Detecting and Correcting Electrocardiogram Limb-Lead Electrode Swapping via Deep Learning

Ritika Pandey[1], Matthew A. Reyna[2], Zuzana Koscova[2], Gari D. Clifford[3], and Reza Sameni[2,4]

[1]School of Medicine, Emory University, Atlanta, GA, USA
[2]Department of Biomedical Informatics, Emory University, Atlanta, GA, USA
[3]Department of Biomedical Engineering, Anesthesiology, and Critical Care, Johns Hopkins University, Baltimore, MD
[4]Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Electrocardiogram (ECG) electrode swaps are common in high-acuity care and out-of-clinic settings, reported in up to 4% of critical care recordings and resulting in diagnostic errors. Automatic methods for detecting limb electrode swaps remain limited, particularly for left arm and left leg (LA-LL) swaps, as the most common error. Using the physical relationship between ECG electrodes and lead voltages, we developed a generative model to synthesize limb electrode swaps and trained one-dimensional convolutional neural network-based model for binary electrode swap detection, and multiclass permutation classification and correction. The model were trained using synthetic swaps from the CODE-15% dataset and externally evaluated on PTB–XL dataset. The binary classifier achieved an area under the receiver operating characteristic curve (AUROC) of 0.99 on held-out CODE-15% and unseen synthetically electrode-swapped PTB-XL data, while the multiclass classifier achieved a macro-AUROC of 0.99 on both datasets. Applied to unmodified PTB-XL ECGs, the binary classifier identified 4.3% as potentially swapped, similar to population-level reported clinical prevalence. Our models demonstrate accurate detection and classification of limb electrode swaps, including LA-LL swaps, with strong generalizability to unseen datasets. Automated detection of electrode swaps could reduce diagnostic error and enable real-time ECG quality control.

1. Introduction

The electrocardiogram (ECG) is routinely used to diagnose cardiovascular pathologies, but electrode misplacement can alter lead morphology and cause downstream diagnostic errors. Complete electrode swaps have been estimated to occur in approximately 0.4% of emergency department and up to 4% of intensive care recordings and can mimic conditions including myocardial infarction, dextrocardia, and extreme axis deviation [1, 2]. Automated detection methods remain limited, particularly for detecting left arm and left leg electrode swaps [1, 3].

In this work, using the physical relationship between electrode and lead voltages, we develop and train neural network-based classifiers to detect and correct limb lead electrode swaps.

2. Methods

2.1. Theory of limb-lead electrode swapping and correction

Standard 12-lead ECGs are acquired using ten electrodes: three limb electrodes placed on the right arm (RA), left arm (LA), and left leg (LL); six precordial electrodes; and a right-leg (RL) electrode used as the electrical reference. Focusing on the limb electrodes, we denote the electrical potentials at the RA, LA, and LL electrodes by $e_{\rm RA}$, $e_{\rm LA}$, and $e_{\rm LL}$, respectively, and the common reference by e_0. The limb electrode voltages relative to this reference are $v_{\rm RA} := e_{\rm RA} - e_0$, $v_{\rm LA} := e_{\rm LA} - e_0$, and $v_{\rm LL} := e_{\rm LL} - e_0$. Both the electrode potentials and lead signals are functions of time, but we drop the time argument for notational simplicity. Using Einthoven's equations, the limb leads are

$$\begin{aligned} v_{\rm I} &:= v_{\rm LA} - v_{\rm RA} = e_{\rm LA} - e_{\rm RA}, \\ v_{\rm II} &:= v_{\rm LL} - v_{\rm RA} = e_{\rm LL} - e_{\rm RA}, \\ v_{\rm III} &:= v_{\rm LL} - v_{\rm LA} = e_{\rm LL} - e_{\rm LA} = v_{\rm II} - v_{\rm I}. \end{aligned} \tag{1}$$

