

Luiz Gustavo Caobianco

# **Caracterização Computacional de Patologias Laríngeas Analisadas Acusticamente com Elementos Energéticos**

Monografia apresentada ao Departamento de Ciências de Computação e Estatística do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos necessários para aprovação na disciplina Projeto Final.

Orientador: Prof. Dr. Rodrigo Capobianco  
Guido

São José do Rio Preto - SP

2016

Luiz Gustavo Caobianco

# **Caracterização Computacional de Patologias Laríngeas Analisadas Acusticamente com Elementos Energéticos**

Monografia apresentada ao Departamento de Ciências de Computação e Estatística do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos necessários para aprovação na disciplina Projeto Final.

---

Prof. Dr. Rodrigo Capobianco Guido

---

Luiz Gustavo Caobianco

Banca Examinadora:

Prof. Dr. Aleardo Manacero Junior

Prof. Dr. Geraldo Francisco Donegá

Zafalon

São José do Rio Preto - SP

2016

Dedico a minha família e amigos.

# Agradecimentos

Este trabalho não seria possível sem o apoio imensurável de pessoas que participaram do meu cotidiano durante todo o processo pelo qual passei. A priori, agradeço a todos que me deram algum tipo de incentivo, e fizeram com que eu acreditasse que era possível realizar o sonho de ingressar na UNESP.

Começo meus agradecimentos dedicando meu trabalho a minha irmã, Ana Paula. Desde minhas primeiras lembranças, sua companhia foi indispensável, e pra sempre será. Além disso, quero que saiba que todas as conquistas por mim realizadas devem a você uma grande porcentagem de todo o esforço. É fácil seguir um caminho depois que alguém já o desbravou. Quando cheguei a pensar que não conseguiria desenvolver esse projeto devido aos inúmeros problemas encontrados ao longo do desenvolvimento, pude contar com seu apoio. Que fique registrado que este trabalho só foi encerrado graças a seu apoio. Serei eternamente grato.

Dedico esse trabalho aos meus pais, Paulo e Lúcia, que foram os responsáveis por tudo que sou hoje. Ingressar na UNESP foi fácil pois tive a grande sorte de ter vocês, que trabalharam incessantemente para que eu tivesse uma formação diferenciada. Sei que nunca agradecerei o suficiente o sacrifício feito por vocês para que eu pudesse subir esse degrau, mas quero que saibam que empreguei tudo que aprendi até hoje neste trabalho, e espero que tenha ficado a altura de pessoas excepcionais como vocês. Quero, do fundo do coração, agradecer a vocês e à Ana, pois sei que sem vocês nada disso seria possível.

Grande parte desse trabalho devo ao meu orientador, professor Rodrigo Capobianco Guido, que soube me guiar na odisséia que é o desenvolvimento de um trabalho de conclusão de curso. Agradeço não só pela orientação, mas pelo modelo que foi para mim. Neste meu sonho de um dia ser professor, espero que consiga alcançar uma pequena porcentagem do que o senhor é. Se esse sonho for realizado é porque tive alguém em quem me espelhar. Agradeço por ter sido um grande mestre, e espero que este trabalho tenha ficado a altura de um grande professor como o senhor.

*“The only thing we have to fear is fear itself”*

**Franklin D. Roosevelt**

# Resumo

O objetivo deste trabalho é especificar e implementar um algoritmo para o pré-diagnóstico das patologias laríngeas edema de Reinke e nódulos nas pregas vocais, sendo que a estratégia particular consiste em examinar pacientes por meio da análise dos respectivos sinais de voz digitalizados. Um total de 42 pacientes foram analisados pelo sistema desenvolvido, entre saudáveis e patológicos, implicando na acurácia global de 52.38% que, apesar da baixa, identificou com 100% de precisão os pacientes patológicos, lançando apenas falsa suspeita naqueles verdadeiramente saudáveis. Assim, frente à expectativa inicial de identificação plena dos pacientes afetados patologicamente, os objetivos foram plenamente satisfeitos.

Palavras-chave: Processamento de Voz. Patologias no Trato Vocal. Reconhecimento de Padrões.

# Abstract

This work aims to specify and to implement an algorithm for the prediagnosis of voice pathologies, namely Reinke's edema and nodules in the vocal folds, being the specific strategy the analysis of the digitized signals from the subjects. A total of 42 voices, considering healthy (HS) and pathologically-affected (PA) people, were analysed, conducting to 52.38% of global accuracy. Despite of being modest, the system correctly classified all the PA subjects, misclassifying just the HS. Thus, upon the initial perspective of a perfect evaluation for PA people, the author figure out that his objective was reached.

**Keywords:** Speech processing. Vocal-tract pathologies. Pattern Recognition.

# Lista de Figuras

Figura 2.1 - Corte transversal de uma laringe, adaptado de [5] . . . . .	8
Figura 2.2 - Sinal de voz, visualizado no aplicativo <i>Audacity</i> . . . . .	8
Figura 2.3 - O formato de sinal <i>wav</i> , retirado de [15]. . . . .	10
Figura 3.1 - Ondas de voz . . . . .	16
Figura 3.2 - Gráfico de valores dos pacientes com edema, gerado através da média processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda. . . . .	17
Figura 3.3 - Gráfico de valores dos pacientes com nódulo, gerado através da média processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda. . . . .	17
Figura 3.6 - Gráfico de valores dos pacientes com nódulo, gerado através da variância processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda. . . . .	18
Figura 3.4 - Gráfico de valores dos pacientes saudáveis, gerado através da média processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda. . . . .	18
Figura 3.5 - Gráfico de valores dos pacientes com edema, gerado através da variância processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda. . . . .	18
Figura 3.7 - Gráfico de valores dos pacientes saudáveis, gerado através da variância processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda. . . . .	19
Figura 4.1 - Gráfico de tempo de execução, no eixo Y, por tamanho da amostra, no eixo X. Percebe-se a relação praticamente linear. . . . .	24
Figura 4.2 - matriz com erros falsos positivos. . . . .	28



Figura 4.3 - matriz com erros falsos negativos . . . . .	29
Figura A.1 - Valores dos sinais saudáveis, selecionados para o grupo modelo, que geraram o pior resultado . . . . .	35
Figura A.2 - Valores dos sinais saudáveis, selecionados para o grupo teste, que geraram o pior resultado. . . . .	36
Figura A.3 - Valores dos sinais de pacientes com nódulo, selecionados para o grupo modelo, que geraram o pior resultado. . . . .	37
Figura A.4 - Valores dos sinais de pacientes com nódulo, selecionados para o grupo teste, que geraram o pior resultado. . . . .	38
Figura A.5 - Valores dos sinais de pacientes com edema, selecionados para o grupo modelo, que geraram o pior resultado. . . . .	39
Figura A.6 - Valores dos sinais de pacientes com edema, selecionados para o grupo teste, que geraram o pior resultado. . . . .	40
Figura A.7 - Valores dos sinais de pacientes saudáveis, selecionados para o grupo modelo, que geraram o melhor resultado. . . . .	41
Figura A.8 - Valores dos sinais de pacientes saudáveis, selecionados para o grupo teste, que geraram o melhor resultado. . . . .	42
Figura A.9 - Valores dos sinais de pacientes com noódulo, selecionados para o grupo modelo, que geraram o melhor resultado. . . . .	43
Figura A.10 -Valores dos sinais de pacientes com noódulo, selecionados para o grupo teste, que geraram o melhor resultado. . . . .	44
Figura A.11 -Valores dos sinais de pacientes com edema, selecionados para o grupo modelo, que geraram o melhor resultado. . . . .	45
Figura A.12 -Valores dos sinais de pacientes com edema, selecionados para o grupo teste, que geraram o melhor resultado. . . . .	46

# Lista de Tabelas

Tabela 4.1 - Tamanhos dos vetores, tempos de execução medidos, tempos de execução obtidos através da complexidade do algoritmo e o erro entre o tempo esperado e o tempo obtido. . . . .	25
--	----

# Sumário

<b>Lista de Figuras</b>	<b>8</b>
<b>Lista de Tabelas</b>	<b>10</b>
<b>1 Introdução</b>	<b>2</b>
1.1 Formulação do Problema, Objetivos e Justificativa . . . . .	2
<b>2 Fundamentação Teórica</b>	<b>4</b>
2.1 Detecção não-invasiva de patologias na laringe . . . . .	4
2.2 O Aparelho Fonador . . . . .	7
2.3 O Sinal de Voz . . . . .	8
2.3.1 Digitalização do Sinal de Voz . . . . .	9
2.4 O Formato WAV . . . . .	9
2.5 Extrator de Características . . . . .	11
2.6 Classificador . . . . .	12
2.7 Matriz de Confusão . . . . .	13
2.8 Validação Cruzada . . . . .	14
<b>3 O Sistema Proposto</b>	<b>15</b>
<b>4 Testes e Resultados</b>	<b>21</b>
4.1 Base de Dados . . . . .	21
4.2 Procedimentos . . . . .	22
4.3 Complexidade do Algoritmo . . . . .	23
4.4 Primeiros Resultados . . . . .	26
4.4.1 Melhor Resultado . . . . .	26
4.4.2 Pior Resultado . . . . .	27
4.5 Otimização dos Resultados . . . . .	27

4.5.1	Classificação dos Erros . . . . .	27
4.6	Resultados Finais . . . . .	30
4.6.1	Melhor Resultado . . . . .	30
<b>5</b>	<b>Conclusões</b>	<b>32</b>
	<b>Apêndice</b>	<b>34</b>
<b>A</b>	<b>Valores gerados pelo programa</b>	<b>34</b>

# Capítulo 1

## Introdução

### 1.1 Formulação do Problema, Objetivos e Justificativa

A voz é um dos instrumentos mais presentes na vida de todos, seja como ferramenta de trabalho para muitas profissões ou como simples meio de comunicação entre todos. A produção de um sinal de voz envolve diferentes órgãos do corpo humano, como pulmão, laringe, faringe, dentre outros. Algumas anomalias, causadas por diversas razões como mau uso ou motivos neurológicos, podem ocasionar distorções ou até mesmo perda da voz, principalmente quando o órgão afetado é a laringe.

Por muito tempo, patologias como edemas e nódulos foram identificadas por meio de exames como videolaringoscopia ou videoestroboscopia, procedimentos que podem ser complicados para pacientes que têm muito reflexo ao toque na garganta. Ademais, tais exames são considerados invasivos e demandam a presença de equipamentos sofisticados para a captura da imagem. Registre-se, ainda, a apreensão que muitos indivíduos possuem em investigar a sua saúde e descobrir uma patologia, visitando assim os profissionais da área mediante quadros graves e que tornam-se irreversíveis na maioria das vezes.

Tendo em vista que o Processamento de Sinais Digitais (PDS) pode auxiliar na análise de sinais de voz, fazendo uso de sistemas computacionais para a procura de padrões, o objetivo do presente trabalho é o de, após realizar um estudo dos conceitos pertinentes, elaborar e implementar um algoritmo para o pré-diagnóstico de patologias na laringe, evitando assim a

execução de exames invasivos. Em relação aos exames realizados por otorrinolaringologistas, as técnicas de PDS apresentam vantagens por não necessitar de dispositivos altamente sofisticados. É necessário apenas um microfone, para obtenção do sinal, e algum dispositivo que gerencie a captura do sinal e possa se comunicar com o *software* responsável por executar o pré-diagnóstico.

O potencial computacional tem crescido cada vez mais. Até mesmo os *smartphones* vendidos hoje possuem poder de processamento capaz de executar algoritmos como o proposto neste trabalho. Métodos que utilizam o PDS oferecem uma alternativa mais barata, e talvez o aspecto mais importante, mais portátil quando comparados ao ferramental necessário para os exames. Esses motivos são os principais que impulsionam e motivam o desenvolvimento deste trabalho.

## Capítulo 2

# Fundamentação Teórica

*Neste capítulo, são discutidos todos os aspectos teóricos que envolvem o desenvolvimento deste trabalho, desde as patologias que estão sendo analisadas até os conceitos matemáticos empregados para extração de características e classificação dos sinais de voz.*

### 2.1 Detecção não-invasiva de patologias na laringe

Costa [1], em sua tese de doutorado, realizou a caracterização e classificação de sinais de vozes normais e patológicas<sup>1</sup> por meio da quantificação de recorrência e análise dinâmica não linear. O autor utilizou-se destes métodos fazendo validação cruzada e obteve acurácia que variou entre 95.44% e 100% para vozes saudáveis e patológicas; 100% para vozes saudáveis e afetadas por paralisia; e na faixa de 94,75% e 100% para vozes saudáveis e afetadas por edema, assim como entre saudáveis e nódulos. Estes resultados mostraram que o método aplicado pelo autor é eficaz, e pode ser utilizado no desenvolvimento de sistemas cujos âmbitos coincidem com os deste trabalho.

Viera *et al* [2] discutem a utilização de vários métodos para a classificação de sinais de voz saudáveis e afetados por patologias laríngeas. Na discriminação entre os sinais, o método que mais se destacou foi o comprimento médio das linhas diagonais. Na classificação entre as pa-

---

<sup>1</sup>Ainda que os termos “voz normal” e “voz patológica” possam parecer estranhos, a literatura da área adota-os comumente.

tologias analisadas pelos autores, o melhor resultado foi encontrado quando aplicado o método tempo de recorrência do “tipo 2.” Os autores mostram os resultados, divididos em duas classes de sinal: longo intervalo de tempo e curto intervalo de tempo. As patologias analisadas foram paralisia, edema e nódulo. A acurácia encontrada pelo autor, quando comparado sinais saudáveis e patológicos, permaneceu no intervalo de 88,27% a 94,85%. Na comparação entre saudável e paralisia, a precisão foi de 88,65% a 96,27%. Quando comparado saudável e edema, a acurácia encontrou-se na faixa de 80,96% a 92,33%. Entre saudável e nódulo, o autor encontrou uma precisão de 90,82% a 96,00%. Todos os menores valores destes intervalos de precisão citados foram encontrados quando o autor utilizou o curto intervalo de tempo.

*Souza et al*[3], dividiu sua pesquisa em dois métodos. No primeiro, denominado método das sub-bandas, o autor aplicou 12 descritores de textura sobre cada sub-banda obtida, totalizando para cada sinal de sinal 48 características a serem analisadas. No segundo, denominado método dos descritores, o autor extraiu 13 descritores de Haralick, obtendo 52 características para analisar. Os melhores resultados foram alcançados empregando redes neurais *Multilayer Perceptron* (MLP) em conjunto com o algoritmo *Particle Swarm Optimisation*, obtendo um intervalo de acurácia entre 79,43% e 91,27%, provando que esses métodos também são válidos para aplicação na área deste trabalho.

