An examination of lead importance in 12-lead electrocardiogram

Carl Harris[1], **Ankush Jindal**[1], **and Logan Falzarano**[1]

[1]Johns Hopkins University

ABSTRACT

12-lead electrocardiogram (ECG) is a non-invasive, cost-effective method for screening and diagnosing cardiac conditions. Existing work has made significant strides in automatic diagnosis from these signals, but the relative importance of each component lead to classification decisions remains poorly-tackled in existing literature. If high-fidelity predictions can be made from a small number of leads (e.g., using 1 to 4 leads rather than 12), the cost and time required for automated ECG diagnosis could be markedly reduced, which may benefit resource-poor environments and critical care settings. To examine this question, we train a series of models to predict common cardiac conditions (e.g., atrial fibrillation, T-wave changes, etc.) using an expert-annotated dataset of 45,152 ECGs recorded from two hospitals in China and accessed via Physionet. Our primary model architecture will be a CNN trained on raw waveforms, and we will compare the performance of this model to an XGBoost classifier trained on expert features (e.g., rr-interval, qrs-axis) and generic time series features (e.g., signal mean, variance) extracted from the waveforms. We will then conduct a Shapley values analysis to examine the relative importance of each component lead (e.g., I, aVR, V1) to classification decisions. We use the results from our Shapley values analysis to select optimal coalitions of leads for 4- and 8-lead ECG, and externally validate our models on PTB-XL, a publicly available dataset containing 21,799 ECGs.

1 Motivation

What is the problem that you are trying to solve? What is the goal of the project and why is it necessary? Cardiovascular diseases (CVDs) stand as a leading cause of mortality worldwide, exerting a substantial burden on healthcare systems and societies at large. Within the spectrum of CVDs, timely and accurate diagnosis of cardiac conditions is pivotal for effective management and intervention. The 12-lead electrocardiogram (ECG) has long been hailed as a cornerstone in the diagnostic armamentarium, offering valuable insights into cardiac function and rhythm. However, existing literature highlights a critical gap in understanding the relative importance of individual leads in ECG interpretation.

Moreover, the conventional approach to ECG analysis necessitates the acquisition and interpretation of 12 leads, a practice that may pose logistical challenges and resource constraints, particularly in resource-limited settings and critical care environments. The need for cost-effective, time-efficient, and accessible diagnostic modalities underscores the urgency for innovative approaches that optimize the use of ECG data while maintaining diagnostic fidelity.

Furthermore, advancements in machine learning and computational techniques have unlocked unprecedented opportunities for enhancing healthcare delivery and clinical decision-making. Leveraging these technologies, the proposed project endeavors to develop robust predictive models capable of extracting actionable insights from ECG waveforms. By harnessing the power of convolutional neural networks (CNNs) and ensemble learning algorithms such as XGBoost, the research aims to transcend traditional diagnostic paradigms, offering scalable and efficient solutions for cardiac condition classification.

Importantly, the potential impact of the project extends beyond academic discourse, resonating profoundly with healthcare providers, policymakers, and patients alike. By elucidating the relative importance of individual leads in ECG classification, clinicians can tailor diagnostic strategies to patient-specific needs, fostering personalized and evidence-based care pathways. Moreover, the deployment of streamlined diagnostic protocols holds the promise of accelerating clinical workflows, minimizing diagnostic delays, and ultimately improving patient outcomes.

2 Significance

How solving the problem will help society and what would be the impact?

1. Cost-Effective Diagnosis - The ability to make accurate predictions from a reduced number of leads presents a paradigm shift in ECG diagnosis, potentially alleviating the financial burden associated with extensive diagnostic procedures. By streamlining the diagnostic process, healthcare systems can optimize resource allocation, making cardiac care more accessible and affordable for patients across diverse socioeconomic backgrounds.

2. Time Efficiency - Efficient diagnostic protocols translate to faster treatment initiation and improved patient outcomes. By leveraging a smaller number of leads for ECG analysis, healthcare providers can expedite the diagnostic workflow, enabling timely interventions and reducing the risk of adverse cardiac events, particularly in critical care settings where time is of the essence.

