

CENTRO UNIVERSITÁRIO FEI  
RODRIGO GALUZZI GARCIA PIVA

**UM NOVO BIOMARCADOR PARA AUXILIAR NA INTERPRETAÇÃO DOS  
EXAMES DE ESPIROMETRIA ATRAVÉS DA APLICAÇÃO DA INTELIGÊNCIA  
ARTIFICIAL E DAS FUNÇÕES DAS WAVELETS**

São Bernardo do Campo

2022

RODRIGO GALUZZI GARCIA PIVA

**UM NOVO BIOMARCADOR PARA AUXILIAR NA INTERPRETAÇÃO DOS  
EXAMES DE ESPIROMETRIA ATRAVÉS DA APLICAÇÃO DA INTELIGÊNCIA  
ARTIFICIAL E DAS FUNÇÕES DAS WAVELETS**

Tese de Doutorado apresentada ao Centro Universitário FEI para obtenção do título de Doutor em Engenharia Elétrica, orientado pelo Prof. Dr. Renato Camargo Giacomini.

São Bernardo do Campo

2022

Galuzzi Garcia Piva, Rodrigo.

UM NOVO BIOMARCADOR PARA AUXILIAR NA  
INTERPRETAÇÃO DOS EXAMES DE ESPIROMETRIA ATRAVÉS  
DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL E DAS  
FUNÇÕES DAS WAVELETS / Rodrigo Galuzzi Garcia Piva. São  
Bernardo do Campo, 2022.

102 p. : il.

Tese - Centro Universitário FEI.

Orientador: Prof. Dr. Renato Camargo Giacomini.

1. Espirometria. 2. Wavelets. 3. Inteligencia Artificial. I. Camargo  
Giacomini, Renato, orient. II. Título.

**Aluno:** Rodrigo Galuzzi Garcia Piva

**Matrícula:** 517302-6

**Título do Trabalho:** Um novo biomarcador para auxiliar na interpretação dos exames de espirometria através da aplicação da inteligência artificial e das funções das wavelets.

**Área de Concentração:** Processamento de Sinais e Imagens

**Orientador:** Prof. Dr. Renato Camargo Giacomini

**Data da realização da defesa:** 29/04/2022

**ORIGINAL ASSINADA**

**Avaliação da Banca Examinadora**

O candidato apresentou seu trabalho de doutoramento, incluindo cada um dos tópicos abordados no texto escrito. Foi então arguido por cada membro da banca, recebendo sugestões, críticas e dúvidas. Respondeu sobre os pontos levantados e comprometeu-se a fazer as correções que se fizeram necessárias. Foi considerado aprovado por unanimidade e deverá apresentar a versão final no prazo máximo regulamentar.

São Bernardo do Campo,     /     /     .

**MEMBROS DA BANCA EXAMINADORA**

|                                    |             |
|------------------------------------|-------------|
| Prof. Dr. Renato Camargo Giacomini | Ass.: _____ |
| Prof. Dr. Renato Aparecido Aguiar  | Ass.: _____ |
| Prof. Dr. Fabio Delatore           | Ass.: _____ |
| Prof. Dr. Regis Reis               | Ass.: _____ |
| Prof. Dr. José Antonio Fonseca     | Ass.: _____ |

A Banca Examinadora acima-assinada atribuiu ao aluno o seguinte:

APROVADO ☒

REPROVADO ☐

**VERSÃO FINAL DA TESE**

**ENDOSSO DO ORIENTADOR APÓS A INCLUSÃO DAS  
RECOMENDAÇÕES DA BANCA EXAMINADORA**

Aprovação do Coordenador do Programa de Pós-graduação

\_\_\_\_\_  
Prof. Dr. Carlos Eduardo Thomaz

Aos que me apoiaram nessa difícil jornada.

## **AGRADECIMENTOS**

Ao Prof. Dr. Renato Camargo Giacomini, por aceitar o desafio de me orientar nessa longa jornada.

Ao Prof. Dr. Aldo Artur Belardi, por todo o apoio, dedicação e energia para a realização de todas as etapas necessárias desse projeto. Pela amizade, conselhos, aulas e ensinamentos em todos esses anos.

Ao amigo de longa jornada, Sr. Dorival Sortino, pela inicial oportunidade concedida a mim, sem ela eu não teria produzido esse estudo.

A Profa. Dra. Arianne Soares do Nascimento, pelo auxílio e aconselhamentos na produção deste trabalho.

Aos amigos Rodrigo Prior Bechelli, Renato Barelli, Douglas Rizzo e André Mendes, por compartilharem as alegrias e desafios da ciência.

Ao amigo Wilson Renato Francisco, pela disponibilidade e auxílio no decorrer desse trabalho.

Ao amigo Daniel Alberto Castrucci, pelo apoio, orientação, em todos os momentos que precisei.

À equipe da pós-graduação: Adriana, Dreycer, Luana, Márcia, Carina, Raquel e Ricardo sempre auxiliando da melhor maneira possível.

Ao pessoal que cuida do laboratório de elétrica da FEI, principalmente o Sr. Acácio que sempre me auxiliou.

À equipe do CGI da FEI, pela eficiência, atenção e disposição.

Aos funcionários da Biblioteca da FEI, pela paciência e vontade de ajudar, sempre.

A Elâine de Araujo Piva, por sempre estar ao meu lado e pelo reconhecimento, pelo amor, incentivo, força e apoio incondicional.

Ao meu pai Sr. Alcides Aparecido Piva(in memoriam), que sempre esteve ao meu lado em todas as iniciativas.

E por fim, mas não menos importante, dedico este trabalho à minha querida mãe Sra. Tania Galuzzi Garcia que sempre acredita nos empreendimentos da minha vida.

E a todos que não foram citados, mas que auxiliaram direta ou indiretamente para a realização deste trabalho.

Assim sendo, **O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.**

“A caverna que tememos entrar guarda o tesouro  
que tentamos buscar”

Joseph Campbell

“Muitos homens devem a grandeza da sua vida  
aos obstáculos que tiveram que vencer. ”

Charles Spurgeon

“A humildade é a chave de nossa libertação.”

Chico Xavier

## RESUMO

Nas últimas décadas, o volume de informações armazenadas digitalmente cresce em taxas expressivas. Soluções tradicionais relacionadas ao tratamento de séries temporais e inteligência artificial são amplamente aplicadas no contexto na extração de informação deste tipo de dado. Um dos sistemas mais robustos e de baixo custo para a aquisição de sinais de função pulmonar, são os espirometros. O processamento dos sinais de espirometria, através de inteligência artificial, utilizando um conjunto de *wavelets* para uma análise espacial e temporal da capacidade pulmonar, apresenta uma alternativa para classificação e ordenação de um conjunto de exames. Neste trabalho é proposto uma metodologia para classificar e representar uma relação entre exames de espirometria juntamente com conjuntos de exames posteriormente realizados para diferentes grupos de voluntários. Também é interessante mencionar que um novo marcador em exames de espirometria é de grande importância, visto que na nossa pesquisa bibliográfica poucos trabalhos foram encontrados utilizando um modelo matemático mais robusto na busca por novos marcadores e, procurando ter uma maior eficiência, utilizando-se modelos matemáticos e séries temporais, além da aplicação desses resultados em uma estrutura de armazenamento de dados de forma que possam ser estudadas no futuro, com a análise das *wavelets* que também são utilizadas por diversos pesquisadores como uma alternativa da série de fourier clássica, auxilia neste caso o médico pneumologista a ter um diagnóstico mais confiável e preciso.

Assim, para a inteligência artificial ter sucesso, precisamos da inteligência e de um artefato, e o computador tem sido o artefato preferido.

Este trabalho aplica métodos matemáticos em sistemas multivariáveis, e através da análise estatística em conjunto com a inteligência artificial, produz artefatos de forma a caracterizar exames de prova de função pulmonar em três níveis, classificando os pacientes em normal, obstrutivo ou restritivo. O resultado final é um sistema computacional que demonstra graficamente os valores encontrados de um banco de dados, comparando-os com o resultado do exame obtido a ser analisado no momento.

Através desse conjunto de ferramentas, utilizou-se um banco de dados, gentilmente cedido pela Dra. Silvia Rodrigues, proporcionando realizar o treinamento do modelo e, utilizando ferramentas de I.A., extraíndo a covariância dos coeficientes produzidos pelas *wavelets*, possibilitando a classificação em três grupos distintos (clustering), e através da regressão de Ridge, criar parâmetros como o índice de precisão de uma nova amostra analisada.

**Keywords:** Espirometria, Wavelets, Inteligencia Artificial

## ABSTRACT

In the last decades, the volume of information stored digitally grows at expressive rates. Traditional solutions related to the treatment of time series and artificial intelligence are widely applied in the context of extracting information from this type of data. One of the most robust and low-cost systems for acquiring signs of pulmonary function are spirometers. The processing of spirometry signals, through artificial intelligence, using a set of *wavelets* for a spatial and temporal analysis of lung capacity, presents an alternative for classifying and ordering a set of exams. In this work, a methodology is proposed to classify and represent a relationship between spirometry tests together with sets of tests performed later for different groups of volunteers. It is also interesting to mention that a new marker in spirometry tests is of great importance, since in our bibliographic research few works were found using a more robust mathematical model in the search for new markers and, seeking to have greater efficiency, using models mathematical and time series, in addition to the application of these results in a data storage structure so that they can be studied in the future, with the analysis of wavelets that are also used by several researchers as an alternative to the classic Fourier series, helps in this case the pulmonologist to have a more reliable and accurate diagnosis.

So for artificial intelligence to succeed, we need intelligence and an artifact, and the computer has been the artifact of choice.

This work applies mathematical methods in multivariable systems, and through statistical analysis in conjunction with artificial intelligence, produces artifacts in order to characterize pulmonary function tests at three levels, classifying patients as normal, obstructive or restrictive. The final result is a computer system that graphically demonstrates the values found from a database, comparing them with the exam result obtained to be analyzed at the moment.

Through this set of tools, a database was used, kindly provided by Dr. Silvia Rodrigues, providing the training of the model and, using A.I. tools, extracting the covariance of the coefficients produced by the wavelets, allowing the classification into three distinct groups (clustering), and through Ridge regression, creating parameters such as the precision index of a new analyzed sample.

Keywords: Spirometry, Wavelets, Artificial Intelligence

## LISTA DE SIMBOLOS

$\psi_{a,b}(x)$  “Wavelet” mãe

a Parâmetro escala para “Wavelets” contínuas

b Parâmetro deslocamento para “Wavelets” contínuas

$\psi_{j,k}(x)$  “Wavelet” mãe substituindo  $a = 2^{-j}$  e  $b = k2^{-j}$

j Níveis de subespaço para “Wavelets” discretas

k Deslocamento em cada nível para “Wavelets” discretas

$\psi^{(H)}(x)$  “Wavelet” de Haar em relação ao eixo x

$\phi^{(H)}(x)$  Função escala para a “Wavelet” de Haar eixo x

$\psi^{(H)}(y)$  “Wavelet” de Haar em relação ao eixo y

$\phi^{(H)}(y)$  Função escala para a “Wavelet” de Haar eixo y

$\psi_{j,k}^{(H)}(x)$  “Wavelet” de Haar discreta em relação ao eixo x

$\psi_{j,k}^{(H)}(y)$  “Wavelet” de Haar discreta em relação ao eixo y

$\psi_{j,k}^{(H)}(x,y)$  “Wavelet” de Haar resultante da combinação dos eixos x e y

$\omega$  “Wavelet” de Meyer

$\phi$  “Wavelet” de Battle-Lemarié

$\psi(x)$  “Wavelet” Chapéu Mexicano e Shannon

Simbologia e nomenclatura para pneumologia:

BTPS – Condições corporais, temperatura corporal, pressão ambiente, saturada com vapor d’água.

CV – Capacidade vital

CPT – Capacidade pulmonar total

CVI – Capacidade vital inspiratória

CVF – Capacidade vital forçada

CVFI – Capacidade vital forçada inspiratória

PFE – Pico de fluxo expiratório.

VVM – Ventilação voluntária máxima.

TEF – Tempo de expiração forçada.

## LISTA DE ILUSTRAÇÕES

|           |                                                                                                                  |    |
|-----------|------------------------------------------------------------------------------------------------------------------|----|
| Figura 1  | – Wavelet de HAAR . . . . .                                                                                      | 29 |
| Figura 2  | – Wavelet de Meyer . . . . .                                                                                     | 30 |
| Figura 3  | – Wavelet de Battle-Lemarié . . . . .                                                                            | 31 |
| Figura 4  | – Wavelet Chapéu Mexicano . . . . .                                                                              | 31 |
| Figura 5  | – Wavelet de Shannon . . . . .                                                                                   | 31 |
| Figura 6  | – Decomposição de 4 níveis de Wavelet. . . . .                                                                   | 32 |
| Figura 7  | – Estimativa dos coeficientes da regressão de <i>ridge</i> . . . . .                                             | 33 |
| Figura 8  | – Modelo de Testes de Turing . . . . .                                                                           | 35 |
| Figura 9  | – Busca Subida de Encosta . . . . .                                                                              | 37 |
| Figura 10 | – Gráfico de Dispersão . . . . .                                                                                 | 40 |
| Figura 11 | – Gráfico de Histograma . . . . .                                                                                | 41 |
| Figura 12 | – Tela de Cadastro . . . . .                                                                                     | 44 |
| Figura 13 | – Software do Espirometro Durante o Período do Doutorado, Tela de exame<br>de Capacidade Vital Forçada . . . . . | 44 |
| Figura 14 | – Clustering dos dados femininos e masculinos . . . . .                                                          | 49 |
| Figura 15 | – Modelo de análise interpretativa atual . . . . .                                                               | 49 |
| Figura 16 | – Código da Aplicação do Q-Learn . . . . .                                                                       | 50 |
| Figura 17 | – Equação do Q-Learn . . . . .                                                                                   | 50 |
| Figura 18 | – Matriz de Estados Obtida . . . . .                                                                             | 51 |
| Figura 19 | – Fluxograma de Construção do Software . . . . .                                                                 | 54 |
| Figura 20 | – Etapa de visualização do exame analisado . . . . .                                                             | 55 |
| Figura 21 | – Tela do Software do Relatório do Equipamento MultiSpiro . . . . .                                              | 56 |
| Figura 22 | – Tela do Software do Relatório do Equipamento Vyntus . . . . .                                                  | 57 |
| Figura 23 | – Tela do Software do Relatório do Equipamento Koko . . . . .                                                    | 58 |
| Figura 24 | – Imagem do Equipamento Desenvolvido . . . . .                                                                   | 59 |
| Figura 25 | – Tela do Software do Relatório do Equipamento Desenvolvido . . . . .                                            | 60 |
| Figura 26 | – Grafico com valores de resposta da familia de Wavelets . . . . .                                               | 61 |
| Figura 27 | – Grafico com valores de resposta da familia de Wavelets detalhados . . . . .                                    | 62 |
| Figura 28 | – Demonstração da linha de corte com o Teste de Tukey . . . . .                                                  | 63 |

|                                                                                                                       |    |
|-----------------------------------------------------------------------------------------------------------------------|----|
| Figura 29 – Etapa de análise do exame realizado . . . . .                                                             | 64 |
| Figura 30 – Interpretações fornecidas para o sistema . . . . .                                                        | 64 |
| Figura 31 – Amostra do Sinal a Ser Processado . . . . .                                                               | 65 |
| Figura 32 – Níveis de Decomposição Apresentados Durante o Estudo . . . . .                                            | 66 |
| Figura 33 – Apresentação do Software com a Leitura do Banco de Dados . . . . .                                        | 67 |
| Figura 34 – Apresentação do Software com a Leitura do Banco de Dados e a Aplicação<br>da Regressão de Ridge . . . . . | 68 |
| Figura 35 – Apresentação do Software com a Leitura de um novo exame considerado<br>NORMAL . . . . .                   | 68 |
| Figura 36 – Apresentação do Software com a Leitura de um novo exame considerado<br>RESTRITIVO . . . . .               | 69 |
| Figura 37 – Apresentação do Software com a Leitura de um novo exame considerado<br>OBSTRUTIVO . . . . .               | 69 |

## SUMÁRIO

|              |                                                                 |    |
|--------------|-----------------------------------------------------------------|----|
| <b>1</b>     | <b>INTRODUÇÃO</b>                                               | 16 |
| 1.1          | OBJETIVOS DA PESQUISA                                           | 17 |
| 1.2          | JUSTIFICATIVA DO TRABALHO                                       | 18 |
| 1.3          | ESTRUTURA DO TRABALHO                                           | 19 |
| <b>2</b>     | <b>ESTADO DA ARTE</b>                                           | 20 |
| <b>3</b>     | <b>BASE TEÓRICA</b>                                             | 27 |
| 3.1          | WAVELETS                                                        | 27 |
| 3.2          | CLASSIFICADOR DE RIDGE                                          | 33 |
| 3.3          | INTELIGENCIA ARTIFICIAL E ENGENHARIA DE SOFTWARE                | 35 |
| 3.4          | ANÁLISE ESTATÍSTICA                                             | 39 |
| 3.5          | EQUIPAMENTO DESENVOLVIDO DURANTE O ESTUDO                       | 43 |
| <b>4</b>     | <b>METODOLOGIA PROPOSTA</b>                                     | 46 |
| 4.1          | DADOS OBTIDOS                                                   | 46 |
| 4.2          | CRIAÇÃO DO BANCO DE DADOS                                       | 46 |
| 4.3          | INTERPRETAÇÕES                                                  | 47 |
| 4.4          | SELEÇÃO DA FAMÍLIA DE WAVELETS                                  | 48 |
| 4.5          | EXPERIMENTO DO Q-LEARN NA ATRIBUIÇÃO DO DIAGNÓSTICO             | 49 |
| 4.6          | TESTE DO CLASSIFICADOR                                          | 51 |
| 4.7          | PROCESSAMENTO DE DADOS                                          | 51 |
| 4.8          | PROCEDIMENTO EXPERIMENTAL                                       | 54 |
| <b>5</b>     | <b>RESULTADOS</b>                                               | 56 |
| 5.1          | EQUIPAMENTOS EXISTENTES E SUAS INTERPRETAÇÕES                   | 56 |
| <b>5.1.1</b> | <b>EQUIPAMENTO MULTISPIRO - CREATIVE BIOMEDICS</b>              | 56 |
| <b>5.1.2</b> | <b>EQUIPAMENTO VYNTUS - SENTRY SUITE</b>                        | 57 |
| <b>5.1.3</b> | <b>EQUIPAMENTO KOKO - NSPIRE</b>                                | 58 |
| <b>5.1.4</b> | <b>RESULTADOS DO EQUIPAMENTO DESENVOLVIDO</b>                   | 59 |
| 5.2          | SELEÇÃO DA FAMÍLIA DE WAVELETS                                  | 61 |
| 5.3          | AVALIAÇÃO DAS CURVAS DE ESPIROMETRIA                            | 63 |
| 5.4          | CLASSIFICAÇÃO INDIVIDUAL DOS DADOS                              | 65 |
| 5.5          | APLICAÇÃO DOS DADOS OBTIDOS APÓS O TÉRMINO DA CRIAÇÃO DO MODELO | 66 |

|          |                                                                                                                    |            |
|----------|--------------------------------------------------------------------------------------------------------------------|------------|
| 5.6      | DISCUSSÃO . . . . .                                                                                                | 72         |
| <b>6</b> | <b>CONCLUSÃO . . . . .</b>                                                                                         | <b>75</b>  |
| 6.1      | TRABALHOS FUTUROS . . . . .                                                                                        | 76         |
|          | <b>REFERÊNCIAS . . . . .</b>                                                                                       | <b>77</b>  |
|          | <b>APÊNDICE A – TERMO DE CONFIDENCIALIDADE PARA CRIAÇÃO<br/>DO BANCO DE DADOS . . . . .</b>                        | <b>85</b>  |
|          | <b>APÊNDICE B – MODELO DE EXAME UTILIZADO PARA CRIAÇÃO<br/>DO BANCO DE DADOS . . . . .</b>                         | <b>87</b>  |
|          | <b>APÊNDICE C – CURSOS CONCLUÍDOS DURANTE O ESTUDO . . . .</b>                                                     | <b>89</b>  |
|          | <b>APÊNDICE D – AULAS MINISTRADAS DURANTE O DOUTORAMENTO</b>                                                       | <b>91</b>  |
|          | <b>APÊNDICE E – PRODUÇÃO LITERÁRIA E PRODUÇÃO CORRELA-<br/>CIONADAS . . . . .</b>                                  | <b>93</b>  |
|          | <b>APÊNDICE F – SUBMISSÃO DE ARTIGO PUBLICADO SOBRE PO-<br/>LISSENOGRAFIA . . . . .</b>                            | <b>96</b>  |
|          | <b>APÊNDICE G – CARTA DE ACEITE DE ARTIGO PUBLICADO SOBRE<br/>O TEMA . . . . .</b>                                 | <b>99</b>  |
|          | <b>APÊNDICE H – CARTA DE ACEITE DE ARTIGO PUBLICADO SOBRE<br/>APLICAÇÃO DE WAVELETS EM EXAMES DE EEG . . . . .</b> | <b>101</b> |

## 1 INTRODUÇÃO

Nas últimas décadas, o volume de informações armazenado digitalmente cresce em taxas nunca antes vistas e boa parte destas informações são geradas pelo armazenamento amplo de imagens, vídeos e documentos digitalizados, além da tecnologia blockchain (LIU; KO, 2022).

Soluções tradicionais relacionadas ao tratamento das características inerentes às imagens, como por exemplo, metodologias de visão computacional e inteligência artificial são amplamente aplicadas no contexto de extração de informação destas imagens (DESERNO; ANTANI; LONG, 2009).

Uma alternativa é a extração de informações não descritivas a partir de sinais analógicos como um caminho adicional para a compreensão das informações contidas em um conjunto de dados, e determinar uma compreensão de mais alto nível sobre a imagem ou grupo de imagens (LIANG et al., 2010). O retorno desta informação pode ser usado como fonte de recurso para a base de conhecimento para um possível treinamento de um sistema de aprendizado de máquina (RUI; HUANG, 1998), buscando reduzir as discontinuidades existentes entre o que pode ser visualizado em um gráfico e o que pode ser percebido neste mesmo panorama, como visto em (LI et al., 2010).

O processamento dos sinais de espirometria, através de inteligência artificial, utilizando um conjunto de *wavelets* para uma análise espacial e temporal da capacidade pulmonar, apresenta uma alternativa para classificação e ordenação de um conjunto de exames.

Dessa forma, Um dos sistemas mais robustos e de baixo custo utilizados para a aquisição de sinais de função pulmonar, são os espirômetros. Algumas vantagens que podem ser descritas sobre esse tipo de dispositivo são: a mobilidade em relação à dimensão dos equipamentos; o baixo consumo de energia que possibilita sua operação através de baterias; o uso de conexões sem fio que ampliam a gama de possíveis testes; a rápida curva de aprendizado relacionada à montagem do equipamento, uma vez que, em poucos ciclos de utilização, é possível aprender a montar e capturar sinais e realizar exames de função pulmonar.

Partindo de bancos de dados de exames, e posteriormente utilizando-se um banco de dados para comparação, o projeto busca, através da obtenção de indicadores, desenvolver uma interface humano-computador (IHC) para classificar estes exames.

## 1.1 OBJETIVOS DA PESQUISA

Neste trabalho é proposta uma metodologia para classificar e representar uma relação entre exames de espirometria para diferentes grupos de voluntários, utilizando os sinais elétricos de um espirômetro e realizar a extração de características aplicando funções wavelets, como indicativo de uma compreensão de mais alto nível sobre o exame realizado.

Utilizando o retorno desta informação como fonte de recurso para a melhoria da busca individual e como base de conhecimento para o treinamento de um sistema de aprendizado de máquina, buscamos reduzir as incertezas existentes no diagnóstico.

Também é interessante mencionar que um novo marcador em exames de função pulmonar é de grande importância, visto que na nossa pesquisa bibliográfica poucos trabalhos foram encontrados utilizando um modelo matemático avançado, através das séries temporais ou mesmo da implementação de inteligência artificial, que pode no futuro reduzir a dimensionalidade e tratar os indivíduos de várias regiões etnográficas.

Desta forma este trabalho pretende responder às seguintes perguntas:

- a) Existe a possibilidade de se traçar uma relação entre os dados dos exames de forma a conseguir interpretá-los?
- b) Existe a possibilidade de se traçar uma relação entre os exames de espirometria já realizados e confrontá-los com exames a serem realizados?
- c) É possível desenvolver um modelo de teste e processamento de sinais utilizando-se de inteligência artificial para complementar os métodos utilizados em sistemas de espirometria atualmente utilizados?
- d) Através das informações obtidas seria possível melhorar o desempenho do modelo de interpretação utilizado atualmente?
- e) Obter um novo modelo de processamento de sinais destinado à aplicação na área da saúde pulmonar?

## 1.2 JUSTIFICATIVA DO TRABALHO

De acordo com a pesquisa bibliográfica realizada para a base desse trabalho, poucos artigos ou periódicos demonstram a utilização de séries temporais e de inteligência artificial em busca de bio-marcadores, e com este estudo procuramos obter uma maior eficiência utilizando modelos matemáticos e séries temporais, com posterior aplicação desses resultados em uma estrutura de armazenamento de dados de forma que possam ser estudadas posteriormente, auxiliando o médico pneumologista e ter um diagnóstico mais confiável e preciso.

