

Biomarcador vocal para a estimativa de Saturação de Oxigênio no sangue (SpO₂) usando Redes Neurais

Natália Hitomi Koza
Ricardo Mikio Morita

Orientador: Prof^o Dr. Marcelo Finger

IME-USP
2022

Introdução

- Motivação
 - COVID-19: Insuficiência respiratória → internação
 - Detecção através de amostras da voz?
 - Grupo SPIRA: 91.66% de acurácia. **Problema de Classificação Binário**
- Qual seria o desempenho de uma Rede Neural treinada para estimativa do nível de SpO2? **Problema de Regressão**



Conceitos

Biomarcadores

- Mede algum parâmetro biológico
- Ex.: termômetro medindo temperatura para verificar febre

SpO2 (%)

- Saturação de oxigênio no sangue
- Oxímetro
- Bom indicador de insuficiência respiratória*

MFCC

- *Mel-Frequency Cepstral Coefficients*
- Obtida a partir da transformada de Fourier do sinal
- Muito utilizado como *feature* em análises de áudio/fala

Metodologia

Coleta dos áudios

Pré-processamento

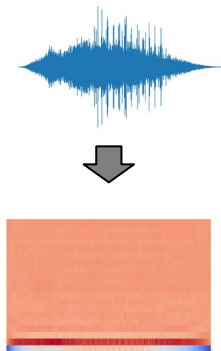
Modelo

Análise

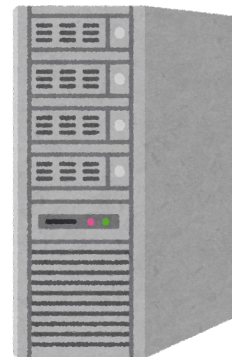
Dados foram coletados por enfermeiros



Conversão do áudio e geração das features



