

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Classificação de Sinais de Voz Afetada por Patologia
nas Pregas Vocais Utilizando Reconstrução do
Espaço de Fases

João Vilian de Moraes Lima Marinus

Tese submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Modelos Computacionais e Cognitivos

Prof. Dra. Joseana Macêdo Fechine Régis de Araújo

Prof. Dr. Herman Martins Gomes

(Orientadores)

Campina Grande, Paraíba, Brasil

©João Vilian de Moraes Lima Marinus, 05/09/2019

M339c

Marinus, João Vilian de Moraes Lima.

Classificação de sinais de voz afetada por patologia nas pregas vocais utilizando reconstrução do espaço de fases / João Vilian de Moraes Lima Marinus. – Campina Grande, 2019.

136 f. : il. color.

Tese (Doutorado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2019.

"Orientação: Profa. Dra. Joseana Macêdo Fachine Régis de Araújo, Prof. Dr. Herman Martins Gomes".

Referências.

1. Modelos Computacionais e Cognitivos. 2. Processamento Digital de Sinais de Voz. 3. Redes Neurais Convolucionais. 4. Classificação – Vozes Afetadas por Patologias. 5. Bases de Dados de Distúrbios da Voz. I. Araújo, Joseana Macêdo Fachine Régis de. II. Gomes, Herman Martins. III. Título.

CDU 004.81(043)

**"CLASSIFICAÇÃO DE SINAIS DE VOZ AFETADA POR PATOLOGIA NAS PREGAS
VOCAIS POR RECONSTRUÇÃO DO ESPAÇO DE FASES"**

JOÃO VILIAN DE MORAES LIMA MARINUS

TESE APROVADA EM 05/09/2019

JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, Dra., UFCG
Orientador(a)

HERMAN MARTINS GOMES, Ph.D, UFCG
Orientador(a)

LEANDRO BALBY MARINHO, Dr., UFCG
Examinador(a)

LEONARDO VIDAL BATISTA, Dr., UFPB
Examinador(a)

FRANCISCO MARCOS DE ASSIS, Dr., UFCG
Examinador(a)

FRANCISCO MADEIRO BERNARDINO JUNIOR, Dr., UNICAP
Examinador(a)

CAMPINA GRANDE - PB

*Dedico esse trabalho ao meu filho Lucas Oliveira Marinus e a minha esposa Sanny
Emanuelle Oliveira.*

“A persistência é o menor caminho do êxito.”

Charles Chaplin

Resumo

Nos últimos anos, várias pesquisas na área de Processamento Digital da Voz vêm sendo realizadas com o objetivo de avaliar a qualidade da voz do paciente e auxiliar um especialista no diagnóstico de patologias nas pregas vocais. A análise acústica da voz pode ser uma ferramenta eficiente para o diagnóstico de patologias e possui como vantagem o fato de não ser invasiva. Neste contexto, esta tese teve como objetivo principal a investigação e a criação de métodos para a classificação de vozes afetadas por patologias nas pregas vocais. Para tanto, objetivou-se verificar a utilização da análise não linear do sinal de voz para caracterizar patologias das pregas vocais, a partir de imagens obtidas de técnicas da Teoria do Caos. Para este fim, foram estudadas 5 classes de patologias das pregas vocais: Paralisia, Edema, Nódulo, Pólipo e Queratose. Adicionalmente, foi estudada uma classe denominada Lesão Benigna nas pregas vocais, consistindo no agrupamento de sinais de voz afetada por nódulo, pólipo e cisto. Na pesquisa, foram utilizadas duas bases de dados: *Massachusetts Eye and Ear Infirmary* (MEEI) e *Saarbruecken Voice Database* (SVD). A etapa de pré-processamento consistiu no aumento da quantidade de sinais utilizando o método de *Time Stretching*, a segmentação e o janelamento dos sinais. Na fase de extração de características, foram obtidas imagens de cada segmento do sinal a partir das trajetórias do espaço de fase reconstruído do sinal. As imagens foram utilizadas para treinar duas Redes Neurais Convolucionais (*Convolutional Neural Network* - CNN), uma com e outra sem uma camada *bottleneck*. A partir da camada *bottleneck*, foram obtidos vetores de características, que foram usados para treinar uma Máquina de Vetores de Suporte (*Support Vector Machine* - SVM). Os resultados da SVM foram comparados com os resultados da CNN sem a camada *bottleneck*. Foram realizadas 14 classificações: Normal *versus* Patologia; 10 classificações pareadas entre as 5 classes de patologias; e 3 classificações entre as classes Paralisia, Edema e Queratose *versus* Lesão Benigna. A classificação Normal *versus* Patologia proporcionou 100% de acerto, tanto para CNN quanto para SVM. A classificação Nódulo *versus* Pólipo proporcionou acurácias acima de 90%, e as outras classificações proporcionaram resultados entre 70 e 90%. Foi observado que, de uma forma geral, classificações utilizando aumento de dados no conjunto de treinamento tiveram resultados melhores que classificações sem a utilização de tal aumento, exceto em classificações envolvendo a classe Pólipo. Na maioria dos casos, a utilização de tamanho de segmentos de 10 ciclos de *pitch*, para a formação das imagens, proporcionou melhores resultados que o tamanho clássico de 20 ms. No geral, a classificação utilizando *bottleneck* e SVM proporcionou resultados superiores àqueles utilizando apenas CNN. A abordagem proposta se mostrou promissora para a área de reconhecimento de patologias nas pregas vocais pela voz, uma vez que proporcionou bons resultados ao classificar diferentes tipos de patologia, que é uma tarefa árdua devido ao caráter ruidoso do sinal de voz afetado por patologia.

Palavras-chave: Processamento digital de sinais de voz, Classificação de vozes afetadas por patologias, Bases de Dados de Distúrbios da Voz, Reconstrução do Espaço de Fase, Redes Neurais Convolucionais.

Abstract

In recent years, several researches in the area of digital voice processing have been carried out with the objective of evaluating the quality of the patient's voice and assisting a specialist in the diagnosis of vocal fold pathologies. The acoustic analysis of the voice can be an efficient tool for the diagnosis of pathologies and has the advantage that it is not invasive. In this context, this thesis has as main objective the investigation and creation of methods for the classification of voices affected by pathologies in the vocal folds. In this thesis, the objective was to verify the use of non-linear analysis of voice signal to characterize vocal fold pathologies from images obtained from Chaos Theory techniques. For this purpose, 5 classes of vocal fold pathologies were studied: Paralysis, Edema, Nodule, Polyp and Keratosis. In addition, a class called Benign Lesion in the vocal folds, consisting of voice signals affected by nodule, polyp and cyst, was studied. In this research, two databases were used: Massachusetts Eye and Ear Infirmary (MEEI) and Saarbruecken Voice Database (SVD). The preprocessing step consisted in increasing the number of signals using the Time Stretching method, and the segmentation and the windowing of the signals. In the feature extraction step, images of each segment of the signal were obtained from the trajectories of the reconstructed phase space of the signal. The images were used to train two Convolutional Neural Networks (CNN), one with and one without a bottleneck layer. From the bottleneck layer, feature vectors were obtained to train a Support Vector Machine (SVM). The SVM results were compared with the CNN results without the bottleneck layer. There were 14 classifications: Normal versus Pathology; 10 pairwise classifications among the 5 classes of pathologies; and 3 classifications between the classes Paralysis, Edema and Keratosis versus Benign Lesion. The Normal versus Pathology classification provided 100% accuracy for both CNN and SVM. The Nodule versus Polyp classification provided accuracy above 90%, and the other classifications provided accuracy between 70 and 90%. It was observed that, in a general way, classifications using data augmentation in the training set had better results than classifications without the use of data augmentation, except in classifications involving the Polyp class. In most cases, the use of segment sizes of 10 pitch cycles for image generation provided better results than the classic size of 20 ms. Overall, classification using bottleneck and SVM provided higher results than those using only CNN. The proposed approach proved to be promising for the area of vocal fold pathology recognition, since it provided good results in classifying different types of pathology, which is an arduous task due to the noisy character of the voice signal affected by pathology.

Keywords: Digital voice signal processing, voices affected by pathologies classification, Disordered Voice Databases, Phase Space Reconstruction, Convolutional Neural Networks.

Agradecimentos

Agradeço a Deus, aos meus pais Nustenil e Maria, aos meus irmãos Marinus e Nustenil Segundo, aos meus professores Joseana Macêdo Fachine Regis de Araujo e Herman Martins Gomes, ao Instituto Federal de Alagoas e a todos os que contribuíram para realização dessa etapa na minha vida.

Agradeço também ao meu filho Lucas, razão da minha vida, motivador maior de todos os meus esforços para ser um profissional e uma pessoa melhor.

Agradeço especialmente a Sanny Emanuelle, minha eterna companheira, por ter sido a melhor esposa que eu poderia ter. Agradeço por todo seu companheirismo, dedicação, esforço, amor, motivação. Agradeço por estar ao meu lado em todos os momentos da minha pesquisa, seja para me dar ideias, seja para comemorar resultados alcançados, seja para me consolar em momentos de frustração. Esse doutorado também é seu. Te amo.

Conteúdo

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 1.1 | Caracterização do Problema | 2 |
| 1.2 | Objetivos | 4 |
| 1.2.1 | Objetivo Geral | 4 |
| 1.2.2 | Objetivos Específicos | 4 |
| 1.3 | Hipóteses | 5 |
| 1.4 | Abordagem | 5 |
| 1.5 | Estrutura do Documento | 6 |
| 2 | Fundamentação Teórica | 7 |
| 2.1 | Fisiologia e Patologias das Pregas Vocais | 7 |
| 2.1.1 | Sistema de Produção da Voz | 8 |
| 2.1.2 | Patologias das Pregas Vocais | 9 |
| 2.2 | Pré-processamento do Sinal de Voz | 14 |
| 2.3 | Modelo Linear para a Produção da Voz | 15 |
| 2.4 | Métodos Não Lineares para Extração de Características | 16 |
| 2.4.1 | Espaço de Fase | 16 |
| 2.5 | Classificação | 20 |
| 2.5.1 | Redes Neurais Convolucionais | 20 |
| 2.5.2 | Máquinas de Vetores de Suporte | 24 |
| 2.6 | Considerações Finais | 25 |
| 3 | Revisão Sistemática da Literatura | 26 |
| 3.1 | Resultados da Pesquisa Sistemática | 26 |

| | | |
|----------|---|------------|
| 3.2 | Patologias Estudadas | 28 |
| 3.3 | Bases de Dados | 29 |
| 3.4 | Extração de Características | 32 |
| 3.5 | Seleção de Características | 33 |
| 3.6 | Classificação | 33 |
| 3.7 | Considerações Finais | 50 |
| 4 | Materiais e Métodos | 51 |
| 4.1 | Bases de Dados | 51 |
| 4.2 | Abordagem Proposta | 55 |
| 4.2.1 | Pré-processamento e Extração de Características | 56 |
| 4.2.2 | Arquiteturas da CNN | 59 |
| 4.2.3 | Classificação | 60 |
| 4.2.4 | Análise Estatística | 64 |
| 4.3 | Considerações Finais | 65 |
| 5 | Resultados e Discussão | 66 |
| 5.1 | Análise da Trajetória do Espaço de Fase dos Sinais de Voz | 66 |
| 5.2 | Detecção de Patologia | 68 |
| 5.3 | Identificação de Patologias | 70 |
| 5.4 | Considerações Finais | 78 |
| 6 | Considerações Finais e Sugestões para Pesquisas Futuras | 79 |
| 6.1 | Sumário da Pesquisa | 79 |
| 6.2 | Contribuições da Pesquisa | 81 |
| 6.3 | Propostas para Pesquisas Futuras | 82 |
| A | Metodologia da Revisão Sistemática | 92 |
| B | Descrição das Bases de Dados Utilizadas | 97 |
| C | Códigos-fonte | 112 |
| D | Resumo dos Resultados | 115 |

| | |
|--|------------|
| E Implementações de outros Trabalhos | 119 |
| F Comparação com Resultados Obtidos na Literatura | 135 |

Lista de Siglas e Abreviaturas

| | |
|-------|--|
| 2D | 2 dimensões |
| ACM | <i>Association for Computing Machinery</i> - Associação para Maquinário de Computação |
| CNN | <i>Convolutional Neural Network</i> - Rede Neural Convolutacional |
| GMM | <i>Gaussian Mixture Models</i> - Modelo de Mistura de Gaussianas |
| GRBAS | <i>Grade, Roughness, Breathiness, Asthenia, Strain</i> - Grau, Aspereza, Soprosidade, Astenia, Tensão; |
| HMM | <i>Hidden Markov Models</i> - Modelos de Markov Escondidos |
| IEEE | <i>Institute of Electrical and Electronics Engineers</i> - Instituto de Engenheiros Eletricistas e Eletrônicos |
| MEEI | <i>Massachusetts Eye and Ear Infirmary</i> - Enfermaria de Olho e Ouvido de Massachusetts |
| RBF | <i>Radial Basis Function</i> - Função Base Radial |
| ReLU | <i>Rectified Linear Unit</i> - Unidade Linear Retificada |
| SVD | <i>Saarbruecken Voice Database</i> - Base de Dados de Voz Saarbruecken |
| SVM | <i>Support Vector Machines</i> - Máquinas de Vetores de Suporte |
| TS | <i>Time Stretching</i> - Alongamento de Tempo |

Lista de Figuras

| | | |
|------|---|----|
| 1.1 | Processo de classificação do sinal de voz. | 5 |
| 2.1 | Visão lateral de uma prega vocal. | 8 |
| 2.2 | Prega vocal afetada por Edema de Reinke. | 10 |
| 2.3 | Prega vocal afetada por Paralisia unilateral à esquerda. | 11 |
| 2.4 | Prega vocal afetada por Queratose (leucoplasia). | 12 |
| 2.5 | Prega vocal afetada por Nódulos. | 13 |
| 2.6 | Prega vocal afetada por Pólipos. | 13 |
| 2.7 | Prega vocal afetada por Cisto. | 13 |
| 2.8 | Modelo Linear para Produção da Fala. | 15 |
| 2.9 | Espaço de fase de um pêndulo sob o efeito de atrito. | 17 |
| 2.10 | Trajетórias dos espaços de fase da função seno (a) sem ruído e (b) com ruído aleatório. | 18 |
| 2.11 | Curva de Informação Mútua do sinal CAC10AN (Edema) da base de dados MEEI. | 20 |
| 2.12 | Funcionamento da camada convolucional com tamanho do <i>kernel</i> = 3 e <i>stride</i> = 2. | 21 |
| 2.13 | (a) Entrada e (b) saída de uma camada <i>maxpooling</i> com tamanho do <i>kernel</i> = 2 e <i>stride</i> = 2. | 23 |
| 2.14 | Camada <i>bottleneck</i> ao centro da figura. | 24 |
| 3.1 | Quantidade de pesquisas para cada grupo de bases de dados, em função da base de dados utilizada. | 31 |

| | | |
|-----|--|----|
| 4.1 | Time Stretching das primeiras 500 amostras do sinal (a) 1-a_n.wav da base SVD, com (b) $\alpha = 0,8$, (c) $\alpha = 0,9$, (d) $\alpha = 1,1$ e (e) $\alpha = 1,2$ | 57 |
| 4.2 | Trajectoria do Espaço de Fase Reconstruído dos primeiros 20 ms de (a) um sinal de voz normal (AXH1NAL), (b) um sinal de voz afetada por Edema (CAK25AN), (c) um sinal de voz afetada por Paralisia (RAN30AN), (d) um sinal de voz afetada por Nódulo (MXN24AN), (e) um sinal de voz afetada por Pólipo (MPB23AN) and (f) um sinal de voz afetada por Queratose (EMP27AN). Todos os sinais pertencem à base MEEI. | 58 |
| 4.3 | Método da contagem de caixas com $N = 2$ e pixel preenchido = 255. | 59 |
| 4.4 | (a) Trajetória do espaço de fase reconstruído dos primeiros 20 ms de um sinal de voz afetada por Edema (CAC10AN) da base MEEI e a imagem após a redução de dimensionalidade utilizando (b) $N = 10$, e (c) $N = 15$ | 59 |
| 4.5 | Arquitetura da CNN proposta. | 60 |
| 4.6 | Sistema de classificação hierárquica do sinal de voz. | 60 |
| 5.1 | Trajectoria do Espaço de Fase Reconstruído dos primeiros 20 ms de (a) um sinal de voz normal (1-a_n), (b) um sinal de voz normal (2-a_n), (c) um sinal de voz afetada por Paralisia (929-a_n) e (d) um sinal de voz afetada por Edema (1963-a_n). Todos os sinais pertencentes à base SVD. | 66 |
| 5.2 | Trajectoria do Espaço de Fase Reconstruído de um sinal de voz normal (1-a_n) da base SVD, com janela de tamanho (a) 20 ms e (b) 10 ciclos de <i>pitch</i> | 67 |
| 5.3 | Trajectoria do Espaço de Fase Reconstruído dos primeiros 20 ms de (a) um sinal de voz afetada por Paralisia (929-a_n) e (d) um sinal de voz afetada por Paralisia (130-a_n). Todos os sinais pertencentes a base SVD. | 67 |
| 5.4 | Trajectoria do Espaço de Fase Reconstruído dos primeiros 20 ms de duas elocuições de um mesmo sujeito com voz normal: (a) sinal 10-a_n e (b) sinal 2200-a_n da base SVD. | 68 |
| 5.5 | Saída da última camada convolucional para janelas de (a) e (b) dois sinais de voz normal e (c) e (d) dois sinais de voz afetada por patologia, todos da base MEEI. | 69 |

| | | |
|-----|---|----|
| 5.6 | Saída da última camada convolucional para janelas de (a) e (b) dois sinais de voz afetada por Paralisia e (c) e (d) dois sinais de voz afetada por Edema, todos da base MEEI. | 71 |
| 5.7 | Saída da última camada convolucional para janelas de (a) e (b) dois sinais de voz afetada por Paralisia e (c) e (d) dois sinais de voz afetada por Queratose, todos da base MEEI. | 72 |
| A.1 | Etapas da revisão sistemática. | 93 |

Lista de Tabelas

| | | |
|-----|--|-----|
| 3.1 | Resultados da busca e seleção de pesquisas, com as quantidades dos artigos mantidos. | 27 |
| 3.2 | Patologias das pregas vocais estudadas. | 29 |
| 3.3 | Quantidade de pesquisas agrupadas por base de dados. | 30 |
| 4.1 | Total de sessões de patologias das pregas vocais de cada base de dados. . . | 52 |
| 4.2 | Total de imagens por classe e base de dados. | 56 |
| 5.1 | Melhores Resultados da Classificação Normal <i>versus</i> Patologia. | 69 |
| 5.2 | Valor P do teste t de Student dos resultados MCC da classificação Normal <i>versus</i> Patologia. | 70 |
| 5.3 | Melhores Resultados de cada Classificação na base MEEI. | 73 |
| 5.4 | Melhores Resultados de cada Classificação na base SVD. | 73 |
| 5.5 | Valores P do teste de Shapiro-Wilk para verificação da normalidade dos resultados MCC das classificações entre patologias. | 74 |
| 5.6 | P-value do teste t de Student dos resultados MCC das classificações entre patologias. | 76 |
| 5.7 | Resultados obtidos nas implementações dos trabalhos para a base MEEI. . . | 77 |
| 5.8 | Resultados obtidos nas implementações dos trabalhos para a base SVD. . . | 77 |
| B.1 | Descrição das bases de dados MEEI e SVD. | 97 |
| B.2 | Descrição do subconjunto MEEI. | 99 |
| B.3 | Descrição do subconjunto SVD. | 100 |
| B.4 | Sinais de voz do subconjunto MEEI. | 102 |
| B.5 | Sinais de voz do subconjunto SVD. | 104 |

| | | |
|------|---|-----|
| D.1 | Melhores Resultados da Classificação Paralisia <i>versus</i> Edema. | 115 |
| D.2 | Melhores Resultados da Classificação Paralisia <i>versus</i> Queratose. | 115 |
| D.3 | Melhores Resultados da Classificação Paralisia <i>versus</i> Pólipo. | 116 |
| D.4 | Melhores Resultados da Classificação Paralisia <i>versus</i> Nódulo. | 116 |
| D.5 | Melhores Resultados da Classificação Paralisia <i>versus</i> Lesão Benigna. . . . | 116 |
| D.6 | Melhores Resultados da Classificação Edema <i>versus</i> Queratose. | 116 |
| D.7 | Melhores Resultados da Classificação Edema <i>versus</i> Pólipo. | 117 |
| D.8 | Melhores Resultados da Classificação Edema <i>versus</i> Nódulo. | 117 |
| D.9 | Melhores Resultados da Classificação Edema <i>versus</i> Lesão Benigna. | 117 |
| D.10 | Melhores Resultados da Classificação Queratose <i>versus</i> Pólipo. | 117 |
| D.11 | Melhores Resultados da Classificação Queratose <i>versus</i> Nódulo. | 118 |
| D.12 | Melhores Resultados da Classificação Queratose <i>versus</i> Lesão Benigna. . . | 118 |
| D.13 | Melhores Resultados da Classificação Nódulo <i>versus</i> Pólipo. | 118 |
| F.1 | Resumo dos resultados de trabalhos publicados de 2011 até 2018 ordenados por classificação e por acurácia. N.I. significa "Não Informado". | 136 |

Lista de Quadros

| | | |
|-----|---|----|
| 3.1 | Resumo das Pesquisas. | 35 |
| 4.1 | Classificações realizadas na pesquisa. | 61 |
| A.1 | Artefatos gerados pela revisão sistemática. | 93 |

Lista de Códigos-Fonte

| | | |
|-----|---|-----|
| C.1 | Código-fonte da função que obtém o primeiro mínimo da curva de informação mútua média | 112 |
| C.2 | Código-fonte da função que realiza o Time Stretching para o aumento de dados | 112 |
| C.3 | Código-fonte da função que implementa o método da contagem de caixas | 113 |
| E.1 | Código-fonte da implementação do trabalho de Sodre, Rosa e Dassie-Leite (2017). | 119 |
| E.2 | Código-fonte da implementação do trabalho de Dahmani e Guerti (2017). | 123 |
| E.3 | Código-fonte da implementação do trabalho de Fang et al. (2018). | 126 |
| E.4 | Código-fonte da implementação do trabalho de Ali et al. (2018). | 131 |

Capítulo 1

Introdução

Atualmente, cerca de 25% da população exerce alguma profissão que exige vocalização em excesso ou atuam em ambientes com alto nível de ruído, sendo forçados a elevar o nível da voz (AL-NASHERI et al., 2018) como, por exemplo, professores, advogados e cantores, dentre outros. A extensão do grupo de risco atesta a importância de pesquisas na identificação de patologias nas pregas vocais.

O diagnóstico inicial de patologias nas pregas vocais é usualmente realizado de maneira subjetiva, em que o médico escuta o paciente e realiza uma análise subjetiva do som produzido, baseada em sua experiência (ALI et al., 2015; CRESPO et al., 2018). O diagnóstico preciso é realizado a partir de métodos laboratoriais, tais como laringoscopia (exame com um instrumento de fibra óptica), glotografia (monitoramento do movimento das pregas vocais a partir de uma câmera), videoestroboscopia (iluminação estroboscópica da laringe, útil para visualização dos movimentos), eletromiografia (observação indireta do estado funcional da laringe) e videoquimografia (avalia a vibração das pregas vocais a partir de uma câmera) (KUKHARCHIK et al., 2007). Entretanto, tais métodos são invasivos, pois causam desconforto ao paciente, podendo aumentar sua resistência durante o exame, além do alto custo de manutenção, uma vez que os equipamentos têm que ser esterilizados após cada utilização, consumindo tempo e recursos (CORDEIRO; FONSECA; RIBEIRO, 2014).

1.1 Caracterização do Problema

Nos últimos anos, várias pesquisas na área de Processamento Digital da Voz vêm sendo realizadas com o objetivo de avaliar a qualidade da voz do paciente e auxiliar especialistas no diagnóstico de patologias nas pregas vocais (HEMMERLING; SKALSKI; GAJDA, 2016; LOPES et al., 2017; MUHAMMAD et al., 2018). A análise acústica da voz pode ser uma ferramenta eficiente para o diagnóstico de patologias e possui como vantagem o fato de não ser invasiva.

A área de pesquisa em Processamento Digital de Sinais de vozes afetadas por patologias pode ser dividida em três principais campos de estudo (MARTÍNEZ et al., 2012): (i) a detecção de patologias, que consiste na diferenciação do sinal de voz entre voz normal e voz afetada por patologia; (ii) a verificação da qualidade da voz, por meio de uma escala perceptual, como a escala GRBAS (*Grade, Roughness, Breathiness, Asthenia, Strain* - Grau, Aspreza, Soprosidade, Astenia, Tensão) e (iii) a identificação de uma ou mais patologias específicas. A pesquisa ora descrita foca na detecção e na identificação de patologias nas pregas vocais. Portanto, nos parágrafos que se seguem, são discutidos apenas os itens (i) e (iii).

No campo de estudo de detecção de patologias, várias pesquisas conseguiram resultados expressivos. Por exemplo, Hariharan et al. (2013) conseguiram obter taxas de acerto acima de 99% para sinais de voz de baixa qualidade, utilizando pacotes wavelet e k-means. Hemmerling, Skalski e Gajda (2016) obtiveram acurácia de 100%, utilizando um vetor de características composto por 28 parâmetros e “Florestas Aleatórias” (*Random Forest*), como classificador.

Esforços foram conduzidos no sentido de destacar uma patologia em relação às demais como, por exemplo, o Edema de Reinke (CORDEIRO et al., 2015; SODRÉ; ROSA; DASSIE-LEITE, 2017), a Paralisia nas pregas vocais (MUHAMMAD et al., 2016; SODRÉ; ROSA; DASSIE-LEITE, 2017; LOPES et al., 2017; AL-NASHERI et al., 2018) e Nódulos nas pregas vocais (CORDEIRO et al., 2015; SODRÉ; ROSA; DASSIE-LEITE, 2017; ALI et al., 2018). Apesar do progresso na área da classificação de patologias da voz reportado anteriormente, ainda é possível vislumbrar a necessidade de pesquisas, com vistas ao amadurecimento de metodologias e técnicas referentes à extração e à seleção das melhores ca-

racterísticas para modelar o problema. Aliado a isto, destaca-se que ainda é uma tarefa árdua diferenciar uma patologia das demais, devido ao caráter ruidoso dos sinais de vozes afetadas por patologias, à similaridade entre as patologias, dentre outros (VERDOLINI; ROSEN; BRANSKI, 2014).

