DeRC_ECG: Combined UNet-ResNet Framework for Automated Denoising and Classification of Noisy Paper-Based ECGs

Weijie Sun[1], Siqi Cao[2], Sunil Kalmady [1, 3], Md Saiful Islam [3],
Tayyib Ul Hassan [4], Abram Hindle [1], Russ Greiner [1, 5], Padma Kaul [3]

[1]Department of Computing Science, University of Alberta, Edmonton, Canada
[2]Department of Computer Science Johns Hopkins University, Baltimore, USA
[3]Canadian VIGOUR Center, University of Alberta, Edmonton, Canada
[4] National University of Sciences and Technology (NUST), Islamabad, Pakistan
[5]Alberta Machine Intelligence Institute, Edmonton, Canada

Abstract

As part of the George B. Moody PhysioNet Challenge 2024, we developed a computational approach, leveraging a combination of UNet and ResNet models, to analyze and classify electrocardiogram (ECG) images. Our team, DeRC_ECG, introduced an innovative method that applies UNet for denoising ECG images, followed by the use of ResNet18 Networks and other CNN-based models to accurately classify ECGs into various diagnostic categories. For the classification task, our approach achieved a macro F-measure of 0.359 on the official test set.

1. Introduction

We participated in the 2024 George B. Moody PhysioNet Challenge, which invited teams to develop automated, open-source methods for digitizing and classifying electrocardiograms (ECGs) captured from images or paper printouts [1, 2]. While ECG images or printouts remain common, ECG classification algorithms typically rely on ECG time series data recorded digitally from the ECG devices to identify cardiac abnormalities. Our team focused solely on the ECG image classification task using the UNet-ResNet framework.

2. Methods

Our Challenge entry addresses this problem by first denoising ECG images using a UNet model, followed by classification using CNN-based models, including ResNet, VGG and, ViT. We trained on the labels and ECG images from the PTB-XL dataset [3, 4],. We validated our models using 5-fold cross-validation on the internal test set, then submitted our best-performing models to the Challenge and validated on challenge ECG image data [5].

2.1. Data Process

We ran the ECG-Image-Kit [6, 7] process on the PTB-XL dataset [3, 4] to generate ECG images at both 500 Hz and 100 Hz resolutions. This generation process created noisy ECG images by applied random distortions, grid overlays, simulated handwriting, and slight rotations – distortions designed to mimic real-world scenarios where ECGs are captured from printed or scanned sources, preserving the challenges of noise and artifacts in the data. To enhance the dataset further, we modified the ECG-Image-Kit to generate clean ECG images. These clean images focused solely on the ECG waveforms by removing all background elements, grids, and any superimposed text. Specifically, we excluded text related to the lead name, axes labels and units of the ECG plot. This preprocessing step aimed to create a set of images that highlight the ECG signal without distractions, ensuring that our models could effectively learn from both noisy and clean data.

2.2. UNet Denoising

We focused on denoising the ECG images before feeding them into the classification model, aiming to improve overall performance. For input data, we used noisy ECG images with RGB channels, resized to 896×1152. The ground truth data consisted of clean, grayscale ECG images of the same size. To further reduce background noise, we applied a sigmoid function to the pixel values, and retained only pixels with a value less than 250, effectively isolating the ECG lines from the background. We implemented a UNet[8] architecture for the denoising task, which consists of a contracting path (encoder) and an expansive path (decoder). The encoder was a series of DoubleConv layers, each containing two convolutional layers followed by batch normalization and ReLU activation.

Max-pooling layers are used to progressively downsample the input, reducing spatial dimensions while increasing feature depth. Specifically, the encoder includes convolutional layers with 64, 128, 256, 512, and 1024 filters. The decoder mirrors the encoder's structure, using transposed convolutions to upsample the feature maps. At each upsampling step, the feature maps are concatenated with the corresponding encoder features to preserve spatial context. The DoubleConv layers then refine these feature maps. The final output layer consists of a 1x1 convolution, reducing the feature map to the desired number of output channels (1 grayscale channel), followed by a sigmoid activation to produce the final denoised ECG images. This architecture allows for precise localization and reconstruction of the ECG waveforms, effectively reducing noise. To train our UNet model effectively, we employed a custom loss function that combines Dice Loss and Focal Loss, designed to address the unique challenges of ECG image segmentation.