*Arias-Londono et al*[4] extraiu inúmeras características em seu trabalho, dentre elas diversos tipos de entropia: aproximada, da amostra, aproximada de *kernel* Gaussiano, do processo de Markov, além de utilizar validação cruzada. Em seu trabalho, comprovou-se que sistemas automáticos podem chegar a uma considerável acurácia. Publicado em 2011, o autor alcançou uma acurácia de classificação entre sinais patológicos e saudáveis de 98,23%.

*Godino et al*[13] utilizou vetores multidimensionais de características com análise a curto intervalo de tempo. Neste artigo, os autores empregam a *Distribuição F de Fisher-Snedecor* e a *Discriminante Linear de Fisher*, e mostram que é possível utilizar tanto vetores de coeficiente *mel cepstral*[17], quanto a primeira derivada deles. Uma abordagem para encontrar estes coeficientes é através da Transformada Rápida de Fourier não-parametrizada. Os autores ignoraram a segunda derivada, utilizando portanto apenas a derivada de ordem 1 nos vetores de coeficiente. Além disso, também foram utilizados 6 modelos de mistura gaussiana. Neste projeto, o autor



utilizou 24 coeficientes melceptrais, e a precisão encontrada foi de 94,07%, com uma margem de erro de 3,28%.

*Wang et al*[14] utilizou o mesmo processo de [13], fazendo uso de coeficientes melceptrais e modelos de mistura gaussiana. Porém, neste artigo, foram utilizados 16 coeficientes melceptrais, enquanto que em [13] foram utilizados 24. A quantidade de modelos de mistura gaussiana também diferiu. Neste projeto, junto aos 16 coeficientes melceptrais, os autores utilizaram 16 modelos de mistura gaussiania, sendo que em [13] apenas 6 haviam sido utilizados. A precisão deste projeto encontra-se na faixa de 96,1%, com margem de erro de 2,51%.

*Costa et al*[16] empregeram a análise de quantificação de recorrência, e compararam vozes patológicas e saudáveis, ao longo de um intervalo de tempo, utilizando sinais de vogal sustentada /a/. Neste projeto, os autores combinaram a Entropia de Shannon com 12 coeficientes LPC no classificador Quadratic Discriminant Analysis, ou análise discriminante quadrática. Este processo levou os autores à uma precisão de 94,19%, com margem de erro de 0,87%.

*Henríquez et al*[18] discutiram 6 medidas caóticas não-lineares, baseadas na teoria dinâmica não-linear, com duas categorias de sinal: saudáveis e patológicos. As medidas deste estudo foram as entropias de Rényi, tanto de primeira quanto de segunda ordem. Os autores também avaliaram a entropia de Shannon. Quanto ao classificador, os autores empregaram um método baseado em rede neural padrão. Dentre todos os sinais testados, de duas bases que foram utilizadas, a precisão foi de 82,47% a 99,69%.

*Boyanov et al*[19] utilizaram uma base de sinal própria, onde 400 pessoas foram gravadas em uma câmara anecoica, e o sinal gravado foi a vogal /a/ sustentada. Cada paciente foi gravado uma vez, periodicamente de quatro em quatro meses, em um total de 3 gravações. Portanto, a base utilizada por estes autores consistia de 1200 sinais, sendo 3 sinais por paciente. Nestes sinais, foram aplicados médias e desvios padrão, e utilizando o Teste T de Student, o trabalho desenvolvido encontrou uma precisão de 89% a 97%.

## 2.2 O Aparelho Fonador

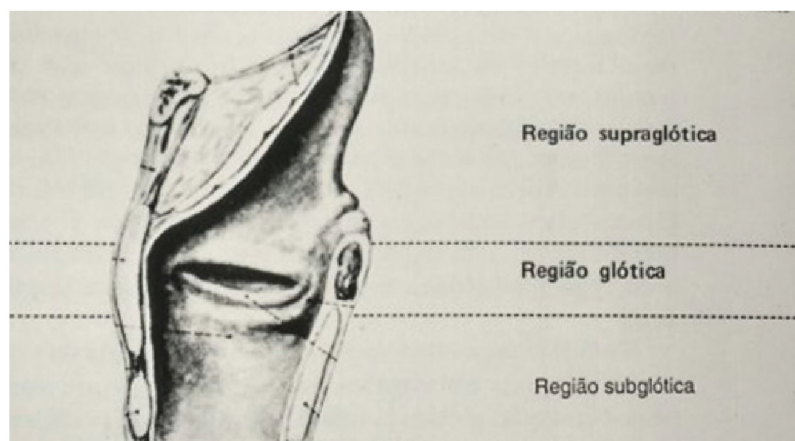
O processo de fala envolve diferentes órgãos do corpo humano, e é um movimento sincronizado entre eles que resulta na produção do sinal analógico de voz. Esse processo envolve a respiração, quando o pulmão é preenchido de ar, e ao ser liberado, passa pela laringe, onde se encontram as pregas vocais, e então pelo trato vocal, conforme em [5]: “[...] a laringe é que vai produzir o som para ser articulado na faringe e na boca e, desta forma, ter o meio de comunicação tão necessário, que é a linguagem”.

Infecções na laringe podem ser causadas por inúmeros motivos. Elas são mais comuns entre indivíduos do sexo masculino e adultos. Quando trata-se de indivíduos que abusam do uso da voz ou são tabagistas, a incidência é ainda maior. Outros fatores que parecem causar laringite crônica são:

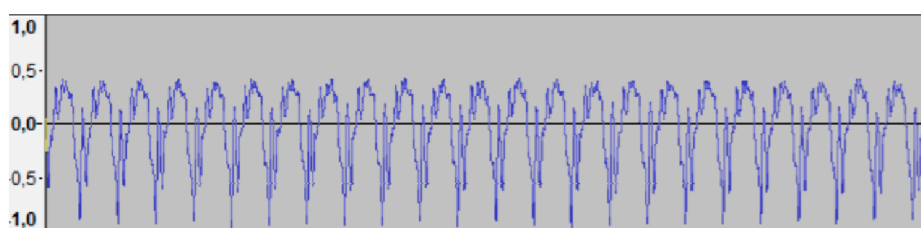
- repetidas infecções de vizinhança;
- uso indevido da voz, ou em condições adversas;
- álcool e fumo;
- contato frequente com ambientes poluídos.

Esta etiologia pode ser encontrada no texto [5]. O principal sintoma relatado pelos pacientes acometidos por essa patologia é a disfonia ou rouquidão. As modificações na voz são causadas pela necessidade de maior esforço para movimentar a prega vocal. Além disso, o doente com laringite crônica apresenta uma voz melhor pela manhã, e que se agrava durante o dia pela tosse e pelo uso da voz. A dor da laringe, dispnéias e disfagia são mais raros, e tipicamente aparecem como resultado de outros problemas, como quando o processo inflamatório produz miosites ou artrite[6].

Uma laringe simplificada pode ser vista na figura 2.1. Para o intuito deste trabalho, uma descrição profunda da laringe e seus componentes não será útil, e por isso a figura 2.1 representa apenas uma descrição geral.



**Figura 2.1** – Corte transversal de uma laringe, adaptado de [5]



**Figura 2.2** – Sinal de voz, visualizado no aplicativo *Audacity*.

## 2.3 O Sinal de Voz

A voz é um sinal que, em sua origem não processada, é analógico. Isso significa que é possível representar a voz em um gráfico cujo os valores de abscissa representam um determinado instante no tempo. O problema é que, na sua essência, sinais analógicos contêm infinitos valores dentro de um intervalo. possível ver na figura 2.2 um sinal de voz cujos valores de amplitude são utilizados para desenhar a onda ao longo de um determinado tempo.

Esse problema conflita com a característica intrínseca dos computadores de incapacidade de processar dados de tamanho infinito. Portanto, técnicas devem ser aplicadas para que um sinal analógico, no caso a voz, seja então passível de ser operado nos computadores. As técnicas aplicadas para a digitalização desse sinal, ou seja, para que esse sinal seja grande o suficiente de maneira que não perca suas características, mas que não seja infinitamente grande, serão discutidas na seção 2.3.1

### 2.3.1 Digitalização do Sinal de Voz

Como dito na seção 2.3, o sinal de voz precisa passar, antes de qualquer método matemático, por métodos para transformá-lo em um sinal digital. Nessa seção, será discutido os dois processos pelo qual o sinal deve passar para poder ser considerado digital: discretização e quantização.

Para discretizar o sinal, primeiro é necessário entender quantas amostras são necessárias neste processo. A definição 1 mostra qual deve ser a taxa mínima de amostragem que um sinal pode ter sem que perca suas características.

**Definição 1.** *Considerando um sinal de frequência  $B$ , a quantidade de amostras por tempo,  $f_s$ , deve ser:*

$$f_s > 2B \quad [7][8]$$

Após o processo de amostragem, o sinal agora pode ser representado por uma quantidade finita de pontos, definida pelo teorema 1. Porém, os valores destes pontos ainda possuem infinitas casas decimais, oriundos da característica analógica do sinal.

Para sanar este problema, é aplicado o processo de quantização. Este processo consiste em dividir a amplitude do sinal analisado em vários intervalos de mesmo tamanho. Estes intervalos serão usados para auxiliar no arredondamento das amostras. Deve-se, amostra por amostra, arredondar o valor de uma amostra que tem muitas casas decimais para o limite inferior do intervalo em que a amostra está contida.

Feito isso, o sinal encontra-se digitalizado, e então é capaz de ser processado por computadores digitais, e pode-se garantir que suas características foram preservadas.

## 2.4 O Formato WAV

Um arquivo que esteja no formato *.wav* obedece à uma estrutura particular que pode ser separada em dois campos: *headers* e *raw data*. Como é possível ver na figura 2.3, o formato



**Figura 2.3** – O formato de sinal wav, retirado de [15].

é dividido em muitas outras categorias, entretanto o foco será o campo *data*. Os outros campos possuem informações gerais acerca do sinal digitalizado, conforme detalhado em [10].

O campo *data* é fundamentalmente importante pois seus valores são importados para o programa, permitindo então aplicar os métodos matemáticos da seção ???. Estes valores são números inteiros que correspondem às amplitudes digitalizadas. Assim, é preciso garantir que estes valores foram corretamente importados para o programa, pois qualquer erro nesta fase prejudicaria todo o processo aplicado, invalidando os resultados do projeto.

Este formato armazena os dados em blocos (*chunks*, na figura 2.3), e é uma variação do formato RIFF. Um dos problemas, causados pelo uso de inteiros de 32 bits para gravar o tamanho do arquivo, é o tamanho máximo. Arquivos WAV não podem ultrapassar 4GB, porém esta limitação não oferecerá problemas ao desenvolvimento do projeto, uma vez que os sinais duram em torno de 5 segundos.

Como é possível ver na figura 2.3, o arquivo é dividido em um *subchunk*, em tradução livre sub-bloco, FMT e um *subchunk* DATA. O sub-bloco FMT é o responsável por descrever o formato das informações encontradas no sub-bloco DATA. Os campos da figura 2.3 serão brevemente explicados:

- *ChunkID* Contém as letras RIFF em formato ASCII;

- *ChunkSize* É o tamanho de todo o arquivo em bytes, excluindo-se os 8 bytes dos campos não inclusos: *ChunkID* e *ChunkSize*;
- *Format* Contém as letras WAVE;
- *Subchunk1ID* Contém as letras “fmt”;
- *Subchunk1Size* É o tamanho do resto do sub-bloco;
- *AudioFormat* Usado para mostrar se o sinal está com alguma forma de compressão de dados;
- *NumChannels* Define se o sinal é Mono, Stereo;
- *SampleRate* Taxa de amostras;
- *ByteRate* Este campo é definido pela fórmula:  $(SampleRate * NumChannels * BitsPerSample)8$ ;
- *BlockAlign* Número de bytes, para uma amostra, incluindo todos os canais;
- *BitsPerSample* Número de bits que constitui cada amostra;
- *Subchunk2ID* Contém as letras “data”;
- *Subchunk2Size* Número de bytes do campo *data*;
- *Data* Os valores brutos dos dados, e são os dados sobre os quais os métodos serão aplicados.

## 2.5 Extrator de Características

O extrator de características é uma ferramenta que será utilizada no projeto. Esta ferramenta consiste em receber uma grande quantidade de dados que podem ser informações de diferentes origens, por exemplo, imagens e sinais acústicos. Neste projeto, as informações providas ao extrator são os dados provenientes do campo *Data*, da figura 2.3.

A extração é feita para que o classificador receba apenas uma quantidade reduzida de dados. Além de reduzir e fixar a quantidade das informações, um aspecto importante da extração de característica é a escolha da característica a ser extraída. Diferentes características podem ser extraídas de um sinal acústico, e o extrator aqui construído irá utilizar características energéticas.

Uma vez que o extrator de características encerra seu processo, todas as informações que foram utilizadas como base para a produção destas características podem ser desprezadas, uma vez que o classificador irá operar diretamente nas informações geradas pelo extrator. Desta maneira, o extrator, além de diminuir a quantidade de dados brutos e ser o meio no qual a extração da característica escolhida é feita, também prov, ao classificador, todos os dados necessários para que o classificador possa atuar.

O processo de extração é feito para que todos os sinais tenham uma característica em comum, que seja mensurável, e independente de fatores externos. Por exemplo, sinais podem ser gravados por homens e mulheres, e claramente tm muitas diferenças. Porém, aplicando-se um extrator, ambos os sinais podem ser comparados com uma mesma característica, seja ela qual for.

## **2.6 Classificador**

Um classificador é uma das principais partes de um processo onde algo precise ser identificado, seja ele o estilo música, o idioma. Desta maneira, o classificador tem como dados de entrada características extraídas no extrator, e sua saída é a classificação daquilo que está sendo testado.

Neste projeto, o classificador é a ferramenta responsável por classificar todos os sinais. Um classificador consiste de métodos matemáticos que serão aplicados aos valores providos pelo extrator de características. Neste processo, o classificador irá comparar informações entre os sinais que estão sendo testados e aqueles que foram definidos como modelos.

