

A hierarchical transformer for electrocardiogram classification and diagnosis

Um transformer hierárquico para classificação e diagnóstico de eletrocardiograma

Un transformer jerárquico para clasificación y diagnóstico de electrocardiograma

Pedro Robles Dutenhfner¹, Turi Andrade Vasconcelos Rezende¹, Gisele Lobo Pappa², Gabriela Miana de Matos Paixão³, Antônio Luiz Pinho Ribeiro⁴, Wagner Meira Jr.²

1 Aluno de graduação, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brasil.

2 Professor Doutor, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brasil.

3 Doutor, Centro de telessaúde do Hospital das Clínicas da Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brasil.

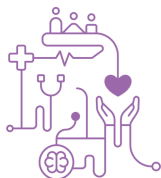
4 Professor Doutor, Centro de telessaúde do Hospital das Clínicas da Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brasil.

Autor correspondente: Pedro Robles Dutenhfner
E-mail: pedroroblesduten@gmail.com

Resumo

Objetivo: O eletrocardiograma (ECG) é uma importante ferramenta utilizada para avaliar condições cardíacas, e o avanço da inteligência artificial permitiu progresso na análise automática de ECGs. Com o objetivo de melhorar o diagnóstico automático e suporte à decisão médica, este artigo apresenta um novo modelo de transformer hierárquico (HiT) para a classificação de ECGs de 12 derivações. **Método:** O modelo HiT integra blocos convolucionais e de transformers - especialmente projetados com mecanismos de atenção local - direcionando o aprendizado de características locais e globais dos sinais de ECG. **Resultados:** Utilizando um subconjunto do CODE, uma abrangente base de dados de ECG do Brasil, o modelo foi desenvolvido para reconhecer seis condições cardíacas e apresentou valor médio de f1-score acima de 0.84, superando o estado da arte para os mesmos dados. **Conclusão:** Assim, demonstra-se o potencial de um transformer hierárquico multinível para um diagnóstico automático mais acurado de doenças cardíacas.

Descritores: Redes neurais; Diagnóstico Automático; Eletrocardiograma



Abstract

Objective: Electrocardiogram (ECG) is an important tool to assess cardiac conditions. The advancement of artificial intelligence has enabled progress in the automatic analysis of ECGs. Aiming to improve the predictive performance of automatic diagnosis, this paper presents a new hierarchical transformer model (HiT) for classifying 12-lead ECGs. **Method:** The HiT model integrates convolutional and transformer blocks - specifically designed with local attention mechanisms - to guide the learning of local and global features of ECG signals. **Results:** Using a subset of CODE, a broad ECG database from Brazil, the model was developed for classifying six cardiac conditions and achieved an average f1-score over 0.84, surpassing the state of the art for the same data. **Conclusion:** Therefore, this work demonstrates the potential of a multi-level hierarchical transformer for more accurate automatic diagnosis of heart diseases.

Keywords: Neural networks; Automatic Diagnosis; Electrocardiogram

Resumen

Objetivo: El electrocardiograma (ECG) es una herramienta importante utilizada para evaluar las condiciones cardíacas. El avance de la inteligencia artificial ha permitido progresos en el análisis automático de los ECG. Con el objetivo de mejorar el diagnóstico automático, este artículo presenta un nuevo modelo de transformer jerárquico (HiT) para la clasificación de ECG de 12 derivaciones. **Método:** El modelo HiT integra bloques convolucionales y de transformers, diseñados con mecanismos de atención local, para promover el aprendizaje de características locales y globales de los ECG. **Resultados:** Utilizando un subconjunto de CODE, una amplia base de datos de ECG de Brasil, el modelo fue desarrollado para la clasificación de seis condiciones cardíacas y logró un f1-score promedio superior a 0.84, superando el estado del arte para los mismos datos. **Conclusión:** Por consiguiente, este trabajo demuestra el potencial de un transformer jerárquico multinivel para un diagnóstico automático más preciso de enfermedades cardíacas.