Next, let $\mathbf{e} := [e_{\rm RA}, e_{\rm LA}, e_{\rm LL}]^\top$ denote the vector of electrode node potentials and $\mathbf{v} := [v_{\rm I}, v_{\rm II}]^\top$ the lead voltages for Leads I and II. Although $v_{\rm III}$ is theoretically a linear combination of $v_{\rm I}$ and $v_{\rm II}$ following (1), real ECG-recordings contain lead-specific noise and measurement error. Depending on the source of the noise (whether common-mode, differential-mode or independent noise), lead $v_{\rm III}$ may be cleaner or noisier than the individual leads

I and II. Therefore, all three limb leads are retained for training the electrode swapping detection model; however, the 2-lead formulation is used for deriving the lead correction algorithm, as it allows for a closed-form solution (after detecting the swapped electrodes).

Using (1), the lead-space vector is related to the electrode-space vector by

$$\mathbf{v} = \mathbf{A}\mathbf{e}, \quad \mathbf{A} := \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{2\times 3}$ is the electrode-to-lead transformation matrix. Because $\mathbf{A}$ is not invertible, electrode potentials can be recovered up to an arbitrary offset using the Moore-Penrose pseudoinverse:

$$\hat{\mathbf{e}} = \mathbf{A}^{+}\mathbf{v}, \quad \mathbf{A}^{+} = \frac{1}{3}\begin{bmatrix} -1 & -1 \\ 2 & -1 \\ -1 & 2 \end{bmatrix}. \tag{3}$$

where $\hat{\mathbf{e}}$ is an estimate of the electrode potentials (up to a constant voltage offset); $\mathbf{A}\mathbf{A}^{+} = \mathbf{I}_2$ is identity and $\mathbf{A}^{+}\mathbf{A}$ is a projection (idempotent). With this formulation, electrode swapping corresponds to connecting electrodes to incorrect anatomical locations while the acquisition system follows the standard differential lead measurement and the software continue to apply their standard construction and diagnosis algorithms. The three electrodes RA, LA, and LL can have up to $3! = 6$ permutations. In practice, the LA and LL electrodes have been reported as the most commonly swapped, comprising an estimated 58% of electrode swaps in critical care [4]. Swapping the LA and LL electrodes yields $\tilde{v}_{\mathrm{I}} = v_{\mathrm{RA}} - v_{\mathrm{LL}} = v_{\mathrm{II}}$, $\tilde{v}_{\mathrm{II}} = v_{\mathrm{RA}} - v_{\mathrm{LA}} = v_{\mathrm{I}}$ and $\tilde{v}_{\mathrm{III}} = v_{\mathrm{LL}} - v_{\mathrm{LA}} = -v_{\mathrm{III}}$. Thus, an LA-LL swap interchanges Leads I and II and negates Lead III, potentially producing misleading but physiologically plausible morphology.

Mathematically, electrode swapping can be generally modeled with a permutation matrix $\mathbf{P} \in \mathbb{R}^{3\times 3}$ (a 3×3 binary matrix with a single one in each row and column), which acts on the electrode space vector $\mathbf{e}$. $\mathbf{P} = \mathbf{I}_3$ corresponds to the normal electrode configuration; other choices of $\mathbf{P}$ result in electrode swapping. Using the properties of the Moore–Penrose pseudoinverse, we can show that if the electrode permutation $\mathbf{P}$ is known, the electrode-swapped lead vector $\tilde{\mathbf{v}}$ can be mapped back to the correct lead configuration $\mathbf{v}$ via

$$\mathbf{A}\mathbf{P}^{\top}\mathbf{A}^{+}\tilde{\mathbf{v}} = \mathbf{v} \tag{4}$$

Therefore, the problem of correcting the limb leads reduces to finding the electrode permutation matrix.

2.2. A deep learning pipeline for detecting electrode swapping

We used synthetically swapped electrode data to build a deep learning pipeline to detect the swapped electrodes.

