

Diagnóstico assistido por ML para detecção e classificação de sopros cardíacos usando fonocardiogramas digitais

Adrian Alejandro Chavez Alanes
Programa de Pós Graduação em Telecomunicações
Instituto Nacional de Telecomunicações
adrian@mtel.inatel.br

Resumo—Este trabalho apresenta um sistema baseado em aprendizado de máquina para a detecção e classificação de sopros cardíacos em fonocardiogramas digitais. Foram utilizados dados do PhysioNet 2022, com pré-processamento do sinal, segmentação manual e extração de características por Mel Frequency Energy. Dois enfoques foram implementados utilizando arquiteturas CNN compactas, alcançando acurácias superiores a 93% nos testes. Os modelos foram treinados e implantados na plataforma Edge Impulse, possibilitando execução eficiente em dispositivos móveis e microcontroladores, facilitando o diagnóstico remoto.

Index Terms—Heart Murmur Detection, Phonocardiogram, Machine Learning, Convolutional Neural Networks, Tiny Machine Learning

I. INTRODUÇÃO

As doenças cardiovasculares representam a principal causa de mortalidade mundial, com impacto especialmente crítico em populações pediátricas [1]. A detecção precoce de anomalias cardíacas, como os sopros, é fundamental para o diagnóstico e tratamento eficaz dessas doenças [2]. Tradicionalmente, a ausculta manual com estetoscópio é o método mais utilizado para triagem, porém apresenta limitações relacionadas à subjetividade e à experiência do profissional de saúde, o que pode comprometer a acurácia diagnóstica [1].

Com os avanços da fonocardiografia digital e das técnicas de aprendizado de máquina, tem-se observado um crescente interesse no desenvolvimento de sistemas automatizados capazes de detectar e classificar sopros cardíacos a partir de fonocardiogramas digitais [3], [4]. O surgimento de grandes bases de dados, como o CirCor DigiScope e o PhysioNet, com múltiplas gravações anotadas, tem permitido o treinamento e avaliação de modelos mais robustos [1].

Apesar dos avanços, desafios relacionados à variabilidade dos sinais, ruídos ambientais, e à escassez de dados anotados dificultam a adoção ampla dessas tecnologias [4]. Além disso, a possibilidade de capturar fonocardiogramas com dispositivos móveis, como smartphones, amplia o acesso a exames de qualidade, especialmente em áreas remotas [5].

Neste contexto, este trabalho visa contribuir para a detecção e classificação automatizada de sopros cardíacos por meio do uso de aprendizado de máquina aplicado a fonocardiogramas digitais.

II. TRABALHOS RELACIONADOS

Nos últimos anos, o uso de aprendizado de máquina (ML) e aprendizado profundo (DL) para detecção e classificação automática de sopros cardíacos tem recebido crescente atenção [1], [4]. As abordagens vão desde extração manual de características até modelos end-to-end baseados em redes neurais convolucionais e recorrentes.

Ganguly e Sharma [2] exploraram técnicas tradicionais de extração de atributos temporais, espectrais e estatísticos em fonocardiogramas (PCGs), obtendo resultados promissores. Sideras e Boura [3] ressaltam a importância da seleção e balanceamento de amostras para evitar vieses nos classificadores.

Com bases robustas como CirCor DigiScope e PhysioNet 2022, modelos DL que combinam CNNs e RNNs têm sido desenvolvidos para capturar características temporais e espectrais dos sinais [1], [4]. Nie et al. [6] propuseram um modelo multitarefa 2D CNN para detecção de sopros e estimativa de frequência cardíaca, com alto desempenho clínico.

Estudos focados em classificação multiclasse, como Zhao et al. [7], e técnicas de aumento de dados para superar limitações de anotações, como Singh et al. [8], contribuem para maior robustez. Redes recorrentes aplicadas a PCGs, segundo Kim e Lee [9], oferecem alternativas para captura temporal dos sinais.

