Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge 2024

Matthew A Reyna[1], Deepanshi[1], James Weigle[1], Zuzana Koscova[1], Andoni Elola[2], Salman Seyedi[1], Kiersten Campbell[1], Gari D Clifford[1,3], Reza Sameni[1,3]

[1]Department of Biomedical Informatics, Emory University, USA
[2]Department of Electronic Technology, University of the Basque Country UPV/EHU, Spain
[3]Department of Biomedical Engineering, Georgia Institute of Technology, USA

Abstract

[The abstract is limited to 25 lines.]

The George B. Moody PhysioNet Challenge 2024 invited teams to develop algorithmic approaches for digitizing and classifying electrocardiograms (ECGs) from images or scanned paper printouts.

Paper ECGs have existed for decades, capturing the variability and evolution of cardiovascular diseases (CVDs) across demographics, geography, and time. Paper ECGs and ECG images remain common in cardiac care. However, ECG-based classification algorithms typically require digital time-series representations of ECG data, so existing algorithms cannot classify them, and new algorithms cannot learn from them. Therefore, digitizing ECG images is important for improving the accessibility and quality of cardiac care.

To support this goal, the Challenge also introduced a synthetic ECG image generator with various realistic distortions, such as wrinkles, creases, shadows, rotations, and handwriting, to allow teams to create arbitrary large and diverse training sets for creating generalizable approaches. To date, several dozen teams have participated in the Challenge, representing diverse approaches from both academia and industry worldwide.

[The manuscript is limited to 4 pages, including everything, i.e., the title, authors and affiliations, abstract, text, figures, tables, and references.]

1. Introduction

The electrocardiogram (ECG) is an accessible, non-invasive pre-screening tool for cardiovascular diseases (CVDs). Invented by Willem Einthoven in 1895, the ECG has evolved significantly, with General Electric introducing portable devices in 1927 and paper-printing ECGs by 1948 [1]. Modern advances include digital ECG devices and algorithmic interpretation, which have improved accessibility to CVD-based diagnosis.

Despite the rise of digital ECGs, paper ECGs still exist and remain prevalent, especially in the Global South [2]. These ECGs reflects the diversity and evolution of CVDs across demographics, geography, and time. However, ECG diagnosis algorithms generally expect ECG time-series instead of images, limiting the utility of the paper ECGs. Moreover, images of paper ECGs often have distortions and other artifacts, such as creases, tears, fading ink, and stains on the paper as well as shadows, skewing, and blurriness from the images.

Therefore, digitizing ECGs to provide ECG time-series data and providing low-cost interpretation to aid ECG-based diagnosis are vital for improving global cardiac care access. The 2024 Challenge invites teams to digitize and classify paper ECGs.

2. Methods

Algorithms for digitizing and classifying ECG images typically apply classical image processing and, more recently, deep learning techniques. Some approaches attempt to digitize the images and use the extracted time series for classification, and other approaches attempt to classify the images directly without using the underlying time series.

Classical image processing techniques include grayscale thresholding for grid removal, pixel scanning for ECG digitization, and template-based optical character recognition (OCR) for patient data extraction [3]; heuristics derived from pixel intensities for region-of-interest identification [4]; the Hough transform for skew correction, color-based segmentation for grid removal, and median filtering for noise removal [5]. Deep learning techniques include a dense neural network for grid removal [6] and U-Net architecture for segmentation [7].

Deep learning methods have the potential to be less sensitive than classical imaging processing approaches to paper distortions and image noise and artifacts. However, these methods are limited by a lack of diverse noise artifacts in training datasets and a scarcity of ground truth ECG time-series data.

2.1. Challenge Data

The Challenge data include data from multiple sources, including public and private databases of ECG waveforms, ECG images, and/or ECG-based diagnoses or labels.

The public training set contains waveforms and labels from the PTB-XL dataset [8, 9], which has 21,799 12-lead ECG recordings, with ECG images. The hidden validation set contains a subset of the waveforms and labels from the PTB-XL dataset with ECG images that we printed and scanned. The hidden validation set contains a subset of the waveforms and labels from the PTB-XL dataset and a separate private source of 12-lead ECG records that we printed and photographed.

