VinDigitizer: An Image Processing Approach to Digitize Paper ECG Records

Cuong V Nguyen[1], Hieu X Nguyen[1], Nhat A Duong[1], Cuong D Do[1,2]

[1] College of Engineering and Computer Science, VinUniversity, Hanoi, Vietnam
[2] VinUni-Illinois Smart Health Center, VinUniversity, Hanoi, Vietnam

Abstract

The dominance of traditional paper-based ECG records in clinical settings poses a barrier to integrating AI for cardiovascular disease detection. The shift towards digital ECG is imperative yet hindered by the lack of reliable digitization tools. For the PhysioNet Challenge 2024, our ViNUniTeam proposes a hybrid digitization approach merging Deep Learning and image processing algorithms. Our methodology comprises three key phases: lead segmentation, signal extraction, and scale conversion. Initially, we employ a pre-trained YOLOv8 single-shot detector to capture snapshots of 12 ECG leads. Then, each snapshot undergoes Otsu's technique for RGB to binary conversion, Hough transform for grid detection, and Viterbi's algorithm for sampling and extracting signal values in pixels. Finally, we convert these pixel values to voltage values by estimating grid resolution within the snapshot. To address uneven lengths in the extracted lead signals, we use cubic spline interpolation to horizontally scale the signals to align with a predetermined target length while preserving their waveform. Our method ranked 9/16 *in the official phase, with SNR of -0.066.*

1. Introduction

The electrocardiogram (ECG) is a vital tool for pre-screening cardiovascular diseases (CVDs), one of the leading causes of death worldwide. In recent years, algorithmic methods for interpreting ECGs have emerged, enhancing the accessibility of ECG-based CVD diagnosis. Those methods, especially deep learning (DL) based [1–5], relied heavily on digital time-series ECG data. However, paper-based ECGs have been standard for nearly a century and remain prevalent, creating barriers to accessing data and developing automatic diagnosis tools. Worldwide, over 300 million ECGs are performed annually; in low-income and middle-income countries, accurate diagnosis is particularly challenging due to the scarcity of experienced cardiologists [6]. Therefore, digitizing paper-based ECGs is crucial for capturing the diversity of ECG data and lowering the cost of CVD diagnosis, thereby improving global cardiac care accessibility.

For the PhysioNet Challenge 2024 [7, 8], we introduce VinDigitizer, a digitization approach based mainly on image processing algorithms. The paper is structured as follows. Section 2 outlines the VinDigitizer methodology, detailing its three core phases: lead segmentation, signal extraction, and scale conversion. Section 3 reports on our performance during the official challenge phase, including an analysis of the digitizer's strengths and limitations. Section 4 summarizes our key insights and introduces potential directions for future research.

2. Methodology

VinDigitizer consists of three main stages, as shown in Figure 1: detecting signal rows on the paper ECG, extracting signal from the background, and finally assigning numerical values to each signal point. This section describes the inner functionality and key techniques of each stage.

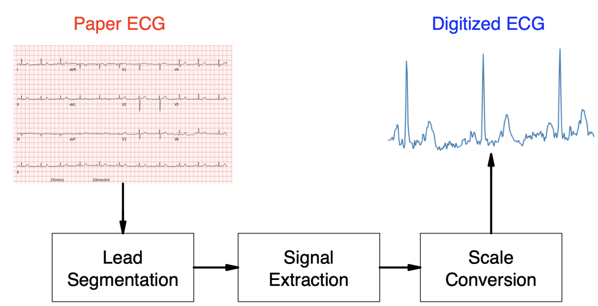

Figure 1. Proposed digitization pipeline.

2.1. Lead Segmentation

The initial phase involves isolating ECG lead-containing image sections. Standard paper ECGs typically present all 12 leads in 2.5-second segments, arranged in three rows of four leads each. To achieve this segmentation, we trained a YOLOv8 [9] model on a dataset of 300 labeled paper ECGs, enabling it to identify four distinct signal rows. Figure 2 illustrates the outcome of this segmentation step.