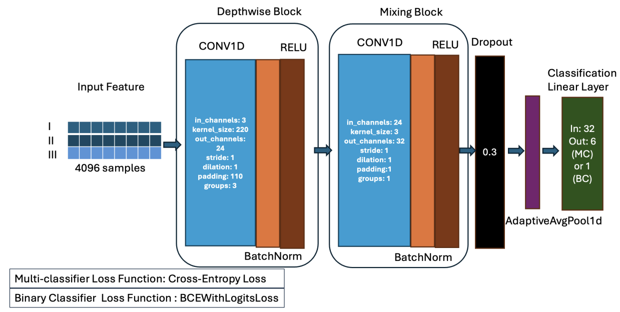


Figure 1: One-dimensional depthwise separable CNN architecture for the binary and multiclass classifier.

2.2.1. Data generation and preprocessing

We generated synthetic limb-lead swapped ECGs from the CODE-15% dataset [5], which contains 345,779 12-lead ECGs from 233,770 patients with a sampling frequency of 400 Hz. ECGs are stored as 4096×12 numeric matrices, corresponding to 7–10 second recordings zero-padded to 4096 samples. For the limb-lead model, only Leads I, II, and III were retained, yielding a 3×4096 input matrix.

Synthetic swapped ECGs were generated using the lead-to-electrode-space transformation

$$\tilde{\mathbf{v}} = \mathbf{A}\mathbf{P}\mathbf{A}^{+}\mathbf{v}, \tag{5}$$

where $\mathbf{v}$ represents the original ECG, $\mathbf{A}$ and $\mathbf{A}^{+}$ are as defined in (2) and (3), and $\mathbf{P}$ is a permutation matrix specifying the electrode swap.

For binary classifier training, 60 k CODE-15% ECGs were used to generate one unswapped and one randomly selected swapped sample each (120 k in total). For multiclass training, 30 k ECGs were used to generate all six permutations (180 k total; 30 k per class).

After synthetic data generation, ECGs were preprocessed with a 0.5 Hz high-pass filter and mean subtraction to remove baseline drift. The CODE-15% dataset was then split into 70% training, 15% validation, and 15% unseen test sets.

2.2.2. Binary classification task

For acute care settings, we designed a binary classifier, intended to function as a real-time detection system (which could potentially be implemented in an ECG monitor firmware/software and flag misplaced electrodes to clinicians in real-time). Therefore, in order to have a computationally efficient model, we used a model architecture comprising a one-dimensional-based convolutional neural network shown in Figure 1. The model input consisted of a 3×4096 sample representing Leads I, II and III from 10 s ECG recordings. The model was trained using `BCEWithLogitsLoss` [6]. Hyperparameter tuning was performed over the learning rate, batch size, dropout rate, kernel size, and channel dimensions using Optuna. The final model trained on 30 epochs with a learning rate of 9×10^{-3}, batch size of 64, dropout rate of 0.3, a first-layer kernel size of 220 with multiplier 8, a second-layer kernel size of 3, and 32 mixing output channels.

Model generalization was evaluated using the PTB–XL dataset [7], which contains 21,799 12-lead ECGs from 18,885 patients sampled at 500 Hz and stored as 5000×12 waveforms. All PTB–XL ECGs were resampled from 500 Hz to 400 Hz and zero-padded to dimensions 3×4096 to match the model input format.

Using the same transformation procedure as for model development, approximately 42,000 synthetic binary and multiclass ECGs were generated from PTB–XL to evaluate performance on unseen ECG data with known electrode-swap labels. Generalizability to real-world ECGs was additionally assessed by applying the model to approximately 21,000 unmodified PTB–XL recordings. Because ground-truth electrode-swap labels were unavailable for these recordings, ECGs were ranked by their predicted probability of electrode misplacement. A physician randomly reviewed several cases to assess model accuracy. For each tracing, the predicted electrode-swap permutation was applied to generate a corrected ECG, and the resulting waveform was qualitatively assessed for restoration of physiologically plausible morphology.

2.2.3. Multiclass classification problem

For retrospective data analysis, a multiclass classification approach is more appropriate because it enables correction of existing ECG datasets. Using (4), as long as the permutation matrix is known (or estimated), correction of electrode swaps and perfect recovery of the corrected lead measurements is mathematically guaranteed.