Diante do exposto, acrescenta-se que a maioria das pesquisas em detecção e/ou classificação de patologias da voz se concentra em uma única base de dados (COSTA, 2008; MARINUS, 2010; BRANDT, 2012; COSTA et al., 2012; MUHAMMAD; MELHEM, 2013; ESPINOLA, 2014; MUHAMMAD; MELHEM, 2014; ALI et al., 2016; CORDEIRO et al., 2017; ALI et al., 2018) a base de dados da *Massachusetts Eye and Ear Infirmary* (MEEI) (KAY-ELEMETRICS, 1994). Algumas pesquisas utilizaram outras bases de dados (MARTÍNEZ et al., 2012; CORDEIRO; FONSECA; RIBEIRO, 2014; MUHAMMAD et al., 2018) e, ultimamente, mais de uma base de dados para comparação (MUHAMMAD et al., 2016; AL-NASHERI et al., 2017, 2018). A maioria das pesquisas que compara mais de uma base de dados utiliza uma de cada vez. Soma-se ao fato de que nem todas as bases utilizadas são disponíveis ao público, tornando impossível a reprodutibilidade de pesquisas. Diante do exposto, surge a questão: as técnicas de extração de características e de classificação estudadas para modelar patologias da voz são dependentes de uma base de dados?

Um ponto em comum em relação às bases de dados de distúrbios da voz é a quantidade limitada de sinais de voz com uma determinada patologia. Por exemplo, a base MEEI só possui 03 sinais afetados por sulco vocal, dificultando, assim, estudos sobre essa patologia. Devido a isso, esta pesquisa focou apenas nas patologias das pregas vocais com as maiores quantidades de sinais disponíveis nas bases de dados utilizadas. As patologias são: Edema nas pregas vocais, Paralisia nas pregas vocais, Queratose nas pregas vocais, Nódulo nas pregas vocais e Pólipo nas pregas vocais. Também foi pesquisada uma classe de patologias denominada Lesão Benigna nas pregas vocais, que consiste no agrupamento das patologias Nódulo e Pólipo, junto com Cistos nas pregas vocais. Ainda assim, a quantidade de sinais de voz é limitada em cada classe. Nesta pesquisa, foi investigada uma abordagem para aumento de dados com o objetivo de tentar reduzir essa limitação.

Deve-se destacar também as mudanças fisiológicas causadas pelas patologias. Tradicionalmente, utiliza-se o modelo linear de produção da voz nas análises acústicas da voz. Esse modelo considera que o sinal de voz é invariante ao tempo, assumindo características

estacionárias em curtos intervalos de tempo (SOTOMAYOR, 2003). Titze (1995) sugeriu a adoção de tipos de sinais de voz para a análise acústica. O tipo 1 corresponde aos sinais quase periódicos. O tipo 2 corresponde aos sinais com fortes modulações ou sub-harmônicas e o tipo 3 corresponde aos sinais irregulares e aperiódicos. A análise linear do sinal de voz é recomendada apenas para sinais do tipo 1. Para sinais dos tipos 2 e 3, Titze (1995) recomenda a utilização de métodos de dinâmica não linear. As desordens vocais ocasionadas por patologias devem ocasionar mudanças de comportamento nas características temporais e espectrais do sinal de voz (COSTA, 2012). Dependendo do grau de severidade da patologia, o sinal de voz afetada por patologia pode ser do tipo 2 ou do tipo 3.

Diante do exposto, o objetivo desta pesquisa de doutorado consistiu no estudo de diferentes métodos, a fim de propor uma nova abordagem para melhorar a classificação de sinais de voz afetada por patologia.

1.2 Objetivos

Nesta seção, estão descritos o objetivo geral da presente pesquisa e os objetivos específicos.

1.2.1 Objetivo Geral

O principal objetivo da presente pesquisa foi realizar uma análise acústica do sinal de voz afetado por patologia utilizando modelos não lineares, com o objetivo de identificar a ocorrência de patologia e classificar o sinal de voz afetada por patologia do tipo Edema, Paralisia, Queratose, Nódulo, Pólipo ou Lesão Benigna.

1.2.2 Objetivos Específicos

Diante do objetivo geral, foram determinados os objetivos específicos a seguir.

- Investigar métodos para aumento de dados das bases de dados disponíveis;
- Investigar métodos de pré-processamento que possam melhorar a classificação das diferentes classes de patologias nas pregas vocais;
- Investigar e criar métodos para extração de características baseados em modelos não lineares que modelem classes de patologias nas pregas vocais;
- Investigar e criar métodos de classificação que separem classes de patologias nas pregas vocais.

1.3 Hipóteses

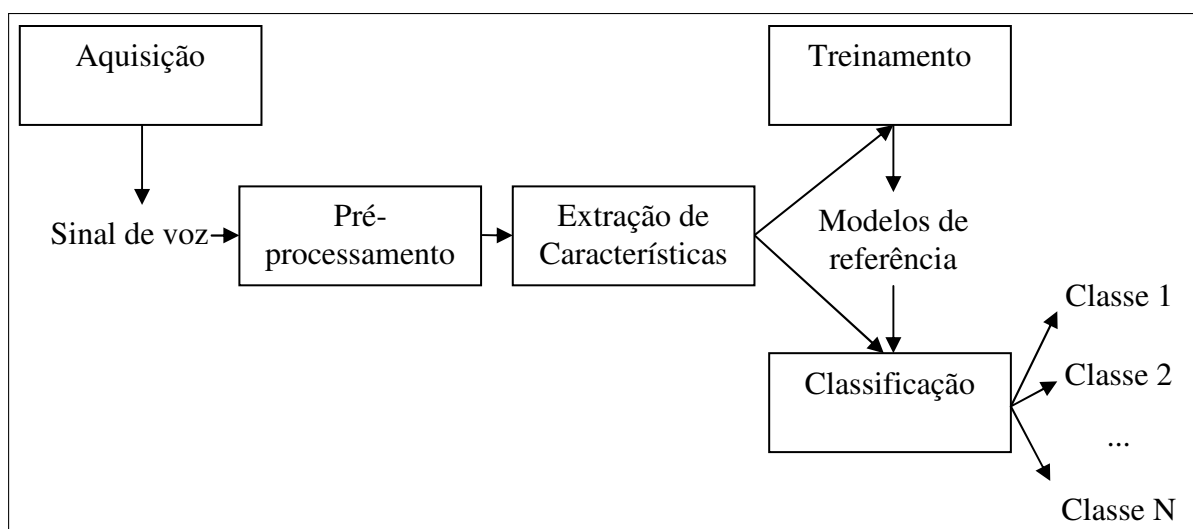
A partir dos objetivos, as hipóteses principais desta pesquisa são apresentadas a seguir.

- **Hipótese 1:** O uso de técnicas de aumento de dados proporciona melhores resultados em termos de MCC (Coeficiente de Correlação de Matthews - *Matthews Correlation Coefficient*) do que a não utilização.
- **Hipótese 2:** Os resultados em termos de MCC de classificações de patologias das pregas vocais são melhores quando se usa janelas de tamanho variável do que quando se usam janelas de tamanho fixo para o método de extração de características utilizado.
- **Hipótese 3:** O uso de *bottleneck*+SVM para classificação proporciona melhores resultados em termos de MCC que o uso de CNN para o método de extração de características utilizado.

1.4 Abordagem

O processo de classificação do sinal de voz pode ser dividido em quatro etapas principais: aquisição, pré-processamento, extração de características e treinamento/classificação, conforme mostrado na Figura 1.1.

Figura 1.1: Processo de classificação do sinal de voz.



Fonte: Autoria própria.

Nesta pesquisa, não foi abordada a fase de aquisição. Os sinais de voz foram adquiridos a partir de bases de dados públicas.

O sinal de voz é obtido na fase de aquisição, em que a voz é gravada e digitalizada. Em seguida, é realizado o pré-processamento do sinal de voz, como pré-ênfase e janelamento, em que o sinal é preparado para a etapa seguinte. Na etapa de extração de características, são computadas medições sobre o sinal de voz, de modo a caracterizá-lo para fins de classificação. A partir das características obtidas, um classificador é treinado para obter modelos de referência de uma ou mais classes. Na etapa de classificação, as características extraídas do sinal são comparadas com os modelos de referência, para assim obter a classe à qual o sinal de voz pertence.

1.5 Estrutura do Documento

O restante do documento é organizado como se segue. No Capítulo 2, é apresentada a fundamentação teórica da pesquisa, incluindo a descrição da fisiologia de produção da voz e de diversas patologias que afetam as pregas vocais.

Também é apresentada a descrição dos modelos matemáticos empregados nos experimentos de classificação de sinais de voz, os modelos matemáticos para extração de características (Reconstrução do Espaço de Fase) e para classificação utilizados (Redes Neurais Convolucionais e Máquina de Vetores de Suporte).

No Capítulo 3, é apresentada uma revisão sistemática da literatura na área de classificação de patologias das pregas vocais pelo sinal de voz.

No Capítulo 4, é apresentada uma descrição mais detalhada da abordagem proposta, incluindo informações sobre as bases de dados utilizada nos experimentos.

No Capítulo 5, são apresentados e analisados os resultados obtidos no processo de classificação e no Capítulo 6 são apresentadas as considerações finais e sugestões para pesquisas futuras.

No Apêndice A, encontra-se a metodologia utilizada para a revisão sistemática da literatura. No Apêndice B, encontram-se as tabelas com a identificação dos sinais de voz utilizados na base de dados, além de algumas características dos locutores, tais como sexo e faixa etária. No Apêndice C, encontram-se códigos-fonte de funções implementadas. No Apêndice D, encontram-se tabelas com resultados obtidos na pesquisa e, no Apêndice E, estão códigos-fonte de implementações e resultados de outros trabalhos.

Capítulo 2

Fundamentação Teórica

Neste capítulo, objetiva-se proporcionar um embasamento teórico sobre o problema abordado na pesquisa, que é a análise de sinais de voz afetada por patologias das pregas vocais.

Para esta pesquisa, foram utilizados métodos não lineares para a extração de características, consistindo na obtenção de imagens a partir da trajetória do espaço de fases reconstruído do sinal. Para a classificação, Redes Neurais Convolucionais foram empregadas, tanto como classificador quanto como um segundo passo da extração de características, utilizando uma camada *bottleneck*, para posterior classificação utilizando SVM.

Na primeira seção deste capítulo, é apresentado um estudo sobre o processo de produção da voz, com a descrição do funcionamento dos principais órgãos e uma abordagem sobre patologias nas pregas vocais. Nas seções seguintes, é apresentado um estudo sobre o pré-processamento e a extração de características do sinal de voz, treinamento de classificadores utilizados neste trabalho.

2.1 Fisiologia e Patologias das Pregas Vocais

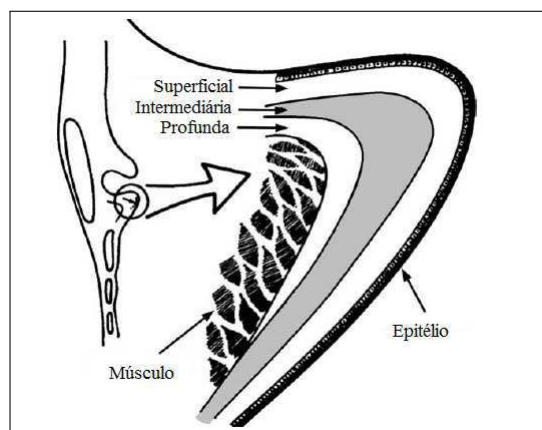
O sistema produtor da voz humana consiste na junção de partes do sistema digestivo e do sistema respiratório, com maior importância na produção da voz humana para esse último. Nas próximas subseções, serão descritos o sistema de produção da voz e as patologias que afetam as pregas vocais.

2.1.1 Sistema de Produção da Voz

Os órgãos que compõem o sistema de produção de voz são: pulmão, traqueia, laringe, faringe, e cavidades nasais e oral. Inicialmente, o ar é inalado nos pulmões na respiração, e é expelido, passando pela traqueia e chegando na laringe. Na laringe, estão localizadas as pregas vocais, que podem vibrar com a passagem do ar vindo dos pulmões. Em seguida, o som formado pelas pregas vocais é modificado pelo trato vocal, que é constituído pela faringe, cavidade oral e cavidades nasais (RIBEIRO, 2003).

As pregas vocais (Figura 2.1) são estruturas compostas por músculo e mucosa. A mucosa pode ser dividida em epitélio e lâmina própria, que é subdividida em camadas superficial (espaço de Reinke), intermediária e profunda (GRAY, 1991). O mecanismo vibratório das pregas vocais origina-se de ondulações nos sentidos horizontal, vertical e longitudinal, determinando o ciclo glótico com suas fases de abdução e adução. O comprimento das pregas vocais varia de tamanho no controle da frequência da voz, e a pressão subglótica varia durante o controle da intensidade da voz. O trato vocal é modificado totalmente durante o controle da qualidade vocal (TUMA et al., 2005).

Figura 2.1: Visão lateral de uma prega vocal.



Fonte: Voice (2017).

A qualidade vocal depende do modo de fechamento e abertura da glote e da vibração das pregas vocais, dentre outros. A vibração das pregas vocais é determinada principalmente por (BHATTACHARYA; KELLEHER; SIEGMUND, 2015):

- A extensão em que as pregas vocais são aduzidas ou abduzidas ou a posição da prega vocal;

- Grau de elasticidade das pregas vocais, que é determinado pela posição e grau de tensão do músculo vocal durante a contração; e
- Nível de pressão do ar por meio das pregas vocais.

Em determinadas condições, que podem ser causadas por diferentes patologias nas pregas vocais, as vibrações das pregas vocais podem se tornar aperiódicas ou irregulares. São elas: assimetria, interferência na homogeneidade, flacidez, tono oscilante e força inconsistente (COSTA, 2008).

2.1.2 Patologias das Pregas Vocais

O surgimento de patologias nas pregas vocais ocorre por diversos motivos, tais como acidentes, mau uso da voz e doenças, dentre outros (LITTLE et al., 2006). As patologias podem ser funcionais, quando ocorrem devido ao mau uso da voz, orgânicas, quando existem alterações na anatomia das pregas vocais, ou organofuncionais, quando são ao mesmo tempo funcionais e orgânicas (BEHLAU, 2005).

Verdolini, Rosen e Branski (2014) dividiram os diferentes distúrbios da voz em 09 categorias, de acordo com a origem do distúrbio e a patologia. As categorias são:

1. Patologias estruturais da laringe;
2. Inflamações na laringe;
3. Trauma na laringe;
4. Condições sistêmicas que afetam a voz;
5. Distúrbios aerodigestivos que não ocorrem na laringe e afetam a voz;
6. Distúrbios psicológicos e psiquiátricos que afetam a voz;
7. Distúrbios neurológicos que afetam a voz;
8. Outros distúrbios que afetam a voz; e
9. Distúrbios da voz sem diagnóstico ou não especificado.

Nesta pesquisa, foram escolhidas 06 patologias para estudo: Edema de Reinke, Paralisia, Queratose, Nódulo, Pólipo e Cisto, sendo que Nódulo, Pólipo e Cisto estão em uma subcategoria denominada **Lesões Benignas Focais da Lâmina Própria** da categoria 1. As cinco primeiras foram escolhidas por serem as patologias com a maior quantidade de sinais nas bases de dados utilizadas na pesquisa dentre as patologias que afetam as pregas vocais e cada uma representa uma classe distinta. Sinais de voz afetada por cisto foram utilizadas apenas em conjunto com sinais de voz afetada por nódulo ou pólipo, em uma classe denominada **lesão benigna**.

Edema de Reinke

O Edema de Reinke, também conhecido como Degeneração Polipoide, é uma lesão difusa, em que se acumula fluido por toda a camada superficial da lâmina própria (BEHLAU, 2005). Geralmente, é bilateral e algumas vezes assimétrico (VERDOLINI; ROSEN; BRANSKI, 2014). As causas podem ser o uso em excesso da voz e exposição ao cigarro, a remédios ou a ocorrência de refluxo gástrico (CIELO et al., 2011).

Dentre as características da voz afetada por Edema de Reinke, tem-se a frequência fundamental normalmente abaixo da média e a presença de rouquidão e aspereza na voz. As características são mais acentuadas em mulheres do que em homens (VERDOLINI; ROSEN; BRANSKI, 2014).

Na Figura 2.2, é apresentada uma prega vocal afetada por Edema de Reinke.

Figura 2.2: Prega vocal afetada por Edema de Reinke.



Fonte: Cóser (2017).

Paralisia nas Pregas Vocais

O nervo vago é o nervo mais longo dentre os que têm origem no crânio humano, indo do tronco cerebral ao abdômen (SOM; CURTIN, 2011). Durante o trajeto, ele se ramifica na laringe formando o nervo laríngeo superior e os nervos laríngeos recorrentes esquerdo e direito. Os últimos são os responsáveis pelo controle das pregas vocais (GARCIA et al., 2009).

A Paralisia nas Pregas Vocais ocorre quando um ou os dois nervos recorrentes são lesionados (BROADHURST, 2015). Portanto, a Paralisia pode ser unilateral ou bilateral, a depender da quantidade de nervos lesionados (VERDOLINI; ROSEN; BRANSKI, 2014). As causas são diversas. Uma das causas mais comuns é o contágio do paciente com uma doença viral, tal como um resfriado. Outras causas podem ser cirurgias no pescoço e problemas no coração (BROADHURST, 2015).

Na Figura 2.3, mostra-se uma prega vocal afetada por Paralisia.

Figura 2.3: Prega vocal afetada por Paralisia unilateral à esquerda.



Fonte: Cóser (2017).

Queratose nas Pregas Vocais

Queratose nas Pregas Vocais é uma mudança na camada epitelial que reveste as pregas vocais, em que ocorre o surgimento de células duplicadas (VERDOLINI; ROSEN; BRANSKI, 2014), sendo assim caracterizada como uma lesão pré-maligna (PINTO et al., 2012). Existem três tipos de Queratose conforme a aparência (PINTO et al., 2012): eritroplasia (cor

avermelhada); leucoplasia (cor esbranquiçada); eritroqueratose (contém lesões das duas cores).

Na Figura 2.4, mostra-se uma prega vocal afetada por Queratose.

Figura 2.4: Prega vocal afetada por Queratose (leucoplasia).



Fonte: Sarmiento (2017).

Lesões Benignas Focais da Lâmina Própria

Lesões Benignas Focais da Lâmina Própria são um dos tipos mais comuns de lesão nas pregas vocais, sendo também o tipo mais controverso na área de estudo de distúrbios da voz. No passado, as diferentes lesões eram diagnosticadas como sendo Nódulo ou Pólipo apenas. Com o surgimento de exames laboratoriais com maior precisão, aumentou-se a capacidade de diferenciação entre as diferentes lesões, que podem ser Nódulos, Pólipos, Cistos, massas fibrosas e lesões reativas. Entretanto, não existe um consenso com relação às definições das lesões, o que ocasiona diagnósticos inconsistentes entre os diferentes centros clínicos (VERDOLINI; ROSEN; BRANSKI, 2014).

Dentre as características mais aceitas, tem-se que Nódulos e Pólipos são massas benignas pequenas na lâmina própria, com a diferença de que Pólipos são mais fluidos e geralmente são unilaterais, enquanto que Nódulos geralmente são bilaterais (VERDOLINI et al., 1998). Os Cistos são pequenas estruturas encapsuladas que podem ocorrer na camada superficial da lâmina própria (Cisto subepitelial) ou em camadas mais profundas (Cisto de Ligamento) (VERDOLINI; ROSEN; BRANSKI, 2014).

Nas Figuras 2.5, 2.6 e 2.7, mostram-se pregas vocais afetadas por diferentes Lesões Benignas.

Figura 2.5: Prega vocal afetada por Nódulos.



Fonte: Cóser (2017).

Figura 2.6: Prega vocal afetada por Pólipos.



Fonte: Cóser (2017).

Figura 2.7: Prega vocal afetada por Cisto.



Fonte: Cóser (2017).

2.2 Pré-processamento do Sinal de Voz

Após a aquisição do sinal de voz, é realizada uma preparação do sinal para a fase de extração de características. Essa preparação é denominada **Pré-processamento**. Nesta pesquisa, o pré-processamento consistiu na segmentação e no janelamento do sinal. Esses métodos de pré-processamento estão descritos nas subseções a seguir.

Segmentação e Janelamento do Sinal de Voz

Após a aquisição, o sinal é dividido em vários segmentos. A segmentação é realizada porque o sinal de voz varia estatisticamente com o tempo (ALENCAR, 2005), mas assume características estacionárias a curtos intervalos de tempo, normalmente entre 16 e 32 ms (RABINER; JUANG, 1993; SOTOMAYOR, 2003). O janelamento do sinal de voz consiste na multiplicação do segmento por uma função janela no domínio do tempo. Portanto, o espectro do sinal janelado é a convolução do espectro do sinal com o espectro da função janela (HAYKIN; VEEN, 2002), ou seja, o janelamento modifica o sinal tanto no domínio do tempo quanto no domínio da frequência.

Dentre as funções de janelamento existentes, a utilizada na pesquisa foi a de *Hamming*. O janelamento de Hamming (Equação 2.1) consiste em aplicar uma função janela cossenoidal ao sinal de voz, mantendo as características espectrais do centro do segmento, e eliminando as transições abruptas das extremidades.

$$J(n) = \begin{cases} 0,54 - 0,46\cos\left(\frac{2\pi n}{N_A-1}\right) & , \text{ se } 0 \leq n \leq N_A - 1; \\ 0 & , \text{ caso contrário} \end{cases} \quad (2.1)$$

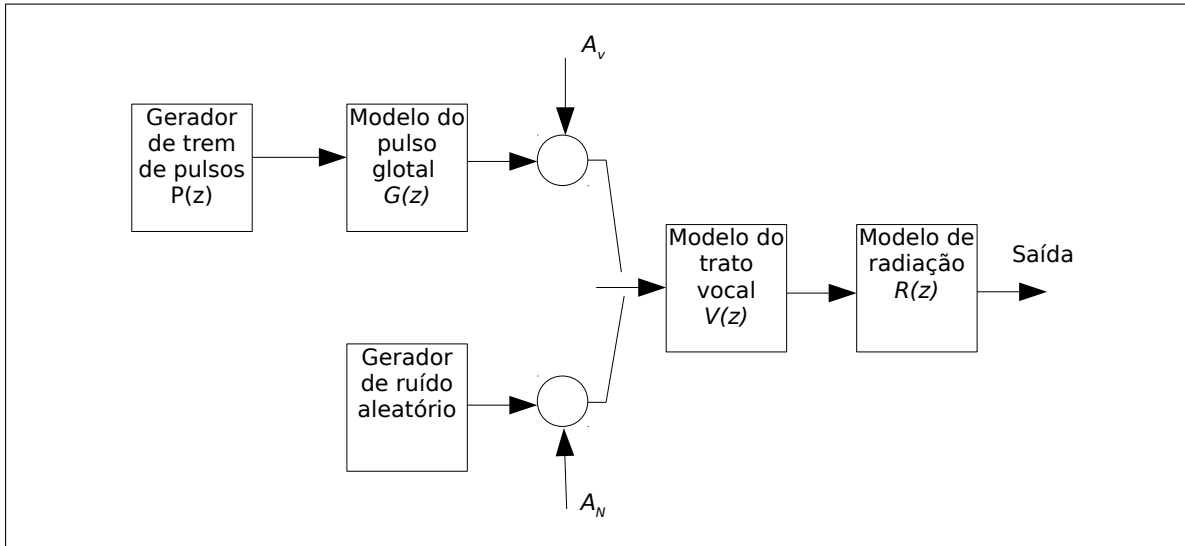
em que $J(n)$ é o valor da janela na posição n e N_A o tamanho da janela.

Como a janela de Hamming proporciona perda espectral nas extremidades, faz-se necessário fazer uso de sobreposição dos segmentos para que essas características não se percam quando as janelas são utilizadas.

2.3 Modelo Linear para a Produção da Voz

Um modelo usualmente aceito para produção da fala é o Modelo Linear (QUATIERI, 2001) (Figura 2.8). Esse modelo descreve a fala como sendo formada por um gerador de trem de pulsos para sons vozeados e por ruído aleatório para sons não vozeados.

Figura 2.8: Modelo Linear para Produção da Fala.



Fonte: Deller, Hansen e Proakis (2000).

O modelo matemático da produção de voz possui dois geradores, um gerador de trem de pulsos para os sinais sonoros e um gerador de ruído aleatório para os sons surdos. Um sinal sonoro, como a vogal sustentada /a/, por exemplo, consiste na convolução do trem de pulsos $p(t)$ com uma resposta ao impulso $h(t)$ (Equação 2.2).

$$s(t) = p(t) * h(t). \quad (2.2)$$

A voz é produzida pela passagem do ar vindo dos pulmões nas pregas vocais, trato vocal e, por fim, cavidades oral e/ou nasal. Portanto, a função de transferência $H(z)$ da resposta ao impulso consiste na Equação 2.3.

$$H(z) = A_v G(z) V(z) R(z), \quad (2.3)$$

em que A_v é o ganho da fonte, $G(z)$ o fluxo glotal, $V(z)$ o modelo do trato vocal e $R(z)$ a radiação do som pelos lábios.

2.4 Métodos Não Lineares para Extração de Características

O sistema de produção de voz descrito na Seção 2.3 considera que o sinal de voz é invariante ao tempo, sendo que o sinal só assume características estacionárias em curtos intervalos de tempo (SOTOMAYOR, 2003). Titze (1995) sugeriu a adoção de três tipos de sinais de voz para a análise acústica. O tipo 1 corresponde aos sinais quase periódicos. O tipo 2 corresponde aos sinais com fortes modulações ou sub-harmônicos, e o tipo 3 corresponde aos sinais irregulares e aperiódicos. A análise linear do sinal de voz é recomendada apenas para sinais do tipo 1. Para sinais dos tipos 2 e 3, Titze (1995) recomenda a utilização de métodos de dinâmica não linear. As desordens vocais ocasionadas por patologias devem ocasionar mudanças de comportamento nas características temporais e espectrais do sinal de voz (COSTA, 2012). Dependendo do grau de severidade da patologia, o sinal de voz afetada por patologia pode ser do tipo 2 ou do tipo 3.

Nas décadas passadas, diferentes estudos verificaram a existência de caos na produção da voz. Um sistema caótico possui um comportamento pseudoaleatório que é ao mesmo tempo determinístico e não linear (JIANG; ZHANG; MCGILLIGAN, 2006). A saída de um sistema caótico é imprevisível e muito sensível aos fatores de controle, o que significa que pequenas alterações em suas condições iniciais podem ocasionar enormes mudanças nos resultados finais. Ainda assim, o comportamento caótico pode ser descrito de modo quantitativo a partir de métodos baseados na dinâmica não linear (TITZE, 1993).

Na Subseção a seguir, é visto o método não linear utilizado na pesquisa para extração de características.