Descriptores: Redes Neuronales; Diagnóstico Automático; Electrocardiograma



Introdução

As doenças cardiovasculares (DCVs) representam a principal causa de morte global, e corresponderam a 32% dos óbitos no mundo em 2019, conforme dados da Organização Mundial da Saúde (OMS) ⁽¹⁾. Nesse contexto, os eletrocardiogramas (ECGs), exames simples e não invasivos, desempenham um papel crucial no diagnóstico e seguimento das doenças cardiovasculares, sendo indicados para a avaliação de cardiopatias. Além do cenário clínico tradicional, os ECGs são fundamentais, também, para avanços na área da saúde digital e telessaúde, especialmente no desenvolvimento de ferramentas automáticas de diagnóstico ⁽²⁾.

Para que ferramentas de diagnóstico automático sejam propostas, o conjunto de dados disponível para estudo é fundamental. A construção da base de dados CODE (*Clinical Outcomes in Digital Electrocardiography*), em 2020, foi um passo importante em prol de avanços no cenário de inteligência artificial em ECGs no Brasil ⁽³⁾. CODE é uma coorte retrospectiva com mais de 2 milhões de ECGs digitais pareados com sistemas de mortalidade e de internação hospitalar de Minas Gerais. Os ECGs foram analisados pela equipe de cardiologistas da Rede de Teleassistência de Minas Gerais, e seis diferentes condições cardíacas normalmente encontradas em exames de ECG foram identificadas: bloqueio atrioventricular de 1 grau (1aAVb), bloqueio de ramo direito (RBBB), bloqueio de ramo esquerdo (LBBB), bradicardia sinusal (SB), fibrilação atrial (AF) e taquicardia sinusal (ST). Essas condições são cruciais na prática clínica cardiológica, sendo indicativas de alterações associadas a riscos aumentados de eventos cardiovasculares como AVC, insuficiência cardíaca e, em casos graves, morte súbita, exigindo intervenções específicas e monitoramento contínuo para prevenir riscos à saúde dos pacientes.

Dada a importância do auxílio do diagnóstico automático para condições cardíacas como as identificadas no CODE, este trabalho tem como objetivo propor um novo modelo baseado em redes neurais profundas para identificação dessas seis condições. O modelo será proposto de forma a explorar melhor o conhecimento dos especialistas na construção do modelo da rede.

O primeiro estudo que usou o CODE para diagnosticar essas seis diferentes condições cardíacas ⁽⁴⁾ propôs uma Rede Neural Convolutacional Residual (ResNet). Essa rede apresentou resultados preditivos melhores do que residentes de cardiologia



do 4º ano, residentes de emergência do 3º ano, e estudantes de medicina do 5º ano, como discutido no trabalho original.

Assim como observado em outras áreas da inteligência artificial (IA), os resultados obtidos pela ResNet ⁽⁴⁾ demonstraram o potencial preditivo de redes convolucionais (CNNs). Entretanto, o surgimento do modelo Transformer ⁽⁵⁾, mudou o curso da IA, superando o desempenho das CNNs em diversas tarefas ^(6, 7). Enquanto CNNs capturam características locais dos dados, transformers têm seu processo de aprendizado baseado no mecanismo de atenção, que permite a extração de informações contextuais globais dos dados, explicando seu êxito em imagens e textos ^(6, 8).

No traçado do ECG, cada batimento possui uma onda P, um complexo QRS e uma onda T. A onda P mostra a contração dos átrios, enquanto o complexo QRS registra a contração dos ventrículos. A onda T aparece após o complexo QRS, indicando a repolarização dos ventrículos. Em se tratando de classificação de ECGs, desejamos que, no escopo de um batimento, o modelo seja capaz de aprender relações entre essas ondas. Entretanto, além da informação detalhada no nível de onda, o modelo também deve ser capaz de compreender a estrutura de um batimento completo, assim como as relações globais e distâncias entre batimentos.

Assim, visando incorporar esse conhecimento prévio do problema à arquitetura da rede, introduzimos o método HiT (Hierarchical Transformer), um método hierárquico baseado em transformers para classificação de ECGs. No HiT, camadas convolucionais e de blocos de transformer são intercalados. As camadas convolucionais aplicam filtros aos dados de entrada para extrair características locais do batimento e reduzir a dimensionalidade dos dados por meio da agregação de informações. Já os blocos de transformer, usados para o aprendizado de representações, consideram informações contextuais sobre a relevância de cada parte da entrada, e usam um mecanismo de *self-attention* localmente restrito ⁽⁹⁾. Essa combinação permite um aprendizado hierárquico e multinível, direcionando o modelo a extrair informações do ECG em diferentes níveis de detalhe e escala temporal.

