

Diagnóstico Hierárquico em Cardiologia Utilizando um Modelo de Mistura de Especialistas e Conhecimento Prévio de Cardiologistas

Diogo Chaves¹, Pedro Dutenhefner¹, Gisele L. Pappa¹
Antônio Ribeiro¹, Gabriela Paixão¹, Wagner Meira Jr.¹

¹Universidade Federal de Minas Gerais

(diogochaves,pedroroblesduten,glpappa,meira)@dcc.ufmg.br;tom@hc.ufmg.br;gabimiana@gmail.com

ABSTRACT

Although several predictive models exist for identifying cardiovascular disorders in electrocardiograms (ECGs), many disregard the well-established knowledge from medical literature, forcing the models to relearn known patterns. We propose a Mixture of Experts (MoE) model that incorporates cardiologists' expertise to hierarchically organize the expert models. We classify six physician-defined electrocardiographic abnormalities, covering both rhythm and conduction disorders. Trained on the CODE-15 dataset, our model outperformed a widely used baseline in the literature for this task, achieving an F1-score of 0.84 compared to 0.77 for the reference model. These results highlight the relevance of architectures guided by medical knowledge and demonstrate significant improvements over established approaches, underscoring the potential of our proposal to advance ECG classification performance.

KEYWORDS

Deep Learning, Mixture of Experts (MoE), Hierarchical Classification, Electrocardiogram (ECG)

1 INTRODUÇÃO

O eletrocardiograma (ECG) é um exame não invasivo amplamente utilizado para identificar distúrbios cardíacos, sendo fundamental no diagnóstico clínico, dado que doenças cardiovasculares permanecem entre as principais causas de morte em 2024 [7]. Nos últimos anos, o avanço do aprendizado profundo possibilitou o desenvolvimento de modelos robustos para a análise automática de ECGs [8].

Entretanto, abordagens tradicionais de aprendizado profundo apresentam limitações, como a classificação independente das condições, sem explorar a organização hierárquica dos diagnósticos médicos [4]. Para superar esse desafio, propomos um modelo hierárquico baseado em *Mixture of Experts* (MoE), no qual uma rede de (*gating*) encaminha cada exame para especialistas dedicados a grupos específicos de anormalidades. Assim, a classificação incorpora conhecimento clínico e explora as relações entre as classes.

As principais contribuições deste trabalho são:

- (1) um mapeamento das abordagens de classificação de ECGs, evidenciando os benefícios da organização hierárquica;

- (2) a proposta de uma nova arquitetura hierárquica baseada em Mixture of Experts, que introduz um mecanismo de *gating* capaz de explorar a estrutura do problema e direcionar dinamicamente os sinais para especialistas dedicados, resultando em maior especialização e ganho de desempenho;
- (3) uma análise experimental comparando a classificação hierárquica com métodos planos, destacando ganhos em métricas como precisão, revocação e F1-score;
- (4) a investigação das representações internas do *gating network*, confirmando a captura da estrutura hierárquica.

Os resultados mostram que a abordagem hierárquica baseada em MoE supera a classificação plana, proporcionando previsões mais precisas e interpretáveis para apoio ao diagnóstico clínico.

Todas as implementações e execuções dos experimentos foram realizadas pelo aluno Diogo Tuler Chaves, sob orientação dos professores Wagner Meira Jr. e Gisele L. Pappa, do Departamento de Ciência da Computação da UFMG, e com o apoio dos professores Gabriela Paixão e Antônio Ribeiro, do Departamento de Medicina da UFMG. A concepção do projeto e a análise dos resultados foram desenvolvidas em colaboração entre o aluno e os professores, contando também com a participação do mestrando Pedro Dutenhefner, do Programa de Pós-Graduação em Ciência da Computação. Este trabalho já resultou na publicação do artigo [13] na conferência Computing in Cardiology.

2 TRABALHOS RELACIONADOS

A análise automática de sinais eletrocardiográficos (ECG) tem recebido crescente atenção da comunidade científica. Nos últimos anos, redes neurais convolucionais unidimensionais (1D-CNNs) destacaram-se como abordagens particularmente eficazes [1].