The **Dice Loss** ($\text{Loss}_{\text{Dice}}$) is particularly useful for handling imbalanced data, where the area of interest (e.g., ECG lines) is much smaller compared to the background. Dice Loss measures the overlap between the predicted segmentation map and the ground truth. It is defined as:

$$\text{Loss}_{\text{Dice}} = 1 - \frac{2 \cdot \sum(\text{p}_{\text{true}} \cdot \text{p}_{\text{pred}}) + \eta}{\sum \text{p}_{\text{true}} + \sum \text{p}_{\text{pred}} + \eta}$$

where p_{true} and p_{pred} represents the ground truth and predicted probabilities, respectively. The η term is added to avoid division by zero. This loss function is effective in maximizing the overlap between the predicted and true labels, ensuring that the model focuses on accurately segmenting the ECG lines.

We incorporate the **Focal Loss** ($\text{Loss}_{\text{Focal}}$)to further address the class imbalance issue, particularly in cases where the majority of pixels belong to the background (white), and only a small fraction represents the ECG lines (black). The Focal Loss modifies the traditional binary cross-entropy (BCE) loss to down-weight easy-to-classify examples, emphasizing harder examples. The BCE and the Focal Loss are defined as:

$$\text{BCE} = -\frac{1}{N} \sum_{\text{p} \in \text{Pixel}} \left(\text{p}_{\text{true}} \log(\text{p}_{\text{pred}}) + (1 - \text{p}_{\text{true}}) \log(1 - \text{p}_{\text{pred}}) \right)$$

$$\text{Loss}_{\text{Focal}} = \alpha \left(1 - \exp(-\text{BCE})\right)^{\gamma} \cdot \text{BCE}$$

where α and γ are hyperparameters that control the contribution of hard-to-classify examples. This loss function enhances the model's ability to learn from difficult cases, thereby improving segmentation performance on the minority class.

Finally, the **Combined Loss** ($\text{Loss}_{\text{Combine}}$) is computed as the sum of Dice Loss and a weighted Focal Loss:

$$\text{Loss}_{\text{Combine}} = \text{Loss}_{\text{Dice}} + \beta \cdot \text{Loss}_{\text{Focal}}$$

where β is a scaling factor that adjusts the influence of the Focal Loss relative to the Dice Loss. This combined loss function ensures that the model is not only accurate in segmenting the ECG lines but also robust to the class imbalance present in the data, leading to improved overall performance.

For training, we used the Adam optimizer with a learning rate of 1e-4 and a ReduceLROnPlateau scheduler, which reduced the learning rate to a minimum of 1e-6 after 7 epochs of no improvement in validation loss. We also implemented early stopping with a patience of 10 epochs to prevent overfitting, retaining the model weights from the epoch with the lowest validation loss.

2.3. Classification Task

For the classification task, we used both clean and denoised ECG images to classify 11 diagnostic labels. We experimented with several pre-trained models, including ResNet18, ResNet50, ResNet152[9], VGG11[10], and ViT-ECG[11], adapting each model to suit our single-channel ECG input and multi-label classification requirements. To accommodate the grayscale ECG images, we modified the first convolutional layer of each model to accept a single-channel input instead of the standard three channels. Specifically, for ResNet18, the first convolutional layer was reconfigured to process single-channel ECG inputs by averaging the weights across the original three input channels. This adjustment allowed the model to effectively process grayscale images while maintaining the integrity of the pre-trained weights. Additionally, the final fully connected layer of each model was adapted to output predictions for the 11 diagnostic labels. This modification ensured that the model could perform multi-label classification, assigning the appropriate set of diagnoses to each ECG image. We conducted 5-fold cross-validation (5-CV) on the clean ECG images to evaluate the performance of the different models. The evaluation was based on the average Area Under the Receiver Operating Characteristic (AUROC) score across all folds. ResNet18 consistently delivered the best performance, making it our model of choice for the final submission. During training, we utilized the Binary Cross-Entropy with Logits Loss, a loss function well-suited for multi-label classification tasks. The training process for the classification model followed the same approach as described in the UNet model section. We utilized the Adam optimizer with an initial learning rate of 1×10^{-4}, along with a learning rate scheduler and early stopping to prevent overfitting, ensuring optimal model performance.