Para o desenvolvimento do modelo, utilizamos para treinamento a amostra pública de 15% da base de dados CODE ⁽³⁾, e avaliamos o desempenho do modelo com o conjunto CODE-TEST, ambos publicamente acessíveis. O modelo proposto



obteve resultados superiores à Resnet apresentada anteriormente ⁽⁴⁾, se consolidando como o novo modelo estado-da-arte desenvolvido e avaliado na base de dados CODE. Além da comparação de resultados de performance preditiva, um médico cardiologista avaliou as classificações resultantes de ambos os modelos, contribuindo para a interpretação dos resultados e entendimento de como cada modelo, em termos médicos, se sobressai em relação ao outro.

Trabalhos Relacionados

O campo de IA e aprendizado profundo aplicado ao contexto de ECGs representa um cenário de pesquisa em crescimento contínuo. Diversos trabalhos pioneiros utilizaram-se de redes neurais para classificação de ECGs ⁽¹⁰⁾. Por exemplo, os autores em ⁽¹¹⁾ desenvolveram uma rede convolucional de uma dimensão para classificar 12 classes de ritmo em 91.232 ECGs de uma única derivação e demonstraram boa performance preditiva para a tarefa abordada.

Como já mencionado, outro avanço significativo na análise automatizada de ECG foi alcançado por meio de uma ResNet apresentada em ⁽⁴⁾, que demonstrou desempenho superior a médicos residentes em cardiologia na detecção de seis tipos de anormalidades em traçados de ECG de 12 derivações. Este progresso é evidenciado pelo estudo conduzido com mais de 2 milhões de exames rotulados da base CODE. Os resultados mostram que o desempenho da rede, com pontuações F1 acima de 80% e especificidade superior a 99%, representa o estado da arte para a base de dados e tarefa em questão.

Trabalhos recentes se dedicaram a desenvolver modelos de transformer para eletrocardiogramas ⁽¹²⁾. Para a construção destes modelos, em geral, é utilizado uma camada de projeção convolucional, que projeta o sinal original para uma espaço vetorial de representação e em seguida essa representação passa por blocos consecutivos de transformer, que por fim é utilizada para a classificação do ECG.

Outro modelo robusto proposto recentemente é o BAT (*Beat-aligned transformer*) ⁽¹³⁾, baseado no *Swin Transformer* ⁽⁹⁾. Os autores desenvolveram um modelo hierárquico que envolve o pré-processamento do sinal para dividir e alinhar batimentos a partir da compressão e/ou dilação de cada batimento. Após essa divisão os segmentos são associados ao grau de modificação sofrido e transformado pelo modelo hierarquicamente para enfim ser utilizado para classificação. Diferentemente



de nosso modelo, o BAT utiliza uma série de pré-processamentos para identificar os batimentos e transformá-los, de forma que todos os batimentos tenham o mesmo tamanho. Além disso, o bloco de agregação utilizado se baseia apenas em uma projeção linear formada por uma convolução, normalização e MaxPool.

Base de dados

Em termos de eletrocardiogramas, existem bases de dados públicas frequentemente empregadas para testar e avaliar novos métodos baseados em ECGs (14, 15). No entanto, essas bases de dados tem poucos exemplos e sua qualidade pode ser questionável. Apesar da base CODE completa não estar publicamente disponível, um subconjunto dela está aberto para uso geral. A amostragem CODE-15 (3) representa 15% da base de dados original do CODE e inclui 345.779 exames de 233.770 pacientes, quantidade superior à encontrada nas demais bases públicas.

Para avaliar o desempenho do modelo, utilizamos o conjunto CODE-TEST (4). Este conjunto contém 827 exames e foi rotulado seguindo um rigoroso processo de consenso entre dois ou três especialistas em cardiologia. Os diagnósticos incluem as seis anormalidades cardíacas previamente listadas: 1sAVb, RBBB, LBBB, SB, AF e ST.