Além das CNNs tradicionais, estratégias de *ensemble* têm se mostrado úteis para lidar com a diversidade e complexidade dos padrões de ECG. Abordagens híbridas como CNN-LSTM, que exploram tanto representações espaciais quanto temporais, atingiram desempenho superior ao de 1D-CNNs [6]. Mais recentemente, a literatura passou a explorar arquiteturas modernas e estratégias multimodais. Dutenhefner et al. (2024a) introduziram o HiT, um transformer hierárquico que combina convoluções e mecanismos de atenção local [3].

Nesse cenário, destaca-se também o interesse crescente em arquiteturas do tipo *Mixture of Experts* (MoE). O MoE combina uma rede de (*gating*) com especialistas independentes, promovendo maior eficiência e modularidade no aprendizado. Embora amplamente utilizado em processamento de linguagem natural e, mais recentemente, em transformers e CNNs de larga escala [10], sua aplicação no domínio do ECG ainda é pouco explorada. Alguns trabalhos

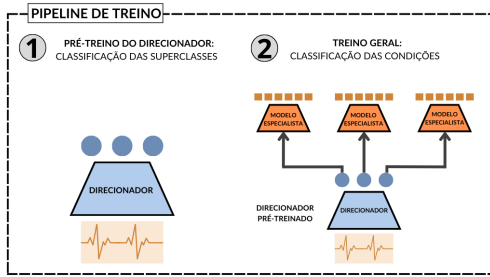


Figura 1: Processo de treinamento do modelo MoE.

iniciais sugerem seu potencial: por exemplo, a classificação de batimentos a partir da combinação de diferentes atributos [2], o uso de redes modulares com aprendizado negativamente correlacionado para identificação de arritmias [12], e sistemas baseados em especialistas para segmentação de ondas e detecção de padrões raros. Apesar desses avanços, observa-se que a maioria dessas iniciativas permanece restrita a tarefas pontuais, com pouca ênfase na interpretabilidade clínica e na exploração hierárquica dos diagnósticos.

Propomos um framework hierárquico baseado em Mixture of Experts (MoE) para classificação de ECGs, onde especialistas são organizados por grupos de anormalidades. Essa estrutura reflete o raciocínio clínico, aumentando a acurácia e a interpretabilidade do diagnóstico automático.

3 MÉTODO PROPOSTO

O framework *Mixture of Experts* (MoE) é uma abordagem robusta que integra múltiplos modelos especializados, ou modelos especialistas, cada um focado em um aspecto distinto de um problema [9]. Dentro desse framework, uma rede de *gating* atribui dinamicamente os dados de entrada ao modelo especialista mais adequado, ou a uma combinação de modelos especialistas, com base nas características dos dados. Isso permite que o MoE divida tarefas complexas em sub-tarefas menores e mais gerenciáveis, com cada modelo especialista se especializando em uma área específica.

Em nosso trabalho, aproveitando o conhecimento prévio sobre a organização hierárquica das condições cardíacas, categorizamos as seis classes presentes no conjunto de dados em três superclasses distintas: *bloqueios*, *distúrbios de ritmo* e *normal*. A categoria *bloqueios* inclui bloqueio atrioventricular de primeiro grau (1dAVb), bloqueio de ramo direito (RBBB) e bloqueio de ramo esquerdo (LBBB), anormalidades caracterizadas por atrasos ou interrupções nas vias de condução elétrica do coração, resultando em intervalos prolongados e/ou alterações nas morfologias de onda do sinal de ECG. A superclass *distúrbios de ritmo* compreende bradicardia sinusal (SB), fibrilação atrial (AF) e taquicardia sinusal (ST), condições que envolvem anormalidades do ritmo cardíaco do coração, resultando em frequência cardíaca sinusal lentificada, atividade atrial desorganizada, rápida e irregular e frequência cardíaca sinusal acelerada, respectivamente. A categoria *normal* inclui casos sem anormalidades elétricas significativas.

Essa categorização permite refletir com mais precisão as relações médicas subjacentes entre as condições. Consequentemente, o modelo MoE foi projetado para incluir uma rede de *gating* e três

redes modelos especialistas, cada uma inicialmente atribuída a uma dessas superclasses.