Estes métodos matemáticos dependem diretamente da abordagem decidida. Diversas técnicas podem ser utilizadas, e entre as técnicas mais comuns, estão distância Euclidiana e rede neurais. Tais técnicas são responsáveis por definir o que está sendo analisado. Ou seja, em um programa cujo objetivo é discernir vozes saudáveis de patológicas, o classificador é o responsável pelo veredito: o sinal analisado é saudável ou patológico.

Então, o classificador é construído no âmbito de identificar, ou classificar, o que está sendo analisado pelo programa, e deve retornar um parecer.

## **2.7 Matriz de Confusão**

Uma matriz de confusão é uma matriz quadrática, de ordem  $N$ , onde  $N$  é a quantidade de classes que o classificador pode escolher para aquilo que está sendo testado. Esta matriz inicia preenchida de zeros.

As linhas da matriz de confusão são as classificações originais, ou seja, as classificações às quais as informações analisadas realmente pertencem. Quando a informação que está sendo testada pertence a uma certa classe, deve-se incrementar um a linha correspondente a mesma classe.

As colunas de uma matriz de confusão representam a classificação dada pelo programa. Ou seja, quando o classificador encerra, ele deve classificar a informação testada em alguma das classes possíveis. Cada linha representa uma classe, e quando o classificador decide a qual classe pertence aquela informação, deve-se então incrementar o valor da coluna correspondente a classe escolhida pelo classificador.

Desta forma, a diagonal principal desta matriz corresponde aos acertos do classificador, e é importante maximizar estes valores. As outras posições na matriz são os erros, e, no caso de reconhecimento de patologias, estes erros podem ser classificados em 2 categorias: falso



positivo e falso verdadeiro.

## **2.8 Validação Cruzada**

O processo de validação cruzada é muito utilizado quando temos dois grupos de informação: modelos e testes. O processo de validação cruzada consiste em embaralhar as informações entre estes grupos. Quando um classificador começa a executar, as informações devem ser separadas em dois grupos, para que o classificador possa escolher um sinal dentre os testes, e compará-lo com os sinais modelos.

Esta técnica contribui na fase de testes, uma vez que o teste de todas as possibilidades pode ser inviável. Essa contribuição vem do fato que, quando as informações são separadas, não se sabe qual é a melhor maneira de separá-las. Portanto, separando aleatoriamente, tem-se um primeiro conjunto de informações sobre o qual o classificador pode operar.

Ao final da classificação, a utilização da validação cruzada faz com que todas as informações que serão utilizadas pelo classificador sejam misturadas, e então sorteadas novamente, de maneira que o classificador opere em conjuntos de informações diferentes.

Portanto, quando utiliza-se validação cruzada, os grupos modelo e teste alteram frequentemente, em busca da melhor separação possível. A melhor divisão pode ser encontrada de acordo com a quantidade de acertos feitos pelo classificador quando utilizado as informações sorteadas.

A cada embaralhamento de informações, novos conjuntos de dois grupos são formados, e os acertos e erros destes conjuntos devem ser comparados, de maneira a encontrar a melhor divisão.

# Capítulo 3

## O Sistema Proposto

*Neste Capítulo, o sistema proposto está detalhado, em todos os seus passos. Sua leitura deve possibilitar compreender como o sistema foi estruturado e qual a idéia principal por trás dele.*

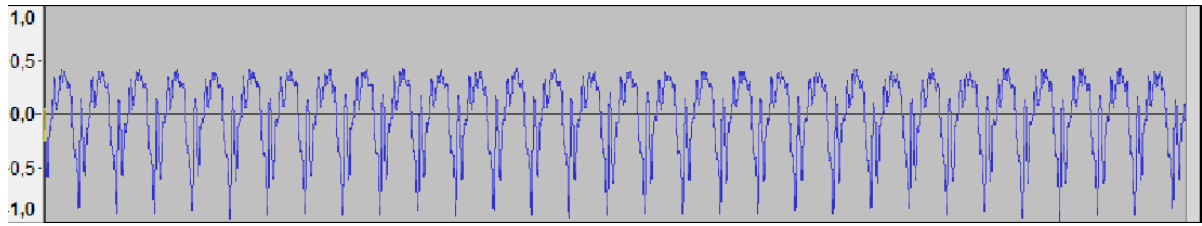
O sistema proposto neste trabalho possui os seguintes módulos que operam sequencialmente:

- módulo  $M_1$ : entrada de dados;
- módulo  $M_2$ : extração de características;
- módulo  $M_3$ : classificação;
- módulo  $M_4$ : testes.

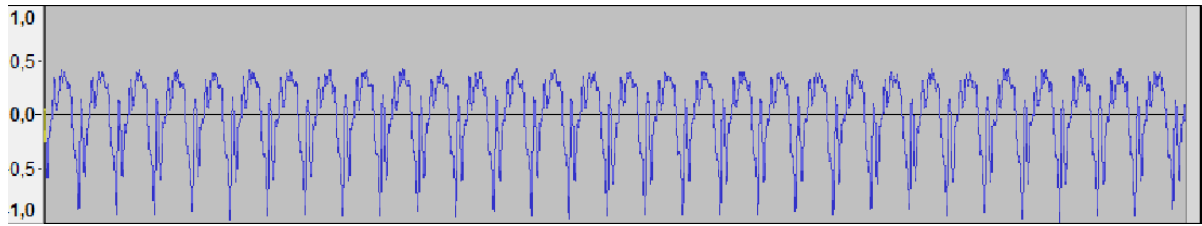
Particularmente com relação ao módulo  $M_1$ , tomou-se o cuidado de, inicialmente, garantir que os valores digitalizados dos sinais WAVE estavam sendo importados corretamente. Assim, um breve procedimento de conferência foi realizado:

1. o sinal escolhido é exportado, por meio do aplicativo *Audacity*, em um arquivo texto cujo conteúdo está representado nos campos da figura 2.3;
2. o sinal escolhido é lido com a biblioteca de *software* fornecida pelo orientador deste trabalho e os valores das amostras digitalizadas são impressos em um arquivo de texto. Apenas o *raw data*, que corresponde ao campo *data* da figura 2.3, é impresso.
3. os valores do arquivo em pleno texto gerado pelo *Audacity* são comparados com aqueles obtidos com o uso da biblioteca, um a um, garantindo que são exatamente iguais.

A título de exemplo, é possível observar na figura 3.1-(a) e 3.1-(b), que as ondas geradas são perfeitamente iguais.



(a) Onda de voz gerada com os valores extraídos pela biblioteca de *software*



(b) Onda de voz gerada no aplicativo *Audacity*

**Figura 3.1** – Ondas de voz

Uma vez que este processo garante a consistência dos dados importados pela biblioteca utilizada, esta passou a ser acrescida de um considerável volume de código-fonte em linguagem C/C++ desenvolvido por este autor, visando implementar os módulos  $M_2$ ,  $M_3$  e  $M_4$ .

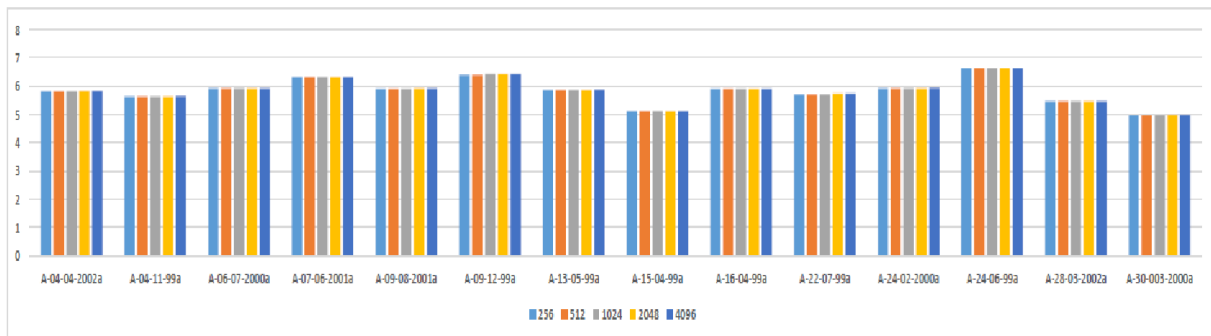
Com relação ao módulo  $M_2$ , responsável pela extração de características, o objetivo é o de homogeneizar os sinais de voz de modo que eles sejam providos para o *classificador* com a mesma quantidade de características. Assim, cada sinal de voz original, de tamanho consideravelmente alto e variável, é convertido em sinais relativamente pequenos e de tamanhos fixos, chamados vetores de características. Além disso, tais vetores contêm informações úteis para a etapa classificadora.

Particularmente,  $M_2$  foi desenvolvido de modo a calcular a energia normalizada de cada janela de tamanho  $P$  de cada sinal de voz, sobrepondo janelas consecutivas em 50%. Os valores de  $P$  utilizados, em processos independentes, foram 256, 512, 1024, 2048, 4096. Nas situações em que as amostras finais de um sinal de voz contivessem tamanhos insuficientes para caracterizar uma janela, o trecho final sofria um descarte.

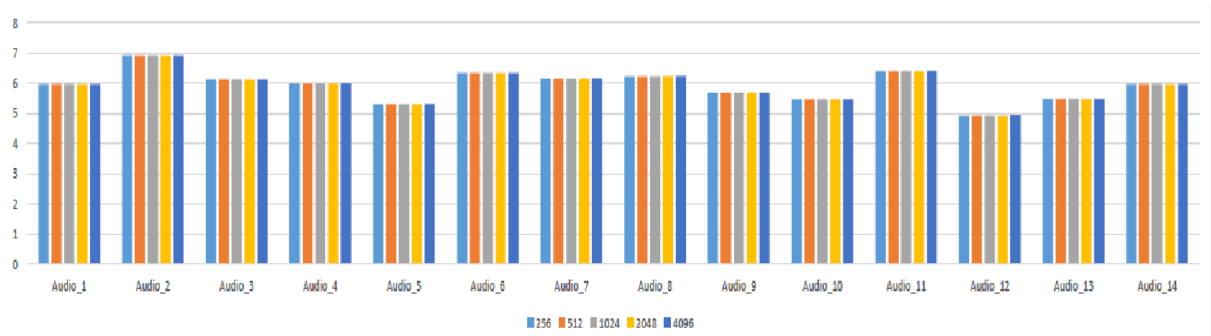
Em seguida, todas as energias normalizadas foram convertidas em dois valores, apenas:

*média e variância.* Diante do exposto, cada um dos cinco valores de  $P$  utilizados proporcionou uma média e uma variância, implicando em um vetor de características de tamanho 10. Resultados anteriores, contidos em textos referenciados na revisão bibliográfica, trazem resultados documentando que a média, e principalmente a variância, possuem potencial para indicar a presença de patologias em sinais de voz, justificando assim o uso de tais medidas estatísticas. A presença de patologias na laringe, usualmente faz com que o locutor seja impedido de sustentar a constância das características vocálicas necessárias para caracterizar um padrão normal de voz. Assim, variâncias consideravelmente altas nas energias dos sinais de voz indicam a presença de patologias.

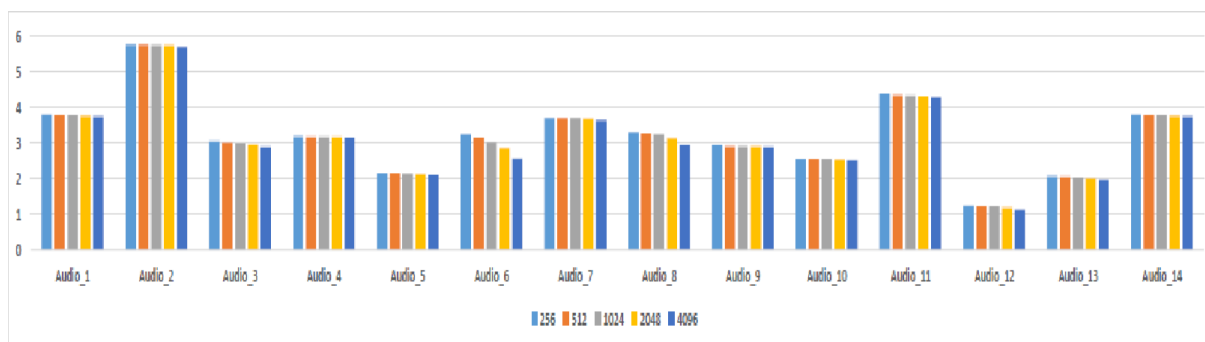
Complementarmente, os dez valores contidos nos vetores de características foram armazenados em escala logarítmica de modo a possuírem magnitudes suficientes para a visualização gráfica confortável, vindo a auxiliar na observação de padrões, conforme as figuras 3.2, 3.3, 3.4, 3.5, 3.6 e 3.7.



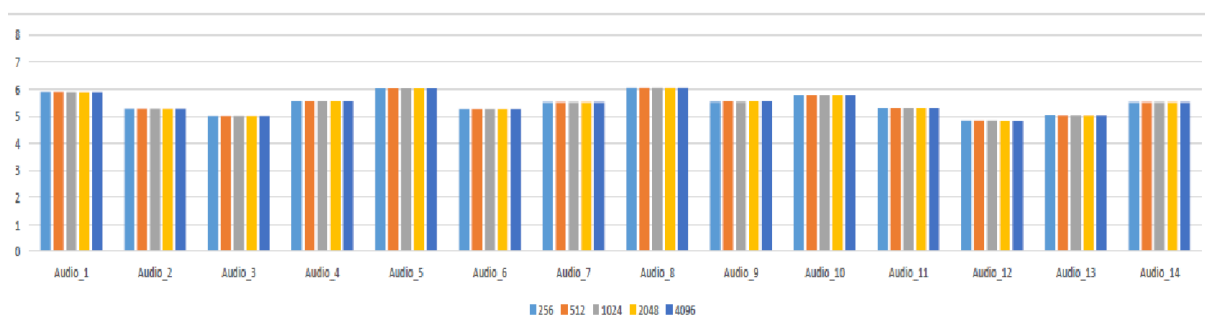
**Figura 3.2** – Gráfico de valores dos pacientes com edema, gerado através da média processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda.



