Transforming ECG Images to Waveforms with UNet-LSTM and Multilabel Classification with ResNet-LSTM: A Multimodal Approach by PulsePlex

MD Kamrujjaman Mobin

Shahjalal University of Science and Technology, Sylhet, Bangladesh

Abstract

In the 2024 George B. Moody PhysioNet/Computing in Cardiology Challenge, our team, PulsePlex, worked on digitizing and classifying electrocardiograms (ECGs) from images and paper printouts, aiming to enhance cardiac care accessibility in underserved regions. For the digitization task, we developed a hybrid model combining a UNet and LSTM to convert ECG images into waveform representations. Our model achieved a leaderboard score of 0.081 on hidden validation data and a final test signal-to-noise ratio (SNR) of -0.247 ± 0.096, securing 11th place out of 16 teams. For classification, we used a multimodal fusion approach, with a ResNet extracting image features and an LSTM processing the digitized ECG signals for multilabel classification of cardiac conditions. This resulted in a leaderboard score of 0.317 on hidden validation data and a test macro F-measure of 0.222 ± 0.000 on hidden test data, placing us 10th out of 16 teams in the Official Rankings. These outcomes underscore the challenges of ECG digitization and classification from non-digital sources while highlighting the promise of multimodal approaches in advancing automated ECG analysis.

1. Introduction

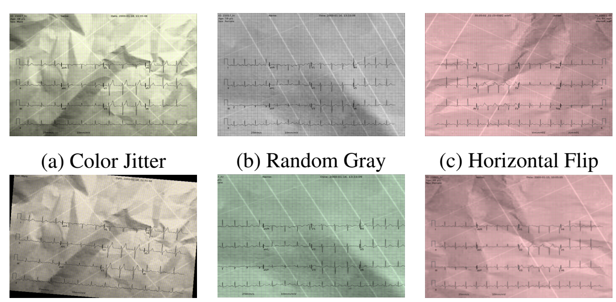
Cardiovascular disease is the leading cause of death worldwide, posing significant health and economic challenges in the United States and globally [1]. The electrocardiogram (ECG), invented by Willem Einthoven in 1895 [2], remains a vital non-invasive diagnostic tool. However, paper-based ECGs are still prevalent, particularly in resource-limited regions of the Global South. These legacy ECGs, crucial for understanding cardiovascular diseases in diverse populations, face significant challenges in digitization and classification, which are essential for improving care in underserved communities.

The 2024 George B. Moody PhysioNet Challenge[3–5] invited participants to develop algorithms for digitizing ECG images and multilabel classification of cardiac conditions. In response, we created a hybrid UNet-LSTM model for digitization and employed a multimodal ResNet-LSTM approach for classification, integrating both image and signal data to improve accuracy. Our work aims to bridge the gap between legacy ECG data and modern digital healthcare solutions.

2. Methodology

2.1. Image Synthesis and Pre-processing

We used the PTB-XL[6] and PTB-XL+[7] datasets to generate 21,799 synthetic ECG images, split into training (17,440) and validation (4,359) sets. Augmentations from ECG-Image-Kit[8, 9] included wrinkles, contrast adjustment (45), sharpness (10), and random grid color. Figure 1 highlights the effects of these augmentations. The Challenge aimed to digitize ECG images by reconstructing waveforms and classifying them into 11 categories [3, 4].



(a) Color Jitter (b) Random Gray (c) Horizontal Flip

(d) Random Rotation (e) Random Crop (f) Gaussian Blur

Figure 1. Images with Augmentation Techniques

To mitigate overfitting, images were resized to 425x650 pixels, converted to RGB, and normalized to a range of [-1, 1]. Augmentations applied included adjustments to brightness, contrast, and saturation (0.2 each), hue (0.1), grayscale (30% probability), horizontal flips (50% probability), 8-degree rotations, cropping (80-100% scale), and Gaussian blur (5x5, sigma 0.1-2). Examples of these augmentations are illustrated in the Figure 1.

Each image was paired with its corresponding ECG signal, flattened for digitization, and labeled with disease information for classification.