Nossa abordagem MoE emprega a arquitetura de rede neural convolucional residual (*ResNet*) [11] tanto para a rede de *gating* quanto para os modelos especialistas. A *ResNet* utilizada representa o estado da arte consolidado para a tarefa de classificação de ECGs no período em que esta pesquisa foi conduzida, sendo amplamente adotada como baseline em estudos anteriores [1, 11]. Essa escolha permite uma comparação consistente com trabalhos prévios e fornece um ponto de referência confiável para avaliar melhorias introduzidas pela arquitetura MoE. Embora a mesma arquitetura seja utilizada em múltiplos modelos especialistas, o mecanismo de *gating* garante que cada modelo especialista se especialize na análise de padrões distintos de ECG, enquanto a classificação final é obtida agregando as previsões ponderadas conforme os pesos atribuídos pelo *gating*. Esta configuração possibilita uma avaliação objetiva do impacto da nova arquitetura hierárquica, diferenciando os efeitos do MoE daqueles atribuíveis à própria escolha da rede base.

3.1 Processo de treinamento

Para desenvolver nosso modelo MoE, primeiro treinamos a rede de *gating* separadamente para classificar os sinais de ECG nas três superclasses predefinidas. Essa etapa inicial permitiu que a rede de *gating* aprendesse a distinguir efetivamente entre essas categorias amplas de anormalidades eletrocardiográficas.

Após essa primeira fase, integramos a rede de *gating* pré-treinada ao framework MoE, composto por três redes modelos especialistas. Essas redes especialistas foram inicializadas com pesos aleatórios, permitindo que desenvolvessem especializações únicas durante o treinamento subsequente. Na configuração MoE, a rede de *gating* e modelos especialistas foram treinadas simultaneamente na tarefa de classificação de ECGs. Durante o treinamento, as previsões da rede de *gating* foram usadas para ponderar as saídas (*logits*) de cada modelo especialista, garantindo que o modelo especialista mais relevante tivesse maior influência na decisão final de classificação. Essa agregação ponderada determinava a saída final e influenciava as atualizações dos gradientes durante a retropropagação, ajustando dinamicamente a contribuição de cada modelo especialista com base nas previsões da rede de *gating*. O pipeline geral de treinamento é ilustrado na Figura 1.

Métricas	ResNet	MoE	MoE Gate	MoE GateC
Acurácia	0.9883	0.9865	0.9874	0.9914
Precisão	0.9692	0.9557	0.9610	0.9368
Sensibilidade	0.6620	0.6202	0.6369	0.7833
F1-score	0.7602	0.7187	0.7451	0.8496

Tabela 1: Métricas médias (agregando todas as classes) para a tarefa de classificação de seis condições cardíacas.

Classes	Acurácia		Precisão		Sensibilidade		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

Tabela 2: Métricas de classificação por classe para ResNet e MoE GateC.

3.2 Colapso dos modelos especialistas

A inatividade de modelos especialistas, ou problema dos *modelos especialistas inativos* (*dying experts*), é comum em Mixture of Experts. Um especialista é considerado "inativo" se recebe, em média, menos de 1% de importância, definida pela soma das saídas da rede de *gating* ao longo do treinamento [10]. Isso ocorre devido ao desequilíbrio no *gating*, onde certos especialistas são favorecidos desde o início, criando um ciclo em que modelos subutilizados não aprendem efetivamente. Esse fenômeno desperdiça recursos e reduz a diversidade e robustez do modelo.

Para evitar isso, adotamos a restrição de importância média proposta em [10]. A importância média de cada especialista E_i é calculada como a média da importância atribuída até um instante t . A restrição desativa um especialista caso sua importância média ultrapasse um limite m_{mean} , garantindo que nenhum especialista domine o modelo e promovendo a participação de todos, aumentando a robustez e eficácia do MoE.

