DiffWave para Melhoria da Qualidade de Fonocardiogramas Digitais com Foco no Apoio ao Diagnóstico Clínico

Adrian Alejandro Chavez Alanes
Instituto Nacional de Telecomunicações – INATEL
Santa Rita do Sapucaí, Brasil
adrian@mtel.inatel.br

***Abstract*—As doenças cardiovasculares permanecem a principal causa de mortalidade global, destacando a necessidade de métodos de triagem e monitoramento cada vez mais precisos. A ausculta cardíaca é amplamente utilizada, mas sofre com limitações de qualidade de sinal e subjetividade na interpretação. Este trabalho propõe o uso de técnicas modernas de melhoria de áudio (*audio enhancement*) aplicadas a fonocardiogramas digitais, com foco no desenvolvimento de versões compactas da arquitetura DiffWave. O objetivo é fornecer sinais acústicos mais claros e confiáveis, preservando os sons cardíacos essenciais (S1, S2 e sopros) e reduzindo interferências externas. A proposta se diferencia de abordagens anteriores ao buscar modelos leves, adequados à execução em dispositivos móveis, permitindo suporte à ausculta. Neste estudo será utilizado o CirCor DigiScope Dataset, avaliando o desempenho por meio de métricas objetivas de redução de ruído e fidelidade espectral. Espera-se que a solução contribua para aumentar a confiabilidade da ausculta em contextos clínicos, especialmente em regiões com recursos limitados.**

I. INTRODUÇÃO

As doenças cardiovasculares representam aproximadamente 32% de todas as mortes globais, totalizando quase 18 milhões de óbitos anuais [1]. No Brasil, assim como em diversos países, as condições cardiovasculares constituem um dos principais problemas de saúde pública, sendo responsáveis por internações recorrentes, alto custo hospitalar e perda de qualidade de vida.

A detecção precoce de anomalias cardíacas, como sopros, arritmias e alterações valvares, é essencial para o encaminhamento clínico adequado e para a prevenção de complicações graves, como insuficiência cardíaca ou eventos isquêmicos. No entanto, a prática da ausculta cardíaca permanece altamente dependente da experiência do profissional, do ambiente em que o exame é realizado e da qualidade do dispositivo utilizado. Mesmo com estetoscópios digitais, a presença de ruídos ambientais e artefatos de captação continua sendo um obstáculo importante.

A fonocardiografia digital ampliou as possibilidades de análise ao permitir o registro e processamento dos sons cardíacos. Nos últimos anos, modelos de aprendizado profundo têm sido aplicados com sucesso à classificação automática de sopros e à segmentação do ciclo cardíaco. Entretanto, ainda há uma lacuna significativa na aplicação de modelos voltados exclusivamente à melhoria da qualidade do áudio. Métodos tradicionais de filtragem podem suprimir murmúrios patológicos ou distorcer os sons cardíacos, enquanto abordagens de aprendizado profundo previamente exploradas (como U-Net e LU-Net) apresentam limitações relacionadas ao uso de ruído sintético e ao alto custo computacional.

Diante desse cenário, este trabalho propõe a adaptação da arquitetura recente de melhoria de áudio DiffWave [2], para o domínio dos fonocardiogramas. A inovação consiste em desenvolver versões mais compactas e ajustadas ao espectro acústico cardíaco (20–800 Hz), capazes de reduzir ruídos externos e preservar S1, S2 e sopros. Além disso, busca-se demonstrar a viabilidade de execução desses modelos em dispositivos móveis, possibilitando suporte à ausculta sem substituir a decisão clínica humana.

II. TRABALHOS RELACIONADOS

A literatura recente em fonocardiografia digital mostra um avanço concreto em métodos de melhoria de sinal e separação de fontes, com impacto direto na ausculta clínica assistida por computador. Um estudo clássico em cenário de competição clínica avaliou quatro algoritmos de melhoria de áudio aplicados diretamente antes da classificação e observou queda de desempenho por supressão inadvertida de sopros; por outro lado, quando a melhoria de áudio por subtração espectral com estimação de Wiener foi usado como pré-processamento para segmentação, a acurácia global do sistema aumentou (sensibilidade 96%, especificidade 74%) [3].

Em paralelo, revisões abrangentes sistematizam que a remoção de ruído permanece etapa essencial do fluxo de análise de sons cardíacos, destacando filtros passa-banda, subtração espectral e seleção de segmentos de melhor qualidade como procedimentos recorrentes [4], [5].

Com aprendizado profundo, surgiram arquiteturas específicas para o domínio do fonocardiograma (PCG) com ruído realista. A LU-Net, proposta como encoder–decoder leve, foi treinada com mistura de ruídos respiratórios e ambiência hospitalar, reportando melhora média de SNR e aumento substancial de desempenho em porções ruidosas com cerca de 1,3 milhões de parâmetros [6].