Tabela 1 – Distribuição das seis condições cardíacas, idade e sexo nos conjuntos de dados utilizados.

	Treino/Val (n = 345779)	Teste (n = 827)
Anormalidade		
1dAVb	5716 (1.7%)	28 (3.4%)
RBBB	9672 (2.8%)	34 (4.1%)
LBBB	6026 (1.7%)	30 (3.6%)
SB	5605 (1.6%)	16 (1.9%)
AF	7033 (2.0%)	13 (1.6%)
SF	7584 (2.2%)	36 (4.4%)
Faixa de idade		
16-25	32820 (9.5%)	43 (5.2%)
26-40	66729 (19.3%)	122 (14.8%)
41-60	100072 (28.9%)	340 (41.1%)
61-80	112181 (32.4%)	278 (33.6%)

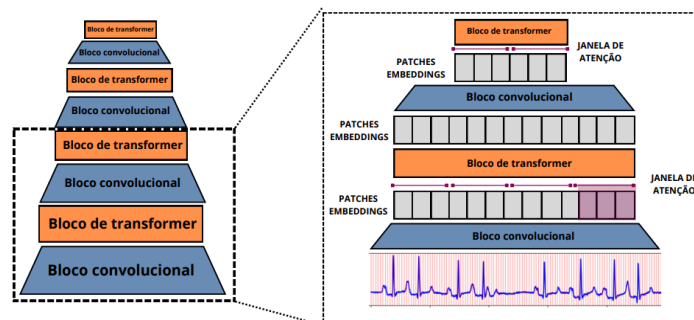


≥81	33957 (9.8%)	44 (5.3%)
Sexo		
Masculinho	206576 (59.7%)	321 (38.8%)
Feminino	139203 (40.3%)	506 (61.2%)

Métodos

Neste trabalho, propomos um transformer hierárquico, especialmente construído para classificação de ECGs, denominado HiT (*Hierarchical Transformer*). O modelo foi desenvolvido para extrair representações do ECG de forma hierárquica, aumentando gradativamente a abrangência temporal das representações. Dessa forma, estabelecemos um processo de aprendizado que envolve a extração de características em múltiplos níveis, a fim de que nos estágios iniciais haja um foco em detalhes de cada onda e distâncias associadas em um curto espaço de tempo, como morfologia de onda P e distâncias de segmento PR (a linha que une o fim da onda P com o início do complexo QRS). Já nos estados superiores, são trabalhadas informações de alto nível de abstração e larga escala temporal, permitindo, por exemplo, entendimento sobre intervalos maiores como o QT (representado por uma linha reta entre o início do complexo QRS e o final da onda T) e até distâncias entre batimentos. O modelo completo e sua estrutura hierárquica pode ser visto na figura 1.

Figura 1 – Modelo HiT de 4 níveis, com blocos convolucionais e de transformer intercalados.



Cada estágio do modelo HiT possui um bloco convolucional, que foca na extração de características e agregação de representações e um bloco de transformer com atenção restrita, que aprende relações contextuais entre parcelas do dado dentro de uma janela controlada.

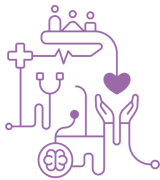
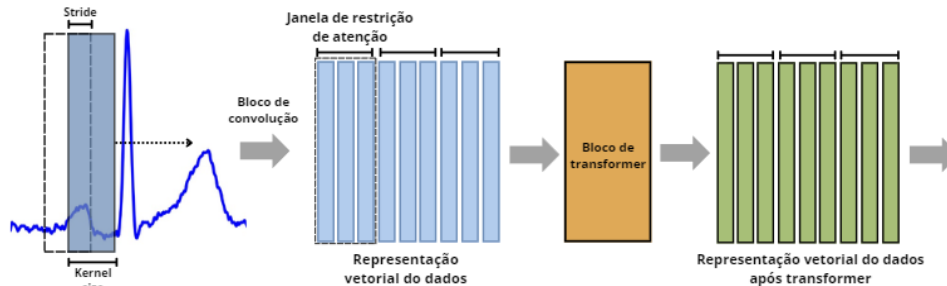
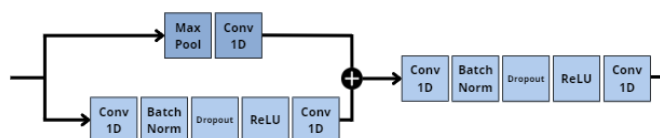