A captura via dispositivos móveis também tem sido estudada. Luo et al. [5] avaliaram a qualidade dos sinais capturados por smartphones em usuários não especializados, confirmando a viabilidade para análise clínica, apesar de limitações demográficas.

Desafios persistem, como a variabilidade interpacientes, a necessidade de modelos eficientes para dispositivos limitados e a melhoria da interpretabilidade clínica. O Tiny Machine Learning (TinyML) surge como solução promissora, permitindo modelos compactos e eficientes em microcontroladores como Arduino e ESP32 [10], [11], conciliando precisão e portabilidade.

Este trabalho segue essa perspectiva, utilizando a plataforma Edge Impulse para desenvolver modelos executáveis em dispositivos móveis e exportáveis para ambientes embarcados, garantindo portabilidade e baixo consumo energético sem comprometer a performance [12].

III. METODOLOGIA

A. Conjunto de Dados

O conjunto de dados utilizado neste trabalho foi o PhysioNet 2022 [1], [3], [4], um repositório amplamente conhecido na área da saúde. Este dataset foi criado a partir da coleta de sons cardíacos utilizando estetoscópios digitais em ambientes clínicos na região Nordeste do Brasil. Contém ao todo 5.272 gravações no formato WAV, provenientes de 1.568 pacientes, com múltiplas gravações feitas em diversos pontos do tórax de cada paciente.

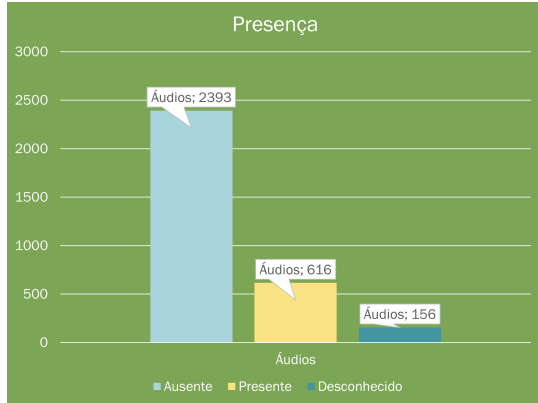


Figura 1. Distribuição geral das gravações no dataset PhysioNet 2022, categorizadas como “Ausente”, “Presente” e “Desconhecido”.

A distribuição das gravações, referente ao total de áudios classificados como “Ausente”, “Presente” e “Desconhecido”, está ilustrada na Figura 1, que mostra a quantidade de áudios para cada categoria.

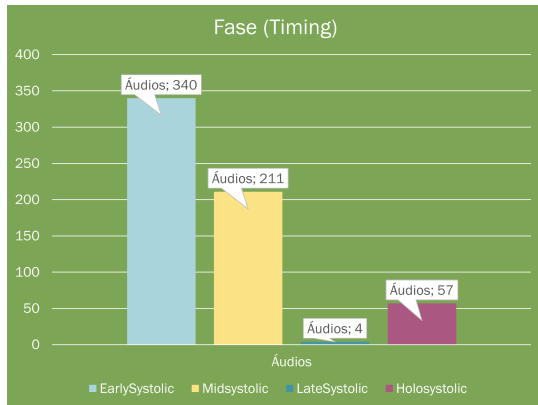


Figura 2. Distribuição das gravações com sopros presentes segundo as classes de timing sistólico: “Early”, “Mid”, “Late” e “Holo”.

Focando nas gravações com sopros presentes, o dataset também disponibiliza rótulos adicionais baseados na fase e no timing dos sopros sistólicos, classificados em quatro tipos principais: “Early”, “Mid”, “Late” e “Holo”. A distribuição desses áudios conforme essas classes está apresentada na Figura 2.

B. Pré-processamento

Os áudios originais do dataset apresentavam uma considerável presença de ruído e informações não relevantes para a análise dos sons cardíacos, o que poderia comprometer o desempenho dos modelos. Estudos indicam que os sons cardíacos possuem componentes significativos principalmente na faixa de 20 a 150 Hz, embora os sopros possam conter frequências até cerca de 500 Hz [1], [2].