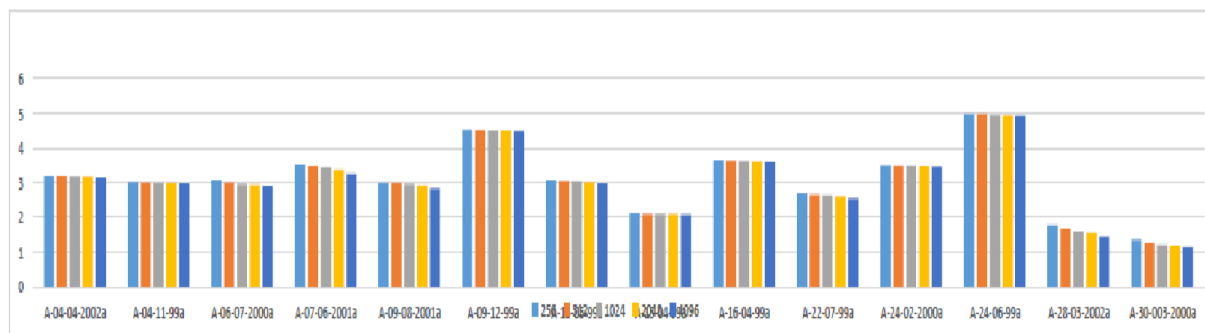
**Figura 3.3** – Gráfico de valores dos pacientes com nódulo, gerado através da média processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda.



**Figura 3.6** – Gráfico de valores dos pacientes com nódulo, gerado através da variância processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda.

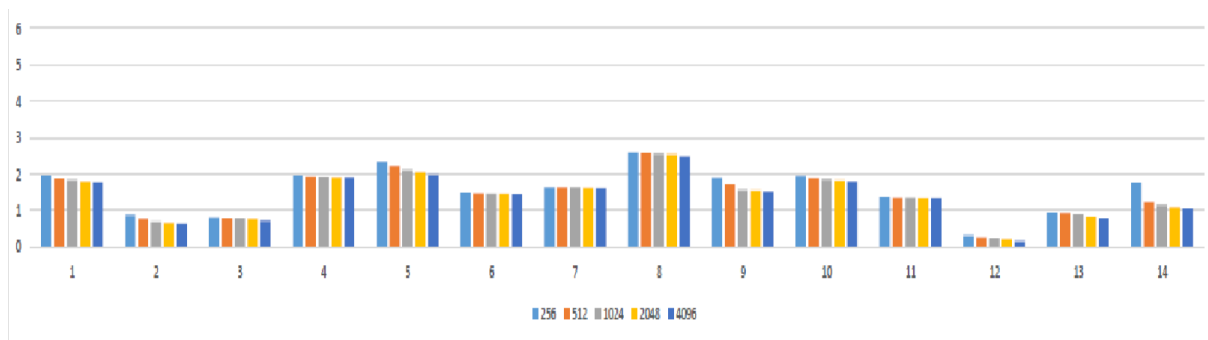


**Figura 3.4** – Gráfico de valores dos pacientes saudáveis, gerado através da média processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda.



**Figura 3.5** – Gráfico de valores dos pacientes com edema, gerado através da variância processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda.

Particularmente nas figuras 3.5, 3.6 e 3.7 é possível observar os gráficos gerados com os valores de variância processados. Porém, a análise visual possibilita uma conclusão. Ao visualizar estes valores, é possível ver na figura 3.7 que os valores de variância dos pacientes saudáveis são significativamente menores que os valores de variância dos pacientes com edema, figura 3.5, e para os pacientes com nódulo, 3.6.



**Figura 3.7** – Gráfico de valores dos pacientes saudáveis, gerado através da variância processada, agrupados por sinal, onde cada barra de uma certa cor corresponde a um tamanho de janela, de acordo com a legenda.

Em seguida, passou-se a desenvolver o módulo  $M_3$ . Para ser possível explicar o quesito de classificação aplicado, primeiro deve-se ter em mente como os vetores serão agrupados. Dentre os 42 sinais, foram sorteados aleatoriamente 7 de cada classe e, então, estes foram colocados no grupo *modelo*. Este grupo consiste dos vetores aos quais os valores analisados serão comparados. Portanto, tem-se 21 vetores: 7 deles de pacientes com nódulo, 7 com edema de Reinke e 7 saudáveis. Restaram então 21 sinais, que foram agrupados no grupo *teste*, que consiste dos vetores que serão testados.

Uma das técnicas de classificação que será aplicada é a Distância Euclidiana (DE) [12]. Esta métrica é uma medida utilizada para encontrar a distância entre dois pontos. Primeiramente, foi medida a DE de cada vetor sob teste em relação a *todos* os vetores do grupo *modelo*. Assim, deve ser encontrado o menor valor possível, que indica a maior similaridade. Quando encontrado o menor valor, deve-se então rotular o sinal que esta sendo testado como pertencente à mesma classe do sinal *modelo* cujo menor valor originou a menor DE.

Em outras palavras, a fase de teste consiste no seguinte algoritmo:

1. escolher um vetor para teste, dentre os 21 sinais do grupo *teste*.
2. encontrar a menor DE entre os valores deste vetor e *todos* aqueles do grupo *modelo*.
3. ao encontrar a menor DE, rotular o sinal respectivo: Edema, Nódulo ou Saudável.

Cada classificação executada neste algoritmo foi comparado a classificação original do sinal de teste, gerando uma matriz de confusão: se a classificação do *software* proposto for correta, deve-se acrescentar um ao valor da diagonal principal. Caso a classificação esteja errada, deve-

se incrementar um ao valor da linha correspondente.

O procedimento foi repetido para todos os sinais que se encontram no grupo de teste. Este processo faz surgir uma matriz de confusão, onde é possível analisar a quantidade de acertos e erros, globalmente. Cada matriz de confusão[11] utilizada para a verificação possui ordem 3, cada uma correspondente ao grupo respectivo: normal, nódulo e edema. Em princípio, o objetivo foi o de encontrar a matriz de confusão cuja soma dos valores da diagonal principal seja o maior possível, garantindo a maior quantidade de acertos.

Observa-se que o procedimento de validação cruzada foi aplicada com o objetivo de tentar encontrar a melhor matriz de confusão possível. Os sinais escolhidos e separados nos grupos foram aleatoriamente trocados, conforme detalhado adiante. Nota-se que, com a quantidade de sinais que foram utilizados, o processo de teste para todas as possibilidades de agrupamento dos sinais seria altamente dispendioso. Especificamente, os 42 sinais geram  $\frac{14!}{7!(14-7)!}^3$  possibilidades de combinações, o que requereu um conjunto de testes aleatórios em lugar da execução de todas as possibilidades.

# Capítulo 4

## Testes e Resultados

*Neste capítulo serão apresentados os processos empregados aos sinais. Estes processos foram aplicados, provendo ao classificador os dados de entrada necessários para o pré-diagnóstico dos pacientes analisados. Os testes feitos e suas abordagens, bem como os resultados dos mesmos, também são apresentados.*

### 4.1 Base de Dados

Conforme mencionado anteriormente, a base de sinais que foi utilizada é composta de 42 sinais. Eles estão divididos em 2 grandes grupos: patológicos e saudáveis. No grupo de sinais patológicos há duas classificações: pacientes com edema e pacientes com nódulo. Cada um desses grupos é composto por 14 sinais. Ou seja, a base de sinais que foi utilizada para os testes contém dados de 14 pacientes saudáveis, 14 pacientes com nódulo e 14 pacientes com edema.

Cada sinal consiste em um arquivo de formato WAVE. Em todos os 42 sinais, a locução de cada paciente foi dura 5 segundos. Nesse período de tempo, os pacientes sustentaram o som da vogal /a/. Quando submetidos ao processo de amostragem a 22050 am/seg e de quantização com 16 bits, cada sinal gerou aproximadamente 130 mil valores para o extrator de características.



## 4.2 Procedimentos

Os procedimentos adotados neste trabalho podem ser divididos em duas fases: sorteio dos vetores entre os grupos *modelo* e *teste* e classificação de todos os sinais sorteados no grupo *teste*.

A abordagem aleatória foi adotada para a separação dos sinais porque, como visto anteriormente, seria impraticável testar todas as possibilidades. Portanto, para cada teste, os sinais foram divididos aleatoriamente entre os grupos de modelo e de teste. Os grupos *modelo* e *teste* foram compostos por 21 sinais cada, ou seja, a base de dados foi dividida em 50% para modelos 50% para testes. Desses 21 sinais, 7 eram pacientes saudáveis, 7 pacientes com nódulos e 7 pacientes com edema.

Assim, a primeira fase de cada teste foi o sorteio dos sinais para a composição do grupo de modelos. Dentre grupo de pacientes saudáveis foram sorteados aleatoriamente 7 pacientes, para que seus sinais compusessem o grupo modelo. Feito isso, os outros 7 sinais que não foram sorteados eram automaticamente colocados no grupo de testes. Este processo repetiu-se para os 14 pacientes com nódulos e para os 14 pacientes com edemas, obedecendo ao mesmo critério: os sorteados eram colocados como modelos, e não sorteados colocados como testes.

Após a fase de sorteio, todos os dados estavam divididos entre os grupos modelo e teste, iniciando assim os testes. Nessa fase, cada sinal é representado por 10 valores de médias e variâncias processadas, providos pelo extrator de características, totalizando 210 valores (10 para cada sinal sorteado) no grupo *modelo* e 210 valores no grupo *teste*.

Para que o processo de teste seja melhor explicado deve-se imaginar um sinal  $X$  colocado no grupo de teste. Como dito anteriormente, esse sinal  $X$  é representado por 10 valores gerados pelo processo de extração de características. Cada um desses 10 valores do sinal  $X$  foi comparado a todos os valores de todos os sinais no grupo *teste*. Este processo consistiu na aplicação do critério de DE. Durante os testes, cada vez que uma matriz de confusão terminava de ser preenchida, ela era comparada a duas matrizes de confusão: *melhor matriz de confusão*

e a *pior matriz de confusão*, possibilitando redefinir quais foram os melhores e piores resultados.

A *melhor matriz de confusão* representa a melhor matriz encontrada. O critério utilizado para a melhor matriz foi a soma da diagonal principal, ou seja, a soma de todos os acertos. A melhor matriz é substituída ao final de uma iteração se a soma da diagonal principal pertencente a matriz da iteração atual for maior que a soma da diagonal da melhor. Quando isso é feito, o algoritmo é capaz de informar exatamente quais os grupos de sinal estavam no grupo teste e no grupo modelo.

Isso foi feito para que, além de ser encontrada a melhor solução, os vetores que geraram-na também fossem encontrados e analisados. Portanto a cada iteração obtia-se uma matriz de confusão, e a ela comparava-se a melhor matriz, sendo substituída, se necessário.

O tratamento da melhor matriz repetiu-se no tratamento da pior matriz de confusão. A cada nova matriz, obtida no final de uma iteração de classificação, era executada uma comparação com a pior matriz. Porém, o critério utilizado para a substituição foi a soma dos elementos que não pertenciam a diagonal principal.

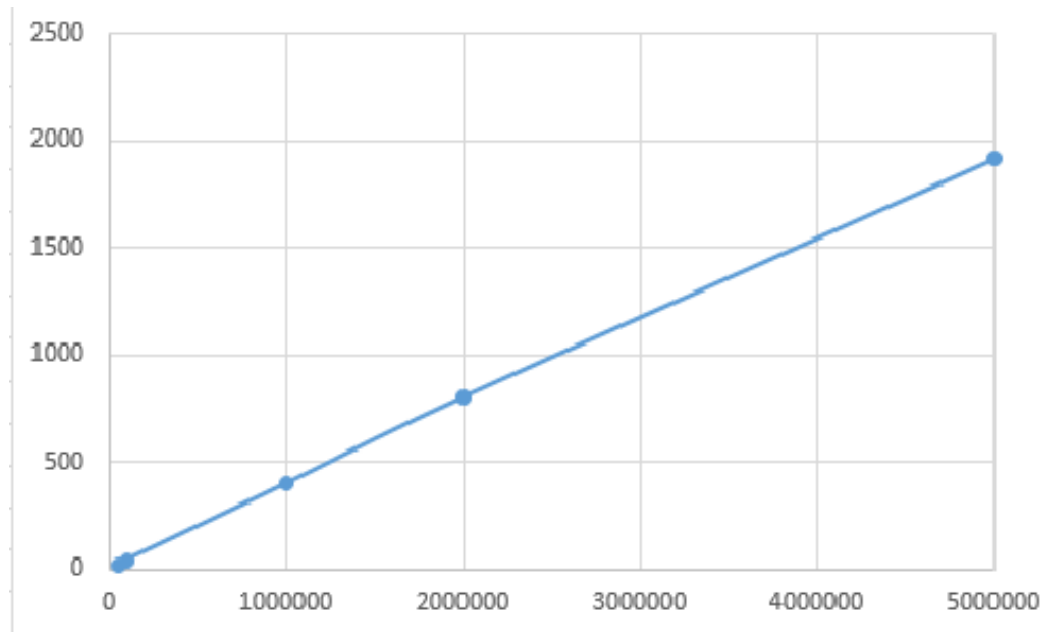
### 4.3 Complexidade do Algoritmo

Devido a enorme quantidade de possibilidades para serem testadas, como visto anteriormente, cabe fazer uma análise da complexidade do algoritmo desenvolvido. Essa análise foi importante para a decisão da quantidade de testes a serem executados.

O valor escolhido para a quantidade de testes deveria ser grande o suficiente para que os testes não representassem uma parte muito pequena do intervalo total, mas não tão grande que o tempo de execução fosse exagradamente demorado, tornando a execução dos testes e a obtenção dos resultados mais lenta que o necessário.

Para isso, o algoritmo foi submetido a uma série de testes. Esses testes consistiram em exe-

cutar o programa com diferentes valores para a quantidade de iterações. Ao final do programa, seu tempo de execução era exibido. Com diversas quantidades de testes, foi possível a plotagem de um gráfico, que pode ser visto na figura 4.1.



**Figura 4.1** – Gráfico de tempo de execução, no eixo Y, por tamanho da amostra, no eixo X. Percebe-se a relação praticamente linear.

Na figura 4.1 é possível ver o tempo de execução do algoritmo crescer conforme a amostra de testes aumenta. É possível observar que a figura obtida é suficientemente próxima à uma reta. Essa análise empírica foi utilizada para comparar o tempo de execução real do programa com o tempo de execução previsto utilizando a complexidade do algoritmo obtida por meio dessa análise.

Esse processo é importante para verificar a veracidade do tempo de execução previsto, e, da mesma forma, embasar a decisão da quantidade máxima de testes executados, uma vez que essa decisão está fortemente ligada ao tempo de execução para um valor escolhido.