Também é interessante salientar que até o momento não existem espirometros de fluxo no mercado nacional, sendo que estes equipamentos chegam a um valor muito alto no mercado brasileiro, em torno de trinta mil reais com base nos preços de 2022.

É importante mencionar que na área de pneumologia, percebemos uma carência dos estudos na busca de novos marcadores para um diagnóstico mais preciso, e geralmente não estão aliados a modelos matemáticos mais complexos.

Assim, percebemos que existe um hiato neste tema, que pode ser preenchido com este presente estudo, e que pode ser utilizado em diversas outras áreas da medicina, colaborando com a saúde global.

O uso da inteligência artificial, muito utilizada atualmente em diversos segmentos, tem se tornado um poderoso aliado colaborando com tomadas de decisões em diversas áreas, como segurança, meio ambiente, tecnologias assistivas e também na área da saúde, permitindo que os sistemas se tornem mais autônomos e a influência emocional ou ambiental seja eliminada em diversos casos.

A classificação de dados hoje, com o uso de tecnologias assistivas, tem melhorado a qualidade de vida do ser humano, e permite que se faça o mesmo trabalho em menos tempo do que anteriormente sem essa colaboração.

Novos métodos, novos estudos e a concomitância com equipamentos cada vez mais precisos e miniaturizados, associadas as novas capacidades de processamento de dados, fizeram uma pequena revolução tecnológica em diversos segmentos.

Assim, o uso de todas essas ferramentas (análise de dados, séries temporais e inteligência artificial) disponíveis atualmente, são ferramentas poderosas para uma análise mais precisa, permitindo assim uma melhor qualidade, expectativa e saúde dos indivíduos.

### 1.3 ESTRUTURA DO TRABALHO

Este trabalho está dividido nos seguintes capítulos:

- a) O Capítulo 1 apresenta uma descrição introdutória sobre a motivação do trabalho e os aspectos relevantes para a pesquisa.
- b) O Capítulo 2 apresenta o estado da arte encontrado nesse estudo.
- c) O Capítulo 3 apresenta os fundamentos teóricos que serão utilizados ao longo do trabalho.
- d) O Capítulo 4 apresenta a metodologia, com todas as etapas, processos e equipamentos utilizados e desenvolvidos para a parte experimental da pesquisa.
- e) O Capítulo 5 apresenta os resultados da pesquisa, e
- f) O Capítulo 6 uma discussão sobre os temas abordados e suas conclusões.

## 2 ESTADO DA ARTE

Neste capítulo foram avaliados os principais artigos que contribuíram para a fundamentação deste trabalho, comparando um conjunto específico de textos que tratam da aplicação de wavelets via inteligência artificial em espirometria.

Conforme descrito já em (SPRIGGS, 1978), uma das áreas que vem sendo estudada desde a década de 1970, e atualmente devido ao envelhecimento populacional, é a área médica de pneumologia, onde o principal exame realizado é a prova de função pulmonar, conhecida por espirometria. Para realizar este exame, o médico utiliza um equipamento conhecido como espirômetro, que faz a análise do volume expirado e inspirado pelo paciente e a partir destes dados executa uma série de cálculos numéricos para se obter um resultado de como está a saúde pulmonar da pessoa. Temos relatos desde a época do ano de 1700 que o ser humano tenta obter dados da capacidade pulmonar de seu sistema respiratório.

De acordo com (HANKINSON; GARDNER, 1982), com o intuito de testar e qualificar espirômetros, selecionou-se um conjunto de vinte e quatro formas de onda, a partir de um conjunto maior de manobras de capacidade vital forçada. Parâmetros da espirometria foram determinados em cada forma de onda; e por conseguinte, as exigências de desempenho puderam ser facilmente estabelecidas. Estas formas de onda padronizadas foram inseridas em vários softwares de espirometria e usadas para avaliar o equipamento, através da calibração por uma seringa mecânica de três litros. Essa calibração é usada na rotina diária de exames de espirometria, realizando o teste do software que executa este exame. Dessa sua pesquisa resulta o padrão de normalidade utilizado desde 1982 em pacientes americanos.

Estudando a pesquisa de (FINKELSTEIN et al., 1995), percebemos a importância do exame de espirometria também na área de transplantes, pois foi possível perceber que existe diferença significativa nos resultados dos exames de espirometria em valores que podem ser usados como marcadores de uma possível rejeição do órgão transplantado.

Os métodos diagnósticos atuais são baseados na análise dos sinais obtidos, aplicando-se métodos convencionais como análise de rampas descendentes e ascendentes do sinal adquirido (KABIR; IFTEQAR; BHAT, 2013).

Em (AKAY, 1997), é mostrada a aplicação de wavelets na área de biomedicina. Mostra-se que os resultados obtidos têm sido satisfatórios em análise de sinais e detecção prematura de artefatos em exames de diagnóstico por imagem. É também descrito neste estudo qual seria o avanço e em que segmentos este modelo matemático seria utilizado.

As funções wavelets, aliadas à inteligência artificial trazem para a área da saúde uma boa perspectiva para aprimorar novas técnicas de coleta de dados, processá-los e trazer um resultado diagnóstico final preciso (YUEFAN; SEN; ZHENGTAO, 2021). Alguns pesquisadores já utilizam esses métodos para alcançar seus resultados em diagnóstico médico e processamento de sinais de exames (SERHAL; ADBALLAH; MARION, 2022).

Em (COSTA; JARNARNI, 2001), é realizado um estudo que verifica métodos de diagnósticos em exames de espirometria, estabelecendo parâmetros para realizar uma interpretação confiável, formato ainda utilizado atualmente.

Utilizando-se séries temporais o que se deseja é aproximar uma função por uma combinação linear de funções ortogonais (SEROV, 2017), visando gerar uma função periódica de período  $2\pi$  de quadrado integrável.

Na pesquisa de (RODRIGUES; PEREIRA, 2001), é apresentada a resposta em exames de espirometria com o uso do bronco dilatador, demonstrando em seus resultados que não ocorreu diferença significativa em parâmetros utilizados no diagnóstico de patologias pulmonares, mas sim, em outros valores que geralmente não são interpretados.

Em (LADOSKY et al., 2001) são comparados os valores encontrados no exame de espirometria em previstos diferentes. Previsto é um termo utilizado para definir os valores esperados na prova de função pulmonar do paciente, de forma a comparar estes valores obtidos no exame, com um previamente estabelecido, de forma a classificar o indivíduo em relação a população de mesmas características em que se encontra. Dessa forma, o estudo mostra que esses valores dependem basicamente do local onde está sendo realizado o exame (população local) e não do previsto utilizado.

As wavelets vêm sendo aplicadas na área médica, como em (ADELI; ZHOU; DADMEHR, 2003), que utiliza as transformadas de wavelets a fim de analisar pacientes com epilepsia. Ele se utilizou de um método para decompor os dados de frequências de picos de EEG de 3Hz (Hertz), e fazendo do uso da família de wavelets de Daubechies, observou que seria necessário realizar um estudo com maior profundidade, pois a idéia inicial era predizer se o paciente poderia ter uma crise epilética ou não.

No estudo de (AKPINAR-ELCI; FEDAN; ENRIGHT, 2006), é mostrado o quanto é difundido e utilizado o exame de função pulmonar, além de mostrar o quão importante é a compreensão do paciente na execução do exame. Pacientes obstrutivos por algumas vezes não realizam a manobra expiratória dentro dos valores de tempo aceitáveis, gerando uma frustração em não completar o exame, tanto pelo técnico quanto ao próprio paciente. Assim, este tipo de

situação colabora com erros no estudo, devido ao uso de testes de má qualidade nas amostras. Então em seu estudo propõe o uso da variável VEF6 (Ventilação Expiratória Forçada nos 6 primeiros segundos) como critério de classificação em patologias pulmonares.

(FERRARI; HOTO; MACULAN, 2006), utilizou quatro sistemas de redes neurais a fim de determinar se o paciente possui doença pulmonar obstrutiva crônica; baseando-se em vários métodos classificadores, entre eles as wavelets, obtendo uma possibilidade de prever se o paciente possui tal enfermidade.

Conforme mostrado em (CARTA et al., 2007) é possível realizar um comparativo entre leituras espirométricas de pressão diferencial e pressão linear, obtendo resultados satisfatórios nos dois casos, ou seja, os valores são próximos entre si.

O estudo de (KALICKA; SLOMINSKI; KUZIEMSKI, 2008), modelou a curva de espirometria através do método de regressão matemática não linear, obtendo sucesso com este modelamento.

Em (PINHEIRO et al., 2009), comparou-se os valores do previsto brasileiro com os sugeridos pela Global Lung Function Initiative 2012. Alguns parâmetros utilizados para diagnóstico obtiveram valores diferentes, alertando sobre a escolha correta para a realização do exame.

Em (SAHIN et al., 2010) apresenta o uso de máquinas de suporte vetorial multiclasse (SVMs) para o diagnóstico de padrões de espirometria (normal, restritivo, obstrutivos). As decisões SVM foram fundidas utilizando códigos de correção de erros (CRC). O SVM multiclasse juntamente com o CRC foi treinado em três parâmetros de espirometria, (volume expiratório forçado em 1s-VEF1, capacidade vital forçada – CVF e razão FEV1/FVC). A precisão total de acurácia dessa SVM foi de 97,32%. Os resultados obtidos confirmaram a validade das SVM para ajudar na tomada de decisão clínica.

Como em (KAVITHA; SUJATHA; RAMAKRISHNAN, 2010), diversos algoritmos têm sido estudados com o intuito de estabelecer padrões de normalidade, neste caso diagnosticar obstrução de vias aéreas através da máquina de suporte vetorial.

Conforme estudado em (LAY-EKUAKILLE; VENDRAMIN; TROTTA, 2010), é possível recriar a curva espirométrica, utilizando coeficientes de equações matemáticas não lineares, onde neste estudo utilizaram um algoritmo conhecido como CGA, ou "controlled genetic algorithm". Conclui-se que é possível a utilização de cálculos matemáticos em curvas de espirometria.

Através de processamento matemático, foi possível prever as condições de saúde em transplantados pulmonares, como descrito em (WANG, 2011). No campo de caracterização de

pacientes foi possível estabelecer padrões de reconhecimento de padrões baseando-se nos valores do volume expiratório pulmonar dos pacientes.

Verificamos que em (DESHMUKH; PANDITRAO, 2012) é descrito um método para a construção de um espirometro baseado em um termistor que esfria de acordo com o fluxo que passa por ele. Conseguiram obter resultados satisfatórios pelo fato de conseguir visualizar a curva espirométrica e também se o paciente tossiu durante o exame.

A pesquisa de (KHUBANI, 2012), estabeleceu parâmetros para se prever o reconhecimento de padrões utilizando a máquina de suporte vetorial, a fim de auxiliar nas investigações em espirometria. A máquina de suporte vetorial constrói um hiperplano ou conjunto de hiperplanos em um espaço tridimensional, que pode ser usado para a classificação, ou em outras tarefas, reunindo dados de diferentes hospitais. Os dados adquiridos foram então utilizados para prever a precisão de reconhecimento de padrões espirométricos.

Uma alternativa mostrada em (ASHER, 2012), utiliza wavelets objetivando a classificação dos valores de dióxido de carbono presentes na corrente sanguínea, através do exame de capnografia, que possui um grande potencial em analisar indivíduos com doença pulmonar obstrutiva crônica ou indivíduos com restrição pulmonar.

No campo de caracterização de pacientes, como estudado em (MHETRE; KHUBANI, 2012), o mesmo conseguiu estabelecer padrões de reconhecimento de valores baseando-se nos valores do volume expiratório pulmonar dos pacientes.

Conforme (DEVI; VENKATESAN, 2013), o teste de espirometria é um teste essencial no caso de doenças respiratórias e doenças pulmonares relacionadas. Este trabalho trata da previsão de FEV1, que é um dos fatores mais importantes e um dos valores que decidem o diagnóstico sobre doenças respiratórias. Essa previsão de valores de FEV1 melhorou o método de diagnóstico da espirometria. Quando os dados são incompletos ou mal gravados, a precisão do diagnóstico das anomalias pode ser melhorada usando métodos baseados em SVM.

Considerando a área da pneumologia, verificamos que os pesquisadores estão desenvolvendo métodos numéricos avançados, como o estudado em (BETANCOURT et al., 2013), que utiliza cálculos com wavelets no intuito de classificar o grau de severidade pulmonar em pacientes asmáticos.

Um trabalho realizado por (LOPATA et al., 2013), demonstra os requisitos de espirometros existentes no mercado e quais são as necessidades de leitura e aquisição dos dados necessários para realizar um exame aceitável e o seu correto diagnóstico, além de alertar para a necessidade de correção de uma parcela do exame conhecida como espaço morto.

No estudo de (LABATE et al., 2013) se utiliza a decomposição de wavelets com o objetivo de retirar de sinais de ECG um sinal que expressa a respiração do indivíduo. Sua conclusão mostra que é importante a escolha correta da família de wavelets que irão ser utilizadas no estudo. A wavelet de Meyer possui derivação infinita, rápida atenuação e espectro finito. Os cálculos numéricos se beneficiam deste modelo, principalmente em grandes áreas da engenharia.

No estudo de (ZHANG; DENG; HAN, 2013), é discutido as propriedades das wavelets de Meyer em sinais de imagens obtendo duas expressões centrais desenvolvendo um software para a aplicação de processamento de sinais.

Em (VALENTINUZZI; JOHNSTON, 2014) é mostrado a história da espirometria, e apresenta diversos equipamentos que eram utilizados para analisar o comportamento fisiológico e dinâmico pulmonar, equipamentos que não possuíam na época qualquer cálculo matemático ou sistema digital de análise dos dados obtidos.

Também no estudo de (MYTHILI; SUJATHA, 2015) considera útil o uso de classificadores SVM para realizar o diagnóstico da espirometria, uma ferramenta valiosa no diagnóstico de patologias pulmonares, além de melhorar a precisão na caracterização de seu resultado.

Um estudo feito por (MAIA, 2015), avaliou em seu estudo da detecção da leishmaniose através de exames de espirometria, obtendo resultados satisfatórios em seu trabalho, atentando que as pessoas portadoras dessa patologia apresentaram comportamento semelhante a pessoas com restrição pulmonar nos exames, concluindo que a espirometria é uma ótima ferramenta para determinar patologias viscerais.

Em (KASSEM; HAMAD; EL MOUCARY, 2015) foi desenvolvido um modelo de equipamento com o intuito de registrar os valores de pico de fluxo expiratório (FEV) em portadores de asma. O equipamento comparava valores de fluxo pulmonar, correlacionando os valores lidos com parâmetros aceitos na sociedade médica, dando um resultado instantâneo do paciente, que pode ser interpretado pelo médico que acompanha.

Conforme (MORETO TRINDADE; LINS FAGUNDES DE SOUSA; LUIS PEREIRA ALBUQUERQUE, 2015), “A espirometria é a principal ferramenta funcional respiratória e deve persistir por várias décadas.” Este mesmo autor menciona que a morfologia das curvas devem constar do diagnóstico funcional, e o nosso trabalho apresenta resultados numéricos para auxiliar neste estudo morfológico e que podem ser num futuro utilizado para uma maior precisão do diagnóstico.

Conforme descrito em (LEE, J. et al., 2016), parâmetros da espirometria convencionais têm demonstrado fraca correlação com os sintomas e estado de saúde de doença pulmonar

obstrutiva crônica (DPOC). Embora seja bem conhecido, o padrão da curva de fluxo-volume expiratório (CVF) representa uma disfunção ventilatória, e pequenas tentativas têm sido feitas para derivar os parâmetros quantitativos por meio da análise da curva. Neste estudo, o objetivo foi obter parâmetros úteis da capacidade vital forçada expiratória via análise gráfica e tentou validá-los em pacientes com DPOC.

Um estudo realizado em (SOLEIMANI et al., 2016) teve como objetivo realizar a medição de parâmetros de espirometria através da gravação de imagens do comportamento do tórax do indivíduo, obtendo sete parâmetros necessários a avaliação pulmonar.

Em (ULBRICHT, 2016), foi estudado o comparativo de previstos de forma a analisar as diferentes variações étnicas e concluiu que devido uma utilização errônea do previsto utilizado como parâmetro comparativo do exame, pode-se diagnosticar a pessoa incorretamente, ou seja, o mesmo indivíduo, dependendo do padrão de previsto utilizado, pode ser considerado como obstrutivo, restritivo ou normal.

A aplicação de modelos para a modificação de coeficientes gerados após a transposição de um sinal para o domínio das funções wavelets, é amplamente utilizado. Um exemplo deste tipo de aplicação, pode ser observado no trabalho de (DAUBECHIES, 1999) com modelos de compressão de imagens, onde os valores de seleção acima de um limite aplicados em imagens em escala de cinza, definindo os coeficientes restantes para zero, a fim de tornar a compressão da imagem, mas cuidadosamente preservando a quantidade total de energia no sinal.

Os estudos de (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) apresentam uma abordagem para suavização de wavelets baseada em regressão linear, para criar um limite para minimizar coeficientes pequenos no domínio wavelet, de forma a recompor um sinal amortizado. Eles também descrevem um método de uso de regressão (*lasso*) que faz a redução dos coeficientes do sinal com um pacote de software chamado SURE (*Stein Unbiased Risk Estimation*) (DONOHO; JOHNSTONE, 1994), que define coeficientes para patamares mínimos, ou até mesmo para zero, no domínio das wavelets, e em seguida é revertido novamente para o domínio original do sinal com uma transformada inversa de wavelet.

Em (PIVA; BECHELLI; BELARDI, 2018) foi demonstrado que o sinal de processamento wavelet define coeficientes que reproduzem bem o sinal original. Assim, tais coeficientes podem ser utilizados para classificar e caracterizar determinadas regiões da coleta de dados que, no caso de uma nova inserção de dados, será criado um valor auxiliar de diagnóstico (biomarcador) para determinar e auxiliar as decisões nas situações em que o diagnóstico não é tão evidente.

O algoritmo de agrupamento K-means é um algoritmo de aprendizado não supervisionado muito simples. Ele fornece uma maneira muito fácil de classificar um determinado conjunto de dados em um certo número de clusters, ou seja, um conjunto de dados como  $x_1, x_2, x_3, \dots, x_n$  são que são agrupados em K clusters. A idéia principal por trás deste algoritmo é definir centros K, um para cada cluster atentando-se que os centros do cluster K devem ser selecionados aleatoriamente (ALAM; RAHMAN; HOSSAIN, 2019).

A Transformada de Haar é uma boa escolha quando a detecção de bordas e caracterização de contornos são alguns dos objetivos principais (PAT; RAMAMOCHAN, 1987). Demonstrar também quando o algoritmo wavelets é aplicado no sinal original, características significativas começam a surgir quando o sinal varia repentinamente, criando a possibilidade de recuperar essas discrepâncias com marcadores (SHUVRA; DREPTTE; TAKALA, 2019). Combinado com modelos gaussianos, essa detecção de borda é mais precisa em biosinais (ROMERO, 2019).

### 3 BASE TEÓRICA

#### 3.1 WAVELETS

Conforme visto em (MORETIN, 1999), a análise das wavelets já era utilizada por vários pesquisadores como uma alternativa à série de Fourier clássica. As wavelets de J. S. Liénard e de X. Rodet se relacionam ao tratamento numérico de sinais sonoros e as de J. Morlet foram desenvolvidas para estocar e interpretar os sinais sísmicos, sendo que o físico A. Grossmann juntamente com Morlet foram os responsáveis pelo desenvolvimento da transformada das wavelets contínuas. O uso das wavelets aumentou a partir de 1985, quando foi dado um novo impulso a esta teoria através da contribuição matemática e especialistas em processamento de sinais.

Usualmente, nos deparamos com sinais no ambiente, que podem ser interpretados como sinais elétricos, por exemplo. Uma pequena movimentação de terra em nosso planeta pode ser analisada através de sensores que captem essa movimentação, gerando um sinal elétrico que pode ser interpretado e gerado um gráfico, conhecido como sinal sísmico. Um sinal mio-elétrico comum é o estudo das ondas cerebrais, que são sinais de baixa amplitude e devem ser lidos e posteriormente filtrados a fim de se obter alguma característica que se possa analisar este sinal; esses sinais são conhecidos como EEG, ou eletro-encefalografia (SMITH, 2005) e (CHEN; LIANG; HO, 2012).

Na área médica, diagnosticar e prevenir malefícios respiratórios devido a diversas variáveis como stress, tabagismo, má-formação congênita, ambiente social e problemas respiratórios são frequentes, devendo utilizar uma metodologia para avaliar de modo não invasivo a qualidade respiratória do paciente.

Atualmente é possível analisar a saúde pulmonar através da medição do ar que entra e sai dos pulmões. Hoje, a tecnologia permitiu que se utilize um equipamento portátil, de baixo custo, que realiza com precisão esta análise, conhecido como espirometro. Este equipamento também gera um sinal elétrico, que é interpretado por um computador, gerando gráficos e valores que podem ser lidos por este computador.

A transformada wavelet vem sendo utilizada como uma ferramenta alternativa para processamentos digitais de imagens e de sinais, pelo fato de gerar matrizes esparsas. Estas matrizes geram uma quantidade de valores nulos consideravelmente alta, permitindo que o tempo de processamento caia de uma forma expressiva, como visto no artigo publicado pelo autor (NAMI, 2010), onde o tempo de processamento usual, que era em torno de 4 a 5 minutos, fora

reduzido para 3 a 4 segundos, assim como também ampliando sua acurácia, e os resultados aplicando-se métodos convencionais que eram por volta de 60%, passaram a ter uma taxa de acurácia de 86%, aumentando sua precisão, e não comprometendo a qualidade do resultado final. Assim, esta função tem sido utilizada nos mais diversos campos que não possuem uma solução no domínio de valores reais, possibilitando um acréscimo no campo de pesquisa onde a rapidez computacional é requerida (ALDROUBI; UNSER, 1995).

Definimos a Função Wavelet de um sinal como sendo a somatória temporal de todo o sinal  $f(t)$  multiplicada pela versão escalada e deslocada da função wavelet  $\psi_{a,b}(x)$ , descritos conforme a Equação 1, assim podemos definir uma wavelet mãe a partir da seguinte função:

$$\psi_{a,b}(x) = |a|^{-1/2} \psi\left(\frac{x-b}{a}\right), a, b \in \mathbb{R}, a \neq 0 \quad (1)$$

Os parâmetros  $a$  e  $b$  variam continuamente, como visto em (GRAPS, 1995), ou seja,  $a$  é parâmetro de escalamento, e  $b$  sendo parâmetro de deslocamento, ou translação. Dessa forma, assume-se valores para  $a = 2^{-j}$  e para  $b = k2^{-j}$  denominando-se wavelets e definidas como se segue, quando utilizada na forma bidimensional:

$$\psi_{j,k}^{(H)}(x) = 2^{j/2} \psi(2^j x - k) \quad j, k \in \mathbb{Z} \quad (2)$$

O parâmetro de escalamento influi no acréscimo ou decréscimo da largura e amplitude da wavelet. e o parâmetro translacional responde por percorrer o sinal em função do tempo (MORETTIN, 1999).

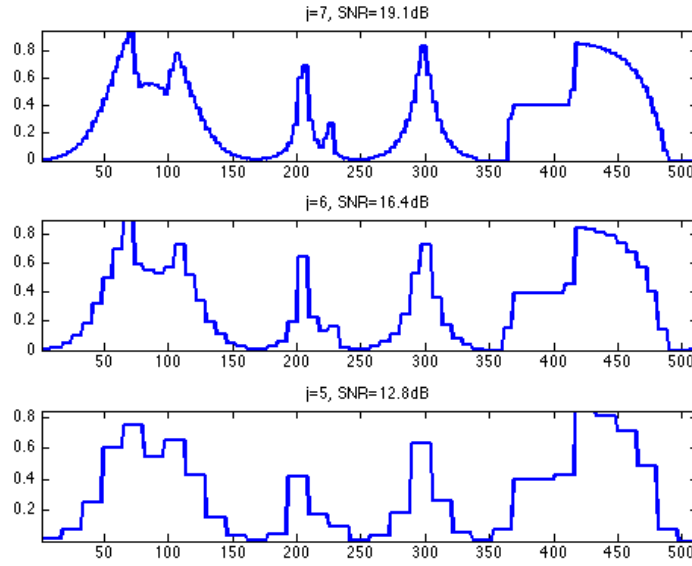
As wavelets de Haar são definidas no seguinte intervalo, sendo  $\phi$  a wavelet pai:

$$\phi^{(H)}(x) = \begin{cases} 1 & 0 \leq x < 0.5, \\ 0 & \text{Para Outros Intervalos} \end{cases} \quad e \quad (3)$$

Considerando que um pequeno parâmetro de escalamento representa um sinal reduzido da função, de forma análoga, é possível detectar em um sinal componentes de rápidas mudanças de sinais em alta frequência (acima de 256Hz). Entretanto, componentes de baixa frequência podem ser destacadas do sinal ajustando uma função wavelet "esticada", ou inserindo parâmetros escalares maiores (VALENS, 1999b).

Então, podemos definir as wavelets como a estudada por Haar (MORETTIN; PINHEIRO; VIDAKOVIC, 2017), onde descreve a função wavelets como uma base ortonormal definida em um domínio  $[0,1]$ , podendo dessa forma ser expressa graficamente. A Figura 1 ilustra a aplicação desta wavelet a um sinal genérico:

Figura 1: Wavelet de HAAR



Fonte: (CHEN; LIANG; HO, 2012)

Adotando valores especiais para  $a$  e  $b$ , com  $j$  e  $k$  pertencentes ao conjunto dos números inteiros relativos, sendo  $j$  e  $K$  os limites de passo que a wavelet percorre no domínio da frequência, assim descrevemos que as wavelets são funções matemáticas diferentes que por exemplo, as transformadas de Fourier, esta última, utilizando-se de ondas senoidais como base; ou seja, percorre o sinal apenas no domínio da frequência, e utilizando as funções wavelets podemos percorrer o sinal dentro do domínio da frequência (baixas ou altas frequências), deslocamento e amplitude.