3. Accessibility - Reducing the reliance on a 12-lead ECG setup enhances the accessibility of cardiac diagnostics, especially in underserved communities and remote regions

where access to specialized medical equipment is limited. By requiring fewer resources and infrastructure, the project aims to democratize access to quality cardiac care, bridging healthcare disparities and improving health equity on a global scale. helping to improve patient comfort.

4. Improved Patient Care - Accurate and timely diagnosis of cardiac conditions is paramount for delivering effective treatment and mitigating disease progression. By deciphering the relative importance of each lead in ECG classification, clinicians can tailor diagnostic approaches to individual patient needs, fostering personalized care pathways and optimizing treatment outcomes while minimizing unnecessary interventions and associated risks. Additionally, reducing the number of leads would reduce the number of electrodes required to capture an ECG, helping to improve patient comfort.

5. Knowledge Advancement - The project's exploration of novel model architectures, feature extraction techniques, and lead selection strategies contributes to the advancement of knowledge in ECG analysis and cardiovascular medicine. By elucidating the intricate relationship between ECG leads and diagnostic accuracy, the research paves the way for evidence-based clinical practices, empowering healthcare providers with actionable insights and refined diagnostic tools to enhance patient care and clinical decision-making.

3 Materials

Datasets

Training dataset We train our models using a publicly available arrhythmia dataset[1,2], accessed via Physionet. This dataset contains 45,152 patient ECGs annotated with cardiovascular conditions (e.g., atrial flutter, T-wave changes, etc.) identifiable from ECG and annotated by experienced cardiologists.

Validation dataset In our final project, we will validate out models (trained on the arrhthmia dataset) on PTB-XL[3], a publicly available ECG dataset with 21,799 ECGs and annotations by two independent cardiologists.

4 Methodology

4.1 Models

We adapt a convolutional neural network (CNN) architecture as our primary deep learning model, based on the architecture described in Attia et al. (2019)[4]. The model architecture is shown in Figure 2. Currently, we use six blocks of (convolution, batch-norm, ReLU activation, and max pooling), followed by two sense layers and an output layer. For our convolution layers, we use filter sizes of $[16,16,32,32,64,64]$ and corresponding kernel widths of $[5,5,5,3,3,3]$. Our dense layers have a hidden size of 64 and 32, and our output size is 1. We train the model with binary cross-entropy loss.

We train a separate model for each condition, and the output is a probability $\hat{y} \in [0,1]$ estimating the probability of a given condition. We train the model to predict conditions with an overall incidence of ≥ 1 percent in the training dataset.

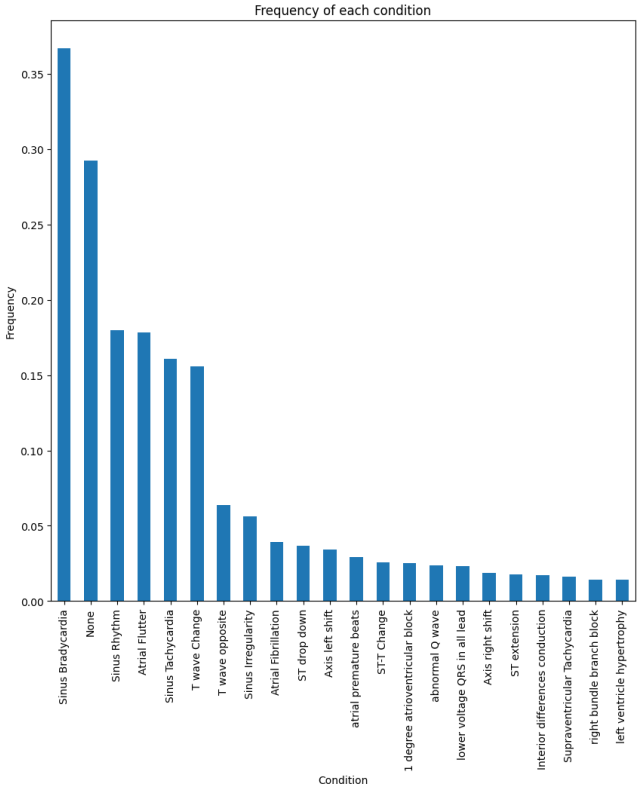

Figure 1. Condition frequency.