Com base nessas evidências, aplicou-se um filtro passa-banda entre 20 e 500 Hz para preservar as informações importantes dos sinais cardíacos enquanto se reduz o ruído de alta frequência.

As Figuras 3 e 4 ilustram, respectivamente, um exemplo de áudio original contendo ruído e outro áudio após aplicação do filtro passa-banda, evidenciando a remoção de componentes fora da faixa cardíaca.

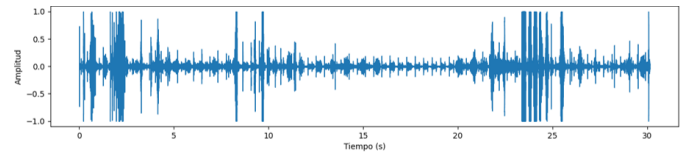


Figura 3. Áudio original contendo ruídos e componentes não relacionados ao som cardíaco.

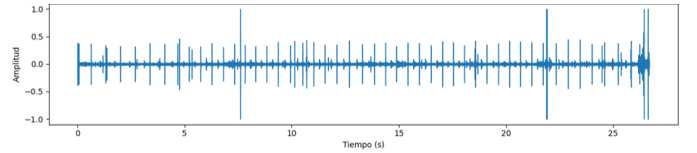


Figura 4. Áudio após aplicação do filtro passa-banda entre 20 e 500 Hz.

Para a segmentação dos áudios, foram avaliadas diferentes durações de janelas temporais. Optou-se por utilizar janelas de 5 segundos, que representam um compromisso técnico entre a quantidade suficiente de exemplos para aprendizado robusto e a riqueza informativa de cada segmento, garantindo um balanceamento adequado entre volume de dados e qualidade dos sinais analisados.

A segmentação foi realizada manualmente para cada áudio utilizando a ferramenta *Split* da plataforma Edge Impulse, que permite a visualização da forma de onda e a seleção precisa dos trechos mais relevantes para treinamento. A Figura 5 apresenta a interface desta ferramenta durante o processo de segmentação.

Além do filtro passa-banda aplicado, técnicas avançadas de processamento de sinais cardíacos, como descritas por Wang et al. [13], têm contribuído para melhorar a extração de características relevantes em fonocardiogramas. Para a criação dos modelos, arquiteturas CNN otimizadas para dispositivos portáteis, conforme proposto por Patel e Kumar [14], oferecem um bom equilíbrio entre desempenho e eficiência computacional, fundamentando a escolha das redes compactas utilizadas neste trabalho.

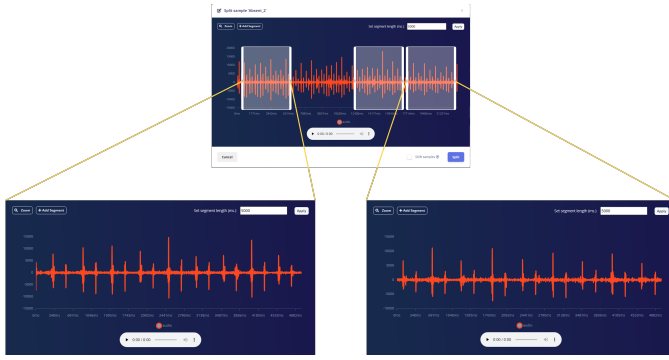


Figura 5. Ferramenta *Split* da plataforma Edge Impulse utilizada para segmentação manual dos áudios.

Este pré-processamento garante que os dados utilizados para o treinamento dos modelos estejam mais limpos e representem com maior fidelidade os sinais cardíacos relevantes, contribuindo para o aumento da precisão dos modelos.