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.135

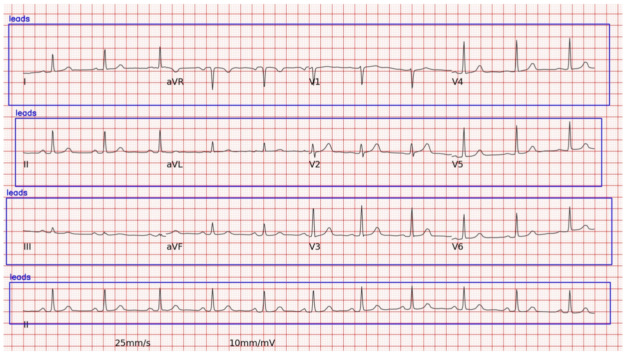

Figure 2. Lead Segmentation: capturing signal rows on paper ECG with YOLOv8.

For YOLOv8 model training, we utilized the PTB-XL dataset [10] to generate 300 synthetic paper ECG images via the ECG-Image-Kit's synthetic ECG image generator [11, 12]. These synthetic images were then manually annotated with bounding boxes enclosing the lead rows using the LabelImg tool[1].

2.2. Signal Extraction

The next phase involves extracting numerical values that represent the signal waveform within each image. Figure 3 illustrates this signal extraction process. It begins with the color conversion of RGB to grayscale, which quantifies the intensity of each pixel. Subsequently, we transform the grayscale image into a binary representation with Otsu's thresholding [13]. In this binary image, intensity values below the computed threshold are black (0), and the remaining are white (255).

$$I = 0.299 \times R + 0.587 \times G + 0.114 \times B \qquad (1)$$

Next, the background grid from the binary image is iteratively removed, following the procedure described in [14]. Vertical and horizontal lines forming the grid are detected using white-point column density. If a grid is detectable in the binary image, the threshold used for grayscale-to-binary conversion is reduced by 5%. A new binary image is then generated using this reduced threshold. This process is repeated until the grid becomes undetectable, resulting in a signal-only binary image.

Finally, we extract the pixel values representing the signal waveform from the clean binary image. We employ the process introduced in [14]. This method interprets vertical white pixel clusters as nodes and defines the signal as the least-cost path from left to right across the image. Thus, signal extraction equates to identifying the optimal cross-image path. The algorithm begins by identifying candidate nodes at each column. Then, it analyzes each column, evaluating potential connections to points in the nearest preceding non-empty column, and calculating the cost of each connection using a cost function that considers both Euclidean distance and path angle alignment. The path with the minimal total cost is selected as the extracted signal, and an array of vertical pixel values along this path is returned.

2.3. Scale Conversion

This phase aims to convert these signal values from pixel scale to voltage units. The pixel-to-voltage conversion is accomplished using reference squares on the paper ECG, where each square of length 10mm corresponds to 1mV and 0.4 seconds (25mm/s).

First, vertical and horizontal lines are extracted using Hough transformation [15]. Grid resolution (in pixels) is then estimated by subtracting the coordinates of two adjacent lines. The scaling factor is calculated by dividing the grid length in millimeters by the grid resolution and multiplying by the voltage-millimeter scale, as shown in Equation 2. This scaling factor can be applied directly to the pixel-scale signal to obtain the voltage signal.

$$scalingFactor = \frac{gridSizeMM}{gridSizePixel} \times \frac{1mV}{10mm} \qquad (2)$$

Finally, the resulting signal is centered by subtracting the median of the signal from each signal point.

2.4. Post-processing

Since we digitized the entire signal rows on the ECG image, the leads on each row must be separated. The first three rows on the ECG image contain four leads of equal length; thus, each row signal is divided into four segments, with each segment assigned to a lead. To ensure that each lead signal has the same target length (the number of samples specified in the header divided by four), we apply cubic spline interpolation [16] to scale them horizontally.

The final step is to position the 12 leads as they appear on the ECG image. We initialize 12 zero-filled arrays with lengths equal to the number of samples in the header, divided into four segments to accommodate a lead signal. Each lead signal is then placed in its corresponding segment, with the other segments remaining zero. These positioned leads are then converted into Waveform Database (WFDB) format to save on disk.