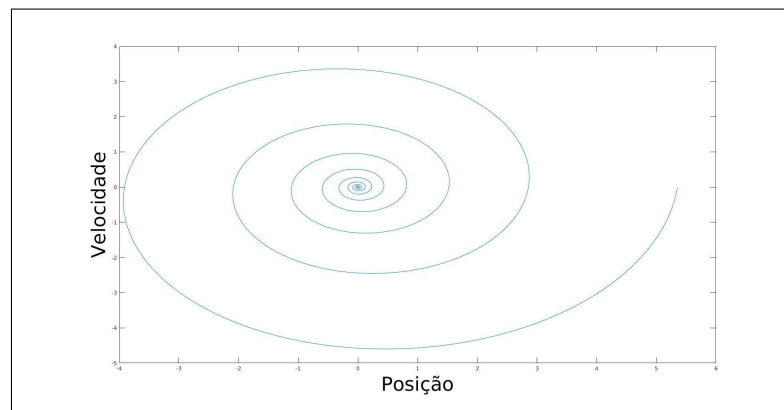
2.4.1 Espaço de Fase

Um sistema dinâmico consiste em um modelo matemático empregado para representar diversos sistemas (SAVI, 2006; SOUZA, 2008) e é caracterizado por estados que mudam com o tempo (MONTEIRO, 2002).

O espaço de fase consiste na representação matemática de um sistema dinâmico e na sua tradução em uma imagem contendo todas as informações do sistema em um dado ins-

tante (GLEISER, 2002). Com o espaço de fases, é possível investigar a evolução de um sistema dinâmico ao longo do tempo a partir das variáveis de estado (OTT, 2002). Por exemplo, em um sistema de duas variáveis (x, y) , o espaço de fase consiste em um gráfico cartesiano em que cada ponto contém a situação do sistema em cada instante. Por exemplo, um pêndulo sob o efeito de atrito com posição x e velocidade y irá ter o espaço de fase como mostrado na Figura 2.9.

Figura 2.9: Espaço de fase de um pêndulo sob o efeito de atrito.



Fonte: Autoria própria.

No geral, tem-se disponível em séries temporais como os sinais de voz a evolução no tempo de apenas uma variável observável, representada por $x(t)$. É possível analisar séries temporais a partir de métodos obtidos da teoria dos sistemas dinâmicos (SANTOS, 2019). O teorema de imersão de Takens (TAKENS, 1981) estabelece que é possível reconstruir um espaço de fase difeomorficamente equivalente ao original a partir do conhecimento de uma única variável do sistema, convertendo uma sequência de medidas escalares $x(n)$ em vetores de estado (TAKENS, 1981; KANTZ; SCHREIBER, 2004).

A reconstrução do espaço de fase é possível pois a história temporal de um sinal contém informações sobre as variáveis de estado não observáveis que podem ser usadas para prever um estado presente (SAVI, 2006).

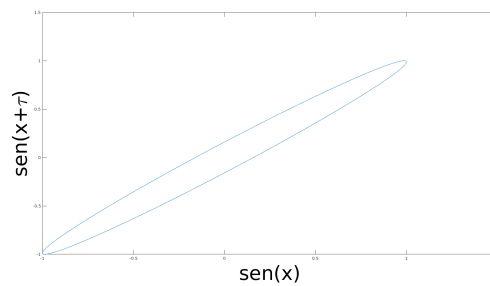
O método mais utilizado na literatura para a reconstrução do espaço de fase é o Método das Coordenadas Defasadas. Esse método permite a reconstrução do espaço de fase m -dimensional similar ao espaço de fase original a partir de uma única variável de estado (PAC-KARD et al., 1980). Portanto, vetores $\vec{\xi}_i$ m -dimensionais são reconstruídos a partir da série temporal $x(t)$ (Equação 2.4).

$$\vec{\xi}_t = \{x(t), x(t + \tau), \dots, x(t + (m - 1)\tau)\}, \quad (2.4)$$

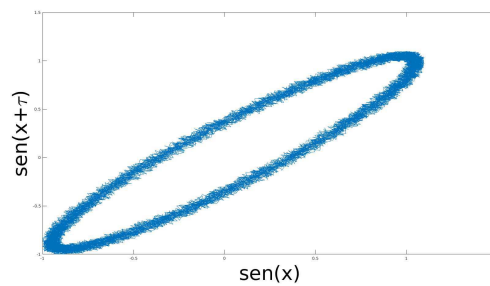
em que $x(t)$ é a série temporal, m representa a dimensão de imersão, τ é o tempo de defasagem ótimo e $\vec{\xi}_t$ representa o estado no tempo t .

A sequência dos estados consiste na trajetória do espaço de fase. Na Figura 2.9, a trajetória do espaço de fase é espiral. O comportamento da trajetória do espaço de fase reconstruído representa as dinâmicas vocais. Quanto mais regular é o espaço de fase, mais o sinal se aproxima de ser um sinal periódico. Por exemplo, na Figura 2.10 são apresentadas duas trajetórias de espaço de fase, uma de um sinal periódico e outra com o mesmo sinal adicionado de ruído aleatório.

Figura 2.10: Trajetórias dos espaços de fase da função seno (a) sem ruído e (b) com ruído aleatório.



(a)



(b)

Fonte: Autoria própria.

Para o cálculo do tempo de defasagem ótimo (τ), é necessário obedecer a algumas características para que o τ seja adequado para recuperar a dinâmica do sistema. Caso o tempo de defasagem seja muito curto, não seria possível capturar a dinâmica dos dados, enquanto que, caso seja muito longo, a reconstrução se tornaria completamente independente em um sentido estático (ABARBANEL, 2012).

Um dos modos de calcular o τ é o método da Informação Mútua Média, que é baseado na Teoria da Informação (FRASER; SWINNEY, 1986). O método consiste em encontrar um intervalo de tempo τ que minimize a informação mútua entre vetores vizinhos, para garantir a reconstrução dos vetores com o menor nível de informação redundante possível, mas ainda assim correlacionados.

O cálculo da informação mútua consiste em verificar o quanto de informação existe em uma medida no tempo t quando se observa outra medida no mesmo sinal no tempo $t + \tau$. A informação mútua é uma generalização da função de autocorrelação para sistemas não lineares (SAVI, 2006).

Dado um histograma de b intervalos de classe, denominados *bins*, criado para estimar a distribuição de probabilidade dos dados de um sinal $x(t)$, o cálculo da informação mútua média entre $x(t)$ e $x(t + \tau)$ está descrito na Equação 2.5.

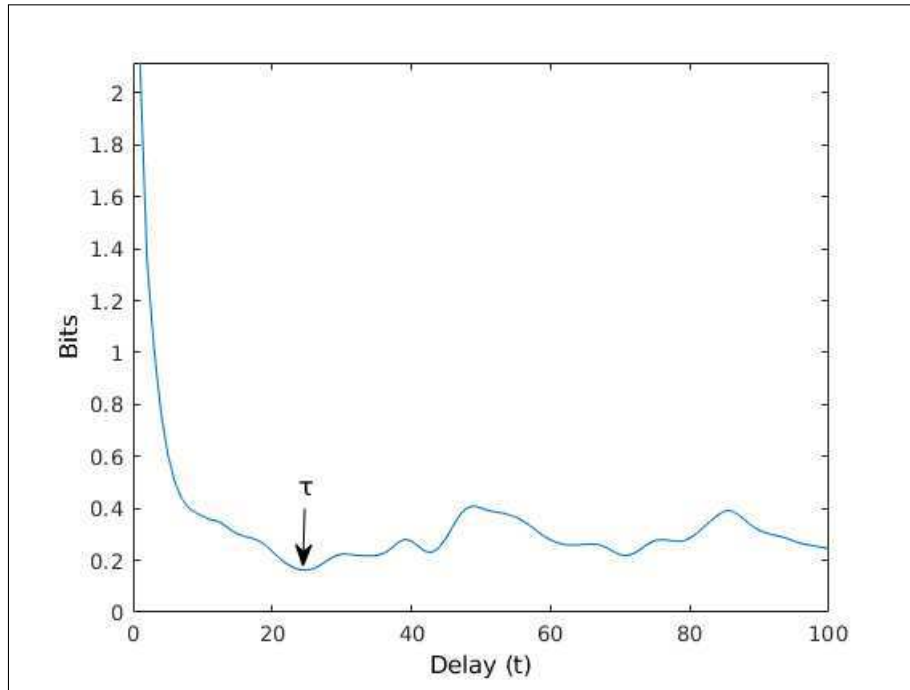
$$I(\tau) = \sum_{i=1}^b \sum_{j=1}^b P_{i,j}(x(t), x(t + \tau)) \log_2 \left[\frac{P_{i,j}(x(t), x(t + \tau))}{P_i(x(t))P_j(x(t + \tau))} \right], \quad (2.5)$$

em que $P_i(x(t))$ é a probabilidade de que o sinal $x(t)$ assuma um valor no i -ésimo intervalo do histograma, $P_j(x(t + \tau))$ é a probabilidade de que $x(t + \tau)$ esteja no j -ésimo intervalo e $P_{i,j}(x(t), x(t + \tau))$ é a probabilidade de que $x(t)$ esteja no i -ésimo intervalo e de que $x(t + \tau)$ esteja no j -ésimo intervalo, simultaneamente.

O tempo de defasagem ótimo τ consiste no primeiro mínimo da curva de informação mútua média obtida a partir da Equação 2.5. Na Figura 2.11 é apresentado um exemplo de uma curva de informação mútua média obtida a partir de um sinal de voz afetada por Edema.

O método mais tradicional para o cálculo da dimensão de imersão m é o método dos Falsos Vizinhos Próximos (KENNEL; BROWN; ABARBANEL, 1992). O método consiste em contar quantos estados do espaço de fase são vizinhos em uma dimensão $m - 1$ e não são mais vizinhos na dimensão m , os chamados Falsos Vizinhos. O valor de m é incrementado a partir de 1 até que a quantidade de falsos vizinhos seja igual a zero. A dimensão de imersão será o primeiro m em que não existam mais falsos vizinhos. Para esta pesquisa, não foi utilizado nenhum método para o cálculo da dimensão de imersão. A dimensão de imersão foi fixada em $m = 2$ para todos os sinais (PINHO et al., 2016), assim todas as imagens obtidas com a trajetória de espaço de fase são 2D.

Figura 2.11: Curva de Informação Mútua do sinal CAC10AN (Edema) da base de dados MEEL.



Fonte: Autoria própria.

2.5 Classificação

Após a etapa de extração de características dos sinais de voz, os vetores de características formam um conjunto de treinamento, que servirá para realizar o aprendizado de um classificador, conseqüentemente, diferenciação entre as diversas classes. No contexto desta pesquisa, objetivou-se diferenciar, mediante o classificador, diferentes patologias das pregas vocais. Nas subseções a seguir, são abordados os dois classificadores utilizados na pesquisa, Redes Neurais Convolucionais e Máquinas de Vetores de Suporte.

2.5.1 Redes Neurais Convolucionais

Uma Rede Neural Convolutiva (*Convolutional Neural Network* - CNN) é uma rede que contém dois tipos especiais de camadas: camada convolutiva e camada *pooling*. Uma arquitetura típica de uma CNN consiste em uma camada de entrada, um ou mais pares de camadas convolucionais e *poolings*, que são conectadas alternadamente, e uma ou mais camadas densas que são conectadas no fim da rede. A CNN combina três ideias para garantir

um nível de invariância ao deslocamento e à distorção: campos receptivos locais, pesos compartilhados e subamostragem espacial (GOODFELLOW; BENGIO; COURVILLE, 2016).

Cada neurônio da camada convolucional tem como entrada uma pequena seção retangular da camada anterior e aplica a essa entrada um filtro. Esses filtros são replicados ao longo do espaço de entrada para detectar um tipo específico de padrão local. Os parâmetros da camada são a dimensão da seção retangular (tamanho do *kernel*), número de filtros, *stride* e *padding*. *Stride* indica o tamanho do deslocamento do filtro e *padding* consiste em aumentar a imagem adicionando pixels com valor igual a zero na borda da imagem (GOODFELLOW; BENGIO; COURVILLE, 2016). A partir de uma entrada de tamanho $w \times h \times d$, em que w é a largura, h a altura e d a quantidade de canais, a camada convolucional gera uma saída de dimensão $w_1 \times h_1 \times d_1$, cujos termos podem ser calculados pelas Equações 2.6 à 2.8.

$$w_1 = \frac{w - K + 2P}{S} + 1, \quad (2.6)$$

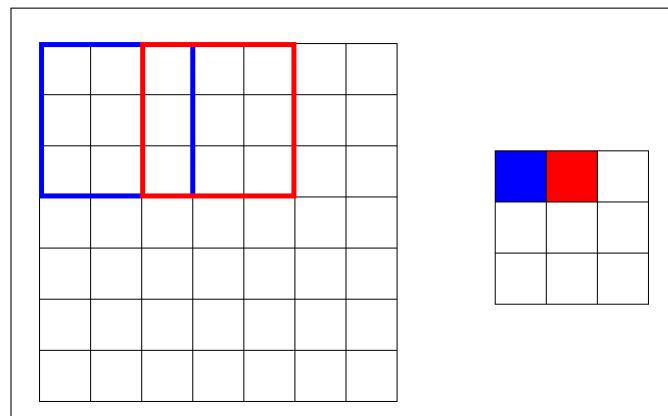
$$h_1 = \frac{h - K + 2P}{S} + 1, \quad (2.7)$$

$$d_1 = N, \quad (2.8)$$

em que K é o tamanho do *kernel*, P o *padding*, S o *stride* e N o número de filtros.

Um exemplo do funcionamento de uma camada convolucional é ilustrado na Figura 2.12.

Figura 2.12: Funcionamento da camada convolucional com tamanho do *kernel* = 3 e *stride* = 2.



Fonte: Autoria própria.

A camada convolucional consiste em um conjunto de operações lineares. É comum aplicar uma função de ativação não linear na camada para introduzir não linearidade ao sistema. No passado, funções como tangente ou sigmoide foram habitualmente utilizadas, mas recentemente foi verificado que a função de ativação *ReLU* (*Rectified Linear Unit* - Unidade Linear Retificada) funciona melhor como função de ativação, pois permite um treinamento mais rápido da rede sem diferença significativa de acurácia (NAIR; HINTON, 2010). A função ReLU está descrita na Equação 2.9.

$$f(x) = \max(0, x), \quad (2.9)$$

em que x é a saída de um neurônio. Como a função zera saídas negativas, pode-se dizer que apenas alguns neurônios são ativados, isso faz com que a rede seja mais esparsa e com isso o custo computacional diminui.

O objetivo da camada *pooling* é obter invariância espacial por meio da redução da resolução dos mapas de características. Essa camada tem como entrada a saída de uma camada convolucional e proporciona uma versão de menor dimensão por meio de um filtro de ativação de uma pequena região de tamanho $n \times n$. Os parâmetros da camada *pooling* são o tamanho da região, o *stride* e o *padding* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A camada irá produzir, a partir de uma entrada de tamanho $w_1 \times h_1 \times d_1$, uma saída $w_2 \times h_2 \times d_1$, dimensões essas que podem ser calculadas pelas Equações 2.10 a 2.12.

$$w_2 = \frac{w_1 - K + 2P}{S} + 1, \quad (2.10)$$

$$h_2 = \frac{h_1 - K + 2P}{S} + 1, \quad (2.11)$$

$$d_2 = d_1, \quad (2.12)$$

em que K é o tamanho do *kernel*, P o *padding* e S o *stride*. O método mais comum de funcionamento de uma camada *pooling* consiste em obter o valor máximo da região, conforme ilustrado na Figura 2.13. Dentre outros métodos, tem-se a obtenção do valor médio da região, por exemplo. Nesta pesquisa, foi utilizado o método do valor máximo.

Figura 2.13: (a) Entrada e (b) saída de uma camada *maxpooling* com tamanho do *kernel* = 2 e *stride* = 2.

| | | | | | |
|---|---|---|---|---|---|
| 9 | 8 | 2 | 3 | | |
| 6 | 9 | 4 | 2 | | |
| 2 | 5 | 1 | 5 | 9 | 4 |
| 0 | 3 | 6 | 7 | 5 | 7 |
| | | | | | |
| | | | | | |

(a) (b)
Fonte: Autoria própria.

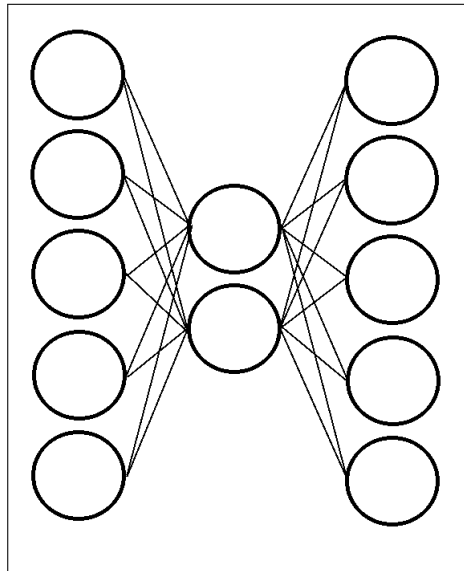
Após o conjunto de camadas convolucionais e *pooling*, uma ou mais camadas densas são adicionadas para obter a classificação das entradas a partir das saídas das camadas anteriores (GOODFELLOW; BENGIO; COURVILLE, 2016). Essas camadas geralmente são formadas por neurônios do tipo perceptron, formulados por Rosenblatt (1958). Atualmente, a função de ativação mais utilizada para a camada de saída, função que é utilizada nesta pesquisa, é a função *softmax*. A função *softmax* é um tipo de função sigmoide que transforma as saídas para cada classe em valores entre 0 e 1 e as divide pela soma das saídas, fazendo com que as saídas da rede sejam tratadas como probabilidades de a entrada pertencer à determinada classe. A Equação 2.13 descreve a função *softmax* para a saída do neurônio i .

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}, \text{ para } i = 1, \dots, J, \quad (2.13)$$

em que J é a quantidade de saídas.

Um outro modo de utilização da CNN além da classificação propriamente dita, é a obtenção de características a partir de uma camada *bottleneck*. Uma camada *bottleneck* consiste em uma camada densa com um tamanho inferior ao das outras camadas densas, como se fosse um “gargalo” (Figura 2.14). As características consistem nas ativações dessa pequena camada (GREZL; FOUSEK, 2008).

A rede com a camada *bottleneck* é treinada normalmente e, após o treinamento, todas as entradas de treinamento e teste são avaliadas pela rede, e são obtidos os novos vetores de características a partir da camada *bottleneck*. Esses vetores de características são utilizados para treinar e testar um outro classificador. Nesta pesquisa, foram utilizadas Máquinas de Vetores de Suporte como classificador, as quais serão descritas na seção a seguir.

Figura 2.14: Camada *bottleneck* ao centro da figura.

Fonte: Autoria própria.

2.5.2 Máquinas de Vetores de Suporte

As Máquinas de Vetores de Suporte (*Support Vector Machine* - SVM) são classificadores reconhecidos pela alta capacidade de predição em várias aplicações de processamento digital de sinais (VAPNIK, 1998). O treinamento de uma SVM consiste em construir um hiperplano ótimo dividindo os vetores de característica de duas classes.

Dado um conjunto de treinamento de N elementos $\{x_i, y_i\}_{i=1}^N$, em que x_i é o i -ésimo padrão de entrada e y_i é o i -ésimo padrão de saída, a SVM realiza o mapeamento para um espaço de características de alta dimensão utilizando uma função núcleo (*kernel*). O objetivo é fazer com que os vetores de características sejam linearmente separáveis. O hiperplano é definido pela Equação 2.14.

$$f(x) = \vec{w}\vec{x} + b, \quad (2.14)$$

em que \vec{w} e b são parâmetros obtidos ao se resolver o problema de otimização da Equação 2.15.

$$\begin{aligned} &\text{Minimizar } \frac{1}{2}\vec{w}^T\vec{w} + C\sum_{i=1}^N \varepsilon_i \\ &\text{Sujeito a } y_i(\vec{w}^T\theta(x_i) + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, \end{aligned} \quad (2.15)$$

em que $\theta(x_i)$ é a função *kernel* e $C > 0$ é o parâmetro de penalidade do erro ε_i .

Existem várias funções *kernel*, como por exemplo as funções linear, polinomial, base radial (*Radial Basis Function* - RBF), sigmoide, dentre outras. Nesta pesquisa, foi utilizada a função RBF (MUSAVI et al., 1992).

2.6 Considerações Finais

Neste capítulo, foi apresentado um estudo sobre a fisiologia das pregas vocais e sobre as patologias das pregas vocais abordadas nesta pesquisa: Edema, Paralisia, Queratose, Nódulo, Pólipo e Cisto. Além disto, foi apresentado um estudo sobre os métodos para extração de características e classificadores utilizados na pesquisa: Reconstrução do Espaço de Fase, Redes Neurais Convolucionais e Máquinas de Vetores de Suporte.

No próximo capítulo, será apresentada uma revisão bibliográfica sistemática sobre processamento digital de sinais de vozes afetadas por patologias nas pregas vocais, com ênfase nas pesquisas que buscaram identificar uma patologia específica.

Capítulo 3

Revisão Sistemática da Literatura

No presente capítulo, objetiva-se apresentar uma revisão bibliográfica sistemática sobre processamento digital de sinais de vozes afetadas por patologias nas pregas vocais, com ênfase nas pesquisas que buscaram identificar uma patologia específica.

A Revisão Sistemática foi realizada de acordo com a metodologia de Keele (2007). Tal metodologia é baseada em uma estratégia de busca que objetiva encontrar o maior número de pesquisas relevantes a um dado contexto. A descrição da metodologia utilizada encontra-se no Apêndice A deste documento.

O presente capítulo se encontra estruturado como se segue. Os resultados da pesquisa estão descritos na Seção 3.1. As Seções 3.2 a 3.6 contêm as descrições das pesquisas encontradas agrupadas por patologias estudadas, bases de dados, técnicas para extração e seleção de características e estratégia de classificação, respectivamente.

3.1 Resultados da Pesquisa Sistemática

Durante a elaboração do protocolo de pesquisa, foram definidos os critérios de exclusão de artigos a seguir.

1. O artigo não possui relevância ao tema abordado.
2. O artigo não faz diferenciação entre patologias.
3. A patologia estudada não ocorre nas pregas vocais (Ex: Mal de Parkinson).

Após a definição do protocolo de pesquisa (Apêndice A), a busca foi realizada nas fontes de pesquisa selecionadas, com a seleção de 543 pesquisas. Em seguida, foram verificados os títulos e resumos das pesquisas e foram eliminadas 407 pesquisas que atenderam aos critérios de exclusão 1 e 3. Por fim, das 136 pesquisas restantes, 99 foram eliminadas no último passo de seleção dos artigos, pelo critério de exclusão 2. O resumo da seleção dos 37 artigos restantes em relação à fonte de pesquisa encontra-se na Tabela 3.1.

Tabela 3.1: Resultados da busca e seleção de pesquisas, com as quantidades dos artigos mantidos.

| Fonte de Pesquisa | Pesquisas encontradas | Pesquisas Selecionadas - Resumo | Pesquisas Selecionadas - Final |
|-------------------|-----------------------|---------------------------------|--------------------------------|
| Total | 543 | 136 | 37 |
| IEEE | 127 | 70 | 19 |
| ACM | 6 | 0 | 0 |
| Science Direct | 133 | 52 | 18 |
| Springer | 277 | 14 | 0 |

Fonte: Autoria própria.

Pode-se observar que as 37 pesquisas selecionadas se encontram nas fontes de pesquisa IEEE e Science Direct. A fonte de pesquisa Springer apresentou pesquisas na área de processamento digital de sinais de vozes afetadas por patologias, com 14 pesquisas encontradas entre 2011 e dezembro de 2018, mas todas as pesquisas foram excluídas pelo critério de exclusão 2, ou seja, elas não buscaram identificar uma patologia específica que afeta o sinal de voz.

Nota-se, também, que a quantidade de pesquisas que abordam a diferenciação de sinais de vozes entre voz normal e voz com patologia (99), sem realizar a identificação da patologia é mais de 2 vezes superior à quantidade de pesquisas que buscam identificar uma patologia específica (37). Essa diferença pode ser explicada pela maior facilidade em realizar apenas uma classificação global do sinal de voz, normal ou com patologia, sem identificar uma determinada patologia.

3.2 Patologias Estudadas

O objetivo desta seção é apresentar quais patologias que afetam as pregas vocais foram estudadas nos últimos anos. Das 37 pesquisas selecionadas, 05 organizaram as diversas patologias em grupos e as outras 32 estudaram patologias isoladamente.

Shi, Dongkai e Ke (2011) dividiram as patologias em dois grupos: patologias orgânicas, que são aquelas que causam alterações anatômicas nas pregas vocais e patologias funcionais, que são aquelas que não causam alterações.

Uloza et al. (2011) e Vaiciukynas et al. (2012) estudaram outro tipo de agrupamento. As patologias das pregas vocais foram divididas em dois grupos, conforme a forma da lesão. As patologias que causam lesões localizadas, tais como Nódulo, Pólipo e Cisto foram agrupadas na classe “Lesão Nodular” e as patologias que causam lesões por toda a prega vocal, tais como papiloma, Queratose ou carcinoma, foram agrupadas na classe “Lesão Difusa”.

Cordeiro et al. (2015) e Cordeiro et al. (2017) agruparam as patologias Edema e Nódulo em um único grupo denominado “doenças fisiológicas” e compararam com outro grupo denominado “doenças neuromusculares”, que contém apenas sinas de voz afetada por Paralisia nas pregas vocais.

Dentre as 32 pesquisas que estudaram patologias isoladamente, as patologias mais estudadas foram Paralisia, Pólipos, Nódulos e Edema, com 19, 19, 17 e 9 pesquisas, respectivamente. Essas patologias possuem mais sinais de vozes nas bases de dados. Além disso, as referidas patologias são as de maior incidência na população (REMACLE et al., 2017) sendo, assim, de mais fácil obtenção.

Na Tabela 3.2, estão as quantidades de pesquisas que estudaram cada patologia. Pode-se observar que a quantidade de pesquisas apresentadas na tabela é superior a 37 (quantidade de pesquisas estudadas). Isto se deve ao fato de que a maioria das pesquisas estudou mais de uma patologia, como, por exemplo, a pesquisa de Al-Nasheri et al. (2017), que teve como objetivo diferenciar sinais de vozes afetadas por Cisto, Pólipo e Paralisia.

Tabela 3.2: Patologias das pregas vocais estudadas.

| Patologia | Quantidade de Pesquisas |
|----------------------------|-------------------------|
| Adutor | 6 |
| Cicatriz nas pregas vocais | 1 |
| Cisto | 7 |
| Corpectomia | 1 |
| Disfonia Espasmódica | 4 |
| Disfonia Funcional | 1 |
| Edema de Reinke | 9 |
| Falsa Prega Vocal | 1 |
| Hiperfunção | 2 |
| Laringite | 1 |
| Nódulo | 17 |
| Paquidermia | 1 |
| Paralisia | 19 |
| Paresia | 1 |
| Pólipo | 19 |
| Queratose | 9 |
| Sulco Vocal | 2 |
| Tumor nas pregas vocais | 1 |

Fonte: Autoria própria.

3.3 Bases de Dados

O objetivo desta seção é apresentar quais são as bases de dados de vozes afetadas por patologias utilizadas na literatura. Para responder esta questão de pesquisa, foram analisadas todas as 136 pesquisas que estudaram a análise do sinal de voz afetada por patologia, incluindo aquelas que não diferenciaram as patologias e foram excluídas pelo critério 2.