Pela linearidade explícita na figura 4.1, os tempos previstos foram calculados usando proporção entre a quantidade de amostras anterior e o tempo de execução da mesma. Os valores que deram origem a figura 4.1, bem como os valores previstos e o erro entre tempo de execução previsto e tempo de execução medido podem ser vistos na tabela 4.1.

Quantidade de Testes	Tempo de Execução	Tempo Esperado	Erro
50000	19,317108	-	-
100000	39,600114	-	-
1000000	396,333647	-	-
2000000	798,793876	792,667294	0,0076698
5000000	1909,128714	1996,98469	-0,0460189
50000000	18987,98072	19091,28714	-0,0054406

**Tabela 4.1** – Tamanhos dos vetores, tempos de execução medidos, tempos de execução obtidos através da complexidade do algoritmo e o erro entre o tempo esperado e o tempo obtido.

Na tabela 4.1, a coluna *Quantidade de Testes* representa quantos testes foram executados, a coluna *Tempo de Execução* representa o tempo medido, em segundos, da execução da função de classificação dos sinais, a coluna *Tempo Esperado* representa o tempo previsto de execução, considerando uma complexidade linear. A coluna *Erro* representa o *Erro Relativo*[20] entre o tempo de execução esperado e o tempo de execução obtido.

Os valores na coluna *Erro* não estão representados em porcentagem. Ou seja, o valor  $-0.0054406$  representa um erro de aproximadamente 0,5% do valor obtido em relação ao valor previsto. Alguns valores na tabela 4.1 são negativos pois o tempo gasto na execução foi menor que o tempo previsto.

A complexidade do algoritmo apresentou valores suficientemente próximos dos valores esperados, se considerado uma complexidade  $O(N)$ , como pode ser visto na tabela 4.1. Comprovado a veracidade desses valores, a complexidade  $O(n)$  desempenhou um importante papel na decisão da quantidade de testes. Como visto, a quantidade total de testes passa de 40 bilhões. Portanto, para ser executado todos os testes possíveis, em um processo *single-threaded*, demoraria cerca de 185 dias para o final de execução.

## 4.4 Primeiros Resultados

Foram implementados, em linguagem C/C++, 400 milhões de testes, buscando a melhor matriz de confusão com base no processo explicitado na seção ???. Estes testes foram implementados com 4 *threads*, designando 100 milhões de testes para cada uma.

Cada *thread* trabalhou de maneira independente, produzindo seus próprios arquivos. Isso foi feito para evitar qualquer tipo de problema de concorrência e para tornar desnecessária a aplicação de algum mecanismo para controle de sincronismo (semáforos, por exemplo). Com isso, o foco das *threads* foi tornar o processo de teste mais rápido, e de fato apresentou um tempo de teste aproximadamente quatro vezes mais rápido que se um único programa fosse executado com 400 milhões de testes.

### 4.4.1 Melhor Resultado

Este resultado foi encontrado após 400 milhões de testes. A matriz de confusão obtida ao final da execução é:  $\begin{pmatrix} 0 & 7 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 7 \end{pmatrix}$

A matriz reflete um resultado preocupante: 3 classificações do tipo falso negativa. Para a classificação entre *saudáveis* e doentes, o resultado implicou em 52.38% de acurácia. Para o pré-diagnóstico dos 7 pacientes com nódulo, o algoritmo apresentou 57.14% de precisão. Já para a classificação dos 7 pacientes com nódulo, houve 100% de acurácia.

O importante deste resultado é notar que os elementos definidos como falsos negativos implicaram que o algoritmo identificou um paciente doente como saudável. Esse problema foi levado em conta para o desenvolvimento de um critério para refinar os resultados.

#### 4.4.2 Pior Resultado

O pior resultado, encontrado ao longo da execução de 400 milhões de testes está representado pela matriz  $\begin{pmatrix} 7 & 0 & 0 \\ 7 & 0 & 0 \\ 7 & 0 & 0 \end{pmatrix}$

Apesar de apresentar 100% de precisão para a classificação dos 7 pacientes saudáveis, as categorias nódulo e edema apresentaram 0% de precisão. E mais importante, para as duas categorias, todos os erros cometidos implicaram em falsos negativos.

Nesta execução, o algoritmo ainda não apresentava critérios para decidir qual tipo de erro era mais danoso, entre falsos negativos e falsos positivos. Portanto, o fato de todos os erros terem sido falsos negativos é uma consequência da aleatoriedade do programa.

A análise destas matrizes permite ver que os resultados encontrados podiam ser otimizados, de maneira que falsos negativos fossem evitados. O processo de otimização dos resultados obtidos, bem como os resultados propriamente otimizados, são discutidos adiante.

### 4.5 Otimização dos Resultados

#### 4.5.1 Classificação dos Erros

Os erros cometidos pelo algoritmo proposto podem ser classificados em duas categorias: *falsos positivos* e *falsos negativos*. Essa distinção foi importante como critério de desempate na comparação entre matrizes.

Um *falso positivo* é uma classificação, seja ela nódulo ou edema, para a qual o paciente na verdade era saudável. Ou seja, um paciente saudável é sorteado para o teste e é classificado

como patológico, independente da patologia erroneamente identificada.

O resultado desse pré-diagnóstico pode levar o paciente a procurar um médico e submeter-se ao exame necessário, descobrindo então que a classificação dada pelo algoritmo havia sido incorreta. Essa situação, por mais desagradável, não oferece risco a saúde do paciente.

Apesar de, por várias vezes ao longo deste trabalho, ser explicado que o objetivo do mesmo é um *pré-diagnóstico*, um paciente diagnosticado como falso negativo é um sério problema. *Falso negativo* significa que o programa classificou o paciente doente, esteja ele com nódulo ou edema, como saudável. Isso pode levar o paciente a crer que está bem, apesar de encontrar-se doente.

Para a minimização desse dano, o algoritmo foi programado de maneira que, quando um paciente pertence a classe doente e ele fosse classificado como saudável, esse erro tivesse um peso maior do que se fosse saudável e classificado como doente. Isso garante que a melhor e pior matriz levassem em conta os falsos positivos e falsos negativos. Na matriz de confusão, o falso positivo é representado pelos elementos da segunda e terceira linha, ambos da primeira coluna. As figuras 4.2 e 4.3 exemplificam o caso.

Na figura 4.2, os elementos em vermelho representam os elementos de uma matriz de confusão que equivalem a resultados falsos positivos. Já na figura 4.3, os elementos em vermelho representam os resultados falsos negativos.

	Saudáveis	Nódulos	Edemas
Saudáveis	$a$	$b$	$c$
Nódulos	$d$	$e$	$f$
Edemas	$g$	$h$	$i$

**Figura 4.2** – matriz com erros falsos positivos.

	Saudáveis	Nódulos	Edemas
Saudáveis	$a$	$b$	$c$
Nódulos	$d$	$e$	$f$
Edemas	$g$	$h$	$i$

**Figura 4.3** – matriz com erros falsos negativos

Devido a gravidade de resultados falsos negativos, foi gerado um critério de desempate que considera falsos negativos piores que falsos positivos para encontrar a melhor matriz de confusão. Foi aplicado um fator multiplicador aos elementos de falsos positivos e, da mesma maneira, um fator multiplicador foi aplicado aos elementos falsos negativos.

Isso garante que, quando em busca da melhor matriz de confusão, o algoritmo evite considerar casos cujos resultados foram falsos negativos, fazendo com que os elementos, destacados em vermelho na figura 4.3, fiquem o mais próximo possível de 0.

Os elementos da matriz de confusão que não pertencem a diagonal principal e não estão contemplados nas figuras 4.2 e 4.3 são diagnósticos incorretos. Assim, o paciente realmente possuía nódulo e foi classificado como edema, ou o paciente possuía edema e foi classificado como nódulo. Estes 2 erros foram considerados equivalentes e, portanto, não foram utilizados como critérios de desempate.

Com essas observações feitas, os testes foram executados novamente, e em maior escala. Utilizando todos os recursos de desempate descritos ao longo da seção 4.5.1, foram executados 800 milhões de testes. O pior e melhor resultado, bem como suas matrizes de confusão, serão discutidos a seguir.



## 4.6 Resultados Finais

O pior resultado é aquele que cometeu maior quantidade de falsos negativos. Na matriz  $\begin{pmatrix} 0 & 7 & 0 \\ 7 & 0 & 0 \\ 7 & 0 & 0 \end{pmatrix}$  é possível ver a quantidade de erros no pior resultado encontrado, implicando em 0% de precisão em todas as classificações. A primeira linha, que representa as classificações dadas aos sinais de pessoas saudáveis, teve todos os sinais classificados como nódulo. Como discutido anteriormente, esses erros não foram utilizados em critérios de desempate. Porém, nas linhas seguintes da matriz de confusão, é possível ver que ambas patologias, nódulo e edema, foram classificadas como normais, totalizando 14 falsos negativos. Esse resultado é claramente o pior possível a ser obtido, uma vez que além de ter 0% de precisão, cometeu todos os erros graves possíveis.

Tal resultado era esperado, devido ao forte peso atribuído na soma dos erros quando um falso negativo era identificado. Assim, o algoritmo apresentou o comportamento esperado.

### 4.6.1 Melhor Resultado

Após a execução dos 800 milhões de testes, a melhor matriz de confusão obtida foi  $\begin{pmatrix} 0 & 7 & 0 \\ 0 & 6 & 1 \\ 0 & 5 & 2 \end{pmatrix}$ .

O critério de desempate desenvolvido foi, assim, eficaz. Como pode ser visto na matriz, todos os erros considerados falso negativo foram extinguidos, cumprindo com o objetivo de minimizar a gravidade de erros que fossem cometidos. Os sinais normais foram todos classificados como doentes com nódulo, as este erro consiste em um falso positivo e, como dito anteriormente, não apresenta grave perigo a saúde de um paciente.

Quanto a precisão de classificação entre saudáveis e patológicos, o algoritmo acertou 14 dos 21 sinais analisados, o que representa uma precisão de 66,67%. Essa precisão foi medida considerando apenas dois tipos de sinais: saudáveis e doentes.

O algoritmo desenvolvido classificou corretamente todos os sinais patológicos como patológicos. Isso é de grande importância, visto que nenhum paciente doente seria prejudicado por uma ausência de diagnóstico. Dos 14 pacientes doentes, apenas 3 tiveram suas patologias classificadas erroneamente. De 7 pacientes com nódulo, apenas um foi classificado como edema, e dentre os 7 pacientes com edema, 2 foram classificados com nódulo.

Portanto, o algoritmo apresentou as seguintes precisões: 66,67% entre as classificações saudável e patológico, 52,34% se levado em consideração diagnósticos errados entre os tipos de doença, 85,71% para a classificação de sinais com nódulo, e 71,42% para a classificação de sinais com edema.

## Capítulo 5

### Conclusões

Este trabalho apresentou uma discussão, ainda que breve, sobre os problemas dos métodos de diagnósticos disponíveis para a identificação das patologias na laringe. Os métodos são desconfortáveis para os pacientes, e isso foi um dos motivos que levou a criação deste trabalho. Em seguida, foram apresentadas pesquisas sobre o reconhecimento não-invasivo de patologias na laringe. Além disso, todos os conceitos teóricos abrangidos neste projeto foram exibidos também durante a fundamentação teórica.

No que tange ao algoritmo proposto, antes de aplicação de qualquer processamento, foi feita a digitalização dos sinais. Este processo obedeceu aos teoremas que o regem, como o Teorema da Amostragem. Prosseguindo, o extrator de características construído foi aplicado a todos os sinais da base de dados. Ao final do extrator de características, 420 vetores de características de tamanho 10 foram providos ao classificador, que foi responsável por preencher a matriz de confusão de acordo com seus acertos e erros.

O classificador, que empregou a Distância Euclidiana para classificar cada sinal colocado no grupo teste, obedeceu a um esquema de validação cruzada, implicando em 1 bilhão e 200 milhões de sorteios, um sorteio para cada iteração de teste executada.

Ao longo dos testes, as matrizes de confusão foram discutidas com o intuito de comprovar que, sem o critério de desempate explanado, o algoritmo cometia erros graves e que podiam ser refinados para que o resultado do mesmo fosse o melhor o possível.

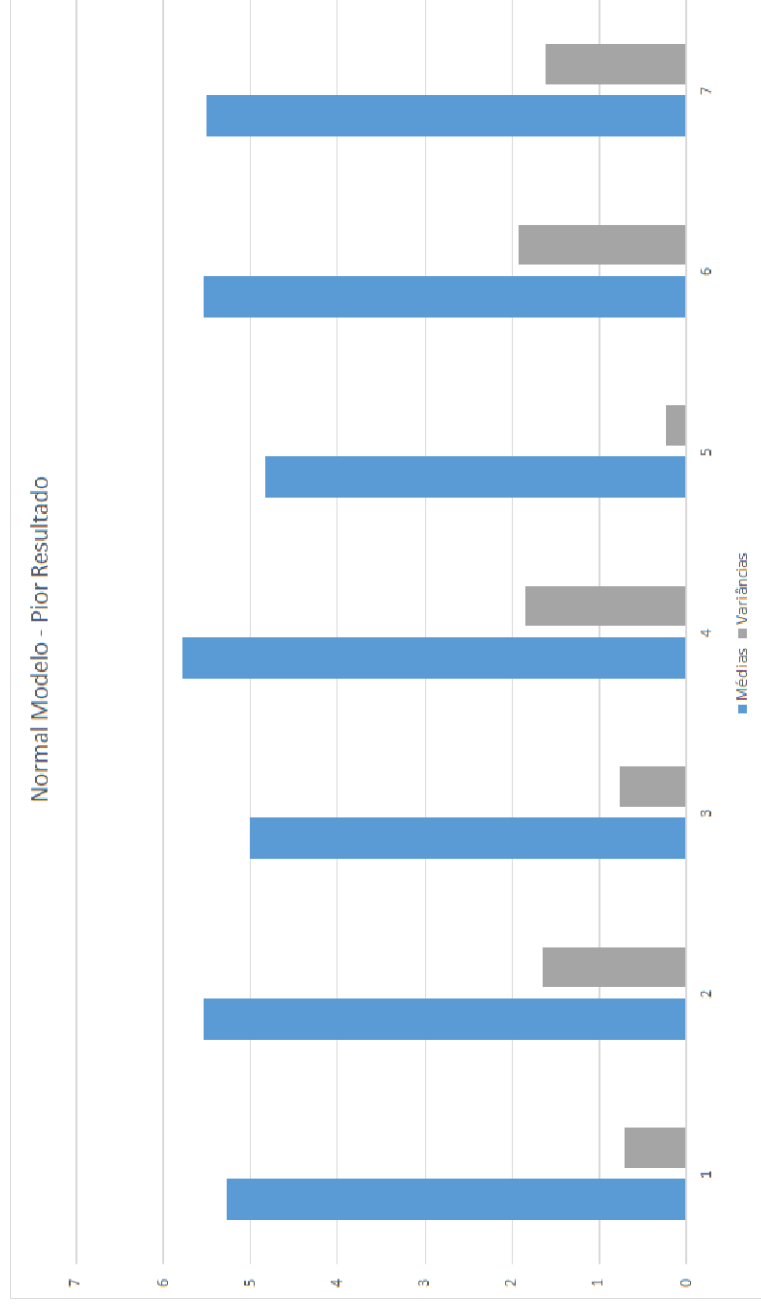
Por fim, foram mostradas a melhor e a pior matriz de confusão geradas. Na pior matriz, foi possível observar que o critério de desempate solucionou o problema de falsos negativos. Particularmente, na melhor matriz de confusão obtida, o resultado foi de 66,67% entre as classificações saudável e patológico. Neste resultado, apenas 3 pacientes tiveram suas patologias confundidas pelo algoritmo, mas ainda assim eles foram classificados como oriundos de indivíduos doentes. Para os 14 sinais patológicos, o algoritmo apresentou 100% de precisão, ou seja, nenhum paciente doente foi classificado como normal. Desse modo, satisfeitas estão as expectativas.

