

Introdução

No cenário da pandemia de COVID-19, a tecnologia emergiu como uma aliada fundamental, catalisando projetos inovadores como o **SPIRA** [1] (Sistema de detecção Precoce de Insuficiência Respiratória por meio de análise de Áudio).

O projeto visa desenvolver uma ferramenta de **triagem remota** acessível, utilizando modelos de Inteligência Artificial para identificar biomarcadores de insuficiência respiratória por análise de audios. O objetivo é permitir o monitoramento de pacientes via smartphone, auxiliando na redução da sobrecarga hospitalar.

Uma das estratégias notáveis do projeto SPIRA para garantir a robustez de seu modelo foi a inserção artificial de ruído hospitalar nos dados de áudio, forçando o modelo a se tornar invariante às condições ambientais. Contudo, essa abordagem levantou uma questão de pesquisa pertinente: seria a inserção de ruído a melhor forma de mitigar o viés ambiental? Uma estratégia alternativa, e talvez mais intuitiva, seria a de remover o ruído existente por meio de algoritmos de filtragem.

Objetivos

O objetivo deste trabalho é investigar a eficácia e os desafios da aplicação de filtros de ruído como método de pré-processamento para a classificação de insuficiência respiratória a partir da voz. Espera-se, com isso, não apenas avaliar comparativamente as duas estratégias, mas também oferecer subsídios práticos e teóricos para o desenvolvimento de sistemas mais confiáveis e robustos em ambientes reais de uso.

Modelos de Classificação

Modelos utilizados para *fine-tuning* nos nossos dados:

CNN10 - Rede Neural Convolucional

- **Tipo:** CNN da família das PANNs (Pre-trained Audio Neural Networks).
- **Pré-treinamento:** AudioSet.
- **Domínio:** Tempo-Frequência (*Log-mel Spectrogram*)

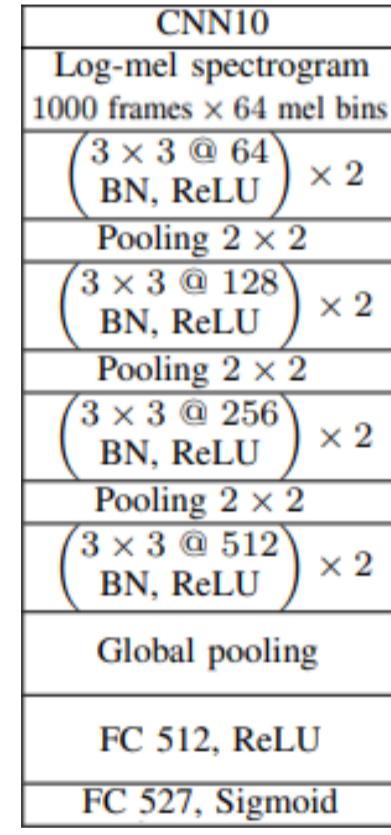


Figura: Arquitetura da CNN10 [2]

AudioMAE - Transformer (MAE)

- **Tipo:** Arquitetura baseada em Transformer (Masked Autoencoder).
- **Pré-treinamento:** AudioSet auto-supervisionado.
- **Domínio:** Tempo-Frequência (*Log-mel Spectrogram*)

