Deep learning using electrocardiographs

An electrocardiograph (ECG) is a low-cost and widely available test used to measure the electrical activity of the heart to help diagnose cardiac conditions. Rich in prognostic data, millions of ECGs are recorded annually and they have fuelled artificial intelligence (AI) algorithms to identify heart conditions, including atrial fibrillation, and impaired left ventricular systolic dysfunction for which a clinical trial is in progress. The Lancet Digital Health reports two studies using ECG data, which demonstrate that AI models are powerful tools for patient screening for cardiovascular conditions and diseases beyond the heart.

Hongling Zhu, Cheng Cheng, and colleagues apply a deep learning algorithm to ECGs obtained in Tongji Hospital, China, to diagnose 20 different arrhythmias or conduction heart disorders. The model exceeds the performance of 53 physicians clinically trained in ECG interpretation and the authors show the model outperforms a comparable deep learning tool described by Hannun and colleagues. This algorithm was able to accurately diagnose multiple heart conditions from complex ECG data, indicating that computers may aid clinical decision-making and potentially provide care for those with limited access to cardiologists.

Joon-myoung Kwon, Younghoon Cho, and colleagues apply a deep learning algorithm using ECGs and demographic information obtained in hospitals in South Korea. This algorithm demonstrated good performance in diagnosing anaemia in patients, a task beyond the abilities of physicians. While previous algorithms for predicting cardiovascular events have been reported using unexpected data types such as body CT scans or fundoscopic images, this is the first study, to our knowledge, to show ECG data can be applied in deep learning algorithms to diagnose systemic disease. The authors developed a sensitivity map to identify regions of the ECG which may be most critical to diagnose anaemia. In developing a transparent model, the authors hope that this study will aid further research to understand the potential relationship between electrophysiology and systemic diseases.

These studies used ECGs recorded in health-care facilities; however, recent advances have enabled ECG recordings using consumer wearable devices, which could help screen broader populations. For example, two clinical trials to test the efficacy of wearables to detect cardiovascular risks for use in monitoring and diagnosing patients with COVID-19 infection are underway. Continuous monitoring and rapid, real-time analysis of ECG data using wearable devices presents an exciting prospect for generating the much-needed digitalised large-scale datasets. However, a recent study published in Nature Medicine by Han and colleagues showed that common perturbations to single-lead ECG recordings, imperceptible to the human eye, could disrupt deep learning algorithms causing a misdiagnosis rate of 74%. These findings question the safety of using deep learning to analyse millions of ECG recordings generated in clinics or by consumer wearable devices.

It is seldom known what influences AI decision making, and this could cause problems when the tools are used in the clinic. Han and colleagues highlight adversarial examples (ie, intentional inputs to AI models that cause mistakes) could impact the robustness of medical devices that rely on ECGs and cause intentional bias in clinical trials. In a non-peer-reviewed study, adversarial examples were introduced to three AI tools designed for scanning medical images causing a misclassification of the images by altering just a few pixels. To overcome adversarial examples, it is crucial to understand the limitations and vulnerabilities of the deep learning algorithms. Han and colleagues highlight that ECG deep learning models should generalise well to new data, undergo adversarial training to protect against adversarial examples and receive certification for robustness with mathematical proofs as performed in other industries (eg, aviation).

The studies published in The Lancet Digital Health include data from multiple centres and are externally validated to show the models can generalise well to new data. However, like many deep learning tools, ethnic or regional diversity was not considered in their development. Generalisation to adversarial examples may require further validation of these models in different environments, countries, and devices to ensure safe deployment across populations and to protect patients from unintended harm. ■ The Lancet Digital Health