C. Criação dos Modelos

Para o treinamento dos modelos, foi utilizado como entrada o bloco de processamento Mel Frequency Energy (MFE), que extrai características espectrais do áudio, configurado para janelas de 2.400 ms com passo de 200 ms. Esse processo gera um conjunto robusto de 7.616 features por amostra, que servem como base para as arquiteturas subsequentes de redes neurais.

Duas abordagens foram implementadas para a detecção e classificação dos sopros cardíacos:

- **Primeira abordagem:** Uso de dois modelos sequenciais. O primeiro modelo realiza uma detecção binária, classificando os áudios em "Normal" ou "Sopro", utilizando um conjunto de 581 áudios balanceados automaticamente entre treino, validação e teste. O segundo modelo, acionado apenas quando o primeiro identifica sopro, realiza uma classificação multi-classe entre três tipos de sopros sistólicos "Early", "Mid" e "Holo", utilizando um conjunto de 236 áudios balanceados, tendo a classe MidSystolic como referência.
- **Segunda abordagem:** Modelo único de classificação multi-classe com quatro categorias: "Normal" e os três tipos de sopros sistólicos "Early", "Mid" e "Holo". Para evitar desequilíbrio, a quantidade de áudios da classe Normal foi reduzida aleatoriamente, garantindo um balanceamento adequado para o treinamento.

Cada camada de convolução 2D descrita na tabela a seguir inclui uma operação de pooling (tipicamente max pooling), conforme implementado na plataforma Edge Impulse, que reduz a dimensionalidade e ajuda a prevenir o overfitting.

A Tabela I apresenta a estrutura detalhada das arquiteturas empregadas em cada modelo, incluindo o tipo das camadas, número de filtros, tamanho dos kernels e taxas de dropout utilizadas.

Tabela I
ARQUITETURAS DOS MODELOS TREINADOS

Modelo	Camada	Filtros	Kernel	Dropout
Binário	Input	-	-	-
	Reshape	-	-	-
	Conv2D + Pool	8	3×3	0.25
	Conv2D + Pool	16	3×3	0.25
	Conv2D + Pool	32	3×3	0.25
	Conv2D + Pool	64	4×4	0.25
	Flatten	-	-	-
	Output	2	-	-
Multi-3 classes	Input	-	-	-
	Reshape	-	-	-
	Conv2D + Pool	8	3×3	0.25
	Conv2D + Pool	16	3×3	0.25
	Conv2D + Pool	32	3×3	0.35
	Conv2D + Pool	64	4×4	0.35
	Conv2D + Pool	64	4×4	0.25
	Flatten	-	-	-
Multi-4 classes	Input	-	-	-
	Reshape	-	-	-
	Conv2D + Pool	8	3×3	0.25
	Conv2D + Pool	16	3×3	0.35
	Conv2D + Pool	32	3×3	0.35
	Conv2D + Pool	64	4×4	0.25
	Conv2D + Pool	64	4×4	0.25
	Flatten	-	-	-
	Output	4	-	-

IV. RESULTADOS

A. Resultados do Treinamento

Durante o treinamento, os modelos apresentaram desempenho robusto, com altas acurácias e métricas consistentes na validação. As Tabelas II, III e IV mostram as matrizes de confusão para os três modelos treinados.

a) *Modelo Binário:* O modelo binário alcançou acurácia de 96,4%, com baixa taxa de falsos positivos e negativos, refletida na Tabela II.

Tabela II
MATRIZ DE CONFUSÃO NA VALIDAÇÃO — MODELO BINÁRIO

Classes	Normal	Sopro
Normal	97,4%	2,6%
Sopro	4,8%	95,2%

b) *Modelo Multi-3 Classes:* A acurácia do modelo com três classes foi de 97,5%, com excelente discriminação entre sopros Early, Holo e Mid. Veja a Tabela III.

