Leveraging Cardiologists Prior-Knowledge and a Mixture of Experts Model for Hierarchically Predicting ECG Disorders

Diogo Tuler[1], Pedro Robles Dutenhefner[1],Jose Geraldo Fernandes[1], Turi Rezende[1], Gabriel Lemos[1],Gisele L. Pappa[1], Gabriela Paixão[1],Antônio Ribeiro[1], Wagner Meira Jr.[1]

[1] Universidade Federal de Minas Gerais , Belo Horizonte, Brazil

Abstract

Automatic methods to identify cardiovascular disorders based on Electrocardiograms (ECGs) have been used for decades and gained special attention with the breakthrough of deep learning. This paper proposes a new model based on a Mixture of Experts (MoE) to identify 6 physician-defined clinical labels spanning rhythm and conduction disorders. The MoE framework combines the opinions of multiple models ("experts") to make weighted decisions. Unlike typical MoE scenarios, where the gating network finds ways to weigh and combine the models, our approach leverages prior knowledge from cardiology specialists, who hierarchically divide the space, enabling us to tailor the MoE architecture to incorporate this information explicitly. The disorders are organized in a 2-level hierarchy, where specialists first separate normal ECGs from those rooted in rhythm and morphological abnormalities. The MoE is trained on the CODE-15 dataset, and the results represent the new state-of-the-art in this task.

1. Introduction

An Electrocardiogram (ECG) is a non-invasive exam considered one of the most effective methods for identifying heart disorders nowadays. As heart diseases remain a leading cause of death in 2024 [1], there is an urgent need to develop efficient automatic methods for identifying cardiovascular disorders using ECG exams. Machine learning-based methods have gained significant attention with the breakthrough of deep learning, which has led to the development of robust models for automatically detecting these conditions [2].

Previous deep learning approaches for ECG analysis typically classify diseases by treating each class independently [3], without leveraging structured or prior medical knowledge of class organization. This work incorporates a hierarchical structure, organizing six clinical conditions found in ECGs in three superclasses: "blockage", "rhythm" and "no disorders". Based on this hierarchy, this work proposes a mixture of experts (MoE) framework for the hierarchical classification of electrocardiograms (ECGs). The MoE architecture comprises two main components: a gating network and multiple expert networks. The gating network decides how much influence each expert should have on the final classification, by dynamically assigning weights to the outputs of each expert based on the input data.

The results show that, by routing the ECG data through the gating network, the MoE framework ensures that each ECG exam is analyzed by the most relevant expert network, leading to more precise and reliable classifications, surpassing the approach based on flat classes.

2. Related work

In the realm of electrocardiogram (ECG) signal classification, deep learning has demonstrated significant potential. One-dimensional convolutional neural networks (1D-CNNs) have achieved remarkable success, such as 99% accuracy in cardiac arrhythmia classification [4]. Additionally, deep neural networks applied to 12-lead ECG data have outperformed cardiology residents in diagnosing heart conditions [5]. Ensemble methods, combining various architectures such as CNN-LSTM and RRHOS-LSTM, have further improved ECG classification accuracy to 95.81% [6], addressing challenges like class imbalance.

The Mixture of Experts (MoE) model, introduced by [7], utilizes a gating network to dynamically assign inputs to specialized models for enhanced accuracy and efficiency. Initially used in natural language processing, MoE frameworks have been adapted for transformer models and, in some works, incorporated into convolutional neural networks (CNNs) for image classification [8]. Our work builds on these advances by integrating sparse MoE framework into CNNs models for ECG classification, aiming to improve diagnostic accuracy and interpretability by leveraging specialized expert models for distinct cardiac conditions.

3. Mixture of Experts (MoE)

A Mixture of Experts (MoE) framework is a robust approach that integrates multiple specialized models, or experts, each focusing on a distinct aspect of a problem [9]. Within this framework, a gating network dynamically assigns input data to the most appropriate expert or combination of experts based on the data's characteristics. This allows the MoE to break down complex tasks into smaller, more manageable sub-tasks, with each expert specializing on a particular area.