The ECG waveforms are standard 12-lead ECGs that are 10 seconds long with sampling frequencies of either 250 Hz or 500 Hz. We encoded the ECG waveforms in a WFDB format using 16-bits of signal quantitization.

We generated ECG images using ECG-Image-Kit [10, 11]. This package also allows the generation of ECG images with synthetic artifacts that resemble the real-world artifacts in the validation and test sets. Teams could include this code or other code in their entries to augment the training set to improve the robustness of their model.

Each ECG record has one or more labels from the following classes: (1) acute myocardial infarction, (2) atrial fibrillation or atrial flutter, (3) bradycardia, (4) conduction disturbances, (5) hypertrophy, (6) normal ECG, (7) old myocardial infarction, (8) premature atrial complex, (9) premature ventricular complex, (10) ST/T changes, and (11) tachycardia.

The labels for records from the PTB-XL database are taken directly from the PTB-XL database, but we used the the 12SL statement codes from the PTB-XL+ database to define separate acute MI and old MI classes [8, 9]. The labels for records from the separate database are derived directly from 12SL statement codes and matched to the above classes.

2.2. Challenge Objective

For the 2024 Challenge, we asked participants to design and implement open-source algorithms that could digitize the ECG and/or classify paper ECGs:

1. Digitize the ECGs, i.e., turn images of ECGs (scanned from paper) into waveforms (time-series data) representing the same ECGs;
2. Classify the ECGs (either from the image, or from the converted time-series data that you extract from the image).

Teams could complete either or both tasks. The winners of each of the two parts of the Challenge are the teams whose algorithms achieved the highest performance on the hidden test set.

2.2.1. Challenge Timeline

This year's Challenge was the 25$^{\text{th}}$ George B. Moody PhysioNet Challenge [12]. As in previous years, the Challenge had an unofficial phase and an official phase. The unofficial phase (25 January 2024 to 10 April 2024) introduced the teams to the Challenge. We publicly shared the Challenge objective, training data, example algorithms, and evaluation metric and invited the teams to submit their code for evaluation, scoring at most five entries from each team on the hidden validation set. Between the unofficial and official phases, we took a hiatus (11 April 2024 to 23 May 2024) to improve the Challenge. The official phase (24 May 2024 to 19 August 2024) continued the Challenge. We updated the Challenge data, example algorithms, and evaluation metric and again invited teams to submit their code for evaluation, scoring at most ten entries from each team on the hidden validation set.

After the end of the official phase, each team chose a single entry from their team for us to evaluate on the test set. The winners of the Challenge were the teams with the best scores on the test set. We announced the results at the end of the Computing in Cardiology (CinC) 2024 conference, where the teams presented, defended, and published their work. Only teams that presented and published their work at the conference were eligible for rankings and prizes. We will publicly release the algorithms after the end of the Challenge and the publication of these papers.

The Challenge Organizers also held a hackathon at Data Science Africa in Nyeri, Kenya, from 3 June 2024 to 6 June 2024 and at CinC 2024 on 8 September 2024.

2.2.2. Challenge Evaluation

The evaluation metric for the ECG digitization task is the signal-to-noise ratio (SNR) of the reconstructed signal. Let $x = (x_i)_{i=1}^n$ be a signal in an ECG image, and let $y = (y_i)_{i=1}^n$ be a signal digitized from the ECG image. Since small horizontal and vertical translations are common but typically do not affect the interpretation of an ECG signal, we shifted y to maximize its cross-correlation with x as long as the shift was no more than ±0.5 s and/or ±1 mV cross-correlation. We then computed

$$\text{SNR} = 10 \log_{10} \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n x_i^2}. \tag{1}$$

Higher SNR values are better, indicating that the model outputs better capture the ECG waveform with less noise.

We computed the mean of the SNR values across all records and all channels in each record. The team with the highest mean SNR across the test set images wins the digitization task.