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.393

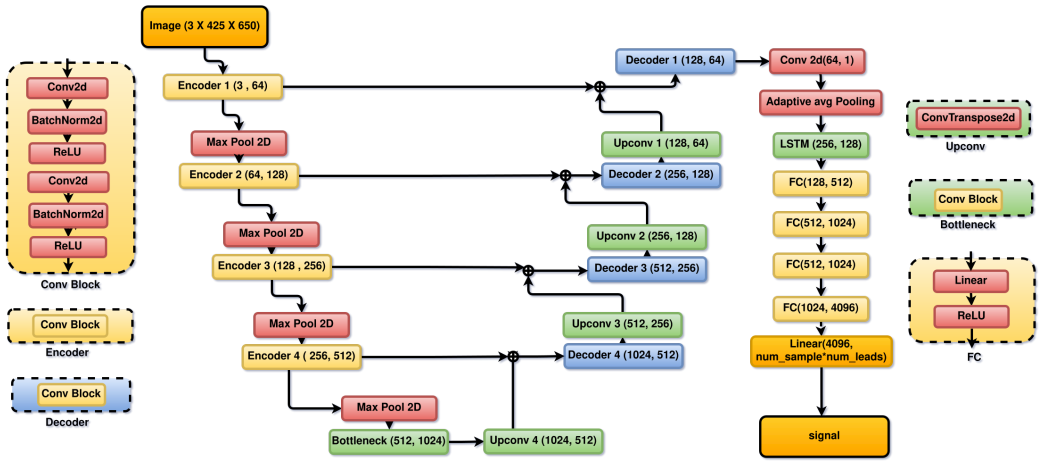


Figure 2. Architecture of the Unet-LSTM based hybrid digitization model. The meaning of the parameters in brackets of Encoder and Decoder is (input filter, output filter).

2.2. ECG Image Digitization

2.2.1. Digitization Model Architecture

The proposed digitization model transforms ECG images into waveforms, inspired by the UNet architecture for image segmentation. Its encoder-decoder structure and skip connections preserve high-resolution features, as shown in **Figure 2**.

- **Encoder:** The encoder consists of four blocks, each with two 2D convolutional layers using 64, 128, 256, and 512 filters. All layers have a 3×3 kernel, padding of 1, and a stride of (1, 1). A max-pooling layer is applied after each block for downsampling, halving the spatial dimensions to reduce complexity while retaining key features.
- **Bottleneck:** At the core of the model is the bottleneck, which connects the encoder and decoder. This component captures a condensed and abstract representation of the ECG image through a single convolutional block featuring 1024 filters. The block uses a 3×3 kernel with padding of 1 and a stride of (1, 1).
- **Decoder:** The decoder is structured symmetrically to the encoder and consists of four upconvolutional blocks, each followed by a convolutional block that includes a pair of 2D convolutional layers. The upconvolutional blocks employ 2D transposed convolution for effective upsampling and concatenate the resulting feature maps with corresponding encoder outputs through skip connections. These blocks utilize 512, 256, 128, and 64 filters, with each block incorporating upsampling followed by two convolutional layers.
- **Recurrent Neural Network (RNN):** The decoder's output is reshaped and fed into a Long Short-Term Memory (LSTM) network with two layers, featuring 128 input features and 64 hidden units, to capture the temporal dependencies of ECG signals.
- **Output Layer:** The LSTM output is processed through fully connected layers that map the 128-dimensional features to 512, 1024, and 4096 units respectively. A final linear layer reshapes the output to match the desired ECG waveform.
- **Loss Function:** In this model, A custom loss function based on Signal-to-Noise Ratio (SNR) is used to evaluate ECG signal reconstruction quality:

$$\text{Loss} = -\text{SNR} = 10 \times \log_{10}\left(\frac{\sum_{n=1}^{N}(\mathrm{y}_n - \mathrm{x}_n)^2}{\sum_{n=1}^{N}(\mathrm{y}_n)^2}\right)$$

where y_n is the ground truth and x_n is the reconstructed signal. The goal is to maximize SNR by minimizing noise, reducing discrepancies between predicted and actual signals. This encourages the model to reduce discrepancies between predicted and actual signals.

2.2.2. Digitization Model Training

The ECG signal reconstruction model is trained using a custom negative SNR loss function, optimized with Adam (learning rate 1×10^{-3}, weight decay 1×10^{-4}) for up to 20 epochs. The batch size is 4. A learning rate scheduler reduces the learning rate by a factor of 0.1 every 4 epochs. Training progress is reported every 50 iterations, and early stopping is applied if validation loss doesn't improve for 7 consecutive epochs. The model with the best validation performance is saved to ensure optimal results.