In our work, leveraging prior knowledge of the hierarchical organization of cardiac conditions, we categorized the six classes present in the dataset into three distinct superclasses: *blockage*, *rhythm*, and *normal*. *Blockage* includes 1dAVb, RBBB, and LBBB. These conditions are characterized by delays or interruptions in the electrical conduction pathways of the heart, which can lead to prolonged intervals or altered waveforms in the ECG signal. The *rhythm* superclass comprises SB, AF and ST. These conditions primarily involve abnormalities in the heart's electrical rhythm, resulting in slower or faster than normal heart rates or irregular atrial activity. The *normal* category includes cases without significant electrical abnormalities.

This categorization allows us to reflect the underlying medical relationships between the conditions more accurately. Consequently, the MoE model is designed to include one gating network and three expert networks, with each expert network initially assigned to specialize in one of these superclasses.

Our MoE approach employs the same residual convolutional neural network (ResNet) architecture [5] for both the gating and experts networks. This ResNet architecture has consistently demonstrated high predictive performance in scenarios involving flat classes, making it a reliable choice for ECG classification tasks. Although identical architectures are used across multiple experts, the gating mechanism ensures that each expert network specializes in analyzing distinct types of ECG patterns with unique characteristics. The final classification is achieved by aggregating the predictions from all experts, weighted according to the gating network's output. This output assigns higher weights to predictions from experts that are deemed more relevant for the specific input data. By prioritizing the most appropriate expert insights, this weighted combination enhances both the accuracy and reliability of the ECG classification process.

Training process: To develop our MoE model, we first trained the gating network separately to classify the ECG signals into the three predefined superclasses. This initial training step allowed the gating network to learn how to effectively distinguish between these broad categories of cardiac conditions.

Following this first phase, we integrated the pre-trained

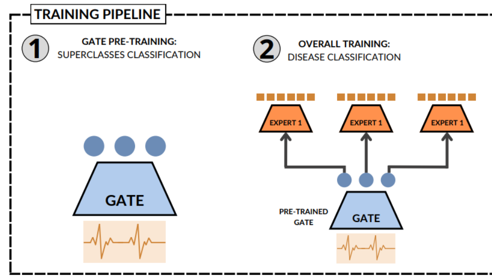

Figure 1. MoE training process.

gating network into the MoE framework, which has three expert networks. These expert networks were initialized with random weights, allowing them to develop unique specializations during subsequent training. In the MoE configuration, the gating and expert networks were trained simultaneously on the ECG classification task. During training, the gating network's predictions were used to weight the logits output from each expert, ensuring that the most relevant expert had a greater influence on the final classification decision. This weighted aggregation determined the final output and influenced the gradient updates during backpropagation, dynamically adjusting the contribution of each expert based on the gating network's predictions. The overall training pipeline can be seen in Figure 1.

Experts Collapse: One challenge we faced was the issue of *dying experts*.

As outlined in [8], an expert is considered "dead" if it receives less than 1% average importance, which is determined by summing the gate outputs for that expert across all batches. Expert death often occurs due to an imbalance in the gating mechanism, where the gate overly favors certain experts, especially early in training, due to initial parameter settings. This can lead to a positive feedback loop where favored experts improve rapidly because they receive more data, while others are underutilized and fail to learn effectively. This imbalance not only wastes computational resources but also reduces the diversity and robustness of the model, making it less capable of generalizing to unseen data.

To avoid this problem, in our work, we adopt the mean importance constraint proposed in [8], which is a mechanism designed to balance the utilization of experts based on their importance over time. Specifically, the mean importance of each expert E_i is calculated as the average importance assigned to that expert up to a given time step t. This importance is determined by summing the gate outputs for E_i across all input batches. The constraint then operates by zeroing out the weight of an expert, effectively deactivating it, if its mean importance surpasses a predefined threshold m_{mean}. By implementing this constraint, we ensure that no single expert dominates the model and helps to maintain active participation from all experts and enhances the overall robustness and effectiveness of the MoE model.

Metrics	ResNet	MoE	MoE Gate	MoE GateC
Accuracy	0.9883	0.9865	0.9874	**0.9914**
Precision	**0.9692**	0.9557	0.9610	0.9368
Recall	0.6620	0.6202	0.6369	**0.7833**
F1-score	0.7602	0.7187	0.7451	**0.8496**

Table 1. Mean metrics (aggregating all classes) for the classification task of six cardiac conditions.