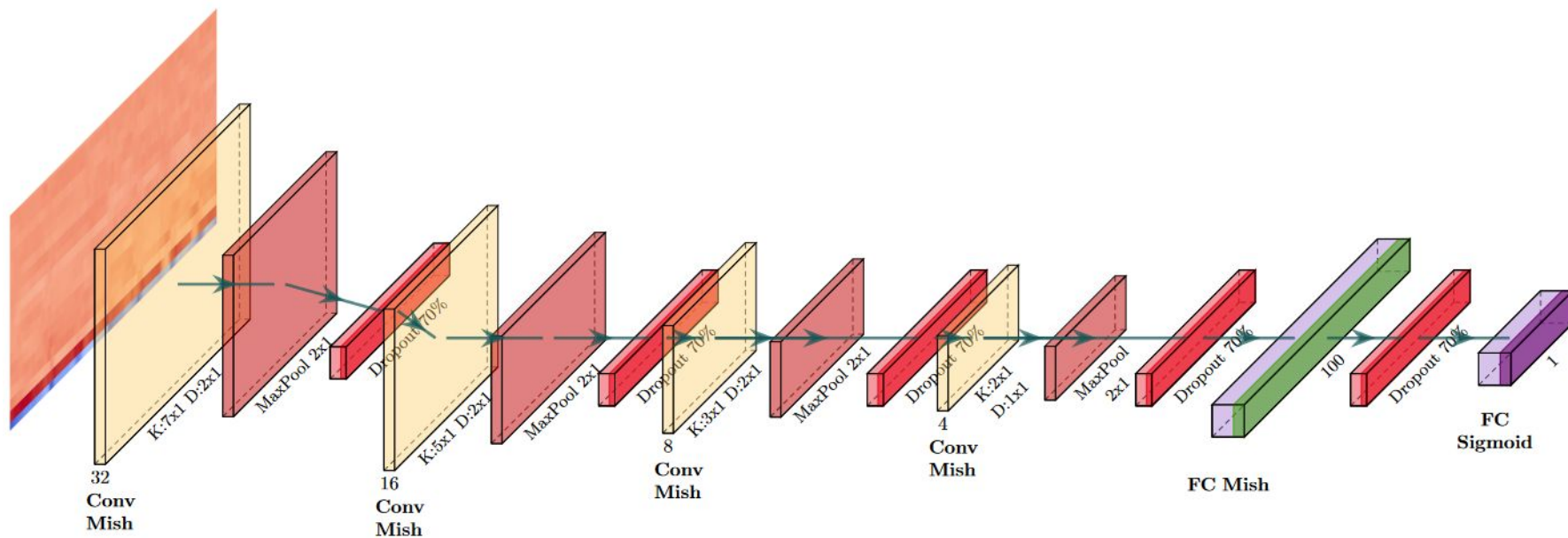
Modelagem e treino da rede



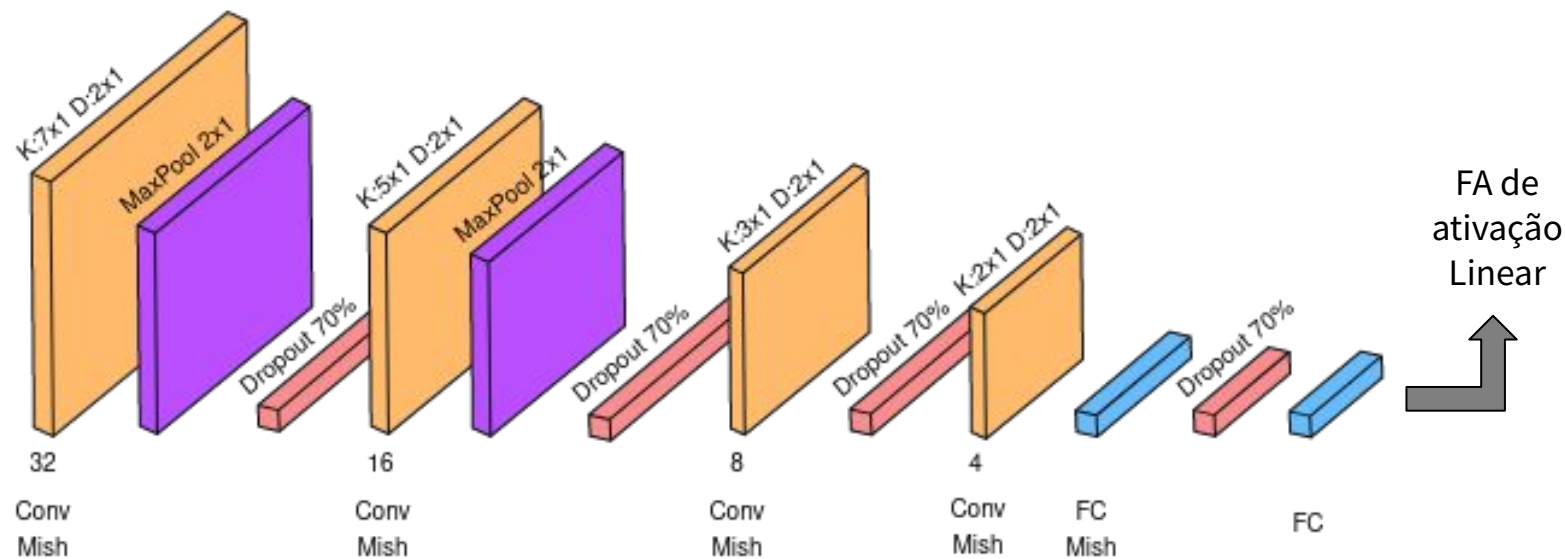
Análise dos experimentos realizados



Modelo



Modelo



Modelo - Função de Loss

- Avalia desempenho da Rede
 - Projeto SPIRA : **Classificação Binária**
 - Binary Cross Entropy Loss (BCE Loss)
 - Nosso Trabalho: **Regressão**
 - Mean Absolute Error Loss (MAE Loss)
 - Mean Squared Error Loss (MSE Loss)
 - Root Mean Squared Error Loss (RMSE Loss)

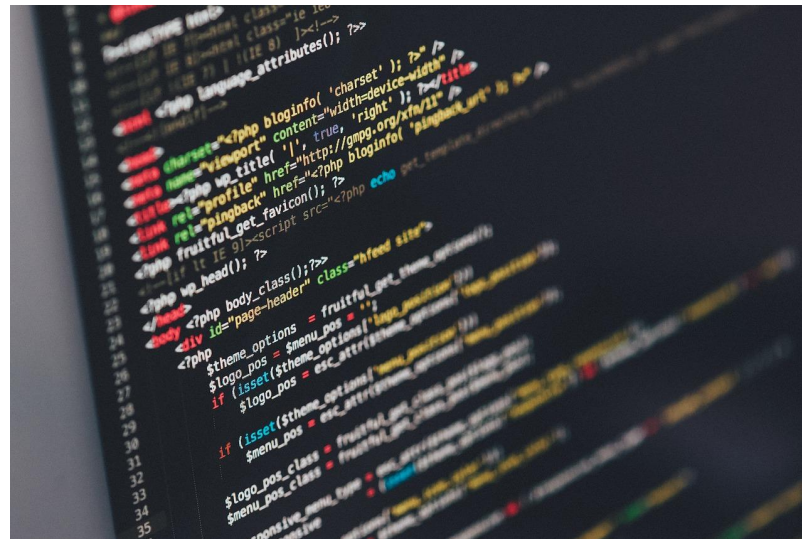
$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

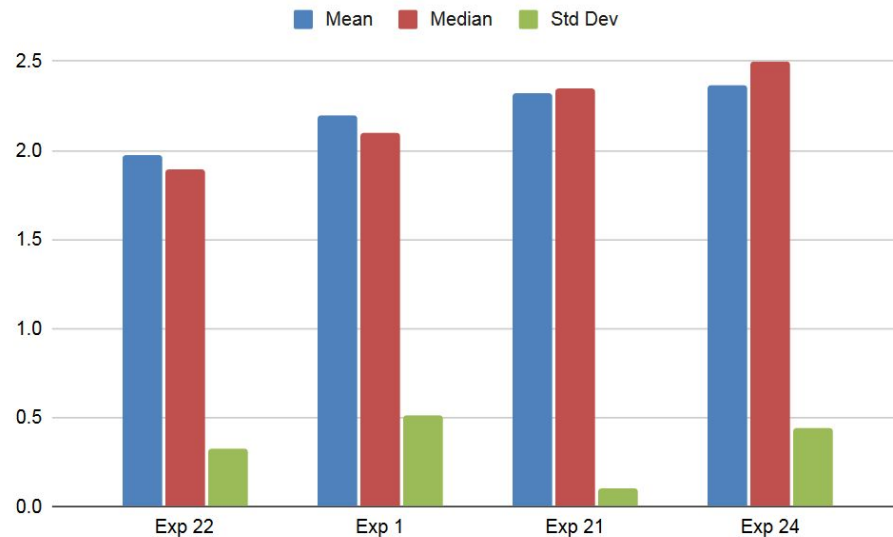
Resultados - O que testamos?

- **Learning Rate:** [10E-05; 10E-02]
- **Loss Function:** MSE vs MAE vs RMSE
- **Scheduler:** Noam vs não usar
- **Áudio:** Mono vs Estéreo
- **Parâmetros** para calcular **MFCC**
- **Ruídos:** Com ou Sem Adição
- **MFCC** vs **Mel espectrograma**



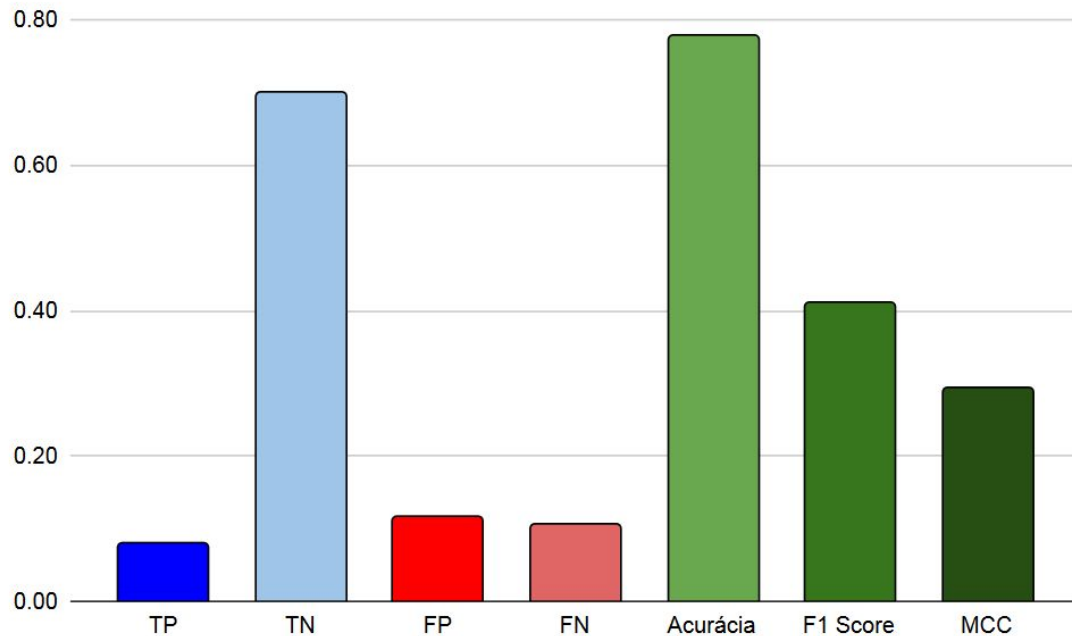
Resultados

- Mais de 90 Experimentos com configurações distintas, cinco repetições cada
- Estimativa próxima de dois pontos do valor real de SpO2
- Mas quais são os resultados clínicos?