As pesquisas foram agrupadas conforme as bases de dados utilizadas. As pesquisas que utilizaram bases de dados não disponíveis ao público foram colocadas no grupo “bases particulares”. Essas bases de dados geralmente foram obtidas pelos autores em algum hospital local ou, então, por intermédio de outros autores que publicaram pesquisas que utilizaram es-

sas bases anteriormente. Em seguida, foram consideradas as pesquisas que utilizaram bases de dados liberadas ao público, seja gratuitamente, seja comercialmente. As pesquisas foram organizadas em diferentes grupos, sendo um grupo para cada base de dados. Na Tabela 3.3, é apresentada a quantidade de pesquisas de cada grupo.

Tabela 3.3: Quantidade de pesquisas agrupadas por base de dados.

| Bases de Dados | Pesquisas não excluídas pelo critério 2 | Pesquisas excluídas pelo critério 2 |
|--------------------|---|-------------------------------------|
| Bases particulares | 18 | 49 |
| Base MEEI | 22 | 46 |
| Base SVD | 7 | 18 |

Fonte: Autoria própria.

Observa-se que a base de dados comercial *Disordered Voice Database, Model 4337*, desenvolvida pela *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (KAY-ELEMENTRICS, 1994), denominada de “base MEEI”, é bastante comum na literatura, tendo sido utilizada em 68 das 136 pesquisas realizadas entre 2011 e 2018, representando 50% das pesquisas.

Uma alternativa à base MEEI, que é uma base comercial, é a criação de uma base de dados própria. Trata-se de uma atividade custosa, que pode demorar anos, mas que está sendo bastante utilizada. Essas bases de dados geralmente são fechadas e dificilmente são disponibilizadas para outros pesquisadores, tornando assim difícil a reprodução da pesquisa. Alguns pesquisadores conseguiram contato com autores que utilizaram bases próprias e tiveram acesso a essas bases. Nesses casos, as pesquisas foram agrupadas em “bases particulares”, compondo 56 das 136 pesquisas realizadas.

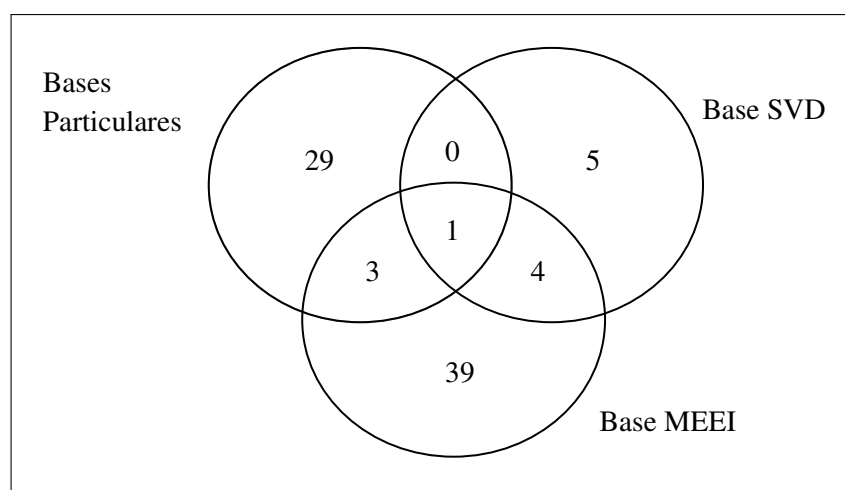
Nos últimos anos, o Instituto de Fonética da Universidade de Saarland construiu uma base de dados de voz com distúrbio, a base de dados *Saarbrücker (Saarbrücker Voice Database - SVD)* (BARRY; PÜTZER, 2014). Devido à base de dados ser recente, uma quantidade menor de pesquisas a utilizou. Conforme observado na Tabela 3.3, apenas 25 pesquisas de 136 utilizaram a base SVD, e apenas 07 atenderam ao critério 2, ou seja, a maioria realizou apenas a classificação normal *versus* patologia, sem especificar a patologia.

Por fim, pode-se observar que o total de pesquisas apresentado na tabela extrapola a quantidade total de pesquisas realizadas. Isto ocorre porque algumas pesquisas utilizam mais de uma base de dados como, por exemplo, a pesquisa de Muhammad et al. (2016), que utiliza as bases SVD e MEEI para fins de comparação. Na Figura 3.1, é apresentada a intersecção

entre grupos de pesquisa. Dessas pesquisas, duas realizam classificação inter base. Ali et al. (2017) utilizaram as bases de dados MEEI, SVD e uma local para realizar classificação entre voz normal e voz afetada por patologia. Eles treinaram o classificador com sinais de uma base e testaram com sinais de outra classe. O melhor resultado foi obtido usando a base SVD para treinamento e a base MEEI para teste, com acurácia de 81,96%. Muhammad et al. (2017) utilizaram as mesmas bases de dados de Ali et al. (2017) para realizar a classificação entre Cisto, Paralisia e Pólipo. Eles empregaram duas bases para treinamento e a última para teste. O melhor resultado foi de 85% de acurácia para diferenciar entre as três classes utilizando as bases MEEI e local para treinamento e a base SVD para teste.

Algumas pesquisas realizaram comparações entre as bases MEEI e SVD, utilizando o mesmo procedimento com as duas. Em todas, os resultados na base MEEI foram superiores aos da base SVD para diferenciar voz normal de voz afetada por patologia. Al-nasheri et al. (2014) obtiveram 97% de acerto na base MEEI e 72% na base SVD. Ali et al. (2017) obtiveram 94% de acerto na base MEEI e 80% de acerto na base SVD. Muhammad et al. (2017) obtiveram 99% de acerto na base MEEI e 93% na base SVD. Uma justificativa dada pelos autores é a de que o ambiente de gravação de voz normal foi diferente do ambiente de gravação de voz afetada por patologia na base MEEI e, portanto, os modelos de classificação estavam usando a informação do ambiente de gravação para diferenciar entre as duas classes.

Figura 3.1: Quantidade de pesquisas para cada grupo de bases de dados, em função da base de dados utilizada.



Fonte: A autoria própria.

3.4 Extração de Características

Nesta seção, objetiva-se identificar quais são as técnicas utilizadas atualmente para caracterizar um sinal de voz afetada por patologia. O campo de estudos de diferenciação de patologias que afetam as pregas vocais é amplo e as pesquisas não são conclusivas. As pesquisas possuem diferenças significativas em relação às técnicas utilizadas para extração de características.

Nas 37 pesquisas selecionadas, as características mais utilizadas foram Coeficientes Mel-cepstrais, utilizados em 10 pesquisas; Coeficientes da Transformada Wavelet, utilizados em 5 pesquisas; e Coeficientes Cepstrais, utilizados em 4 pesquisas. Ainda assim, cada pesquisa difere da outra, em função do modo para a obtenção das características.

Dentre as pesquisas que utilizaram coeficientes mel-cepstrais, Uloza et al. (2011) adotaram entre 14 e 35 coeficientes. Jothilakshmi (2014) agrupou 13 coeficientes mel-cepstrais com suas primeira e segunda derivadas, criando um vetor de características com 39 coeficientes. Mojaly, Muhammad e Alsulaiman (2014) utilizaram 12 e 39 coeficientes. Salma et al. (2014) utilizaram 13 coeficientes. Cordeiro et al. (2015) variaram a quantidade de coeficientes mel-cepstrais de 8 a 24 e, para cada variação, agruparam com sua derivada, a energia e a energia da derivada. Cordeiro et al. (2017) variaram a quantidade de coeficientes mel-cepstrais de 8 a 24 e, para cada variação, agruparam com sua derivada. Salma, Asma e Aicha (2017) obtiveram os coeficientes mel-cepstrais do produto multiescalar do sinal, não do próprio sinal.

Considerando pesquisas que utilizam Coeficientes Wavelets, Shi, Dongkai e Ke (2011) usaram a Transformada Wavelet Contínua. Carvalho, Cavalcante e Cortez (2011) aplicaram a Transformada Wavelet Discreta a janelas de curtíssima duração (6 ms). Saeedi e Almasganj (2013) e Salehi (2015) utilizaram algoritmos genéticos para a escolha da função base da Transformada Wavelet. Mehdizadehfar, Almasganj e Torabinezhad (2017) utilizaram a entropia de *Shannon* de pacotes Wavelet.

Em relação aos coeficientes cepstrais, Cordeiro, Fonseca e Ribeiro (2013) utilizaram apenas o primeiro pico da envoltória cepstral, enquanto que Marinus et al. (2013) utilizaram 12 coeficientes cepstrais de janelas de 20 ms, Jothilakshmi (2014) comparou o uso de 12 coeficientes cepstrais de janelas de 20 ms com 39 coeficientes mel-cepstrais e suas derivadas

obtidos das mesmas janelas e Ali et al. (2016) utilizaram 11 coeficientes cepstrais de janelas de 10 ms.

3.5 Seleção de Características

Conforme a seção anterior, diversas pesquisas utilizaram mais de uma técnica para extração de características. Em função da quantidade de coeficientes escolhidos ou de agrupamentos de vetores de características, algumas pesquisas utilizaram técnicas para seleção de características, a fim de reduzir a dimensionalidade dos vetores de características obtidos. Das 27 pesquisas encontradas, 04 pesquisas utilizaram alguma técnica nesse sentido.

Uloza et al. (2011) compararam 10 técnicas para a extração de características, com o tamanho do vetor de cada uma delas variando de 11 a 100. Os autores utilizaram algoritmos genéticos para reduzir a dimensionalidade de cada técnica separadamente e também para reduzir o vetor de características gerado pelo agrupamento de todas as técnicas.

Markaki e Stylianou (2011) utilizaram uma técnica denominada “Máxima relevância” para escolher as características mais relevantes do vetor. A relevância foi obtida pela técnica de informação mútua.

Mojaly, Muhammad e Alsulaiman (2014) utilizaram três técnicas para a seleção de características, para fins de comparação. Foram utilizados os testes estatísticos t de Student e de Kruskal-Wallis e também algoritmos genéticos. Os melhores resultados foram obtidos utilizando 10 coeficientes selecionados pelo teste Kruskal-Wallis. Os autores não especificaram como a seleção foi utilizada.

Por fim, Muhammad et al. (2016) utilizaram análise de componentes principais para reduzir a dimensionalidade do vetor de características. Eles utilizaram 15, 25 e 35 componentes principais, sendo o melhor resultado obtido a partir de 35 componentes principais.

3.6 Classificação

Nesta seção, objetiva-se identificar quais as técnicas utilizadas atualmente para classificação de padrões de sinais de vozes afetadas por patologias. Apesar da grande quantidade de classificadores descritos na literatura, foi verificado que a maioria das pesquisas encontradas

na revisão sistemática utiliza o classificador SVM (*Support Vector Machine* - Máquina de Vetores de Suporte).

Das 37 pesquisas encontradas, duas utilizaram métodos estatísticos para verificar se existem diferenças significativas entre as patologias estudadas. Kumar, Bhat e Mukhi (2011) utilizaram ANOVA para realizar um estudo sobre sinais de vozes afetadas por Nódulo. Choi et al. (2012) utilizaram vários testes estatísticos, como Kruskal-Wallis e Pearson, para mostrar diferenças entre sinais de vozes afetadas por sulco vocal e por cicatriz nas pregas vocais.

Dentre as pesquisas restantes, Árvores de Decisão, Quantização Vetorial e CNN (*Convolutional Neural Network* - Redes Neurais Convolucionais) foram utilizadas, uma técnica para cada pesquisa. Em 03 pesquisas foram utilizados HMM (*Hidden Markov Models* - Modelos de Markov Escondidos). Em cinco pesquisas foram utilizadas Redes Neurais, porém uma delas não indicou a topologia utilizada (MUHAMMAD et al., 2011). Em quatro pesquisas foi utilizada Análise Discriminante Linear, uma delas utilizou também Análise Discriminante Quadrática para comparação (COSTA et al., 2012). Em seis pesquisas foram utilizados GMM (*Gaussian Mixture Models* - Modelos de Misturas de Gaussianas) e 21 pesquisas utilizaram SVM, 56,8% do total, mostrando uma tendência para utilização desse classificador.

Uma das contribuições desta pesquisa é a investigação do uso de CNN para classificação de patologias da voz, já que, apesar de CNN ser um classificador amplamente utilizado em reconhecimento de padrões, o uso ainda é incipiente na área de identificação de patologias da voz.

Por fim, está contido no Quadro 3.1 um resumo das pesquisas obtidas a partir da revisão sistemática, contendo na última linha um resumo da presente pesquisa.

Quadro 3.1: Resumo das Pesquisas.

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|--------------------------|---|------------------------|--|-----------------|--|
| Muhammad et al. (2018) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Cordectomia <i>versus</i> disfonia hiper-funcional <i>versus</i> paquidermia <i>versus</i> laringite <i>versus</i> Paralisia <i>versus</i> Pólipo | SVD | Espectrograma de oitava. | CNN. | <ol style="list-style-type: none"> 1. 98,40% 2. 96,30% |
| Al-Nasheri et al. (2018) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Cisto <i>versus</i> outras patologias 3. Paralisia <i>versus</i> outras patologias 4. Pólipo <i>versus</i> outras patologias | MEEI, SVD e Particular | Pico máximo, valor de atraso do pico e entropia. | SVM. | <ol style="list-style-type: none"> 1. 99,96% (MEEI), 92,79% (SVD), 92,79% (Par) 2. 99,54% (MEEI), 99,53% (SVD), 96,02% (Par) 3. 98,07% (MEEI), 94,71% (SVD), 91,37% (Par) 4. 98,45% (MEEI), 95,44% (SVD), 92,65% (Par) |

Continua na página seguinte

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|-----------------------------------|--|---------|---|-----------------------------------|---|
| Ali et al. (2018) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Pólipo <i>versus</i> Adutor 3. Pólipo <i>versus</i> Queratose 4. Pólipo <i>versus</i> Nódulo 5. Adutor <i>versus</i> Nódulo 6. Adutor <i>versus</i> Queratose 7. Queratose <i>versus</i> Nódulo 8. Paralisia <i>versus</i> outros | MEEI | Coefficientes FCB. | GMM. | <ol style="list-style-type: none"> 1. 99,72% 2. 97,54% 3. 99,08% 4. 96,75% 5. 98,65% 6. 95,83% 7. 97,10% 8. 99,13% |
| Lopes et al. (2017) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Normal <i>versus</i> Pólipo 3. Normal <i>versus</i> Cisto 4. Normal <i>versus</i> Paralisia 5. Normal <i>versus</i> sulco vocal 6. Normal <i>versus</i> VDDGER 7. Nódulo <i>versus</i> Paralisia 8. Nódulo <i>versus</i> sulco vocal 9. Cisto <i>versus</i> Paralisia | Própria | Frequência fundamental, jitter, shimmer e GNE. | Análise Discriminante Quadrática. | <ol style="list-style-type: none"> 1. 71,76% 2. 75,14% 3. 73,22% 4. 79,64% 5. 78,39% 6. 76,00% 7. 76,79% 8. 72,50% 9. 72,50% |
| Dahmani e Guerti (2017) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> Paralisia <i>versus</i> disfonia espasmódica | SVD | Coefficientes Mel-cepstrais, Pitch, Jitter e Shimmer. | Rede Bayesiana. | <ol style="list-style-type: none"> 1. 90,00% |
| Sodré, Rosa e Dassie-Leite (2017) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. 23 patologias | MEEI | Descritor MDVP. | Rede MLP. | <ol style="list-style-type: none"> 1. 99,40% 2. 50,40% |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|-----------------------------|--|------------------------|---|-----------------|--|
| Al-nasheri et al. (2017) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Cisto <i>versus</i> outras patologias 3. Paralisia <i>versus</i> outras patologias 4. Pólipo <i>versus</i> outras patologias | MEEI, SVD e Particular | Descritor MDVP. | SVM. | <ol style="list-style-type: none"> 1. 89,71% (MEEI), 99,68% (SVD), 72,53% (Par) 2. 88,89% (MEEI), 97,50% (SVD), 82,86% (Par) |
| | | | | | <ol style="list-style-type: none"> 3. 65,56% (MEEI), 79,17% (SVD), 57,14% (Par) 4. 30,00% (MEEI), 82,08% (SVD), 60,00% (Par) |
| Muhammad et al. (2017) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Cisto <i>versus</i> Paralisia <i>versus</i> Pólipo | MEEI, SVD e Particular | Padrão Derivativo Entrelaçado (<i>Interlaced Derivative Pattern</i>) <i>versus</i> Coeficientes Mel-cepstrais <i>versus</i> descritor MDVP. | SVM | <ol style="list-style-type: none"> 1. 99,04% (MEEI), 93,20% (SVD), 91,50% (Par) 2. 94,84% (MEEI), 95,84% (SVD), 95,36% (Par) |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|--|---|------------|---|---|---|
| Mehdizadehfar, Almasganj e Torabinezhad (2017) | <ol style="list-style-type: none"> 1. Cisto <i>versus</i> disfonia funcional 2. Cisto <i>versus</i> falsa prega vocal 3. Cisto <i>versus</i> massa fibrosa 4. Cisto <i>versus</i> Paralisia 5. Disfonia funcional <i>versus</i> massa fibrosa 6. Disfonia funcional <i>versus</i> Paralisia 7. Disfonia funcional <i>versus</i> disfonia espasmódica 8. Disfonia funcional <i>versus</i> refluxo 9. Falsa prega vocal <i>versus</i> disfonia espasmódica 10. Falsa prega vocal <i>versus</i> refluxo 11. Paralisia <i>versus</i> disfonia espasmódica 12. Paralisia <i>versus</i> refluxo | Particular | Entropia do Pacote Wavelet. | HMM. | <ol style="list-style-type: none"> 1. 53,33% 2. 70,83% 3. 66,66% 4. 70,00% 5. 63,33% 6. 64,16% 7. 83,33% 8. 66,66% 9. 53,33% 10. 70,83% 11. 83,33% 12. 66,66% |
| Cordeiro et al. (2017) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Normal <i>versus</i> Nódulo e Edema <i>versus</i> Paralisia | MEEI | Coeficientes Mel-cepstrais <i>versus</i> coeficientes LSF | SVM <i>versus</i> GMM <i>versus</i> Análise Discriminante Quadrática. | <ol style="list-style-type: none"> 1. 98,70% 2. 84,40% |

Continua na página seguinte

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|-----------------------------|--|------------------------|---|---|--|
| Forero et al. (2016) | 1. Normal <i>versus</i> Paralisia <i>versus</i> Nódulo | Particular | Fechamento de fase, abertura de fase, quociente de abertura, quociente de fechamento, quociente de amplitude, quociente de velocidade, diferença entre harmônicos, fator de riqueza dos harmônicos, jitter, shimmer <i>versus</i> Coeficientes Mel-cepstrais. | Rede MLP <i>versus</i> HMM <i>versus</i> SVM. | 1. 97,20% |
| Al-Nasheri et al. (2017) | 1. Normal <i>versus</i> patologia 2. Cisto <i>versus</i> outras patologias 3. Paralisia <i>versus</i> outras patologias 4. Pólipo <i>versus</i> outras patologias | MEEI, SVD e particular | Picos e Defasagens. | SVM. | 1. 99,81% (MEEI), 90,98% (SVD), 93,85% (Par) 2. 98,74% (MEEI), 98,94% (SVD), 95,19% (Par) 3. 97,29% (MEEI), 92,25% (SVD), 90,14% (Par) 4. 97,08% (MEEI), 92,34% (SVD), 90,78% (Par) |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|-----------------------------|--|------------|--|---|--|
| Muhammad et al. (2016) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Cisto <i>versus</i> outras patologias 3. Paralisia <i>versus</i> outras patologias 4. Pólipo <i>versus</i> outras patologias | MEEI e SVD | Medidas estatísticas da área glotal: média, variância, taxa de variância, assimetria e curtose. | SVM. | <ol style="list-style-type: none"> 1. 99,22% (MEEI), 99,22% (SVD) 2. 94,70% (MEEI), 94,70% (SVD) 3. 89,40% (MEEI), 93,90% (SVD) 4. 94,10% (MEEI), 96,90% (SVD) |
| Cordeiro et al. (2015) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia fisiológica (Nódulo e Edema) <i>versus</i> Paralisia | MEEI | Coefficientes Mel-cepstrais e suas derivadas, Coeficientes LSF e DLSF e Coeficientes MLSF e DMLSF. | SVM <i>versus</i> GMM <i>versus</i> Análise Discriminante Quadrática. | <ol style="list-style-type: none"> 1. 77,90% |
| Salehi (2015) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Pólipo <i>versus</i> Queratose 3. Queratose <i>versus</i> Edema 4. Edema <i>versus</i> Paralisia 5. Paralisia <i>versus</i> Nódulo 6. Nódulo <i>versus</i> Paresia | MEEI | Wavelets adaptativas. | SVM. | <ol style="list-style-type: none"> 1. 98,30% 2. 93,50% 3. 83,43% 4. 74,19% 5. 81,53% 6. 65,57% |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|------------------------------------|--|------------|---|---------------------|--|
| Kumar e Daoudi (2015) | <ol style="list-style-type: none"> 1. Nódulo <i>versus</i> outras patologias 2. Adutor <i>versus</i> outras patologias 3. Queratose <i>versus</i> outras patologias | MEEI | Decomposição Matching Pursuit <i>versus</i> Parâmetros obtidos pelo software PRAAT. | Árvores de Decisão. | <ol style="list-style-type: none"> 1. 78,30% 2. 71,42% 3. 67,76% |
| Ali et al. (2016) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Nódulo <i>versus</i> outras patologias 3. Pólipo <i>versus</i> outras patologias 4. Adutor <i>versus</i> outras patologias 5. Queratose <i>versus</i> outras patologias 6. Paralisia <i>versus</i> outras patologias | MEEI | Espectro Auditivo <i>versus</i> Coeficientes Cepstrais. | GMM. | <ol style="list-style-type: none"> 1. 99,41% 2. 85,53% 3. 89,33% 4. 93,30% 5. 85,53% 6. 78,42% |
| Cordeiro, Fonseca e Ribeiro (2014) | <ol style="list-style-type: none"> 1. Edema <i>versus</i> Nódulo | Particular | Frequência do primeiro pico LPC espectral, Largura de Banda do primeiro pico LPC espectral, Pitch e Jitter. | SVM. | <ol style="list-style-type: none"> 1. 84,60% |

Continua na página seguinte

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|--------------------------------------|--|-------|---|------------------------|---|
| Salma et al. (2014) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> hiperfunção 2. Normal <i>versus</i> Paralisia 3. Paralisia <i>versus</i> hiperfunção | MEEI | <p>Coeficientes Mel-cepstrais <i>versus</i> Coeficientes Mel-cepstrais e Frequência fundamental <i>versus</i> Coeficientes Mel-cepstrais, suas primeira e segunda derivadas e Energia <i>versus</i> Coeficientes Mel-cepstrais e Quociente de abertura.</p> | SVM. | <ol style="list-style-type: none"> 1. 99,56% <i>versus</i> 100,00% 2. 100,00% <i>versus</i> 100,00% 3. 97,70% <i>versus</i> 100,00% <p>Obs: Não apresentou resultados de acurácia, apenas taxas de acerto individuais de cada classe</p> |
| Mojaly, Muhammad e Alsulaiman (2014) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Nódulo <i>versus</i> outras patologias 3. Pólipo <i>versus</i> outras patologias 4. Adutor <i>versus</i> outras patologias 5. Queratose <i>versus</i> outras patologias | MEEI | <p>Coeficientes Mel-cepstrais <i>versus</i> Coeficientes LPC <i>versus</i> Coeficientes RASTA-PLP.</p> | SVM <i>versus</i> GMM. | <ol style="list-style-type: none"> 1. 99,90% 2. 95,28% 3. 99,90% 4. 99,90% 5. 99,90% |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|---------------------------|---|------------|--|-----------------------|--|
| Muhammad e Melhem (2014) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Nódulo <i>versus</i> Pólipo 3. Nódulo <i>versus versus</i> Queratose 4. Nódulo <i>versus</i> Adutor 5. Pólipo <i>versus</i> Queratose 6. Pólipo <i>versus</i> Adutor 7. Adutor <i>versus</i> Queratose 8. Paralisia <i>versus</i> outras patologias | MEEI | Características de áudio do MPEG-7. | SVM. | <ol style="list-style-type: none"> 1. 99,41±0,07% 2. 97,62±0,36% 3. 95,75±0,70% 4. 97,47±0,37% 5. 94,91±0,86% 6. 94,55±0,70% 7. 93,19±0,39% 8. 98,34±0,03% |
| Jothilakshmi (2014) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Paralisia cerebral <i>versus</i> disartria <i>versus</i> deficiência auditiva <i>versus</i> laringectomia <i>versus</i> Paralisia nas pregas vocais <i>versus</i> retardo mental <i>versus</i> quadriparesia <i>versus</i> voz balbuciante <i>versus</i> AVC <i>versus</i> tumor no trato vocal | Particular | Coeficientes Mel-cepstrais <i>versus</i> Coeficientes Cepstrais. | HMM <i>versus</i> GMM | <ol style="list-style-type: none"> 1. 94,40% 2. 95,74% |
| Saeedi e Almasganj (2013) | <ol style="list-style-type: none"> 1. Disfonia espasmódica <i>versus</i> Pólipo <i>versus</i> Queratose <i>versus</i> Edema <i>versus</i> Paralisia <i>versus</i> Nódulos | MEEI | Transformada Wavelet Discreta. | SVM. | <ol style="list-style-type: none"> 1. 100,00% |

