

Detecção de sopro cardíaco utilizando uma rede neural convolucional

João H. V. M. Melo, Magno T. M. Silva, Renato Candido

Resumo— Este trabalho apresenta um sistema para detecção automática de sopro cardíaco a partir de sinais de fonocardiograma utilizando técnicas de aprendizado de máquina. A abordagem proposta combina a extração de características acústicas com uma rede neural convolucional para realizar a classificação entre sinais normais e sinais com sopro cardíaco. Os resultados indicam que a metodologia proposta é promissora para aplicações de triagem cardíaca assistida por computador.

Palavras-Chave— Aprendizado de máquina, classificação de sopros cardíacos, rede neural convolucional, fonocardiograma, coeficientes cepstrais de frequência mel.

Abstract— This work presents an automatic heart murmur detection system based on phonocardiogram signals using machine learning techniques. The proposed approach combines acoustic feature extraction with a convolutional neural network to classify normal and murmur heart sounds. The results indicate that the proposed methodology is a promising approach for computer-aided cardiac screening applications.

Keywords— Machine learning, heart murmur classification, convolutional neural network, phonocardiogram, mel-frequency cepstral coefficients.

I. INTRODUÇÃO

Uma das etapas preliminares na avaliação do sistema cardiovascular é a análise dos sons cardíacos [1], [2]. Apesar de ter mais de 200 anos, o estetoscópio analógico depende da audição subjetiva e da experiência do médico [3]. Diante disso, o fonocardiograma (*phonocardiogram* – PCG) é uma alternativa para a detecção de problemas cardíacos. Diferente da ausculta tradicional (auditiva), o PCG transforma vibrações mecânicas em sinais elétricos e, posteriormente, em registros visuais que podem ser armazenados e analisados. Ele permite a correlação precisa entre os sons (S1, S2, sopros) e as fases da sístole e diástole, servindo como uma ferramenta de documentação clínica superior à memória auditiva [4].

Neste artigo, aborda-se a detecção de sopro cardíaco com o uso de sinais de PCG dos bancos de dados *The Pascal Classifying Heart Sound* [5] e de Yaseen e Kwon [10]. Como modelo, considerou-se uma rede neural convolucional (*convolutional neural network* – CNN) [6], alimentada com coeficientes cepstrais de frequência mel (*Mel-Frequency Cepstral Coefficients* – MFCCs) [7]–[9]. O artigo está organizado da seguinte forma. Na Seção II, descrevem-se os bancos de dados utilizados. Na Seção III, aborda-se a etapa de pré-processamento do sinal de fonocardiograma e na Seção IV,

a arquitetura da CNN proposta. As Seções V e VI contêm respectivamente os resultados de simulação e as principais conclusões do trabalho.

II. BANCOS DE DADOS

O banco de dados PASCAL [5] é dividido em dois conjuntos principais (A e B). O conjunto A foi coletado via aplicativo móvel (iStethoscope) em ambiente não controlado e o conjunto B foi obtido em ambiente clínico utilizando estetoscópios digitais. Esses conjuntos categorizam os sons em normal, *murmur* (sopro), *extrasystole* e *extra heart sound*. Neste estudo, utilizaram-se apenas os sinais das categorias normal e sopro do conjunto B. Essa escolha justifica-se pela maior quantidade de arquivos já que o conjunto B contém 320 arquivos de áudios no formato wav, categorizados como normal e 95 categorizados como sopro, amostrados com 4 kHz. Para a avaliação do modelo, adotou-se a divisão de treino e teste, utilizando 80% dos dados para treino e 20% para teste. Essa estratégia permite avaliar a capacidade de generalização e resiliência do classificador. Também foi utilizado o banco de dados Yaseen [10], contendo 200 arquivos classificados como normal e 800 como diferentes tipos de sopro (AS, MR, MS e MVP), amostrados com 8 kHz.