Assim sendo, as wavelets de Haar podem ser definidas através da seguinte equação, ou seja a função  $f(x,y)$  /citecheremkhin2019, que pode ser descrita por uma função de ponderação (pulsada) :

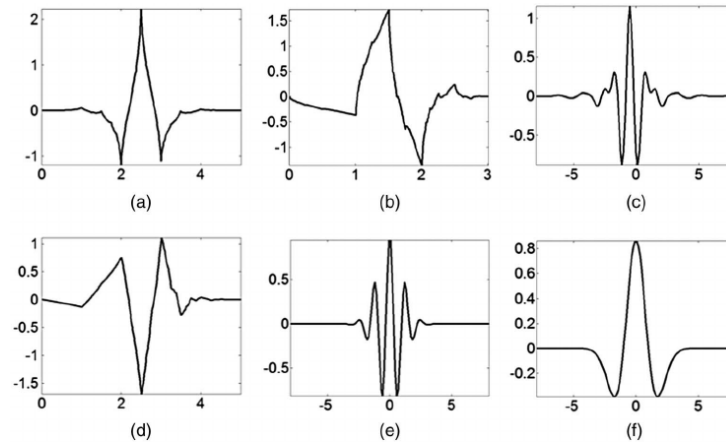
$$f(x,y) = \sum_{k=-\infty}^{\infty} c\varphi(x,y) + \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} f(x,y) \psi_{j,k}(x,y) \quad (4)$$

Além de Haar temos outros tipos de wavelets tais como Battle-Lemarié, Chapéu Mexicano e Shannon (BELARDI, 2013), além de Meyer (ZHANG; DENG; HAN, 2013).

Representamos as Wavelets de Meyer na figura 2:

Esta wavelet pode ser definida pelas seguintes funções:

Figura 2: Wavelet de Meyer



Fonte: (ZHANG; DENG; HAN, 2013)

$$f(\omega) = \begin{cases} 0 & , \omega \leq 0 \\ e^{-1/\omega^2} & , \omega < 0 \end{cases} \quad (5)$$

$$e^{-i\omega/2} = \sqrt{\phi(\omega/2)^2 - \phi(\omega)^2} \quad (6)$$

$$\phi(\omega) = \sqrt{h(\omega)h(-\omega)} \hat{\psi}(\omega) \quad (7)$$

$$h(\omega) = \frac{f\left(\frac{4\pi}{3} - \omega\right)}{f\left(\omega - \frac{2\pi}{3}\right) + f\left(\frac{4\pi}{3} - \omega\right)} \quad (8)$$

Na Figura 3 representamos a Wavelet de Battle-Lemarié:

$$\varphi(x) = \begin{cases} \frac{1}{2}x^2 & 0 \leq x < 1, \\ -x^2 + 3x - \frac{3}{2} & 1 \leq x < 2, \\ \frac{1}{2}(x-3)^2 & 2 \leq x < 3 \end{cases} \quad (9)$$

Podemos formular essa função pela seguinte equação:

Esta função wavelet pode ser formulada pela seguinte equação:

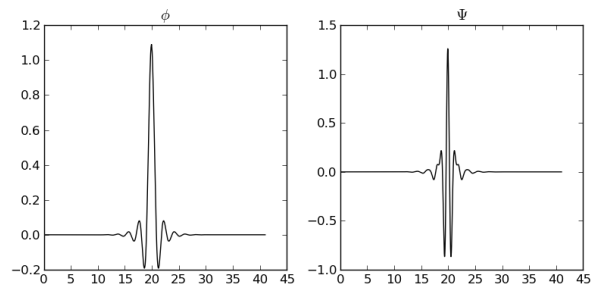
$$\psi(x) = (1 - x^2)e^{-x^2/2} \quad (10)$$

Na Figura 4 representamos a wavelet Chapéu Mexicano:

A Wavelet de Shannon é definida pela seguinte equação:

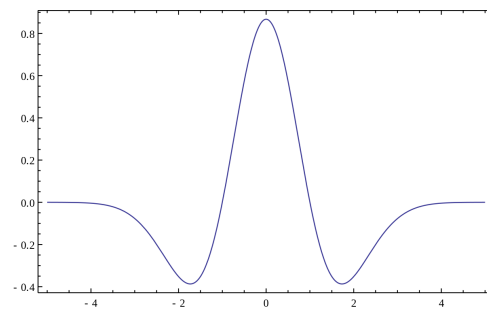
$$(\psi(x) = \frac{\sin(\frac{\pi x}{2})}{\frac{\pi x}{2}} \cos(\frac{3\pi x}{2}) \varphi(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x}, & x \neq 0 \\ 1, & x = 0 \end{cases} \quad (11)$$

Figura 3: Wavelet de Battle-Lemarié



Fonte: (BELARDI, 2013)

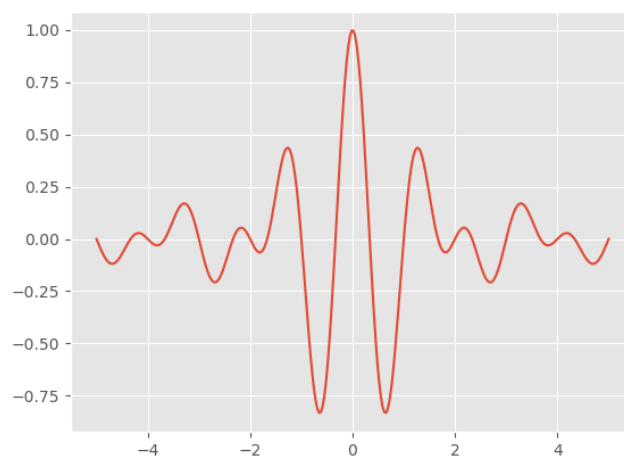
Figura 4: Wavelet Chapéu Mexicano



Fonte: (BELARDI, 2013)

Esta wavelet pode ser demonstrada pela Figura 5.

Figura 5: Wavelet de Shannon



Fonte: (BELARDI, 2013)

A Equação 12, define de forma geral o modo de operação das wavelets. Diferente da teoria da transformada de *Fourier*, que utiliza ondas senoidais como base, a teoria das wavelets

permite o uso, e eventualmente novos desenvolvimentos, de uma variedade de wavelets mãe conforme a necessidade da aplicação (VALENS, 1999a). As wavelets mãe, definem a função sobre o qual o sinal será decomposto, utilizando a analogia com a teoria de *Fourier*, no lugar das ondas senoidais.

$$\psi(x) = \begin{cases} +1 & 0 \leq x < 1/2, \\ -1 & 1/2 \leq x < 1, \\ 0 & \text{restante.} \end{cases} \quad (12)$$

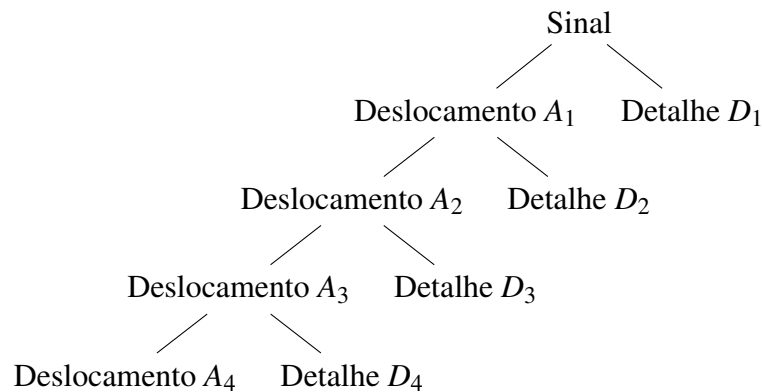
A estrutura de formação das wavelets permite que tanto os sinais de alta frequência como os de baixa frequência possam ser aplicados de forma iterativa ao sinal. Dessa maneira, a decomposição pode ser aplicada em diversos níveis, aumentando a quantidade de detalhes do sinal e realçando as altas frequências.

Podemos definir "níveis" de decomposição de uma forma mais comum, como sendo a quantidade de iterações que essa função vai percorrer. Quando utilizamos o modelo com apenas uma iteração, o software que contém esse modelo, executa apenas uma vez o processamento. É sabido que o sinal vai se decompondo no domínio do tempo, assim, o sinal original passa a ser um modelo anterior ao que é utilizado numa segunda iteração.

Ou seja, se pensarmos em um modelo computacional de aplicação deste sinal, o código do software executa diversas vezes a mesma rotina de forma que a cada iteração, os dados utilizados são os da iteração anterior, e não mais o sinal original.

Na Figura 6, o sinal é decomposto em 4 níveis conforme segue:

Figura 6: Decomposição de 4 níveis de Wavelet.



Fonte: Autor.

### 3.2 CLASSIFICADOR DE RIDGE

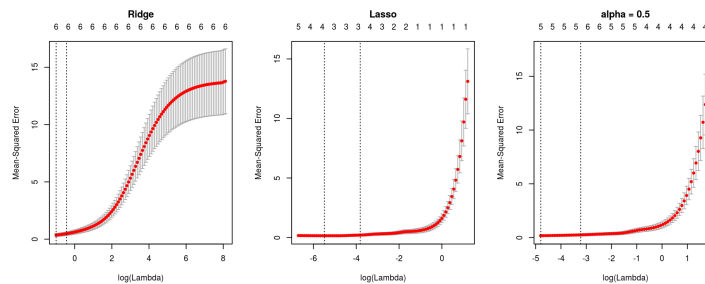
Utilizamos a regressão de Ridge, pois em nossos estudos, conforme visto em (ALDROUBI; UNSER, 1995), se comporta adequadamente em diversas situações onde transformadas de wavelets são utilizadas na área médica.

A Regressão de *Ridge*, descrita pela Equação 13, conforme visto em (JEREMIA; NURROHMAH; FITHRIANI, 2020) é uma técnica que trabalha a redução dos coeficientes  $\beta$  de uma equação linear, tendo como base a multiplicação destes coeficientes por um fator  $\lambda$  com o objetivo de minimizar o erro associado ao sinal original. Com esse processo de modificação das variáveis, a Regressão de *Ridge* reduz a multicolinearidade, que é uma situação onde um forte relacionamento entre as variáveis independentes torna difícil a estimativa de efeitos destas mesmas variáveis sobre uma variável dependente, ou de forma mais precisa, a variância entre alguns coeficientes obtidos na aplicação das wavelets (saídas "a e b" do modelo), se tornam grandes, numa ordem de grandeza que não pode ser previamente estabelecida, pois depende do sinal original (o sinal original varia sua amplitude, pois é oriundo do valor de fluxo lido pelo espirômetro).

$$\hat{\beta}_{ridge} = (X'X + \lambda I)^{-1} X'y \quad (13)$$

A influência do  $\hat{\beta}_{ridge}$ , pode ser interpretado através da Figura 7, que representa o termo de regularização  $l_2$  como a linha em vermelho que tende a sua estabilização conforme visto nesta figura. O ponto  $\hat{\beta}$  representa o erro quadrático mínimo. O erro aumenta exponencialmente a medida que se aproxima do termo de regularização. Uma vez que os termos se cruzam, ambos serão minimizados, determinando o fator  $\hat{\beta}_{ridge}$ .

Figura 7: Estimativa dos coeficientes da regressão de *ridge*.



Fonte: "Adaptado de "Hastie, Tibshirani e Wainwright (2015).

Regressão de *Ridge* é conhecida pela resiliência ao sobre-ajuste (*overfitting*), ou seja, o ajuste do ruído em detrimento do ajuste do sinal. Averso a isso, desejamos que um ruído não deve ser suavizado, e sinais vizinhos onde existe um sinal real não possuem uma transição gradiente.

Quando o sinal possui uma amostragem pequena, em relação aos parâmetros que o modelo possui, é possível que ocorra uma suavização do próprio ruído. A regressão de Ridge possui um ajuste dessa suavização, tornando um modelo mais ideal para o objeto deste estudo, acompanhando a tendência do sinal do experimento.

Obteve-se dessa forma um resultado positivo que, conjuntamente com algoritmos de aprendizado de máquina, resultou em um único ponto de convergência (JAMES et al., 2013), possuindo também um bom desempenho contra vetores de entrada que possuem tamanhos maiores que o tamanho das amostras de entrada (vetores de alta dimensionalidade). Este grande vetor é processado por regressão de *ridge* para identificar uma quantidade de variáveis que contribuem para o sinal e variáveis que podem ser reduzidas (*ridge*) ou mesmo descartadas como em outros métodos como por exemplo o (*lasso*) (ZHAO; OGDEN; REISS, 2012).

Nesse tipo de situação, uma regressão pelo métodos dos mínimos quadrados passa a ter coeficientes muito altos. No caso do classificador de *ridge* (EHSANES; K., 2019), conforme a Equação 13, o fator  $\alpha$  é configurado como parâmetro de entrada do treinamento do classificador para que seja controlado o resultado final da predição.

Sendo  $w = (w_1, \dots, w_p)$ , um conjunto de coeficientes que multiplicam o vetor de entrada do classificador,  $w_0$  o coeficiente linear que intercepta a função em  $y$  para  $x = 0$ ,  $X$  representa o vetor de entrada das medidas que precisam ser classificadas,  $y$  representa a categoria (*label*) em que o vetor  $X$  é classificado e  $\hat{y}(w, x)$  representa o resultado esperado após o treinamento do classificador, temos:

$$\hat{y}(w, x) = w_0 + w_1x_1 + \dots + w_px_p \quad (14)$$

Para uma regressão através do método dos mínimos quadrados, temos:

$$\min_w ||Xw - y||_2^2 \quad (15)$$

E a regressão de *ridge* é determinada por Jeremia, Nurrohmah e Fithriani (2020), conforme a equação abaixo:

$$\min_w ||Xw - y||_2^2 + \alpha ||w||_2^2 \quad (16)$$

A Equação 16, pode ser interpretada com uma regressão linear pelo método dos mínimos quadrados, apresentado na equação 15, com um fator adicional que impõe uma penalização através do conjunto de coeficientes.

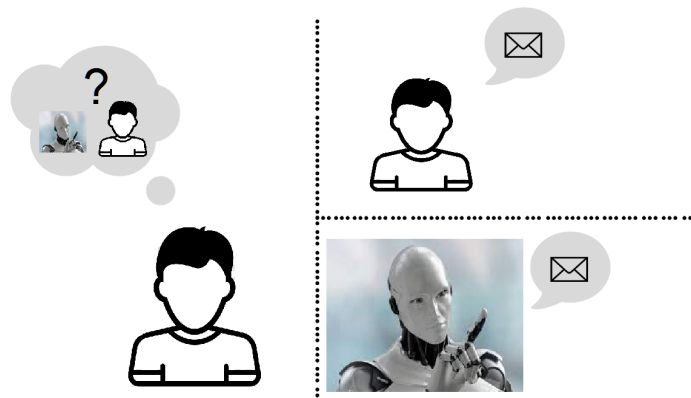
Para controlar os resultados, vários níveis de alpha em um método de classificação de aprendizado de máquina de pesquisa de grade *gridsearch* podem ser aplicados para determinar o resultado máximo de classificação.

É importante salientar que a regressão de ridge, juntamente com as wavelets, são as mais eficazes para este tipo de estudo como o proposto por nós, pois possibilitam a classificação no domínio da frequência, amplitude e deslocamento, e serão fundamentais para conseguirmos obter um resultado na segregação dos dados analisados, conseguindo obter um alto índice de precisão na definição das classificação dos dados analisados.

### 3.3 INTELIGENCIA ARTIFICIAL E ENGENHARIA DE SOFTWARE

Ao dizer que a inteligência artificial é interessante, precisamos dizer o que ela é. Podemos visualizar definições de IA, dispostas ao longo de duas dimensões, que iremos comentar a seguir, na Figura 8.

Figura 8: Modelo de Testes de Turing



Fonte:Autor

As definições do lado esquerdo medem o sucesso em termos de fidelidade ao desempenho humano, enquanto as definições do lado direito medem o sucesso comparando-o a um conceito ideal de inteligência, chamado de racionalidade. Um sistema é racional se “faz a coisa certa”, dado o que ele sabe.

O teste de Turing, proposto por (TOURING, 1950), foi projetado para fornecer uma definição operacional satisfatória de inteligência. O computador passará no teste se um

interrogador humano, depois de propor algumas perguntas por escrito, não conseguir descobrir se as respostas escritas vêm de uma pessoa ou de um computador. Os detalhes do teste e também se um computador seria de fato inteligente se passasse nele, podem ser encontrados neste artigo citado acima.

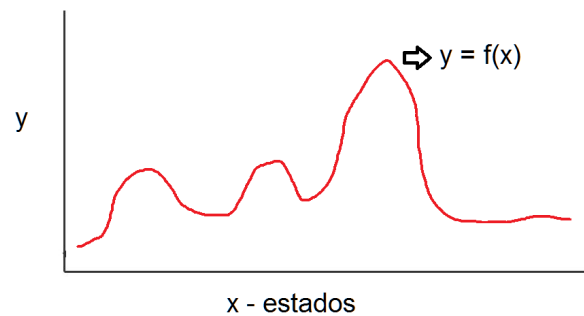
Mas observamos que programar um computador para passar no teste exige muito trabalho. Os filósofos demarcaram a maioria das ideias importantes sobre a IA, mas o salto para uma ciência formal exigiu certo nível de formalização matemática em três áreas fundamentais: lógica, computação e probabilidade. A teoria da decisão, que combina a teoria da probabilidade com a teoria da utilidade, fornece uma estrutura formal e completa para decisões (econômicas ou outras) tomadas sob a incerteza, ou seja, em casos nos quais as descrições probabilísticas captam de forma apropriada o ambiente do tomador de decisões.

A visão do cérebro como um dispositivo de processamento de informações, uma característica importante da psicologia cognitiva, tem suas origens nos trabalhos de William James (1842-1910) (MCGRANAHAN, 2017).

Podemos colocar as representações ao longo de um eixo de complexidade crescente e poder de expressividade — atômico, fatorado e estruturado. Para ilustrar essas ideias, consideremos um componente do agente em particular, como aquele que lida com “O que minhas ações realizam”. Os algoritmos de busca são projetados para explorar sistematicamente espaços de busca, como exemplo árvores binárias e seus métodos de busca, grafos e seus respectivos métodos, assim como buscas simples de dados em sistemas informatizados, onde podemos citar filas, listas, buscas sequenciais em vetores. Esse caráter sistemático é alcançado mantendo-se um ou mais caminhos na memória e registrando-se as alternativas que foram exploradas em cada ponto ao longo do caminho e quais delas não foram exploradas. Quando um objetivo é encontrado, o caminho até esse objetivo também constitui uma solução para o problema (FRANZONI; PETTINI, 2004).

O algoritmo de busca de subida de encosta (versão encosta mais íngreme) é mostrado na Figura 9. Ele é simplesmente um laço repetitivo que se move de forma contínua no sentido do valor crescente, isto é, encosta acima. O algoritmo termina quando alcança um “pico” em que nenhum vizinho tem valor mais alto. O algoritmo não mantém uma árvore de busca e, assim, a estrutura de dados do nó atual só precisa registrar o estado e o valor de sua função objetivo. A busca de subida de encosta não examina antecipadamente valores de estados além dos vizinhos imediatos do estado corrente.

Figura 9: Busca Subida de Encosta



Fonte: autor

O problema de observabilidade parcial, onde a percepção do agente não é suficiente para definir o estado exato. Como observado, se o agente estiver em um dos vários estados possíveis, uma ação pode levar a um dos diversos tipos de resultados possíveis, mesmo se o ambiente for determinístico. O conceito fundamental necessário para resolver problemas parcialmente observáveis é o estado de crença, que representa a crença atual do agente sobre os possíveis estados físicos em que poderia estar, dada a sequência de ações e percepções até aquele ponto. Com o cenário mais simples para o estudo dos estados de crença, que é quando o agente não tem sensores; então adicionamos um sensoriamento parcial, bem como ações não determinísticas (DESERNO; ANTANI; LONG, 2009).

Quando essas informações não podem ser obtidas com essa percepção, ocorre um problema de sensoriamento, ou, em alguns casos, um problema conformante. Acreditamos em algumas vezes que este sensoriamento não irá conseguir elucidar determinado problema se não conseguir identificar o estado que se encontra, mas mesmo assim, problemas onde não existe este sensoriamento podem se tornar solúveis na maioria das vezes (EAKINS, 2002).

Definimos a IA como o estudo da ação racional, o que significa que o planejamento é a elaboração de um plano de ação para atingir determinados objetivos e é uma parte crítica da IA (DIGNUM, 2019). Além disso, os agentes sem sensoriamento podem ser surpreendentemente úteis, principalmente porque eles não dependem de sensores funcionando corretamente. Em sistemas de manufatura, por exemplo, foram desenvolvidos muitos métodos engenhosos para orientar corretamente as peças a partir de uma posição inicial desconhecida usando uma sequência de ações completamente sem atividade sensorial.

A revolução da tecnologia da informação significa, entre outras coisas, que o software tornou-se uma parte cada vez maior dos produtos. Softwares são encontrados em produtos que

vão desde torradeiras até naves espaciais. Isto significa que uma grande quantidade de softwares tem sido e estão sendo desenvolvidos.

Desenvolvimento de software não é fácil, mas sim um processo que demanda tempo e dedicação. O rápido crescimento da área também fez com que numerosos projetos de software tenham problemas em termos de falta de funcionalidade, excesso de custos, prazos não cumpridos e de má qualidade. Estes problemas ou desafios foram identificados já na década de 1960, e em 1968 o termo “engenharia de software” foi cunhado com a intenção de criar uma disciplina de engenharia focada no desenvolvimento de sistemas de software intensivo (GUNNING; STEFIK; CHOI, 2019).

Engenharia de software é formalmente definido pelo IEEE como “engenharia de software, a aplicação de uma abordagem sistemática e disciplinada, quantificando o desenvolvimento, operação e manutenção do software (DIRECTORS, 2019).

Quando as propostas de melhoria forem identificadas, é necessário determinar como são, e se houver. Muitas vezes não é possível apenas alterar o software existente sem ter mais informações sobre o efeito real da melhoria proposta. Em outras palavras, é necessário avaliar as propostas antes de tomar qualquer grande mudança. A engenharia de software é um assunto interdisciplinar. Estende-se desde técnico como bancos de dados e sistemas operacionais, através de questões de linguagem, por exemplo, sintaxe e semântica, a questões sociais e psicologia (WEIB, 2001).

No entanto, devemos, ao estudar e fazer investigação em engenharia de software, destinando-se a tratar como uma disciplina científica. Isto implica usando métodos científicos para pesquisa e na tomada de decisões em relação a mudanças na forma como desenvolvemos. Para realizar pesquisa científica em engenharia de software, temos que compreender os métodos que estão disponíveis para nós, suas limitações e quando eles podem ser aplicados. Engenharia de software decorre da comunidade técnica. Assim, ele é natural em olhar para os métodos utilizados na investigação em, por exemplo, design de hardware e codificação teórica, mas com base na natureza da engenharia de software, devemos olhar em outras disciplinas também (RUSSELL; NORVIG; DAVIS, 2010).

Eles foram inicialmente apresentados como engenharia de software conforme visto por (BASILI, 2003).

A introdução para a área é baseada na introdução de um processo para experimentação. Os passos básicos no processo podem ser utilizados para outros tipos de estudos empíricos também.

O foco é, no entanto, em fornecer orientações e suporte para a realização de experimentos em engenharia de software.

Experimentos de engenharia de software são muitas vezes quase experiências, ou seja, experimento em que, por exemplo, não tem sido possível atribuir participantes nas experiências a grupos aleatórios. Quase experimentos são importantes, e eles podem fornecer resultados valiosos. O processo apresentado neste livro é destinado a ambos os experimentos 'verdadeiras' e quase experimentos. O último é particularmente suportado por uma discussão aprofundada das ameaças às experiências (RUSSELL; NORVIG; DAVIS, 2010).

### 3.4 ANÁLISE ESTATÍSTICA

Quando executamos um determinado projeto, seja ele qual for, após a coleta de dados experimentais na fase de operação, ou melhor dizendo, fase de execução, queremos ser capazes de tirar conclusões com base nesses dados. Para ser capaz de tirar conclusões válidas, devemos interpretar os dados da experiência. Interpretação quantitativa pode ser efetuada em três passos(ZOU, 2020).

No primeiro passo, os dados são caracterizados utilizando estatística descritiva, que nos permite visualizar a tendência central, a dispersão, etc. Na fase dois, pontos de dados falsos ou anormais são excluídos, reduzindo, assim, o conjunto de dados a um conjunto de pontos de dados válidos. No terceiro passo, os dados são analisados por testes de hipóteses, onde as hipóteses do experimento são avaliadas estatisticamente, a um dado nível de significância(LEE, J. K. et al., 2020).