Figura 2 – Primeiro estágio do HiT: o bloco convolucional projeta o sinal para um espaço de representação de vetores. Em seguida, o bloco de transformer modifica as representações de acordo com um contexto restrito à uma janela de atenção.



Para construção dos blocos convolucionais, nos baseamos na ResNet proposta anteriormente ⁽⁴⁾. Desenvolvemos um bloco residual não-linear, que contém uma *branch* principal com camadas de convolução 1D, seguidas por *batch normalization*, ReLU e *dropout*, somada a uma *branch* residual com uma camada de *MaxPool* e uma convolução 1D. Além disso, acrescentamos mais uma camada de convolução 1D, normalização, ReLU e dropout, porém desta vez sem resíduo. O bloco de convolução pode ser visto na Figura 3.

Figura 3 – Arquitetura completa do bloco convolucional proposto.



Já os blocos de Transformer são notáveis por seu mecanismo de *self-attention*, que possibilita o aprendizado de relações contextuais entre diferentes partes do dado. Esse processo envolve receber representações vetoriais do dado e transformá-las, levando em conta o contexto fornecido por outras partes. No contexto da análise de ECGs, isso significa aprender a representação vetorial de uma onda específica, enriquecendo-a com informações sobre as demais ondas do sinal.

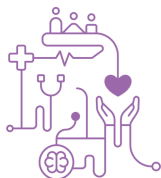
Como já mencionado, utilizamos blocos de transformer baseados no *Swin Transformer* ⁽⁹⁾, que, em comparação com o transformer tradicional, possui o mecanismo de *self-attention* restrito à janelas locais, de forma a somente capturar as



relações entre ondas dentro de janelas de tamanho pré-determinado. Esta restrição tem o objetivo de focar a atenção do modelo em relações locais, de forma a garantirmos que as transformações das representações vetoriais inicialmente considerem as interações de pequena abrangência temporal no sinal, e de maior abrangência à medida que o sinal progride pelos estágios do modelo e é agregado pelas camadas convolucionais. Esse processo é realizado repetidamente em 4 estágios. Assim, no modelo utilizamos uma configuração de bloco convolucional de forma que o dado seja dividido por 4 após cada estágio. Além disso, determinamos uma janela de atenção restrita de tamanho 10.

De forma intuitiva, esse processo pode ser visto da seguinte forma: por meio da convolução, extraímos representações vetoriais do sinal, de modo que o número de vetores resultantes seja 4 vezes menor que o número de medidas no dado original. Por exemplo, um batimento de 320 medidas é reduzido, após passar pelo primeiro bloco, para uma representação de 80 vetores de tamanho 96. Em seguida, o bloco de transformer modifica essas representações de forma condicionada ao contexto restrito pela janela de atenção de tamanho 10. O primeiro estágio do modelo trabalha em uma restrição bem menor que um batimento, direcionado um aprendizado minucioso e detalhado sobre cada onda, como distâncias curtas entre ondas e pequenos detalhes morfológicos do sinal. Em seguida, utilizamos um bloco convolucional para agregar as informações vetoriais após o bloco de transformer; que novamente reduz o número de vetores em 4 vezes. No segundo estágio, mantendo o mesmo tamanho de janela, o mecanismo de atenção possui a abrangência próxima ao limite de um batimento. Isto é, o segundo bloco de transformer recebe informações agregadas, já em um maior nível de abstração, e aprende a transformá-las a partir do contexto restrito a um batimento. Seguindo este mesmo processo, no terceiro estágio, trabalhamos com representações ainda mais agregadas e temos uma janela que abrange dois ou mais batimentos. Já no quarto e último estágio, o modelo realiza suas operações em um nível de abstração e escala temporal que envolve múltiplos batimentos. As representações finais são utilizadas para realizar o diagnóstico de doenças.