4 ANÁLISE EXPERIMENTAL

4.1 Configurações dos experimentos

O modelo proposto foi desenvolvido com o conjunto CODE-15 [11], que contém 345.779 exames de 233.770 pacientes, classificados em seis distúrbios eletrocardiográficos de relevância clínica. A avaliação foi realizada no conjunto CODE-TEST, composto por 827 exames rotulados por consenso entre especialistas em cardiologia, onde realizamos 1.000 iterações de reamostragem bootstrap para garantir robustez [5]. O limiar de classificação para cada classe de saída foi determinado utilizando a partição de validação para todos os modelos, e os resultados foram comparados.

Foram utilizadas as métricas de acurácia, precisão, revocação e F1-score, além da análise das saídas da *gating network* do MoE, garantindo uma avaliação detalhada do desempenho hierárquico em comparação aos métodos tradicionais.

4.2 Resultados

Para demonstrar os benefícios da arquitetura Mixture of Experts (MoE), comparamos os resultados obtidos pelo MoE com os resultados da ResNet padrão, treinada com classes planas. Além disso, para avaliar os benefícios da incorporação do conhecimento médico específico, testamos o MoE sem a fase de pré-treinamento da rede de *gating*. Nessa configuração, a rede de *gating* foi treinada de forma independente para aprender a estrutura hierárquica das classes durante o processo geral de treinamento. Comparando essas abordagens, buscamos entender o impacto do conhecimento especializado no desempenho do modelo e a importância do pré-treinamento para guiar as previsões do *gating*.

A Tabela 1 apresenta o desempenho do modelo MoE em diferentes condições: sem pré-treinamento da rede de *gating* (**MoE**), com pré-treinamento da rede de *gating* (**MoE Gate**), e uma variação que inclui a restrição de Importância Média (**MoE GateC**).

Comparando os resultados da Tabela 1, observamos que a média do F1-score não melhorou com MoE e MoE Gate em relação à ResNet. Para investigar melhor esse comportamento, plotamos a porcentagem de cada classe classificada por cada especialista através do *gating*, conforme mostrado na Figura 2.

Na Figura 2(a), observa-se que, no MoE, o *gating* atribui um valor apenas a um dos especialistas, explicando os resultados semelhantes ao baseline, pois apenas um modelo está sendo utilizado. Em contrapartida, na Figura 2(b), o *gating* no MoE Gate sem restrição utiliza mais de um especialista, mas, devido ao desequilíbrio de dados, o especialista de ritmos "morre". Como resultado, o especialista de bloqueios fica responsável por classificar todas as seis anormalidades, aproximando-se novamente do comportamento do modelo baseline.

Essa análise mostra claramente que, ao adicionar restrições, os resultados melhoram significativamente, pois diferentes especialistas agora conseguem classificar exemplos de suas respectivas classes de especialização (veja Figura 2(c)). Conforme esperado, as doenças relacionadas a bloqueios são predominantemente classificadas pelo especialista de bloqueios, enquanto as doenças relacionadas a ritmos são classificadas pelo especialista de ritmos.

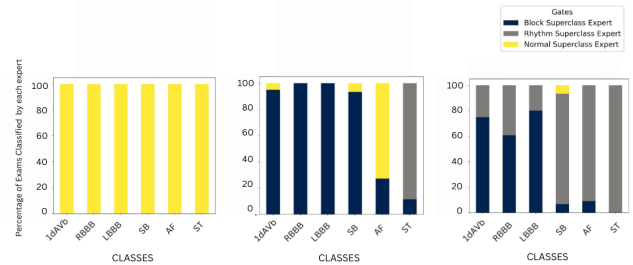


Figura 2: Porcentagem de cada classe classificada por cada especialista da gate, da esquerda para a direita: (a) MoE, (b) MoE Gate, (c) MoE GateC.

Para investigar se o espaço latente gerado pela camada antes da saída do *gating* preserva a estrutura das superclasses, realizamos uma análise de redução de dimensionalidade.

A Figura 3(a) apresenta a projeção t-SNE do espaço latente produzido pelo *gating* para todos os exames, com cores representando as superclasses reais. Observamos que, mesmo após o treinamento do MoE, as amostras de cada superclasse continuam a se agrupar em regiões específicas. Isso sugere que o *gating* mantém informações estruturais relevantes sobre as superclasses. A Figura 3(b) exibe a mesma projeção, mas utilizando um subconjunto balanceado entre as classes. Nesse caso, a separação entre os grupos se torna ainda mais evidente, indicando que o desbalanceamento dos dados impacta a organização das classes no espaço latente.