# Apêndice A

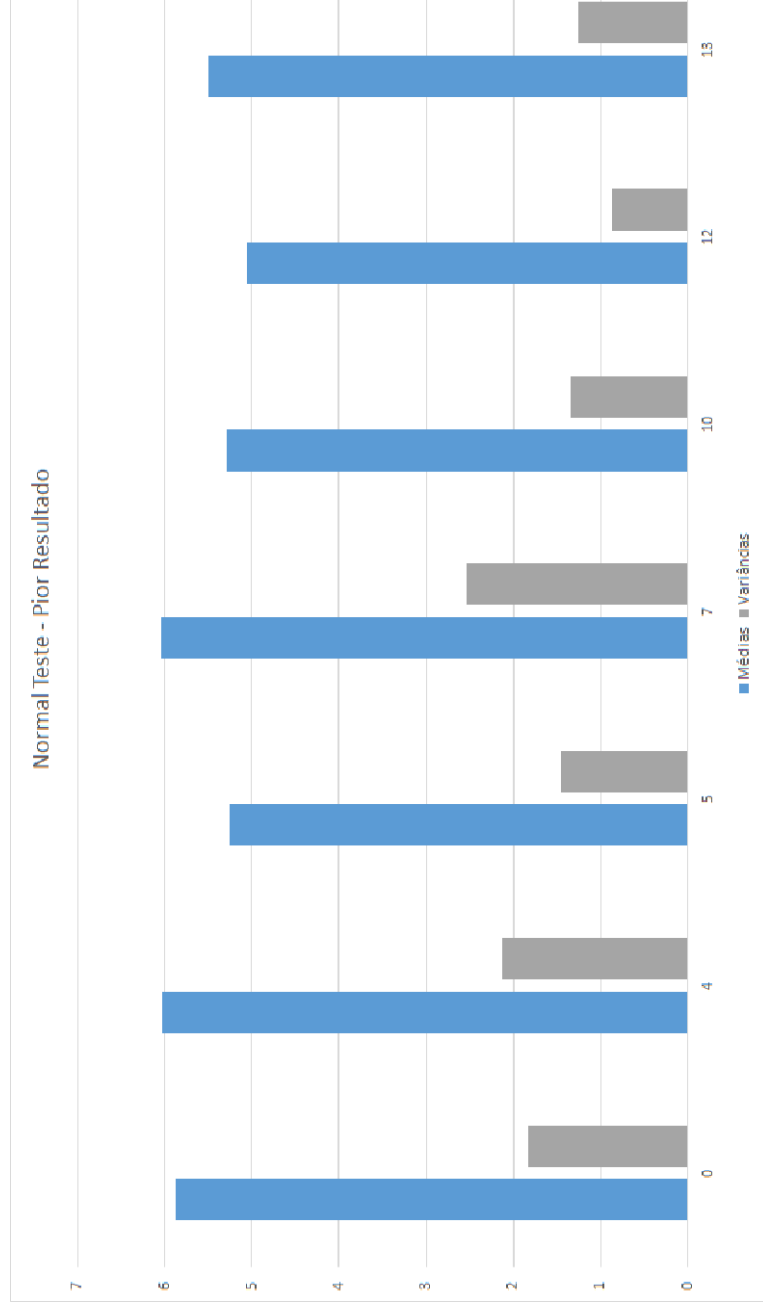
## Valores gerados pelo programa

Os sinais nesta seção, em vez de serem compostos por 10 valores como utilizado ao longo deste trabalho, serão representados por apenas dois valores: um que representa a média de todas as médias processadas do sinal, representado em todas as figuras pela coluna azul, e um que representa a média de todas as variâncias do sinal, representado em todas as figuras desta seção pela coluna laranja.

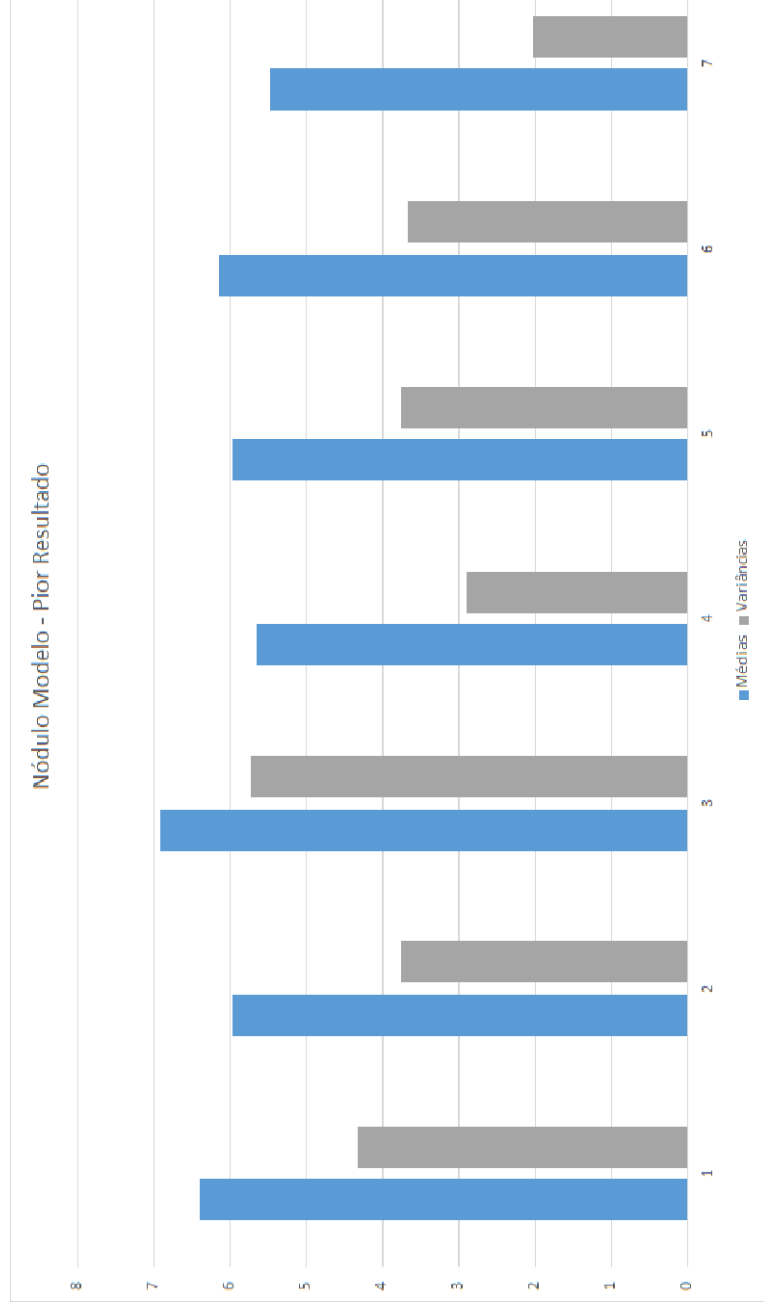
Isso foi feito na tentativa de transformar os gráficos para que fossem mais fáceis de serem visualizados. Este processo não apresentou perda significativa das características dos valores. Todas as 5 médias processadas e 5 variâncias processadas, ao serem transformadas em uma média de médias processadas e uma média de variâncias processadas tiveram seus desvios-padrão medidos, e todas as amostras encontravam-se dentro do desvio.



**Figura A.1** – Valores dos sinais saudáveis, selecionados para o grupo modelo, que geraram o pior resultado

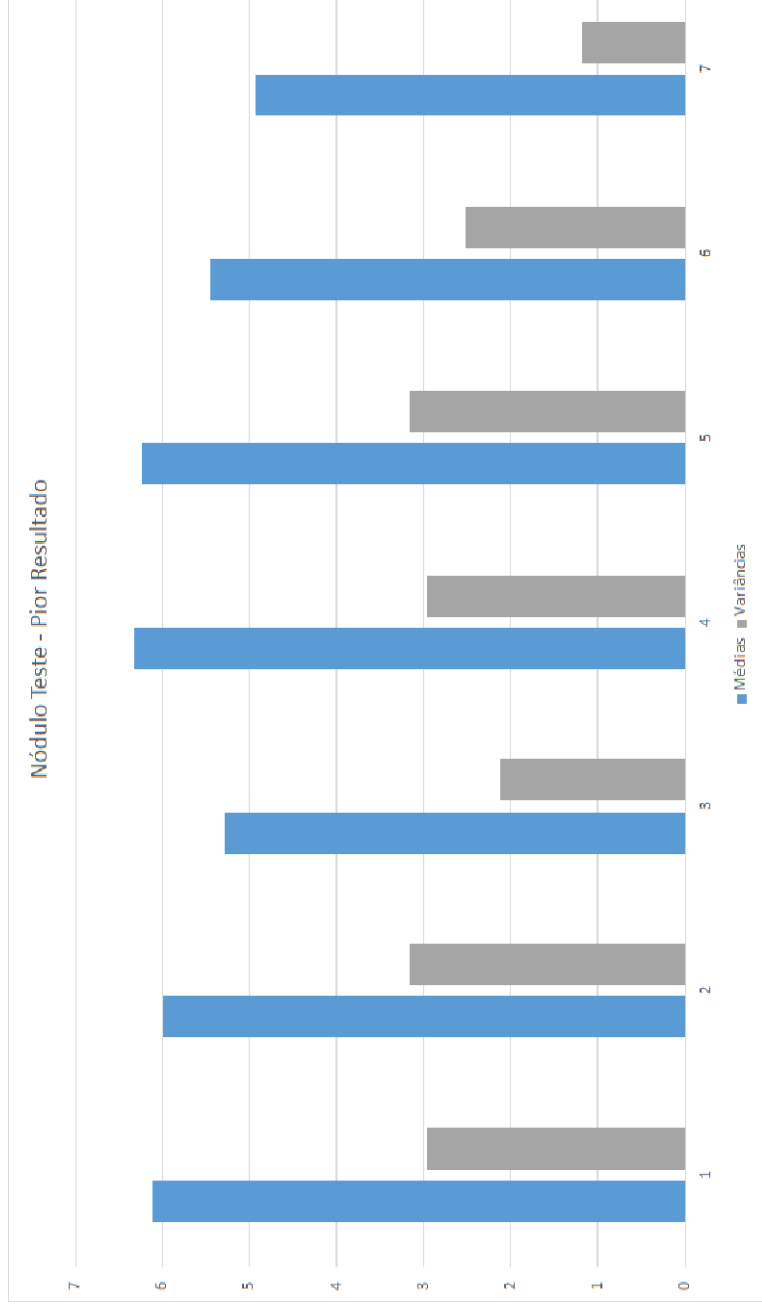


**Figura A.2** – Valores dos sinais saudáveis, selecionados para o grupo teste, que geraram o pior resultado.

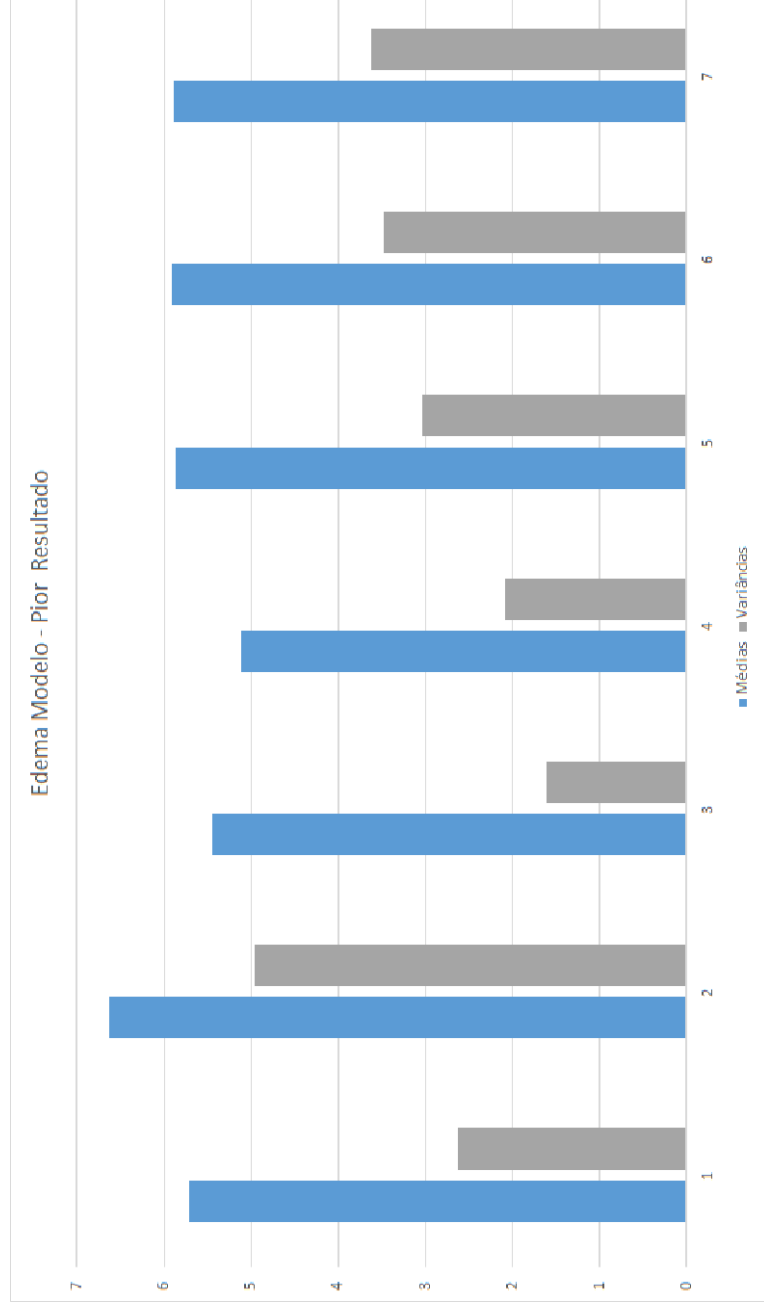


**Figura A.3** – Valores dos sinais de pacientes com nódulo, selecionados para o grupo modelo, que geraram o pior resultado.