Em linha semelhante, um U-Net para melhoria de áudio de PCG corrompido por ruídos do mundo real (fala infantil, tosse, espirro, amassados) emprega STFT curta e reconstrói amplitude e fase, superando baselines como limiarização em wavelets e autoencoder de melhoria de áudio em métricas objetivas e avaliação qualitativa [7]. Outra abordagem recente explora também revisões de técnicas de aprendizado profundo aplicadas ao processamento de PCG, enfatizando como novas arquiteturas podem preservar murmúrios e padrões sutis em cenários ruidosos [8].

Além da melhoria de áudio, há interesse crescente na separação de fontes coração–pulmão como forma de melhoria de ausculta. Nesse contexto, arquiteturas de separação temporal originalmente aplicadas a áudio geral, como Demucs, têm sido destacadas como modelos de referência para processar sinais biomédicos de forma eficiente [9]. De forma complementar, trabalhos recentes em classificação baseada em representações tempo-frequenciais reforçam que modelos leves podem atuar como pré-processamento promissor em pipelines de PCG [10].

Em síntese, os resultados convergem para três recomendações: projetar melhoria de áudio que preserve murmúrios e componentes fisiológicos de interesse; treinar com ruídos realistas e protocolos controlados de mistura; e priorizar arquiteturas compactas com perdas no domínio temporal e espectral e integração embarcada.

III. PROBLEMA

A prática clínica da ausculta enfrenta limitações significativas. Ruídos ambientais comuns em hospitais (conversas, alarmes, movimentação de equipamentos), artefatos provocados pelo atrito da campana, respiração do paciente e até interferência elétrica prejudicam a clareza do exame. Esses fatores reduzem a confiabilidade do diagnóstico auditivo e dificultam o processo de ensino da semiologia cardíaca.

Embora estetoscópios digitais capturem sinais com maior fidelidade, essas fontes de ruído permanecem presentes e afetam tanto a interpretação por profissionais quanto o desempenho de algoritmos de classificação automática treinados em condições ideais. Métodos tradicionais de filtragem apresentam limitações, podendo atenuar murmúrios patológicos junto com o ruído, enquanto abordagens de aprendizado profundo previamente exploradas, como U-Net e LU-Net, utilizam muitas vezes ruídos sintéticos e demandam elevado custo computacional.

Portanto, o desafio é propor uma solução que realize a melhoria dos fonocardiogramas preservando sopros e sons cardíacos fisiológicos, mas que ao mesmo tempo seja leve o suficiente para execução em dispositivos móveis. A ausência de arquiteturas adaptadas especificamente a este cenário evidencia a necessidade de desenvolver versões de modelos que trabalham com difução, como DiffWave [2], capaz de oferecer suporte à ausculta e aumentar a confiabilidade do exame sem comprometer a interpretação clínica.

IV. HIPÓTESE

A hipótese central deste trabalho é que arquiteturas modernas de melhoria de áudio, em especial o DiffWave [2], podem ser adaptadas ao domínio biomédico de forma a reduzir ruídos externos sem comprometer a integridade dos sons cardíacos (S1, S2 e sopros).

Parte-se do pressuposto de que redes neurais treinadas com pares de fonocardiogramas ruidosos e limpos são capazes de aprender representações discriminativas que diferenciam padrões fisiológicos de ruídos ambientais. Enquanto estudos anteriores utilizaram arquiteturas como U-Net ou LU-Net, frequentemente com ruídos sintéticos, este trabalho propõe desenvolver uma opção usando DiffWave, ajustada ao espectro restrito do PCG e adequadas.

Assim, acredita-se que essas variantes leves sejam capazes de fornecer sinais cardíacos igual ou mais claros e confiáveis, beneficiando tanto a prática clínica. Pretende-se demonstrar sua viabildiade em smartphones, isso configuraria um avanço prático sem substituir a decisão clínica humana.

V. METODOLOGIA

A metodologia proposta está organizada em múltiplas etapas, detalhadas a seguir.

A. Coleta de Dados

Serão utilizados registros do CirCor DigiScope Dataset [11], um dos maiores bancos públicos de fonocardiogramas digitais disponíveis. Esse conjunto contém mais de 4.000 gravações coletadas em pacientes pediátricos e adultos, obtidas em diferentes pontos de ausculta.

Além da captura bruta dos sinais, o dataset inclui anotações de especialistas em cardiologia sobre a presença e o tipo de sopros, assim como informações sobre ruídos presentes durante o exame. Essa riqueza de informações permite tanto o treinamento supervisionado de classificadores quanto a construção de pares limpos e ruidosos para o presente trabalho.

B. Construção dos Pares Noisy–Clean

Para o treinamento supervisionado de modelos de melhoria, cada segmento de áudio limpo será associado a uma versão ruidosa correspondente.

- Clean: obtido a partir de sinais de boa qualidade do dataset, processados com filtro passa-banda, normalização de amplitude e supressão suave de ruído estacionário. Sopros e componentes fisiológicos (S1/S2) serão preservados integralmente.
- Noisy: obtido de duas formas: (i) gravações já degradadas naturalmente no dataset (respiração intensa, fala, fricção da campana); (ii) adição de ruídos artificiais como ruído branco ou voz ambiente. Diferentes níveis de SNR serão aplicados para garantir robustez.