Erro absoluto dos melhores experimentos

Resultados



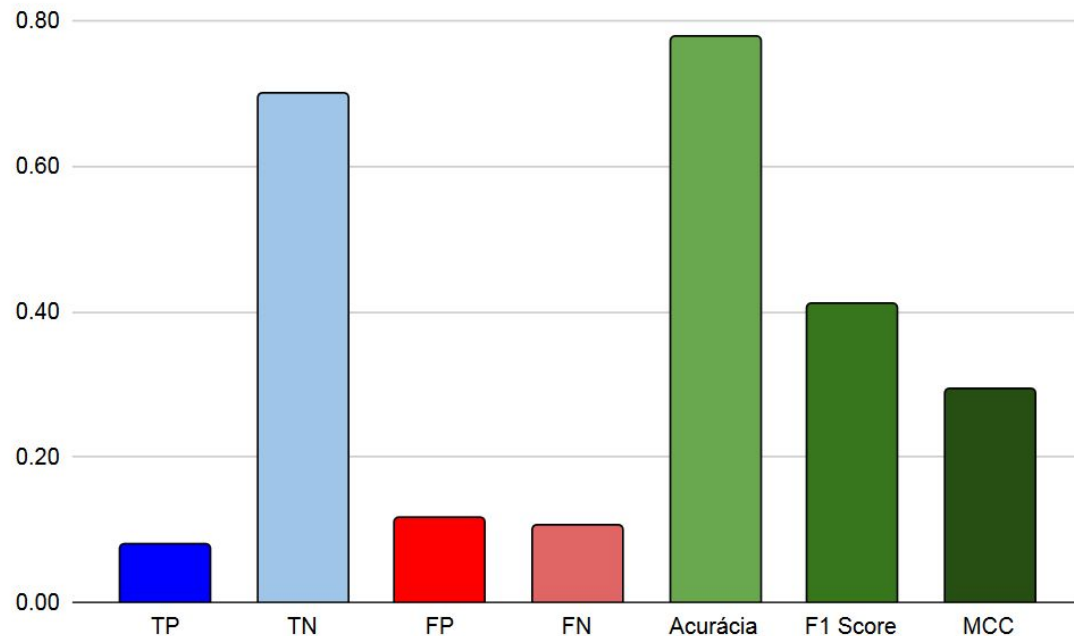
Matriz de Confusão dos melhores experimentos

Acurácia ≈ 0.8

Mede o quão bem as classes foram corretamente classificadas

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Resultados



Matriz de Confusão dos melhores experimentos

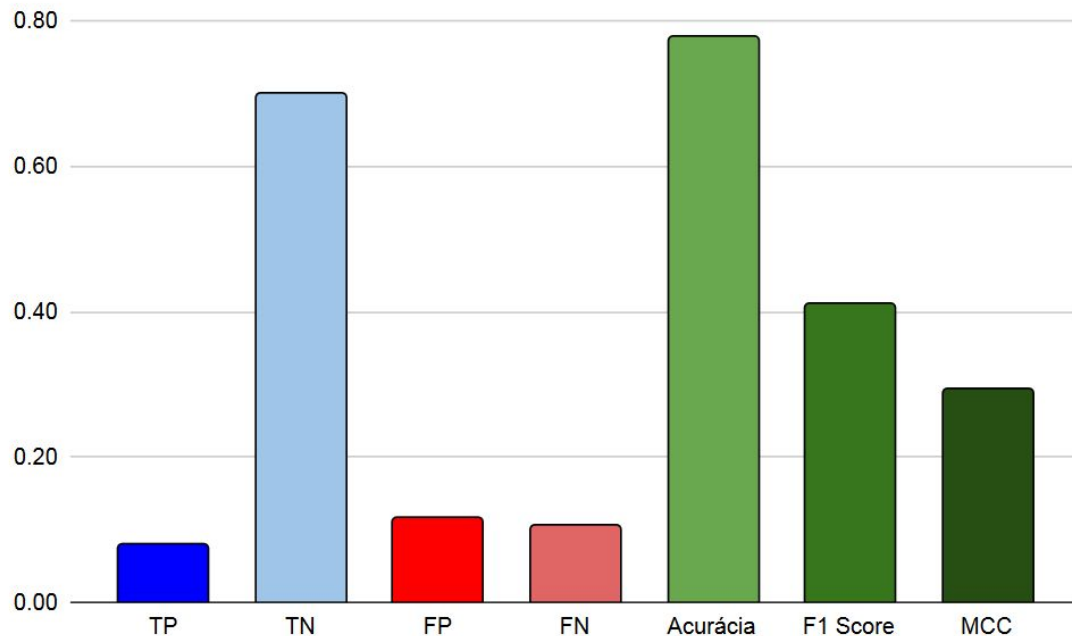
F1 Score ≈ 0.4
média harmônica da
precisão e **recall**

$$\frac{TP}{TP + FP}$$

$$\frac{TP}{TP + FN}$$

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

Resultados



MCC $\approx$ 0.35

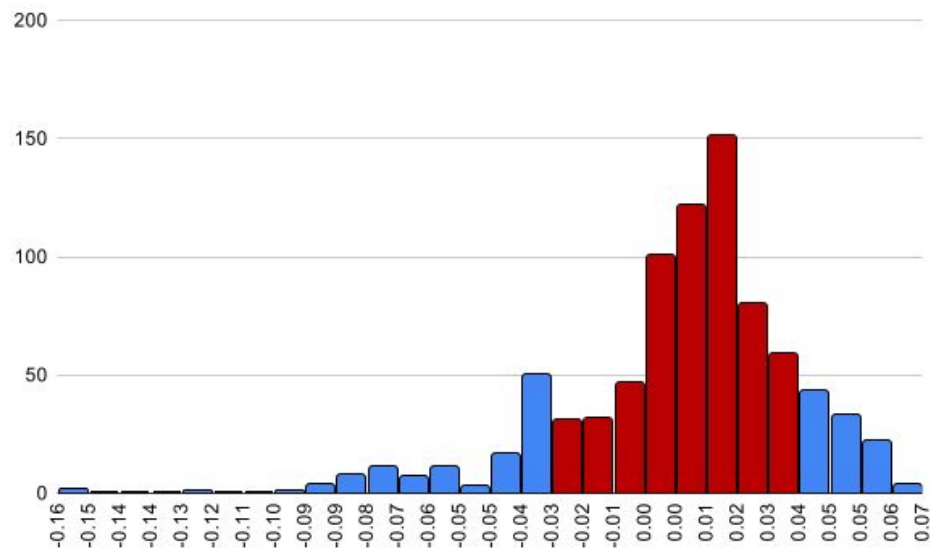
Leva em
consideração
todos os
parâmetros

