Automated Deep Learning Based Digitization and Classification of Paper Electrocardiograms

Haobo Zhu, Mohammad Atwany, Alexander Sharp, Abhirup Banerjee*

Department of Engineering Science, University of Oxford, Oxford, UK

Aims: For over 100 years, the electrocardiogram (ECG) has been a fundamental investigatory tool in the management of cardiovascular disease (CVD). Digitization and classification of these traditionally paper-based records is vital to: i) retrospectively understand the evolution of CVD within various populations; ii) prospectively improve global accessibility to high quality care. Whilst significant strives have been made in classification of extracted ECG signals, this is limited by a paucity of work exploring ECG digitization.

Methods: Our approach seeks to address this issue by employing novel deep learning (DL) methods classify ECGs from images of paper records. Starting with the baseline random forest model trained on all synthetic images in the PTB-XL dataset, we applied the ResNet-18, ResNet-50, and Swin Transformer models, directly classifying the images. We trained the DL models on 987 images in directory '00000' of the PTB-XL and tested them on all images in the other 21 directories (total 19825 images).

Results: Our result during the unofficial phase of the Challenge (as the 1st of 5 entries) for classification F-measure was 0.528 (MultiMeDIA_OX; rank 22), using the random forest model; we achieved the baseline digitization performance with reconstruction signal-to-noise ratio (SNR) of -18.12 (rank 51). In our analysis using three DL models, Swin Transformer achieved the highest F-measure score of 0.793, while ResNet-18 and ResNet-50 attained F-measure scores of 0.781 and 0.779, respectively. The recall by Swin Transformer far exceeds the other two models (0.832 vs 0.795 and 0.777).

Conclusions: Swin Transformer has shown excellent potential in classifying ECG images, providing the foundation for a transformer-based image classification model. Our future developments will focus on automated ECG digitization using a novel automated pipeline involving removal of gridlines, extraction of single-leads, and 1-dimensional signal generation, for subsequent classification using transformer-based models.

Table 1: Performance of different models on test dataset (19825 images).

Model	F-measure	Accuracy	AUROC	Precision	Recall
ResNet-18	0.781	0.808	0.888	0.768	0.795
ResNet-50	0.779	0.809	0.887	**0.780**	0.777
Swin	**0.793**	**0.813**	**0.891**	0.759	**0.832**