III. ETAPA DE PRÉ-PROCESSAMENTO DO SINAL

No conjunto de dados Yaseen, os sinais foram reamostrados para uma frequência de 4 KHz. Para sincronizar os registros, empregou-se a energia de Shannon na extração da envoltória do sinal e na localização das bulhas cardíacas (S1 e S2, principais eventos presentes em sinais de fonocardiograma) [11]. A energia de Shannon foi escolhida por sua capacidade de realçar picos de média e baixa intensidade, evidenciando os principais eventos cardíacos e atenuando ruídos de fundo presentes nos registros, característica particularmente importante para os sinais do banco de dados PASCAL [12], [13]. A Fig. 1 apresenta um exemplo de um sinal de PCG e sua respectiva envoltória obtida por meio da energia de Shannon.

O comprimento dos áudios apresenta variação entre os registros, podendo assumir valores de 1 a 10 segundos. Para padronizar o vetor de entrada da rede neural, os sinais de ambos os bancos de dados foram limitados a segmentos de 2 segundos, duração suficiente para abranger aproximadamente dois ciclos cardíacos e evitar que o zero *padding* influenciasse significativamente o processo de treinamento. Para a seleção dos segmentos, foram empregados algoritmos de detecção de máximos locais (*peak detection*) para identificar o instante correspondente ao primeiro evento cardíaco de maior energia. A partir desse ponto de referência, o áudio foi segmentado, utilizando-se apenas o trecho subsequente de 2 segundos como

entrada para o treinamento do modelo, conforme ilustrado na Fig. 2. Depois de sincronizar e limitar os sinais de áudio, os MFCCs foram calculados considerando 15 coeficientes, janela e passo de 2048 e 512 amostras respectivamente, dando origem a uma matriz de dimensões 15×16 , como ilustrado na Fig. 3.

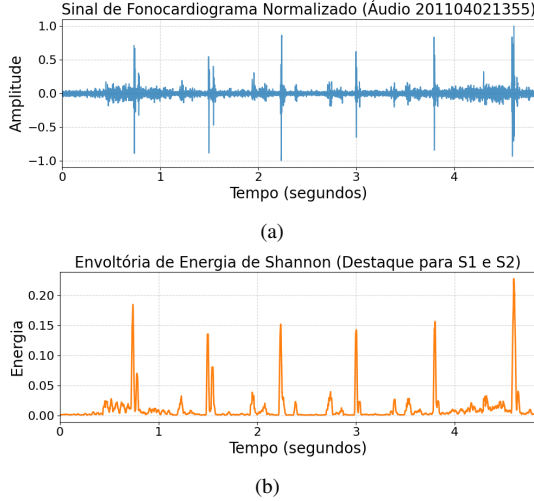


Fig. 1. Comparação temporal evidenciando picos sistólicos e diastólicos: (a) Sinal de fonocardiograma normalizado no domínio do tempo; (b) Envoltória da Energia de Shannon destacando as bulhas cardíacas principais (S1 e S2).

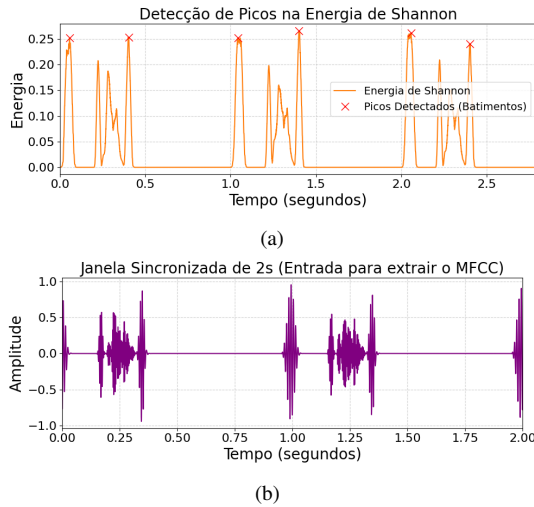


Fig. 2. Processo de sincronização e segmentação do sinal de fonocardiograma: (a) Detecção dos picos cardíacos utilizando a envoltória gerada pela energia de Shannon; (b) Janela sincronizada de 2,0 segundos, segmentada a partir do pico detectado e pronta para a extração dos MFCCs.

