

Classificação de insuficiência cardíaca em ecocardiogramas utilizando Redes Neurais Profundas

Luiz Fernando Oliveira Maciel¹, Alexei Manso Correa Machado^{1,2}

¹Departamento de Ciência da Computação

Pontifícia Universidade Católica de Minas Gerais (PUC Minas), Belo Horizonte, Brasil

²Departamento de Anatomia e Imagem - Faculdade de Medicina

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil

lfomaciel@sga.pucminas.br, alexeimcmachado@gmail.com

Abstract. This article presents a deep learning model based on MobileNetv4 for detecting heart failure in echocardiograms based on left ventricle ejection fraction (LVEF) estimation. The pipeline consists of two stages: a regression model predicting ventricular volume coordinates, and a classification model that computes LVEF from the predicted volumes. A 5-fold cross-validation was performed, yielding a coefficient of determination R^2 of 0.83 ± 0.01 (95% CI: [0.83, 0.84]) and an accuracy of 0.74 ± 0.01 (95% CI: [0.74, 0.75]).

Resumo. Este artigo apresenta um modelo de detecção de insuficiência cardíaca em ecocardiogramas baseado na arquitetura MobileNetv4, a partir do cálculo da fração de ejeção do ventrículo esquerdo (LVEF). O pipeline é composto de duas etapas: uma de regressão das coordenadas referentes ao volume ventricular, e outra de classificação, que calcula a LVEF a partir dos volumes encontrados. Experimentos baseados em validação cruzada com 5 divisões de treinamento e teste resultaram em um coeficiente de determinação R^2 de $0,83 \pm 0,01$ (IC 95%: [0,83, 0,84]) e acurácia de $0,74 \pm 0,01$ (IC 95%: [0,74, 0,75]).

1. Introdução

Doenças cardiovasculares são uma das principais causas de morte no mundo. Em 2021, a doença cardíaca isquêmica foi a principal causa de morte global, representando 13% do total. Apesar de sua gravidade, cerca de 90% das doenças cardiovasculares podem ser prevenidas com acompanhamento médico, diagnóstico precoce e tratamento [McGill et al. 2008]. Isso destaca a importância de métodos eficientes, acessíveis e interpretáveis para o diagnóstico dessas condições.

A ecocardiografia, exame não invasivo que permite a visualização em tempo real do coração, é um dos métodos mais comuns e acessíveis para avaliar a função cardíaca. A fração de ejeção do ventrículo esquerdo (LVEF), um indicador crucial da função cardíaca que avalia o desempenho do coração em bombear sangue, pode ser determinada a partir de um ecocardiograma [Omerovic and Jain 2024]. Valores anormais de LVEF podem indicar insuficiência ou alguma cardiomiopatia.

Este trabalho propõe um *pipeline* baseado na arquitetura *MobileNetV4* para classificar insuficiência cardíaca em ecocardiogramas, a partir da regressão do volume ventricular e cálculo da fração de ejeção, equilibrando interpretabilidade e eficiência com a simplicidade computacional do *MobileNetV4*. O trabalho foi feito exclusivamente pelo primeiro autor sob a orientação do segundo.

2. Trabalhos Relacionados

[Ghorbani et al. 2020] propuseram o *EchoNet* para segmentação cardíaca, cálculo de métricas e classificação de fenótipos (idade, sexo, peso e altura), alcançando alta precisão e interpretabilidade, mas com variação conforme a tarefa. Em continuidade, [Ouyang et al. 2020] apresentaram o *EchoNet-Dynamic*, que segmenta o ventrículo esquerdo, estima a *LVEF* e diagnostica insuficiência cardíaca, utilizando pequenos clipes de batimentos cardíacos para estimar a fração de ejeção, superando a avaliação humana ($R^2 = 0,81$, $AUC = 0,97$). Já [Paul and Bhuiyan 2024] compararam CNNs pré-treinadas para segmentação e classificação cardíaca, obtendo 0,77 de acurácia com *EfficientNetB7* e 0,72 com o *MobileNetV2*, antecessor da arquitetura utilizada neste estudo.