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

$\left\{ \begin{array}{ll}]0,1] & \text{Predição boa} \\ 0 & \text{Jogar uma moeda} \\ [-1,0[& \text{Predição ruim} \end{array} \right.$

Resultados

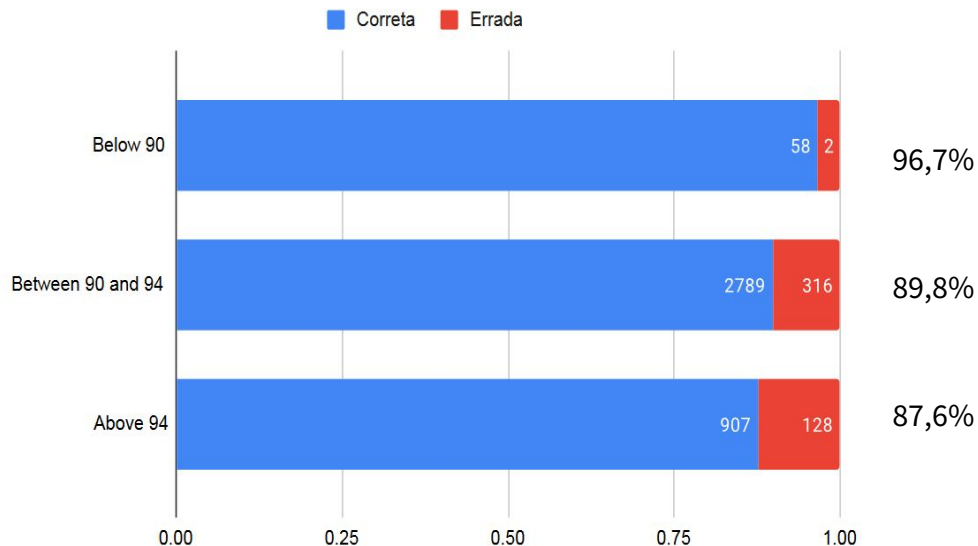
- Erro Relativo Percentual
 - $= (\text{Alvo} - \text{Previsão}) / \text{Alvo}$
- 80% das estimativas estão no intervalo $[-3.34\%; 4.01\%]$, representadas pelas colunas em vermelho.



Erro relativo percentual de SpO2

Estimativa de SpO2 para Insuficiência Respiratória

- $SpO_2 < 92 \Rightarrow IR$
- Maioria dos dados estão no intervalo de dois pontos da SpO2 Real
- Acima de 94 e abaixo de 90 teríamos boa precisão?
- Previsão IR confiável (96,7%) para $\hat{SpO}_2 < 90$. Confiança de 89% para $\hat{SpO}_2 \geq 90$.
- Comparável com previsão por classificação por CNN (acurácia 91%)



Taxa de Acerto médio das melhores estimativas

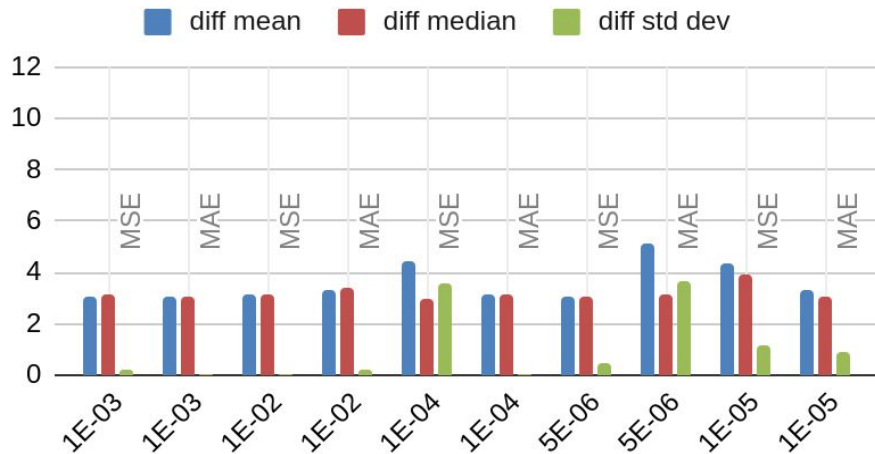
Resultados - Hiperparâmetros

- Learning rate de $10E-03$ e scheduler Noam
- Função de perda MSE
- MFCC
- Áudio estéreo
- Parâmetros para calcular MFCC
 - Janela de 400 frames
 - Hop length de 160 frames
 - Janela de 1200 amostras para a FFT
 - 40 MFCC
 - 40 Mel Bands
- Não adicionar ruídos