Tabela III
MATRIZ DE CONFUSÃO NA VALIDAÇÃO — MODELO MULTI-3 CLASSES

Classes	EarlySys	HoloSys	MidSys
EarlySys	96,3%	3,5%	0,2%
HoloSys	2,9%	97,1%	0,0%
MidSys	0,0%	0,9%	99,1%

c) *Modelo Multi-4 Classes*: Incluindo a classe Normal, o modelo apresentou acurácia de 96,0%, com resultados detalhados na Tabela IV.

Tabela IV
MATRIZ DE CONFUSÃO NA VALIDAÇÃO — MODELO MULTI-4 CLASSES

Classes	EarlySys	HoloSys	MidSys	Normal
EarlySys	94,2%	2,4%	0,2%	2,6%
HoloSys	3,0%	95,6%	0,9%	0,6%
MidSys	0,0%	0,5%	99,5%	0,0%
Normal	4,1%	1,1%	0,0%	94,7%

B. Resultados no Ambiente Edge Impulse

No ambiente Edge Impulse, os modelos mantiveram alta performance no conjunto de teste, conforme demonstrado pelas matrizes de confusão nas Tabelas V, VI e VII. As latências e consumo de recursos são detalhados logo após.

a) *Modelo Binário*: A acurácia do teste foi de 92,63%, com latência total de 1.148 ms e uso de RAM de 80,7 KB. A matriz de confusão é mostrada na Tabela V.

Tabela V
MATRIZ DE CONFUSÃO NO TESTE — MODELO BINÁRIO ABORDAGEM 1

Classes	Normal	Sopro	Uncertain
Normal	93,3%	3,4%	3,2%
Sopro	4,8%	92,0%	3,2%

b) *Modelo Multi-3 Classes*: A acurácia foi de 95,52%, com latência de 1.110 ms e RAM de 85,5 KB. Matriz de confusão na Tabela VI.

Tabela VI
MATRIZ DE CONFUSÃO NO TESTE — MODELO 2 ABORDAGEM 1

Classes	EarlySys	HoloSys	MidSys	Uncertain
EarlySys	90,4%	6,4%	0,0%	3,0%
HoloSys	0,3%	92,0%	1,9%	0,0%
MidSys	0,0%	1,3%	98,2%	0,4%

c) *Modelo Multi-4 Classes*: Com acurácia de 94,55% e latência similar, a matriz de confusão está na Tabela VII.

Tabela VII
MATRIZ DE CONFUSÃO NO TESTE — MODELO ABORDAGEM 2

Classes	EarlySys	HoloSys	MidSys	Normal	Uncertain
EarlySys	88,9%	1,6%	1,2%	3,2%	5,2%
HoloSys	0,3%	92,0%	3,4%	0,3%	4,0%
MidSys	0,0%	0,8%	99,0%	0,0%	0,2%
Normal	2,3%	1,1%	0,0%	95,1%	1,5%

C. Implementação em Dispositivo Móvel

Os modelos desenvolvidos podem ser facilmente implementados em qualquer dispositivo móvel com acesso à internet por meio da função de *deployment* integrada da plataforma Edge Impulse. Essa funcionalidade permite a exportação rápida e prática dos modelos, disponibilizando um QR code que pode ser escaneado para executar o sistema diretamente no celular.

Para a avaliação prática, utilizou-se o microfone integrado do dispositivo móvel para captura dos sons cardíacos. Embora

essa abordagem simplifique o processo, o ideal é a utilização de estetoscópios digital conectados diretamente ao smartphone para garantir maior fidelidade do sinal e menor interferência.

A Figura 6 mostra a interface visual do aplicativo no celular após o deployment, evidenciando a simplicidade e facilidade de uso do sistema. Cada um dos modelos propostos conta com sua própria interface dedicada.