3. Redes neurais convolucionais

Redes neurais convolucionais (CNN) são um tipo específico de rede neural artificial normalmente utilizadas para tarefas de reconhecimento de padrões em imagens [O’Shea and Nash 2015], sendo a diferença principal entre as redes convolucionais e as redes neurais clássicas a especificidade dos problemas que buscam resolver.

O *MobileNetV4* é uma arquitetura de CNN desenvolvida por [Qin et al. 2024], cujo foco é a eficiência computacional. Ela possui uma série de otimizações, como o uso de blocos arquiteturais leves e uma busca arquitetural otimizada. Essa abordagem torna o *MobileNetV4* uma solução eficaz para a tarefa de regressão deste estudo.

4. Materiais e Métodos

A base de dados utilizada é a *EchoNet-Dynamic* [Ouyang et al. 2020], com 10030 ecocardiogramas de 112x112 *pixels*, e informações como a quantidade de frames e as coordenadas dos volumes final diastólico (VFD) e sistólico (VFS) extraídas de 42 pontos-chave nas imagens (84 números no total). Os *frames* do fim da diástole e sístole foram extraídos, divididos em treinamento/validação (17.438 imagens) e teste (2.552 imagens), com validação cruzada em 5 divisões (80% treino, 20% validação). Foram removidos 35 exames (70 imagens) por apresentarem erros.

A implementação usou *Python* e *Pytorch*, com pesos pré-treinados do *MobileNetV4* (*timm*). O treinamento foi realizado em *CPU Intel i5-4460* com 16 GB de RAM, sem *GPUs*. O *pipeline* possui duas etapas, primeiro é feita a regressão dos 84 valores referentes às coordenadas do volume cardíaco, seguida pelo cálculo da *LVEF* para classificar o exame como normal, leve ou anormal. A abordagem baseada em regressão permite interpretar e visualizar os volumes estimados no exame, algo valioso no contexto clínico.

Observando a figura 1, à esquerda temos uma visão do pipeline completo, e à direita, a estrutura da arquitetura de regressão, compreendendo um bloco com o *MobileNetV4* e outro com novas camadas ao fim. A entrada é uma imagem contendo 3 canais e 224x224 *pixels*. A primeira dimensão em todas as camadas representa o tamanho dos lotes utilizados no treinamento.

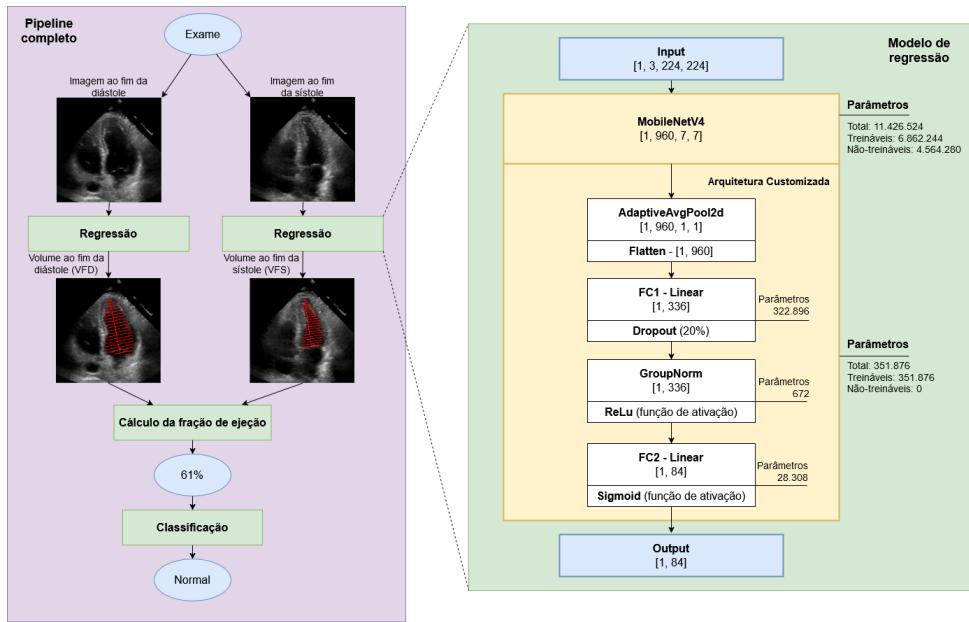


Figura 1. Visualização do pipeline (esquerda), com suas principais etapas, e da rede proposta (direita), com o nome e tamanho da saída das camadas e a quantidade de parâmetros treináveis e não-treináveis

