From Paper to Digital: ECG Processing with U-Net Digitization and ResNet Classification

Xiankai Yu, Yangcheng Huang, Jian Wu, Jiahao Wang, Wenjie Cai

University of Shanghai for Science and Technology, Shanghai, China

Abstract

Introduction: Electrocardiograms (ECGs) exist in both paper and digital formats. While digital ECGs offer ease of analysis and categorization, paper ECGs are still widely used. The PhysioNet Challenge 2024 aims to digitize and classify ECGs captured from images or paper printouts. Method: YOLOv8 Tiny was utilized for image correction. The YOLO identified the tilt angle between the upper and lower lead names. And the angle was used to correct the image. For the digitization task, a ResUNet model with an integrated CBAM module was employed. A novel method was proposed to process input and output images, followed by a column-by-column scanning of the binary output image to derive one-dimensional signals. For the classification task, ResNet50 was utilized. ECGs were classified based on the predicted values of 11 output classes. Result: Our team, USST_ Med, received a challenge score of an SNR of 2.202 for the digitization task (ranked 5th out of 16 teams) and a macro F-measure score of 0.393 (ranked 8th out of 16 teams) for the classification task on the hidden test set. Conclusion: The proposed approaches had good digitalization and classification effects on paper ECGs.

1. Introduction

ECG, as an important tool for diagnosing cardiovascular diseases, mainly comes in two forms: paper ECG and digital ECG. The digital ECG format offers researchers a more convenient means of processing and analyzing data. This format enables the identification of disease categories associated with ECGs through various methods. With the development of artificial intelligence, many automatic ECG classification algorithms have emerged [1–3]. These classification algorithms are primarily based on ECGs recorded in digital format, but such data remains relatively limited. More and more ECGs are now being saved in digital form, but paper ECGs still prevail on a global scale. In addition to the ECGs signals themselves, paper ECGs may also contain other information such as age, gender, height, weight, and even specific diagnosis information. Therefore, the digitization of ECGs is crucial for capturing diverse ECG data and enhancing global accessibility to cardiac care. Several researchers have conducted research on ECG digitization [4–6], with some leveraging deep learning models for signal extraction [7].

The George B. Moody PhysioNet Challenge 2024 [8–10] aims to digitize and classify ECGs captured from images or paper printouts. The aim of this study is twofold: first, to develop an algorithm to segment the signal component from paper ECGs and extract it as a one-dimensional signal; and second, to introduce a deep learning model to classify paper ECGs.

2. Methods

The overall flowchart of this study was shown in Figure 1. It was divided into two tasks: the digitization task and the classification task. The first two steps in both tasks are data preprocessing and model training. In the preprocessing stage, we need to correct the image. For the digitization task, we need to generate label images. For the classification task, we converted each class to either 0 or 1. Finally, we extracted the signal and output the classification.

2.1. Datasets

The PTB-XL dataset [11] contains 21,799 clinical 12 lead ECG records with a length of 10 seconds and sampling rates of 100Hz and 500Hz. The PhysioNet Challenge 2021 Dataset [12] contains a total of 88,253 12 lead ECG recordings, including China Physiological Signal Challenge in 2018 (CPSC 2018), St Petersburg INCART 12 lead Arrestmia Database, Physiokalisch Technische Bundesanstalt (PTB), Georgia database, an undisclosed American database, Shaoxing People's Hospital (Chapman-Shaoxing) and UMich Database. The digitization task only used the PTB-XL dataset. Both datasets were used for the classification task.