Esses resultados indicam que o espaço latente do *gating* não apenas separa os especialistas, mas também preserva a estrutura hierárquica das anormalidades do ECG. Dessa forma, o *gating* não age apenas como um mecanismo de direcionamento, mas organiza as informações de forma coerente com a taxonomia das alterações eletrocardiográficas. Esse achado destaca o impacto positivo da incorporação de conhecimento prévio na arquitetura MoE e reforça a importância do pré-treinamento da gate.

Os resultados obtidos pelo MoE GateC para cada uma das seis classes são apresentados na Tabela 2. O MoE GateC demonstrou desempenho superior na maioria das métricas, com um F1-score mais equilibrado devido às restrições aplicadas, permitindo uma melhor detecção de classes como "1dAVb", anteriormente negligenciadas.

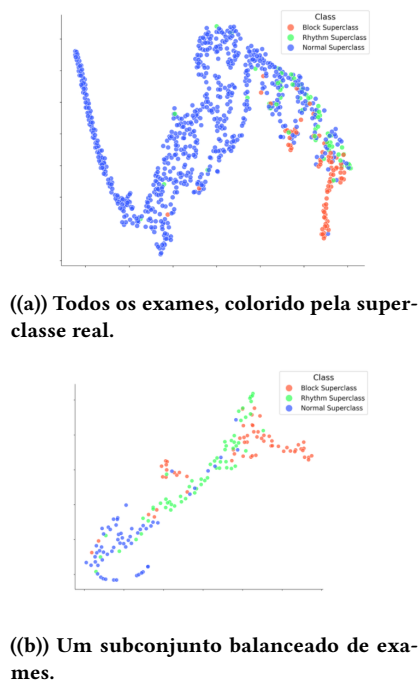


Figura 3: Projeção t-SNE do espaço latente gerado pelo *gating*

Um avanço significativo é observado na sensibilidade, que é substancialmente maior no MoE GateC. A sensibilidade, ou sensibilidade, mede a proporção de casos positivos reais corretamente identificados pelo modelo. Em aplicações médicas, uma alta sensibilidade é especialmente crítica, pois minimiza a probabilidade de não identificar corretamente pacientes com uma determinada condição. Essa melhoria na sensibilidade é crucial na área médica, onde as consequências de falsos negativos - não identificar uma alteração - são frequentemente muito mais graves do que as de falsos positivos.

É importante ressaltar que a arquitetura proposta do MoE representa uma abordagem geral e extensível, capaz de incorporar diferentes modelos como especialistas no processo de classificação. Dessa forma, a estrutura hierárquica não está restrita a um tipo específico de rede, podendo integrar desde redes convolucionais profundas até arquiteturas baseadas em transformers, conforme as características do problema ou do conjunto de dados. Essa flexibilidade evidencia o caráter modular e adaptativo do MoE, destacando seu potencial como uma solução robusta para tarefas de classificação de ECGs e, potencialmente, para outros domínios biomédicos que se beneficiem de estruturas hierárquicas de decisão.

O tempo de inferência do MoE é equivalente ao do modelo ResNet padrão. Já custo computacional total é maior devido à presença de múltiplos especialistas, mas essa sobrecarga é facilmente gerenciável, pois a arquitetura MoE foi projetada para paralelização, permitindo que os especialistas sejam avaliados simultaneamente sem impactar significativamente a velocidade de predição.

5 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, apresentamos um modelo MoE que integra conhecimento prévio de cardiologistas para aprimorar a classificação de

anormalidades do ECG. Ao organizar hierarquicamente as alterações, incorporar uma rede de *gating* inovadora e aplicar restrições para lidar com o desequilíbrio de dados, nossa abordagem demonstra melhorias significativas em métricas-chave.