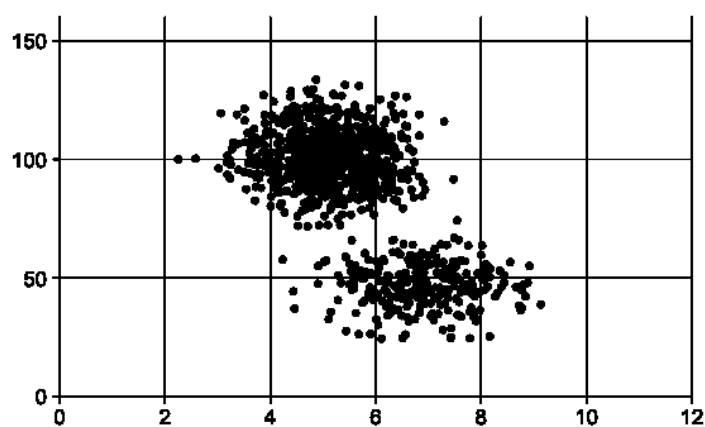
O valor médio é significativo para as escalas de intervalo e proporção. Assim, mediana é calculada classificando-o em amostras ascendente (ou descendente). A mediana pode ser definida como a média aritmética de dois valores médios. Esta última operação requer que a escala tenha pelo menos um intervalo. Se a escala é ordinal, um dos dois valores médios podem ser selecionados pela escolha aleatória, ou a mediana pode ser representado como um par de valores(GRAMI, 2020).

A variância é a média do quadrado da distância entre a média da amostra. Em particular, a variância da amostra é uma estimativa imparcial e consistente da variância da variável estocástica. A variância é significativo para as escalas intervalares e relação.

Se tivermos mais de duas variáveis, podemos aplicar a análise multivariada, incluindo técnicas tais como regressão múltipla, análise do componente principal (PCA), conjunto análise e análise discriminante (GRAMI, 2020).

O gráfico de dispersão é boa opção para avaliar as dependências entre as variáveis, como exemplo a Figura 10. Ao examinar o gráfico de dispersão, é possível visualizar espalhamentos ou concentração dos dados como realmente são, e se há uma tendência de relação linear. Valores atípicos (outliers) podem ser identificados, e a correlação também pode ser observada.

Figura 10: Gráfico de Dispersão



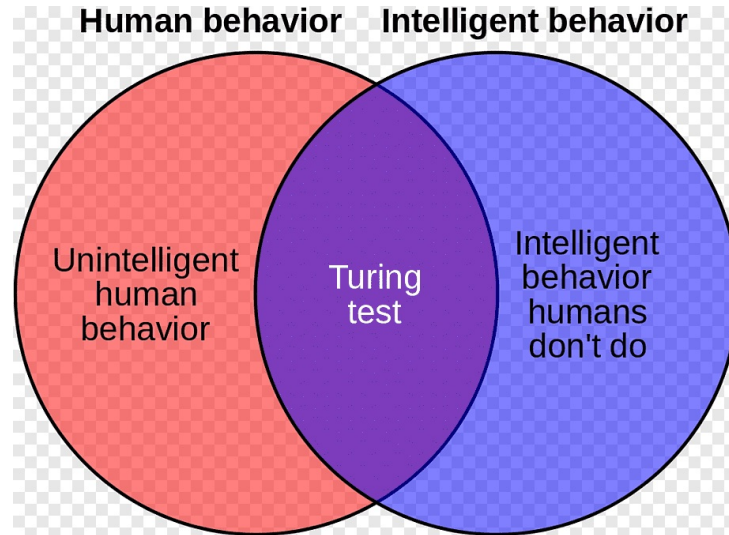
Fonte: autor

Os valores fora das caudas inferior e superior são chamados valores aberrantes, e são mostrados explicitamente no gráfico de caixa. Um histograma pode ser usado para dar uma visão geral da densidade de distribuição das amostras a partir de uma variável. Um histograma é constituído por barras com alturas que representam a frequência (ou a frequência relativa) de um valor ou de um intervalo de valores.

O histograma é, portanto, uma representação gráfica de uma frequência. Uma distribuição de particular interesse é a distribuição normal, uma vez que é um aspecto que deve ser levado em conta na análise dos dados. Assim, uma trama poderia fornecer uma indicação primeiro se os dados se assemelham a uma distribuição normal ou não. Também é possível testar os dados de normalidade.

A Figura 11 demonstra um modelo de histograma criado no software MatLab (THE MATHWORKS, 2017).

Figura 11: Gráfico de Histograma



Fonte: autor

O t-teste é um teste paramétrico utilizado para comparação de duas amostras independentes. O projeto deve ser um fator com dois níveis. O t-teste pode ser executado com base em uma série de pressupostos diferentes, mas uma alternativa frequentemente utilizada é descrito em (BOROKOV, 2019). O F-teste é um teste paramétrico que pode ser utilizado para comparar a variação de duas amostras independentes .

O teste de Wilcoxon é uma alternativa não paramétrica para o teste t emparelhado. Os únicos requisitos são de que é possível determinar qual medida de um par é o maior e que é possível classificar as diferenças. O teste, que se baseia em fileiras, não é descrito em detalhes aqui.

O teste de sinal é, como o teste de Wilcoxon, uma alternativa não paramétrica para o emparelhado t-teste. O teste de sinal pode ser usado em vez do teste de Wilcoxon quando não é possível ou necessário para classificar as diferenças, uma vez que se baseia apenas no sinal da diferença dos valores em cada par. Por exemplo, não é necessário o uso de Wilcoxon quando é possível mostrar um significado com o teste do sinal. Isto é devido que o teste de sinal tem uma potência inferior. Além disso, o teste de sinal é mais fácil de executar (BASILI, 2003).

Análise de variância pode ser usada para analisar as experiências de um número de diferentes desenhos. O nome, análise de variância, é usado porque o método é baseado na observação para a variabilidade total dos dados e a divisória de acordo com a variabilidade de

diferentes componentes. Na sua forma mais simples do teste compara a variabilidade devido ao tratamento e a variabilidade devido a erro aleatório. O teste Kruskal-Wallis analisa uma forma de variância por postos é um teste não-paramétrico alternativo para a análise de um fator de variância paramétrica descrito acima. Este teste pode sempre ser usado em vez da análise de variância paramétrica se não é certo que as premissas da ANOVA são cumpridas, como podemos ver em (CORTINA; NOURI, 1999). O teste, que é baseado em fileiras, mas não é descrita em detalhes aqui, melhor descrito em (HENDERSON, 2017).

O teste do qui-2 é baseado em certas suposições, que são susceptíveis de ser preenchido os valores esperados e não são muito pequenos. Uma regra de ouro é que se o número de graus de liberdade ( $f$ ) é igual a 1, o teste Chi-2 não deve ser usado se nenhuma das frequências esperadas são menos de 5. Se  $f > 1$  o teste Chi-2 deve não podem ser utilizados, pois mais de 20 por cento das frequências esperadas são menos do que 5 ou qualquer de eles é inferior a 1 (SAMUSENKO, 2011) e (RUTHERFORD, 2011).

Deve-se observar que, por vezes, o teste é usado embora as frequências esperadas não preenchidas. Nestes casos, este é um risco calculado. Uma forma de obter frequências maior esperadas é combinar categorias relacionadas a novas categorias. No entanto, as novas categorias devem ser significativas.

Cada modelo estatístico está contando com específica suposição a respeito, por exemplo, distribuição, independência e escalas. Se as premissas são invalidadas pelos dados do conjunto, os resultados do teste de hipóteses são inválidos. Por isso, é fundamental para verificar se todos os pressupostos são preenchidos.

A verificação da adequação do modelo é feita de acordo com os pressupostos. Abaixo descrevemos três casos: Se um teste de normalidade assume que os dados são normalmente distribuídos, um teste de Qui-2 podem ser feito para avaliar até que ponto a suposição é satisfeita.

Se o teste assume que os dados são de uma amostra a partir de várias variáveis estocásticas, é necessário verificar se não há nenhuma correlação entre os conjuntos de amostras. Isto pode ser verificado com diagramas de dispersão e através do cálculo de correlação de coeficientes como discutido no início desta seção (YANG; BASU; ZHANG, 2019).

Em muitos modelos estatísticos, resíduos não é um termo que representa o erro estatístico. Muitas vezes, é assumido que o teste de hipóteses de resíduos são normalmente distribuídos. Uma maneira comum de verificar esta propriedade é para traçar os resíduos em um gráfico de dispersão e ver que não existe uma tendência nos dados (a distribuição parece aleatória). Quando os dados do experimento foi analisado e interpretado, precisamos desenhar conclusões sobre

o resultado do experimento. Se as hipóteses são rejeitadas podemos tirar conclusões sobre a influência das variáveis independentes sobre as variáveis dependentes, uma vez que a experiência é válida (RONCHETTI, 2020).

Se, por outro lado, o experimento não pode rejeitar a hipótese nula, não podemos tirar quaisquer conclusões sobre as variáveis independentes na influência sobre as variáveis dependentes. A única coisa que têm demonstrado, neste caso, é que não há nenhuma estatística diferença significativa entre os tratamentos (WAAL; DELDEN; SCHOLTUS, 2019).

Uma vez que se decidiu quais dados serão incluídos na análise dos dados, é hora de tomar um olhar para a análise estatística. A análise estatística é sempre um desafio, e há muitas opiniões diferentes sobre o uso de diferentes métodos estatísticos e quando a utilização de métodos paramétricos e não paramétricos.

Iremos utilizar a análise estatística em nosso trabalho de forma que poderemos conciliar modelos estatísticos conservadores para efeito de comparação ao nosso resultado. Ou seja, a análise estatística é uma ferramenta que pode nos direcionar e possivelmente apurar se nosso modelo está colaborando como proposto, chegando ao objetivo de nosso trabalho. Já é conhecido que os modelos estatísticos foram necessários para a segregação de dados multivariáveis, e sem eles a resposta a diversas questões, mesmo na área médica não seriam obtidos (ASHOFTCH; BRAVO, 2021).

### 3.5 EQUIPAMENTO DESENVOLVIDO DURANTE O ESTUDO

Durante a produção deste doutorado, houve o desenvolvimento de um equipamento para análise pulmonar, que expressa os valores de preditos e resultados como exemplo o FVC e FEV1, semelhantes aos disponíveis no mercado atualmente, sendo que este equipamento levou em consideração a norma da ABNT 60-601 (ABNT, 2010), norma esta relativa a construção e homologação de equipamentos médicos.

O equipamento possui um software instalável no computador, que faz a interpretação dos sinais obtidos do espirometro, de forma a gerar gráficos para que o exame possa ser realizado, interpretado e diagnosticado, auxiliando o médico na tarefa de laudar o exame.

No cadastro de paciente, é possível fazer a seleção de gênero para que os preditos sejam corretamente calculados, de acordo com a figura 12.

Figura 12: Tela de Cadastro

**Cadastro de Pacientes**

Nome: Rodrigo Sobrenome: Piva

Código de Identificação: 6 Data de Nascimento: 07/03/1977 42 Anos

Altura: 180 Centímetros Sexo: ☒ Masculino ☐ Feminino

Peso: 97 Kilos Histórico de Tabagismo: 0 Cigarros / Dia

IMC: 29.94 0 Anos Fumante

Origem Étnica: Caucásiano 0 Maços / Ano

Predito: Pereira (2008) Ocupação: 0

Correção Étnica: 100 Email: 0

Sat. de O<sub>2</sub>: 98 % Médico Solicitante: 0

Empresa: FEI Convenio / Empresa: 0

Endereço: AV PRES HUMBERTO CASTELO BRANCO BD Utilizado: 0

Comentários: 0

Buscar Preditos Cancelar

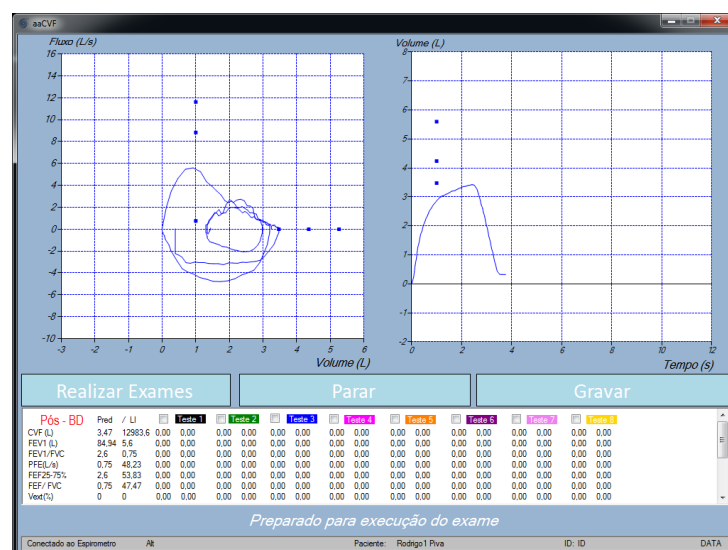
**Cadastrar**

Fonte:Autor

Cabe ressaltar que o software já faz a seleção de gênero, além da idade, altura e peso automaticamente para cálculo dos preditos, como pode ser visto em (PEREIRA, C. A. C., 1996). Os preditos são calculados de forma que o médico tenha mais uma variável de decisão tornando o laudo mais confiável.

A figura 13 mostra o caráter dos gráficos de fluxo em função do volume bem como o volume em função do tempo expressos na tela.

Figura 13: Software do Espirometro Durante o Período do Doutorado, Tela de exame de Capacidade Vital Forçada



Fonte:Autor

Na realidade estes gráficos demonstrados acima, são os que o médico pneumologista, ou o técnico de espirometria visualiza para gerar o diagnóstico, e através da sua experiência, análise visual e possuindo os valores de preditos irá gerar o diagnóstico do exame de espirometria do paciente. Portanto, este doutorado tem como objetivo auxiliar o médico a realizar o diagnóstico mais confiável possível, fornecendo um novo marcador que pode orientá-lo a fim que possa realizar um laudo com mais segurança.

## 4 METODOLOGIA PROPOSTA

Este capítulo demonstra os métodos de como os dados dos exames de função pulmonar foram extraídos, processados, classificados, para posterior apresentação ao usuário que realiza o exame que, de posse dessa informação, possa realizar um exame com maior qualidade e precisão.

### 4.1 DADOS OBTIDOS

A base para a criação dos experimentos é um conjunto de dados provenientes de exames de função pulmonar e, para este estudo, foram utilizados 360 exames de espirometria, separados em masculinos e femininos. Cada exame possui entre 4 e 6 exames diferentes, chamadas de manobras expiratórias. Esses dados foram gentilmente cedidos pelo IAMSPE(Instituto de Assistencia Medica ao Servidor Publico Estadual de São Paulo) e pela Dra. Silvia Carla Sousa Rodrigues, CRM 100632, conforme consta no apendice A a permissão do uso desses dados.

Inicialmente foi utilizada uma parte da base de dados possibilitando efetuar os primeiros levantamentos iniciais e assim validar a metodologia. Um conjunto de dados genérico foi selecionado, com aproximadamente o mesmo tamanho de amostra entre os dois gêneros, de forma que a sua utilização fosse distribuída de maneira uniforme, resultando num total de 1.360 exames do sexo masculino e 1.420 exames do sexo feminino para a criação do modelo e validação dos testes.

### 4.2 CRIAÇÃO DO BANCO DE DADOS

O banco de dados foi criado tendo em vista a coleta de informações e armazenagem dos dados para posterior pós processamento. Ele foi criado em ambiente SQL, na versão "Compact Edition"visto que o acesso a informação é do tipo acesso reduzido, sendo possível a reutilização em qualquer plataforma de programação sem maiores dificuldades.

Como premissa inicial, é necessário que o software que realiza os exames de espirometria alimente esse banco de dados com os dados necessários e armazene essa informações em tabelas adequadas para que possa ser utilizado pelo modelo projetado. Este modelo irá servir de base para a aplicação de inteligencia artificial, de forma a permitir a tomada de decisão em agrupar o resultado da análise nos três grupos distintos.

### 4.3 INTERPRETAÇÕES

É interessante apresentar neste momento que possíveis classificações dos exames, utilizados na atualidade de acordo com a bibliografia estudada, existem, sendo elas:

- a) Previsões FEV1 (Fluxo Expiratório Forçado no primeiro segundo) / FVC (Capacidade vital forçada) e FEV3 (fluxo expiratório forçado durante os tres primeiros segundos) / FVC são necessárias.
- b) Os intervalos de confiança FEV1 / FVC e FEV3 / FVC são necessários.
- c) Os resultados de FEV1 / FVC e FEV3 / FVC não estão disponíveis. (Não foi possível obter tais resultados)
- d) Nenhuma obstrução das vias aéreas (VEF1 / CVF e VEF3 / CVF dentro dos limites normais).
- e) Obstrução das vias aéreas sugerida (VEF3 / CVF abaixo dos limites normais).
- f) Obstrução leve das vias aéreas (VEF1 / CVF abaixo dos limites normais).
- g) Obstrução moderada das vias aéreas (VEF1 / CVF abaixo dos limites normais).
- h) Obstrução severa das vias aéreas (VEF1 / CVF abaixo dos limites normais).
- i) A obstrução pode estar subestimada (tempo expiratório abaixo dos limites normais).
- j) Obstrução das vias aéreas superiores sugerida (VEF.5 / VEF1 abaixo dos limites normais).
- k) Ar aprisionamento pode estar presente (FIVC / FVC acima dos limites normais).
- l) Curto tempo de exalação pode esconder obstrução leve.
- m) Restrição torácica também pode estar presente (FVC e FIVC abaixo dos limites normais). Sugira volumes pulmonares e DLCO.
- n) Não melhorou após broncodilatador (CVF, VEF.5, VEF1 ou FEF25-75)
- o) O resultado da CVF não é utilizado na interpretação pós-broncodilatador. Como o tempo pós / pré-expiratório não está dentro dos limites, a mudança na CVF pode ser devida unicamente à mudança no tempo expiratório e não ao fluxo aumentado.
- p) O resultado do FEF25-75 não é utilizado na interpretação pós-broncodilatador porque o tempo pós / pré-expiratório e / ou a CVF não estão dentro dos limites. Uma mudança no tempo expiratório ou na CVF pode alterar o FEF25-75 na ausência de qualquer alteração no próprio fluxo.
- q) Os resultados pré e pós-CVF, VEF.5, VEF1 e FEF25-75 não estão disponíveis para interpretação pós-broncodilatador.

r) Espirometria normal.

Estas interpretações são englobadas em nosso estudo em três classes: Obstrutiva, Restritiva e Normal. É obtido assim, o reconhecimento de padrões das diversas interpretações nesses 3 grupos distintos, com a aplicação das wavelets e do uso de inteligência artificial, obtendo o resultado esperado o qual não foi encontrado em nenhum dos artigos estudados durante nossa pesquisa bibliográfica.

Assim, o banco de dados do experimento conta com os parâmetros necessários para a realização da interpretação atualmente utilizada e adicionando o processamento de sinais de forma a gerar um resultado que se torne uma variável adicional para a determinação do resultado final.

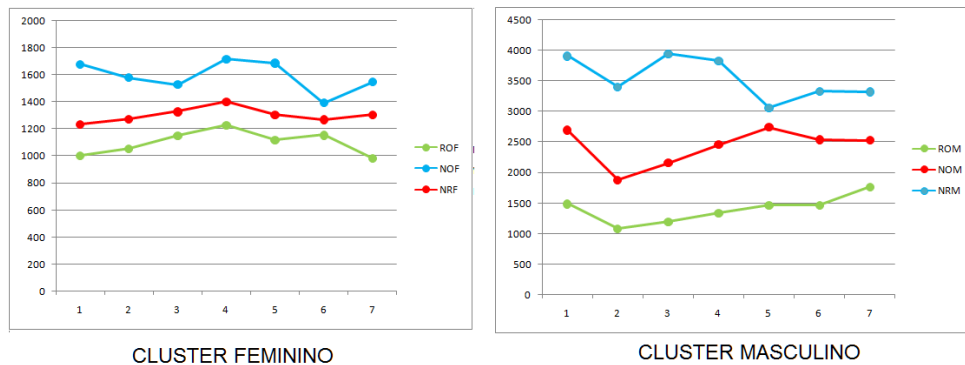
#### 4.4 SELEÇÃO DA FAMÍLIA DE WAVELETS

Conforme visto em (PIVA, Rodrigo G. G., 2017), é possível fazer a extração dos dados da leitura desse tipo de exame, e possibilitar a aplicação de uma função temporal, para analisar os resultados dos exames masculinos e femininos em normal, obstrutivo e restritivo. Através do uso do modelo proposto, é possível a extração desses dados diretamente pelo banco de dados, visto que o software utiliza um banco de dados "open access" que permite qualquer usuário utilizar a metodologia proposta neste trabalho.

Conforme as premissas descritas no capítulo 3.3, todo o desenvolvimento deste software seguiu as normas de engenharia de software bem como foi possível fazer a segregação de pacientes em idade, gênero e altura, valores considerados básicos para a obtenção dos preditos para a realização do exame.

A figura 14 apresenta a separação dos dados separados pelo gênero. O Eixo vertical demonstra os valores de covariância, e o eixo horizontal demonstra a amostra analisada anteriormente no estudo. A faixa ROM Demonstra os indivíduos classificados como restritivos, a Sigla NOM, obstrutivos e a Sigla NRM os indivíduos normais da amostra analisada.

Figura 14: Clustering dos dados femininos e masculinos



Fonte:Autor

Assim, é possível obter informações que auxiliam na classificação dos exames que são realizadas atualmente, conforme mostrado na figura 15, e que necessitam de melhorias visto a tecnologia disponível atualmente, como filtros e melhorias digitais.

Figura 15: Modelo de análise interpretativa atual

| Tabela 1- DPOC X espirometria |                                                              |                                                 |
|-------------------------------|--------------------------------------------------------------|-------------------------------------------------|
| Parâmetro funcional           | Achado típico                                                | Observação                                      |
| VEF1 /CVF                     | < 0,70                                                       | Caracteriza a presença de obstrução             |
| VEF1                          | < 80% previsto                                               | Avalia a intensidade da doença                  |
| CVF                           | < 70 % previsto                                              | Provável hiperinsufilação                       |
| Prova broncodilatadora        | Negativa: VEF1 pós-bd com variação <12% do pré-bd e < 200 ml | Quando positiva pensar na possibilidade de asma |
| VEF1 /CVF                     | ≥ 0,70 com CVF reduzida                                      | Interrogar doença restritiva                    |

Fonte:Autor

Conforme já mencionado, de acordo com os resultados obtidos em (PIVA, R. G. G., 2017), iremos utilizar a família de Wavelets de Meyer, pois foi a que melhor representa o sinal proveniente do exame de espirometria, gerando extremos bem definidos, enquanto que nas outras famílias de wavelets a representação não proporcionava a mesma aderência no sinal original.

#### 4.5 EXPERIMENTO DO Q-LEARN NA ATRIBUIÇÃO DO DIAGNÓSTICO

O Q-Learn, modelo de inteligência artificial escolhido e utilizado em nosso estudo, permitiu que pudessemos percorrer a área gráfica total da análise interpretativa, pontualmente, de forma a estabelecermos de forma autônoma com que o software interprete o resultado.

De uma forma abrangente, estabelecemos as seguintes premissas:

- O valor obtido da covariância através aplicação da família selecionada estabeleceu o parâmetro do eixo Y.

- b) O valor obtido da amostra no universo de exames estudados, estabeleceu o valor do eixo X daquela amostra.
- c) A Regressão de Ridge criou as retas de aproximação, de forma a estabelecer regiões onde um novo exame quando for inserido se localizar poderá fornecer o diagnóstico daquele paciente em questão.
- d) A matriz do Q-Learn possui 5 dimensões, fazendo que o valor da covariância encontrado também se tornasse um peso relacional, atribuindo esse valor em um nível de profundidade sempre diferente dos demais.
- e) O grafico interpretativo possui uma dimensão de 1024x768 pontos, de forma a expressar na tela a real posição do exame estudado.

Com essas premissas definidas, criou-se o modelo de implementação do Q-Learn, conforma a Figura 16, que demonstra como é a criação do código em ambiente computacional.

Figura 16: Código da Aplicação do Q-Learn

```

81 for vezes in range(10):
82     for linha in range(1024):
83         for coluna in range(768):
84             pos = [linha, coluna]
85             for acaoIdx in range(4):
86
87                 acao = ['B', 'C', 'D', 'E'][acaoIdx]
88
89                 novaPos, recomp = mover(pos, acao)
90                 Q[pos[0], pos[1], acaoIdx] = Q[pos[0], pos[1], acaoIdx] + alpha * (recomp +
91                                     gamma*np.max( Q[novaPos[0], novaPos[1]] ) - Q[pos[0], pos[1], acaoIdx] )
92

```

Fonte: Autor

A equação do Q-Learn, vista em (WATKINS C. J. C. H.; DAYAN, 1992) foi um dos maiores avanços em Inteligencia Artificial. Ela determina uma função Q, que é um par de 'estado e ação' e considera as transições ocorridas a cada iteração, o que a faz ser diferente de funções temporais. Ela é definida conforme a equação vista na Figura 17 abaixo:

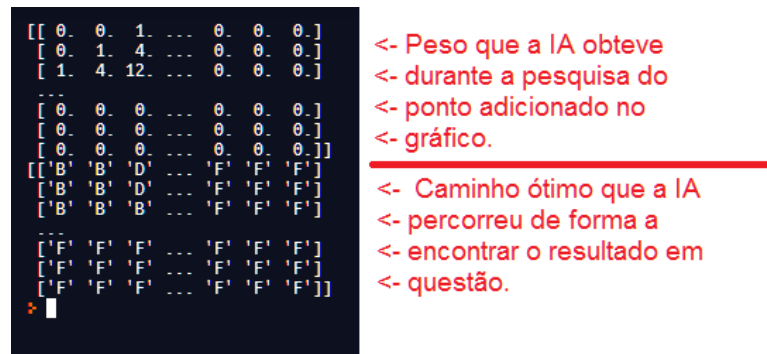
Figura 17: Equação do Q-Learn

$$\underbrace{NewQ(s, a)}_{\text{New Q value for that state and that action}} = \underbrace{Q(s, a)}_{\text{Current Q value}} + \underbrace{\alpha}_{\text{Learning Rate}} \left[ \underbrace{R(s, a)}_{\text{Reward for taking that action at that state}} + \underbrace{\gamma}_{\text{Discount rate}} \underbrace{\max_{a'} Q'(s', a')}_{\text{Maximum expected future reward given the new s' and all possible actions at that new state}} - Q(s, a) \right]$$

Fonte: (RUSSELL; NORVIG; DAVIS, 2010)

Dessa forma, foi possível obter a matriz de estados, que percorre o ambiente gráfico totalmente, de forma a determinar em qual ponto o valor inserido para aquele exame em estudo se encontra. Parte da matriz esta demonstrada na Figura 18:

Figura 18: Matriz de Estados Obtida



Fonte: Autor

É interessante informar também, que o tempo de execução para a verredura foi em torno de 2 minutos, sendo esse o tempo para a primeira execução. Com o algoritmo proposto, o tempo para inserção e detecção de uma única amostra fica abaixo de 3 segundos, sendo perfeitamente possível a utilização desse tipo de modelo em rotinas de trabalho.