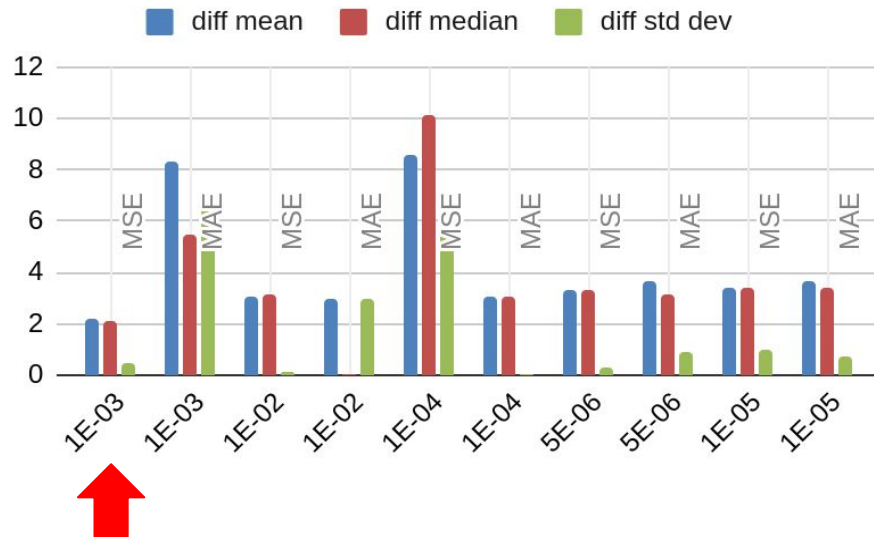
Resultados - Hiperparâmetros

- Learning rate de $10E-03$ e scheduler Noam
- Função de perda MSE
- MFCC

MAE vs MSE sem scheduler



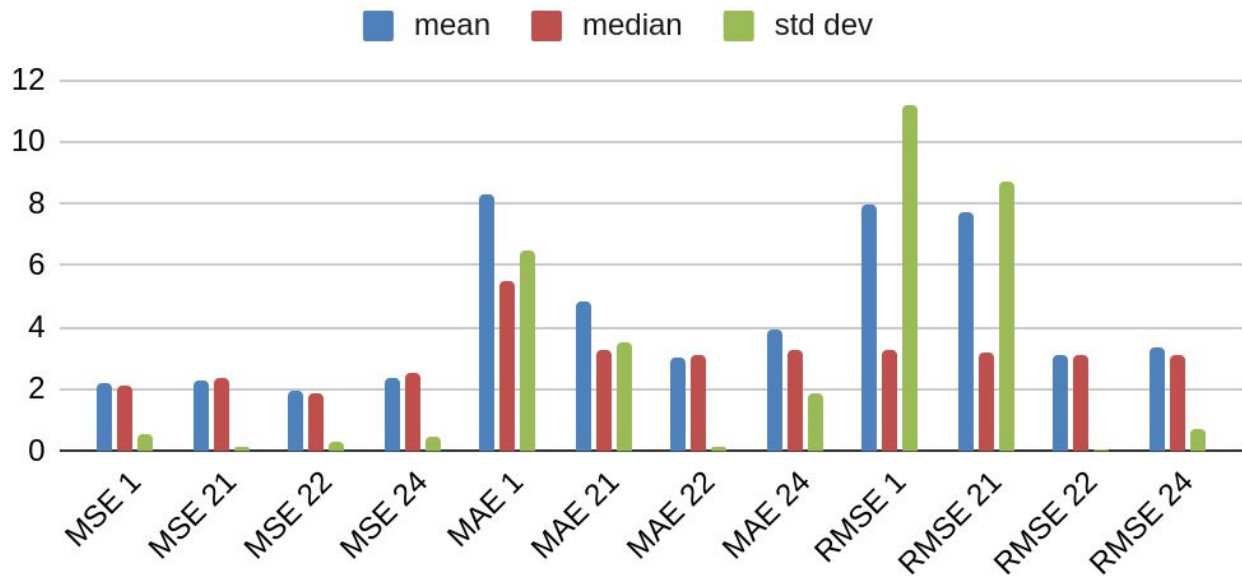
MAE vs MSE com scheduler Noam



Resultados - Hiperparâmetros

- Learning rate de 10E-03 e scheduler Noam
- Função de perda MSE
- MFCC

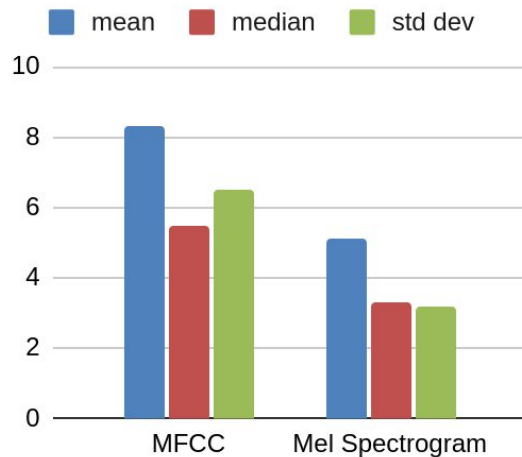
MAE vs MSE vs RMSE



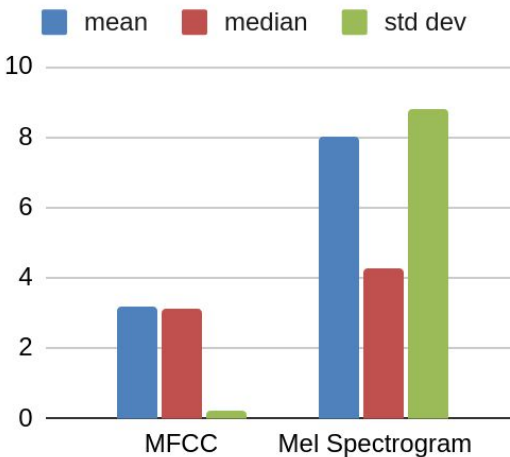
Resultados - Hiperparâmetros

- Learning rate de 10E-03 e scheduler Noam
- Função de perda MSE
- MFCC

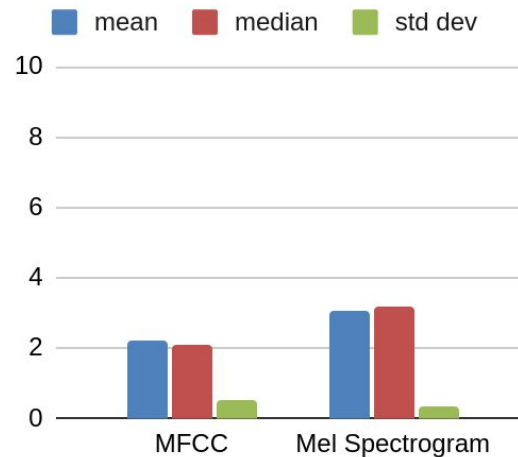
MAE



RMSE

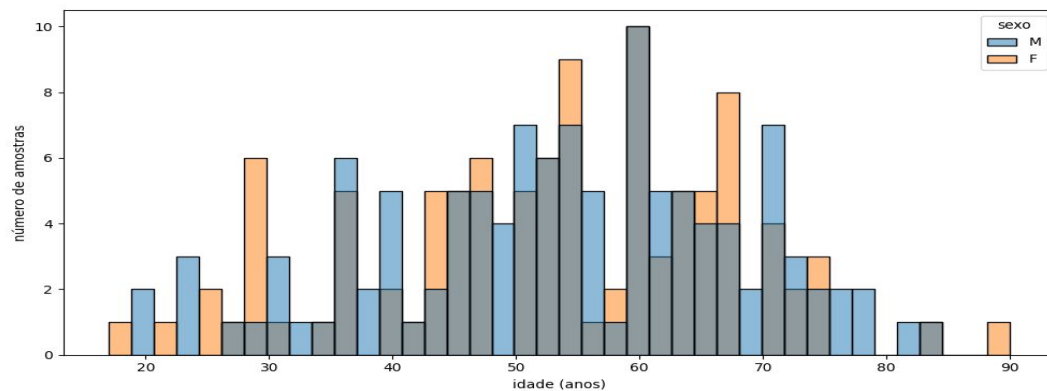
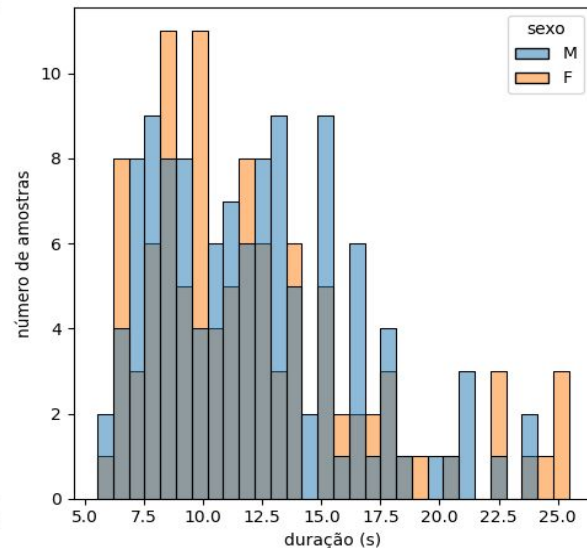
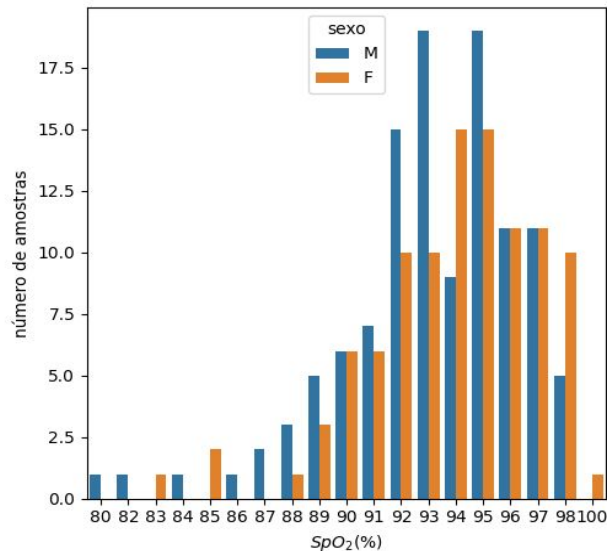