Além de superar o modelo de referência consolidado na literatura, o framework MoE proposto evidencia a vantagem de arquiteturas hierárquicas capazes de direcionar dinamicamente diferentes especialistas, permitindo especialização e flexibilidade na escolha de modelos para cada subtarefa. Essa característica torna a abordagem adaptável a diferentes tipos de redes, hierarquias de classes e conjuntos de dados, ampliando seu potencial de aplicação em problemas biomédicos complexos.

Trabalhos futuros incluem explorar restrições alternativas, integrar diversas arquiteturas como especialistas e na rede de *gating*, e aplicar o modelo a tarefas de ECG com hierarquias mais complexas. Também planejamos análises ablatórias adicionais para quantificar de forma mais abrangente o impacto de cada componente da arquitetura na performance final.

6 AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo CNPq, CAPES e FAPEMIG, além do CIA-Saúde e IAIA - INCT em IA. Também agradecemos ao Centro de Telessaúde de Minas Gerais pelo acesso aos dados e pelas discussões produtivas no âmbito deste trabalho.

REFERÊNCIAS

- [1] Adel. A. Abdullah, Waleed Ali, Talal Abdullah, and Sharaf Malebary. 2023. Classifying Cardiac Arrhythmia from ECG Signal Using 1D CNN Deep Learning Model. *Mathematics* 11 (01 2023), 562. <https://doi.org/10.3390/math11030562>
- [2] Philip De Chazal and Naimul M. Sadr. 2014. ECG Beats Classification Using Mixture of Features. *International Scholarly Research Notices* (2014). <https://doi.org/10.1155/2014/876059>
- [3] Pedro Robles Dutenehner, Turi Andrade Vasconcelos Rezende, Gisele Lobo Pappa, Gabriela Miana de Matos Paixão, Antônio Luiz Pinho Ribeiro, and Wagner Meira Jr. 2024. A hierarchical transformer for electrocardiogram classification and diagnosis. *J. Health Inform.* 16, Especial (2024), 1–11.
- [4] Zahra Ebrahimi, Mohammad Loni, Masoud Daneshalab, and Arash Gharehbaghi. 2020. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X* 7 (2020), 100033.
- [5] Bradley Efron and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
- [6] Ehab Essa and Xianghua Xie. 2021. An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification. *IEEE Access PP* (07 2021), 1–1. <https://doi.org/10.1109/ACCESS.2021.3098986>
- [7] World Heart Federation. 2024. World Heart Report 2024. <https://world-heart-federation.org/resource/world-heart-report-2024>. Accessed: Aug. 20, 2024.
- [8] Essam H Houssein, Moataz Kilany, and Aboul Ella Hassanien. 2017. ECG signals classification: a review. *International Journal of Intelligent Engineering Informatics* 5, 4 (2017), 376–396.
- [9] Michael I Jordan and Robert A Jacobs. 1994. Hierarchical mixtures of experts and the EM algorithm. *Neural computation* 6, 2 (1994), 181–214.
- [10] Svetlana Pavlitska, Christian Hubschneider, Lukas Struppek, and J. Marius Zollner. 2022. Sparsely-gated Mixture-of-Expert Layers for CNN Interpretability. *2023 International Joint Conference on Neural Networks (IJCNN)* (2022), 1–10.
- [11] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixao, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. S. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira Jr., T. B. Schon, and A. L. P. Ribeiro. 2020. Automatic Diagnosis of the 12-Lead ECG Using a Deep Neural Network. *Nature Communications* 11, 1 (2020), 1760. <https://doi.org/10.1038/s41467-020-15432-4>
- [12] Abdulhamit Subasi and M Ismail Gursay. 2017. Classification of ECG arrhythmia by a modular neural network based on Mixture of Experts and Negatively Correlated Learning. *Journal of Medical Systems* (2017). <https://doi.org/10.1007/s10916-012-9893-6>
- [13] Diogo Tuler, Pedro Robles Dutenehner, Jose Geraldo Fernandes, Turi Rezende, Gabriel Lemos, Gisele L. Pappa, Gabriela Paixão, Antônio Ribeiro, and Wagner Meira Jr. 2024. Leveraging Cardiologists Prior-Knowledge and a Mixture of Experts Model for Hierarchically Predicting ECG Disorders. *Computing in Cardiology* (2024).