4.2 Shapley value approximation

Shapley values are a concept from cooperative game theory that provide a method for assigning a value to each player based on their contribution to the total payoff. In the context of machine learning, Shapley values are used to measure the importance of each feature in a prediction model. They answer the question: If we were to distribute the "payout" (i.e., the prediction) among the features, how much should each feature receive fairly based on its contribution? The Shapley value is the average marginal contribution of a feature value across all possible combinations of features. This approach ensures a fair distribution of the prediction among the features, based on the value they provide.

In the case of 12-lead ECG, using Shapley values to determine the importance of each lead poses a computational challenge. The full computation requires assessing the impact of each lead on the diagnosis by considering all possible subsets of leads. This involves training 2^{12} different models for each condition if one were to analyze the leads independently, which is computationally expensive and practically infeasible. The exponential increase in the number of models is due to the power set of the leads, where each subset (including the empty set and the full set) must be evaluated.

To manage this challenge, an approximation method is introduced, based on the method proposed by Strumbelj et al. (2014)[5], which applies Monte-Carlo sampling to approximate the Shapley values. This approximation involves taking

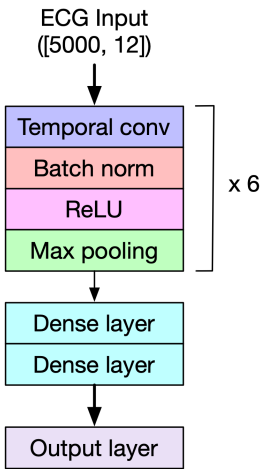

Figure 2. CNN architecture.

random samples of feature coalitions and comparing the prediction of the model with and without a particular feature.

Consider a trained model f that takes as input a 12-lead ECG with dimensionality $x \in [5000, 12]$ and returns an output $\hat{y} = f(x)$, where $\hat{y} \in [0, 1]$ is the estimated probability a given ECG has the condition of interest (e.g., atrial fibrillation). Now we consider the contribution of a given lead number $j \in \{1, 2, ..., 12\}$ to the $\hat{y}$ for a given sample x. To do so, we select a random 12-lead ECG, z, from our dataset. We then "mix" the two ECGs by selecting a random number of ECG leads from z to input into x

For a given feature j, the method creates two versions of the input vector: one with the feature included $x_{+j}^{(m)}$ and one with the feature excluded $x_{-j}^{(m)}$. The model's prediction function, denoted as f, is then applied to both versions. The difference in the predictions, $f(x_{+j}^{(m)}) - f(x_{-j}^{(m)})$, reflects the marginal contribution of the feature j in that particular coalition of features. By averaging these differences over M randomly sampled coalitions, we obtain an estimated Shapley value $\hat{\phi}_j$ for the feature j. This procedure is repeated for each feature to get a comprehensive view of their relative importances. The method thus strikes a balance between accuracy and computational efficiency, allowing for practical estimation of feature importance in high-dimensional datasets such as those with 12-lead ECG.The formal algorithm is described in Algorithm 1.

This approach significantly reduces the computational demand as it doesn't require evaluating all possible subsets, while still providing a robust approximation of each feature's importance to the model's predictions.

Algorithm 1 Approximate Shapley estimation for a single ECG lead

Require: Number of iterations M, instance of interest $\mathtt{x} \in [5000, 12]$, ECG lead index $j \in \{0, 1, ..., 11\}$, data tensor $\mathtt{X} \in [n, 5000, 12]$ (where n is the number of samples), machine learning model $\hat{f}$