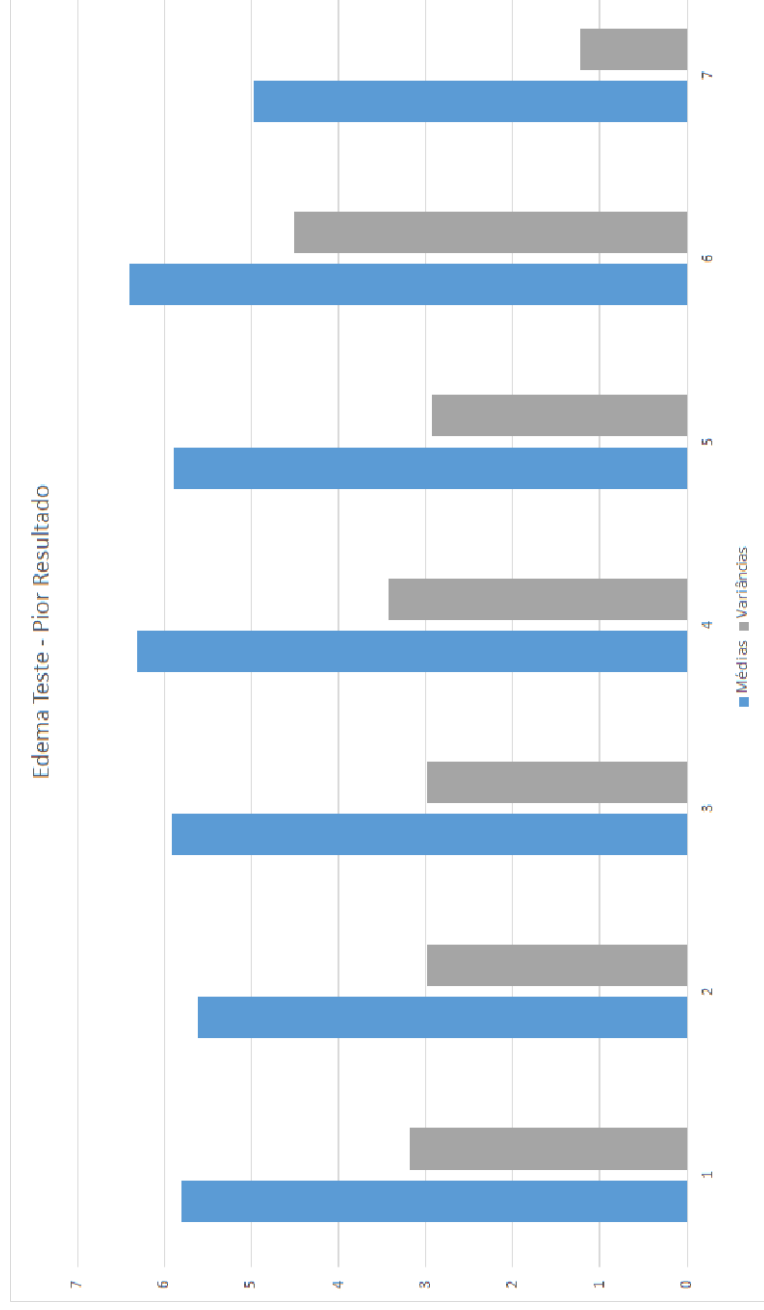




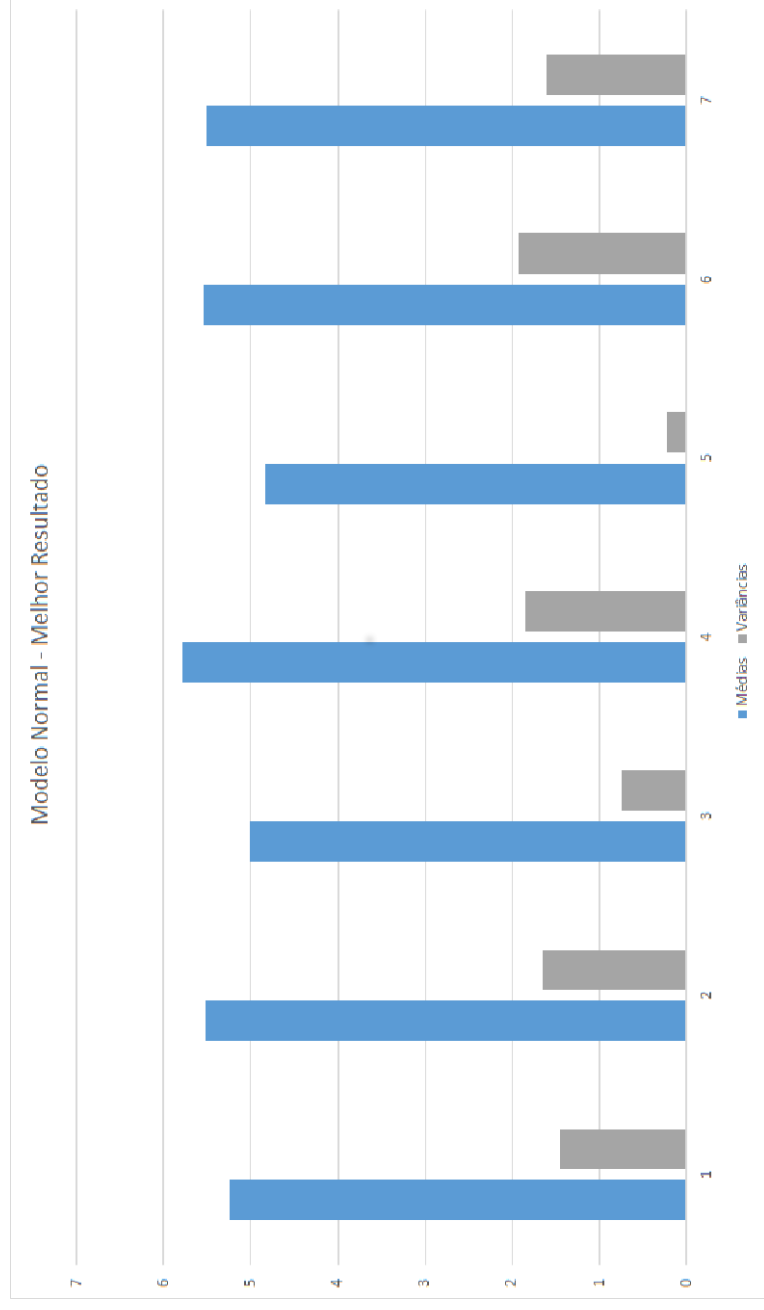
**Figura A.4** – Valores dos sinais de pacientes com nódulo, selecionados para o grupo teste, que geraram o pior resultado.



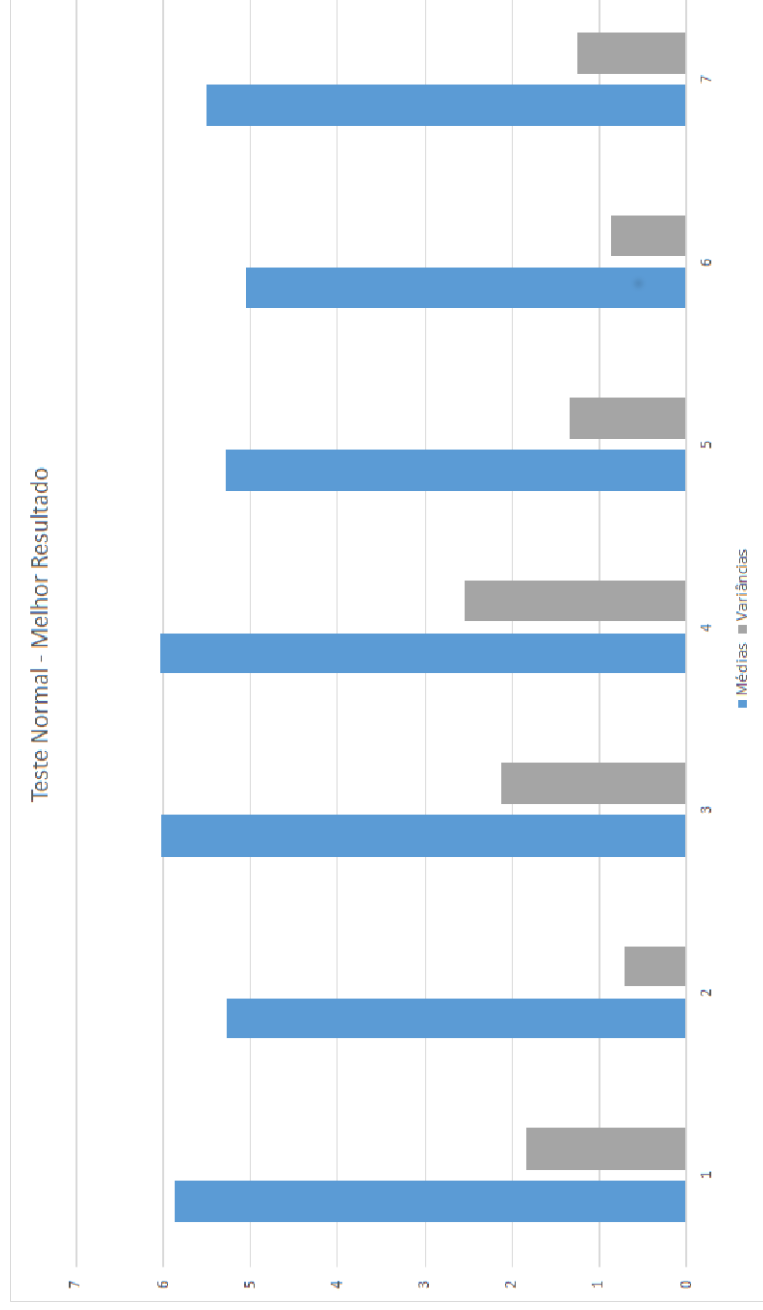
**Figura A.5** – Valores dos sinais de pacientes com edema, selecionados para o grupo modelo, que geraram o pior resultado.



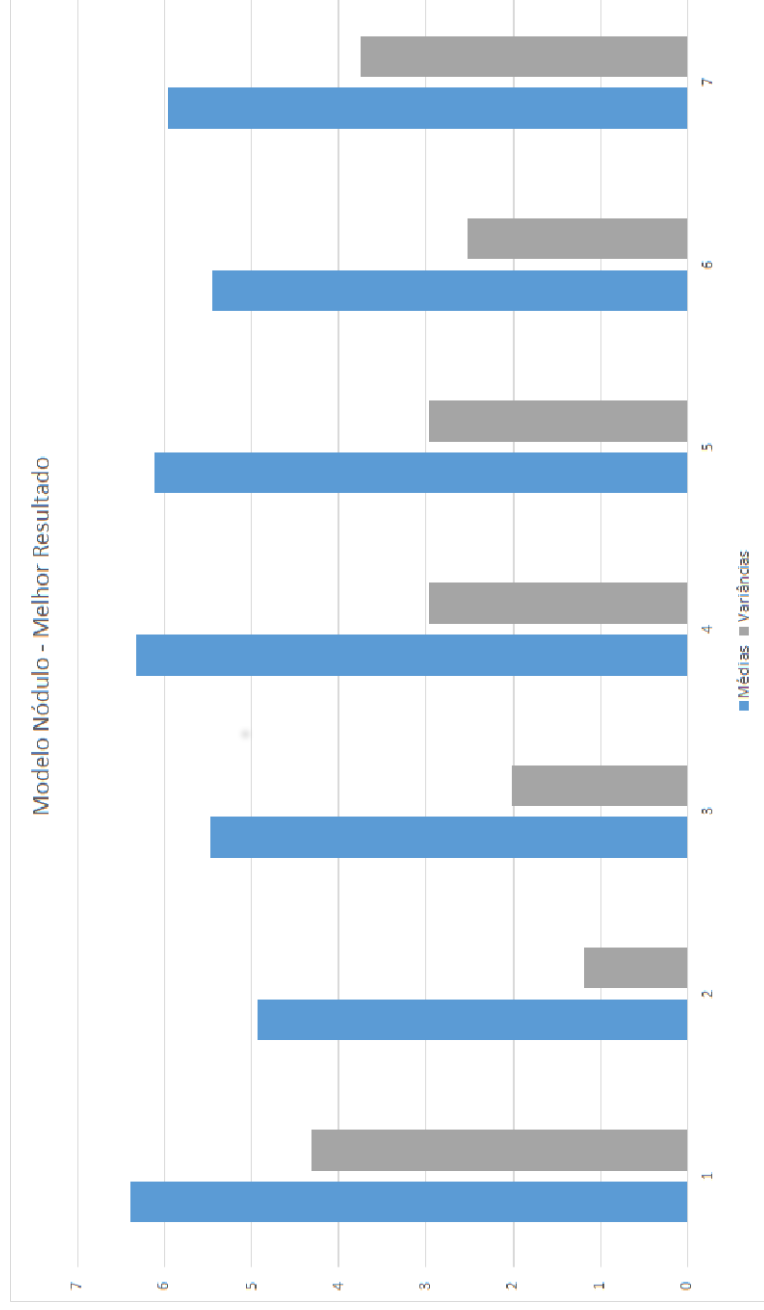
**Figura A.6** – Valores dos sinais de pacientes com edema, selecionados para o grupo teste, que geraram o pior resultado.