Dessa forma, o HiT se diferencia de modelos puramente convolucionais pelo uso do mecanismo de *self-attention*, que permite a melhor extração de características contextuais e relações entre diferentes partes do sinal. Além disso, o modelo proposto também se diferencia de modelos de transformers planos por sua janela de atenção



restrita e arquitetura hierárquica, possibilitando o aprendizado de características locais em diferentes contextos semânticos e como combiná-las globalmente a cada estágio. Tais hipóteses se confirmam nos excelentes resultados de performance preditiva alcançados pelo modelo.

Experimentos e Resultados

Para o treinamento do modelo, utilizamos 95% da base CODE-15 para treino e 5% para validação, sendo tal divisão aleatória e não sobreposta. O treinamento foi realizado empregando o otimizador AdamW ⁽¹⁶⁾ e seguindo um processo escalonamento cossenoidal de taxa de aprendizagem ⁽¹⁷⁾. A taxa de aprendizagem inicial foi definida como 0.0001 e diminuída de forma cossenoidal até 0.00001 durante o treinamento. Além disso, foi adotado o mecanismo de *early stopping*, que consiste em encerrar o treinamento do modelo quando o erro calculado no conjunto de validação não diminui por sete épocas consecutivas. O treinamento foi realizado de forma paralela utilizando 4 GPUs NVIDIA V100.

Para avaliar o desempenho na tarefa de classificação, aplicamos o modelo treinado aos 827 exames do conjunto CODE-TEST. Complementarmente, seguindo o mesmo protocolo de otimização, realizamos o treinamento do modelo ResNet conforme desenvolvido por ⁽⁴⁾. Utilizamos o código do modelo disponibilizado pelos autores. Ambos os modelos foram então testados sob as mesmas condições no conjunto CODE-TEST.

Para a atribuição de uma das condições previstas no ECG, cada classe é considerada de forma independente em uma tarefa binária pelo mesmo modelo, ou seja, a saída do modelo depende da ativação de 6 neurônios. Por exemplo: o modelo prediz se o exame apresenta ou não uma situação de 1dAVb, e assim por diante para as outras 5 classes. Os resultados comparativos dessas avaliações podem ser visualizados nas Tabelas 2 e 3.

Tabela 2 – Média das métricas (agrupando-se todas as classes) na tarefa de classificação de seis condições cardíacas.

Métricas	ResNet	HiT
Acurácia	0.987	0.991
Precisão	0.882	0.909
Revocação	0.684	0.798
F1-score	0.755	0.841

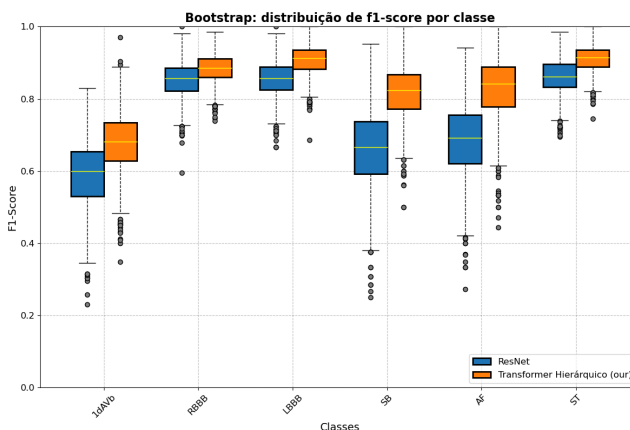


Tabela 3 – Métricas de classificação para cada classe do conjunto de dados utilizado.