Initialize ϕ_j to zero
for $m = 1$ to M **do**
  Draw random instance $\mathtt{z} \in [5000, 12]$ from the data matrix $\mathtt{X}$
  Choose a random permutation $\mathtt{perm}$ of indices $\{0, 1, ..., 11\}$, *excluding* lead j ▷ $\mathtt{perm} \cap j = \varnothing$
  Choose a random number of observations, $\mathtt{r}$, to sample from $\mathtt{perm}$
  $\mathtt{perm} \leftarrow \mathtt{perm}[:\mathtt{r}]$
  $\mathtt{x}_{+j} \leftarrow \mathtt{x}$
  $\mathtt{x}_{+j}[:, \mathtt{perm}] = \mathtt{z}[:, \mathtt{perm}]$ ▷ With j
  $\mathtt{x}_{-j} = \mathtt{x}_{+j}$
  $\mathtt{x}_{-j}[:, j] = \mathtt{z}[:, j]$ ▷ Without j
  $\phi_j^m \leftarrow \hat{f}(\mathtt{x}_{+j}) - \hat{f}(\mathtt{x}_{-j})$ ▷ Marginal contribution
  $\phi_j \leftarrow \phi_j + \frac{\phi_j^m}{M}$ ▷ Update Shapley value estimate
end for

5 Preliminary results

5.1 Full model and single-lead performance

We have currently trained 12-lead and single-lead models. The single-lead models have the same architecture as our main model, but the input size is a single lead (i.e., $x \in \mathbb{R}^{5000}$). This serves as a comparison point to the Shapley approximation described below.

5.2 Shapley approximation

We have implemented the Shapley approximation described above in our dataset. Briefly, we examine the relative importance of each lead using our sampling method applied to our trained 12-lead ECG model. An example of the Shapley value estimates for each lead at shown in Figure 4.

6 Next steps

Hyperparameter optimization We have not conducted hyperparameter optimization yet. In our current formulation, we train a separate model for each of the 19 conditions. We intend to alter our architecture to produce a multi-label output $\hat{\mathbf{y}} \in \mathbb{R}^{19}$ that contains a *vector* of estimated risk probabilities for all conditions. This (1) will expedite training and (2) allow us to more easily conduct hyperparameter optimization. We will use average AU-ROC on the validation set across all models to select our final architecture.

External validation We intend to validate our models on a second dataset (PTB-XL[3]). To do so, we need to ensure that the conditions we predict are the *intersection* of those available in the arrythmia dataset and PTB-XL (i.e., for a given condition we'd like to predict, it must be identified in both datasets). We have not yet confirmed all 19 conditions we are currently predicting are labeled in PTB-XL (though it is likely they are, as ECG abnormalities are quite standardized, particularly for the relatively common conditions we predict here).