Parte dos parâmetros do *MobileNetV4* foram descongelados, mas com uma taxa de aprendizado menor que a dos parâmetros treináveis do bloco customizado, para reter a extração de características do pré-treinamento. O bloco customizado inclui *pooling* adaptativo, *flatten* e duas camadas densas (*FC1* e *FC2*) para regressão das coordenadas. *FC1* é seguida por *Dropout* para regularização, *GroupNorm* para estabilização do treinamento e ativação *ReLU* para não-linearidade. A camada *FC2* reduz a saída para as coordenadas desejadas, e a ativação *Sigmoid* escala os valores entre 0 e 1.

Os volumes cardíacos (*VFS* e *VFD*) são calculados a partir da saída do modelo. Assume-se uma aproximação em que o ventrículo é representado por uma série de discos empilhados ao longo de um eixo longitudinal que o corta em seus extremos. O diâmetro de cada disco é definido pela conexão entre pontos opostos da parede cardíaca, e seu volume é estimado pela multiplicação da área do disco pela altura (definida como o tamanho do eixo longitudinal dividido pela quantidade total de discos, 20). O volume total do ventrículo é aproximado como a soma do volume dos 20 discos.

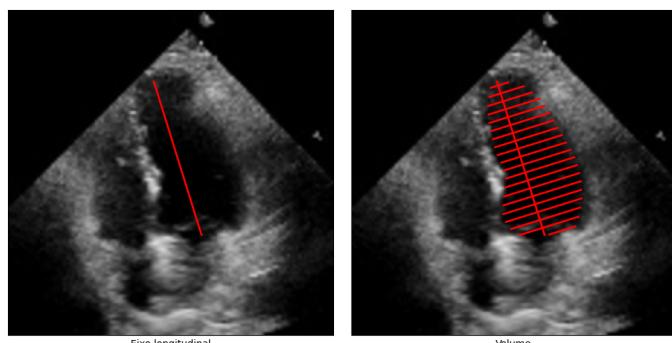


Figura 2. Visualização do eixo longitudinal e volume ventricular de um exame

A fração de ejeção é calculada como $LVEF\% = 100(VFD - VFS)/VFD$. O exame é classificado como normal se $LVEF\% > 50\%$, leve se estiver entre 40% e 49%, e anormal se $LVEF\% < 40\%$ [Paul and Bhuiyan 2024].

Quanto aos hiperparâmetros, a taxa de aprendizado foi 8×10^{-4} , com 5×10^{-5} para camadas descongeladas. O decaimento de peso foi 2×10^{-6} e *dropout* de 0,2. O treinamento ocorreu por 200 épocas, com lotes de 16 exames. Duas camadas pré-treinadas foram descongeladas, e o tamanho das customizadas foi 336. Em cenários reais, técnicas como *early stopping* poderiam evitar processamento desnecessário. Ao todo, o modelo possui 11.426.524 parâmetros, sendo 6.862.244 treináveis.

5. Resultados

Foi feita uma validação cruzada com $K = 5$ divisões (ou dobras). A figura 3 apresenta a evolução das métricas por época para a primeira dobra: perda (MSE), acurácia, coeficiente de determinação R^2 e erro médio normalizado (NME). O eixo horizontal representa as épocas; os eixos verticais, os valores das métricas. As linhas azul, vermelha e verde correspondem às bases de treinamento, validação e teste, respectivamente.

