

Desenvolvimento de um biomarcador vocal para a estimativa de SpO_2



Natália Hitomi Koza
Ricardo Mikio Morita

Orientador: Prof. Dr. Marcelo Finger

Introdução

A pandemia da doença respiratória Covid-19 iniciou-se em meados de 2019 e continua vigente em 2022, motivando diversos estudos acadêmicos na área. Um exemplo seria através de estudos de sons respiratórios, já que essa é uma doença que muitas vezes causa tosse seca, voz aerada e a chamada *hipóxia silenciosa*, em que há a queda da saturação de oxigênio (SpO_2) no sangue sem haver dispnéia, ou seja, sem que o indivíduo apresente dificuldades de respiração.

Objetivo

Tendo em vista os desafios encontrados durante esta crise e nosso interesse na área, nós propomos estimar os níveis de SpO_2 através da análise de voz com o uso de redes neurais, visando o desenvolvimento de um biomarcador de baixo custo e de fácil acesso e manuseio.

Áudio

Os áudios estão disponíveis sob licença CC BY-SA 4.0 no Github do projeto SPIRA (Casanova et al., 2021) e foram coletados em 2020. São 218 gravações de vozes de pacientes e 16 de ruído ambiente.

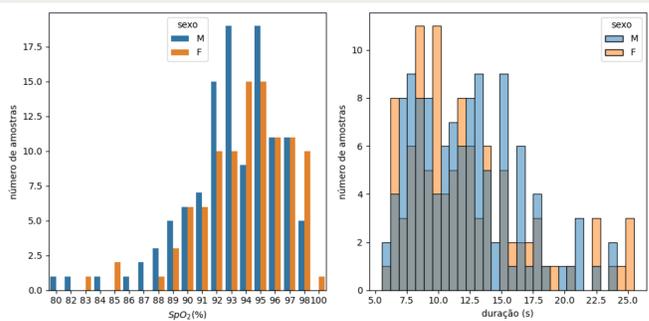


Figure 1: Amostras por valor de SpO_2 e duração

Estes foram gravados através do WhatsApp com extensão OGG e depois convertidos para WAV. Usou-se janelamento de 4 segundos e espaço entre janelas de 1 segundo para amplificar a quantidade de dados, totalizando 1889 amostras. Destas amostras extraímos *features*, no caso *Mel-Frequency Cepstral Coefficients* (MFCC), para serem usadas como entradas da nossa rede.

Modelo

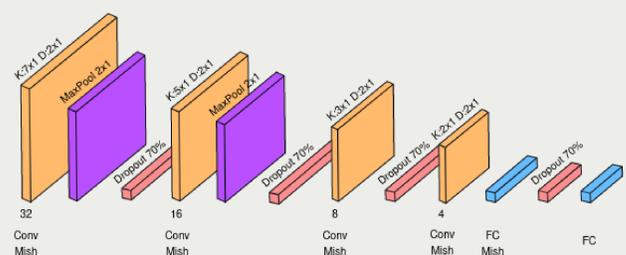


Figure 2: Topologia da Rede Neural

A Rede é composta de quatro camadas convolucionais e duas camadas totalmente conexas. Usamos uma função de ativação linear na saída para tratar nosso problema de regressão.

Experimentos

Foram realizados mais de 90 experimentos, com pelo menos 5 repetições e 500 épocas cada, utilizando sementes aleatórias. Testou-se:

- ▶ Diferentes *learning rates*;
- ▶ *Losses*: MAE vs MSE vs RMSE;
- ▶ *Features*: MFCC vs Mel Espectrograma;
- ▶ Número de camadas e tamanho do *input/output*;
- ▶ Quantidade de *batches*/dados;
- ▶ Parâmetros para extração do MFCC;
- ▶ Adição de ruídos nos áudios;



Figure 3: Divisão dos dados para treino, validação e teste

Resultados

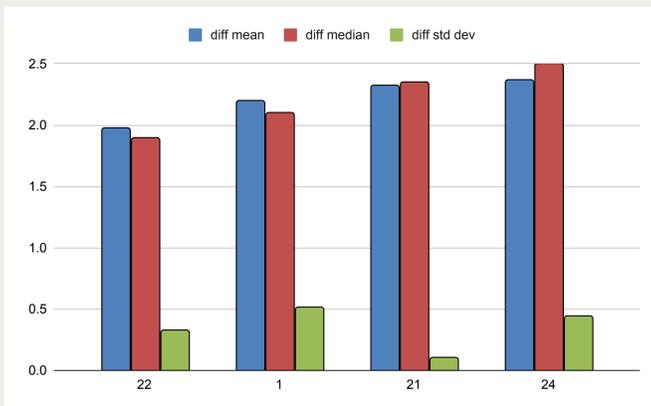


Figure 4: Erros absolutos dos melhores experimentos

Através da média e mediana das diferenças entre SpO_2 prevista e real, obtivemos os resultados da figura 4, que é ao redor de dois pontos em relação ao valor real do paciente.