The multiclass model had a similar architecture in its input and hidden layers, with a six-class output head (corresponding to the six permutations of the LA, LA, and LL electrodes), and was trained using cross-entropy loss. The highest-scoring class determined the predicted electrode permutation. Model performance was evaluated using macro recall, macro AUROC, and a row-normalized confusion matrix.

3. Results

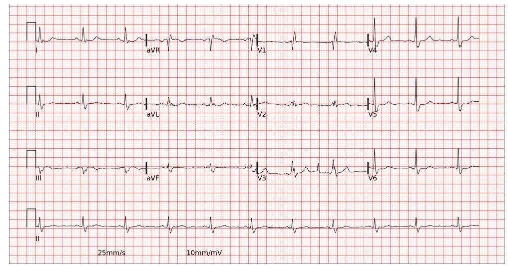
3.1. Binary classification

For the reported results, the classifier output score was thresholded at 0.38, resulting in a sensitivity of 0.98 on the validation set, to classify each ECG as either "swap" or "no swap." The model was evaluated using AUROC, sensitivity, specificity, accuracy, and F1-score at a sensitivity of 0.98.

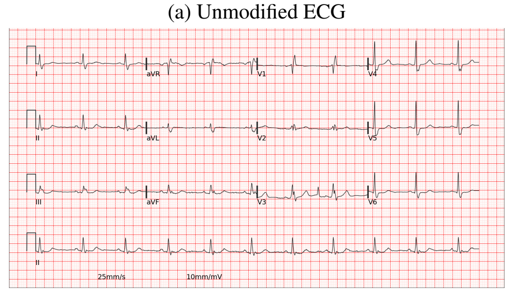
The binary classification performance is summarized in Table 1, with an AUROC of 0.99 on both the CODE 15% and the PTB-XL datsets.

Table 1: Performance of the binary classifier.

Metric	CODE 15% Val.	CODE 15% Test	PTB–XL Test
Accuracy	0.97	0.97	0.96
Sensitivity	0.98	0.98	0.96
Specificity	0.96	0.96	0.96
F1-score	0.97	0.97	0.96
AUROC	0.99	0.99	0.99



(a) Unmodified ECG



(b) Counterfactual ECG

Figure 2: Counterfactual swap transfer for PTB-XL ECG record `6891` (a) Original ECG (swap probability 0.983) (b) Following LA-LL correction (0.001), with restoration of normal Lead II axis.

To evaluate whether inclusion of Lead III improved performance despite its theoretical redundancy, we compared binary classifiers trained using Leads I and II only versus Leads I, II and III across 5 cross validation folds. The 3-lead achieved a slightly higher mean AUROC than the 2-lead model (0.990 ± 0.005 vs 0.987 ± 0.002), supporting the inclusion of Lead III to improve robustness to lead-specific noise and measurement artifacts; although due to the very close performance of the models, the improvement is not necessarily statistically significant.

When applied to 21 k unmodified PTB–XL ECGs, the model identified 4% of tracings as potentially swapped, consistent with the reported prevalence of lead reversal in critical care settings. A physician experienced in ECG reading, confirmed 17/20 of the top scored recordings as lead swapped cases, based on lead polarity. A sample record identified by the model as LA-LL swap is shown in Figure 2.

Table 2: Multiclass classifier performance.

Metric	Code 15%	PTB-XL
Macro recall	0.94	0.85
Macro AUROC	0.99	0.99

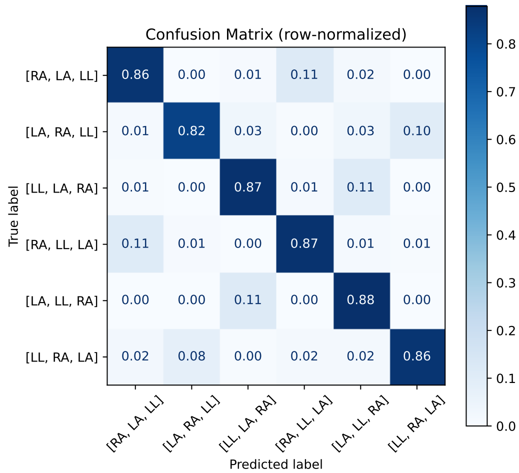


Figure 3: Row-normalized confusion matrix for the multiclass classifier on synthetic PTB-XL ECGs.