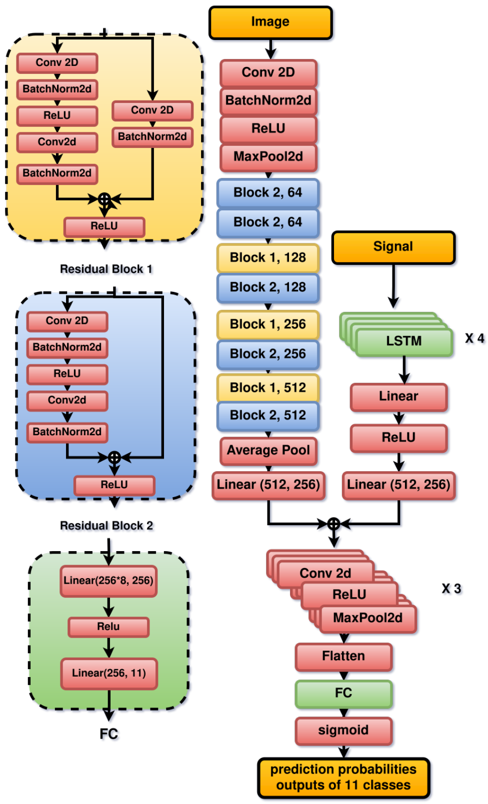


Figure 3. Architecture of the Multimodal ResNet & LSTM Classification model. The meaning of the parameters in each block is the output filters in subsequent layers.

2.3. ECG Image Classification

2.3.1. Classification Model Architecture

The classification model combines image and signal inputs using a ResNet for processing ECG images and an LSTM for handling ECG signals. The features from both modalities are merged before the final classification shown in **Figure 3**.

- **Image Branch (ResNet):** The image branch of the Classification Model uses a ResNet architecture to extract features from 3-channel RGB images sized (3, 425, 550). It starts with a 7×7 convolutional layer with 64 filters and a Max Pooling layer for downsampling. The model has four layers of Residual Blocks of filter sizes 64, 128, 256, and 512. Each Residual layer consists of 2 Residual blocks. Each block contains two 3×3 convolutional layers, ReLU activation, and batch normalization. Downsampling in the Residual Blocks uses a 1×1 convolution layer followed by batch normalization. An Adaptive Average Pooling layer reduces the output to (512, 1, 1), which is then fed into a fully connected layer that outputs a 256-dimensional feature vector for classification.
- **Signal Branch (LSTM):** The signal branch processes ECG signals using an LSTM-based architecture, consisting of four LSTM layers, each with 128 hidden units. Following the LSTM layers, two fully connected layers further reduce the feature vector to 256 dimensions, enabling effective classification of the signals.
- **Combining Image and Signal Features:** The 256-dimensional image and signal features are concatenated to form a 512-dimensional vector.
- **Convolutional Layers for Combined Features:** The combined feature vector is reshaped into a 2D tensor and processed through three convolutional layers. The first layer uses 64 filters, followed by layers with 128 and 256 filters, all with a 3×3 kernel size. Max pooling is applied after each layer to reduce spatial dimensions by half, streamlining feature extraction.
- **Classification Output:** The final feature map is flattened and passed through fully connected layers to 256 units, ultimately outputting 11 probabilities using sigmoid activation for multilabel classification.

2.3.2. Classification Model Training

The Classification Model is trained for 55 epochs using Adam optimizer with an initial learning rate of 1×10^{-3} and weight decay of 1×10^{-4}. A scheduler reduces the learning rate by 0.1 every 7 epochs. Binary Cross-Entropy with Logits is used for the multilabel task, with a sigmoid activation function predicting probabilities for 11 classes. This setup ensures effective classification of ECG images into the specified classes.

3. Results

Our models were evaluated [4] on ECG Image Digitization and Classification tasks. The results for Team PulsePlex's selected entries are shown in Table 1 and 2.

For ECG Image Digitization, a UNet-LSTM model achieved a training signal-to-noise ratio (SNR) of 3.020 ± 0.003. On the leaderboard, its performance dropped to an SNR of -0.081, placing 11th out of 16 teams. In testing, the model performed best on clean color and black-and-white scans, with SNRs of -0.082 and -0.080, respectively. However, it struggled with stained and deteriorated phone-captured scans, black and white scans, as well as screenshots, producing SNRs around -0.303. The overall test SNR was -0.247 ± 0.096.

In the ECG Image Classification task, we combined digitization model-generated signals with input images for classification. The multimodal model achieved a training macro F-measure of 0.476 ± 0.021, while the leaderboard score was 0.317, placing 10th out of 16 teams in the Official Ranking. The macro F-measure was consistent across all test categories, with an overall score of 0.222 ± 0.000.