#### 4.6 TESTE DO CLASSIFICADOR

Para realizar o teste do classificador, após o término da construção do software, realizamos testes com novos pacientes obtendo resultados eficientes com o menor número possível de variáveis tornando o sistema executar o processamento no menor tempo possível.

#### 4.7 PROCESSAMENTO DE DADOS

O Processamento de dados foi realizado da seguinte maneira:

- a) Captura dos dados. A captura dos dados consiste na leitura dos dados dos exames, através de um "fileStream" comum. Os dados não são criptografados, e possuem um canal de exportação para que sejam lidos por outros sistemas, dessa forma, não é necessário nenhuma ferramenta específica para essa leitura. Possuem um cabeçalho padrão, com os dados de cada exame, um breve cadastro do exame, a

marcação do início de cada manobra do exame, a sua duração e os dados lidos do sinal proveniente do sensor de fluxo, armazenados sequencialmente.

- b) Janelamento. O janelamento consiste em verificar qual o tamanho do dado, pois dentro de cada exame, existe uma variável que possui essa informação. No caso deste estudo, os dados continham entre 8000 e 12000 valores. Essa diferença se dá pelo fato de que o paciente pode demorar a estabilizar a respiração, antes do início da manobra de expiração forçada. Como todos os pacientes realizam pelo menos 1 manobra de inspiração e expiração (considera-se este ciclo como 1 respiração), então janelou-se os dados de forma a iniciar a aquisição dos dados 1 ciclo de respiração antes da manobra de expiração forçada. Após essa manobra, o comportamento do sinal permaneceu semelhante em todos os exames, devido aos padrões de realização do exame de espirometria.
- c) Exportação do sinal do exame. Com todas as amostras janeladas, é possível exportar esse sinal para um banco de dados intermediário, criando uma forma prática para manipulação desses valores. Assim, foi criado em um arquivo de texto, para que pudesse ser utilizado nas próximas etapas.
- d) Exportação dos valores encontrados no exame. Após a exportação do sinal do exame, é feita toda a retirada de dados que não serão utilizados, como dados de cadastro, data do exame, variáveis que podem interferir na leitura dos dados realmente necessários, que são os valores unitários de cada segundo que o sensor de fluxo leu durante o exercício da manobra.
- e) Aplicação da função wavelet nos valores de entrada. Com a posse desses dados filtrados, é aplicado via software da transformada de wavelet de Meyer, que irá gerar valores de coeficientes em novos vetores, criando assim uma tabela que contém os coeficientes de todos os exames dessa amostra.
- f) Obtenção do valor da covariância do sinal e aplicação da regressão de ridge. Manipulando os dados processados resultantes da aplicação de wavelets, obteve-se tabelas de covariância, onde aplicando o modelo de clusterização conhecido como K-means, estabeleceu-se regiões onde esses valores se agrupavam. Neste momento, também aplicou-se a regressão de ridge, entre os grupos restritivos e obstrutivos em relação ao grupo normal, e depois mais uma regressão de relação entre os grupos obstrutivos e restritivos.

- g) Inserção no banco de dados. Após esse processamento, armazenou-se essa informação num banco de dados, que pode ser lido, interpretado e utilizado a qualquer momento que o software de interpretação for chamado pelo software principal de realização de exames.
- h) Avaliação do um novo valor com o banco treinado existente. Após a criação desse modelo, foi efetuado testes, onde uma vez inserido um novo valor, registra-se como ele se comporta frente aos valores já existentes no banco de dados, demonstrado esse comportamento no capítulo 5.
- i) Aplicação do modelo proposto. Com base dessa informação, é aplicado o modelo estatístico, baseado nas informações estudadas na seção 3.4, com o intuito de posicionar a amostra na região. Para isso, essa nova amostra passa pela filtragem novamente e seus componentes são verificados de acordo com os dados já processados existente no banco de dados.
- j) Construção das retas de separação . Como podemos ver na seção 3.2 que as regressões lineares são modelos dinâmicos, toda a vez que uma nova amostra é inserida, o software se encarrega de aplicar a regressão de ridge novamente, ou seja, as retas que separam os três grupos se tornam totalmente dependentes dessa nova amostra.
- k) Calculos dos valores médios e índices da taxa de incerteza. Dessa forma, verifica-se a distância, através da aplicação do K-means direcionado para cálculo de distâncias entre centróides, a distância entre o centróide e a localização gráfica da nova amostra, dessa forma, a aplicação do Q-Learn se encarrega de procurar, frente a toda amostra utilizada, qual é essa distância, e retorna esse valor para que seja realizado o cálculo. Como o Q-Learn inicia a varredura em regiões não classificadas, no momento que ele encontra valores conhecidos, atribui esse valor a seu estado anterior e no momento que ele encontra a amostra a ser analisada, consegue determinar qual a distancia e região onde esse valor se encontra.
- l) Exibição gráfica do resultado. Após todo o processamento descrito, é mostrado graficamente, a localização do ponto, a região que ele se encontra, a distância em relação ao centróide da região que se encontra e qual a probabilidade estatística dele estar contido nessa região, através da verificação de distâncias entre o valor analisado e as retas de separação obtidas pela regressão de ridge.

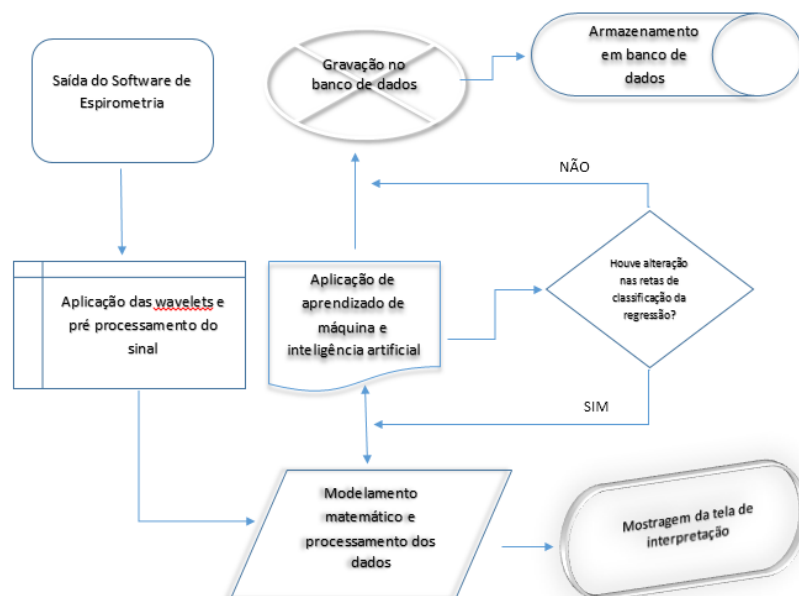
Em todas essas etapas documentadas é possível realizar medições intermediárias, de forma a garantir uma perfeita utilização da engenharia de software nesse caso conforme apresentado em nosso capítulo 3.

#### 4.8 PROCEDIMENTO EXPERIMENTAL

De acordo com a nossa pesquisa bibliográfica, o protocolo de espirometria tem se utilizado somente de alguns valores encontrados no exame, como exemplo o pico expiratório forçado (PFE), ou o valor de fluxo máximo lido pelo equipamento, e a capacidade vital forçada (CVF), que é o volume em litros máximo, lido e calculado pelo equipamento, assim como outros comentados anteriormente. A nossa proposta será de adicionar um novo biomarcador adicional podendo ser inserido nos diversos softwares existentes no mercado.

O modelo criado esta apresentado na figura 19, onde contempla a exportação dos dados do software, a utilização do modelo de wavelet, a construção de um sistema de inteligência artificial para auxiliar o profissional da saúde a laudar o exame, dessa forma colaborando para um diagnóstico mais confiável e com um maior nível de precisão. Os eixos dessa figura, na abscissa vertical representa o paciente analisado, e na abscissa horizontal, o valor de sua covariância.

Figura 19: Fluxograma de Construção do Software

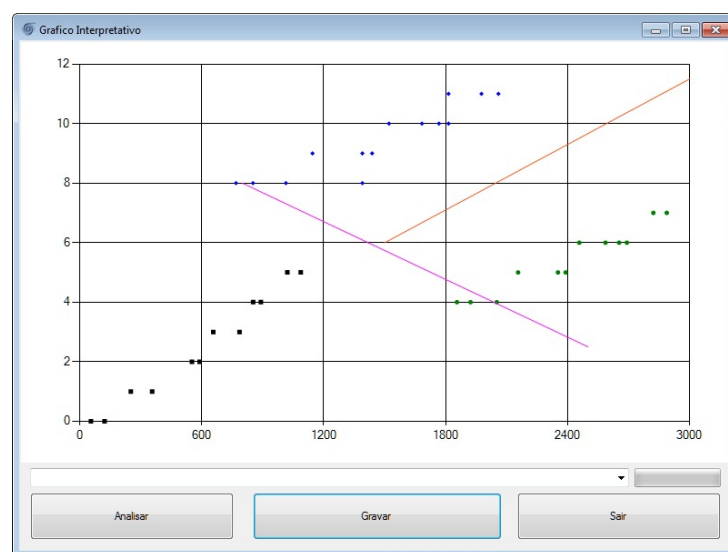


Fonte:Autor

Para isso, é necessário a utilização de um sistema que podemos demonstrar em blocos, onde as etapas conforme o capítulo 3 mostra, utilizando parâmetros encontrados na engenharia de software, de forma a construir o software conforme a figura 20, que apresenta um gráfico expressando duas retas, onde a reta da cor lilás é a regressão linear dos pacientes normais daqueles que possuem alguma patologia, e a reta em vermelho representa a regressão linear dos pacientes com patologias restritivas e obstrutivas.

As linhas obtidas neste gráfico, é o resultado da regressão de Ridge, através da aplicação da inteligência artificial que, em conjunto com essa regressão encontra em modelos dinâmicos de banco de dados informações de valor médio para a apresentação do modelo. Este modelo será atualizado toda a vez que uma nova informação for inserida no banco de dados, e se uma nova interpretação for realizada, é dada a opção ao médico que irá laudar o exame, verificar como é o comportamento daqueles pacientes em relação àquela região populacional.

Figura 20: Etapa de visualização do exame analisado



Fonte: Autor

O conjunto de variáveis que são possíveis extrair dos exames de espirometria, se utilizados conjuntamente, possibilita ter uma grande chance de sucesso na realização de um diagnóstico real, e com a metodologia proposta, é possível chegar a um resultado mais confiável se houver dúvida, sejam elas ocasionados pelo paciente não ter colaborado corretamente na realização do exame, sejam valores que estejam fora dos valores previstos.

Partindo desse conjunto de requisitos, foi desenvolvido um processo de software automatizado para gerar todas as etapas necessárias para a construção do modelo, cujo resultados são apresentados no próximo capítulo.

## 5 RESULTADOS

### 5.1 EQUIPAMENTOS EXISTENTES E SUAS INTERPRETAÇÕES

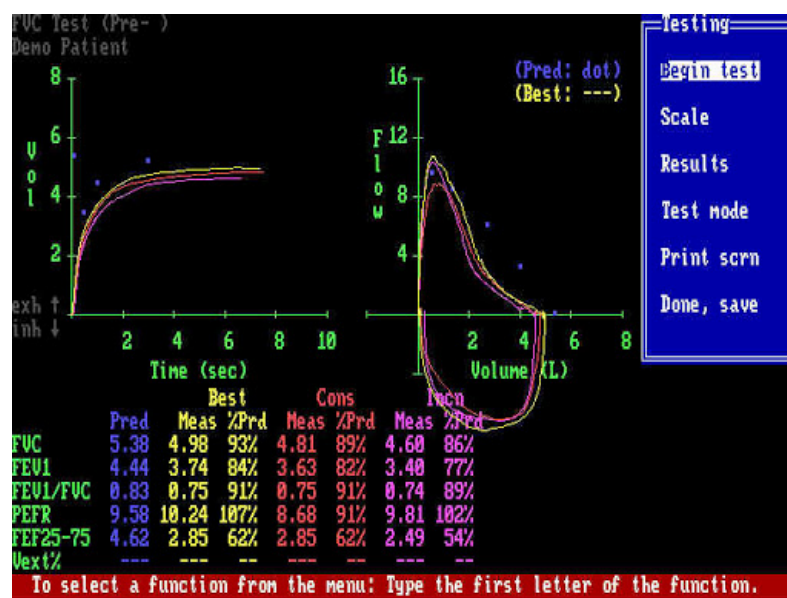
Inicialmente é interessante demonstrar como os equipamentos existentes no mercado tratam a interpretação e diagnóstico dos exames de pneumologia.

#### 5.1.1 EQUIPAMENTO MULTISPIRO - CREATIVE BIOMEDICS

Este equipamento foi desenvolvido a mais de 20 anos, sendo sua produção encerrada a aproximadamente 10 anos. Ele possuía diversos parâmetros para análise do exame, mas não possuía um sistema de interpretação. Posteriormente foi adicionado um padrão de interpretação rudimentar, não tendo uma análise computacional do exame.

A tela do software é demonstrada na Figura 21, onde denota-se a carência de recursos computacionais. Este equipamento ainda é encontrado no mercado, mas vem sendo substituído por equipamentos mais novos e com mais recursos.

Figura 21: Tela do Software do Relatório do Equipamento MultiSpiro

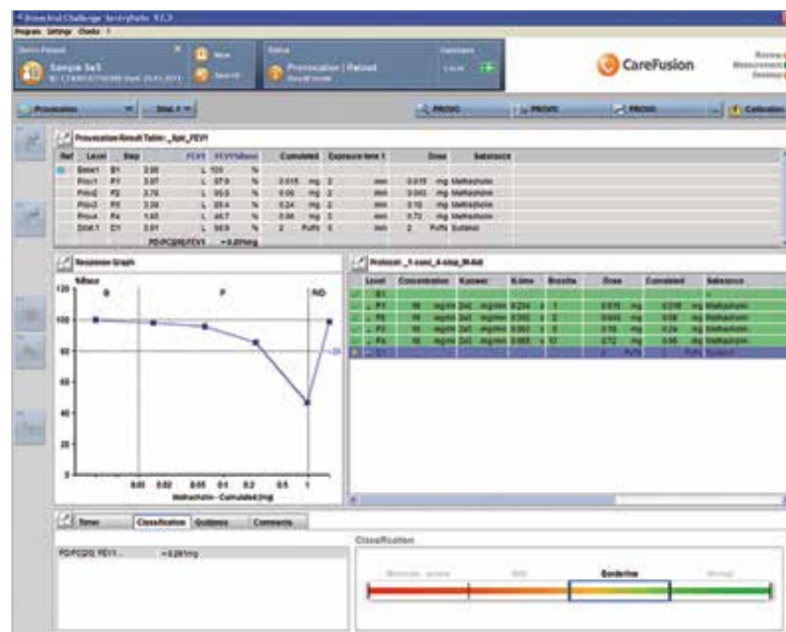


Fonte: Autor

### 5.1.2 EQUIPAMENTO VYNTUS - SENTRY SUITE

Este equipamento é um equipamento de pouco tempo no mercado brasileiro, visto os outros que estão a 20 anos ou mais no mercado. A tela dos exames e relatório são demonstrados na Figura 22. Este equipamento não possui os preditos nacionais, deixando a desejar no quesito confiabilidade de diagnóstico, lembrando que os preditos estabelecidos nos exames são regionais, portanto, os valores encontrados como por exemplo, Capacidade Vital Forçada ou FVC, o padrão brasileiro difere do padrão americano. Ele possui o mesmo modelo de interpretação rudimentar que o citado acima. Mas possui uma confiabilidade maior pois os valores lidos são mais precisos (em torno de 2 por cento de taxa de erro de leitura).

Figura 22: Tela do Software do Relatório do Equipamento Vyntus



Fonte: Autor

### 5.1.3 EQUIPAMENTO KOKO - NSPIRE

Por mais que esse equipamento também está no mercado aproximadamente a 20 anos, vem sofrendo constantes atualizações ao longo desse tempo. Já houveram 3 atualizações importantes de hardware, e o software do equipamento sofreu diversos upgrades, fazendo com que ele se torne o padrão-ouro do mercado nacional e em alguns outros países. É um espirometro de fluxo como os anteriores, mas possui um sistema de interpretação mais inteligente, levando em consideração diversos estudos como o do (PEREIRA, C. A., 2002), mas leva em consideração estudos com valores pontuais, não existindo nele um modelo matemático que analise a curva de espirometria em sua totalidade. Também não existe filtros de aperfeiçoamento da curva, como o filtro da Kalman, também utilizado como visto em (WELLS, 1996), de forma a superar descontinuidades no sinal amostrado.

Figura 23: Tela do Software do Relatório do Equipamento Koko

**Pulmonary Function Report**

Pre vs. Post Report

| Patient Information                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |              |                                      |                 |      |      |      |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|--------------------------------------|-----------------|------|------|------|
| Name:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | ID:          | Birthdate:                           |                 |      |      |      |
| Height at test (in):                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Sex:         | Smoking history (pk-yrs): 5          |                 |      |      |      |
| Weight at test (lb):                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Age at test: | Predicted set: Pereira 2008 (Brazil) |                 |      |      |      |
| Comments:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |              |                                      |                 |      |      |      |
| Diagnosis:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |              |                                      | Interpreted by: |      |      |      |
| Interpretation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |              |                                      |                 |      |      |      |
| <p>NORMAL SPIROMETRIC VALUES indicate the absence of any significant degree of obstructive pulmonary impairment and/or restrictive ventilatory defect. Bronchodilator therapy was administered followed by repeat spirometric testing. Post-bronchodilator testing failed to demonstrate a significant change in FVC, FEV1, or FEF 25-75. This indicates that this patient may not benefit from continued bronchodilator therapy. This interpretation is valid only upon physician review and signature.</p> |              |                                      |                 |      |      |      |
| <div style="display: flex; justify-content: space-between;"> <div> <p>Site:</p> <p>Physician:</p> <p>Technician: Edmilson</p> </div> <div> <p>Effort protocol: ATS/ERS 2005</p> <p>Bronchodilator:</p> <p>Estimated Lung Age: 39</p> </div> <div> <p>Test date/time: 17/07/18 08:35:30</p> <p>Pre-BD Number of efforts performed: 3</p> <p>Post-BD Number of efforts performed: 3</p> </div> </div>                                                                                                          |              |                                      |                 |      |      |      |
| Results                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |              |                                      |                 |      |      |      |
| Result                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Pred         | Pre                                  | %Prd            | Post | %Prd | %Chg |
| FVC (L)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 3,52         | 3,43                                 | 97%             | 3,42 | 97%  | 0%   |
| FEV1 (L)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 2,90         | 2,95                                 | 102%            | 2,99 | 103% | 1%   |
| FEV1/FVC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 0,82         | 0,86                                 | 104%            | 0,87 | 106% | 2%   |
| FEF25-75% (L/s)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 2,83         | 3,87                                 | 136%            | 4,22 | 149% | 9%   |
| PEFR (L/s)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 7,46         | 7,32                                 | 98%             | 7,26 | 97%  | -1%  |
| Vext %                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | ---          | 1,70                                 | ---             | 2,66 | ---  | 56%  |

Fonte: Autor

O sistema de interpretação deste equipamento está de certa forma sendo abandonado, pois durante este tempo no mercado, diversas análises não foram bem sucedidas, tendenciando o médico a diagnosticar por vezes resultados equivocados. Alguns estudos tem sido realizados, como por exemplo o de (WANG et al., 2016), onde a análise dos valores de FEV6/FVC tem sido um bom parâmetro para a análise e diagnóstico de pacientes que realizam a espirometria. Mas este equipamento não possui essa funcionalidade, ficando um pouco atrás na vanguarda tecnológica em exames de função pulmonar.

#### 5.1.4 RESULTADOS DO EQUIPAMENTO DESENVOLVIDO

O equipamento desenvolvido durante esse estudo foi o que permitiu realizar diversas implementações que só seriam possíveis se este equipamento estivesse operacional. O motivo é que a análise de interpretações foi muito proveitosa em ser implantada em equipamentos que pudessem interagir com ela, mesmo sendo uma parte dedicada do software em questão. Todos os outros equipamentos, o software tem seu código fonte fechado, e não é possível integrar um modelo de interpretação da forma que foi realizada no equipamento desenvolvido.

De uma certa forma, o equipamento realiza exames de forma satisfatória. Assim a Figura 24 mostra o equipamento finalizado, conectado a um notebook. Este equipamento foi alvo de desenvolvimento por 5 anos, chegando a fase final de prototipagem e sendo possível a sua produção.

Figura 24: Imagem do Equipamento Desenvolvido

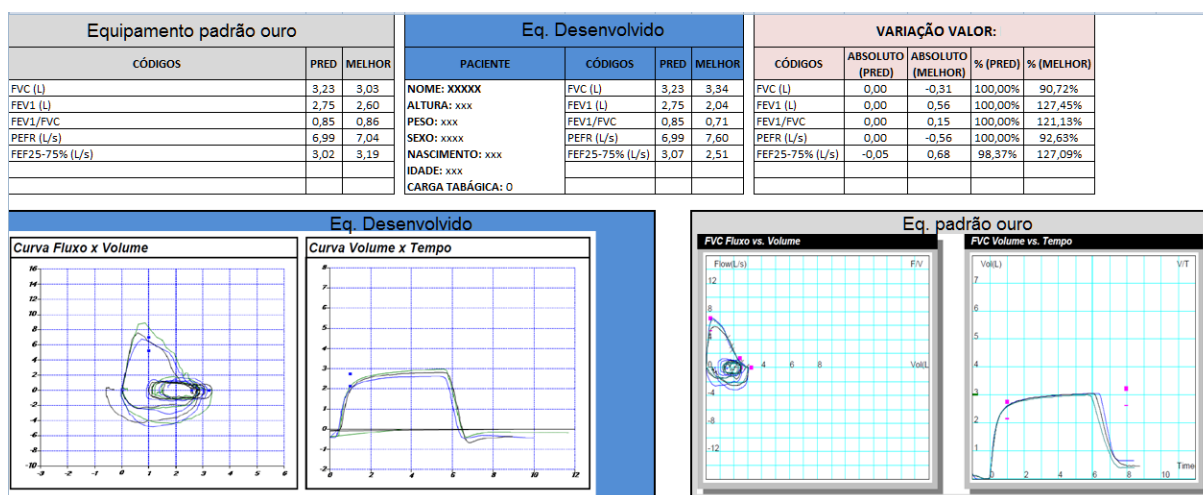


Fonte: Autor

Foram realizados vários testes de prototipagem, e testes de qualidade do sinal amostrado. O equipamento usa um sensor de pressão diferencial e um conversor analógico-digital de 12 bits para realizar a conversão, além dos modelos de interpretação e criação de relatórios de forma aberta, para que possam ser utilizados com outras bases de dados, base essa que também é feita em ambiente do tipo SQL, proporcionando uma maneira inovadora para tratamento de dados, pois os outros equipamentos citados possuem banco de dados proprietários, e não relacionais, trazendo problemas de segurança e confiabilidade no armazenamento dos dados.

De uma forma abrangente, o exame executado neste equipamento produz resultados satisfatórios, como os mostrados pela Figura 25. A figura mostra o exame realizado no equipamento padrão ouro e no equipamento desenvolvido, sendo que a morfologia da curva de espirometria e os valores encontrados estão próximos entre si. Para criar esse comparativo, foi realizado um exame no mesmo dia e mesmo horário, com os dois equipamentos, de forma a não causar nenhuma discrepância devido a variantes externas que pudessem comprometer a confiabilidade do resultado.

Figura 25: Tela do Software do Relatório do Equipamento Desenvolvido



Fonte: Autor

É possível notar que os valores apresentados no equipamento desenvolvido e no padrão ouro tiveram valores que não diferiram em 1%. Com o acréscimo do modelo apresentado nesse projeto, além das funcionalidades como integração com outros sistemas e possibilidade de customização do software de forma a satisfazer o cliente que o adquiriu, além da confiabilidade de um banco de dados em uma programação mais sólida e segura, fazem um ótimo concorrente aos equipamentos do mercado.