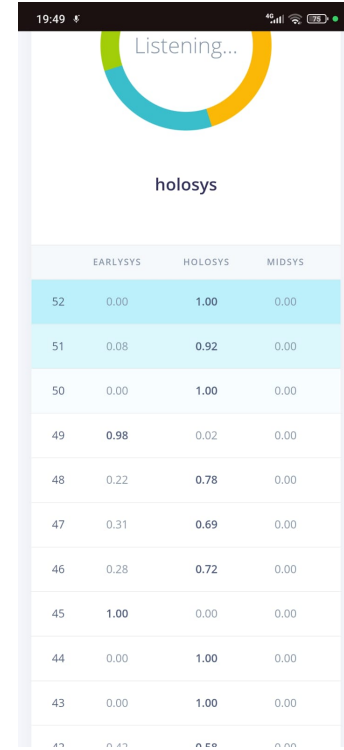


Figura 6. Interface do sistema de calificação de tipo de sopros cardíacos implementado em dispositivo móvel via Edge Impulse.

V. DISCUSSÃO

A comparação dos resultados obtidos com a proposta atual, resumida na Tabela VIII, evidencia que, embora a acurácia do modelo seja ligeiramente inferior a de arquiteturas mais complexas reportadas na literatura, o desempenho permanece competitivo e adequado para aplicações práticas. Por exemplo, McDonald et al. [1] apresentam modelos baseados em redes neurais recorrentes combinadas com modelos ocultos semi-Markovianos (HSM), alcançando alta performance às custas de maior complexidade computacional. De modo semelhante, Luo et al. [4] aplicam um método semi-supervisionado com Mean Teacher, conseguindo acurácias elevadas, porém exigindo maior poder de processamento.

Modelos baseados em ensembles de CNN, como o proposto por Elola et al. [15], demonstram robustez no desempenho, porém com arquiteturas pesadas, limitando a viabilidade para dispositivos embarcados de baixa potência. Nie et al. [6] apresentam um modelo multitarefa que também requer considerável capacidade computacional.

Tabela VIII
COMPARAÇÃO DE MODELOS PARA DETECÇÃO DE SOPROS CARDÍACOS

Autor	Arquitetura	Complexidade	Accuracy	Conjunto
McDonald et al.	RNN+HSMM	Alta	69.3%	Test
Luo et al.	CNN+Mean Teacher	Média-alta	96.8%	Val
Ganguly & Sharma	ML Clássico	Baixa	92.0%	Val
Elola et al.	CNN	Alta	80.4%	Test
Nie et al.	2D CNN MTL	Média	95.0%	Test
Proposta Atual I	CNN Compacta	Baixa	92,63% 95,52%	Test
Proposta Atual II	CNN Compacta	Baixa	94,55%	Test

Por outro lado, a simplicidade arquitetural do modelo aqui proposto, detalhado na Tabela VIII, destaca-se pela utilização de uma CNN compacta com apenas cinco camadas. Apesar de alcançar uma acurácia um pouco menor (entre 92,63% - 95,53% no teste), às abordagens permitem a implementação eficiente em dispositivos móveis e microcontroladores como Arduino e ESP32, conforme indicado na Tabela IX. Essa portabilidade representa um avanço importante para a democratização do diagnóstico precoce de sopros cardíacos, alinhando-se com as tendências do Tiny Machine Learning (TinyML) [10]–[12].

Tabela IX
COMPATIBILIDADE TINYML E PLATAFORMAS DE IMPLEMENTAÇÃO

Autor	TinyML	Plataforma
McDonald et al. '24	✗	PC / GPU
Luo et al. '24	✗	PC
Elola et al. '23	✗	PC
Nie et al. '24	✗	PC
Proposta Atual	✓	Smartphone/ESP32/Arduino

Além disso, a plataforma Edge Impulse facilita o deployment dos modelos em múltiplas plataformas, preservando baixo consumo energético e reduzindo latência na inferência, aspectos cruciais para aplicações clínicas móveis e embarcadas [5].

Em síntese, a principal contribuição deste trabalho reside na combinação de desempenho competitivo, simplicidade estrutural e eficiência computacional, viabilizando soluções acessíveis e portáteis para diagnóstico assistido de sopros cardíacos.