4. Computational Results

Experimental Setup: The MoE was developed using the publicly available CODE-15 dataset, that represents 15% of the original CODE database [5] and includes 345,779 exams from 233,770 patients. In this dataset, ECGs are classified according to the same six types of disorders we previously introduced. These conditions are vital in clinical cardiology practice and can be associated with increased risks of cardiovascular events such as stroke, heart failure, and, in severe cases, sudden cardiac death. The model's performance was evaluated in the CODE-TEST set [5]. This set has 827 exams and was labeled following a rigorous consensus process involving two or three cardiology specialists.

When evaluating the models on CODE TEST, we performed 1,000 bootstrap resampling iterations for robustness [10]. Each output node class threshold was determined using the validation partition for all models, and the results were compared.

To show the benefits of employing a mixture of experts (MoE) architecture, we compared the results obtained using the MoE model against those achieved with the standard ResNet model trained on flat classes. Additionally, to assess the benefits of incorporating domain-specific medical knowledge, we evaluated MoE without the pre-training phase for the gating network. In this scenario, the gating network was allowed to learn the hierarchical structure of the classes independently during the overall training process. By comparing these approaches, we can understand the impact of domain-specific knowledge on the performance of the model and the importance of pre-training in guiding the gate predictions.

Results: Table 1 shows the performance of MoE under different conditions: without pre-training the gating network **(MoE)**, pre-training the gating network **(MoE Gate)**, and a second variation pre-training the gating network with the Mean Importance Constraint **(MoE Gate C)**).

Comparing the results in Table 1, we observe that the mean F1 did not improve with MoE and MoE Gate if compared to the ResNet. To further investigate, we plotted the percentage of each class classified by each expert through the gate to understanding the internal workings of the model in Figure 2.

We observed that in the MoE (Figure 2(a)), the gate assigns a value to only one of the experts, which explains the results close to the baseline, as only one model is used. In contrast, in the MoE Gate without constraints (Figure 2(b)), the gate uses more than one expert. However, due to data imbalance, the rhythm expert "dies". As a result, the block expert is responsible for classifying all six diseases, which, again, is very similar to what the baseline model does.

This analysis clearly shows that when we add constraints, the results improved significantly as different experts can now classify examples from their specialization class (see Figure 2(c)). As expected, block-related diseases are primarily assigned by the block expert, while rhythm-related are predominantly classified by the rhythm expert.

The results obtained by Moe GateC for each of the six classes are reported in Table 2. MoE GateC showed superior results for most metrics. F1, in particular, is more balanced due to the applied constraints, which enables better detection of classes like "1dAVb", which were previously neglected.

A significant improvement is observed in recall, which is substantially higher in MoE GateC. Recall, or sensitivity, measures the proportion of actual positive cases the model correctly identifies. In medical applications, high recall is particularly critical because it minimizes the likelihood of missing true positive cases, such as patients with a specific condition. This enhancement in recall is crucial in the medical field, where the consequences of false negatives – failing to identify a condition – are often far more severe than those of false positives.

5. Conclusion

This work introduces a Mixture of Experts (MoE) model that utilizes cardiologists' prior knowledge to advance ECG disorder classification. By hierarchically organizing disorders, incorporating a gating network, and applying constraints to address data imbalance, our approach achieves state-of-the-art results on the CODE-15 dataset. The MoE framework enhances both F1 scores and recall, a critical metric in medicine that ensures fewer true cases are missed, facilitating better detection and timely patient intervention.

Future work will explore alternative constraints and integrate various models into the experts and gating network. Additionally, we aim to apply our model, or its adaptations, to other ECG tasks with more complex hierarchies.