Esse processo garante diversidade de exemplos e prepara o modelo para operar em condições adversas típicas de ambientes hospitalares.

C. Pré-processamento

Os sinais serão segmentados em janelas de 5 segundos, tempo suficiente para capturar múltiplos ciclos cardíacos. Cada segmento passará por:

- Normalização temporal.
- Filtragem digital passa-banda (20–800 Hz) para eliminar frequências irrelevantes.
- Balanceamento entre segmentos limpos e ruidosos para assegurar diversidade no treino e validação.

D. Modelagem

Serão implementadas duas arquiteturas de referência para melhoria de áudio por difução: DiffWave [2].

Este será configurado em versão simplificada, explorando menor profundidade de camadas e parâmetros otimizados ao espectro cardíaco (20–800 Hz). Será treinado em pares de sinais limpos e ruidosos, com funções de perda que combinem métricas no domínio temporal e espectral. Estratégias de compactação como quantização serão aplicadas para viabilizar posterior execução embarcada.

E. Treinamento e Validação

O conjunto de dados será dividido em 70% para treinamento, 20% para validação durante o ajuste dos modelos e 10% para teste final. Durante o treinamento, os modelos receberão como entrada segmentos ruidosos e terão como alvo os segmentos correspondentes considerados limpos, aprendendo a reduzir interferências e preservar os sons cardíacos principais.

Para aumentar a robustez, serão aplicadas técnicas de aumento de dados, como variação de volume, pequenas mudanças de velocidade e adição de diferentes tipos de ruído.

A validação será realizada de forma objetiva, comparando os sinais de saída com suas referências limpas por meio de métricas consagradas em processamento de áudio:

- SNRi (Signal-to-Noise Ratio improvement): Mede o quanto o modelo aumenta a relação sinal–ruído em relação ao sinal original.
- SI-SDR (Scale-Invariant Signal-to-Distortion Ratio): Avalia a proximidade entre o sinal processado e o de referência, independentemente da escala.
- LSD (Log-Spectral Distance): Verifica a fidelidade espectral, importante para preservar características acústicas de S1, S2 e sopros.

Esses indicadores permitirão quantificar de maneira clara a diferença entre o áudio de entrada e o áudio aprimorado, demonstrando a eficácia do modelo.

F. Implementação

O modelo com melhor desempenho será exportado para execução em dispositivos móveis utilizando TensorFlow Lite. Serão avaliados tempo de inferência e espaço em memória, verificando a viabilidade de aplicação. Além disso, será analisada a preservação clínica do sinal processado, assegurando que os componentes acústicos relevantes permaneçam intactos durante a execução embarcada.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (cvds)," https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29, 2021.

[2] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, "Diffwave: A versatile diffusion model for audio synthesis," in *International Conference on Learning Representations (ICLR)*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2009.09761

[3] M. H. Asmare, G. D. Clifford, and D. B. Springer, "Can heart sound denoising be beneficial in phonocardiogram classification tasks?" in *43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2021, pp. 657–660. [Online]. Available: https://doi.org/10.1109/EMBC46164.2021.9630454

[4] X. Ren, J. Pan, H. Wang, W. Chen, and S. Li, "A comprehensive survey on heart sound analysis in the deep learning era," *Biomedical Signal Processing and Control*, vol. 83, p. 104643, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2301.09362

[5] S. N. Ali, S. B. Shuvo, and T. Hasan, "A review on deep learning methods for heart sound signal analysis," *Frontiers in Artificial Intelligence*, vol. 7, p. 1434022, 2024. [Online]. Available: https://doi.org/10.3389/frai.2024.1434022

[6] S. N. Ali, S. B. Shuvo, M. I. S. Al-Manzo, A. Hasan, and T. Hasan, "An end-to-end deep learning framework for real-time denoising of heart sounds for cardiac disease detection in unseen noise," *IEEE Access*, vol. 11, pp. 86 123–86 140, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3292551

[7] S. Mukherjee, A. Singh, A. Dey, and S. Banerjee, "A novel u-net architecture for denoising of real-world noise corrupted phonocardiogram signal," *arXiv preprint arXiv:2310.00216*, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2310.00216

[8] J. Wang, Q. Liu, H. Xu, and L. Chen, "Deep learning in heart sound analysis: From techniques to clinical applications," *Frontiers in Physiology*, vol. 15, p. 11461928, 2024. [Online]. Available: https://doi.org/10.34133/hds.0182

[9] A. Defossez, G. Synnaeve, and Y. Adi, "Real time speech enhancement in the waveform domain," *arXiv preprint arXiv:2006.12847*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2006.12847

[10] Y. Zhang, H. Li, W. Chen, and S. Wang, "A deep-learning approach to heart sound classification based on combined time-frequency representations," *Technologies*, vol. 13, no. 4, p. 147, 2025. [Online]. Available: https://doi.org/10.3390/technologies13040147

[11] D. Springer, E. K. J. Tang, J. D. P. Howard, A. Manfredi, and P. Lio, "The circor digiscope phonocardiogram dataset: From murmur detection to clinical outcomes," https://physionet.org/content/circor-heart-sound/1.0.3/, 2022.