Classes	Acurácia		Precisão		Revocação		F1 Score	
	ResNet	HiT	ResNet	HiT	ResNet	HiT	ResNet	HiT
1dAVb	0.981	<u>0.983</u>	<u>1.000</u>	0.938	0.429	<u>0.536</u>	0.600	<u>0.682</u>
RBBB	0.989	<u>0.990</u>	<u>0.963</u>	0.861	0.765	<u>0.912</u>	0.852	<u>0.886</u>
LBBB	0.990	<u>0.994</u>	0.923	<u>1.000</u>	0.800	<u>0.833</u>	0.857	<u>0.909</u>
SB	0.989	<u>0.993</u>	<u>0.818</u>	0.778	0.563	<u>0.875</u>	0.667	<u>0.824</u>
AF	0.989	<u>0.995</u>	0.625	<u>0.909</u>	0.769	<u>0.769</u>	0.690	<u>0.833</u>
ST	0.989	<u>0.993</u>	0.967	<u>0.970</u>	0.784	<u>0.865</u>	0.866	<u>0.914</u>

A fim de garantir significância estatística, avaliamos os resultados dos modelos em 1000 subconjuntos amostrados com reposição de CODE-TEST, seguindo uma metodologia de bootstrap ⁽¹⁸⁾. A distribuição dos resultados para a métrica F1 pode ser vista na figura 4.

Figura 4 – Distribuição de F1-score para cada classe resultante do bootstrap.



Foram usados os testes de Smirnov-Kolmogorov e Mann-Whitney para comparar as distribuições de F1-score de cada modelo e cada classe. Os p-valores obtidos são menores que 0.001, sendo suficientemente pequenos para descartar a hipótese de que as distribuições são iguais.

Por fim, solicitamos um médico cardiologista especialista que avaliasse as predições de ambos os modelos, o que é detalhado no estudo de caso a seguir.

Estudo de caso: Fibrilação Atrial (FA)

Um aspecto importante da avaliação dos modelos é entender porque o modelo proposto, HiT, supera o modelo usado como linha de base, a ResNet. A Figura 4



mostra que fibrilação atrial é o diagnóstico onde o HiT alcançou melhores resultados quantitativos em relação à ResNet. Nesta seção buscamos explicar as origens desses ganhos através de uma análise de contraste entre os erros e acertos dos modelos.

Considerando os casos de fibrilação atrial na base de dados de teste, avaliamos três cenários: (i) ambos os modelos erraram o diagnóstico; (ii) HiT acertou o diagnóstico mas a ResNet errou; e (iii) HiT errou e ResNet acertou o diagnóstico. Para o cenário onde ambos os modelos erraram, um médico cardiologista avaliou 12 exames associados a fibrilação atrial com erro de diagnóstico. Interferência foi o problema que dificultou o diagnóstico em 8 dos 12 exames, seguido de identificação errônea de ondas P (2 exames), do intervalo PR (1 exame) e do intervalo QRS (1 exame). A partir dessa amostra, podemos concluir sobre a inabilidade de ambos os modelos em lidar com interferências e outros desafios correlatos.

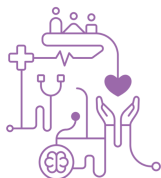
Já para o cenário onde o HiT diagnosticou corretamente a AF mas a ResNet errou, o médico cardiologista avaliou sete exames a fim de identificar eventuais vantagens do modelo proposto em relação a ResNet. Em quatro desses exames, foi possível identificar que o HiT foi capaz de detectar um tipo específico de condição do sinal associado a extrassístoles ventriculares, enquanto a ResNet não. Os outros três casos mostram uma maior resiliência do modelo proposto, pois ele foi capaz de diagnosticar corretamente diante de interferência na onda P, presença de marcapasso e inversão de eletrodos.

Para o último cenário, onde o HiT errou o diagnóstico, mas a ResNet acertou, o médico cardiologista avaliou 3 exames e identificou que o HiT não conseguiu diagnosticar fibrilação atrial corretamente diante de ondas P de pequena amplitude. Em um dos casos, a presença de um marcapasso explica o erro do modelo proposto.

Assim, embora haja cenários frequentemente associados a interferências onde ambos os modelos erram, a avaliação de cenários reforça a percepção quantitativa de que o HiT foi mais efetivo em cenários específicos de interferências e se mostrou mais robusto a outros problemas de aquisição. Por outro lado, se mostrou mais suscetível a questões envolvendo a amplitude da onda P.