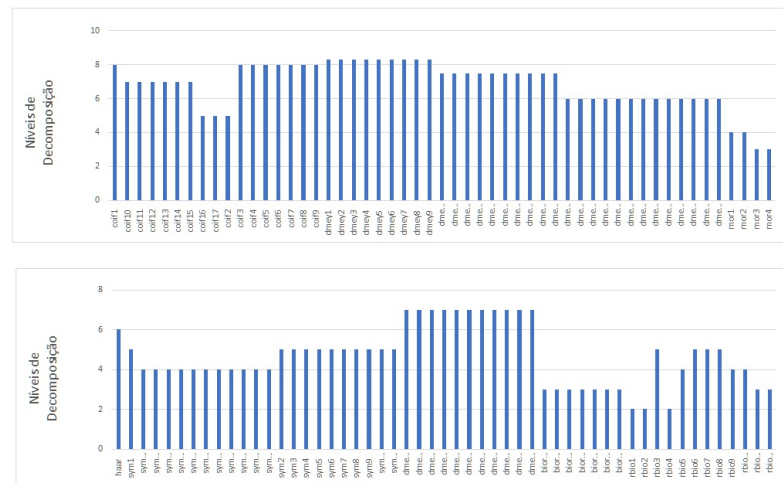
## 5.2 SELEÇÃO DA FAMÍLIA DE WAVELETS

Com o intuito de verificar qual seria a família de wavelet que melhor representa este tipo de sinal, e consequentemente, não utilizar as wavelets que tivessem resultados não satisfatórios, aplicou-se o modelo de diversas famílias no restante das amostras. Após essa análise, foi utilizado um subconjunto de aquisições, e decidiu-se seguir com este padrão, definindo qual a melhor família de wavelet a ser utilizada no restante do trabalho.

Com estes resultados obtidos, ou seja, calculado estes níveis de decomposição, os sinais foram processados para as 83 famílias de wavelets. Com isto, gerou-se um conjunto de "features", ou características que foram utilizados como parâmetros de entrada do classificador.

Podemos observar, através da Figura 26 demonstra os vetores criados para esta inserção. Indivíduo representa cada exame analisado, característica a família de wavelet, array é a quantidade de dados analisados, em a sua dimensionalidade, e percentual a taxa de acerto.

Figura 26: Grafico com valores de resposta da família de Wavelets



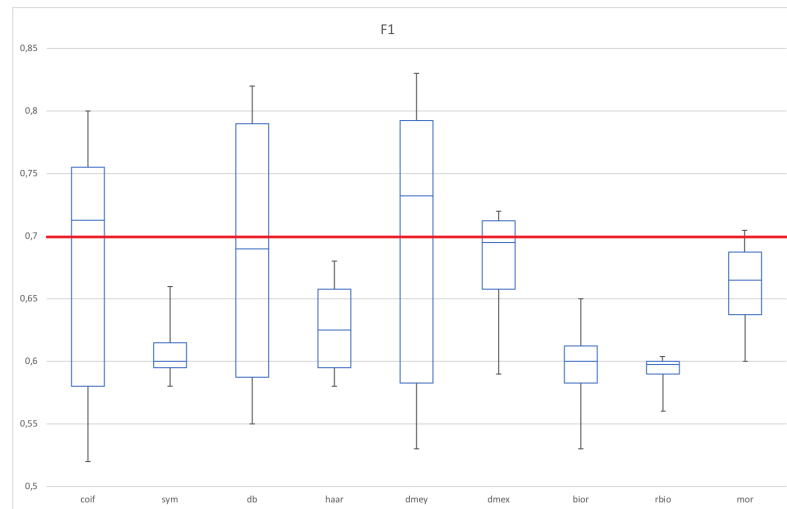
Fonte: Autor

Os resultados que são apresentados na Figura 26, representam qual o valor de F1 para cada uma das famílias de wavelets, bem como suas ordens, no sinal processado. De acordo com essa figura, podemos verificar que existe um padrão visual semelhante nestes resultados, mas que algumas famílias apresentam resultados melhores que outras.

Uma visão mais detalhada é apresentada na Figura 27, onde é possível avaliar os resultados de F1 agrupado por famílias de wavelets. A linha horizontal marca famílias de wavelets com

taxa de classificação abaixo de 0.7 (70%) concluindo também sua avaliação estatística mostrada pela Figura 26.

Figura 27: Grafico com valores de resposta da familia de Wavelets detalhados

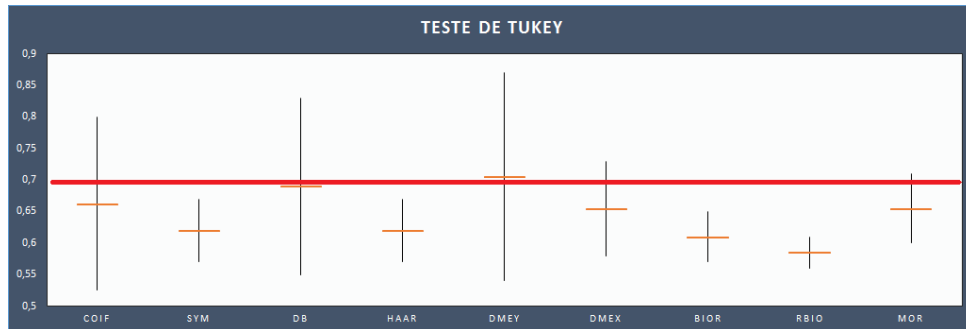


Fonte: Autor

Assim sendo, obtivemos um p-value de 0,0001256, um valor muito baixo, e dessa forma assume-se o valor crítico de 0,05. Considerando que os dados foram testados através da metodologia ANOVA, onde o valor obtido foi de 20,458, é possível perceber que existe uma significativa diferença entre os grupos. Assim, em seguida executamos um teste de diferença mínima significativa, através do Teste de Tukey. A Figura 28 demonstra os resultados comparados de cada família, onde a linha vertical define o valor de 0.05 na análise. Esta diferença média demonstra quais famílias obtiveram um melhor resultado. De acordo com essa métrica, excluimos as famílias que estiveram abaixo dessa linha de corte.

Com esta etapa do experimento concluída, chegamos ao objetivo de definir qual a melhor família de wavelet que apresentou melhor resultado. Comparando-se os resultados das duas figuras prévias, famílias que tiveram taxas inferiores a 70 por cento de no valor de F1 foram excluídas, e as que não passaram no teste de Tukey também foram excluídas. Também devemos notar que algumas famílias não possuem níveis de decomposições iguais, ou seja, suas derivações. Este é um fator que deve ser considerado, pois quanto mais a quantidade de derivações mais dados estatísticos podemos obter.

Figura 28: Demonstração da linha de corte com o Teste de Tukey



Fonte: Autor

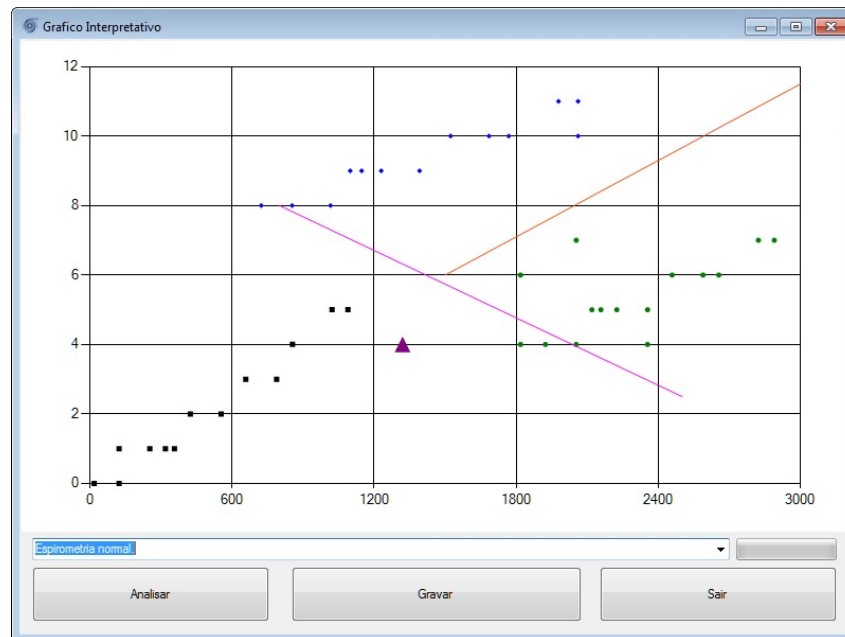
Neste experimento, utilizamos 9 famílias de wavelets, sendo elas a Coiflets (COIF), Daubechies (DB), Discreta de Meyer (DMEY), Haar (HAAR), Symlet (SYM), Chapéu Mexicano (DMEX), Biortogonal(BIOR), Ortogonal Reversa (RBIO) e Morlet (MOR), com diversos níveis de detalhamento.

Por fim, avaliando a Figura 27 e Figura 28, podemos entender que as famílias HAAR, SYM, BIOR e RBIO tiveram sua classificação F1 abaixo de 70%. As famílias COIF, DB, DMEX, MOR foram excluídas pelo teste de Tukey, restando apenas as wavelets de Meyer. Assim, dessas 9 famílias, a DMEY foi a que apresentou o melhor resultado na representação do sinal e dos coeficientes obtidos. E assim definimos qual seria a melhor família de wavelets a ser aplicada neste estudo, obtendo resultados de 78% no teste de F1 e 83% de precisão.

### 5.3 AVALIAÇÃO DAS CURVAS DE ESPIROMETRIA

No Modelo preliminar, a figura 29 demonstra quando um resultado é analisado. A marcação em triângulo no gráfico demonstra o valor encontrado no exame avaliado em relação aos demais. Os pontos exibidos em verde, são os pacientes obstrutivos, em azul os restritivos, e em preto, os normais. As retas demonstram a regressão de ridge, sendo que a reta em roxo demonstra a regressão dos pacientes normais em relação aos obstrutivos e restritivos, e a reta em laranja, a delimitação regional entre os pacientes restritivos e obstrutivos.

Figura 29: Etapa de análise do exame realizado

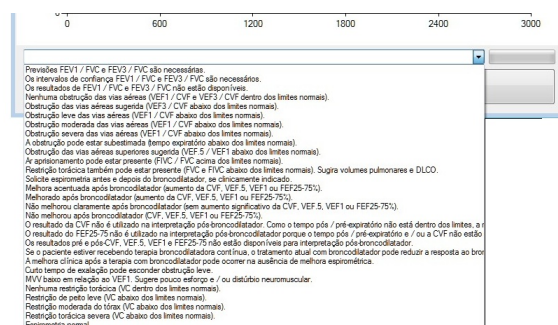


Fonte: Autor

Assim, podemos dizer que quanto mais o exame se afastar das retas "classificadoras" que separam as regiões dos resultados, mais confiável é o resultado. Também podemos dizer qual a taxa de incerteza após calcular o valor médio dos exames daquela região, pois assim conseguimos demonstrar graficamente o valor médio, a taxa de incerteza e o quão distante o valor em análise está distante desse ponto, considerando um método matemático muito robusto e confiável para este tipo de classificação.

A figura 30 demonstra os possíveis diagnósticos inseridos no software, de forma que o médico possa selecionar e aprovar com a sua experiência, e com a ajuda do nosso modelo, ter um resultado mais prático, eficiente e assertivo.

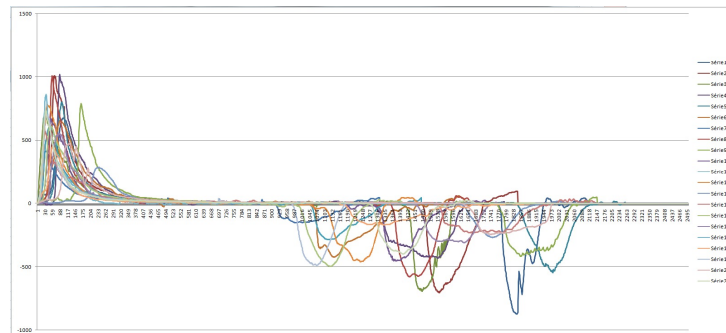
Figura 30: Interpretações fornecidas para o sistema



Fonte: Autor

Com a obtenção de todas as curvas geradas a partir dos sinais característicos, apresentados na Figura 31, demonstra-se uma parte das amostras obtidas com o estudo. Cada curva representa um sinal obtido de um paciente, classificados de acordo com a patologia diagnosticada. O Eixo X representa as marcações de tempo e o eixo Y representa o valor de amplitude do sinal de fluxo lido. De uma forma visual, analisando-se os valores de fluxo entre cada amostra, podemos identificar uma variação entre as curvas. O maior pico de amplitude é o valor de pico expiratório forçado, ou PFE. Para a geração dessas curvas, também foi aplicado um filtro passa banda entre 0,1 e 100Hz com o intuito de facilitar a visualização dos sinais. Para a aplicação do modelo proposto, foi utilizado o sinal na forma RAW, ou seja, completamente sem a aplicação de algum filtro de sinais.

Figura 31: Amostra do Sinal a Ser Processado



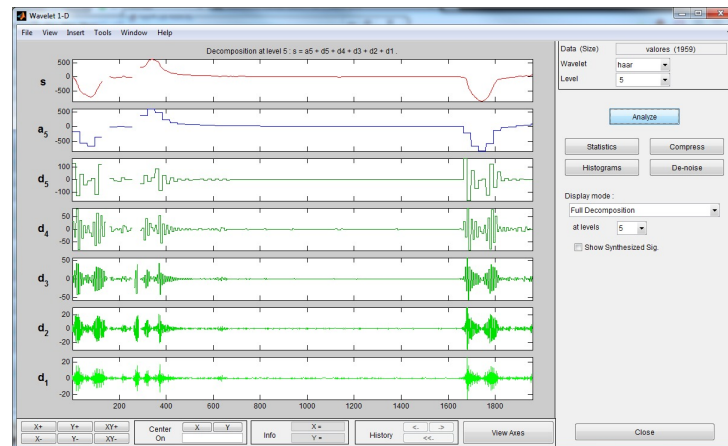
Fonte: Autor

#### 5.4 CLASSIFICAÇÃO INDIVIDUAL DOS DADOS

Na Figura 32 são apresentados os resultados do cálculo que determina quais são os níveis máximos de decomposição que serão aplicados para cada uma das transformadas wavelets contínuas utilizadas.

Na etapa de extração de características utilizamos no experimento um conjunto de 83 famílias e ordens de wavelets. Todos os sinais tiveram a sua decomposição no terceiro nível, ou seja, tres iterações. Não foi aplicado o sinal máximo de decomposição pois o sinal perdeu a sua característica após a terceira decomposição.

Figura 32: Níveis de Decomposição Apresentados Durante o Estudo



Fonte: Autor

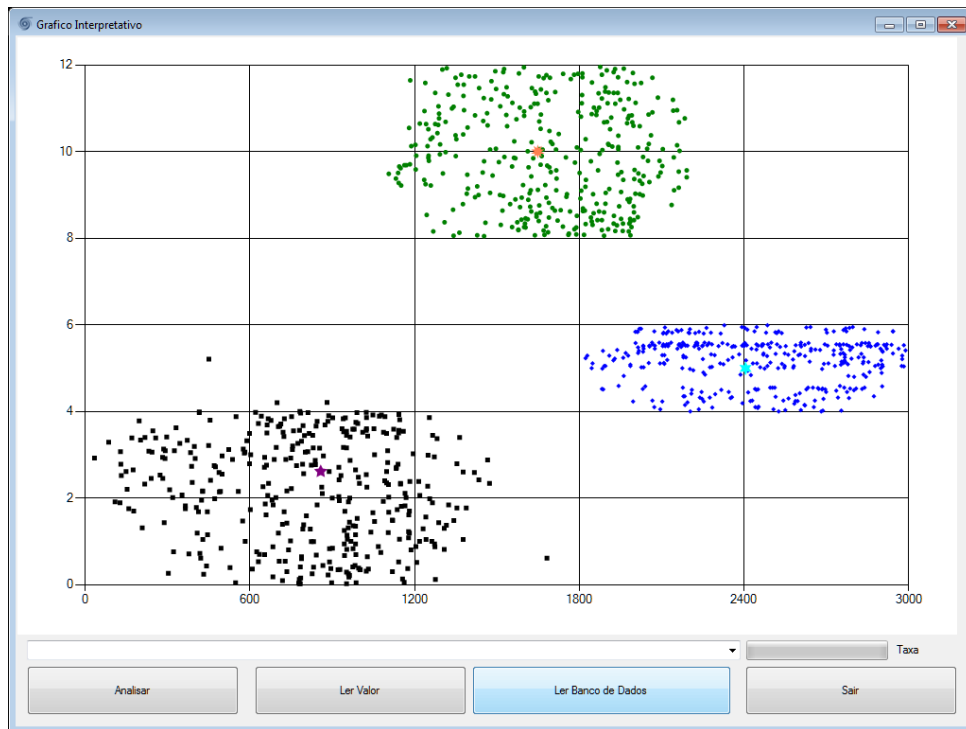
## 5.5 APLICAÇÃO DOS DADOS OBTIDOS APÓS O TÉRMINO DA CRIAÇÃO DO MODELO

Após todos os dados serem processados, caracterizados, discretizados e classificados, montou-se um software de forma a reunir esses dados e realizar a plotagem grafica, bem como a aplicação das retas discretizantes obtidas com a regressão de Ridge, bem como a determinação do centróide da matriz de pontos e o cálculo de assertividade.

Dessa forma, o mesmo software pode ser utilizado para outros estudos, bastando para isso apenas mudar a base de dados dos coeficientes obtidos com a aplicação da wavelet.

Assim, a figura 33 demonstra o software sendo utilizado, após a leitura do banco de dados que contém os coeficientes obtidos.

Figura 33: Apresentação do Software com a Leitura do Banco de Dados



Fonte: Autor

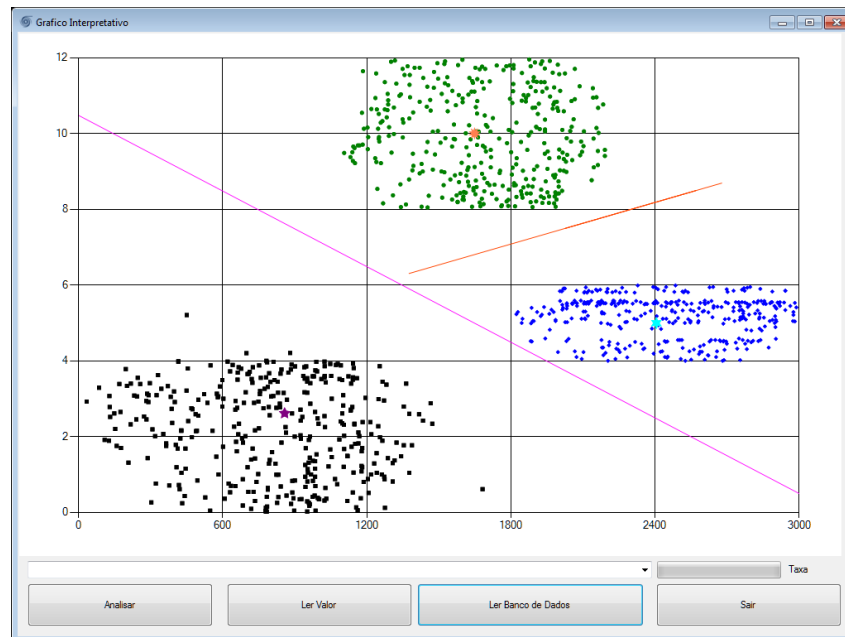
Ao se clicar no botão LER BANCO DE DADOS o software irá criar as retas da regressão de Ridge, que podem ser visualizadas pela figura 34.

É interessante notar que as retas são interdependentes entre si, ou seja, de acordo com o valor que o banco de dados contém, através do modelamento computacional da regressão de Ridge, essas retas podem se alterar também, ou seja, elas conformam e dividem os dados em três categorias dinamicamente.

Nesta etapa é realizado um modelo de clustering, também desenvolvido computacionalmente. Esse modelo divide os dados em 3 grupos, que chamamos de grupo restritivo, obstrutivo e normal.

Como verificado nos estudos desse doutorado, os valores dos coeficientes após a aplicação da wavelet, nesses três grupos, geraram discrepâncias que a regressão de ridge e a clusterização agruparam, e possibilitaram através deste algoritmo também, a geração do centróide que traz o índice de precisão do software. Quanto mais próximo do centróide da região onde um novo exame for analisado estiver, mais próximo da precisão máxima ele estará. Esse fato responde a primeira questão da seção 1, item (a) deste estudo, permitindo relacionar os dados dos exames, conseguindo interpretá-los e classificá-los em 3 grupos distintos.

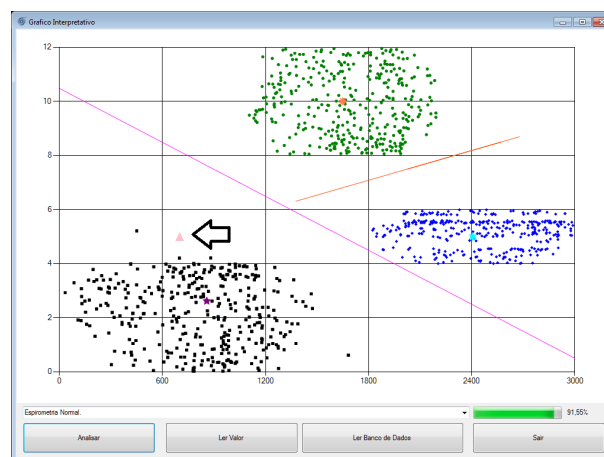
Figura 34: Apresentação do Software com a Leitura do Banco de Dados e a Aplicação da Regressão de Ridge



Fonte: Autor

A figura 35 demonstra o resultado quando inserimos um novo resultado que queremos avaliar, cujo resultado é considerado NORMAL.

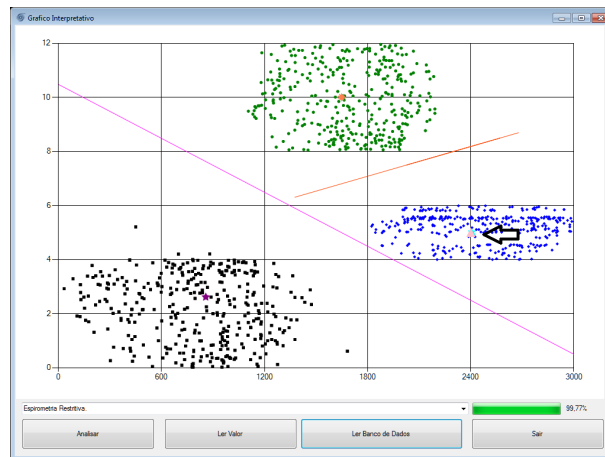
Figura 35: Apresentação do Software com a Leitura de um novo exame considerado NORMAL



Fonte: Autor

Na sequência, a figura 36 demonstra o resultado quando inserimos um novo resultado que queremos avaliar, cujo resultado é considerado RESTRITIVO.

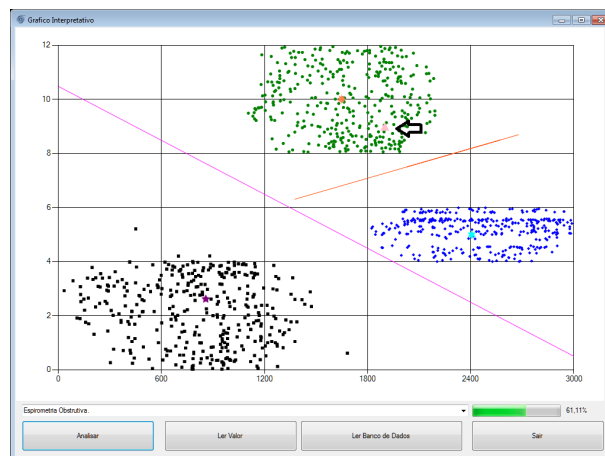
Figura 36: Apresentação do Software com a Leitura de um novo exame considerado RESTRITIVO



Fonte: Autor

E por último, a figura 37 demonstra o resultado quando inserimos um novo resultado que queremos avaliar, cujo resultado é considerado OBSTRUTIVO.

Figura 37: Apresentação do Software com a Leitura de um novo exame considerado OBSTRUTIVO



Fonte: Autor

É possível visualizar em cada figura a apresentação do auxílio em diagnosticar o exame de espirometria, existindo uma barra que possui um valor em percentagem, sendo este o índice de precisão. Por exemplo, o resultado da Figura 37 é próximo do centróide, que apresentou portando uma precisão de 99,77%, ou seja, podemos afirmar com 99,77% de certeza que esse paciente possui características restritivas.

No caso do paciente cujo resultado é mostrado na Figura 36, o índice de precisão foi de 61,11% de precisão, ou seja, podemos dizer que temos 61,11% de certeza que o paciente é obstrutivo.

Então é possível notar que, se o valor de um novo dado inserido se localizar próximo da reta que é o resultado da regressão de ridge entre pacientes restritivos e obstrutivos, o índice de certeza cai consideravelmente. Isso pode ser discutido em trabalhos futuros, para que possa ser melhorado. Dessa forma, o item (b) da seção 1 é elucidado, ou seja, é possível que novos exames sejam adicionados, processados, e confrontados com exames anteriores, dessa forma, criando uma relação entre exames anteriores e novos exames realizados.