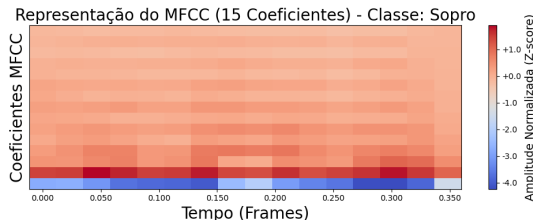


Fig. 3. Imagem gerada após a extração dos MFCCs de um fonocardiograma contendo sopro cardíaco.

IV. REDE NEURAL CONVOLUCIONAL

A arquitetura CNN proposta está mostrada na Fig. 4. Ela é composta por dois blocos de extração de características (camadas convolucionais) seguidos por uma etapa de classificação (um neurônio). A camada de entrada recebe a matriz de

MFCCs com dimensões 15×16 . Nela, é aplicada uma convolução com 8 filtros 3×3 , utilizando *stride* igual a 1 e *padding* de 1. O resultado passa pela função de ativação ReLU, seguida de uma operação de subamostragem (*MaxPooling*) com janela 2×2 e passo 2, responsável por reduzir a dimensionalidade do mapa de características. A segunda camada convolucional recebe os 8 mapas de características gerados e aplica uma nova convolução com 16 filtros 3×3 , passando por uma função de ativação ReLU e por uma segunda camada de *MaxPooling* 2×2 . Após a extração, a estrutura tridimensional resultante é vetorizada. Esse vetor de dimensão 192×1 alimenta um único neurônio de saída com função de ativação sigmoideal.

O processo de aprendizado foi executado ao longo de 1000 épocas, dividindo o conjunto de dados em lotes (*batches*) de tamanho 16. A atualização dos pesos da rede foi realizada com o otimizador Adam (*Adaptive Moment Estimation*) com taxa de aprendizado 0,001, $\beta_1 = 0,9$ e $\beta_2 = 0,999$. Além disso, como a finalidade do problema consiste em uma classificação binária, adotou-se a entropia cruzada binária (*binary cross-entropy* – BCE) como função custo [6].

V. RESULTADOS

Para avaliar o processo de aprendizado e a capacidade de generalização da arquitetura proposta, monitorou-se a evolução da função custo ao longo do treinamento. As curvas da entropia cruzada para os conjuntos de treinamento e validação são apresentadas na Fig. 5. Observa-se que, após um determinado número de épocas, a função custo do conjunto de treinamento continua diminuindo, enquanto a do conjunto de validação deixa de apresentar melhorias, caracterizando o início de *overfitting*. Para mitigar esse efeito, adotou-se a estratégia de *early stopping*, selecionando o modelo correspondente à época que apresentou o menor erro de validação.

O desempenho do classificador foi avaliado por meio das métricas Acurácia, *Recall*, Especificidade, Precisão e F_1 -Score, calculadas respectivamente como $Acurácia = (VP + VN)/(VP + VN + FP + FN)$, $Recall = VP/(VP + FN)$, $Especificidade = VN/(VN + FP)$, $Precisão = VP/(VP + FP)$ e $F1\text{-Score} = 2 \cdot (Precisão \cdot Recall)/(Precisão + Recall)$, em que VP , VN , FP e FN denotam os números de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, respectivamente.

A CNN foi treinada exclusivamente com os sinais das classes normal e sopro do conjunto B do banco de dados PASCAL [5] e posteriormente avaliada tanto em um conjunto de teste proveniente da própria base quanto no conjunto de dados Yaseen. Os resultados obtidos utilizando apenas dados do PASCAL são apresentados na Tabela I. Todas as métricas de desempenho alcançadas pela arquitetura proposta foram superiores a 90%, evidenciando elevado desempenho na distinção entre sinais normais e sinais com sopro.