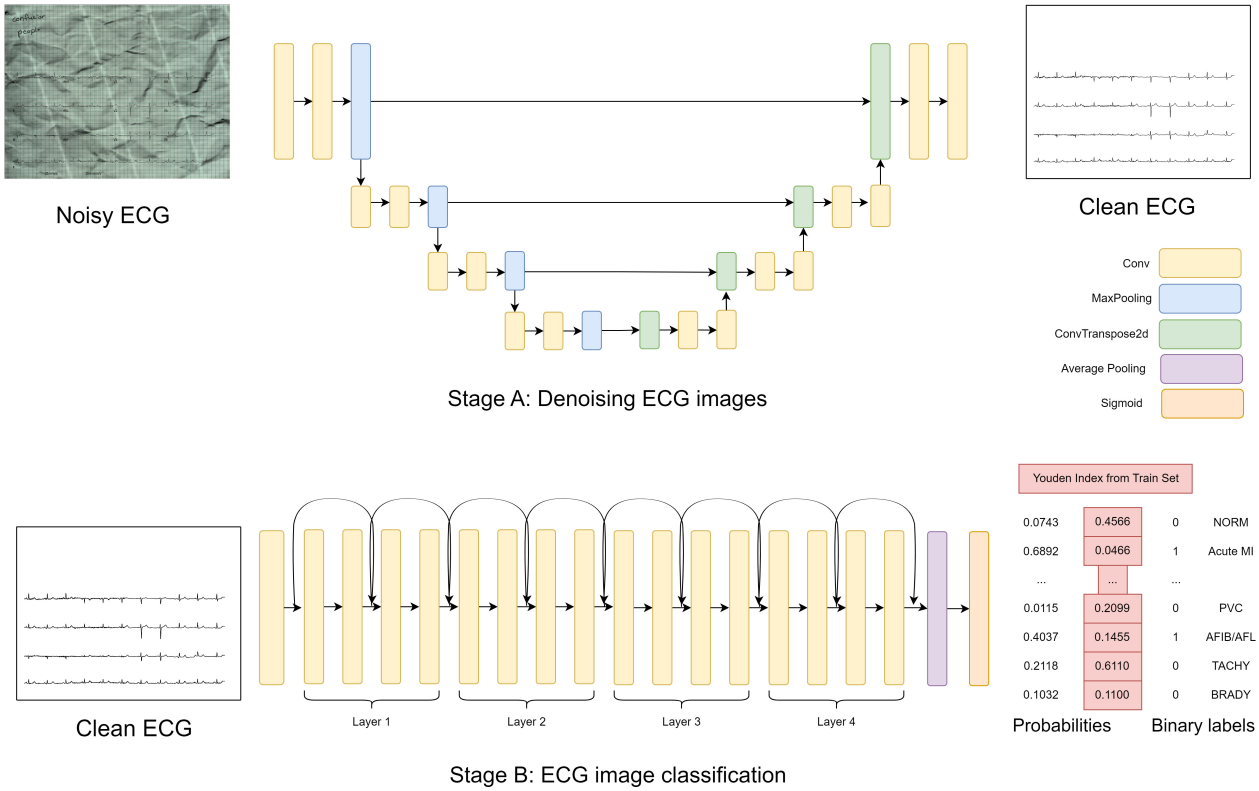

Figure 1. In Stage A, a UNet model denoises the ECG images. In Stage B, the cleaned images are classified into 11 diagnostic labels using a ResNet model. Probabilities from the sigmoid layer are converted to binary outcomes using Youden's Index thresholds.

After training the multi-label classification model, we used the Youden's Index to determine the optimal threshold for each label. Specifically, we trained the model on an 80% training set, and used the remaining 20% as a validation set to obtain the predicted probabilities for each label. To find the best threshold that balances the true positive rate (sensitivity) and the true negative rate (specificity), we set the threshold to Youden's Index, $J = \text{sensitivity} + \text{specificity} - 1$, which identifies the threshold that maximizes both sensitivity and specificity, ensuring that the model performs optimally in distinguishing between the presence and absence of each label.