Mas, ao se observar os resultados dos pacientes incluídos na região normal, verifica-se que os dados se posicionam distanciados da reta que representa a regressão. Assim, é possível obter resultados de precisão muito satisfatórios, pois como exemplo da figura 35, obteve-se 91,55% de segurança em dizer que essa nova amostra analisada possui características normais, demonstrando a solução do item (d) da seção 1, pois valores de precisão acima de 90%, conforme a pesquisa bibliográfica estudada, é um valor acima do que é utilizado atualmente pelos métodos convencionais.

Lembrando o fato desse modelo ser dinâmico, seu modelamento foi desenvolvido de forma a utilizar os coeficientes da aplicação de wavelets no sinal de exame de espirometria armazenados no banco de dados. Dessa forma conseguimos demonstrar que o uso da inteligência artificial, foi uma importante ferramenta utilizada, pois retornou o resultado final de uma forma prática e rápida, colaborando na resposta do item (c) da seção 1 deste trabalho.

Com o estudo realizado, conseguimos definir que a aplicação da regressão de Ridge, permitiu que pudessemos modelar 3 grupos distintos de conjuntos de dados que representam os pacientes em "normal", "obstrutivo" e "restritivo" a um segmento de reta, que delimita a região gráfica visualmente e analiticamente em que região o novo dado a ser analisado se encontra. Essa delimitação foi definida através da regressão de Ridge, ou seja, ela permite ser possível visualizar graficamente as regiões onde os valores de exames utilizados para construir e testar o modelo de inteligência artificial a aprender e classificar uma nova amostra. Sem a aplicação da regressão em questão, o gráfico de diagnóstico não poderia ter as suas regiões delimitadas para os três grupos que temos por objetivo conseguir discretizá-los.

Desta forma, o modelo apresentou a criação de um modelo de inteligência artificial, de forma a decidir o nível de decomposição, que é a quantidade de iterações que este software executa. A I.A. não é a prerrogativa deste estudo, mas sim, uma aplicação onde o construtor

deve decidir qual modelo de I.A. deve utilizar. No caso deste modelo, utilizamos o modelo de Q Learning, escolhido pelo fato de permitir uma rapidez de implantação e estabilidade no nível de iterações desejadas, de acordo com a resolução do sinal original.

Ou seja, quando o sinal original não pode mais ser reconstruído através da inserção dos coeficientes na equação da wavelet reversa, o software para de realizar as iterações. Assim, existe um equilíbrio entre a reconstrução do sinal original, e a geração de matrizes esparsas, aprimorando a análise do sinal frente ao modelo convencional.

Assim, todo o modelamento da aplicação de wavelets (desenvolvimento do algoritmo), criação do modelo de regressão de Ridge, a criação da clusterização, bem como a criação do cálculo dos centróides através do modelo de Kmeans, são algoritmos desenvolvidos via software, ou seja, não são obtidos de outro software de análise e processamento de sinais ou biblioteca externa. Isso não seria possível sem a aplicação da inteligência artificial, percorrendo o gráfico na busca pelo resultado, e sem a aplicação das wavelets, reduzindo a dimensionalidade e tempo computacional, além também na criação de um banco de dados que pudesse ser lido em ferramenta de desenvolvimento em codificação e programação que não demande profundo conhecimento e estudo. Dessa forma, é possível responder o item (e) da seção 1, criando uma nova proposta de processamento e aplicá-la nas provas de função pulmonar.

## 5.6 DISCUSSÃO

A utilização das wavelets sendo uma base para o andamento desse trabalho, se deu através de eventos que ocorreram de maneira próxima e constantes, mas com uma variação de tempo em suas ocorrências, momentos esses que foram o início e fim das manobras no exame de função pulmonar. Esse momento, é diferente em cada exame, ou manobra executada. Além dessa variável de tempo, existe também a variável de amplitude, pois os valores de fluxo lidos em cada manobras são próximos, mas não semelhantes. A aplicação e decomposição do sinal através das wavelets, permitiu ser possível trazer quase na totalidade as manobras do exame de função pulmonar para um mesmo patamar, tanto de amplitude, quanto de temporalidade. Esses vetores criados permitiram a classificação e caracterização dos sinais, criando os grupos distintos para análise.

O desafio de utilizar esses sinais de forma contínua, sem derivações, foi grande devido a poder aplicar este modelo em uma plataforma de exames. Isso só foi possível devido a termos construído um equipamento, tanto hardware quanto software, e de posse do código fonte do software, possível realizar a integração do módulo de interpretação ao software que não possuía essa funcionalidade. Todos os softwares acabam criando sinais com descontinuidades se não houver cooperação do paciente que está realizando o exame, e não possuem métodos para tratar essas descontinuidades, deixando o método de análise e diagnóstico falho nesse panorama.

Com as aquisições obtidas, observou-se que a cada novo exame, o software se encarrega de inserir esse valor encontrado no banco de dados e comparar com os exames anteriores já inseridos. Isso se dá pelo fato do software possuir inteligência artificial que primeiro, faz a varredura dos valores encontrados graficamente, através da aplicação do Q-learning, encontrando qual região aquele valor está, e por último, se encarrega de avaliar e inserir o valor encontrado na região onde ele mais se acomoda. Podemos dizer que essa acomodação é proveniente apenas da leitura do vetor que, após a aplicação da wavelet e pelo fato do software do equipamento ser de própria autoria, permitir uma interpretação convencional.

Assim, é possível auxiliar o profissional da saúde a laudar o exame, se houve acerto no diagnóstico ou não, e dessa forma, realimentar a função de Q-learn através do reforço positivo se houve ou não acerto no diagnóstico, permitindo que o software possua uma tendência a convergir em resultados mais precisos em usos mais intensos, mas claro, dessa forma aumentaria significativamente o processamento e a quantidade de iterações necessárias para o software começar a reproduzir resultados acertos.

O hardware utilizado no encerramento dessa pesquisa, permite que se faça diversos estudos, devido ao alto grau de escalabilidade de seu banco de dados. Devido a todos os outros equipamentos possuírem banco de dados sem a possibilidade de extração de resultados e valores, o utilizado por nós é de fácil obtenção, e sendo protegido por nome de usuário e senha, que pode ser modificado, inserido, alterado sem nenhum onus ao funcionamento do sistema. Assim, é possível a criação de protocolos, de forma a permitir que estudos nessa linha de pesquisa sejam realizados sem nenhuma necessidade de utilização de softwares de terceiros.

Analizando os experimentos que foram elencados no capítulo 2 em nível de complexidade, acreditamos que o nosso em relação a quantidade de exames utilizados, a criação de um sistema de inteligência artificial que busca a informação desejada, bem como a utilização desse modelo em um software de mercado tornou a criação de um sistema mais elaborado. A utilização de 3 grupos distintos, por mais que tenha simplificado a quantidade de resultados possíveis, traz uma possível distinção mais prática ao médico ou técnico que analisa o exame executado. Também é necessário mencionar que, o sistema de I.A. possibilita sempre a inserção de um novo resultado ao banco de dados utilizado, criando uma retroalimentação ativa no modelo desenvolvido. Isso traz uma grande possibilidade de criar sistemas que podem segregar os resultados regionalmente de acordo com a demografia local. Ter um sistema autônomo desses traz mais segurança no diagnóstico, pois é sabido que conforme (PEREIRA, C. A., 2002) os valores preditivos utilizados nos exames são regionalizados.

Como visto em (LUCK, 2014), nossos resultados entram em acordo com os padrões que possuem características distintas entre eles com a aplicação de nosso modelo. Valores de pacientes normais, são diferentes de pacientes obstrutivos, ou de pacientes restritivos. Em todos os artigos que estudamos, não estava claro qual o modelo de processamento foi utilizado, ou como utilizado, diferente do visto em nosso estudo. Todas as etapas de processamento, bem como algoritmos naturais utilizados em nosso estudo são facilmente encontrados na bibliografia estudada. Também a redução de dimensionalidade traz um fator importante na execução do trabalho, trazendo redução computacional e velocidade na entrega do processamento. Um outro ponto importante é mostrar que desenvolvemos um caminho alternativo na busca por um diagnóstico mais eficaz, trazendo também uma outra perspectiva na análise e diagnóstico pulmonar. Em (LEES et al., 2018) também foi utilizado wavelets para criar uma regressão linear baseada em aprendizado de máquina, que apresenta níveis de precisão adequados e por algumas vezes superiores ao encontrado em nosso estudo, mas é possível verificar que houve um aumento de casos de falsos positivos neste estudo.

Assim, com a aplicação de todas as famílias de wavelets apresentadas neste trabalho, nos leva a decidir ser necessário um estudo mais abrangente nos sinais de prova de função pulmonar, e mais ainda, as ramificações que essa extração de características proporciona. Podemos afirmar que os resultados obtidos durante as etapas de classificação atingiu um índice de precisão satisfatório, condizente com os estudos semelhantes ou por vezes até superior a estudos pesquisados em nossa pesquisa bibliográfica.

A utilização de Wavelet de Meyer foi a correta escolha da wavelet utilizada no decorrer de nosso estudo. Também podemos afirmar que os resultados dos dados estatísticos obtidos demonstram o correto desenvolvimento do modelo matemático, obtendo resultados dentro dos padrões aceitáveis nos modelos convencionais. O Teste estatístico aplicado permitiu que fosse certificada a nossa análise, e também permite demonstrar a correlação com os grupos discretizados, podendo dessa maneira selecionar através de um comparativo em qual grupo a amostra que deseja-se analisar naquele instante se encontra com uma precisão de mais de 91% em alguns casos.

E por fim, e não menos importante, a necessidade na busca por novos métodos de diagnósticos médicos, comprovados por modelos estatísticos que apresentam um alto índice de F1, e com boa precisão, tem sido a vanguarda no desenvolvimento e modelamento de sistemas computacionais que segregam, discretizam e processam dados, dados estes que podem ter um conteúdo importante se visto sob outra ótica, como por exemplo, a aplicação de wavelets em ambientes não utilizados anteriormente. Esta foi a proposta que, em conjunto com modelos de aplicação de inteligência artificial, como o Q-Learn, utilizado em nosso estudo permitiu que pudéssemos apresentar com sucesso um novo modelo de biomarcador no diagnóstico de exames de espirometria.

## 6 CONCLUSÃO

Tomando em consideração este trabalho realizado, pode-se observar o uso das transformadas wavelets, do uso de inteligência artificial, e modelos de regressões lineares como regressão de Ridge, além da aplicação do algoritmo de K-means que, aplicado ao conjunto de exames de espirometria, obteve-se resultados com valores acima dos obtidos em outros trabalhos.

Estes resultados foram tratados através de um modelamento estatístico de forma a confirmar a análise e resultados desenvolvidos. É possível notar que o uso adequado da família de wavelets contribui para o modelamento estatístico e juntamente com a iniciativa de utilizar grandes conjuntos de dados, consegue-se analisar os diferentes níveis de coeficientes e decomposição do sinal que fora obtido pelo equipamento.

Com os experimentos realizados, tornou-se possível traçar uma relação entre o comportamento dos sinais oriundos do exame de espirometria e os resultados comparados a diferentes conjuntos de exames realizados por indivíduos, com uma sinificativa relevância estatística.

Estes resultados mostraram que existe uma relação entre os sinais obtidos com o espirometro e o modelo estatístico apresentado neste trabalho.

Também é possível utilizar o modelo proposto em outros segmentos que não somente a análise dos sinais de espirometria. Conseguimos também aumentar o desempenho computacional para auxiliar uma tomada de decisão entre existir uma patologia pulmonar ou não.

O modelo de processamento proposto também pode auxiliar e aumentar a capacidade de realização de exames, pois o diagnóstico de uma possível patologia se torna mais rápido.

Um ponto muito importante a se analisar é que o modelo, devido a criar uma suavização do sinal, elimina discontinuidades presentes em alguns exames, que fazem com que os métodos de diagnósticos utilizados atualmente se tornem falhos, requerendo um maior tempo e critério na elaboração do laudo diagnóstico.

Este trabalho desenvolvido gerou uma contribuição em outras pesquisas, através da integração de software com uma pesquisa em biosinais e criou uma abordagem alternativa para o tratamento e classificação de patologias pulmonares. A proposta desse trabalho também gerou uma criação de um novo modelo estatístico para a avaliação dos resultados.

## 6.1 TRABALHOS FUTUROS

Uma possível continuação para esta pesquisa, seria:

- a) Criar uma análise ampliada sobre os vetores resultantes da regressão de ridge, pois possuem informações que podem acrescentar novos marcadores.
- b) Aplicação do processamento proposto em conjuntos com outros tipos de biosinais, como por exemplo, cânula respiratória ou análise de ventilação volumétrica;
- c) Aplicação dos testes estatísticos, a fim de realizar a análise dos resultados, de uma forma paramétrica.
- d) Minimizar o espectro amostral dos coeficientes resultantes da análise, utilizando por exemplo, métodos Bayseanos.
- e) Aplicar esse modelo em um número maior de dimensões, tentando avaliar se o sistema proposto consegue convergir de uma forma mais precisa que o modelo proposto neste trabalho.
- f) Criar um modelo que permita a criação do mesmo ambiente de análise com dados de outras regiões mundiais.
- g) Estudar e aprimorar a variação que ocorreu aplicando-se a regressão de Ridge, de forma a criar uma melhor precisão e confiança quando um novo resultado analisado estiver próximo da região entre os pacientes obstrutivos e restritivos.
- h) Verificar e aplicar o modelamento do software de forma a estudar e possibilitar que seja utilizado por exemplo, em imagens radiológicas também.

## REFERÊNCIAS

ABNT. **Equipamento eletromédico**. [S.l.]: ABNT NBR, 2010.

ADELI, Hojjat; ZHOU, Ziqin; DADMEHR, Nahid. Analysis Of Eeg Records In An Epileptic Patient Using Wavelet Transform. **Journal of Neuroscience Methods**, v. 123, n. 1, p. 69–87, 2003. ISSN 01650270. DOI: 10.1016/S0165-0270(02)00340-0.

AKAY, Metin. Wavelet applications in medicine. **IEEE Spectrum**, v. 34, n. 5, p. 50–56, 1997. ISSN 00189235. DOI: 10.1109/6.590747.

AKPINAR-ELCI, M.; FEDAN, K. B.; ENRIGHT, P. L. FEV6 as a surrogate for FVC in detecting airways obstruction and restriction in the workplace. **European Respiratory Journal**, v. 27, n. 2, p. 374–377, 2006. ISSN 09031936. DOI: 10.1183/09031936.06.00081305.

ALAM, S.; RAHMAN, M.; HOSSAIN, M. et al. Automatic Human Brain Tumor Detection in MRI Image Using Template-Based K Means and Improved Fuzzy C Means Clustering Algorithm. **Big Data and Cognitive Computing**, 2019.

ALDROUBI, Akram; UNSER, Michael. **Wavelets in Medicine and Biology**. [S.l.: s.n.], 1995.

ASHER, Rebecca J. Capnographic Analysis for Disease Classification by, 2012.

ASHOFTCH, A.; BRAVO, J. M. Data science training for official statistics: A new scientific paradigm of information and knowledge development in national statistical systems. **Metron**, 2021.

BASILI, Victor. **Identifying Relevant Information for Testing Technique Selection: An Instantiated Characterization Schema**. [S.l.: s.n.], 2003.

BELARDI, Aldo Artur. The Method of Moments and the Haar Wavelets Used in a Determination the Eddy Current Matrix. v. 3, n. 4, p. 1–6, 2013.

BETANCOURT, JP et al. Capnogram Feature Extraction Based on Wavelet Decomposition by Segments for Classification of Asthma Severity. **Feature Extraction Based**, p. 1–6, 2013.

BOROKOV, A. A.r. **Mathematical statistics**. [S.l.: s.n.], 2019.

C.H., Spurgeon. **The Concise Oxford Dictionary of the Christian Church**. [S.l.]: Oxford University, 1884.

CAMPBELL, Joseph; MOYERS, Bill D. **The power of myth / Joseph Campbell, with Bill Moyers ; Betty Sue Flowers, editor**. 1st ed. [S.l.]: Doubleday New York, 1988. xix, 233 p., [16] p. of plates : ISBN 0385247745 0385247737.

CARTA, R. et al. A differential pressure approach to spirometry. **Conference Proceedings - IEEE Biomedical Circuits and Systems Conference Healthcare Technology, BiOCAS2007**, p. 5–8, 2007. DOI: 10.1109/BIOCAS.2007.4463295.

CHEN, Shy-Leh; LIANG, Jin-Wei; HO, Keng-Chu. Parametric Identification of Nonlinear Systems by Haar Wavelets: Theory and Experimental Validation. **Journal of Vibration and Acoustics**, v. 134, n. 3, p. 031005, 2012. ISSN 07393717. DOI: 10.1115/1.4006229.

CORTINA, J. M.; NOURI, Hossein. **Effect Size for ANOVA Designs**. [S.l.: s.n.], 1999.

COSTA, D; JARNARNI, E. Bases fundamentais da espirometria. **Rev. bras. fisioter**, v. 5, n. 2, p. 95–102, 2001. ISSN 1413-3555.

DAUBECHIES, Ingrid. **Ten Lectures on Wavelets**. Philadelphia: SIAM, 1999. cap. 1, p. 1–16. DOI: 10.1137/1.9781611970104.

DESERNO, Thomas M; ANTANI, Sameer; LONG, Rodney. Ontology of gaps in content-based image retrieval. **Journal of digital imaging**, v. 22, n. 2, p. 202–15, abr. 2009. ISSN 1618-727X. DOI: 10.1007/s10278-007-9092-x. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3043678/>. Acesso em: 1 jul. 2018.

DESHMUKH, Megha D.; PANDITRAO, Anagha. Design and development of thermistor based sensor for spirometry. **2012 IEEE Students' Conference on Electrical, Electronics and Computer Science: Innovation for Humanity, SCEECS 2012**, p. 1–3, 2012. DOI: 10.1109/SCEECS.2012.6184837.

DEVI, G M.; VENKATESAN, P. Least Squares Support Vector Regression for Spirometric Forced Expiratory Volume (FEV1) Values. **International Journal of Science and Technology**, v. 3, n. 1, p. 74–78, 2013.

DIGNUM, Virginia. AI is multidisciplinary. **AI Matters**, 2019.

DIRECTORS, IEEE Board of. Artificial Intelligence. **IEEE Position Statement**, 2019.

DONOHO, David L.; JOHNSTONE, Iain M. Ideal Spatial Adaptation by Wavelet Shrinkage. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 81, n. 3, p. 425–455, 1994.

EAKINS, John P. Towards intelligent image retrieval. **Pattern Recognition**, v. 35, p. 3–14, 2002. ISSN 00313203. DOI: 10.1016/S0031-3203(01)00038-3.

EHSANES, A.K.; K., Saleh; G. **Theory of Ridge Regression Estimation with Applications**. [S.l.]: John Wiley e Sons, Inc., 2019.

FERRARI, Hevelin Viviane; HOTO, Robinson; MACULAN, Nelson. Uma Comparação Entre Redes Neurais Wavelet , Lms , Mlp E Rbf Para Classificação De Dpoc. February 2016, 2006.

FINKELSTEIN, S.M. et al. Clinical staging of chronic rejection in lung transplantation using home spirometry. **Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings**, v. 17, n. 1, p. 5–6, 1995.

FRANZONI, Roberto; PETTINI, Marco. **Theorem on the Origin of Phase Transitions**. [S.l.: s.n.], 2004.

GRAMI, Ali. Probability, Random Variables, Statistics, and Random Processes, 2020.

GRAPS, Amara. Introduction to wavelets. **IEEE Computational Science & Engineering**, v. 2, n. 2, p. 50–61, 1995. ISSN 10709924. DOI: 10.1109/99.388960. Acesso em: 1 jul. 2018.

GUNNING, D.; STEFIK, M.; CHOI, J. et al. XAI—Explainable artificial intelligence. **Science Robotics**, 2019.

HANKINSON, John L.; GARDNER, Reed M. Standard waveforms for spirometer testing. **Am Rev Respir Dis**, n. 126, p. 362–364, 1982.

HASTIE, T; TIBSHIRANI, R; FRIEDMAN, J. **The elements of statistical learning**. California: Springer, 2009. P. 1–8. Disponível em: <http://link.springer.com/content/pdf/10.1007/978-0-387-84858-7.pdf>. Acesso em: 1 jul. 2018.

HASTIE, Trevor; TIBSHIRANI, Robert; WAINWRIGHT, Martin. **Statistical learning with sparsity The lasso and generalizations**. California: CRC press, 2015.

HENDERSON, Hugh. **SAT Subject Test Physics**. [S.l.]: Kaplan Publishing, 2017.

JAMES, Gareth et al. **An Introduction to Statistical Learning**. California: Springer, 2013. v. 112, p. 215–219.

JEREMIA, N. E.; NURROHMAH, S.; FITHRIANI, I. Robust Ridge regression to solve a multicollinearity and outlier. **Journal of Physics: Conference Series**, 2020.

KABIR, Asad; IFTEQAR, Sarah; BHAT, Abid. Obstructive sleep apnea in adults. **Hospital practice (1995)**, v. 41, n. 4, p. 57–65, 2013. ISSN 21548331. DOI: 10.3810/hp.2013.10.1081.

KALICKA, Renata; SLOMINSKI, Wojciech; KUZIEMSKI, Krzysztof. The modelling of spirometry - The application for diagnostic purposes. **Proceedings of the 2008 1st International Conference on Information Technology, IT 2008**, v. V, May, p. 0–3, 2008. DOI: 10.1109/INFTECH.2008.4621659.

KASSEM, A.; HAMAD, M.; EL MOUCARY, C. A smart spirometry device for asthma diagnosis. **Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS**, 2015-November, p. 1629–1632, 2015. ISSN 1557170X. DOI: 10.1109/EMBC.2015.7318687.

KAVITHA, A; SUJATHA, C M; RAMAKRISHNAN, S. Prediction of Spirometric Forced Expiratory Volume (FEV 1 ) Data Using Support Vector Regression. **Measurement Science Review**, v. 10, n. 2, p. 63–67, 2010. ISSN 13358871. DOI: 10.2478/v10048-010-0011-9.

KHUBANI, J. Spirometric Data Analysis By Support Vector Machine, p. 2–5, 2012.

LABATE, Domenico et al. Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: A comparison. **IEEE Sensors Journal**, v. 13, n. 7, p. 2666–2674, 2013. ISSN 1530437X. DOI: 10.1109/JSEN.2013.2257742.

LADOSKY, Waldemar et al. Comparação entre valores espirométricos de referência obtidos a partir das equações de Knudson e de Pereira Adultos. **Jornal de Pneumologia**, v. 27, n. 6, p. 315–320, 2001. DOI: 10.1590/S0102-35862001000600005.

LAY-EKUAKILLE, A.; VENDRAMIN, G.; TROTTA, A. Spirometric measurement postprocessing: Expiration data recovery. **IEEE Sensors Journal**, v. 10, n. 1, p. 25–33, 2010. ISSN 1530437X. DOI: 10.1109/JSEN.2009.2035743.

LEE, J. K. et al. Estimation of Rock Load of Multi-Arch Tunnel with Cracks Using Stress Variable Method, 2020.

LEE, Jungsil et al. Graphic analysis of flow-volume curves: a pilot study. **BMC Pulmonary Medicine**, BMC Pulmonary Medicine, v. 16, n. 1, p. 18, 2016. ISSN 1471-2466. DOI: 10.1186/s12890-016-0182-8.

LEES, Stephanie et al. A review of rapid serial visual presentation-based brain-computer interfaces. **Journal of Neural Engineering**, v. 15, n. 2, p. 021001, 2018. ISSN 1741-2560. DOI: 10.1088/1741-2562/aa9817. Acesso em: 1 jul. 2018.

LI, Rui et al. Human-centric approaches to image understanding and retrieval. In: 2010 Western New York Image Processing Workshop. 2010 Western N, US: IEEE, nov. 2010. P. 62–65. ISBN 978-1-4244-9298-5. DOI: 10.1109/WNYIPW.2010.5649743.

LIANG, Zhen et al. Content-based image retrieval using a combination of visual features and eye tracking data. **Symposium on Eye-Tracking.**, v. 1, n. 212, p. 41–44, 2010. Disponível em: <http://dl.acm.org/citation.cfm?id=1743675>. Acesso em: 1 jul. 2018.

LIU, G. C.; KO, C. H. Visual Communication Design and Wireless Data Transmission Technology for Blockchain Big Data Information Presentation. **Wireless Communications and Mobile Computing**, 2022.

LOPATA, V. et al. On the spirometry and spirometers standardization. **2013 IEEE 33rd International Scientific Conference Electronics and Nanotechnology, ELNANO 2013 - Conference Proceedings**, p. 316–318, 2013. DOI: 10.1109/ELNANO.2013.6552028.

LUCK, Steven J. **An Introduction to the Event-Related Potential Technique**. Edição: A Bradford Book. Second Edi. California: MIT Press, 2014. P. 2–100. ISBN 9780262621960. DOI: 10.1007/s10409-008-0217-3. Disponível em: <https://mitpress.mit.edu/books/introduction-event-related-potential-technique>. Acesso em: 1 jul. 2018.

MAIA, Isabel A. Avaliação da função pulmonar por espirometria na leishmaniose visceral, 2015.

MCGRANAHAN, Lucas. **Darwinism and Pragmatism: Willian James on Evolution and Self-Transformation**. [S.l.: s.n.], 2017.

MHETRE, Prof Manisha R; KHUBANI, Jitendra. Pattern Recognition and Forced Expiratory Volume ( FEV1 ) Spirometric Data Prediction by Support Vector Machine. v. 28, p. 112–116, 2012.