VI. CONCLUSÃO

Este trabalho desenvolveu e avaliou modelos de aprendizado de máquina para a detecção e classificação de sopros cardíacos a partir de fonocardiogramas digitais, com foco na simplicidade arquitetural e eficiência para deployment

em dispositivos móveis e embarcados. As duas abordagens propostas, utilizando arquiteturas CNN compactas e extração de características via Mel Frequency Energy, mostraram-se eficazes, alcançando acurácias superiores a 92% nos testes realizados no conjunto PhysioNet 2022.

A utilização da plataforma Edge Impulse permitiu a segmentação manual, treinamento e deployment dos modelos, evidenciando a viabilidade de execução em dispositivos com recursos limitados, como smartphones, Arduino e ESP32. A portabilidade e baixo consumo energético são vantagens essenciais para ampliar o acesso ao diagnóstico precoce de anomalias cardíacas, especialmente em regiões remotas.

Futuras pesquisas poderão explorar a ampliação dos conjuntos de dados, o refinamento automático da segmentação e a integração de métodos semi-supervisionados para aumentar ainda mais a robustez e aplicabilidade clínica dos sistemas desenvolvidos.

REFERÊNCIAS

- [1] A. McDonald, M. J. F. Gales, and A. Agarwal, "A recurrent neural network and parallel hidden markov model algorithm to segment and detect heart murmurs in phonocardiograms," *PLOS Digital Health*, vol. 3, no. 11, p. e0000436, 2024.
- [2] A. Ganguly and M. Sharma, "Detection of pathological heart murmurs by feature extraction of phonocardiogram signals," *J. Appl. Adv. Res.*, vol. 2, no. 4, pp. 200–205, 2017.
- [3] A. Sideras and T. Boura, "Heart murmur detection using machine learning," MSc Thesis, National Centre for Scientific Research Demokritos, University of Piraeus, Athens, Greece, 2023.
- [4] Y. Luo and et al., "Phonocardiogram (pcg) murmur detection based on the mean teacher method," *Sensors*, vol. 24, p. 6646, 2024.
- [5] H. Luo and et al., "Smartphone as an electronic stethoscope: factors influencing heart sound quality," *European Heart Journal - Digital Health*, vol. 3, pp. 473–480, 2022.
- [6] J. Nie, R. Liu, B. Mahasseni, E. Azemi, and V. Mitra, "Model-driven heart rate estimation and heart murmur detection based on phonocardiogram," in *2024 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, London, UK, 2024, pp. 1–6.
- [7] Y. Zhao and et al., "Multi-class classification of heart sounds using deep learning," *IEEE Access*, vol. 9, pp. 45 023–45 033, 2021.
- [8] R. Singh and et al., "Data augmentation techniques for heart sound classification," *Biomedical Signal Processing and Control*, vol. 59, p. 101917, 2020.
- [9] J. Kim and H. Lee, "Recurrent neural networks for phonocardiogram classification," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 2, pp. 462–470, 2019.
- [10] C. Banbury, S. Bhattacharyya, and B. L. Evans, "Tinymml: Enabling low-power machine learning applications," *Proceedings of the IEEE*, vol. 109, no. 11, pp. 1939–1953, 2021.
- [11] P. Warden and D. Situnayake, *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*. O'Reilly Media, Inc., 2019.
- [12] Edge Impulse, "Edge impulse - tinymml platform," <https://www.edgeimpulse.com/>, accessed: 2025-07-01.
- [13] L. Wang and et al., "Advanced signal processing techniques for cardiac sound analysis," *Computers in Biology and Medicine*, vol. 150, p. 106123, 2022.
- [14] A. Patel and S. Kumar, "Optimized cnn architectures for portable cardiac sound analysis," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 3, pp. 856–865, 2023.
- [15] V. Elola and et al., "Beyond heart murmur detection: A multi-label classification of phonocardiogram recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 4, pp. 1730–1741, 2023.