3. Results

In addition to the Challenge's private test sets [17], we evaluated our digitization algorithm on a custom validation

[1] https://github.com/HumanSignal/labelImg

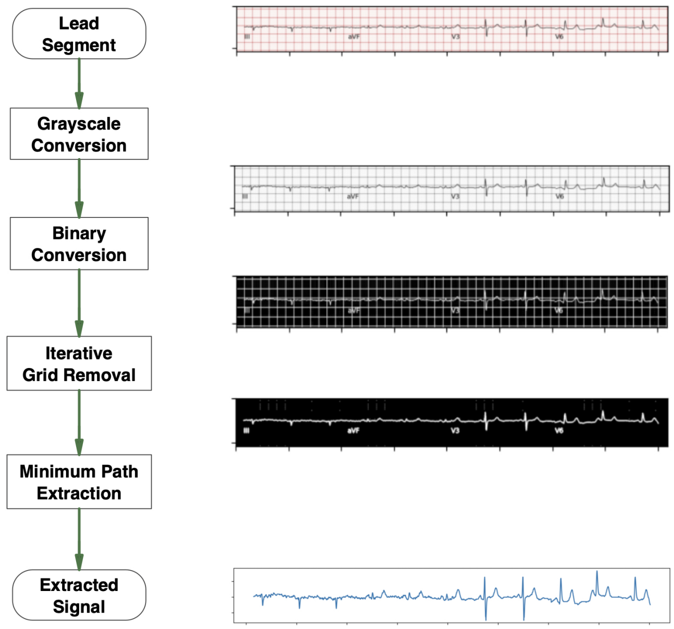

Figure 3. Signal extraction procedure.

dataset for fine-tuning purposes. This dataset comprises 60 synthetic ECG images generated from the PTB-XL dataset using [12], including 30 images generated using default settings without artifacts, and 30 images generated with random artifacts. The settings for the artifacts are as follows: crease angle 90 degree; five and six creases vertically and horizontally, respectively; noise 20; random resolution, padding and grid color; random add header ratio 0.2; random grid present ratio 0.8; wrinkles enabled.

The results of our cross-validation set are as follows: Signal-to-Noise Ratio (SNR) of 0.463, SNR median of 4.281, Kolmogorov-Smirnov (KS) of 0.883, weighted absolute difference (WAD) of 0.440, and adaptive signed correlation index (ASCI) of -0.401. The results demonstrate that artifacts such as folding, creases, and shading negatively impact the digitization outcomes, as illustrated in Figure 4. When excluding artifact-affected images from our validation set, we achieved an SNR of 13.121. Our solution ranked $9/16$ in the digitization leaderboard (Table 1), with specific SNRs for various types of papers shown in Table 2.

Table 1. Signal-to-noise (SNR) ratio and F-measure of our team's model on the hidden data for the digitization and classification tasks, respectively.

Task	**Score**	**Rank**
Digitization SNR	-0.066	9/16
Classification F-measure	n/a	n/a

Table 2. SNR Performance Metrics Across Different Scan Types in the private set.

Leaderboard	-0.066
Color scans of clean papers	-0.069
Black-and-white scans of clean papers	-0.062
Mobile phone photos of clean papers	-0.027
Mobile phone photos of stained papers	-0.027
Mobile phone photos of deteriorated papers	-0.029
Color scans of deteriorated papers	-0.030
Black-and-white scans of deteriorated papers	-0.026
Screenshots of computer monitor	-0.047

4. Conclusion

This study presents VinDigitizer, a solution for digitizing 12-lead ECG signals from paper-based records. Our digitizer operates by first isolating signal rows from the input image, then iteratively removing background grids, before extracting signals by minimum-path detection. Despite its capabilities, VinDigitizer remains susceptible to errors arising from various artifacts related to printing, handling, storage, and scanning processes, such as folding and creases. Future work could explore more advanced image-denoising techniques to address these artifacts and further improve the accuracy and robustness of the digitization process.