Continua na página seguinte

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|------------------------------------|---|------------|-------------------------------------|-------------------------------|---|
| Marinus et al. (2013) | 1. Normal <i>versus</i> patologia 2. Edema <i>versus</i> outras patologias | MEEI | Coefficientes Cepstrais | GMM e MLP | 1. 93,10% <i>versus</i> 94,50% 2. 76,20% <i>versus</i> 85,70% Obs: Não apresentou resultados de acurácia, apenas taxas de acerto individuais de cada classe |
| Muhammad e Melhem (2013) | 1. Normal <i>versus</i> patologia 2. Nódulo <i>versus</i> Pólipo 3. Nódulo <i>versus versus</i> Queratose 4. Nódulo <i>versus</i> Adutor 5. Pólipo <i>versus</i> Queratose 6. Pólipo <i>versus</i> Adutor 7. Adutor <i>versus</i> Queratose | MEEI | Características de áudio do MPEG-7. | SVM. | 1. 99,99% 2. 100,00% 3. 99,97% 4. 100,00% 5. 100,00% 6. 100,00% 7. 100,00% |
| Cordeiro, Fonseca e Ribeiro (2013) | 1. Normal <i>versus</i> patologia 2. Edema <i>versus</i> Nódulo | Particular | Coefficientes Cepstrais. | Análise Discriminante Linear. | 1. 100,00% 2. 87,00% |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|-----------------------------|--|------------|--|--|--------------------------------|
| Costa et al. (2012) | 1. Normal <i>versus</i> Edema 2. Normal <i>versus</i> Paralisia | MEEI | Determinismo <i>versus</i> Comprimento máximo das estruturas diagonais <i>versus</i> Entropia de Shanon <i>versus</i> Inclinação da linha de melhor ajuste <i>versus</i> Laminaridade <i>versus</i> Comprimento do seguimento de linha vertical mais longo <i>versus</i> Média do comprimento de linha vertical. | Análise Discriminante Linear <i>versus</i> Análise Discriminante Quadrática. | 1. 88,89% 2. 100,00% |
| Vaiciukynas et al. (2012) | 1. Normal <i>versus</i> patologia 2. Lesão Difusa <i>versus</i> lesão nodular | Particular | Coefficientes Mel-cepstrais. | SVM <i>versus</i> LS-SVM. | 1. 89,00% 2. 69,00% |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|--------------------------------------|--|------------|--------------------------------|--------------------|--|
| Carvalho, Cavalcante e Cortez (2011) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> Nódulo <i>versus</i> Edema 2. normal <i>versus</i> Nódulo <i>versus</i> Edema <i>versus</i> disfonia neurológica | Particular | Transformada Wavelet Discreta. | Redes Neurais MLP. | <ol style="list-style-type: none"> 1. 97,46% <i>versus</i> 90,37% <i>versus</i> 91,19% 2. 93,33% <i>versus</i> 86,56% <i>versus</i> 87,61% <i>versus</i> 89,40% <p>Obs: Não apresentou resultados de acurácia, apenas taxas de acerto individuais de cada classe</p> |
| Markaki e Stylianou (2011) | <ol style="list-style-type: none"> 1. Normal <i>versus</i> patologia 2. Nódulo <i>versus</i> Pólipo 3. Nódulo <i>versus</i> <i>versus</i> Queratose 4. Nódulo <i>versus</i> Adutor 5. Pólipo <i>versus</i> Queratose 6. Pólipo <i>versus</i> Adutor 7. Adutor <i>versus</i> Queratose 8. Paralisia <i>versus</i> Outras Patologias | MEEI | Modulação espectral. | SVM. | <ol style="list-style-type: none"> 1. 94,08 ± 0,28% 2. 91,66 ± 1,14% 3. 89,11 ± 1,33% 4. 92,09 ± 0,92% 5. 84,77 ± 1,42% 6. 91,23 ± 1,10% 7. 95,77 ± 0,92% 8. 82,14 ± 0,85% |

Continua na página seguinte

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|-----------------------------|---|------------|--|-----------------|--|
| Uloza et al. (2011) | <ol style="list-style-type: none"> 1. normal <i>versus</i> patologia 2. normal <i>versus</i> lesão nodular 3. normal <i>versus</i> lesão difusa 4. lesão nodular <i>versus</i> lesão difusa | Particular | Medidas de perturbação de Pitch e Amplitude <i>versus</i> Características de Frequência <i>versus</i> Características de Mel-frequência <i>versus</i> Características de energia cepstral <i>versus</i> Coeficientes Mel-cepstrais <i>versus</i> Características de autocorrelação <i>versus</i> Relação harmônico ruído no domínio da frequência <i>versus</i> Relação harmônico ruído no domínio cepstral <i>versus</i> Coeficientes LPC <i>versus</i> Coeficientes de predição linear da transformada do cosseno. | SVM. | <ol style="list-style-type: none"> 1. 92,60% 2. 84,90% 3. 94,30% 4. 86,90% |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|-----------------------------|---|------------|-------------------------------|---|--|
| Muhammad et al. (2011) | 1. Cisto <i>versus</i> refluxo gastroesofágico <i>versus</i> Paralisia <i>versus</i> Pólipo <i>versus</i> sulco vocal | Particular | Formantes F1 e F2 | Quantização Vetorial <i>versus</i> Redes Neurais. | 1. 67,86% (mulheres) e 52,50% (homens) |
| Shi, Dongkai e Ke (2011) | 1. Normal <i>versus</i> patologia orgânica <i>versus</i> lesões funcionais 2. patologia orgânica <i>versus</i> lesões funcionais | Particular | Transformada Wavelet Contínua | SVM. | 1. 100,00% <i>versus</i> 95,00% <i>versus</i> 97,50% 2. 100,00% <i>versus</i> 97,50% Obs: Não apresentou resultados de acurácia, apenas taxas de acerto individuais de cada classe |
| Continua na página seguinte | | | | | |

Quadro 3.1 – Continuação

| Pesquisa | Objetivos | Bases | Características | Classificadores | Melhores Resultados (Acurácia) |
|----------------------|--|-------------------|--------------------------------------|------------------------------|--|
| Esta pesquisa | <ol style="list-style-type: none"> 1. Normal <i>versus</i> Patologia 2. Paralisia <i>versus</i> Edema 3. Paralisia <i>versus</i> Queratose 4. Paralisia <i>versus</i> Pólipo 5. Paralisia <i>versus</i> Nódulo 6. Paralisia <i>versus</i> Lesão Benigna 7. Edema <i>versus</i> Queratose 8. Edema <i>versus</i> Nódulo 9. Edema <i>versus</i> Pólipo 10. Edema <i>versus</i> Lesão Benigna 11. Queratose <i>versus</i> Pólipo 12. Queratose <i>versus</i> Nódulo 13. Queratose <i>versus</i> Lesão Benigna 14. Nódulo <i>versus</i> Pólipo | MEEI e SVD | Trajectoria do Espaço de Fase | CNN <i>versus</i> SVM | <ol style="list-style-type: none"> 1. 100,00% (MEEI), 93,00% (SVD) 2. 84,44% (MEEI), 78,41% (SVD) 3. 82,55% (MEEI), 76,86% (SVD) 4. 84,21% (MEEI), 74,85% (SVD) 5. 94,48% (MEEI), 69,40% (SVD) 6. 78,26% (MEEI), 67,99% (SVD) 7. 86,25% (MEEI), 76,32% (SVD) 8. 88,82% (MEEI), 77,65% (SVD) 9. 72,73% (MEEI), 78,00% (SVD) 10. 74,29% (MEEI), 73,33% (SVD) 11. 83,75% (MEEI), 68,18% (SVD) 12. 90,59% (MEEI), 90,40% (SVD) 13. 81,38% (MEEI), 63,33% (SVD) 14. 98,18% (MEEI), 79,66% (SVD) |

3.7 Considerações Finais

Neste capítulo, foi apresentada a revisão sistemática compilada na pesquisa, abrangendo os últimos anos (2011 a 2018) de pesquisas realizadas na área de processamento digital de sinais de vozes afetadas por patologia.

Constatou-se uma tendência dos pesquisadores para a análise de sinais de vozes afetadas por Nódulo, Edema de Reinke, Pólipo e Paralisia, que são as patologias nas pregas vocais com maior incidência na população (REMACLE et al., 2017).

Conforme previsto, enfrentou-se dificuldade em relação ao acesso às bases de dados para avaliações experimentais. Apenas duas bases disponíveis ao público foram identificadas: a base MEEI (KAY-ELEMETRICS, 1994) e a base SVD (BARRY; PÜTZER, 2014), sendo a base comercial a mais utilizada (MEEI).

Para representar o sinal de voz, várias técnicas para a extração de características foram pesquisadas, um indicativo de que não foi encontrado ainda um conjunto de características que melhor represente o sinal de voz afetada por determinada patologia.

Em relação aos classificadores, 58,3% das pesquisas utilizaram SVM, mostrando que existe uma tendência para a utilização deste classificador, já que o mesmo é robusto em problemas com bases de dados pequenas, além de apresentar resultados promissores na área de processamento digital de sinais de voz.

No processo de revisão sistemática, foram encontradas 99 pesquisas relevantes, em 37 das quais buscou-se identificar qual a patologia afeta o sinal de voz, que foi o foco central deste capítulo. A revisão apresentada é passível de refinamentos e extensões futuras, bastando para tanto a inclusão de outras fontes de pesquisa, modificações na expressão de busca e no intervalo de consulta.

No próximo capítulo, são apresentados os Materiais e Métodos desta pesquisa de doutorado.

Capítulo 4

Materiais e Métodos

Neste capítulo, são apresentados os materiais e métodos adotados na pesquisa. Na primeira seção, são descritas as bases de dados utilizadas. A abordagem proposta é apresentada na seção seguinte.

4.1 Bases de Dados

As bases de dados utilizadas nesta pesquisa são a base da Massachusetts Eye and Ear Infirmary (base MEEI) (KAY-ELEMENTRICS, 1994) e a *Saarbruecken Voice Database* (base SVD) (BARRY; PÜTZER, 2014). A base MEEI é uma base comercial constituída por, aproximadamente, 700 sinais de voz da vogal sustentada /a/ e 700 sinais de voz de uma passagem do texto *Rainbow Passage*, obtidos de, aproximadamente, 700 sujeitos. Os sinais foram adquiridos com baixo nível de ruído, distância constante do microfone, taxa de amostragem de 25 mil amostras por segundo para sinais de voz afetada por patologia e 50 mil amostras por segundo para sinais de voz normal, com resolução de 16 bits. Ao todo, 53 sinais da vogal sustentada /a/ e 53 sinais da passagem de texto são de voz normal e o restante de voz afetada por diversas patologias, variando de patologias que afetam as pregas vocais, como Nódulos e Cistos, a patologias neurológicas, como Mal de Parkinson e Gagueira.

A base SVD é aberta e disponível para download por intermédio de uma interface Web. A base de dados consiste em gravações de mais de 2.000 sujeitos. Cada sessão de gravação é definida por uma coleção de gravações das vogais sustentadas /a/, /i/ e /u/, com entonações normal, baixa, alta e baixa-alta-baixa, além da sentença “*Guten Morgen, wie geht es Ihnen?*”

(“Bom dia, como vai você?”, em alemão). Ao todo, são 13 arquivos de áudio por sessão. Além disso, foi armazenado também em cada sessão o sinal do exame de eletroglotografia (EGG), o qual poderá ser útil em análises complementares. O tamanho dos sinais de voz de vogal sustentada varia entre 1 a 3 segundos. Todas as gravações foram amostradas a 50 mil amostras por segundo com resolução de 16 bits. Ao todo, são 869 sessões de voz normal, e 1.356 sessões de voz afetada por uma ou mais das 71 patologias catalogadas. As patologias catalogadas são de diversas origens. Algumas são de origem neurológica, como Mal de Parkinson ou Paralisia Bulbar, outras são de origem psicológica, como Afonia e Microfonia Psicogênicas, e também aquelas que afetam diferentes áreas do trato vocal, como Edema de Reinke (pregas vocais), doença de Forestier (faringe) ou Cisto Valecular (língua e epiglote).

A Tabela 4.1 contém a quantidade de sessões de gravação para construção das bases MEEI e SVD, para diferentes distúrbios da voz, baseada na divisão proposta por Verdolini et al. (VERDOLINI; ROSEN; BRANSKI, 2014). Os distúrbios vocais são divididos em 9 categorias, em que cada uma dessas categorias possui subdivisões.

Tabela 4.1: Total de sessões de patologias das pregas vocais de cada base de dados.

| ID | Categorias | MEEI | | | SVD | | |
|------|--|-------|--------|-------|-------|--------|-------|
| | | Homem | Mulher | Total | Homem | Mulher | Total |
| 1000 | DOENÇAS ESTRUTURAS DA LARINGE | | | | | | |
| 1110 | Lesões Laríngeas Malignas | 1 | 0 | 1 | 22 | 1 | 23 |
| 1120 | Lesões Malignas das Pregas Vocais | 1 | 0 | 1 | 21 | 1 | 22 |
| 1130 | Displasia/carcinoma in situ | 0 | 0 | 0 | 1 | 0 | 1 |
| 1200 | Anormalidades do Epitélio e Lâmina Própria das Pregas Vocais | 67 | 86 | 153 | 66 | 113 | 179 |
| 1210 | Lesões Benignas Focais da Lâmina Própria | 20 | 32 | 52 | 30 | 37 | 67 |
| 1211 | Nódulos de Prega Vocal | 1 | 18 | 19 | 4 | 13 | 17 |
| 1212 | Massa fibrosa | 2 | 2 | 4 | 0 | 0 | 0 |
| 1214 | Pólipo(s) de Prega Vocal | 12 | 8 | 20 | 25 | 19 | 44 |
| 1215 | Cisto de prega vocal | 3 | 4 | 7 | 1 | 5 | 6 |
| 1217 | Lesão Reativa de Prega Vocal | 2 | 0 | 2 | 0 | 0 | 0 |
| 1220 | Edema de Reinke | 12 | 32 | 44 | 7 | 61 | 68 |
| 1230 | Cicatriz de Prega Vocal | 11 | 8 | 19 | 0 | 0 | 0 |

Continua na página seguinte

Tabela 4.1 – Continuação

| ID | Categorias | MEEI | | | SVD | | |
|------|---|-------|--------|-------|-------|--------|-------|
| | | Homem | Mulher | Total | Homem | Mulher | Total |
| 1231 | Cicatriz de Prega Vocal Propriamente Dita | 8 | 8 | 16 | 0 | 0 | 0 |
| 1232 | Sulco de Prega Vocal | 3 | 0 | 3 | 0 | 0 | 0 |
| 1240 | Granuloma de Prega Vocal | 5 | 2 | 7 | 1 | 1 | 2 |
| 1250 | Queratose | 15 | 11 | 26 | 27 | 14 | 41 |
| 1260 | Papilomatose Respiratória Recorrente (PRR) | 3 | 1 | 4 | 1 | 0 | 1 |
| 1270 | Estenose Subglótica | 1 | 0 | 1 | 0 | 0 | 0 |
| 1300 | Anormalidades Vasculares das Pregas Vocais | 1 | 1 | 2 | 0 | 0 | 0 |
| 1310 | Hemorragia de Prega Vocal | 1 | 1 | 2 | 0 | 0 | 0 |
| 1400 | Alterações Congênicas e de Maturação que Afetam a Voz | 11 | 3 | 14 | 25 | 27 | 52 |
| 1410 | Membranas Congênicas (Sinéquia) | 0 | 0 | 0 | 2 | 0 | 2 |
| 1440 | Puberfonia | 1 | 0 | 1 | 10 | 0 | 10 |
| 1450 | Presbifonia | 10 | 3 | 13 | 13 | 27 | 40 |
| 2000 | INFLAMAÇÕES LARÍNGEAS | | | | | | |
| 2100 | Artrite Cricoaritenóidea e Cricotireóidea | 3 | 0 | 3 | 0 | 0 | 0 |
| 2200 | Laringite Aguda | 0 | 2 | 2 | 83 | 57 | 140 |
| 3000 | TRAUMA OU LESÃO DA LARINGE | | | | | | |
| 3100 | Trauma Laríngeo Interno | 2 | 2 | 4 | 0 | 3 | 3 |
| 3110 | Trauma da Mucosa Laríngea (Químico ou Térmico) | 0 | 1 | 1 | 0 | 0 | 0 |
| 3120 | Lesões da Mucosa Laríngea por Intubação/Extubação | 1 | 0 | 1 | 0 | 3 | 3 |
| 3130 | Deslocamento da Aritenóidea | 1 | 1 | 2 | 0 | 0 | 0 |
| 7000 | DISTÚRBIOS NEUROLÓGICOS QUE AFETAM A VOZ | | | | | | |
| 7100 | Lesão do Sistema Nervoso Periférico | 41 | 44 | 85 | 75 | 141 | 216 |
| 7110 | Lesão do Nervo Laríngeo Superior | 0 | 0 | 0 | 1 | 2 | 3 |
| 7120 | Paralisia nas pregas vocais | 33 | 34 | 67 | 74 | 139 | 213 |

Continua na página seguinte

Tabela 4.1 – Continuação

| ID | Categorias | MEEI | | | SVD | | |
|------|--|-------|--------|-------|-------|--------|-------|
| | | Homem | Mulher | Total | Homem | Mulher | Total |
| 7130 | Paresia Unilateral ou Bilateral do Nervo Laríngeo Recorrente | 8 | 10 | 18 | 0 | 0 | 0 |

Na tabela, foram consideradas apenas as categorias de distúrbios que afetam as pregas vocais. Por exemplo, a categoria 5000 (Distúrbios Aerodigestivos não-laríngeos que afetam a voz) não está presente na tabela. Outras categorias possuem apenas algumas subcategorias de distúrbios que afetam as pregas vocais. Apenas essas estão presentes na tabela. Por exemplo, na categoria 7000 (Distúrbios Neurológicos que afetam a voz), apenas algumas subcategorias da subcategoria 7100 (Lesão do Sistema Nervoso Periférico) foram consideradas, como, por exemplo, a 7120 (Paralisia Unilateral do Nervo Laríngeo Recorrente).

Não estão presentes também as subcategorias sem sinal nas bases de dados.

As patologias estão classificadas de acordo com o Manual de Classificação de Distúrbios da Voz (VERDOLINI; ROSEN; BRANSKI, 2014). O código presente nas tabelas é o mesmo do manual e representa as categorias e subcategorias de distúrbios da fala. As categorias principais são divididas de acordo com a casa do milhar. Por exemplo, DOENÇAS ESTRUTURAIS DA LARINGE (ID 1000) e DISTÚRBIOS NEUROLÓGICOS QUE AFETAM A VOZ (ID 7000). Subcategorias de uma categoria principal são divididas de acordo com a casa das centenas, como por exemplo Lesão do Sistema Nervoso Periférico (ID 7100), que é uma subcategoria de DISTÚRBIOS NEUROLÓGICOS QUE AFETAM A VOZ (ID 7000). As subcategorias de uma subcategoria são divididas de acordo com a casa das dezenas, e, por fim, uma subdivisão na casa das unidades. Por exemplo, Pólipo(s) de Prega Vocal (ID 1214) é uma subcategoria de Lesões Benignas Focais da Lâmina Própria (ID 1210) que, por sua vez, é uma subcategoria de Anormalidades do Epitélio e Lâmina Própria das Pregas Vocais (ID 1200), que, por sua vez, é uma subcategoria de DOENÇAS ESTRUTURAIS DA LARINGE (ID 1000).

A seguir, são apresentadas algumas considerações sobre as patologias das bases de dados.

- As subcategorias 1212 (Massa fibrosa - Subepitelial) e 1213 (Massa fibrosa - Ligamento) foram consideradas como sendo uma única (Massa fibrosa), pois as bases de dados não indicam o tipo de Massa fibrosa.

- As subcategorias 1215 (Cisto de prega vocal - Subepitelial) e 1216 (Cisto de prega vocal - Ligamento) foram consideradas como sendo uma única (Cisto de prega vocal), pois as bases de dados não indicam o tipo de Cisto.
- As bases de dados classificaram algumas sessões como “*Scar*”, então foi suposto que essas sessões são de pacientes com cicatriz nas pregas vocais, e os sinais foram classificados como pertencentes à subcategoria 1231 (Cicatriz de Prega Vocal Propriamente Dita).
- As subcategorias 1241 (Granuloma de Prega Vocal não relacionado à Intubação) e 1242 (Granuloma de Prega Vocal Relacionado a Intubação/úlceras de Contato) foram consideradas como sendo uma única (Granuloma de prega vocal), pois a base de dados MEEI não indica o tipo de Granuloma.
- As bases de dados classificaram algumas sessões como “*Hemorrhage*”. Então, foi suposto que esses sinais são de pacientes com hemorragia nas pregas vocais e os sinais foram classificados como pertencentes à subcategoria 1310 (Hemorragia de Prega Vocal).
- As subcategorias 7120 (Paralisia Unilateral do Nervo Laríngeo Recorrente) e 7140 (Paralisia Bilateral do Nervo Laríngeo Recorrente - Periférica) foram consideradas como sendo uma única (Paralisia nas pregas vocais), pois a base de dados SVD não indica o tipo de Paralisia.

Para esta pesquisa, conforme mencionado, foram escolhidas patologias das pregas vocais: Paralisia, Edema, Pólipo, Nódulo e Queratose. Além disto, os sinais de vozes afetadas por Nódulo, Pólipo e Cisto foram agrupadas na classe Lesões Benignas Focais da Lâmina Própria. Para efeitos de simplicidade, a partir desse ponto, essa classe foi denominada *lesão*.

4.2 Abordagem Proposta

A abordagem proposta para a pesquisa está descrita nesta seção. A metodologia proposta para classificação de patologias nas pregas vocais obedece a uma metodologia típica para classificação de padrões em sinais. Inicialmente, o sinal passa pelo pré-processamento. Em

seguida, são extraídas características relevantes do sinal e um classificador é treinado para reconhecer os diferentes padrões. Por fim, foi realizada uma análise estatística para verificação das hipóteses da pesquisa. As subseções a seguir descrevem as etapas da metodologia.

4.2.1 Pré-processamento e Extração de Características

Na etapa de pré-processamento, todos os sinais com taxa de amostragem igual a 50 mil amostras por segundo (sinais da base SVD e sinais de voz normal da base MEEI) foram subamostradas para 25 mil amostras por segundo para padronizar com os sinais de voz afetada por patologia da base MEEI, que possuem esta taxa de amostragem.

Após a subamostragem, foi realizado um aumento do conjunto de dados usando o método *Time Stretching* (TS). Para tanto, foi utilizado $\alpha = 0,8; 0,9; 1; 1,1$ e $1,2$, ou seja, o conjunto de dados foi aumentado em 05 vezes. Um exemplo da aplicação do TS em um sinal de voz normal pode ser visto na Figura 4.1.

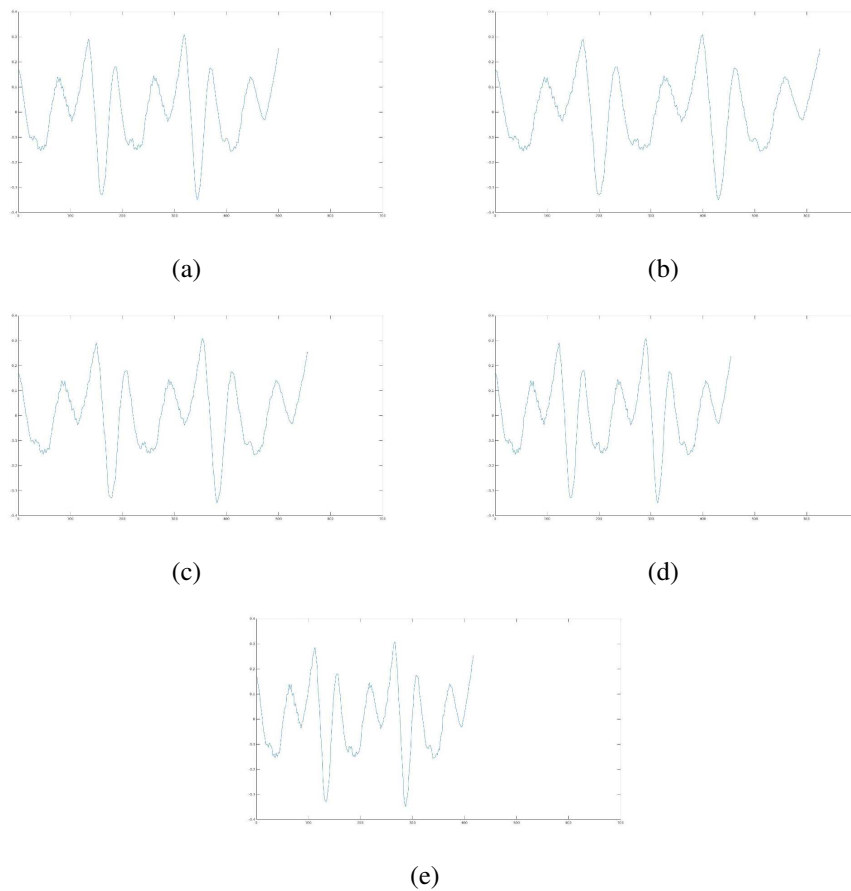
Após o aumento dos dados, foi calculado o τ (mínimo da curva de informação mútua média) de cada sinal e, então, os sinais foram divididos em segmentos de 10 ciclos de *pitch* de duração com sobreposição de 50%. Para efeito de comparação, foi realizado também o janelamento de 20 ms com sobreposição de 50%, tradicionalmente utilizado na literatura. A quantidade de janelas obtidas desconsiderando o uso do TS está na Tabela 4.2.

Tabela 4.2: Total de imagens por classe e base de dados.

| Classe | MEEI | | SVD | |
|-----------|-----------|-------|-----------|-------|
| | 10 Ciclos | 20 ms | 10 Ciclos | 20 ms |
| Normal | 6219 | 15608 | 29566 | 81879 |
| Patologia | 5250 | 16691 | 11583 | 34880 |
| Paralisia | 2249 | 6519 | 6263 | 17939 |
| Edema | 1396 | 4330 | 1888 | 5801 |
| Queratose | 593 | 2462 | 1111 | 4068 |
| Pólipo | 489 | 1980 | 1320 | 4353 |
| Nódulo | 705 | 1894 | 918 | 2302 |
| Lesão | 1289 | 4330 | 2303 | 6845 |

Depois da etapa de janelamento, foi realizada a reconstrução do espaço de fases de cada janela utilizando o τ calculado do sinal e $m = 2$. Após, foi gerado, para cada janela, um gráfico 2D com a trajetória do espaço de fases. Na Figura 4.2 estão contidos exemplos de gráficos das diferentes classes estudadas.

Figura 4.1: Time Stretching das primeiras 500 amostras do sinal (a) 1-a.n.wav da base SVD, com (b) $\alpha = 0,8$, (c) $\alpha = 0,9$, (d) $\alpha = 1,1$ e (e) $\alpha = 1,2$.



