annotated one-third of the test dataset to generate the gold standard. The ECGs annotated by each panel were then individually interpreted by the six cardiologists from the other two panels to evaluate the accuracy of the deep neural network compared with individual experienced cardiologists and compared with the panel gold standard. This approach might reduce the risk of overoptimistic rating of the deep neural network's accuracy due to a low experience level of the human ECG readers.

Published Online
June 4, 2020
[https://doi.org/10.1016/S2589-7500\(20\)30136-9](https://doi.org/10.1016/S2589-7500(20)30136-9)
See [Articles](#) page e348

Finally, the analysis of rhythm and conduction disorders is only one component of ECG interpretation. Other important aspects such as the evaluation of ischaemia-related abnormalities were not part of the study.

Despite these limitations, the work by Zhu and colleagues⁵ provides important evidence that even complex tasks in ECG interpretation such as the identification of multiple concomitant rhythm and conduction abnormalities can be done by deep neural networks with a high level of accuracy, suggesting that computers might eventually outperform experienced physicians in the interpretation of difficult ECGs.

I declare no competing interests.

Copyright © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

Daniel Sinnecker
sinnecker@tum.de

Department of Cardiology, University Hospital Klinikum Rechts der Isar, Technical University of Munich, Munich 81675, Germany

- 1 Fisch C. Centennial of the string galvanometer and the electrocardiogram. *J Am Coll Cardiol* 2000; **36**: 1737–45.
- 2 Pipberger HV, Arms RJ, Stallmann FW. Automatic screening of normal and abnormal electrocardiograms by means of digital electronic computer. *Proc Soc Exp Biol Med* 1961; **106**: 130–32.
- 3 Hwan Bae M, Hoon Lee J, Heon Yang D, et al. Erroneous computer electrocardiogram interpretation of atrial fibrillation and its clinical consequences. *Clin Cardiol* 2012; **35**: 348–53.
- 4 Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019; **25**: 65–69.
- 5 Zhu H, Cheng C, Yin H, et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. *Lancet Digital Health* 2020; published online June 4. [https://doi.org/10.1016/S2589-7500\(20\)30107-2](https://doi.org/10.1016/S2589-7500(20)30107-2).