MSE



Dataset

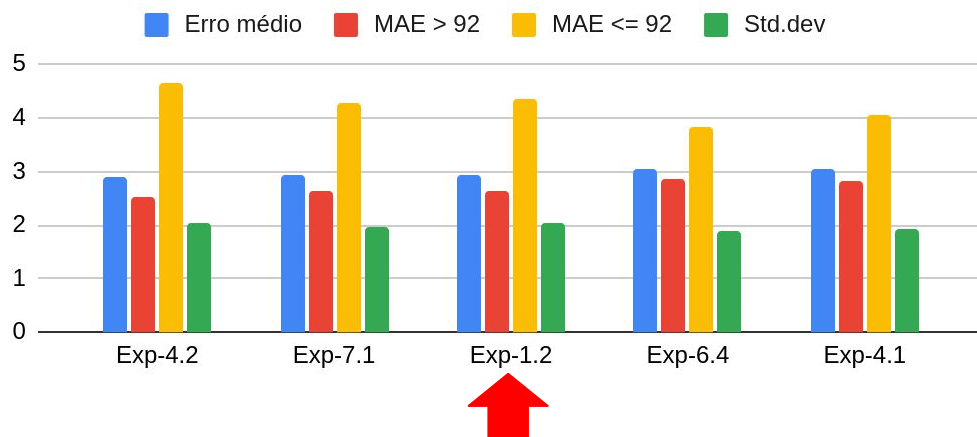
- Desbalanceado
- 16 ruídos hospitalares
- 218 vozes de pacientes
- Janelamento de 4 segundos e salto de 1 segundo



Resultados

- Áudio estéreo
- Parâmetros para calcular MFCC
 - Slaney
 - Janela de 400 frames
 - Hop length de 160 frames
 - Janela de 1200 amostras para a FFT
 - 40 MFCC
 - 40 Mel Bands
- Não adicionar ruídos

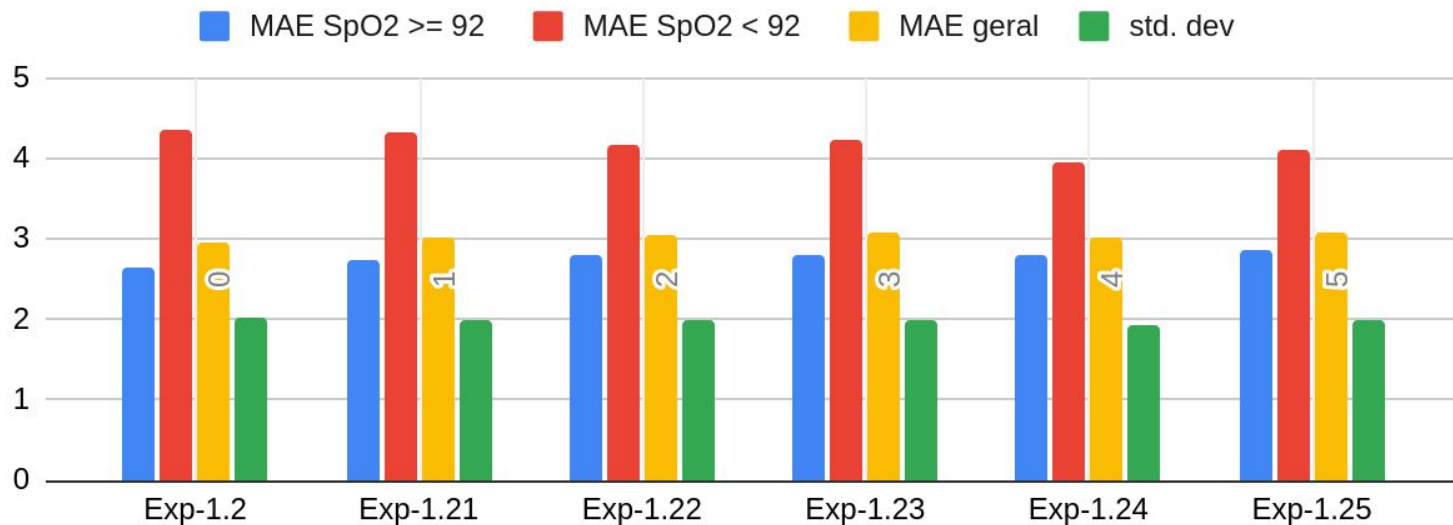
Resultados das 5 melhores configurações testadas