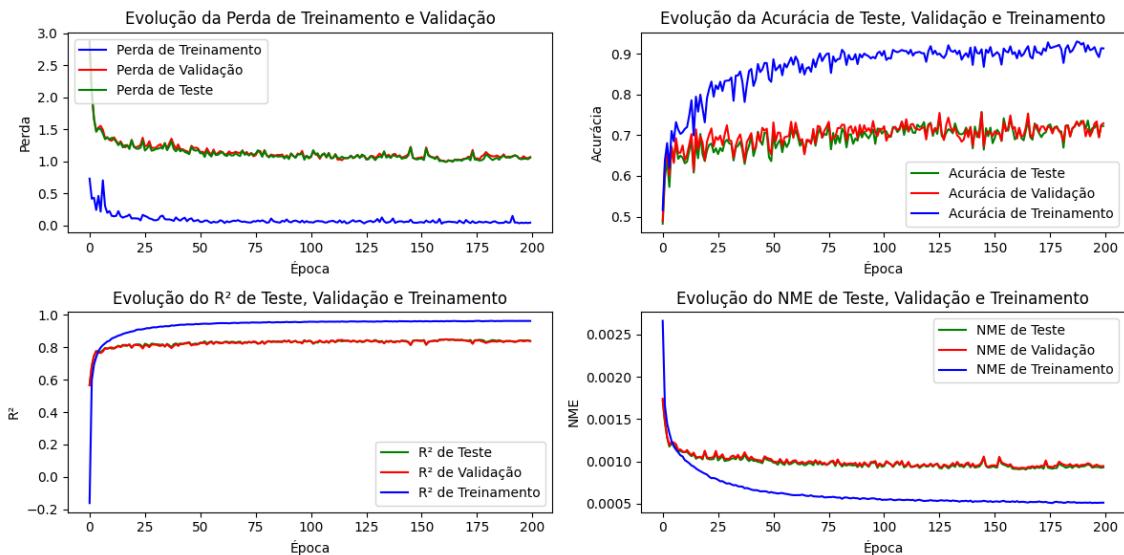


Figura 3. Evolução das métricas de desempenho (perda, acurácia, R^2 e NME) ao longo do treinamento da primeira dobra

Pode-se notar que os resultados entre a base de validação e testes se assemelham entre todos os gráficos, o que é um bom indicador de que o modelo conseguiu generalizar bem para dados fora da base de treinamento.

A tabela 1 apresenta os resultados, incluindo a acurácia, perda (MSE) e coeficiente de determinação R^2 , assim como a média, desvio padrão e intervalo de confiança de 95% para cada uma das métricas. A melhor acurácia (dobra 1 - 0,76) ficou apenas 0,02 acima da média geral, e o desvio padrão da acurácia (0,01) indica baixa variabilidade. As outras métricas também apresentaram pequenas variações, demonstrando consistência nos resultados.

Dobra	Acurácia	Perda (MSE)	R ²
1	0,76	1,10	0,83
2	0,73	1,07	0,84
3	0,74	1,15	0,82
4	0,74	1,06	0,84
5	0,74	1,11	0,83
Média + Desvio padrão	$0,74 \pm 0,01$	$1,10 \pm 0,03$	$0,83 \pm 0,01$
IC 95%	[0,74 , 0,75]	[1,07 , 1,13]	[0,83 , 0,84]

Tabela 1. Resultados da validação cruzada com 5 dobras

Por fim, o tempo de treinamento médio foi de 1.109,95 segundos (aproximadamente 18 minutos e 30 segundos) por época, totalizando uma média de 2 dias, 13 horas, 39 minutos e 50 segundos para treinar um modelo.

5.1. Desempenho por classe

Para uma análise da eficiência do modelo na classificação, foi examinada a matriz de confusão (Tabela 2) e as sensibilidades por classe para o conjunto de testes em cada dobra (Tabela 3). Esses resultados indicam a capacidade do modelo em classificar corretamente cada categoria.

Real	Previsto Dobra 1			Previsto Dobra 2			Previsto Dobra 3			Previsto Dobra 4			Previsto Dobra 5		
	N	L	A	N	L	A	N	L	A	N	L	A	N	L	A
Normal (N)	822	71	25	784	54	21	810	55	31	793	59	23	810	65	28
Leve (L)	29	78	30	33	103	29	34	102	33	31	104	27	31	94	37
Anormal (A)	26	115	80	39	120	93	37	170	68	36	120	83	30	105	76

Tabela 2. Matriz de confusão para cada dobra

Métrica	Dobra 1	Dobra 2	Dobra 3	Dobra 4	Dobra 5
Sensibilidade (Normal)	0.90	0.91	0.90	0.91	0.90
Sensibilidade (Leve)	0.21	0.20	0.20	0.19	0.19
Sensibilidade (Anormal)	0.52	0.48	0.50	0.50	0.50
Acurácia	0.76	0.73	0.74	0.74	0.74