The evaluation metric for the ECG classification task is the macro F-measure. For each record, the ground truth labels and the classifier's labels contain one or more of the 11 possible classes in 2.1. For each class, we computed the per-class F-measure by comparing the ground truth and classifier labels in a one-vs.-rest manner for all records in a database. Higher F-measure values are better, indicating that the model better classifies the ECG image for the class. The macro F-measure is the mean of the per-class F-measures for all classes and all records in a database. The team with the highest macro F-measure on the test set wins the classification task.

3. Challenge Results

The 2024 Challenge is currently ongoing. We will share results from the Challenge after the Computing in Cardiology 2024 conference.

4. Discussion

The 2024 Challenge is currently ongoing. We will share a discussion about the Challenge after the Computing in Cardiology 2024 conference.

5. Conclusions

The 2024 Challenge is currently ongoing. We will share conclusions about the Challenge after the Computing in Cardiology 2024 conference.

Acknowledgements

This research is supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB: R01EB030362); the National Center for Advancing Translational Sciences of the National Institutes of Health (NCATS: UL1TR002378); as well as AliveCor, Amazon Web Services, IEEE SPS, and MathWorks under unrestricted gifts. GDC has financial interests in AliveCor and Mindchild Medical and holds a board position in Mindchild Medical. None of these entities influenced the design of or provided data for this year's Challenge. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the above entities.

References

[1] Yang XL, Liu GZ, Tong YH, Yan H, Xu Z, Chen Q, et al. The history, hotspots, and trends of electrocardiogram. Journal of Geriatric Cardiology 09 2015;12:448–56.

[2] Tison GH, Zhang J, Delling FN, Deo RC. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. Circulation Cardiovascular Quality and Outcomes 2019;12:e005289.

[3] Ravichandran L, Harless C, Shah AJ, Wick CA, Mcclellan JH, Tridandapani S. Novel Tool for Complete Digitization of Paper Electrocardiography Data. IEEE Journal of Translational Engineering in Health and Medicine 2013; 1:1800107 –1800107.

[4] Baydoun M, Safatly L, Abou Hassan OK, Ghaziri H, El Hajj A, Isma'eel H. High precision digitization of paper-based ECG records: a step toward machine learning. IEEE Journal of Translational Engineering in Health and Medicine 2019;7:1–8.

[5] Garg DK, Thakur D, Sharma S, Bhardwaj S. ECG paper records digitization through image processing techniques. International Journal of Computer Applications 2012;48(13):35–38.

[6] Mishra S, Khatwani G, Patil R, Sapariya D, Shah V, Parmar D, et al. ECG paper record digitization and diagnosis using deep learning. Journal of Medical and Biological Engineering 2021;41(4):422–432.

[7] Li Y, Qu Q, Wang M, Yu L, Wang J, Shen L, et al. Deep learning for digitizing highly noisy paper-based ECG records. Computers in Biology and Medicine 2020; 127:104077.

[8] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data 2020;7:154.

[9] Strodthoff N, Mehari T, Nagel C, Aston PJ, Sundar A, Graff C, et al. PTB-XL+, a comprehensive electrocardiographic feature dataset. Scientific Data 2023;10:279.

[10] Shivashankara KK, Deepanshi, Shervedani AM, Reyna MA, Clifford GD, Sameni R. ECG-Image-Kit: a synthetic image generation toolbox to facilitate deep learning-based electrocardiogram digitization. Physiological Measurement May 2024;45(5):055019.

[11] Deepanshi, Shivashankara KK, Clifford GD, Reyna MA, Sameni R. ECG-Image-Kit: A Toolkit for Synthesis, Analysis, and Digitization of Electrocardiogram Images, January 2024. Online at: https://github.com/alphanumericslab/ecg-image-kit.

[12] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

Address for correspondence:

Matthew A Reyna
DBMI, 101 Woodruff Circle, 4th Floor East, Atlanta, GA 30322
matthew.a.reyna@emory.edu