Resultados

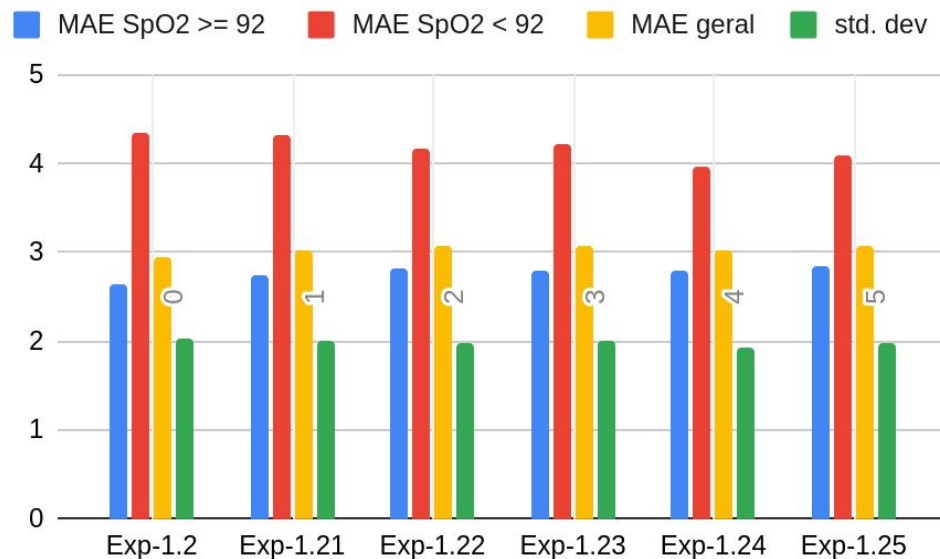
- Vieses
- Ruído vs sem ruído

Inserção de ruído no teste



Resultados

Treino sem inserção de ruído



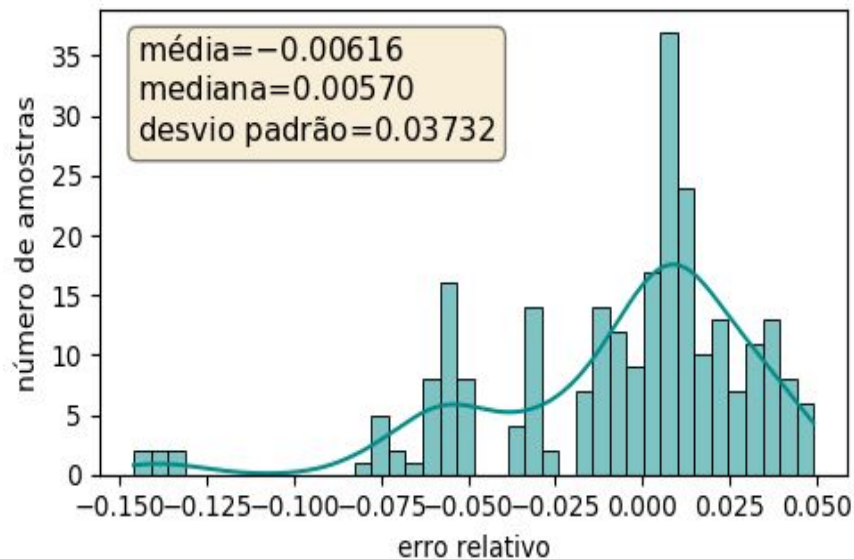
TP = 7%

TN = 48%

FP = 34%

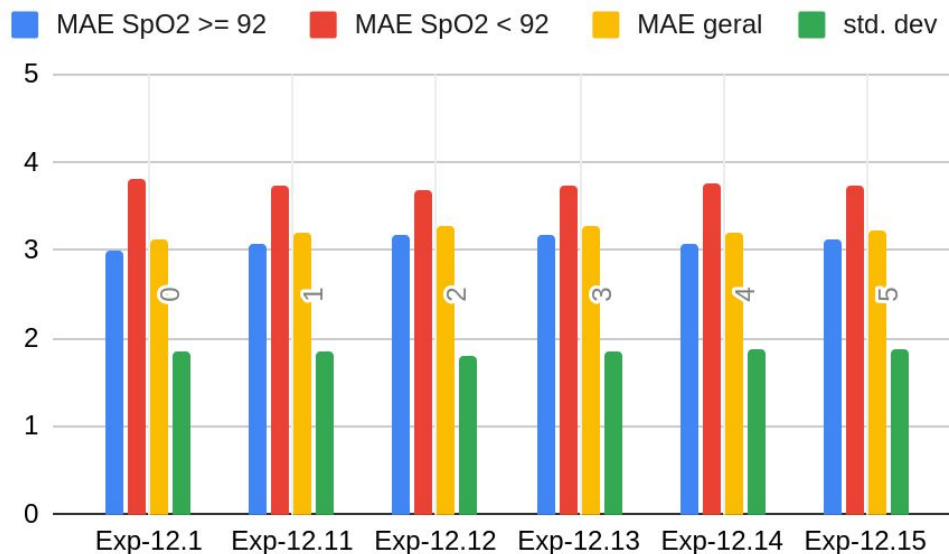
FN = 11%

80% → $|y - \hat{y}|/y \leq 4,27\%$



Resultados

Treino com inserção de 1 amostra de ruído



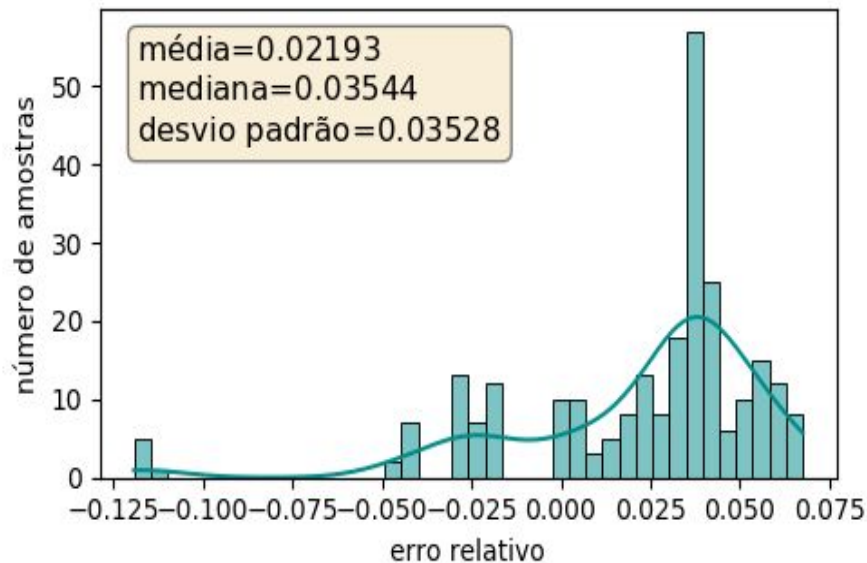
TP = 12%

TN = 25%

FP = 57%

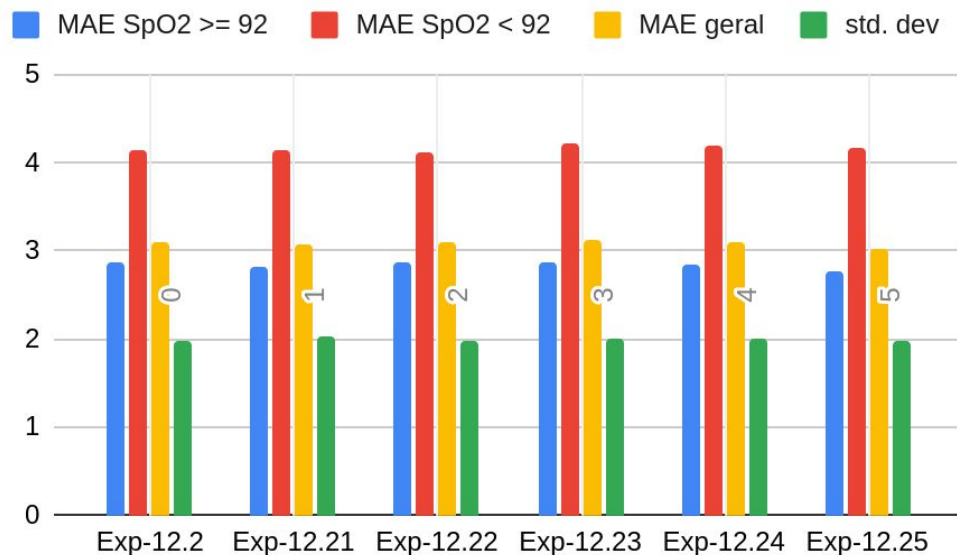
FN = 6%

80% → $|y - \hat{y}|/y \leq 5,35\%$



Resultados

Treino com inserção de 2 amostras de ruído



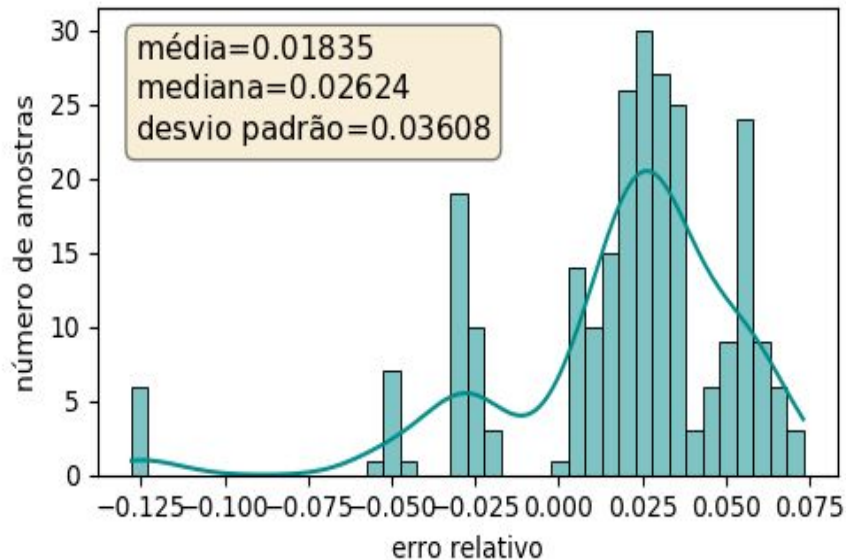
TP = 10%

TN = 40%

FP = 41%

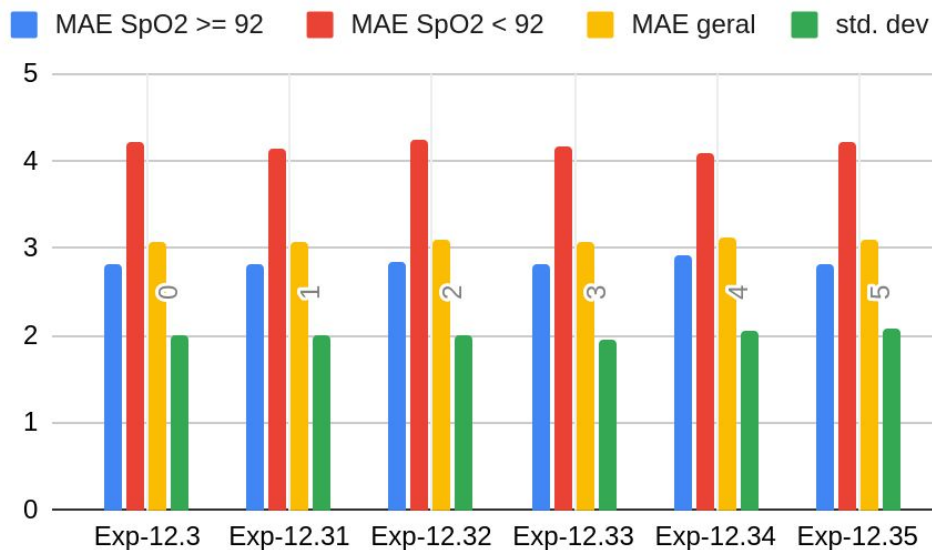