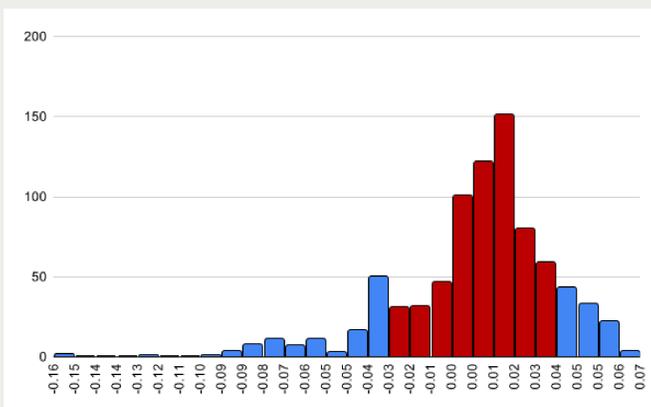


Figure 5: Erro relativo percentual de SpO_2

A figura 5 ilustra o erro relativo entre SpO_2 real e prevista dos melhores experimentos. 80% das estimativas estão no intervalo $[-3.34\%; 4.01\%]$, representadas pelas colunas em vermelho.

Testamos diversas configurações de processamentos de áudio mas o efeito de diversos parâmetros foi inconclusivo. Por isso optamos por utilizar alguns dos parâmetros usados do projeto SPIRA, com exceção da implementação da escala mel: áudio estéreo, *window length* de 400 *frames*, *hop length* de 160 *frames*, janela de 1200 *frames* para o FFT, sem adição de ruídos, 40 MFCCs e 40 *mel bands*.

Considerando $SpO_2 \leq 92\%$ como paciente com insuficiência respiratória (IR) e negativo caso contrário, estimamos a matriz de confusão para a classificação nessas duas classes.

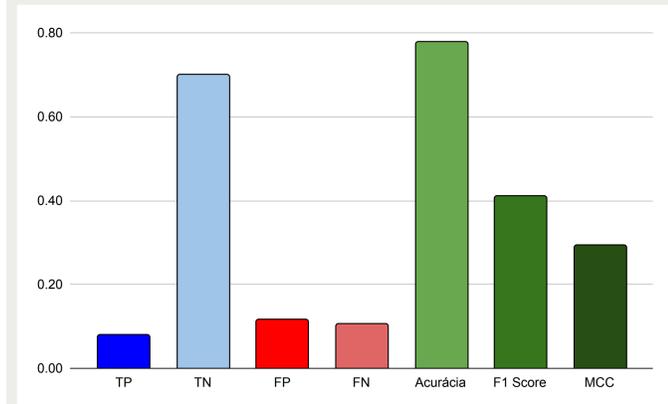


Figure 6: Matriz de Confusão dos melhores experimentos

É possível observar que os modelos tem acurácia em torno de 80%, que são menores do que as obtidas pelo projeto anterior do grupo SPIRA. Nota-se também que o F_1 -Score e MCC foram prejudicados pelos valores de FPs e FNs.

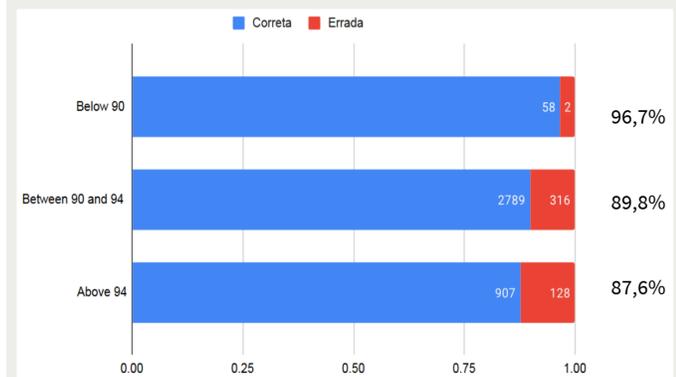


Figure 7: Taxa de acerto médio das melhores estimativas

Na figura 7 temos uma previsão de IR confiável (96,7%) para $\hat{SpO}_2 < 90$ e confiança de 89% para $\hat{SpO}_2 \geq 90$. Isto significa que em certas faixas nossa rede foi comparável com uma CNN que estima IR por classificação de acurácia 91%.

Outra conclusão é de que um profissional da saúde pode confiar mais em nossas estimativas abaixo de 90 daquelas acima de 94, mas seriam necessários mais dados de $SpO_2 < 90$ para podermos verificar a relevância estatística destes fatos.

Conclusão

Apesar de alguns resultados promissores, nosso modelo tem dificuldades na classificação de pacientes com IR. Este é um *trade-off* que esperamos ao trabalhar com um problema de regressão em vez do que com um de classificação binária.

Espera-se que mais dados, principalmente de pacientes com insuficiência, possam melhorar o desempenho da rede neural para ter uma melhor confiabilidade e levar à aplicações práticas no mundo real.

Referências

- ▶ Casanova, Edresson et al. (2021). "Deep learning against covid-19: Respiratory insufficiency detection in brazilian portuguese speech". In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pp. 625–633.