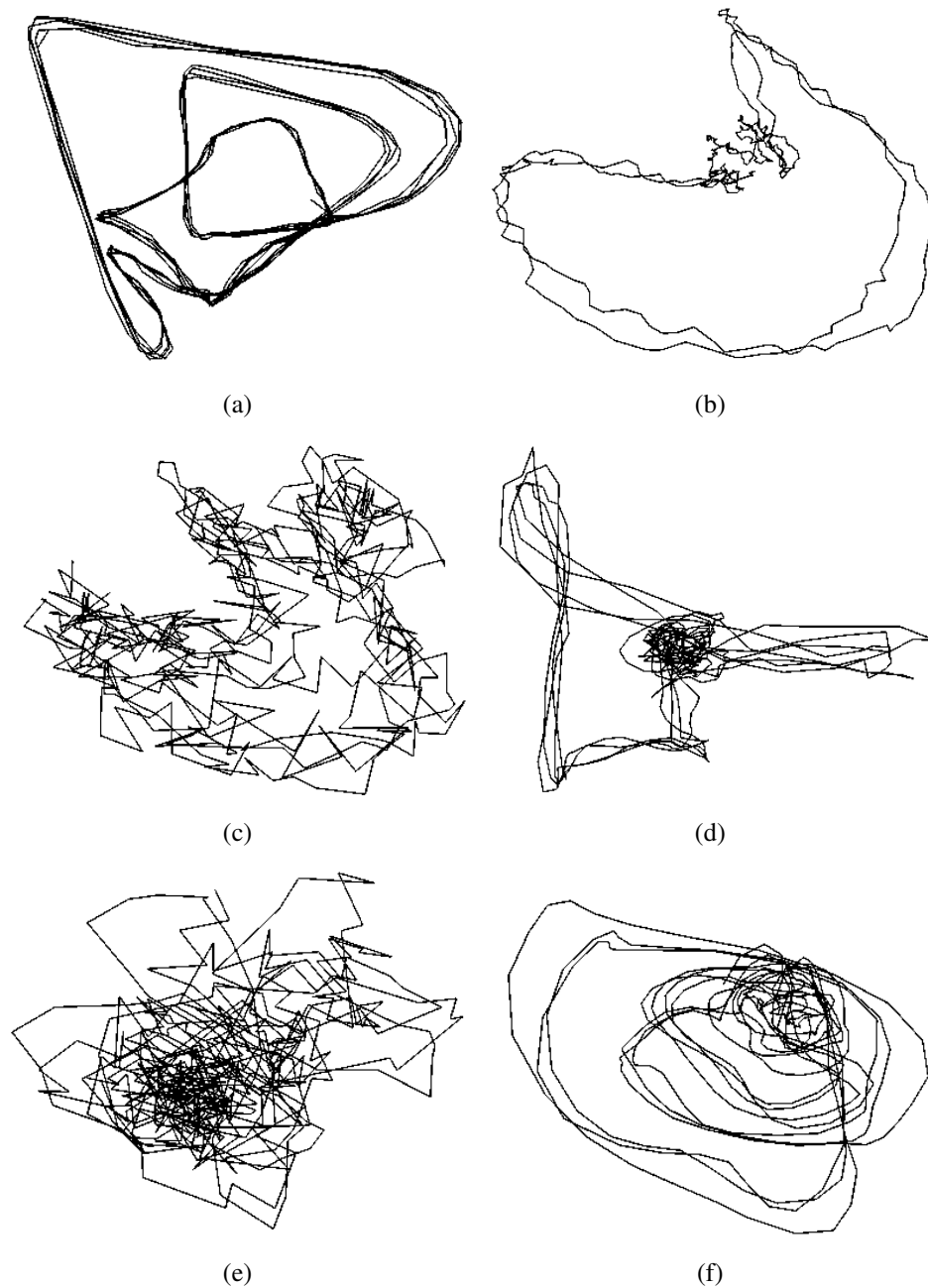
Fonte: Autoria própria.

O tamanho utilizado para cada imagem foi 570×450 pixels, dado que foi possível visualizar diferenças gráficas no tocante as diferentes classes. Para redução da dimensionalidade da imagem, foi utilizada uma variação do método das caixas visitadas (COSTA, 2012). Cada imagem foi dividida em caixas de tamanho $N \times N$ sem sobreposição e, para cada caixa, foi realizada uma contagem de quantos pixels preenchidos cada caixa possui.

Um pixel é considerado preenchido quando ele tem valor igual a 255. Por fim, foi gerada uma nova imagem em que o valor do tom de cinza de cada pixel corresponde à contagem de cada caixa. A Figura 4.3 contém um exemplo do método.

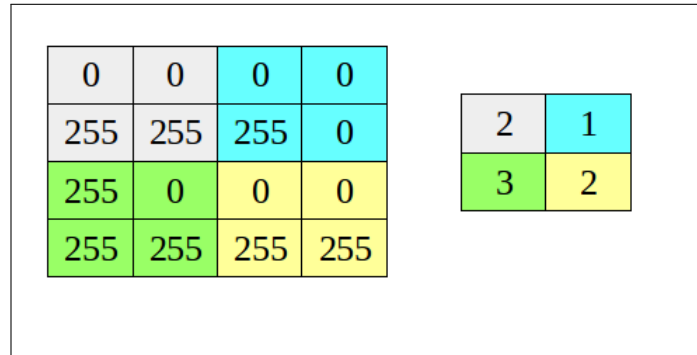
Nesta pesquisa, foram utilizados como tamanho das caixas $N = 10$ e 15 . Na Figura 4.4 pode-se observar o efeito da utilização do método em uma imagem de voz afetada por Edema. Os níveis de cinza da figura foram invertidos para melhor visualização.

Figura 4.2: Trajetória do Espaço de Fase Reconstruído dos primeiros 20 ms de (a) um sinal de voz normal (AXH1NAL), (b) um sinal de voz afetada por Edema (CAK25AN), (c) um sinal de voz afetada por Paralisia (RAN30AN), (d) um sinal de voz afetada por Nódulo (MXN24AN), (e) um sinal de voz afetada por Pólipo (MPB23AN) and (f) um sinal de voz afetada por Queratose (EMP27AN). Todos os sinais pertencem à base MEEI.



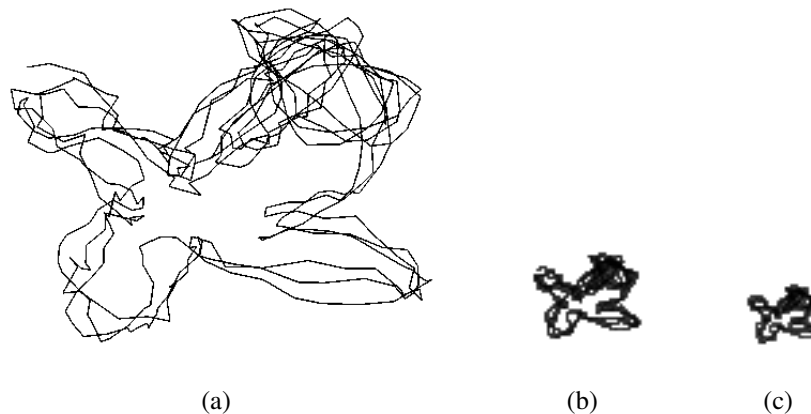
Fonte: Autoria própria.

Figura 4.3: Método da contagem de caixas com $N = 2$ e pixel preenchido = 255.



Fonte: Autoria própria.

Figura 4.4: (a) Trajetória do espaço de fase reconstruído dos primeiros 20 ms de um sinal de voz afetada por Edema (CAC10AN) da base MEEI e a imagem após a redução de dimensionalidade utilizando (b) $N = 10$, e (c) $N = 15$.

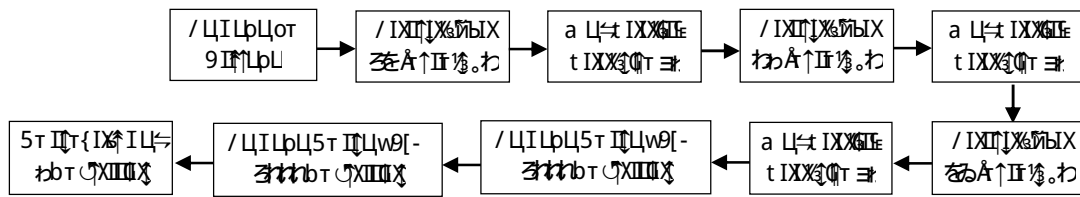


Fonte: Autoria própria.

4.2.2 Arquiteturas da CNN

A CNN foi construída de acordo com a Figura 4.5. Foram utilizadas três camadas convolucionais para descrever as imagens de entrada em um conjunto local de características, que fica mais abstrato à medida que a profundidade da rede cresce. Foi utilizada uma camada *pooling* após cada camada convolucional para reduzir a dimensionalidade. Por fim, foram acrescentadas duas camadas densamente conectadas e uma camada *softmax* com 2 neurônios (um neurônio para cada classe) para a classificação final. Para as camadas convolucionais, foram utilizados 16, 32 e 64 *kernels* de tamanho 3 e a função de ativação Relu. Outras arquiteturas foram avaliadas empiricamente, mas proporcionaram resultados piores.

Figura 4.5: Arquitetura da CNN proposta.



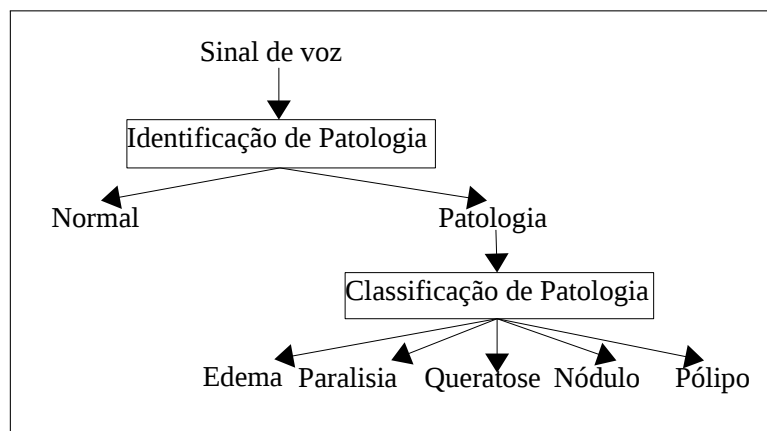
Fonte: Autoria própria.

Para a utilização do *Bottleneck*, foi inserida uma camada densa de tamanho 16 ou 32 entre as duas camadas densamente conectadas. Em experimentos preliminares, percebeu-se que essa faixa de valores para o tamanho da camada *bottleneck* proporcionou resultados satisfatórios. Portanto, foram escolhidos esses dois tamanhos para realização de comparação entre os resultados.

4.2.3 Classificação

Os principais objetivos de um sistema de auxílio ao diagnóstico de patologias das pregas vocais são a identificação da existência de uma patologia e a classificação da patologia. Um método usual para atingir esses objetivos consiste na classificação hierárquica mostrada na Figura 4.6. Esse método consiste em avaliar o sinal de voz em uma classificação Normal *versus* Patologia e, caso o sinal seja classificado como sendo Patologia, realizar uma nova avaliação para especificar qual a patologia que afeta o sinal.

Figura 4.6: Sistema de classificação hierárquica do sinal de voz.



Fonte: Autoria própria.

Para esta pesquisa, optou-se por tornar as duas etapas independentes entre si, ou seja, não foi necessária uma pré-classificação do sinal como Patologia para verificar a que classe de Patologia o sinal pertence. Optou-se também pela realização da classificação dois-a-dois das patologias escolhidas, pois é uma abordagem comum para lidar com conjuntos de dados desbalanceados, que é o caso das bases MEEI e SVD. Uma classificação final pode ser obtida a partir de múltiplas classificações dois-a-dois por meio de uma regra de votação simples (FÜRNKRANZ, 2002), por exemplo. Os tipos de classificações realizadas estão descritos no Quadro 4.1.

Cada avaliação experimental de classificação foi repetida 10 vezes e foi obtida uma média dos 10 resultados. Para cada uma das 10 vezes, os sinais foram divididos aleatoriamente em 3 conjuntos: treinamento, validação e teste. A quantidade de sinais no conjunto de treinamento corresponde a 50% dos sinais da classe com a menor quantidade de sinais. Para o conjunto de validação, a quantidade de sinais corresponde a 10% da classe com a menor quantidade de sinais. Por fim, o conjunto de teste corresponde aos sinais restantes. Por exemplo, na classificação Paralisia *versus* Pólipo da base MEEI, com 67 sinais de Paralisia e 20 sinais de Pólipo, foram utilizados 10 sinais de cada classe para o conjunto de treinamento, 2 sinais de cada classe para o conjunto de validação, e 55 sinais de Paralisia e 8 sinais de Pólipo para o conjunto de testes. Este procedimento foi realizado para balancear a quantidade de sinais de cada classe no treinamento da rede neural. Os sinais artificiais gerados pelo TS foram utilizados apenas nos conjuntos de treinamento.

Quadro 4.1: Classificações realizadas na pesquisa.

| Classificação |
|-----------------------------------|
| Normal <i>versus</i> Patologia |
| Paralisa <i>versus</i> Edema |
| Paralisia <i>versus</i> Queratose |
| Paralisia <i>versus</i> Pólipo |
| Paralisia <i>versus</i> Nódulo |
| Paralisia <i>versus</i> Lesão |
| Edema <i>versus</i> Queratose |
| Continua na página seguinte |

Quadro 4.1 – Continuação

| Classificação |
|--------------------------------|
| Edema <i>versus</i> Pólipo |
| Edema <i>versus</i> Nódulo |
| Edema <i>versus</i> Lesão |
| Queratose <i>versus</i> Pólipo |
| Queratose <i>versus</i> Nódulo |
| Queratose <i>versus</i> Lesão |
| Nódulo <i>versus</i> Pólipo |

Foi utilizado gradiente estocástico descendente com função momentum durante o treinamento do modelo proposto, com taxa de aprendizagem de 0,01 e momentum de 0,9. O tamanho do *batch* foi de $K/10$, em que K é o total de imagens do conjunto de treinamento. O critério de parada do treinamento foi de não haver progresso no erro do conjunto de validação por 5 épocas seguidas.

Para utilização da SVM, foram obtidas as ativações da camada *Bottleneck* da rede treinada de todas as imagens do conjunto de treinamento, e essas ativações foram utilizadas como vetores de características para o treinamento da SVM. O treinamento de cada SVM foi realizado utilizando *grid search*. Por fim, foram obtidas as ativações das imagens do conjunto de teste para serem avaliadas na SVM treinada.

Para avaliar cada classificação, foram utilizadas as medidas de desempenho a seguir.

- Verdadeiros Positivos (VP) - quantidade de sinais pertencentes à classe alvo classificados corretamente. Considera-se a classe alvo, na pesquisa, voz normal, na classificação Normal *versus* Patologia, e voz afetada pela patologia nas classificações Patologia *versus* Outras Patologias.
- Verdadeiros Negativos (VN) - quantidade de sinais não pertencentes à classe alvo classificados corretamente, ou seja, os sinais de voz afetada por patologia, na classificação Normal *versus* Patologia e os sinais de voz de outras patologias nas classificações Patologia *versus* Outras Patologias.

- Falsos Positivos (FP) - quantidade de sinais não pertencentes à classe alvo classificados erroneamente.
- Falsos Negativos (FN) - quantidade de sinais pertencentes à classe alvo classificados erroneamente.
- Sensibilidade (S) - razão entre verdadeiros positivos e o total de sinais pertencentes à classe alvo.

$$S = \frac{VP}{(VP + FN)}. \quad (4.1)$$

- Especificidade (E) - razão entre verdadeiros negativos e o total de sinais não pertencentes à classe alvo.

$$E = \frac{VN}{(VN + FP)}. \quad (4.2)$$

- Acurácia (AC) - razão entre a quantidade de sinais corretamente classificados e o total de sinais.

$$AC = \frac{VP + VN}{(VP + VN + FP + FN)}. \quad (4.3)$$

- Coeficiente de Correlação de Matthews (*Matthews Correlation Coefficient* - MCC) - valor entre -1 e 1, que indica a qualidade de uma classificação binária. MCC = 1 representa predição perfeita, MCC = -1 representa erro total na predição e MCC = 0 indica que a classificação não é melhor do que o aleatório. O MCC é calculado pela Equação 4.4.

$$MCC = \frac{VP \cdot VN - FP \cdot FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}. \quad (4.4)$$

A acurácia é a medida tradicional de avaliação de classificadores. No entanto, ela pode não fornecer uma informação adequada sobre a capacidade discriminativa de um classificador quando as classes são desbalanceadas (CASTRO; BRAGA, 2011), como é o caso desta pesquisa. Para contornar esse problema, optou-se por utilizar também a medida MCC, a qual é amplamente utilizada em pesquisas biomédicas pois ela é robusta para classes desbalanceadas (BOUGHORBEL; JARRAY; EL-ANBARI, 2017).

Todos os experimentos foram realizados utilizando o software MATLAB (MATLAB,

2018). O cálculo da curva de informação mútua média foi realizado utilizando o CRP toolbox (MARWAN, 2013). A reconstrução do espaço de fase foi implementada utilizando o *Chaotic Systems Toolbox* (LEONTITSIS, 2014). As funções de treinamento e teste da CNN foram implementadas utilizando o *Deep Learning Toolbox* (MATLAB..., 2018a) e o SVM foi implementado utilizando o *Statistic and Machine Learning Toolbox* (MATLAB..., 2018b).

4.2.4 Análise Estatística

Em função dos diferentes parâmetros avaliados (base de dados, com ou sem aumento de dados, tipo de classificador, tipo de janela e tamanho da caixa), cada classificação descrita na Tabela 4.1 foi realizada 48 vezes. Para verificação das hipóteses da pesquisa descritas na subseção 1.3, foi realizada uma análise estatística nos 48 resultados MCC obtidos em cada classificação. Foram verificadas ao todo 05 diferentes hipóteses, descritas a seguir:

- Hipótese 1 - Os resultados em termos do MCC, com aumento de dados, são superiores aos resultados sem o aumento de dados;
- Hipótese 2 - Os resultados em termos do MCC, usando janela variável, são superiores aos resultados com janela fixa;
- Hipótese 3 - Os resultados em termos do MCC, usando *bottleneck* de 16 neurônios combinado com SVM, são superiores aos resultados com CNN;
- Hipótese 4 - Os resultados em termos do MCC, usando *bottleneck* de 32 neurônios combinado com SVM, são superiores aos resultados com CNN;
- Hipótese 5 - Os resultados em termos do MCC, usando *bottleneck* de 32 neurônios combinado com SVM, são superiores aos resultados com *bottleneck* de 16 neurônios combinado com SVM.

A análise estatística foi realizada como se segue. Para cada classificação, os 48 resultados MCC foram divididos em dois grupos conforme a hipótese testada, e então foi realizada a comparação dos dois grupos utilizando um teste estatístico. Um teste de comparação de amostras pode ser paramétrico ou não paramétrico. O uso de um teste paramétrico

permite maior robustez na detecção de diferenças entre amostras, mas testes paramétricos só são confiáveis quando os dados obedecem à distribuição normal. Caso os dados não obedecem à distribuição normal, deve-se aplicar um teste não paramétrico (NORMANDO; TJÄDERHANE; QUINTÃO, 2010).

Para auxiliar a escolha do teste estatístico, foi aplicado o teste de Shapiro-Wilk (SHAPIRO; WILK, 1965) com o propósito de verificar se os dados obedecem à distribuição normal. Esse teste tem como hipótese nula os dados aplicados seguirem à distribuição normal, ou seja, caso o valor P resultante seja maior que 0,05, por exemplo, não seria possível rejeitar a hipótese que os dados seguem a distribuição normal com 95% de confiança.

O teste paramétrico utilizado foi o teste t de Student (STUDENT, 1908). O teste t realiza uma comparação entre duas médias. Ele tem como hipótese nula a média das duas amostras serem iguais. O teste t pode ser unicaudal ou bicaudal. O teste unicaudal é aquele em que a hipótese alternativa corresponde à média de uma das amostras ser superior à média da outra. O teste bicaudal é aquele que tem como hipótese alternativa as médias serem diferentes. Para este trabalho, foi utilizado o teste t unicaudal.

4.3 Considerações Finais

Neste capítulo, foram descritos os Materiais e Métodos da pesquisa. Inicialmente, foram descritas as bases de dados. Em seguida, foi apresentada a abordagem proposta, contendo a metodologia geral da pesquisa e a metodologia utilizada nos experimentos.

No próximo capítulo, serão apresentados e analisados os resultados obtidos nos experimentos.

Capítulo 5

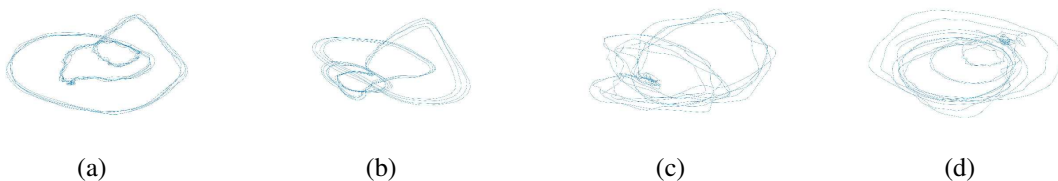
Resultados e Discussão

Neste capítulo, são apresentados e discutidos os resultados da pesquisa. Na Seção 5.1, é apresentada uma análise visual das trajetórias dos espaços de fase dos sinais de voz. Os resultados para identificação de patologias são apresentados e discutidos na Seção 5.2. Por fim, na Seção 5.3 estão apresentados e discutidos os resultados da classificação de patologias.

5.1 Análise da Trajetória do Espaço de Fase dos Sinais de Voz

Inicialmente, foram gerados gráficos das trajetórias dos espaços de fase de cada janela de cada sinal. Como se pode observar na Figura 5.1, visualmente os gráficos de sinais de voz normal são mais regulares que os gráficos de sinais de voz afetada por patologia. Isso ocorre porque sinais de voz normal possuem um comportamento mais próximo de um sinal periódico que sinais de voz afetada por patologia.

Figura 5.1: Trajetória do Espaço de Fase Reconstruído dos primeiros 20 ms de (a) um sinal de voz normal (1-a_n), (b) um sinal de voz normal (2-a_n), (c) um sinal de voz afetada por Paralisia (929-a_n) e (d) um sinal de voz afetada por Edema (1963-a_n). Todos os sinais pertencentes à base SVD.



Fonte: Autoria própria.

Continuando a análise visual, pode-se observar na Figura 5.2 que gráficos de janelas de 10 ciclos de *pitch* são mais densos que gráficos de janelas de 20 ms. Isto é fácil de ser verificado, pois as janelas de 10 ciclos são maiores, portanto, o espaço de fase terá mais estados. Uma janela de tamanho variável dependente do *pitch* pode ser de grande utilidade na formação das imagens, pois além dos estados do espaço de fase, o gráfico irá também conter informações de *pitch*, que é uma característica amplamente utilizada na literatura como parte dos descritores de características (ULOZA et al., 2011; KUMAR; BHAT; MUKHI, 2011; CORDEIRO; FONSECA; RIBEIRO, 2014; DAHMANI; GUERTI, 2017), podendo assim facilitar na classificação das imagens.

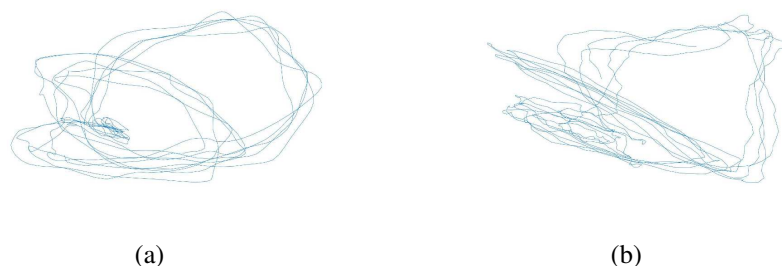
Figura 5.2: Trajetória do Espaço de Fase Reconstruído de um sinal de voz normal (1-a.n) da base SVD, com janela de tamanho (a) 20 ms e (b) 10 ciclos de *pitch*.



Fonte: Autoria própria.

Por fim, observa-se que existe uma grande diferença visual entre sinais com uma mesma patologia (Figura 5.3) ou até mesmo entre sinais de um mesmo sujeito (Figura 5.4).

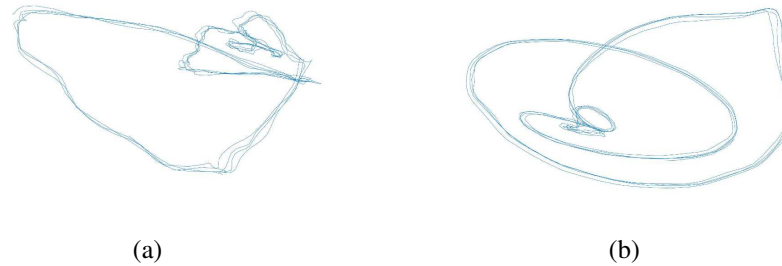
Figura 5.3: Trajetória do Espaço de Fase Reconstruído dos primeiros 20 ms de (a) um sinal de voz afetada por Paralisia (929-a.n) e (d) um sinal de voz afetada por Paralisia (130-a.n). Todos os sinais pertencentes a base SVD.



Fonte: Autoria própria.

As CNN foram utilizadas pela capacidade de extrair características relevantes das imagens, observando as semelhanças e descartando as diferenças entre imagens de uma mesma patologia.

Figura 5.4: Trajetória do Espaço de Fase Reconstruído dos primeiros 20 ms de duas elocuições de um mesmo sujeito com voz normal: (a) sinal 10-a_n e (b) sinal 2200-a_n da base SVD.



Fonte: Autoria própria.

5.2 Detecção de Patologia

A classificação Normal *versus* Patologia teve como objetivo identificar a presença de patologia no sinal de voz, sem identificar qual é a patologia. Ela consiste na primeira etapa de um sistema hierárquico para auxílio ao diagnóstico de patologias da voz. A CNN foi treinada com diversos sinais de voz normal e de voz afetada por patologia. A Figura 5.5 contém a saída da última camada convolucional da rede para 04 janelas de 04 sinais de teste distintos, sendo 02 sinais de voz normal e 02 sinais de voz afetada por patologia. As saídas dessa camada têm dimensão $14 \times 10 \times 64$, que foram reajustadas em imagens de dimensão 112×80 para facilitar a visualização. As saídas para voz normal apresentaram uma quantidade de pixels brancos muito superior às saídas de voz afetada por patologia, facilitando assim a diferenciação entre as duas classes nas camadas densas seguintes.

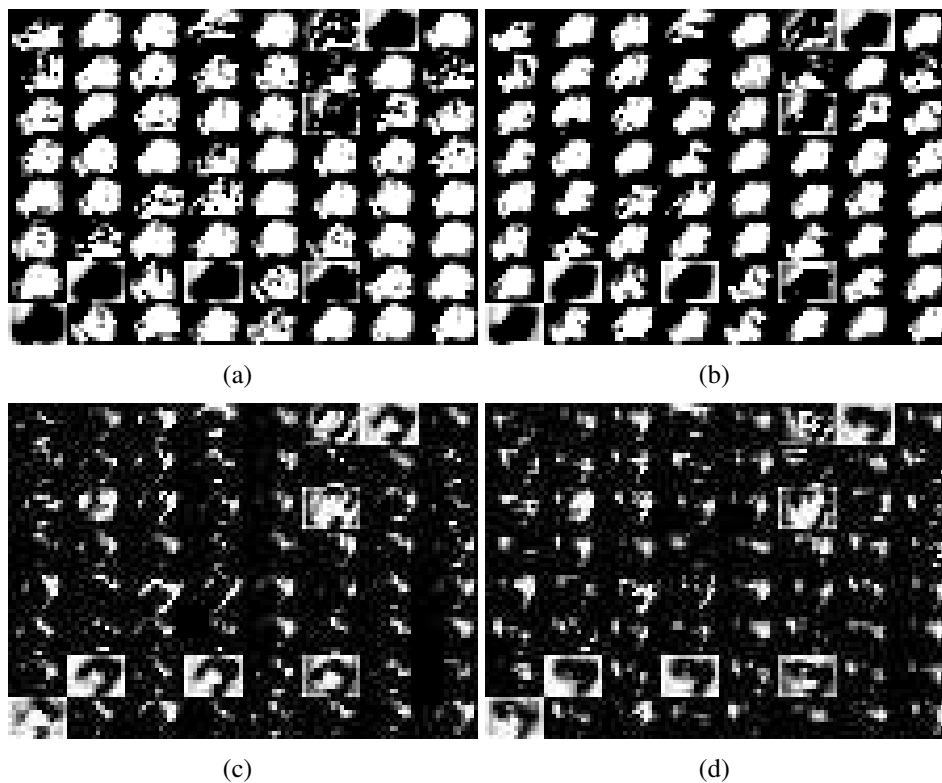
Os melhores resultados para a classificação Normal *versus* Patologia estão contidos na Tabela 5.1. Os melhores resultados obtidos utilizando a base MEEI foram de 100% em todos os casos. Além disso, o melhor resultado usando a base SVD foi acima dos 90%, ao se utilizar *Bottleneck* + SVM e aumento de dados no conjunto de treinamento. Verifica-se também que, na maioria dos casos, o melhor resultado foi obtido com uma janela variável de 10 ciclos de *pitch*.