Com o objetivo de avaliar a capacidade de generalização sob mudança de domínio, o modelo treinado no banco PASCAL foi aplicado diretamente ao banco de dados Yaseen, sem qualquer etapa de retreinamento ou adaptação. Os resultados obtidos são apresentados na Tabela II. Observa-se uma redução significativa do desempenho, com Acurácia de 54,10% e Especificidade de 43,50%, indicando que diferenças nas condições de aquisição, nas características acústicas e na distribuição dos

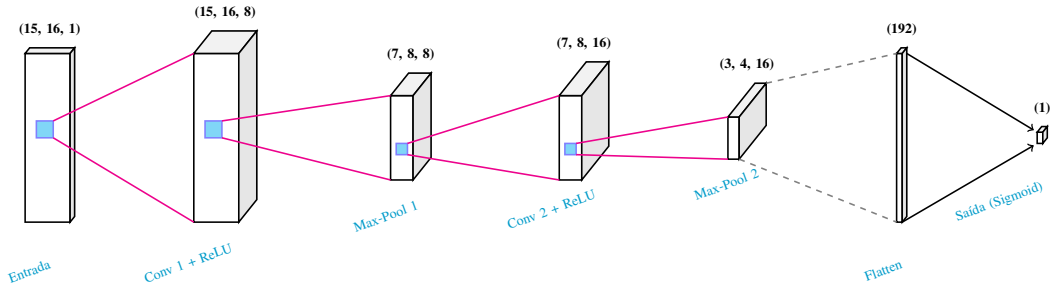


Fig. 4. Arquitetura detalhada da CNN proposta para classificação da matriz MFCC.

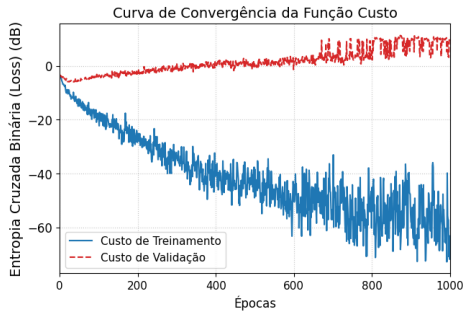


Fig. 5. Curva de convergência da função de custo (Entropia Cruzada Binária) para os conjuntos de treinamento e validação ao longo de 1000 épocas.

dados entre as duas bases afetam significativamente a capacidade de generalização da CNN. Em contrapartida, a Precisão permaneceu relativamente elevada (80,08%), sugerindo que, quando o modelo identifica a presença de sopro, essa decisão tende a ser confiável, embora a elevada incidência de falsos positivos e falsos negativos comprometa o desempenho global.

TABELA I

Desempenho da CNN - PASCAL	
Métrica	Valor (%)
Acurácia	96,08
Recall	90,01
Especificidade	97,50
Precisão	90,91
F ₁ -Score	90,91
Desempenho da MLP de [14] (4 K-Folds)	
Acurácia (%)	63,54

TABELA II

Desempenho da CNN - PASCAL vs Yaseen	
Métrica	Valor (%)
Acurácia	54,10
Recall	56,75
Especificidade	43,50
Precisão	80,08
F ₁ -Score	66,42

Para validar a arquitetura proposta, realizou-se ainda uma comparação com [14], no qual foi empregada uma rede *multi-layer perceptron* (MLP) alimentada por MFCCs e avaliada por meio de validação cruzada com quatro partições [6]. Essa comparação é particularmente relevante por utilizar características acústicas semelhantes às adotadas neste trabalho, diferindo principalmente na arquitetura do classificador. Embora os autores tenham reportado acurácias de até 87,5% em partições específicas da validação cruzada, a acurácia média global obtida foi de 63,54%, conforme apresentado na Tabela I. Em contraste, a CNN proposta atingiu 96,08% de acurácia utilizando apenas a divisão convencional entre treinamento e teste

no banco PASCAL, indicando que a arquitetura convolucional foi mais eficaz na extração de padrões discriminantes dos coeficientes MFCC.