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.134

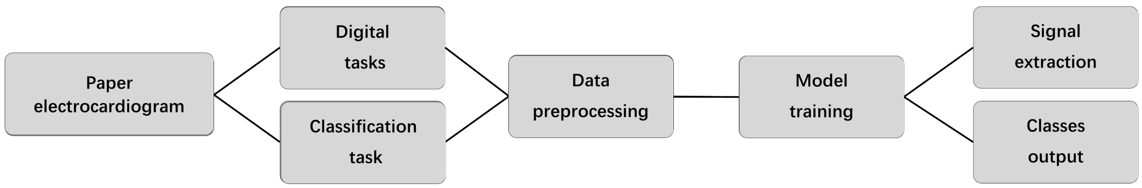


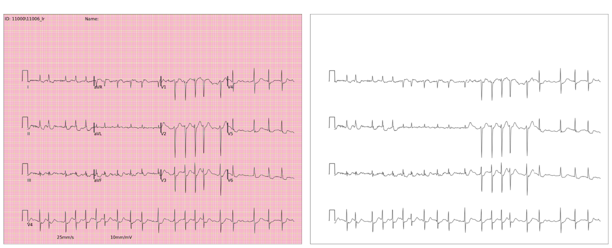
Figure 1. The overall process of this study.

2.2. Data preprocessing

For the classification task, the synthetic ECG image generator generated images from ECG Image Kit [13, 14]. For the digitization task, it was necessary to generate label images corresponding to the input images (Figure 2).

Due to the possibility of image tilt, the first step in processing was to correct the image. The YOLOv8 Tiny model was utilized for object detection, to locate the lead names. After training, the tilt angle between the upper and lower lead names was used as the angle to correct image.

Subsequently, the converted image was prepared for model input. For the digitization task, a threshold was first determined to binarize the labeled image. The input image and labeled image were then converted to grayscale and resized to 1100 × 850 dimensions. Next, the image was divided into four 550 × 425 segments with one row and one column intervals, which were merged into four channel images (Figure 3). The purpose of doing this was to increase the pixel count of the image signal without making the model too complex. In the classification task, the image was cropped to have the same width and height at the center, before being resized to 384 × 384.



a. Input image b. Label image

Figure 2. The input and label of the digitization model.

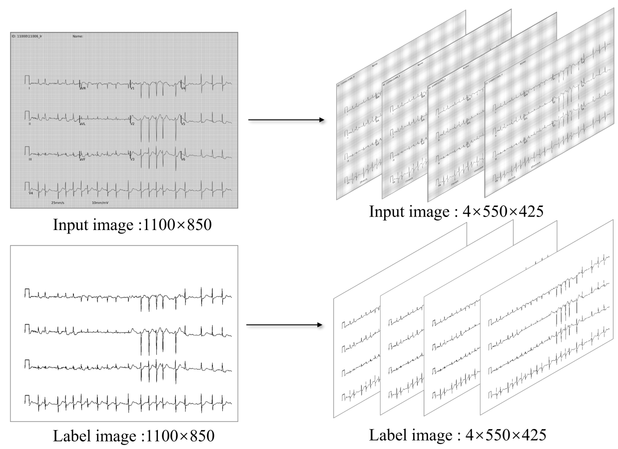


Figure 3. The processing process of original images and labeled images.

2.3. Deep learning models architecture

For the digitization task, we developed the ResUNet architecture [15] (Figure 4). In contrast, for the classification task, we utilized ResNet50 [16] (Figure 5). ResUNet introduces residual blocks into both the encoder and decoder components of UNet, which is a widely adopted network architecture. The network starts with a 32-channel depth in the initial layer and reaches 512 channels in the deepest layer, with both input and output dimensions set at 4 × 550 × 425. Notably, CBAM (Convolutional Block Attention Module) has been incorporated into each module of ResUNet.

On the other hand, ResNet50 consists of a 7x7 convolutional layer, a batch normalization layer, a 3x3 max pooling layer, four residual blocks, a global average pooling layer, and a fully connected layer. The final fully connected layer outputs, 11 classifications, which are activated by Sigmoid function. The ID Block (Identity Block) is a residual block in ResNet50 which is used to handle situations where the input and output dimensions are the same. The Conv Block (Convolutional Block) is another residual block, which is used to handle situations where the input and output dimensions are different.

2.4. Model training and the loss functions

We employed Pytorch as the training framework for our models. The batch size for the digitization model was 16, and the batch size for the classification model was 128.

The learning rate was set to 0.0001, with the digitization model utilizing AMP (Automatic Mixed Precision).