ECG Digitization and Classification Results by Image Type										
Task	Training Score	Color Scans		Phone Photos			Black and White Scans		Screen shots	Test (mean $\pm$ std)
		Clean	Deteriorated	Clean	Stained	Deteriorated	Clean	Deteriorated		
Digitization (SNR)	3.020 $\pm$ 0.003	-0.082	-0.303	-0.303	-0.301	-0.302	-0.080	-0.303	-0.303	**-0.247 $\pm$ 0.096**
Classification (F-measure)	0.476 $\pm$ 0.021	0.222	0.222	0.222	0.222	0.222	0.222	0.222	0.222	**0.222 $\pm$ 0.000**

Table 1. Detailed SNR (digitization) and F-measure (classification) results across various test image types and categories, with corresponding training scores.

Task	Score	Rank
Digitization	-0.081	**11/16**
Classification	0.317	**10/16**

Table 2. PulsePlex's leaderboard scores and rankings on hidden data for digitization (SNR) and classification (F-measure) tasks.

4. Discussion and Conclusion

This study integrates convolutional and recurrent neural networks, employing a UNet-LSTM for ECG digitization and a ResNet-LSTM for cardiac classification.

In the digitization task, we employed a UNet for spatial feature extraction and an LSTM for capturing temporal dependencies, achieving a mean SNR of -0.247 ± 0.096. Clean color and black-and-white scans yielded SNRs of -0.082 and -0.080, while stained, deteriorated scans, phone photos, and screenshots resulted in SNRs as low as -0.303. The low variance in digitization scores highlights our model's consistent performance across diverse image types and categories.

For classification, our multimodal fusion approach combined ResNet for image features and LSTM for digitized signals, achieving a macro F-measure of 0.222 across all input types, as shown in Table 1. Despite the low SNR in digitization, the consistent F-measure underscores the robustness of our approach compared to unimodal methods.

We also explored transformer models for image and signal processing, but simplified versions did not outperform traditional CNNs like ResNet or UNet. Future work will focus on advanced models, including Vision Mamba, Swin Transformer, UNetFormer, and DETR.

Given the periodic nature of ECG data, attention-based models such as transformers and RNNs show promise for enhancing ECG image and signal analysis. The multi-head attention mechanism in transformers effectively captures temporal dependencies, potentially improving both digitization and classification accuracy. Additionally, hybrid models combining CNNs for spatial extraction with transformers for attention may further enhance the robustness and accuracy of ECG analysis.

In conclusion, our multimodal approach demonstrates stability and promise. Future enhancements in digitization and classification models, with improved preprocessing techniques, could yield significant performance gains.

References

[1] Virani S, Alonso A, Aparicio H, Benjamin E, Bittencourt M, Callaway C, Carson. Heart Disease and Stroke Statistics - 2021 Update: A Report From the American Heart Association. Circulation February 2021;143(8):E254–E743.

[2] Kligfield P. The centennial of the Einthoven electrocardiogram. Journal of Electrocardiology 2002;35:123–129.

[3] Reyna MA, Deepanshi, Weigle J, et al. Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge 2024. Computing in Cardiology 2024;51:1–4.

[4] Reyna MA, Deepanshi, Weigle J, 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.

[5] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, Mietus J, Moody G, Peng C, Stanley H. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation 2000; 101(23):e215–e220.

[6] Wagner P, Strodthoff N, Bousseljot R, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). PhysioNet, 2022.

[7] Strodthoff N, Mehari T, Nagel C, Aston P, Sundar A, Graff C, Kanters JK, Haverkamp W, Doessel O, Loewe A, Bär M, Schaeffter T. PTB-XL+, a comprehensive electrocardiographic feature dataset (version 1.0.1). PhysioNet, 2023.

[8] Shivashankara K, Deepanshi, Shervedani A, Reyna M, Clifford G, Sameni R. Ecg-image-kit: a synthetic image generation toolbox to facilitate deep learning-based electrocardiogram digitization. Physiological Measurement 2024; 45:055019.

[9] Deepanshi, Shivashankara K, Clifford G, Reyna M, Sameni R. Ecg-image-kit: A toolkit for synthesis, analysis, and digitization of electrocardiogram images January 2024;.

Address for correspondence:

MD Kamrujjaman Mobin
156/2 Vashantek, Dhaka Cantonment, Dhaka-1206, Bangladesh
kamrujjamanmobin123@gmail.com