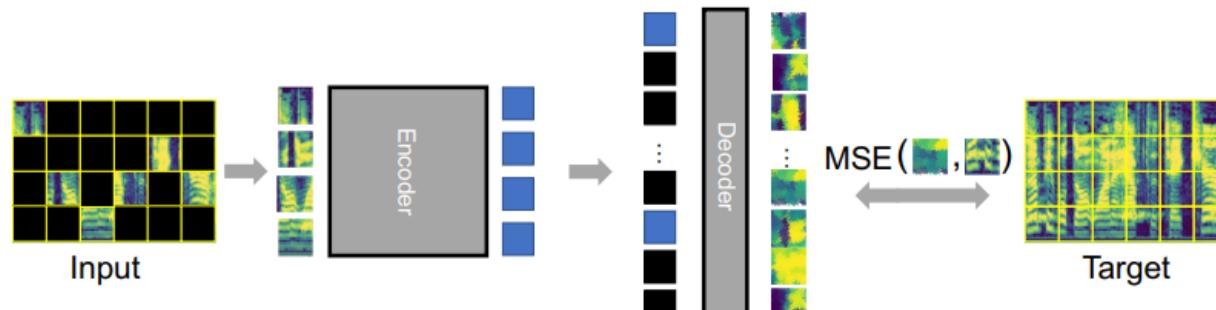


Figura: Esquema do Audio Masked Autoencoder [3]

Filtro Utilizado

Foi utilizada uma ferramenta de **subtração espectral** desenvolvida no contexto do projeto SPIRA que funciona em duas etapas principais:

- A primeira etapa do filtro consiste em um **classificador** que identifica e isola os segmentos do áudio que contêm apenas ruído de fundo operando no domínio do tempo.
- A segunda etapa, o **supressor** de ruído, utiliza o perfil de ruído detectado para realizar uma subtração espectral, operando no domínio tempo-frequência para atenuar as frequências associadas ao ruído no áudio original.

Conjunto de Dados

Conjunto	Homem (Controle)	Mulher (Controle)	Homem (Paciente)	Mulher (Paciente)	Total (Controle)	Total (Paciente)	Total
Treinamento	57	84	76	66	141	142	283
Validação	8	8	8	8	16	16	32
Teste	22	26	28	32	54	54	108

Figura: Dados utilizados

Metodologia

A metodologia consistiu em uma investigação sequencial para avaliar a robustez da filtragem contra o viés ambiental.

1. Estabilização do Filtro:

- Investigação dos parâmetros do filtro para evitar adição de viés por conta de fragmentos ocasionados pela filtragem;

2. Teste de Viés Ambiental:

- Avaliação dos modelos em ambos os cenários: Com e sem adição de ruído antes da filtragem para verificar a persistência ou não do viés ambiental.

Resultados

Os resultados finais, comparando o desempenho no conjunto original (sem ruído) e no conjunto modificado com inserção de ruído, revelaram a falha da estratégia de filtragem.

Modelo	Cenário de Teste	Acurácia	F1	Precisão	Recall	Especificidade
CNN10	Filtragem sem Inserção	0.9907	0.9908	0.9818	1.0000	0.9815
	Filtragem com Inserção	0.5278	0.6792	0.5143	1.0000	0.0556
AudioMAE	Filtragem sem Inserção	0.9907	0.9908	0.9818	1.0000	0.9815
	Filtragem com Inserção	0.5000	0.6667	0.5000	1.0000	0.0000

Figura: Comparação de desempenho dos modelos (CNN10 e AudioMAE) nos conjuntos de teste com e sem inserção de ruído hospitalar antes da filtragem.

Conclusão

Este trabalho demonstrou que a filtragem de ruído é uma estratégia insuficiente para mitigar vieses ambientais severos em datasets de áudios.

A persistência do *shortcut learning* em arquiteturas avançadas indica que a solução para este problema reside em outras abordagens, como na coleta de dados mais controlada ou em outras estratégias de pré-processamento, como a já testada inserção de ruído.

Principais Referências

- [1] FINGER, Marcelo et al. Detecting respiratory insufficiency by voice analysis: the SPIRA project. In: Acoustic communication: an interdisciplinary approach. 2021.
- [2] KONG, Qiuqiang et al. PANNs: Large-scale pre-trained audio neural networks for audio pattern recognition. arXiv preprint, arXiv:1912.10211, 2020.
- [3] HUANG, Po-Yao et al. Masked autoencoders that listen. arXiv preprint, arXiv:2207.06405, 2023.
- [4] PEREIRA, Pedro Leyton. noise-reduce-tool: Uma ferramenta para reduzir ruído nos áudios de voz do SPIRA. 2020. Disponível em: <https://github.com/SPIRA-COVID19/noise-reduce-tool>.