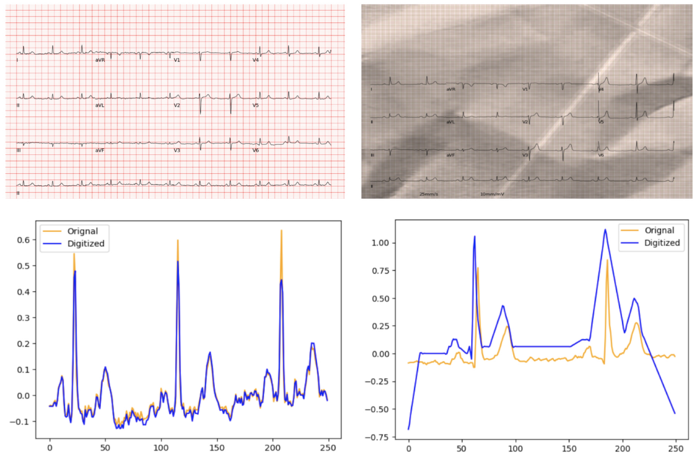

Figure 4. Digitization results for synthetic paper ECG without artifacts (left) and with diagonal folding (right). Plotted signals are of lead I.

References

[1] Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a Deep Neural Network. Nature medicine 2019; 25(1):65–69.

[2] Liu X, Wang H, Li Z, Qin L. Deep learning in ECG diagnosis: A review. Knowledge Based Systems 2021; 227:107187.

[3] Nguyen CV, Do CD. Transfer Learning in ECG Diagnosis: Is It Effective?, 2024. URL `https://arxiv.org/abs/2402.02021`.

[4] Strodthoff N, Wagner P, Schaeffter T, Samek W. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. IEEE Journal of Biomedical and Health Informatics 2020;25(5):1519–1528.

[5] Nguyen CV, Duong HM, Do CD. MELEP: A Novel Predictive Measure of Transferability in Multi-label ECG Diagnosis. Journal of Healthcare Informatics Research 2024; 1–17.

[6] Zhu H, Cheng C, Yin H, Li X, Zuo P, Ding J, Lin F, Wang J, Zhou B, Li Y, et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. The Lancet Digital Health 2020;2(7):e348–e357.

[7] Reyna MA, Deepanshi, Weigle J, Koscova Z, Elola A, Seyedi S, Campbell K, Clifford GD, Sameni R. Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge 2024 2024;51:1–4.

[8] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation 2000;101(23):e215–e220.

[9] Jocher G, Chaurasia A, Qiu J. Ultralytics YOLO, January 2023. URL `https://github.com/ultralytics/ultralytics`.

[10] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Scientific data 2020; 7(1):154.

[11] Shivashankara KK, Shervedani AM, Clifford GD, Reyna MA, Sameni R, et al. ECG-Image-Kit: a synthetic image generation toolbox to facilitate Deep Learning-based Electrocardiogram digitization. Physiological Measurement 2024;45(5):055019.

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

[13] Otsu N, et al. A threshold selection method from gray-level histograms. Automatica 1975;11(285-296):23–27.

[14] Fortune JD, Coppa NE, Haq KT, Patel H, Tereshchenko LG. Digitizing ECG image: a new method and open-source software code. Computer methods and programs in biomedicine 2022;221:106890.

[15] Duda RO, Hart PE. Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM 1972;15(1):11–15.

[16] Hall CA, Meyer WW. Optimal error bounds for cubic spline interpolation. Journal of Approximation Theory 1976;16(2):105–122.

[17] Reyna MA, Deepanshi, Weigle J, Koscova Z, Campbell K, et al. ECG-Image-Database: A Dataset of ECG Images with Real-World Imaging and Scanning Artifacts; A Foundation for Computerized ECG Image Digitization and Analysis, 2024. URL `https://arxiv.org/abs/2409.16612`.

Address for correspondence:

Cuong D. Do
G-310, VinUniversity, Hanoi, Vietnam
cuong.dd@vinuni.edu.vn