3.2. Multiclass classification

The multiclass classifier designed to identify lead swaps achieved macro AUROC of 0.99, macro recall of 0.94 and 0.85 on the validation and synthetic PTB-XL datasets, respectively (Table 2). On PTB-XL, per-class recall ranged from 0.82-0.88 (Figure 3).

4. Discussion and Conclusions

The binary classifier demonstrated strong discrimination on both CODE-15% and synthetic PTB-XL data and identified 4% of unmodified PTB-XL ECGs as potentially swapped, consistent with reported clinical prevalence. Physician overreads of random examples confirmed model accuracy, providing preliminary evidence of generalization to naturally occurring swaps. Clinically, binary detection could enable real-time identification and repositioning of misplaced electrodes during data collection, when integrated into ECG firmware/software.

The multiclass model, which enables predicting the exact electrode permutation, also has a good overall performance, with a macro AUROC of 0.99 and 82-88% per-class recall on the test set. The higher AUROC relative to classification recall likely reflects confusion among physiologically similar swap classes. Prior machine learning approaches have detected limb and precordial electrode misplacement [8, 9]. Our current work focused on limb leads. Although theoretically redundant, Lead III improved performance, likely by providing robustness to lead-specific noise.

Strengths of this study include large-scale training, independent PTB-XL evaluation, and assessment of both binary and multiclass formulations. Limitations include synthetically generated swap labels, restriction to RA/LA/LL permutations, and limited clinical validation. Future work should validate performance on clinically confirmed swaps and more diverse datasets.

Acknowledgments

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).

References

[1] Rudiger A, Hellermann JP, Mukherjee R, Follath F, Turina J. Electrocardiographic artifacts due to electrode misplacement and their frequency in different clinical settings. The American Journal of Emergency Medicine February 2007; 25(2):174–178. ISSN 0735-6757.

[2] García-Niebla J, Llontop-García P, Valle-Racero JI, Serra-Autonell G, Batchvarov VN, Bayés de Luna A. Technical Mistakes during the Acquisition of the Electrocardiogram. Annals of Noninvasive Electrocardiology 2009;14(4):389–403.

[3] Naderi H, Kaplan T, van Duijvenboden S, Pujadas ER, Aung N, Chahal CAA, et al. Deep learning to predict left ventricular hypertrophy from electrocardiograms. Europace 2026; 28(2):euag015.

[4] Jekova I, Krasteva V, Leber R, Schmid R, Twerenbold R, Müller C, et al. Inter-lead correlation analysis for automated detection of cable reversals in 12/16-lead ECG. Computer Methods and Programs in Biomedicine 2016;134:31–41.

[5] Ribeiro A, Ribeiro M, Paixão G, Oliveira D, Gomes P, Canazart J, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications 2020; 11(1):1760.

[6] Ihsanto E, Ramli K, Sudiana D, Gunawan TS. An Efficient Algorithm for Cardiac Arrhythmia Classification Using Ensemble of Depthwise Separable Convolutional Neural Networks. Applied Sciences 2020;10(2):483.

[7] 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.

[8] Rjoob K, Bond R, Finlay D, McGilligan V, Leslie SJ, Rababah A, et al. Machine learning techniques for detecting electrode misplacement and interchanges when recording ECGs: A systematic review and meta-analysis. Journal of Electrocardiology 2020;62:116–123.

[9] Rafiei A, Dwivedi T, Xue J, Schlesinger DE, Katebi N, Albert DE, et al. Toward self-validating ECG systems: A personalized, uncertainty-aware approach for detecting and estimating lead misplacement. Journal of Electrocardiology 2026;95:154204.

Address for correspondence:

Ritika Pandey (ritika.pandey@emory.edu)
Emory University School of Medicine, Atlanta, GA 30322 USA