FN = 9%

80% → $|y - \hat{y}|/y \leq 5,24\%$



Resultados

Treino com inserção de 3 amostras de ruído



TP = 11%

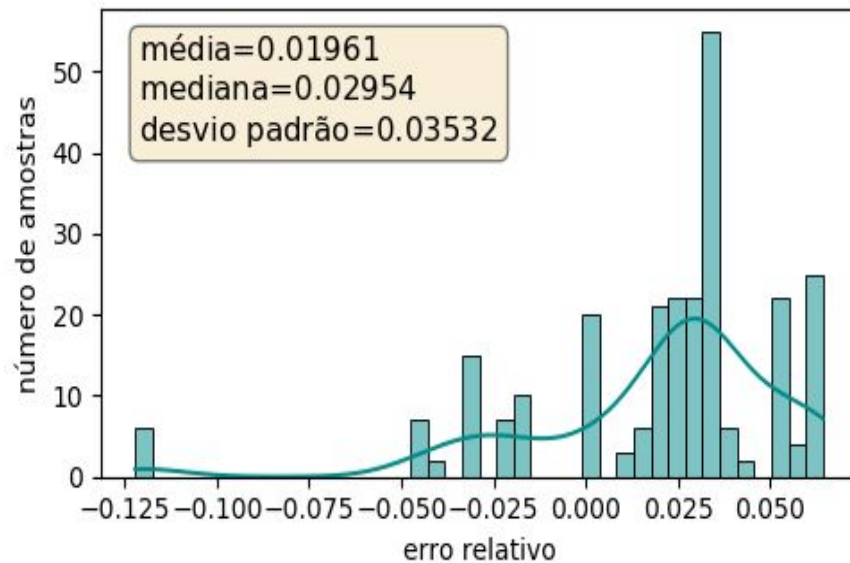
TN = 34%

FP = 47%

FN = 8%

Não adicionar ruídos

80% → $|y - \hat{y}|/y \leq 5,28\%$



Conclusão e Futuro

- Precisamos de mais dados
- Passar por validação clínica
- Criação de aplicação móvel e web



Obrigado!

Referências

- Edresson Casanova et al. “Deep learning against covid-19: respiratory insufficiency detection in brazilian portuguese speech”.
- Ministério da Saúde - SUS. Protocolo de Manejo Clínico da Covid-19 na Atenção Especializada. 2020.
- Haytham Fayek. Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between.