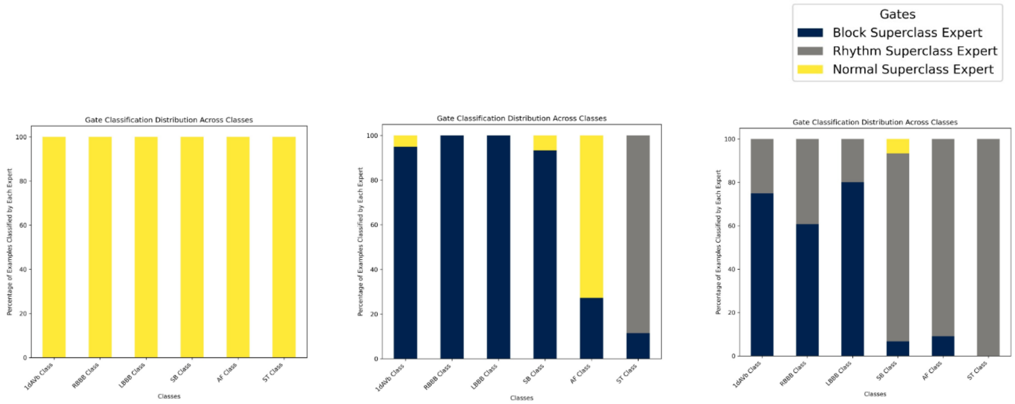

Figure 2. Percentage of each class classified by each expert by the gate from left to right: (a) MoE, (b) MoE Gate, (c) MoE GateC.

Classes	Accuracy		Precision		Recall		F1 Score	
	ResNet	MoE GateC	ResNet	MoE GateC	ResNet	MoE GateC	ResNet	MoE GateC
1dAVb	0.981	**0.991**	**1.000**	0.922	0.425	**0.787**	0.591	**0.847**
RBBB	0.987	**0.991**	**0.971**	0.944	0.656	**0.792**	0.751	**0.859**
LBBB	0.990	**0.992**	**0.981**	0.963	0.725	**0.804**	0.809	**0.875**
SB	0.990	**0.991**	**0.949**	0.919	0.714	**0.778**	0.794	**0.839**
AF	0.991	**0.992**	**0.959**	0.936	0.708	**0.758**	0.796	**0.832**
ST	0.992	**0.992**	**0.956**	0.937	0.744	**0.782**	0.819	**0.846**

Table 2. Classification metrics for each class, comparing the baseline model with the optimized MoE version that pre-trains the gate and applies a mean constraint, identified as the best-performing MoE variant.

Acknowledgments

This work was partially funded by CNPq, CAPES, and FAPEMIG, as well as CIIA-Saúde and IAIA - INCT on AI. We also thank the Telehealth Center of Minas Gerais for access to the data and the productive discussions within the scope of this work.

References

[1] Federation WH. World heart report 2024. https://world-heart-federation.org/resource/world-heart-report-2024. Accessed: Aug. 20, 2024.

[2] Houssein EH, Kilany M, Hassanien AE. Ecg signals classification: a review. International Journal of Intelligent Engineering Informatics 2017;5(4):376–396.

[3] Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ecg arrhythmia classification. Expert Systems with Applications X 2020; 7:100033.

[4] Abdullah AA, Ali W, Abdullah T, Malebary S. Classifying cardiac arrhythmia from ecg signal using 1d cnn deep learning model. Mathematics 01 2023;11:562.

[5] Ribeiro AH, Ribeiro MH, Paixao GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira Jr. W, Schon TB, Ribeiro ALP. Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature Communications 2020;11(1):1760.

[6] Essa E, Xie X. An ensemble of deep learning-based multimodel for ecg heartbeats arrhythmia classification. IEEE Access 07 2021;PP:1–1.

[7] Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE. Adaptive mixtures of local experts. Neural Computation 1991;3:79–87.

[8] Pavlitska S, Hubschneider C, Struppek L, Zollner JM. Sparsely-gated mixture-of-expert layers for cnn interpretability. 2023 International Joint Conference on Neural Networks IJCNN 2022;1–10.

[9] Jordan MI, Jacobs RA. Hierarchical mixtures of experts and the em algorithm. Neural computation 1994;6(2):181–214.

[10] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Chapman & Hall/CRC, 1993.

Address for correspondence:

Diogo Tuler Chaves
DCC–ICEx–UFMG - Pampulha Belo Horizonte, MG 31270-901
Brazil
diogo.tuler@gmail.com