For the classification task, we utilized binary cross entropy (BCE) loss. BCE Loss is a loss function tailored for binary classification problems. It can measure the inconsistency between the probability predicted by the model and the true label. The formula is as following:

$$\text{BCE Loss} = -\left[y \cdot \log(\hat{y}) + (1-y) \cdot \log(1-\hat{y})\right] \quad (1)$$

where y is the target label (actual value), and $\hat{y}$ is the model's predicted probability.

For the digitization task, we utilized Focal Loss [17]. The proportion of signal pixels in the labeled image was very small, while the proportion of background classes was large. Therefore, it is necessary to deal with the problem of class imbalance. The definition of Focal Loss is as follows:

$$\text{Focal Loss}(p_t) = -\alpha_t (1-p_t)^{\gamma} \log(p_t) \quad (2)$$

where p_t represents the model's predicted probability for the true class, α_t is a weighting factor used to balance the importance of positive and negative samples,and γ is a regulatory factor that controls the degree of inhibition of easily classified samples.

Both the digitization model and classification model employed F1 Score as an evaluation metric. The formula is as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (3)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (4)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

2.5. Signal extraction algorithm

The binary image generated by the model was scanned column column by to extract curve information. The base positions were set at the bottom of the pulse columns. The difference in height of each pulse column represented the actual recorded standard voltage difference of 1 mV, from which the voltage value of a pixel could be calculated. To obtain signals, pixel index values within each column were grouped, with emphasis on selecting groups close to the baselines. Then, comparisons were made between neighboring groups to identify individual ECG signal points. By calculating the digital value of each ECG signal point is obtained. Finally, the signal was resampled to the required number of signal sampling points.

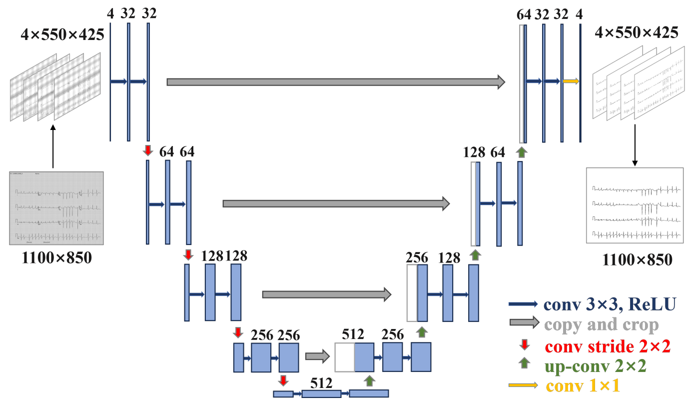


Figure 4. The overall framework of ResUNet.

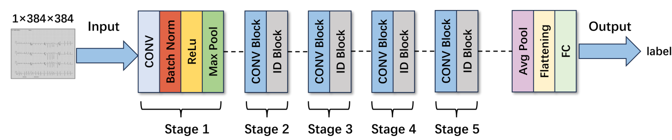


Figure 5. The overall framework of ResNet50.

3. Results

For the digitization task, the training set and validation set were divided in a 9:1 ratio, with 45,000 images in the training set and 5,000 images in the validation set. After 100 epochs of model training, the F1 Score on the local training set reached 0.91 with a corresponding score of 0.89 on the validation set. Upon applying the signal extraction algorithm and utilizing the evaluation code provided by the challenge, an SNR of 5.327 was obtained on the local validation set. For the classification task, the training set and validation set were also divided in a 9:1 ratio, with 58,021 images in the training set and 6,470 images in the validation set. A Macro-F1 score of 0.588 was attained on the local validation set. Due to the issue of category imbalance, an improved Macro-F1 score of 0.634 was ultimately obtained by setting different thresholds for 11 categories. As shown in Table 1, for the the digitization task, the selected entry demonstrated an SNR of 2.202 on the official test set, ranking it in 5th out of 16 teams. For the classification task, our model achieved a macro F-measure of 0.393 on the official test set, placing us 8th out of 16 teams.

Table 1. Performance of proposed models on the local validation set and official test set.