VI. CONCLUSÕES

A arquitetura proposta demonstrou elevado desempenho na detecção de sopros cardíacos a partir de coeficientes MFCC, alcançando acurácia de 96,08% e especificidade de 97,50% no banco de dados PASCAL. Esses resultados evidenciam a eficácia da CNN na extração de características discriminantes para a classificação de fonocardiogramas, superando o desempenho médio reportado por trabalhos correlatos baseados em MLP. Entretanto, a avaliação cruzada entre os bancos PASCAL e Yaseen revelou uma redução significativa no desempenho, evidenciando que a robustez frente à transferência de domínio ainda representa um desafio. Como perspectiva futura, pretende-se investigar estratégias de treinamento com bases de dados mais diversificadas, bem como técnicas de adaptação de domínio, visando aumentar a capacidade de generalização do modelo para diferentes condições de aquisição e populações.

REFERÊNCIAS

[1] T. Chakrabarti *et al.*, “Phonocardiogram signal analysis - practices, trends and challenges: A critical review,” *Proc. Int. Conf. and Workshop Comput. Commun. (IEMCON)*, pp. 1-4, 2015.

[2] S. K. Berkaya *et al.*, “A survey on ECG analysis,” *Biomed. Signal Process. Control*, vol. 43, pp. 216–235, 2018.

[3] A. D. Levin *et al.*, “A Machine Learning Approach to Heart Murmur Detection and Classification,” *Proc. Int. Congr. Image Signal Process., Biomed. Eng. Informatics (CISP-BMEI)*, pp. 521-525, 2020.

[4] M. R. Montinari and S. Minelli, “The first 200 years of cardiac auscultation and future perspectives,” *J. Multidiscip. Healthc.*, vol. 12, pp. 183-189, 2019.

[5] E. Gomes, P. Bentley, M. Coimbra, E. Pereira and Y. Deng, “Classifying heart sounds: Approaches to the PASCAL challenge,” *Proc. Int. Conf. Health Informatics (HEALTHINF)*, pp. 337-340, 2013.

[6] C. M. Bishop, *Deep Learn.: Found. Concepts*. Springer, 2024.

[7] Z. K. Abdul and A. K. Al-Talabani, “Mel Frequency Cepstral Coefficient and its Applications: A Review,” *IEEE Access*, vol. 10, pp. 122136-122158, 2022.

[8] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.

[9] Y. M. Ali *et al.*, “Analysis of Accent-Sensitive Words in Multi-Resolution Mel-Frequency Cepstral Coefficients for Classification of Accents in Malaysian English,” *Int. J. Aut. Mech. Eng.*, vol. 7, pp. 1053-1073, 2013.

[10] S. G.-Y. Yaseen and S. Kwon, “Classification of heart sound signal using multiple features,” *Applied Sciences*, vol. 8, pp. 2444, 2018.

[11] H. Liang, S. Lukkarinen and I. Hartimo, “Heart sound segmentation algorithm based on heart sound envelopogram,” *Computers in Cardiology*, pp. 105-108, 1997.

[12] A. Moukadem *et al.*, “A robust heart sounds segmentation module based on S-transform,” *Biomed. Signal Process. Control*, vol. 8, pp. 273–281, 2013.

[13] Z. Zidelmal *et al.*, “QRS detection using S-Transform and Shannon energy,” *Comput. Methods Programs Biomed.*, vol. 116, pp. 1–9, 2014.

[14] C. Lubis and F. Gondawijaya, “Heart sound diagnose system with BFCC, MFCC, and backpropagation neural network,” *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 508, pp. 012119, 2019.