Como se pode observar, a técnica escolhida para extração de características e classificação é adequada ao problema, possivelmente sendo necessário realizar ajustes na arquitetura da CNN para melhorar a classificação dos sinais da base SVD.

Para verificação das hipóteses da pesquisa, foi aplicado o teste de Shapiro-Wilk nos re-

sultados MCC das 48 classificações Normal *versus* Patologia realizados. O objetivo do teste foi de verificar se os dados seguem a distribuição normal, para indicar a necessidade ou não de utilização de um teste não-paramétrico para comparação dos dados. O valor P obtido no teste de Shapiro-Wilk foi 0,77, portanto não é possível rejeitar a hipótese nula de que os dados seguem a distribuição normal com 95% de confiança.

Figura 5.5: Saída da última camada convolucional para janelas de (a) e (b) dois sinais de voz normal e (c) e (d) dois sinais de voz afetada por patologia, todos da base MEEI.



Fonte: Autoria própria.

Tabela 5.1: Melhores Resultados da Classificação Normal *versus* Patologia.

| Base de Dados | TS | Classificador | Caixa | Janela | Normal (%) | Patologia (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|---------------|------|
| MEEI | Não | CNN | 15 | 20 ms | 100,00 ± 0,00 | 100,00 ± 0,00 | 100,00 ± 0,00 | 1,00 |
| MEEI | Sim | CNN | 15 | 20 ms | 100,00 ± 0,00 | 100,00 ± 0,00 | 100,00 ± 0,00 | 1,00 |
| MEEI | Não | B32+SVM | 10 | 10 Ciclos | 100,00 ± 0,00 | 100,00 ± 0,00 | 100,00 ± 0,00 | 1,00 |
| MEEI | Sim | B16+SVM | 15 | 10 Ciclos | 100,00 ± 0,00 | 100,00 ± 0,00 | 100,00 ± 0,00 | 1,00 |
| SVD | Não | CNN | 10 | 10 Ciclos | 83,09 ± 4,86 | 86,11 ± 7,13 | 83,54 ± 2,53 | 0,55 |
| SVD | Sim | CNN | 15 | 10 Ciclos | 91,79 ± 5,86 | 88,89 ± 5,95 | 91,36 ± 2,12 | 0,71 |
| SVD | Não | B32+SVM | 15 | 10 Ciclos | 87,92 ± 4,38 | 84,72 ± 6,43 | 87,45 ± 2,08 | 0,61 |
| SVD | Sim | B32+SVM | 15 | 10 Ciclos | 93,24 ± 2,97 | 91,67 ± 4,48 | 93,00 ± 1,70 | 0,76 |

A partir deste resultado, optou-se por aplicar o teste t de Student unicaudal nos resultados MCC obtidos, conforme discutido na subseção 4.2.4. Os 48 resultados MCC foram divididos

em dois grupos (conforme os pares de critérios contrastados) para a aplicação do teste. Na Tabela 5.2 estão contidos os valores P dos testes realizados. A hipótese alternativa é que os resultados obtidos usando o critério 1 são superiores aos resultados obtidos usando o critério 2.

Tabela 5.2: Valor P do teste t de Student dos resultados MCC da classificação Normal *versus* Patologia.

| Critério 1 | Critério 2 | p-value |
|--------------------------|----------------------|---------|
| Com aumento de dados | Sem aumento de dados | 0,00 |
| Janela variável | Janela fixa | 0,00 |
| <i>Bottleneck</i> 16+SVM | CNN | 0,00 |
| <i>Bottleneck</i> 32+SVM | CNN | 0,00 |
| <i>Bottleneck</i> 32+SVM | <i>Bottleneck</i> 16 | 0,03 |

Observa-se que é possível rejeitar a hipótese nula de que os resultados obtidos, quando se usa o aumento de dados, são iguais aos resultados quando não se usa aumento de dados, com 95% de confiança. O uso de janelas de tamanho variável também se mostrou superior e o melhor classificador encontrado foi o *bottleneck*, com 32 neurônios na camada escondida, combinado com SVM.

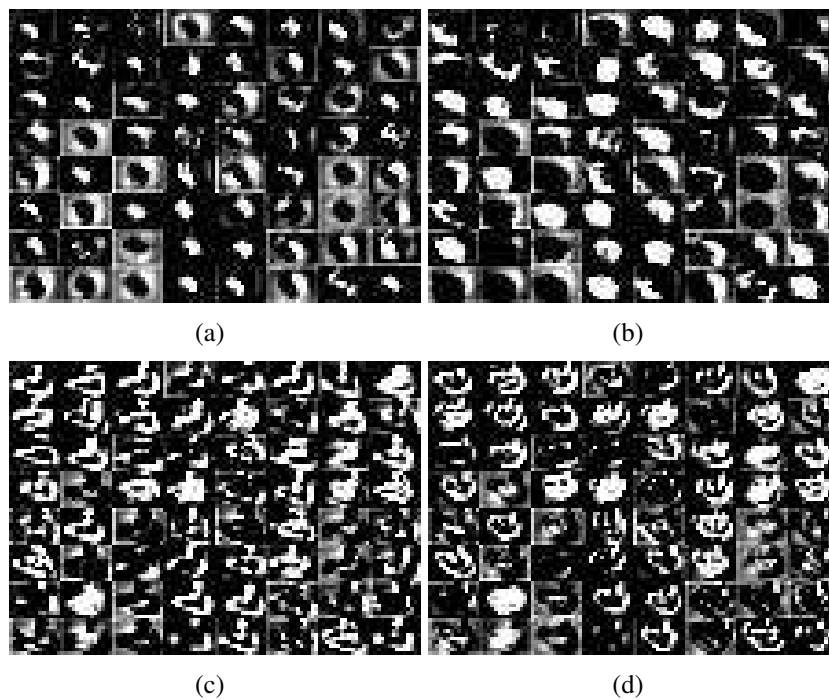
5.3 Identificação de Patologias

Conforme apresentado, foi realizado o estudo de 06 classes de patologias: Paralisia, Edema, Nódulo, Pólipo, Queratose e Lesão Benigna, sendo as 05 primeiras patologias das pregas vocais e a última um agrupamento de patologias contendo sinais de voz afetada por Nódulo, Pólipo e Cisto. Foram realizadas classificações par-a-par entre as 05 patologias estudadas e também entre as classificações Paralisia, Edema ou Queratose *versus* Lesão Benigna. Em um sistema de auxílio ao diagnóstico, um sinal de voz seria submetido a todos os classificadores e o resultado de cada um deles seria combinado por meio de métodos como o de votação simples, que considera apenas a classificação, ou votação ponderada, que considera o valor de saída da rede, para a identificação da patologia. Esta última etapa não faz parte do escopo desta pesquisa.

Nas Figuras 5.6 e 5.7 estão contidas as saídas da última camada convolucional da rede

para 04 janelas de 04 sinais de teste distintos, sendo 02 sinais de cada classe. As saídas dessa camada têm dimensão $14 \times 10 \times 64$, que foram reajustadas em imagens de dimensão 112×80 , colocando as 64 imagens 14×10 lado a lado, para facilitar a visualização.

Figura 5.6: Saída da última camada convolucional para janelas de (a) e (b) dois sinais de voz afetada por Paralisia e (c) e (d) dois sinais de voz afetada por Edema, todos da base MEEI.



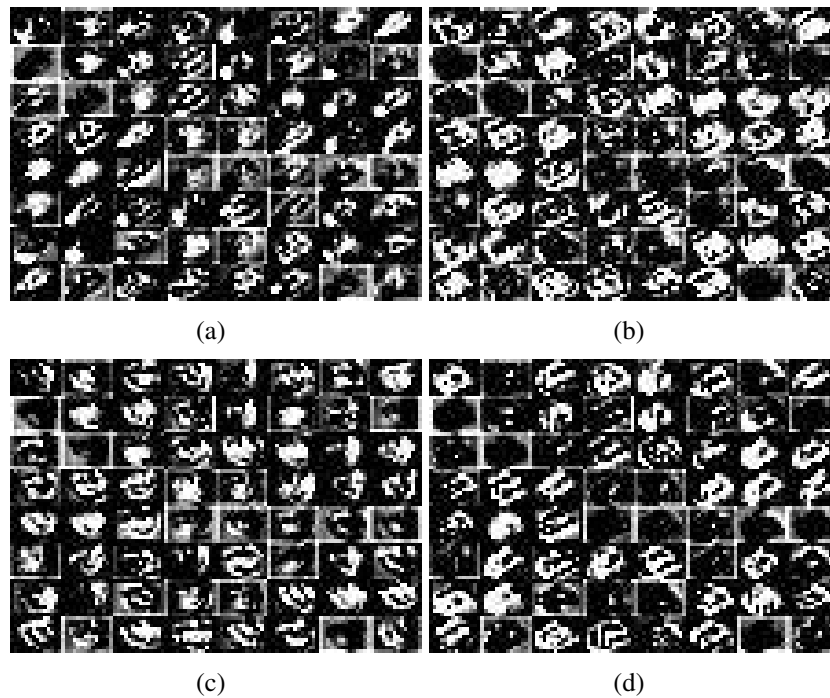
Fonte: Autoria própria.

Como se pode observar, na classificação *Paralisia versus Edema*, é possível verificar uma diferença visual significativa entre os padrões das duas classes, em que as janelas dos sinais de Paralisia consistem em um centro branco com contorno preto, ou as cores invertidas, e as janelas dos sinais com Edema com um padrão parecido com um “cabide”. Na classificação *Paralisia versus Queratose*, o grau de dificuldade é elevado para diferenciação com base na análise visual. Essa dificuldade existe em várias outras das classificações realizadas. Neste sentido, torna-se justificável a utilização de outros classificadores que não as camadas densas do fim da CNN para realizar a diferenciação entre as classes. Nesta pesquisa, foi utilizado SVM.

Nas Tabelas 5.3 e 5.4 estão contidos os melhores resultados conforme a métrica MCC de cada classificação, para cada uma das bases. Em caso de empate, foi escolhido o resultado com a maior acurácia. Persistindo o empate, foi escolhido o resultado com o menor custo

computacional. Tabelas com resultados mais detalhados estão no apêndice D.

Figura 5.7: Saída da última camada convolucional para janelas de (a) e (b) dois sinais de voz afetada por Paralisia e (c) e (d) dois sinais de voz afetada por Queratose, todos da base MEEI.



Fonte: Autoria própria.

Uma das vantagens do método de obtenção de imagens, a partir da trajetória do espaço de fase, foi a possibilidade de utilizar um tamanho de janela variado. Com isto, pode-se agregar mais informações à imagem além daquelas já presentes no espaço de fase. No caso desta pesquisa, a janela continha informações do *pitch* do sinal, aumentando assim a capacidade decisória do sistema. Pode-se observar que na maioria das classificações, o melhor resultado foi obtido com a janela variável de 10 ciclos de *pitch*, em vez da tradicional janela fixa de 20 ms. Mesmo usando menos sinais para treinar a CNN, essa janela mostrou conter mais informações da patologia do que a janela de 20 ms.

Outro fato observado é que a utilização de *Bottleneck* + SVM proporcionou resultados superiores ao uso apenas da CNN na maioria das classificações. Uma possibilidade é que a CNN tenha ótima capacidade de extração de características das patologias em suas camadas convolucionais, mas o SVM é superior aos neurônios densos das últimas camadas da CNN para classificar essas características.

Percebe-se ainda que o aumento dos dados de treinamento proporcionou melhores re-

sultados, exceto em algumas classificações envolvendo Pólipo. As técnicas de aumento de dados permitem a criação de novos dados de treinamento, para que ocorra melhor aprendizado da CNN, mas como estas se baseiam na distorção do sinal para criar um novo, existe a possibilidade de perda de informação. Esse pode ser o caso do Pólipo. Possivelmente, o aumento de dados causou uma perda de informações dessa patologia.

Tabela 5.3: Melhores Resultados de cada Classificação na base MEEI.

| Classe 1 | Classe 2 | TS | Classificador | Caixa | Janela | Classe 1 (%) | Classe 2 (%) | Total (%) | MCC |
|-----------|-----------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| Paralisia | Edema | Sim | B16+SVM | 15 | 10 Ciclos | 88,24 ± 3,77 | 72,73 ± 5,62 | 84,44 ± 2,77 | 0,59 |
| Paralisia | Queratose | Sim | B32+SVM | 10 | 10 Ciclos | 80,83 ± 4,46 | 94,29 ± 3,91 | 82,55 ± 4,07 | 0,55 |
| Paralisia | Pólipo | Sim | CNN | 15 | 10 Ciclos | 85,38 ± 5,70 | 72,00 ± 11,37 | 84,21 ± 5,42 | 0,41 |
| Paralisia | Nódulo | Sim | CNN | 10 | 10 Ciclos | 95,85 ± 3,24 | 80,00 ± 4,47 | 94,48 ± 2,87 | 0,69 |
| Paralisia | Lesão | Não | B32+SVM | 15 | 10 Ciclos | 80,00 ± 6,08 | 72,73 ± 6,62 | 78,26 ± 4,81 | 0,48 |
| Edema | Queratose | Sim | B32+SVM | 10 | 20 ms | 88,00 ± 3,06 | 80,00 ± 4,59 | 86,25 ± 3,68 | 0,63 |
| Edema | Pólipo | Não | B32+SVM | 15 | 10 Ciclos | 89,66 ± 5,67 | 84,00 ± 8,94 | 88,82 ± 5,34 | 0,64 |
| Edema | Nódulo | Sim | CNN | 15 | 10 Ciclos | 72,41 ± 5,53 | 75,00 ± 10,49 | 72,73 ± 3,84 | 0,33 |
| Edema | Lesão | Sim | B32+SVM | 15 | 10 Ciclos | 69,09 ± 7,10 | 80,00 ± 11,32 | 74,29 ± 3,16 | 0,49 |
| Queratose | Pólipo | Sim | B16+SVM | 15 | 10 Ciclos | 85,45 ± 2,49 | 80,00 ± 15,00 | 83,75 ± 4,64 | 0,64 |
| Queratose | Nódulo | Sim | B16+SVM | 10 | 10 Ciclos | 86,67 ± 5,43 | 100,00 ± 0,00 | 90,59 ± 3,83 | 0,81 |
| Queratose | Lesão | Sim | B32+SVM | 10 | 10 Ciclos | 93,33 ± 3,80 | 78,26 ± 6,43 | 81,38 ± 3,83 | 0,60 |
| Nódulo | Pólipo | Não | B32+SVM | 10 | 10 Ciclos | 100,00 ± 0,00 | 96,67 ± 1,87 | 98,18 ± 0,57 | 0,96 |

Tabela 5.4: Melhores Resultados de cada Classificação na base SVD.

| Classe 1 | Classe 2 | TS | Classificador | Caixa | Janela | Classe 1 (%) | Classe 2 (%) | Total (%) | MCC |
|-----------|-----------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| Paralisia | Edema | Sim | B32+SVM | 15 | 10 Ciclos | 78,95 ± 3,32 | 73,33 ± 5,94 | 78,41 ± 2,56 | 0,35 |
| Paralisia | Queratose | Sim | B16+SVM | 15 | 20 ms | 77,73 ± 7,45 | 62,50 ± 7,13 | 76,86 ± 6,76 | 0,22 |
| Paralisia | Pólipo | Sim | B16+SVM | 15 | 10 Ciclos | 75,28 ± 7,14 | 68,89 ± 12,04 | 74,85 ± 5,98 | 0,25 |
| Paralisia | Nódulo | Sim | B16+SVM | 10 | 10 Ciclos | 68,47 ± 7,21 | 96,00 ± 5,48 | 69,40 ± 6,89 | 0,25 |
| Paralisia | Lesão | Sim | B16+SVM | 15 | 10 Ciclos | 68,45 ± 2,61 | 64,29 ± 3,57 | 67,99 ± 1,99 | 0,22 |
| Edema | Queratose | Sim | B32+SVM | 10 | 20 ms | 73,33 ± 5,60 | 87,50 ± 8,15 | 76,32 ± 3,30 | 0,51 |
| Edema | Pólipo | Não | B32+SVM | 10 | 20 ms | 76,00 ± 4,56 | 82,22 ± 8,43 | 77,65 ± 1,32 | 0,53 |
| Edema | Nódulo | Sim | B32+SVM | 15 | 20 ms | 79,02 ± 6,84 | 67,50 ± 16,77 | 78,00 ± 5,53 | 0,31 |
| Edema | Lesão | Sim | CNN | 10 | 10 Ciclos | 67,69 ± 10,39 | 77,65 ± 7,38 | 73,33 ± 4,47 | 0,46 |
| Queratose | Pólipo | Não | B16+SVM | 10 | 10 Ciclos | 75,00 ± 7,65 | 64,29 ± 8,14 | 68,18 ± 5,43 | 0,38 |
| Queratose | Nódulo | Sim | B32+SVM | 10 | 20 ms | 91,00 ± 7,58 | 88,00 ± 7,07 | 90,40 ± 6,54 | 0,73 |
| Queratose | Lesão | Sim | CNN | 10 | 10 Ciclos | 70,00 ± 11,35 | 61,76 ± 6,74 | 63,33 ± 3,82 | 0,25 |
| Nódulo | Pólipo | Não | B32+SVM | 10 | 20 ms | 82,50 ± 0,00 | 79,20 ± 7,40 | 79,66 ± 6,38 | 0,47 |

Verifica-se também que os resultados na base MEEI foram muito superiores aos resultados na base SVD. É possível que na diminuição da taxa de amostragem de 50 mil amostras por segundo para 25 mil amostras por segundo ocorrida nos sinais da base SVD, para igualar com a taxa de amostragem dos sinais da base MEEI, os sinais tenham perdido informações importantes que estão presentes nos sinais da base MEEI que já foram adquiridos a 25 mil amostras por segundo. Outra possibilidade consiste na diferença das condições de aquisição do sinal. No caso das duas bases, diferenças na qualidade dos microfones utilizados.

Para verificação das hipóteses da pesquisa, foi aplicado o teste de Shapiro-Wilk nos resultados MCC obtidos. O objetivo do teste foi verificar se os dados seguem a distribuição

normal, para indicar a necessidade ou não de utilização de um teste não-paramétrico para comparação dos dados. Os valores P obtidos no teste de Shapiro-Wilk estão contidos na Tabela 5.5.

Tabela 5.5: Valores P do teste de Shapiro-Wilk para verificação da normalidade dos resultados MCC das classificações entre patologias.

| Classe 1 | Classe 2 | <i>p-value</i> |
|-----------|-----------|----------------|
| Paralisia | Edema | 0,19 |
| Paralisia | Queratose | 0,07 |
| Paralisia | Pólipo | 0,23 |
| Paralisia | Nódulo | 0,08 |
| Paralisia | Lesão | 0,10 |
| Edema | Queratose | 0,64 |
| Edema | Pólipo | 0,13 |
| Edema | Nódulo | 0,58 |
| Edema | Lesão | 0,80 |
| Queratose | Pólipo | 0,73 |
| Queratose | Nódulo | 0,65 |
| Queratose | Lesão | 0,08 |
| Nódulo | Pólipo | 0,06 |

Como pode-se observar na Tabela 5.5, em nenhuma das classificações foi possível rejeitar a hipótese nula de que os dados seguem a distribuição normal com 95% de confiança. Portanto, conforme discutido na subseção 4.2.4, as hipóteses da pesquisa foram verificadas com o teste t de Student unicaudal nos resultados MCC obtidos. Ao todo, cada uma das 13 classificações entre patologias foi realizada 48 vezes, em função da variação de parâmetros (base de dados, uso ou não de aumento de dados, tipo de classificador, tipo de janela e tamanho da caixa). Para cada classificação, foram aplicados 05 testes distintos. A seguir, apresentam-se as hipóteses de cada um dos testes.

- Teste 1:
 - H_0 : A média dos resultados em termos de MCC com aumento de dados é igual a média dos resultados sem o aumento de dados;
 - H_1 : A média dos resultados em termos de MCC com aumento de dados é superior à média dos resultados sem o aumento de dados;

- Teste 2:
 - H_0 : A média dos resultados em termos de MCC com janela variável é igual a média dos resultados com janela fixa;
 - H_1 : A média dos resultados em termos de MCC com janela variável é superior à média dos resultados com janela fixa;

- Teste 3:
 - H_0 : A média dos resultados em termos de MCC usando *bottleneck* de 16 neurônios combinado com SVM é igual a média dos resultados com CNN;
 - H_1 : A média dos resultados em termos de MCC usando *bottleneck* de 16 neurônios combinado com SVM é superior à média dos resultados com CNN;

- Teste 4:
 - H_0 : A média dos resultados em termos de MCC usando *bottleneck* de 32 neurônios combinado com SVM é igual a média dos resultados com CNN;
 - H_1 : A média dos resultados em termos de MCC usando *bottleneck* de 32 neurônios combinado com SVM é superior à média dos resultados com CNN;

- Teste 5:
 - H_0 : A média dos resultados em termos de MCC usando *bottleneck* de 32 neurônios combinado com SVM é igual a média dos resultados usando *bottleneck* de 16 neurônios combinado com SVM;
 - H_1 : A média dos resultados em termos de MCC usando *bottleneck* de 32 neurônios combinado com SVM é superior à média dos resultados usando *bottleneck* de 16 neurônios combinado com SVM;

Na Tabela 5.6 estão contidos os *p-values* dos testes realizados.

Pode-se observar que, na maioria das classificações, pode-se rejeitar a hipótese nula de que os resultados quando se usa o aumento de dados são iguais aos resultados quando não se usa aumento de dados com 95% de confiança. A hipótese nula não pode ser rejeitada nas classificações envolvendo pólipos, corroborando o que foi visto nas tabelas de resultados.

Tabela 5.6: P-value do teste t de Student dos resultados MCC das classificações entre patologias.

| Classe 1 | Classe 2 | Teste 1 | Teste 2 | Teste 3 | Teste 4 | Teste 5 |
|-----------|-----------|---------|---------|---------|---------|---------|
| Paralisia | Edema | 0,00 | 0,00 | 0,00 | 0,00 | 0,01 |
| Paralisia | Queratose | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 |
| Paralisia | Pólipo | 0,57 | 0,00 | 0,00 | 0,00 | 0,32 |
| Paralisia | Nódulo | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Paralisia | Lesão | 0,03 | 0,00 | 0,00 | 0,00 | 0,02 |
| Edema | Queratose | 0,00 | 0,54 | 0,00 | 0,00 | 0,01 |
| Edema | Pólipo | 0,95 | 0,01 | 0,00 | 0,00 | 0,00 |
| Edema | Nódulo | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Edema | Lesão | 0,22 | 0,00 | 0,00 | 0,00 | 0,06 |
| Queratose | Pólipo | 0,25 | 0,00 | 0,00 | 0,00 | 0,88 |
| Queratose | Nódulo | 0,00 | 0,00 | 0,00 | 0,00 | 0,01 |
| Queratose | Lesão | 0,01 | 0,00 | 0,00 | 0,00 | 0,11 |
| Nódulo | Pólipo | 0,77 | 0,00 | 0,00 | 0,00 | 0,00 |

No geral, o uso de janelas de 10 ciclos de *pitch* também se mostrou superior e o melhor classificador encontrado foi o *bottleneck* com 32 neurônios na camada escondida e SVM.

Para concluir, foi observado que existe dificuldade em realizar uma comparação direta entre os diferentes trabalhos da área, pois são utilizadas diferentes bases de dados e metodologias experimentais em suas classificações. Mesmo nos casos em que são utilizados os mesmos bancos de dados, a quantidade de sinais utilizada é diferente de um trabalho para outro, inclusive do mesmo autor, como os trabalhos de Al-Nasheri et al. (2017) e Al-nasheri et al. (2017). Portanto, para efeito de comparação, foram desenvolvidas implementações próprias de artigos da área com qualis igual ou superior a B1 em Computação. Os códigos-fonte estão no Apêndice E. Os códigos foram aplicados aos subconjuntos utilizados nesta pesquisa para efeito de comparação. Os resultados das avaliações experimentais das implementações em termos de acurácia estão contidos nas Tabelas 5.7 e 5.8.