Task	Local validation set	Official test set	Ranking
Digitization	5.327	2.202	5/16
Classification	0.634	0.393	8/16

4. Discussion

In the digitization task, our model achieved an F1 score of 0.89 on the local validation set, demonstrating its strong segmentation performance in accurately identifying ECG signals. This precision in segmentation minimized erroneous signal classifications, facilitating subsequent signal extraction processes effectively. The signal extraction algorithm achieved a good SNR on the local validation set, indicating that the idea of the signal extraction algorithm was correct. The dataset generation was carefully executed, considering various factors such as wrinkles, creases, adding crop rotation, and crop enhancement. In addition, 4-6 rows of signals were set, and their positions in the image also varied. These measures added diversity to the dataset.

For the classification task, we took into account the situation of signals with different row numbers, and used the PhysioNet Challenge 2021 Dataset as a data supplement. The utilization of the YOLO model for image correction during data preprocessing yielded satisfactory results.

Our research has several limitations. Firstly, class imbalance problem has not been addressed when training models for classification task. The BRADY class had the highest number of 19,055 in the training set, but the Acute MI class had the lowest number of only 230. Secondly, digital signal extraction algorithms have not been fully developed, and have much room for improvement to boost performance. Thirdly, the classification task model was relatively straightforward and it has not taken advantages of the generated signals.

In summary, the proposed deep learning models and data processing methods have shown great effectiveness in the digitization and classification of paper ECGs.

References

[1] Wan X, Liu Y, Mei X, et al. A Novel Atrial Fibrillation Automatic Detection Algorithm Based on Ensemble Learning and Multi-feature Discrimination. Medical & Biological Engineering & Computing 2024;62(6):1809-1820.

[2] Ahmed A A, Ali W, Abdullah T A A, et al. Classifying Cardiac Arrhythmia from ECG Signal Using 1D CNN Deep Learning Model. Mathematics 2023;11(3):562.

[3] Hammad M, Iliyasu A M, Subasi A, et al. A Multitier Deep Learning Model for Arrhythmia Detection. IEEE Transactions on Instrumentation and Measurement 2020;70:1-9.

[4] Ganesh S, Bhatti P T, Alkhalaf M, et al. Combining Optical Character Recognition with Paper ECG Digitization. IEEE Journal of Translational Engineering in Health and Medicine 2021;9:1-9.

[5] Fortune J D, Coppa N E, Haq K T, et al. Digitizing ECG Image: a New Method and Open-source Software Code. Computer Methods and Programs in Biomedicine 2022;221:106890.

[6] Karhe R R, Kale S N. Digitization of Documented ECG Signals Using Image Processing. International Journal of Engineering and Advanced Technology 2019;9(1):1286-1289.

[7] Mishra S, Khatwani G, Patil R, et al. ECG Paper Record Digitization and Diagnosis Using Deep Learning. Journal of Medical and Biological Engineering 2021;41(4):422-432.

[8] 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;51:1-4.

[9] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation; 101(23):e215-e220.

[10] Reyna MA, Deepanshi, Weigle J, Koscova Z, Campbell K, Shivashankara KK, 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.

[11] Wagner P, Strodthoff N, Bousseljot R D, et al. PTB-XL, a Large Publicly Available Alectrocardiography Dataset. Scientific Data 2020;7(1):1-15.

[12] Reyna M A, Sadr N, Alday E A P, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. 2021 Computing in Cardiology (CinC). IEEE 2021;48:1-4.

[13] 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 2024;45:055019.

[14] Deepanshi, Shivashankara KK, Clifford GD, Reyna MA, Sameni R. ECG-Image-Kit: a Toolkit for Synthesis, Analysis, and Digitization of Electrocardiogram Images, January 2024.

[15] Zhang J, Jiang Z, Dong J, et al. Attention Gate ResU-Net for Automatic MRI Brain Tumor Segmentation. IEEE Access 2020;8:58533-58545.

[16] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016:770-778.

[17] Ridnik T, Ben-Baruch E, Zamir N, et al. Asymmetric Loss for Multi-label Classification. Proceedings of the IEEE/CVF International Conference on Computer Vision 2021:82-91.

Address for correspondence:

Wenjie Cai
University of Shanghai for Science and Technology,
516 Jungong Road, Yangpu Distric, Shanghai, China,
wjcai@usst.edu.cn