MORETIN, PA. **Ondas e Ondaletas: da Análise de Fouriera Análise de Ondaletas**. São Paulo: Edusp, 1999. P. 10–31.

MORETO TRINDADE, Alexandre; LINS FAGUNDES DE SOUSA, Thiago; LUIS PEREIRA ALBUQUERQUE, Andre. A interpretação da espirometria na pratica pneumologica: ate onde podemos avançar com o uso dos seus parametros? **Pulmão RJ**, v. 24, n. 1, p. 3–7, 2015.

MORETTIN, P. A.; PINHEIRO, A.; VIDAKOVIC, B. Wavelets in Functional Data Analysis. **Springer International Publishing AG**, 2017.

MORETTIN, P.A. **Ondas e Ondaletas**. [S.l.: s.n.], 1999. P. 225.

MYTHILI, A; SUJATHA, C M. Analysis Of Spirometric Pulmonary Function Data Using Particle Swarm Optimization, 2015.

NAMI, Al. Utilizing wavelet transform and support vector machine for detection of the paradoxical splitting in the second heart sound. **Medical and Biological Engineering and Computing**, v. 48, n. 2, p. 177–184, 2010. ISSN 01400118. DOI: 10.1007/s11517-009-0548-7.

PAT, Y.; RAMAMOCHAN, K. R. Chapter 6 - Fast Discrete Transforms. In: ELLIOTT, Douglas F. (Ed.). **Handbook of Digital Signal Processing**. San Diego: Academic Press, 1987. P. 481–525. ISBN 978-0-08-050780-4. DOI: <https://doi.org/10.1016/B978-0-08-050780-4.50011-4>. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780080507804500114>.

PEREIRA, C. A. Espirometria. **Jornal Brasileiro de Pneumologia**, v. 28, supl 3, s1–s82, 2002. DOI: <http://dx.doi.org/10.1590/S1806-37132007000400008>.

PEREIRA, Carlos A. C. I Consenso Brasileiro sobre Espirometria. **J Pneumol**, v. 22, n. 3, p. 105–164, 1996. ISSN 0102-3586.

PINHEIRO, Murilo Cerqueira et al. Comparação entre os valores de referência para CVF, VEF1 e relação VEF1 /CVF em brasileiros caucasianos adultos e aqueles sugeridos pela Global Lung Function Initiative 2012. v. 35, n. 12, p. 1168–1173, 2009.

PIVA, R. G. G. Auxilio no Diagnostico em Exames de Espirometria com a Aplicação das Wavelets. **Centro Universitário da FEI**, 2017.

PIVA, R. G. G.; BECHELLI, R. P.; BELARDI, A. A. A New Biomarker in Diagnostic in Spirometry Exams with the Application of Wavelets. **Journal of Advances in Applied And Computational Mathematics**, 2018.

PIVA, Rodrigo G. G. Um novo biomarcador em análises de exames de espirometria. Congresso de Novas Tecnologias Físicas da IFSUL, 2017.

RODRIGUES, Roberto Jr.; PEREIRA, Carlos A.C. Resposta a broncodilatador na espirometria : que parâmetros e valores são clinicamente relevantes em doenças obstrutivas ? \*. v. 27, n. 1, p. 35–47, 2001.

ROMERO, Wilfredo Urbina. **Gaussian Harmonic Analysis**. [S.l.]: Springer, 2019.

RONCHETTI, Elvezio. The main contributions of robust statistics to statistical science and a new challenge. **Metron**, 2020.

RUI, Yong; HUANG, Thomas S. Relevance feedback: a power tool for interactive content-based image retrieval. **Circuits and Systems**, v. 8, n. 5, 1998.

RUSSELL, Stuart; NORVIG, Peter; DAVIS, Ernest. **Artificial intelligence: a modern approach**. 3. ed. Upper Saddle River, NJ: Prentice Hall, 2010.

RUTHERFORD, Andrew. **SAT Subject Test Physics**. [S.l.]: John Wiley e Sons, Inc., 2011.

SAHIN, Deniz et al. Diagnosis Of Airway Obstruction Or Restrictive Spirometric Patterns By Multiclass Support Vector Machines. **Journal of Medical Systems**, v. 34, n. 5, p. 967–973, 2010. ISSN 01485598. DOI: 10.1007/s10916-009-9312-7.

SAMUSENKO, Pavel. Inconsistency of chi2 test for sparse categorical data under multinomial sampling, 2011.

SERHAL, H.; ADBALLAH, N.; MARION, J.M. et al. Overview on prediction, detection, and classification of atrial fibrillation using wavelets and AI on ECG. **Computers in Biology and Medicine**, 2022.

SEROV, Valery. **Fourier Series, Fourier Transform and Their Applications to Mathematical Physics**. [S.l.]: SpringerLecture Notes in Mathematics, 2017.

SHUVRA, S.B.; DREPTTE, R. L.; TAKALA, J. **Handbook of Signal Processing Systems**. [S.l.]: Springer, 2019.

SMITH, S J M. EEG in the diagnosis, classification, and management of patients with epilepsy. **Journal of neurology, neurosurgery, and psychiatry**, 76 Suppl 2, p. ii2–7, 2005. ISSN 0022-3050. DOI: 10.1136/jnnp.2005.069245.

SOLEIMANI, Vahid et al. Remote, Depth-based Lung Function Assessment. **IEEE Transactions on Biomedical Engineering**, v. X, n. 10, p. 1–1, 2016. ISSN 0018-9294. DOI: 10.1109/TBME.2016.2618918.

SPRIGGS, E A. The History Of Spirometry. **British journal of diseases of the chest**, v. 72, n. 3, p. 165–180, 1978. ISSN 00070971. DOI: 10.1016/0007-0971(78)90038-4.

THE MATHWORKS, Inc. **MATLAB version 9.3.0.713579 (R2017b)**. [S.l.], 2017.

TOURING, Alan Mathison. **Computing Machinery and Intelligence**. [S.l.]: Mind Magazine, 1950.

ULBRICHT, Leandra. Analise da variancia dos parâmetros de referência para classificação da função pulmonar. October, p. 2–6, 2016.

VALENS, C. **A Really Friendly Guide to Wavelets**. California: UNM Computer Science, 1999.

VALENS, C. A really friendly Guide to Wavelets. UNM Computer Science, 1999.

VALENTINUZZI, Max E.; JOHNSTON, Richard. Spirometry: A historical gallery up to 1905. **IEEE Pulse**, v. 5, n. 1, p. 73–76, 2014. ISSN 21542287. DOI: 10.1109/MPUL.2013.2289511.

WAAL, T.; DELDEN, A.; SCHOLTUS, S. Multi-source Statistics: Basic Situations and Methods, 2019.

WANG, Shengyu et al. FEV1/FEV6 in Primary Care Is a Reliable and Easy Method for the Diagnosis of COPD, 2016.

WANG, Xuewe. (2011). **Event Detection For Post Lung Transplant Based On Home**. 2011. Tese (Doutorado).

WATKINS C. J. C. H.; DAYAN, P. **Machine Learning**. [S.l.: s.n.], 1992.

WEIB, Gerhard. Agent orientation in software engineering. **Cambridge University Press**, 2001.

WELLS, Curt. **Kalman Filter in Finance**. [S.l.: s.n.], 1996.

XAVIER, Chico. **Luz no Lar**. [S.l.]: FEB, 2014.

YANG, Y.; BASU, S.; ZHANG, L. A Bayesian hierarchical variable selection prior for pathway-based GWAS using summary statistics, 2019.

YUEFAN, X.; SEN, Z.; ZHENGTAO, C. Extreme Learning Machine for Heartbeat Classification with Hybrid Time-Domain and Wavelet Time-Frequency Features. **Deep Learning in Bioinformatics and Biomedical Images**, 2021.

ZHANG, Xuying; DENG, Caixia; HAN, Yao. The image space of meyer wavelet transform. **Proceedings of 2013 2nd International Conference on Measurement, Information and Control, ICMIC 2013**, v. 2, n. 50, p. 1136–1139, 2013. DOI: 10.1109/MIC.2013.6758159.

ZHAO, Yihong; OGDEN, R. Todd; REISS, Philip T. Wavelet-Based LASSO in Functional Linear Regression. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 21, n. 3, p. 600–617, 2012. PMID: 23794794. DOI: 10.1080/10618600.2012.679241. eprint: <https://doi.org/10.1080/10618600.2012.679241>. Disponível em: <https://doi.org/10.1080/10618600.2012.679241>. Acesso em: 1 jul. 2018.

ZOU, Yunzhi. Multi-Variable Calculus, 2020.

**APÊNDICE A – TERMO DE CONFIDENCIALIDADE PARA CRIAÇÃO DO BANCO DE  
DADOS**

# TERMO DE CONFIDENCIALIDADE E AUTORIZAÇÃO DO USO DE INFORMAÇÕES

Em 26 de Novembro de 2018, Eu, Dra. Silvia Carla Rodrigues de Sousa, Médica Pneumologista, credenciado pelo CRM sob numero 100632, após análise criteriosa dos documentos anexo a essa carta, constatei que são exames de função pulmonar, também conhecidas como espirometria simples, realizadas no período do ano de 2017.

Constato também que possui em seu conteúdo um breve cadastro para distinção, além de resultados dos exames considerados aceitáveis, além das curvas características de cada exame de fluxo x volume e volume x tempo (FxV e VxT) de cada individuo.

Autorizo o uso dessas informações, que fazem parte de meu arquivo pessoal, de forma confidencial, para serem utilizadas em pesquisa científica no âmbito acadêmico, como solicitado pel Sr. Rodrigo Galuzzi Garcia Piva, aluno de doutorado sob orientação do Prof Dr. Aldo Artur Belardi, com o tema UM NOVO BIOMARCADOR PARA AUXILIAR NA INTERPRETAÇÃO DOS EXAMES DE ESPIROMETRIA ATRAVÉS DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL E DAS FUNÇÕES DAS WAVELETS, permitindo assim que a utilização desses exames realizados possam ser parte integrante de sua tese e levantamento de dados e curvas gráficas necessárias para sua pesquisa científica

Sem Mais.



Silvia Carla Rodrigues Sousa  
CRM 100632

**APÊNDICE B – MODELO DE EXAME UTILIZADO PARA CRIAÇÃO DO BANCO DE  
DADOS**

# Pulmonary Function Report

Pre-Rx Serial Report  
Page 1 of 2

## Patient Information

Name: KoKo, Demo D.

ID: 1234567890

Height (in): 70.0

Sex: Male

Smoking history (pk-ysr): 6

Weight (lb): 200.0

Birthdate: 10/29/52

Predicted set: ITS 1984

Comments:

Diagnosis: Hard Working Instrument

## Interpretation

Site: Local

Protocol:

Test date/time: 03/23/91 01:03:39 PM

Physician:

Agent:

Number of efforts performed: 3

Technician:

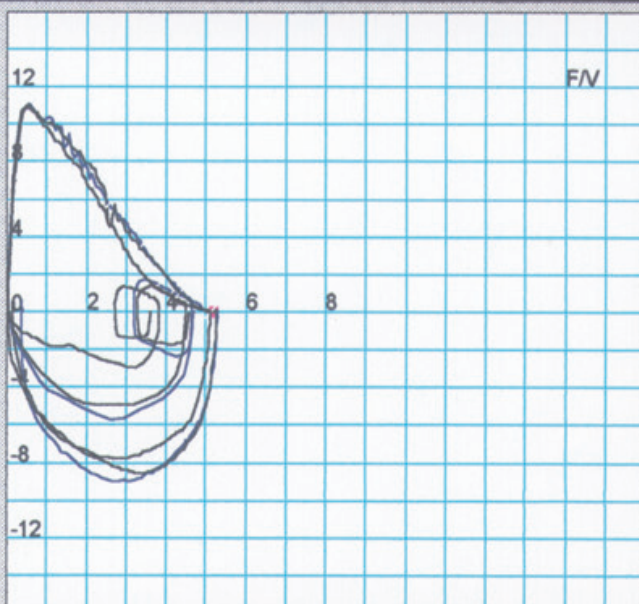
Medication status: Pre

## Results

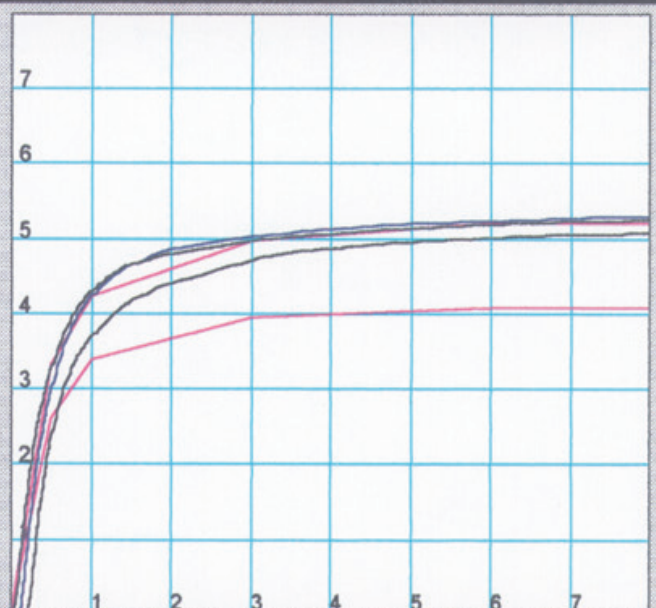
| Result          | Pred | Best  | %Prd   | Cons  | %Prd   | Cons  | %Prd  |
|-----------------|------|-------|--------|-------|--------|-------|-------|
| FVC (L)         | 5.20 | 5.29  | 101.7% | 5.26  | 101.1% | 5.10  | 98.0% |
| FEV1 (L)        | 4.24 | 4.41  | 103.9% | 4.37  | 103.0% | 3.98  | 93.8% |
| FEV1/FVC        | 0.82 | 0.83  | 101.7% | 0.83  | 101.7% | 0.78  | 95.6% |
| FEF25-75% (L/s) | 4.32 | 4.68  | 108.4% | 4.63  | 107.3% | 3.48  | 80.6% |
| PEFR (L/s)      | ---  | 10.67 | ---    | 10.52 | ---    | 10.38 | ---   |

Test comments: Good patient effort and cooperation

## FVC Flow vs. Volume



## FVC Volume vs. Time



## **APÊNDICE C – CURSOS CONCLUIDOS DURANTE O ESTUDO**

Durante o estudo do doutoramento, sentiu-se a necessidade de aprender novas linguagens de programação, bem como novos modos e tecnologias na área de desenvolvimento de software, além do gerenciamento e uso de novas metodologias de desenvolvimento e análise de sistemas. Dessa forma, os seguintes cursos foram realizados:

- a) - Curso de Metodologia Ágil. - Faculdade Anhembí Morumbi - Data: 04/04/2020.  
- Carga Horária: 20 horas.
- b) - Curso de Design Thinking. - Faculdade Anhembí Morumbi - Data: 12/04/2020. -  
Carga Horária: 20 horas.
- c) - Curso de User Experience. - Faculdade Anhembí Morumbi - Data: 20/04/2020. -  
Carga Horária: 20 horas.
- d) - Curso Java Coders Santander. - Escola Digital House Coding School. - Data:  
17/05/2020. - Carga Horária: 240 horas.

## **APÊNDICE D – AULAS MINISTRADAS DURANTE O DOUTORAMENTO**

Concomitantemente, como a produção desse doutorado, foi ministrado aulas para alunos de graduação e pós graduação em áreas correlatas, nas seguintes disciplinas:

- a) Programação Cliente Servidor
- b) Programação Web
- c) Projetos Interdisciplinares
- d) Estrutura de Dados
- e) Estrutura de Dados Avançados
- f) Análise Numérica e Computação Gráfica.
- g) Algoritmos
- h) Engenharia de Software
- i) Programação Orientada a Objetos
- j) Desenvolvimento Web
- k) Programação Gráfica e Visão Computacional.
- l) Algoritmos e Programação
- m) Testes de Software
- n) Gerenciamento de Projetos
- o) Sistemas Digitais
- p) Desenvolvimento Distribuído
- q) Eletromagnetismo
- r) Banco de Dados
- s) Projetos e Desenvolvimento De Sistemas
- t) Gerenciamento de Serviços de TI
- u) Desenvolvimento Móvel e Embarcado
- v) Internet of Things (IoT)
- w) Interface Humano Computador (IHC)

## **APÊNDICE E – PRODUÇÃO LITERÁRIA E PRODUÇÃO CORRELACIONADAS**

Consideramos interessante que, além de demonstrarmos a produção literária produzida desde o início deste doutorado, na sequencia demonstramos também em destaque três produções científicas que são o resultado da aplicação da metodologia proposta neste estudo em outras áreas.

No apêndice F, é mostrado a submissão de um artigo cujo o objetivo é a determinação do início e fim de eventos de apneias e hipopnéias em exames de polissonografia com a aplicação das Wavelets de Haar.

Polissonografia é um teste multiparamétrico que é utilizado no estudo do sono e de suas variáveis fisiológicas. Nosso estudo estabelece um novo método de análise de sinais utilizando-se modelos matemáticos avançados, o que pode possibilitar um diagnóstico mais rápido e assertivo por parte dos profissionais de saúde. O uso de ferramentas matemáticas (análise de dados, séries temporais e inteligência artificial) é proposto de forma metódica. A proposta foi estabelecer um método novo de diagnóstico e análise, através da aplicação de wavelets, em conjunto com um novo software, trazendo uma significativa economia de tempo e desempenho computacional. Neste estudo, conseguimos determinar, através da aplicação de modelos matemáticos avançados, a marcação de eventos respiratórios conhecidos como apnéia e hipopnéia em exames de polissonografia. A taxa de acurácia chegou a 97 por cento, um valor aceitável se pensarmos que utilizamos parâmetros conservadores para a determinação dos gatilhos de disparos na marcação do início e fim do evento respiratório.

No apêndice G, é mostrado a carta de aceite de um artigo gerado pelo conteúdo do estudo desse doutorado, que foi apresentado junto a comunidade científica. Assim, o uso de todas essas ferramentas (análise de dados, séries temporais e inteligência artificial) disponíveis atualmente, são ferramentas poderosas para uma análise mais precisa, permitindo assim uma melhor qualidade de diagnóstico. Com este experimento concluído, chegamos ao objetivo de definir qual a melhor família de wavelet que apresentou melhor resultado. Assim, famílias que tiveram taxas inferiores a 70 por cento de no valor de F1 foram excluídas, e as que não passaram no teste de Tukey também foram excluídas.

Também devemos notar que algumas famílias não possuem níveis de decomposições iguais, ou seja, suas derivações. Este é um fator que deve ser considerado, pois quanto mais a quantidade de derivações mais dados estatísticos podemos obter.

Neste apêndice apresenta-se a carta de aceite de outro artigo, cujo objetivo do estudo é a classificação de um banco de dados de imagens com base na regularização de coeficiente wavelet de regressão de Ridge sobre EEG.

Este artigo apresenta um protocolo de teste de Apresentação Visual Serial Rápida (RSVP) usando P300 Evoked Related Potential (ERP) via EEG comparando diferentes famílias e ordens de wavelets para cada indivíduo para avaliar os melhores resultados de classificação. A taxa média de classificação, obtida em 106 wavelets e onze sujeitos divididos em dois grupos, foi 97% de acurácia para wavelet “daubechies” de ordem “32”. O trabalho também apresenta resultados significativos de classificação de sinais cerebrais para aplicação em bancos de imagens, uma comparação de aplicação entre diferentes tipos de grupos de usuários (especialistas e não especialistas) e diferentes bancos de imagens de pesquisas científicas.

Durante o estudo desse doutorado, também foram produzidas as seguintes produções literárias elencadas abaixo:

- a) Journal Avant. A New Biomarker in Diagnostic in Spirometry Exams with the Application of Wavelets <http://dx.doi.org/10.15377/2409-5761.2018.05.4>
- b) Livro Completo Application the wavelets in engineering ISBN-13: 978-3639665093
- c) Capitulo de Livro "A NEW BIOMARKER FOR ALZHEIMER'S BASED ON THE HIPPOCAMPUS IMAGE THROUGH THE EVALUATION OF THE SURFACE CHARGE DISTRIBUTION" ISBN: 9780429446030
- d) Artigo em revista "Statistic predictive modeling financial market using Haar wavelet" DOI: 10.15413/jbem.2018.0139
- e) Projeto Fapesp/Pipe Desenvolvimento de um equipamento de análise pulmonar - Espirometro (Viabilidade Técnica) Processo 2018/19376-4

**APÊNDICE F – SUBMISSÃO DE ARTIGO PUBLICADO SOBRE POLISSONOGRAFIA**



# An innovative methodology to determine the onset and termination events of apnea and hypopnea in polysomnography exams

Rodrigo Galuzzi Garcia Piva<sup>1\*</sup>, Rodrigo Bechelli<sup>1†</sup>, Rudolf Theoderich Bühler<sup>1†</sup>, Ariane Soares do Nascimento Pereira<sup>2†</sup>, André Luiz Perin<sup>1†</sup> and Renato Camargo Giacomini<sup>1†</sup>

<sup>1\*</sup>Department of Electrical Engineering, Centro Universitário FEI, Av. Humberto de A. C. Branco, 3.972, São Bernardo do Campo, 09850–901, São Paulo, Brasil.

<sup>2</sup>Science and Technology Department, ISPGaya, Av. dos Descobrimentos, 333, Vila Nova de Gaia, 4400-103, Porto, Portugal.

\*Corresponding author(s). E-mail(s): [rggaluzzi@fei.edu.br](mailto:rggaluzzi@fei.edu.br); Contributing authors: [rprior@bechelli.org](mailto:rprior@bechelli.org); [buhler@fei.edu.br](mailto:buhler@fei.edu.br); [aspereira@ispgaya.pt](mailto:aspereira@ispgaya.pt); [aperin@fei.edu.br](mailto:aperin@fei.edu.br); [renato@fei.edu.br](mailto:renato@fei.edu.br);

<sup>†</sup>These authors contributed equally to this work.

## Abstract

Polysomnography is a multi-parametric exam used to study sleep disorders and its physiological variables. This study establishes a new method of analyzing signals using updated mathematical tools, which can enable a faster and more accurate diagnosis results by health professionals. Through the application of Haar wavelets, used to represent the oronasal airflow signal from polysomnography exams, well-defined parameters were obtained and used to classify and generate markings of beginning and end of each apnea and hypopnea event on those exams. Based on this data collection and on the various parameters developed, an automatized algorithm was introduced and implemented, capable to support all the necessary steps to perform the characterization of apnea and hypopnea events.

**APÊNDICE G – CARTA DE ACEITE DE ARTIGO PUBLICADO SOBRE O TEMA**

## ACCEPTANCE LETTER

Dear

Rodrigo Galuzzi Garcia Piva, Rodrigo Prior Bechelli, Rudolph Buhler and Renato Giacomini

**Manuscript No.**

**HJAMR-173-2022**

**Title:** A new biomarker to assist in the interpretation of spirometry exams through the application of artificial intelligence and wavelet functions

Dear

Sir/Madam

On behalf of **Himalayan Journals**, I Samim Ahmed (**Managing Director of Himalayan Journals**) takes pride and pleasure to inform you that our Reviewer has reviewed your article and recommended your Article for publication. In our Upcoming issue (**HJAMR: Volume-3: Issue-1; January-February, 2022**).

Journal:

**Himalayan Journal of Applied Medical Sciences and Research**

**ISSN Print :** 2709-3506 | **ISSN Online :** 2709-3514

**Frequency :** Bi-Monthly

**Language :** English

**Origin :** Kuwait

**Website:** <https://www.himjournals.com/journal/hjamr>

**Indexing:-**

Google Scholar, Index Copernicus, Research Bible, World Cat, Eurasian Scientific Journal Index (ESJI) Citefactor, SHERPA/RoMEO, Scientific Indexing Services (SIS), Road- Directory of Open Access Scholarly Resources, Directory of Research Journals Indexing (DRJI) and others in Progress



*Samim Ahmed*

**Samim Ahmed**  
Managing Director

**APÊNDICE H – CARTA DE ACEITE DE ARTIGO PUBLICADO SOBRE APLICAÇÃO DE  
WAVELETS EM EXAMES DE EEG**

## ACCEPTANCE LETTER

Dear

Rodrigo Galuzzi Garcia Piva, Rodrigo Prior Bechelli, Rudolf Theoderich Böhler, and Renato Giacomini

**Manuscript No.**

**HJAMR-172-2022**

**Title:** Classification a Image Database based on Ridge Regression Wavelet Coefficient Regularization over EEG

Dear

Sir/Madam

On behalf of **Himalayan Journals**, I Samim Ahmed (**Managing Director of Himalayan Journals**) takes pride and pleasure to inform you that our Reviewer has reviewed your article and recommended your Article for publication. In our Upcoming issue (**HJAMR: Volume-3: Issue-1; January-February, 2022**).

Journal:

**Himalayan Journal of Applied Medical Sciences and Research**

ISSN Print : 2709-3506 | ISSN Online : 2709-3514

Frequency : Bi-Monthly

Language : English

Origin : Kuwait

Website: <https://www.himjournals.com/journal/hjamr>

**Indexing:-**

Google Scholar, Index Copernicus, Research Bible, World Cat, Eurasian Scientific Journal Index (ESJI) Citefactor, SHERPA/RoMEO, Scientific Indexing Services (SIS), Road- Directory of Open Access Scholarly Resources, Directory of Research Journals Indexing (DRJI) and others in Progress



*Samim Ahmed*

**Samim Ahmed**  
Managing Director