Tabela 3. Sensibilidades por classe e acurácia para cada dobra

A tabela 3 mostra alta sensibilidade para a classe normal e menor para as classes leve e anormal. Isso possivelmente se deve à distribuição dos valores de *LVEF*: a classe normal abrange uma faixa ampla ($LVEF > 50\%$), enquanto as classes leve (40-49%) e anormal (<40%) têm margens menores. Em um cenário real, espera-se uma maior proporção de exames normais ou leves, presumindo que exames com *LVEF* anormalmente baixos estariam em situação crítica de insuficiência cardíaca.

Na base de dados utilizada, 76,73% dos exames do conjunto de treinamento foram classificados como normais, 9,96% como leves e 13,31% como anormais. Distribuições semelhantes foram observadas nos conjuntos de validação (76,71%, 10,48% e 12,81%) e teste (76,80%, 9,87% e 13,32%).

Como a base é usada apenas no treinamento do modelo de regressão, que estima as coordenadas sem considerar a classificação da função cardíaca, presume-se que o impacto desse desbalanceamento seja limitado. Futuros estudos podem analisar os efeitos do balanceamento artificial dos dados na precisão da regressão e classificação.

6. Conclusões

Este estudo mostrou que a arquitetura *MobileNetV4* é viável para classificar insuficiência cardíaca em ecocardiogramas. Apesar do treinamento em *CPU*, o modelo atingiu $0,74 \pm 0,01$ (IC 95%: [0,74;0,75]) de acurácia com cerca de 2 dias de treinamento e 200 épocas, demonstrando que a arquitetura poderia ser implementada em cenários reais, alcançando resultados competitivos quando comparados a modelos recentes.

A combinação do modelo pré-treinado com camadas customizadas teve um impacto positivo nas métricas de desempenho. Comparando com outras arquiteturas, o resultado foi competitivo: [Paul and Bhuiyan 2024] obtiveram 0,72 de acurácia utilizando a arquitetura *MobileNetV2* (0,02 abaixo do limite inferior do intervalo de confiança), e 0,77 com a arquitetura *EfficientNetB7* (0,02 acima do limite superior do intervalo de confiança); [Ouyang et al. 2020] reportaram um $R^2 = 0,81$ para uma tarefa similar de classificação da função cardíaca, um valor próximo ao R^2 obtido nesse estudo de $(0,83 \pm 0,01)$, estando 0,02 abaixo do limite inferior do intervalo de confiança.

Futuras melhorias incluem testar modelos mais recentes, otimizar hiperparâmetros e explorar novas abordagens para a classificação. Outro ponto de interesse seria uma comparação da eficiência do modelo em diferentes *hardwares*, analisando métricas como a velocidade de treinamento e de predição dos exames.

Agradecimentos — AMC Machado agradece à FAPEMIG pelo auxílio financeiro através dos projetos APQ-02753-24 e APQ-06556-24.

Referências

- Ghorbani, A., Ouyang, D., Abid, A., He, B., Chen, J., Harrington, R., Liang, D., Ashley, E., and Zou, J. (2020). Deep learning interpretation of echocardiograms. *NPJ digital medicine*, 3(1):10.
- McGill, H., McMahan, A., and Gidding, S. (2008). Preventing heart disease in the 21st century. *Circulation*, 117(9):1216–27.
- Omerovic, S. and Jain, A. (2024). Echocardiogram. In *StatPearls*. StatPearls Publishing, Treasure Island (FL).
- OMS (2021). Global health estimates: Leading causes of death.
- O’Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv*.
- Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C., Heidenreich, P., Harrington, R., Liang, D., Ashley, E., et al. (2020). Video-based ai for beat-to-beat assessment of cardiac function. *Nature*, 580(7802):252–6.
- Paul, A. and Bhuiyan, Y. (2024). EchoTrace: A 2D echocardiography deep learning approach for left ventricular ejection fraction prediction. *J. Electron. Electric. Eng.*
- Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., Wang, W., Banbury, C., Ye, C., Akin, B., et al. (2024). Mobilenetv4-universal models for the mobile ecosystem. *arXiv preprint arXiv:2404.10518*.