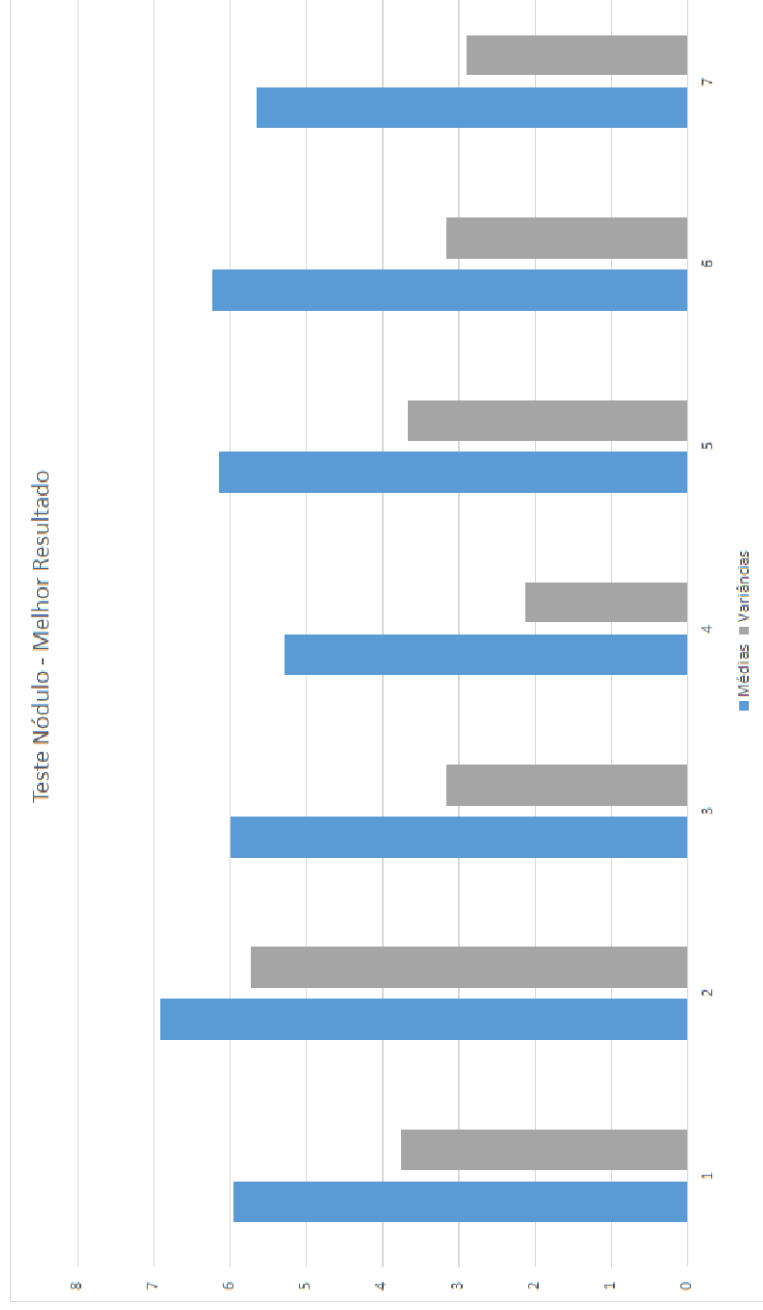
**Figura A.7** – Valores dos sinais de pacientes saudáveis, selecionados para o grupo modelo, que geraram o melhor resultado.



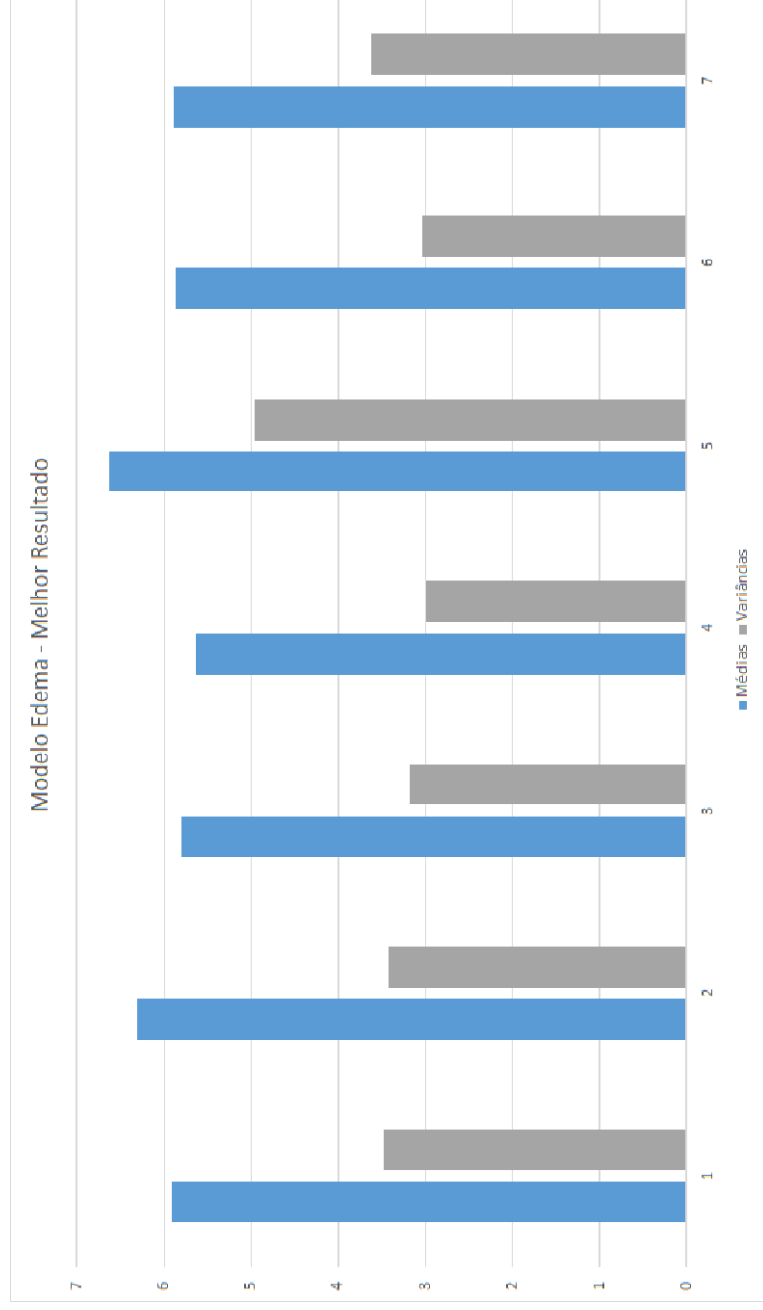
**Figura A.8** – Valores dos sinais de pacientes saudáveis, selecionados para o grupo teste, que geraram o melhor resultado.



**Figura A.9** – Valores dos sinais de pacientes com nóculo, selecionados para o grupo modelo, que geraram o melhor resultado.

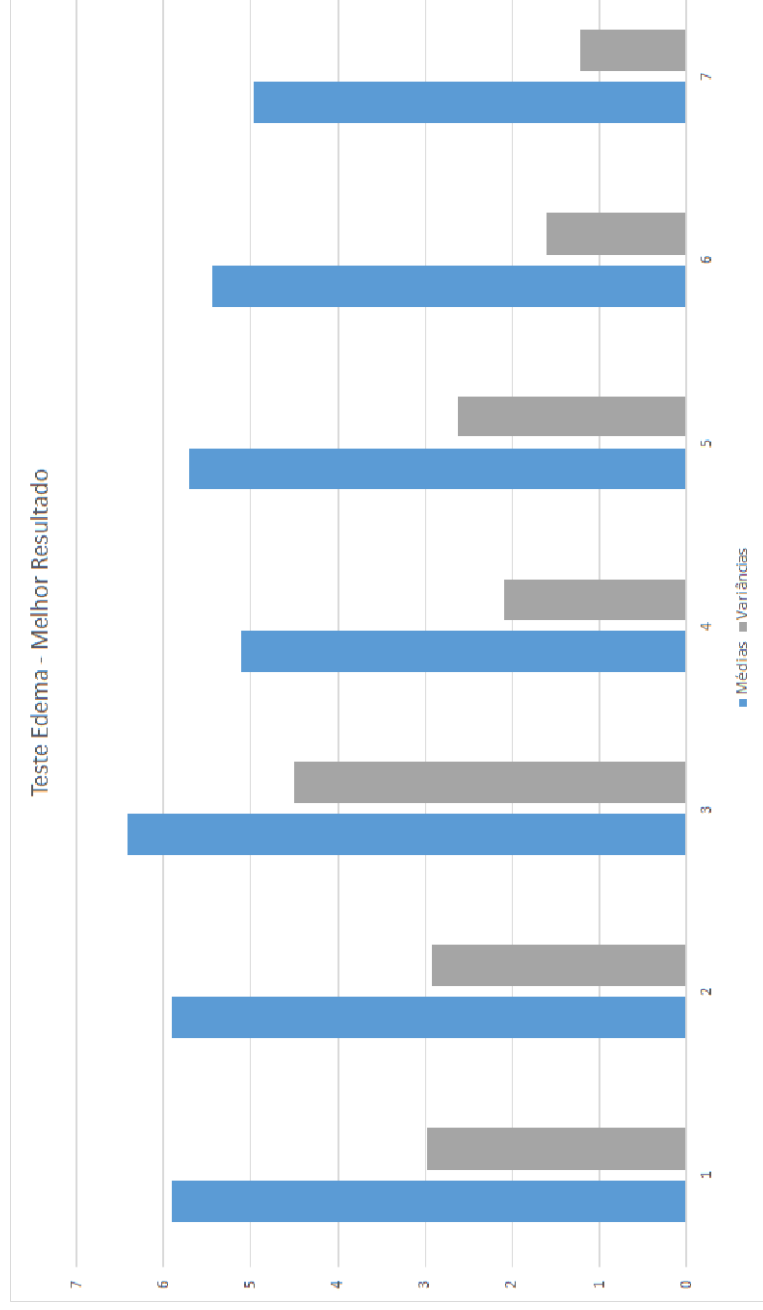


**Figura A.10** – Valores dos sinais de pacientes com nóculo, selecionados para o grupo teste, que geraram o melhor resultado.



**Figura A.11** – Valores dos sinais de pacientes com edema, selecionados para o grupo modelo, que geraram o melhor resultado.





**Figura A.12** – Valores dos sinais de pacientes com edema, selecionados para o grupo teste, que geraram o melhor resultado.

# Referências

- [1] W. C. d. A. Costa, "Análise dinâmica não linear de sinais de voz para detecção de patologias laríngeas," Universidade Federal de Campina Grande. Tese de doutorado, 2012.
- [2] V. Vieira, S. Costa, W. Costa, S. Correia, and F. Assis, "Discriminação de sinais de voz com análise de quantificação de recorrência e redes neurais mlp," XXXI Simpósio Brasileiro de Telecomunicações (SBrT 2013), 2013.
- [3] Taciana A. Souza et al, Voice pathology assessment using wavelet based on texture analysis of recurrence plots, IFPB, João Pessoa, Brazil; em Telecommunications (IWT), 2015 International Workshop, 14-17 junho de 2015. IEEE, DOI: 10.1109/IWT.2015.7224580.
- [4] J. D. Arias-Londono, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Dominguez, "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients," Biomedical Engineering, IEEE Transactions on, vol. 58, 2011.
- [5] I. A. Kuhl, "Laringologia Prática Ilustrada", Segunda Edição, Editora Revinter, ISBN: 85-7309-121-5.
- [6] M. P. Fried, "The Larynx: A Multidisciplinary Approach", Segunda Edição, ISBN: 978-0801680465.
- [7] C. E. Shannon, "Communication in the presence of noise," Proc. IRE, vol. 37, no. 1, pp. 1021, Jan. 1949.
- [8] C. E. Shannon, "Communication in the presence of noise," Proc. IEEE, vol. 72, no. 9, pp. 11921201, Sep. 1984.
- [9] W. d. O. Bussab, P. A. Morettin, Estatística Básica - 8 Ed. 2013.

- [10] Issued as a joint design by IBM Corporation and Microsoft Corporation, “Multimedia Programming Interface and Data Specification v1.0”
- [11] S. Visa, B. Ramsay, A. Ralescu, E. v. d. Knaap, “Confusion Matrix-based Feature Selection”, conferência “Midwest Artificial Intelligence and Cognitive Science”, 2011, Cincinnati, USA, April 16-17.
- [12] E. Deza, M. M. Deza, “Encyclopedia of Distances”, editora Springer, p. 94, 2009.
- [13] Godino-Llorente JI, Gomes-Vilda P, Blanco-Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. IEEE Transactions on Biomedical Engineering. 2006; 53(10):1943-1953.
- [14] Wang X, Zhang J, Yan Y. Discrimination between pathological and normal voices using GMM-SVM approach. Journal of Voice. 2011; 25(1):38-43
- [15] C. S. Sapp WAVE PCM soundfile format, disponível em: <http://sound-file.sapp.org/doc/WaveFormat/wav-sound-format.gif>
- [16] Costa WCA, Vieira VJD, Costa SC, Assis FM, Aguiar Neto BG. Avaliação do uso combinado de medidas de quantificação de recorrência e análise LPC na classificação de vozes patológicas. In: Anais do XXIII Congresso Brasileiro de Engenharia Biomédica (CBEB 2012); Porto de Galinhas, Brasil. 2012
- [17] L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition. Englewood Cliffs, NJ: Prentice-Hall, 1993
- [18] Henríquez P, Alonso JB, Ferrer MA, Travieso CM, Godino-Llorente, Díaz-de-María F. Characterization of healthy and pathological voice through measures based on non-linear dynamics. IEEE Transactions on Audio, Speech and Language Processing. 2009; 17(6):1186-1195
- [19] B. Boyanov and S. Hadjitodorov, “Acoustic analysis of pathological voices: A voice analysis system for screening of laryngeal diseases,” IEEE Eng. Med. Biol. Mag., vol. 16, pp. 7482, Jul./Aug. 1997.

- [20] M.A. Gomes Ruggiero, V. L. da Rocha Lopes. Cálculo Numérico - Aspectos Teóricos e Computacionais, 2ª edição, Editora Pearson, 1997.