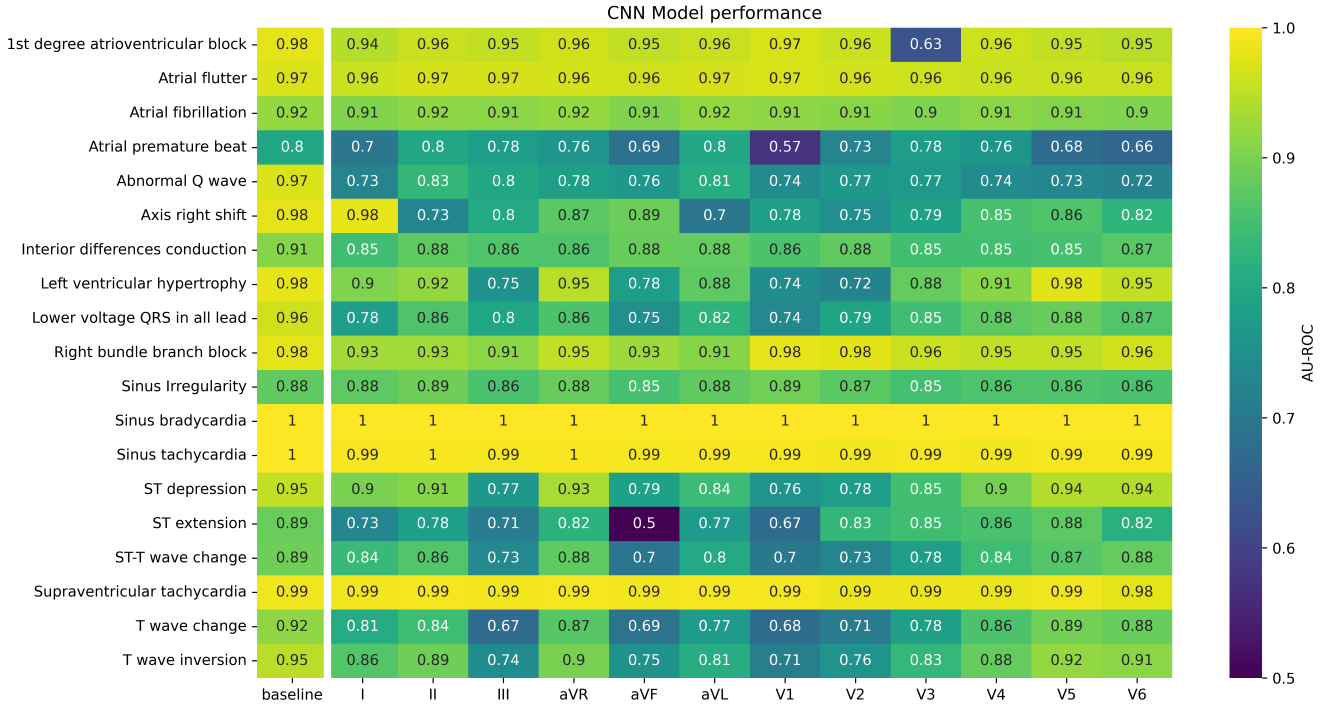

Figure 3. CNN architecture. AU-ROC curves for 12-lead (left most column) and single lead (right 12 columns) models. Conditions are shown in each row, and lead numbers are denoted on the *x*-axis.

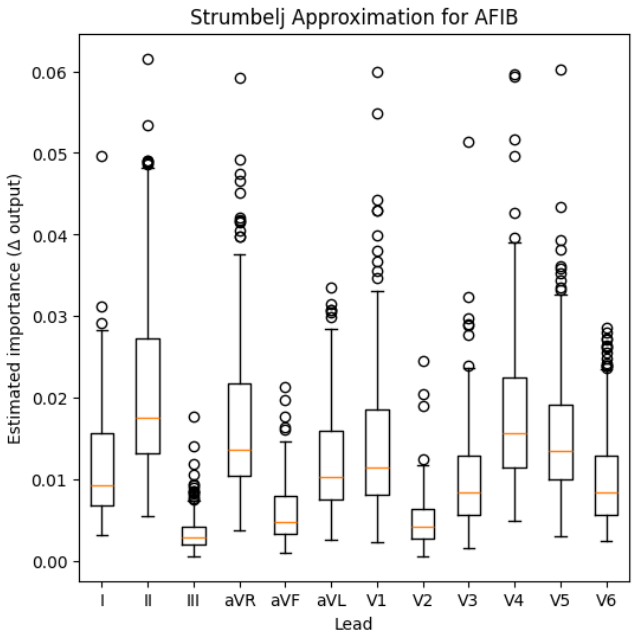

Figure 4. Shapley approximation for AFIB. Each observation represents the change in model output (i.e., risk score upon the addition of the corresponding ECG lead across coalitions) for atrial fibrillation.

Shapley value approximation We have successfully implemented the Shapley approximation method. We intend to systematically compare the single-lead vs. Shapley value models, and identify optimal coalitions. We will then use the Shapley values acquired from our analysis to select the optimal coalition of 2-, 4-, 6-, and 8-lead ECGs. We will then train a model using these coalitions and validate on PTB-XL.

References

1. Zheng, J., Guo, H. & Chu, H. A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0. 0). *PhysioNet 2022* **23** (2022).
2. Zheng, J. *et al.* Optimal multi-stage arrhythmia classification approach. *Sci. reports* **10**, 2898 (2020).
3. Wagner, P. *et al.* Ptb-xl, a large publicly available electrocardiography dataset. *Sci. data* **7**, 1–15 (2020).
4. Attia, Z. I. *et al.* Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat. medicine* **25**, 70–74 (2019).
5. Štrumbelj, E. & Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. information systems* **41**, 647–665 (2014).