Proposta de Trabalho de formatura supervisionado: **Extração de características associadas à respiração a partir de um sinal de áudio de voz**

Aluno: Pedro Leyton Pereira
Supervisor: Marcelo Queiroz
Orientadora: Nina T. S. Hirata

1 Resumo

Esse projeto de TCC envolve o estudo teórico e a implementação de um sistema computacional que extrai informações relevantes sobre a respiração de uma pessoa através de uma gravação de sua voz. Tais informações seriam úteis tanto pelo seu potencial de interpretação por um ser humano quanto por seu potencial emprego em algoritmos de aprendizado de máquina associados à classificação de quadros clínicos com sintomas associados à função respiratória.

2 Motivação

A pandemia de COVID-19 causou e segue causando uma disrupção massiva no sistema de saúde global, principalmente em locais em que o sistema de saúde não estava preparado para aumentar sua capacidade de atendimento em tempos de crise. Especificamente, a triagem de pacientes está sobrecarregada devido ao afluxo de pessoas aos hospitais buscando esclarecimento sobre seu estado de saúde, muitas vezes sem necessidade.

A consulta por telefone ajuda bastante a reduzir este afluxo aos hospitais, mas não resolve dois problemas:

- Não é possível acompanhar a evolução da saúde do paciente. Um paciente assintomático pode em pouco tempo tornar-se um paciente em estado grave.
- A quantidade de pessoas procurando uma triagem ainda é enorme, sobrecarregando os médicos envolvidos neste processo. Quanto mais automática a triagem dos casos clínicos graves, menor o afluxo de pessoas não infectadas aos hospitais, liberando o foco dos médicos aos casos que demandam atenção.

Com essa motivação, formou-se um projeto composto por professores e alunos da USP para formular uma possível solução para este problema: o projeto SPIRA (Sistema de detecção Precoce de Insuficiência Respiratória por meio de análise de Áudio) [3]. O objetivo do projeto é desenvolver um sistema de processamento de áudio remoto que classifique, através de um áudio de voz falando uma frase pré-combinada, se o falante precisa ser internado ou não com suspeita de COVID-19.

Em particular, este trabalho de TCC está planejado como parte do projeto SPIRA, especificamente o tratamento e extração de informação das gravações de voz, através de técnicas de processamento de sinais, para serem usadas como entrada de algoritmos de aprendizado automático.

3 Especificação

3.1 Objetivos

Como dito previamente, o objetivo final do trabalho é criar uma ferramenta para extração automática de informação relativa à respiração através de sinais de áudio de voz de um paciente. No âmbito do projeto SPIRA, essa ferramenta será integrada a um sistema de triagem de pacientes que deveriam receber internação hospitalar.

Tão importante quanto dizer o que o estudo abrange, é importante também restringir o escopo do trabalho. A coleta dos dados utilizados para teste não faz parte do escopo deste projeto, bem como a criação do sistema que executará a análise com dados de pacientes e alertará o médico em tempo real.

O trabalho inclui:

1. a familiarização com conceitos e técnicas de processamento de sinais, particularmente sinais sonoros.
2. a criação de um sistema que receberá como entrada um um sinal de áudio e gerará como saída tanto representações alternativas, como espectrogramas [1] quanto *features* de áudio relacionados ao timbre e respiração [2, 4, 5].
3. auxiliar na construção e testes da ferramenta de aprendizagem de máquina que determinará a partir de tais *features* se o paciente deve ou não se encaminhar ao hospital [3].

Embora o projeto ainda não esteja em um estado no qual podemos dizer exatamente quais *features* e representações serão utilizadas no sistema final, aqui estão algumas das possibilidades que devem ser exploradas:

- contagem e padrão de respirações por minuto do paciente [2].
- coeficientes de frequências mel-cepstrais, ou *MFCCs* [4], utilizados em problemas de classificação de fala [5].
- técnicas de redução de ruído de fundo [1].

3.2 Metodologia

O processo previsto para alcançar os objetivos é o seguinte:

- Familiarização com conceitos e técnicas de processamento de sinais.
 1. Estudar as representações de sinais de áudio no computador.
 2. Estudar diversas técnicas de processamento de sinais, bem como *features* de áudio.
 3. Estudar bibliotecas computacionais de processamento e análise sonoros, principalmente a *librosa*.
- Criação de um sistema para geração de representações alternativas e *features*.
 1. Criar um *jupyter notebook* para documentar o processo de exploração das técnicas de processamento utilizadas nos sinais de áudio dos pacientes. Atualizar esse notebook durante todo o desenvolvimento do projeto.

2. Criar um *dataset* simples, contendo áudios de vozes diversas, que servirá para análise, exploração e teste inicial dos processamentos, enquanto não há um conjunto de dados reais incluindo pessoas saudáveis e doentes no âmbito do projeto SPIRA [3].
 3. Selecionar as *features* e representações que serão geradas na saída do sistema a partir de experimentos piloto de classificação sonora.
 4. Transformar o *notebook* em um sistema *standalone*, para que ele possa ser acessado em servidores através de uma API.
- Auxiliar na construção e testes da ferramenta de aprendizagem de máquina.
 1. Realizar experimentos piloto de classificação para auxiliar no desenvolvimento da primeira versão da ferramenta SPIRA.
 2. Realizar ajustes e melhorias no processamento de áudio e extração de *features* a partir dos experimentos piloto.
 - Redação da monografia e preparação do poster do TCC.

3.3 Cronograma

Passo \ Mês	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro
Entender como áudios são representados no computador	X						
Estudar diversas técnicas de processamento de sinais e som, entendendo quais manipulações geram saídas desejáveis como features para um aprendizado de máquina.	X	X					
Aprender a utilizar bibliotecas de decomposição sonora, principalmente a librosa, para ser possível automatizar a extração de features.	X	X		X			
Criar um jupyter notebook contendo um rascunho de todos os processamentos que ocorrerão nos áudios dos pacientes. Atualizar esse rascunho conforme o desenvolvimento do projeto.	X	X		X			
Criar um dataset simples, contendo áudios de vozes diversas. Ele servirá para análise, exploração e teste inicial dos processamentos enquanto não há um conjunto de dados reais entre pessoas saudáveis e doentes.	X	X					
Decidir todas as features e representações que serão geradas na saída do sistema.				X	X	X	
Realizar experimentos piloto de classificação para auxiliar no desenvolvimento da primeira versão da ferramenta SPIRA.				X	X	X	
Transformar o notebook em um sistema standalone, para que ele possa ser acessado em servidores através de uma API.					X	X	
Realizar ajustes e melhorias no processamento de áudio e extração de features a partir dos experimentos piloto.						X	
Redação da monografia e preparação do poster do TCC.							X

Referências

- [1] S. A. Broughton and K. Bryan. *Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing*. John Wiley & Sons, 2018. 2

- [2] E. Dafna, T. Rosenwein, A. Tarasiuk, and Y. Zigel. Breathing rate estimation during sleep using audio signal analysis. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5981–5984. IEEE, 2015. 2
- [3] M. Finger et al. Estudo SPIRA: Sistema de detecção precoce de insuficiência respiratória por meio de análise de Áudio. 2020. <https://spira.ime.usp.br/>. 1, 2, 3
- [4] R. Hibare and A. Vibhute. Feature extraction techniques in speech processing: a survey. *International Journal of Computer Applications* 107, 2014. 2
- [5] S. S. Ittichaichareon, C. and T. Yingthawornsuk. Speech recognition using MFCC. *International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012)*, 2012. 2