Conclusões

Este trabalho apresentou um novo modelo de transformer hierárquico, denominado HiT, para classificação de 6 cardiopatias, de grande importância clínica,



que podem ser identificadas em eletrocardiogramas. Para a construção do modelo, foram considerados pressupostos médicos que favorecem um processo minucioso de diagnóstico de ECG, envolvendo múltiplos níveis de abstrações e temporalidades. Dessa forma, o modelo desenvolvido intercala blocos convolucionais com blocos de transformer com atenção restrita a uma janela local, direcionando o modelo a extrair e transformar informações locais e em seguida agregá-las globalmente.

O HiT apresentou resultados superiores a ResNet, que era até então o melhor modelo desenvolvido e avaliado nos dados CODE-15. Além disso, com a ajuda de um cardiologista especialista trabalhamos para entendimento de quais características de ECGs foram melhores identificadas pelo nosso modelo. Observamos que o HiT se mostrou robustos a cenários de interferência e problemas de aquisição, além de ter apresentado predições corretas em casos de extrassístoles, marcas passos e inversão de eletrodos. Ademais, é esperado que os resultados apresentados melhorem ainda mais com um maior conjunto de dados para treinamento, como seria o caso da base completa do CODE.

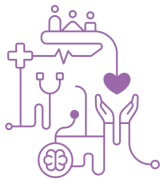
No cenário de saúde digital e telessaúde, é fundamental a contribuição de novos avanços técnicos a fim de estabelecer uma melhor performance preditiva, permitindo que, futuramente, modelos de inteligência artificial possam ser incorporados ao dia-a-dia do cardiologista com boa capacidade preditiva, trazendo amplos benefícios à medicina cardiovascular.

Agradecimentos

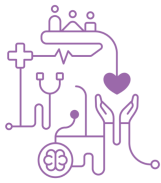
Esse trabalho foi parcialmente financiado pelo CNPq, CAPES e FAPEMIG, além do CIIA-Saúde. Agradecemos também o Centro de Telessaúde de Minas Gerais, pelo acesso aos dados e as discussões produtivas no âmbito deste trabalho.

Referências

1. World Health Organization. Cardiovascular diseases (CVDs) [Internet]. Disponível em: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), acessado em 29 de maio de 2024.
2. Liu X, Wang H, Li Z, Qin L. Deep learning in ECG diagnosis: A review. Knowledge-Based Systems. 2021;227:107187.



3. Ribeiro ALP, Paixao GMM, Gomes PR, Ribeiro MH, Ribeiro AH, Canazart JA, Oliveira DM, Ferreira MP, Lima EM, de Moraes JL, et al. Tele-electrocardiography and bigdata: the CODE (Clinical Outcomes in Digital Electrocardiography) study. *J Electrocardiol.* 2019;57
4. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira Jr W, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun.* 2020;11(1):1760.
5. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30.
6. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929.* 2020.
7. Dong L, Xu S, Xu B. Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2018. p. 5884-5888.
8. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al. Language models are unsupervised multitask learners. *OpenAI blog.* 2019;1(8):9.
9. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision.* 2021. p. 10012-22.
10. Lyon A, Mincholé A, Martínez JP, Laguna P, Rodriguez B. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. *J R Soc Interface.* 2018;15(138):20170821.
11. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* 2019;25(1):65-69.
12. Hu R, Chen J, Zhou L. A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. *Comput Biol Med.* 2022;144:105325.
13. Li X, Li C, Wei Y, Sun Y, Wei J, Li X, et al. Bat: Beat-aligned transformer for electrocardiogram classification. In: *2021 IEEE International Conference on Data Mining (ICDM).* IEEE; 2021. p. 320-9.
14. Wagner P, Strodthoff N, Boussejot R-D, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data.* 2020;7(1):1-15.
15. Moody GB, Mark RG, Goldberger AL. PhysioNet: a web-based resource for the study of physiologic signals. *IEEE Eng Med Biol Mag.* 2001;20(3):70-75.
16. Loshchilov I, Hutter F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101.* 2017.
17. Loshchilov I, Hutter F. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983.* 2016.



CBIS'24

XX Congresso Brasileiro de Informática em Saúde
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

18. Davison AC, Hinkley DV. Bootstrap methods and their application. 1st ed. Cambridge: Cambridge University Press; 1997.