3. Results

We evaluated our models for ECG image classification using 5-fold cross-validation. Performance metrics included macro F1-score, accuracy, and AUROC for each label in the PTB-XL training set.

Table 1 presents a comprehensive comparison of the model's performance across individual labels. It shows the mean and standard deviation for accuracy, AUROC, and F1-score based on 5-fold cross-validation.

Label	Accuracy	AUROC	F1
NORM	0.83 ± 0.03	0.91 ± 0.02	0.82 ± 0.03
Acute MI	0.73 ± 0.09	0.71 ± 0.06	0.04 ± 0.02
Old MI	0.76 ± 0.04	0.82 ± 0.05	0.60 ± 0.05
STTC	0.78 ± 0.03	0.84 ± 0.06	0.62 ± 0.06
CD	0.81 ± 0.03	0.81 ± 0.07	0.61 ± 0.07
HYP	0.81 ± 0.01	0.77 ± 0.12	0.43 ± 0.09
PAC	0.77 ± 0.09	0.80 ± 0.08	0.11 ± 0.07
PVC	0.87 ± 0.05	0.93 ± 0.06	0.43 ± 0.15
AFIB/AFL	0.83 ± 0.06	0.90 ± 0.04	0.43 ± 0.14
TACHY	0.93 ± 0.01	0.97 ± 0.01	0.53 ± 0.04
BRADY	0.85 ± 0.05	0.89 ± 0.06	0.25 ± 0.07

Table 1. Performance metrics for different labels using 5-fold cross-validation (Mean $\pm$ SD)

Our experiments demonstrated that UNet combined with ResNet-18 (UNet-ResNet-18) outperformed its combinations with other models, such as VIT-ECG, VGG11, ResNet-50, and ResNet-152, particularly in terms of F1-score, as shown in Table 2.. Therefore, UNet-ResNet-18 was selected for the challenge submission.

Model	F1-measure
UNet + VIT-ECG	0.40
UNet + VGG11	0.37
UNet + ResNet-50	0.41
UNet + ResNet-152	0.37
UNet + ResNet-18	**0.44**

Table 2. Comparative performance of different CNN architectures based on F1-measure

When evaluated on the official test set [5], our best-performing model, UNet-ResNet-18, achieved a challenge score of 0.359 on the leaderboard, as shown in Table 3.

Task	Score	Rank
Classification	0.359	NA

Table 3. Task performance summary

4. Discussion and Conclusions

In this challenge, we deviated from traditional approaches of digitizing ECG images into waveforms for classification. Instead, we focused on denoising ECG images to create cleaner representations for classification. Our UNet-based denoising method showed promising results, but our exploration of classification models was limited to simple CNN architectures. Future work should explore more advanced classification models to enhance performance.

Significant variability in ECG formats across institutions and devices may limit our method's generalizability. For instance, some ECGs are formatted with four leads per line, while others have three leads per line. Such differences could impact model performance on unseen formats. To address these challenges, we suggest two potential improvements: Instead of denoising the entire ECG image, a patch-based approach using UNet could be employed [12]. By breaking down the ECG images into smaller patches, the denoising process could remain robust, even if the overall layout changes across various formats. Classification by Specific Regions: Instead of classifying the entire ECG image, we could focus on specific regions containing the leads. This approach could enhance model robustness and accuracy.

In conclusion, our study highlights the importance of both denoising and classification in the context of ECG image analysis. While our method shows potential, particularly in its novel approach to handling noise, further research is needed to refine the classification models and improve generalizability to diverse ECG formats. Addressing these challenges will be crucial for enhancing the robustness and applicability of ECG image-based diagnosis in real-world clinical settings.

References

[1] 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.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer, 2015; 234–241.

[9] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv e-prints. arXiv preprint arXiv151203385 2015;10.

[10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-scale Image Recognition. arXiv preprint arXiv14091556 2014;.

[11] Wu B, Xu C, Dai X, Wan A, Zhang P, Yan Z, et al. Visual Transformers: Token-based Image Representation and Processing for Computer Vision, 2020.

[12] 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.

Address for correspondence:

Weijie Sun - weijie2@ualberta.ca
116 St & 85 Ave, Edmonton, AB T6G 2R3, Alberta, Canada