Tabela 5.7: Resultados obtidos nas implementações dos trabalhos para a base MEEI.

| | | Sodré, Rosa e Dassic-Leite (2017) | | | Dahmani e Guerti (2017) | | | Fang et al. (2018) | | | Ali et al. (2018) | | |
|-----------|-----------|-----------------------------------|--------------|-----------|-------------------------|--------------|-----------|--------------------|--------------|-----------|-------------------|--------------|-----------|
| Classe 1 | Classe 2 | Classe 1 (%) | Classe 2 (%) | Total (%) | Classe 1 (%) | Classe 2 (%) | Total (%) | Classe 1 (%) | Classe 2 (%) | Total (%) | Classe 1 (%) | Classe 2 (%) | Total (%) |
| Paralisia | Edema | 60,00 | 51,25 | 56,50 | 65,56 | 72,73 | 67,63 | 74,29 | 56,67 | 67,24 | 80,30 | 32,56 | 61,47 |
| Paralisia | Queratose | 70,83 | 34,00 | 60,00 | 67,00 | 70,00 | 67,39 | 96,92 | 8,00 | 72,51 | 95,45 | 20,00 | 74,73 |
| Paralisia | Nódulo | 79,17 | 42,50 | 70,00 | 78,91 | 82,00 | 79,22 | 86,59 | 25,00 | 73,10 | 98,51 | 15,79 | 80,23 |
| Paralisia | Pólipo | 74,17 | 40,00 | 65,63 | 59,35 | 42,00 | 57,65 | 100,00 | 0,00 | 76,99 | 100,00 | 0,00 | 77,01 |
| Paralisia | Lesão | 78,33 | 52,22 | 67,14 | 52,50 | 74,00 | 58,16 | 72,97 | 37,22 | 58,98 | 73,13 | 16,28 | 50,91 |
| Edema | Queratose | 83,75 | 20,00 | 59,23 | 66,25 | 68,33 | 66,67 | 93,33 | 16,00 | 64,62 | 81,40 | 52,00 | 70,59 |
| Edema | Nódulo | 81,25 | 23,33 | 65,45 | 57,24 | 67,50 | 58,48 | 100,00 | 0,00 | 72,03 | 100,00 | 0,00 | 71,93 |
| Edema | Pólipo | 71,25 | 20,00 | 57,27 | 49,31 | 55,00 | 50,00 | 97,50 | 0,00 | 67,60 | 93,02 | 15,79 | 69,35 |
| Edema | Lesão | 41,25 | 56,25 | 48,75 | 37,27 | 56,67 | 46,00 | 70,00 | 25,00 | 48,17 | 48,72 | 31,58 | 40,26 |
| Queratose | Nódulo | 66,00 | 62,50 | 64,44 | 77,50 | 94,00 | 82,35 | 74,67 | 53,33 | 64,50 | 80,77 | 63,16 | 73,33 |
| Queratose | Pólipo | 55,00 | 43,33 | 50,00 | 45,45 | 55,00 | 48,00 | 100,00 | 0,00 | 57,58 | 43,48 | 41,18 | 42,50 |
| Queratose | Lesão | 22,00 | 84,29 | 58,33 | 68,00 | 63,33 | 64,23 | 18,00 | 92,50 | 65,55 | 31,82 | 92,31 | 70,49 |
| Nódulo | Pólipo | 73,33 | 70,00 | 71,67 | 85,00 | 87,50 | 86,25 | 43,33 | 83,33 | 62,62 | 77,78 | 73,68 | 75,68 |

Tabela 5.8: Resultados obtidos nas implementações dos trabalhos para a base SVD.

| | | Sodré, Rosa e Dassic-Leite (2017) | | | Dahmani e Guerti (2017) | | | Fang et al. (2018) | | | Ali et al. (2018) | | |
|-----------|-----------|-----------------------------------|--------------|-----------|-------------------------|--------------|-----------|--------------------|--------------|-----------|-------------------|--------------|-----------|
| Classe 1 | Classe 2 | Classe 1 (%) | Classe 2 (%) | Total (%) | Classe 1 (%) | Classe 2 (%) | Total (%) | Classe 1 (%) | Classe 2 (%) | Total (%) | Classe 1 (%) | Classe 2 (%) | Total (%) |
| Paralisia | Edema | 92,31 | 13,33 | 72,00 | 50,20 | 59,09 | 51,10 | 98,67 | 4,00 | 75,77 | 80,00 | 37,50 | 69,70 |
| Paralisia | Queratose | 90,36 | 2,86 | 72,86 | 67,13 | 65,00 | 66,99 | 96,77 | 21,90 | 84,02 | 83,33 | 31,25 | 74,47 |
| Paralisia | Nódulo | 97,86 | 0,00 | 85,63 | 46,19 | 72,50 | 47,00 | 96,77 | 6,67 | 87,85 | 80,13 | 56,25 | 77,91 |
| Paralisia | Pólipo | 92,86 | 8,57 | 76,00 | 56,67 | 66,67 | 57,42 | 94,11 | 19,29 | 79,95 | 82,47 | 55,56 | 77,37 |
| Paralisia | Lesão | 88,52 | 12,73 | 66,58 | 40,41 | 71,43 | 44,32 | 89,68 | 10,91 | 68,70 | 100,00 | 0,00 | 73,46 |
| Edema | Queratose | 73,64 | 38,57 | 60,00 | 60,36 | 63,75 | 61,11 | 94,00 | 6,67 | 61,58 | 74,07 | 62,50 | 69,77 |
| Edema | Nódulo | 79,00 | 43,33 | 70,77 | 62,05 | 65,00 | 62,33 | 100,00 | 0,00 | 76,86 | 84,91 | 37,50 | 73,91 |
| Edema | Pólipo | 57,00 | 35,71 | 48,24 | 50,42 | 60,00 | 53,03 | 88,00 | 17,14 | 59,28 | 58,82 | 57,14 | 58,14 |
| Edema | Lesão | 60,00 | 45,00 | 52,11 | 36,67 | 60,59 | 50,69 | 18,00 | 74,55 | 47,19 | 51,02 | 62,26 | 56,86 |
| Queratose | Nódulo | 70,00 | 57,50 | 65,45 | 74,74 | 72,50 | 74,35 | 81,90 | 48,33 | 70,30 | 62,50 | 29,41 | 51,02 |
| Queratose | Pólipo | 38,33 | 60,00 | 50,00 | 37,50 | 57,69 | 50,00 | 0,00 | 100,00 | 54,40 | 87,10 | 54,05 | 69,12 |
| Queratose | Lesão | 30,00 | 56,36 | 47,06 | 56,25 | 66,67 | 64,63 | 8,57 | 96,67 | 65,84 | 35,48 | 75,44 | 61,36 |
| Nódulo | Pólipo | 60,00 | 81,43 | 75,00 | 77,50 | 65,83 | 67,50 | 0,00 | 100,00 | 69,88 | 81,25 | 70,27 | 73,58 |

Como pode-se observar, os resultados obtidos nesta pesquisa foram superiores aos obtidos por essas implementações. Por fim, no Apêndice F está contida uma tabela com um resumo dos resultados obtidos na literatura a partir de 2011 em classificações equivalentes às realizadas nesta pesquisa. Para uma comparação mais realista com o estado da arte, optou-se por incluir na tabela apenas trabalhos publicados em periódicos ou conferências com *qualis* da CAPES acima ou igual à B4 em Ciência da Computação e teses de doutorado.

5.4 Considerações Finais

Neste capítulo, foram apresentados e discutidos os resultados da pesquisa. Inicialmente, foi apresentada uma análise visual das trajetórias dos espaços de fase dos sinais de voz. Em seguida, os resultados para identificação de patologias foram apresentados e discutidos. Por fim, foram apresentados e discutidos os resultados para classificação de patologias.

No próximo capítulo, serão apresentadas as considerações finais da pesquisa e sugestões para pesquisas futuras.

Capítulo 6

Considerações Finais e Sugestões para Pesquisas Futuras

Neste capítulo, são apresentadas as conclusões da pesquisa e sugestões para pesquisas futuras.

6.1 Sumário da Pesquisa

Nesta pesquisa, foi proposta uma abordagem para classificação entre 06 diferentes classes de patologias das pregas vocais: Edema, Paralisia, Queratose, Nódulo, Pólipo e Lesão Benigna. Ao todo, foram realizadas 14 classificações: Normal *versus* Patologia, 10 classificações 2 a 2 entre as classes Edema, Paralisia, Queratose, Nódulo e Pólipo, e 03 classificações Edema, Paralisia ou Queratose *versus* Lesão Benigna.

A abordagem proposta consiste na utilização de imagens obtidas a partir de segmentos do sinal utilizando a técnica de reconstrução do espaço de fase e, a partir dessas imagens, treinamento de um classificador para que este aprenda a diferenciar entre duas classes de patologia.

A seguir, são apresentados os procedimentos realizados durante a fase de obtenção de resultados desta pesquisa.

1. Escolha das classes de patologia das pregas vocais e dos sinais a serem utilizados das bases de dados disponíveis:

- MEEI;
 - SVD;
2. Aumento da quantidade de sinais utilizando o método de *Time Stretching*;
 3. Obtenção do primeiro mínimo da curva de informação mútua média (τ) de cada sinal;
 4. Segmentação dos sinais. Foram utilizados dois tamanhos diferentes de segmentos:
 - 10 ciclos do *pitch* do sinal;
 - 20 ms;
 5. Reconstrução do espaço de fase de cada segmento utilizando o τ encontrado do sinal e $m = 2$;
 6. Geração da imagem da trajetória do espaço de fase reconstruído;
 7. Diminuição das dimensões da imagem utilizando o método de contagem de caixas modificado, com tamanho de caixa $N = 10$ e 15 ;
 8. Divisão dos sinais em conjuntos de treinamento, validação e teste;
 9. Treinamento da CNN. Foram treinadas CNN com e sem uma camada *bottleneck*;
 10. Treinamento da SVM a partir de vetores de características obtidos da camada *bottleneck*;
 11. Teste da CNN sem a camada *bottleneck* e da SVM;
 12. Análise estatística dos resultados obtidos.

Para cada uma das 14 classificações, foi realizado um experimento fatorial com as variáveis a seguir.

- Base de dados utilizada (MEEI ou SVD);
- Tamanho da janela (10 ciclos de *pitch* ou 20 ms);
- Tamanho da caixa (10 ou 15);

- Uso ou não de aumento de dados no conjunto de treinamento;
- Classificador utilizado (CNN ou *Bottleneck* + SVM);
- Tamanho da camada *bottleneck* (16 ou 32).

Para avaliação dos resultados, cada combinação foi realizada 10 vezes com os sinais dos conjuntos de treinamento, validação e teste escolhidos aleatoriamente em cada uma das vezes.

6.2 Contribuições da Pesquisa

A análise não linear já vem sendo utilizada há alguns anos para classificação de patologias das pregas vocais. Neste sentido, a maior contribuição da pesquisa está na utilização de imagens obtidas a partir da trajetória do espaço de fase, permitindo um novo campo de estudos em classificação de patologias das pregas vocais, com a utilização de técnicas de processamento de imagens e visão computacional.

Foi observado que, apesar de ser possível obter 100% de acerto para diferenciar sinais de voz normal de sinais de voz afetada por patologia na base MEEI, o mesmo não foi possível utilizando-se a base SVD, assim como foi reportado na literatura (AL-NASHERI et al., 2014; ALI et al., 2017; MUHAMMAD et al., 2017), o que mostra que a classificação Normal *versus* Patologia ainda necessita de mais pesquisas.

Uma vantagem verificada do método proposto é a utilização de tamanhos de janela variáveis. Nesta pesquisa, comparam-se janelas de tamanho de 10 ciclos de *pitch*, que varia para cada sinal, com janelas fixas de 20 ms. Observou-se que janelas de 10 ciclos proporcionaram resultados melhores que janelas de 20 ms, pois contêm mais informações.

Verificou-se também que o aumento de dados dos conjuntos de treinamento proporcionou melhores resultados no geral, exceto em classificações envolvendo a classe Pólipo.

Para classificação, verificou-se que utilizar a CNN apenas para extrair características da imagem, por meio de uma camada *bottleneck* e usar SVM como classificador, proporcionaram resultados melhores do que os obtidos com a utilização da CNN.

Em relação às bases de dados, a maioria das classificações utilizando a base MEEI proporcionou resultados acima dos 80% e classificações envolvendo a classe Nódulo tiveram

acurácia acima dos 90%, exceto Edema *versus* Nódulo. As classificações na base SVD proporcionaram resultados inferiores às classificações da base MEEI.

Por fim, foram realizadas duas publicações na área com resultados preliminares da pesquisa. Um capítulo de livro contendo uma revisão sistemática da literatura na área (MARINUS; ARAÚJO; GOMES, 2018a) e um artigo apresentado em um congresso internacional com os resultados obtidos utilizando CNN (MARINUS; ARAÚJO; GOMES, 2018b).

6.3 Propostas para Pesquisas Futuras

A seguir, algumas propostas para pesquisas futuras que possam oferecer melhorias na área de estudo de classificação de patologias das pregas vocais.

- O estudo da utilização de diferentes tamanhos de janela para formação da imagem. A depender da quantidade de amostras do sinal utilizadas para formar a imagem, pode-se obter mais informações relevantes da patologia.
- O estudo de sinais de Pólipo, para verificar o porquê de o método de aumento de dados utilizado ocasionar diminuição na taxa de acerto. Essa verificação pode auxiliar na criação de novos métodos para classificação de sinais de Pólipo.
- O estudo de outros métodos para aumento de dados, como interpolação e *pitch scaling*, para verificar quais deles mais se adequam ao problema de classificação de patologias das pregas vocais.
- O estudo de diferentes métodos para redução da dimensionalidade da imagem da trajetória do espaço de fase, ou a utilização da própria imagem.
- O estudo de novas arquiteturas de CNN, com o objetivo de melhorar as taxas de acerto das classificações envolvendo patologias independente da base de dados utilizada.
- A realização de treinamento e validação com sinais de uma base de dados e testar com sinais de outra base.
- A utilização de sinais das duas bases nos conjuntos de treinamento, validação e testes.

- O Desenvolvimento de um sistema de apoio ao diagnóstico de patologias da fala, implementando o método resultante desta tese.

Além disto, pode-se investigar a generalidade de utilização do método em outros problemas envolvendo séries temporais.

Referências Bibliográficas

ABARBANEL, H. *Analysis of observed chaotic data*. [S.l.]: Springer Science & Business Media, 2012.

AL-NASHERI, A. et al. Voice pathology detection using auto-correlation of different filters bank. In: *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*. [S.l.: s.n.], 2014. p. 50–55.

AL-NASHERI, A. et al. An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification. *Journal of Voice*, v. 31, n. 1, p. 113.e9 – 113.e18, 2017.

AL-NASHERI, A. et al. Investigation of voice pathology detection and classification on different frequency regions using correlation functions. *Journal of Voice*, v. 31, n. 1, p. 3 – 15, 2017.

AL-NASHERI, A. et al. Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions. *IEEE Access*, IEEE, v. 6, p. 6961–6974, 2018.

ALENCAR, V. F. S. *Atributos e Domínios de Interpolação Eficientes em Reconhecimento de Voz Distribuído*. Dissertação (Mestrado) — Pontifícia Universidade Católica, 2005. Dissertação de Mestrado.

ALI, Z. et al. Intra- and inter-database study for arabic, english, and german databases: Do conventional speech features detect voice pathology? *Journal of Voice*, Elsevier, v. 31, n. 3, p. 386.e1 – 386.e8, 2017.

ALI, Z. et al. Detection of voice pathology using fractal dimension in a multiresolution analysis of normal and disordered speech signals. *Journal of Medical Systems*, v. 40, n. 1, p. 20, Nov 2015.

ALI, Z. et al. Automatic voice pathology detection with running speech by using estimation of auditory spectrum and cepstral coefficients based on the all-pole model. *Journal of Voice*, Elsevier, v. 30, n. 6, p. 757.e7 – 757.e19, 2016.

ALI, Z. et al. An intelligent healthcare system for detection and classification to discriminate vocal fold disorders. *Future Generation Computer Systems*, Elsevier, v. 85, p. 19–28, 2018.

BARRY, W. J.; PÜTZER, M. *Saarbrücken Voice Database*. 2014. Disponível em <<http://www.stimmdatenbank.coli.uni-saarland.de/>>. Acessado em 01-09-2018.

- BEHLAU, M. *Voz: o livro do especialista*. [S.l.]: Revinter, 2005.
- BHATTACHARYA, P.; KELLEHER, J. E.; SIEGMUND, T. Role of gradients in vocal fold elastic modulus on phonation. *Journal of Biomechanics*, v. 48, n. 12, p. 3356 – 3363, 2015.
- BOUGHORBEL, S.; JARRAY, F.; EL-ANBARI, M. Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PLOS ONE*, Public Library of Science, v. 12, n. 6, p. 1–17, 06 2017.
- BRANDT, R. R. *Classificação de Vozes Patológicas Utilizando Análise Paramétrica e Não Paramétrica*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2012. Doutorado em Engenharia Elétrica.
- BROADHURST, M. S. Vocal fold paralysis. In: *Practical Laryngology*. [S.l.]: CRC Press, 2015. cap. 6, p. 59–72.
- CARVALHO, R. T. S.; CAVALCANTE, C. C.; CORTEZ, P. C. Wavelet transform and artificial neural networks applied to voice disorders identification. In: *IEEE. Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress on*. [S.l.], 2011. p. 371–376.
- CASTRO, C. L. d.; BRAGA, A. P. Aprendizado supervisionado com conjuntos de dados desbalanceados. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, scielo, v. 22, p. 441 – 466, 10 2011. ISSN 0103-1759.
- CHOI, S. H. et al. Nonlinear dynamic-based analysis of severe dysphonia in patients with vocal fold scar and sulcus vocalis. *Journal of Voice*, Elsevier, v. 26, n. 5, p. 566–576, 2012.
- CIELO, C. A. et al. Lesões organofuncionais do tipo nódulos, pólipos e edema de reinke. *Rev CEFAC*, SciELO Brasil, v. 13, n. 4, p. 735–748, 2011.
- CORDEIRO, H. et al. Voice pathologies identification speech signals, features and classifiers evaluation. In: *IEEE. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2015*. [S.l.], 2015. p. 81–86.
- CORDEIRO, H. et al. Hierarchical classification and system combination for automatically identifying physiological and neuromuscular laryngeal pathologies. *Journal of Voice*, Elsevier, v. 31, n. 3, p. 384.e9 – 384.e14, 2017.
- CORDEIRO, H. T.; FONSECA, J. M.; RIBEIRO, C. M. Lpc spectrum first peak analysis for voice pathology detection. *Procedia Technology*, Elsevier, v. 9, p. 1104–1111, 2013.
- CORDEIRO, H. T.; FONSECA, J. M.; RIBEIRO, C. M. Reinke’s edema and nodules identification in vowels using spectral features and pitch jitter. *Procedia Technology*, Elsevier, v. 17, p. 202–208, 2014.
- CÓSER, C. O. *Laringoscopia*. 2017. Disponível em <<http://www.clinicacoser.com/veja-fotos-de/laringoscopia>>. Acessado em 30-05-2017.
- COSTA, C. d. A. W. *Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2012. Doutorado em Engenharia Elétrica.

- COSTA, C. d. A. W. et al. Pathological voice assessment by recurrence quantification analysis. In: IEEE. *Biosignals and Biorobotics Conference (BRC), 2012 ISSNIP*. [S.l.], 2012. p. 1–6.
- COSTA, S. C. *Análise Acústica, baseada no Modelo Linear de produção da fala, para discriminação de vozes patológicas*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2008. Doutorado em Engenharia Elétrica.
- CRESPO, N. et al. Métodos de evaluación de la voz en docentes una revisión sistemática. *REVISTA CIENTÍFICA SIGNOS FÓNICOS*, v. 3, n. 1, 2018.
- DAHMANI, M.; GUERTI, M. Vocal folds pathologies classification using naïve bayes networks. In: *2017 6th International Conference on Systems and Control (ICSC)*. [S.l.: s.n.], 2017. p. 426–432.
- DELLER, J. R.; HANSEN, J. H.; PROAKIS, J. G. *Discrete-time processing of speech signals*. Ieee New York, NY, USA:, 2000.
- ESPINOLA, S. de B. *Análise Acústica para Classificação de Patologias da Voz empregando Análise de Componentes Principais, Redes Neurais Artificiais e Máquina de Vetores de Suporte*. Dissertação (Mestrado) — Universidade Federal de Campina Grande, 2014. Dissertação de Mestrado em Ciência da Computação.
- FANG, S.-H. et al. Detection of pathological voice using cepstrum vectors: A deep learning approach. *Journal of Voice*, Elsevier, 2018.
- FORERO, L. A. et al. Analysis and classification of voice pathologies using glottal signal parameters. *Journal of Voice*, v. 30, n. 5, p. 549 – 556, 2016.
- FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. *Physical review A, APS*, v. 33, n. 2, p. 1134, 1986.
- FÜRNKRANZ, J. Round robin classification. *Journal of Machine Learning Research*, v. 2, n. Mar, p. 721–747, 2002.
- GARCIA, M. d. M. et al. Avaliação por imagem da paralisia de pregas vocais. *Radiologia Brasileira*, Radiologia Brasileira, v. 42, n. 5, p. 321–326, 2009.
- GLEISER, I. *Caos e complexidade: a evolução do pensamento econômico*. Rio de Janeiro: Campus, 2002.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016.
- GRAY, S. D. Basement membrane zone injury in vocal nodules. In: *Vocal Fold Physiology*. Stockholm, Sweden: Singular, 1991. p. 21–27.
- GREZL, F.; FOUSEK, P. Optimizing bottle-neck features for lvcsr. In: *ICASSP*. [S.l.: s.n.], 2008. v. 8, p. 4729–4732.
- HARIHARAN, M. et al. A hybrid expert system approach for telemonitoring of vocal fold pathology. *Applied Soft Computing*, Elsevier, v. 13, n. 10, p. 4148–4161, 2013.

- HAYKIN, S.; VEEN, B. V. *Signals and systems*. Nova York: Wiley, 2002.
- HEMMERLING, D.; SKALSKI, A.; GAJDA, J. Voice data mining for laryngeal pathology assessment. *Computers in biology and medicine*, Elsevier, v. 69, p. 270–276, 2016.
- JIANG, J.; ZHANG, Y.; MCGILLIGAN, C. Chaos in voice, from modeling to measurement. *Journal of Voice*, v. 20, n. 1, p. 2–17, 2006.
- JOTHILAKSHMI, S. Automatic system to detect the type of voice pathology. *Applied Soft Computing*, Elsevier, v. 21, p. 244–249, 2014.
- KANTZ, H.; SCHREIBER, T. *Nonlinear time series analysis*. [S.l.]: Cambridge university press, 2004. v. 7.
- KAY-ELEMETRICS. *Kay Elemetrics Corp. Disordered Voice Database, Model 4337*. 3. ed. [S.l.]: Massachussetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab, 1994.
- KEELE, S. Guidelines for performing systematic literature reviews in software engineering. In: *Technical report, Ver. 2.3 EBSE Technical Report. EBSE*. [S.l.]: sn, 2007.
- KENNEL, M. B.; BROWN, R.; ABARBANEL, H. D. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A, APS*, v. 45, n. 6, p. 3403, 1992.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.
- KUKHARCHIK, P. et al. Vocal fold pathology detection using modified wavelet-like features and support vector machines. In: IEEE. *Signal Processing Conference, 2007 15th European*. [S.l.], 2007. p. 2214–2218.
- KUMAR, A. J.; DAOUDI, K. Discrimination between pathological voice categories using matching pursuit. In: IEEE. *Bioinspired Intelligence (IWOB), 2015 4th International Work Conference on*. [S.l.], 2015. p. 215–218.
- KUMAR, B. R.; BHAT, J. S.; MUKHI, P. Vowel harmonic amplitude differences in persons with vocal nodules. *Journal of Voice*, Elsevier, v. 25, n. 5, p. 559–561, 2011.
- LEONTITSIS, A. *Chaotic Systems Toolbox*. 2014. Disponível em (<https://www.mathworks.com/matlabcentral/fileexchange/1597-chaotic-systems-toolbox>). Acessado em 25-08-2018.
- LITTLE, M. et al. Nonlinear, biophysically-informed speech pathology detection. In: IEEE. *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. [S.l.], 2006. v. 2, p. II–II.
- LOPES, L. W. et al. Accuracy of acoustic analysis measurements in the evaluation of patients with different laryngeal diagnoses. *Journal of Voice*, Elsevier, v. 31, n. 3, p. 382–e15, 2017.

MARINUS, J. V. M. L. *Estudo de Técnicas para Classificação de Vozes Afetadas por Patologias*. Dissertação (Mestrado) — Universidade Federal de Campina Grande, 2010. Dissertação de Mestrado em Ciência da Computação.

MARINUS, J. V. M. L.; ARAÚJO, J. M. F. R.; GOMES, H. M. Uma revisão sistemática sobre análise acústica de sinais de vozes afetadas por patologias nas pregas vocais. In: *Tecnologias Aplicadas às Ciências da Linguagem*. [S.l.]: Editora UFPB, 2018a. cap. 10, p. 193–218.

MARINUS, J. V. M. L.; ARAÚJO, J. M. F. R.; GOMES, H. M. Reconstructed phase space and convolutional neural networks for classifying voice pathologies. In: *SPRINGER. Iberoamerican Congress on Pattern Recognition*. [S.l.], 2018b. p. 792–801.

MARINUS, J. V. M. L. et al. On the use of cepstral coefficients, multilayer perceptron networks and gaussian mixture models for vocal fold edema diagnosis. In: *IEEE. Biosignals and Biorobotics Conference (BRC), 2013 ISSNIP*. [S.l.], 2013. p. 1–6.

MARKAKI, M.; STYLIANOU, Y. Voice pathology detection and discrimination based on modulation spectral features. *IEEE Transactions on Audio, Speech, and Language Processing*, IEEE, v. 19, n. 7, p. 1938–1948, 2011.

MARTÍNEZ, D. et al. Voice pathology detection on the saarbrücken voice database with calibration and fusion of scores using multifocal toolkit. In: *IberSPEECH*. [S.l.]: Springer, 2012. (Communications in Computer and Information Science, v. 328), p. 99–109.

MARWAN, N. Cross recurrence plot toolbox for matlab. *Reference Manual, Version*, v. 5, 2013.

MATLAB. *version 9.4.0 (R2018a)*. Natick, Massachusetts: The MathWorks Inc., 2018.

MATLAB Deep Learning Toolbox. 2018. The MathWorks, Natick, MA, USA.

MATLAB Statistics and Machine Learning Toolbox. 2018. The MathWorks, Natick, MA, USA.

MEHDIZADEHFAR, V.; ALMASGANJ, F.; TORABINEZHAD, F. Investigation of the effects of speech signal length on vocal disorder sorting done via dynamic pattern modeling. *Journal of Voice*, Elsevier, v. 31, n. 4, p. 515.e1 – 515.e8, 2017.

MOJALY, M. A.; MUHAMMAD, G.; ALSULAIMAN, M. Detection and classification of voice pathology using feature selection. In: *IEEE. Computer Systems and Applications (AICCSA), 2014 IEEE/ACS 11th International Conference on*. [S.l.], 2014. p. 571–577.

MONTEIRO, L. H. A. *Sistemas dinâmicos*. [S.l.]: Editora Livraria da Física, 2002.

MUHAMMAD, G. et al. Edge computing with cloud for voice disorder assessment and treatment. *IEEE Communications Magazine*, v. 56, n. 4, p. 60–65, April 2018.

MUHAMMAD, G. et al. Voice pathology detection using interlaced derivative pattern on glottal source excitation. *Biomedical Signal Processing and Control*, Elsevier, v. 31, p. 156–164, 2017.

- MUHAMMAD, G. et al. Automatic voice disorder classification using vowel formants. In: IEEE. *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. [S.l.], 2011. p. 1–6.
- MUHAMMAD, G. et al. Automatic voice pathology detection and classification using vocal tract area irregularity. *Biocybernetics and Biomedical Engineering*, Elsevier, v. 36, n. 2, p. 309–317, 2016.
- MUHAMMAD, G.; MELHEM, M. Mpeg-7 audio features based voice pathology detection. In: IEEE. *EUROCON, 2013 IEEE*. [S.l.], 2013. p. 1617–1624.
- MUHAMMAD, G.; MELHEM, M. Pathological voice detection and binary classification using mpeg-7 audio features. *Biomedical Signal Processing and Control*, Elsevier, v. 11, p. 1–9, 2014.
- MUSAVI, M. T. et al. On the training of radial basis function classifiers. *Neural networks*, Elsevier, v. 5, n. 4, p. 595–603, 1992.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. [S.l.: s.n.], 2010. p. 807–814.
- NORMANDO, D.; TJÄDERHANE, L.; QUINTÃO, C. C. A. A escolha do teste estatístico-um tutorial em forma de apresentação em powerpoint. *Dental Press J. Orthod*, v. 15, n. 1, p. 101–106, 2010.
- OTT, E. *Chaos in dynamical systems*. [S.l.]: Cambridge university press, 2002.
- PACKARD, N. H. et al. Geometry from a time series. *Physical review letters*, APS, v. 45, n. 9, p. 712, 1980.
- PINHO, P. et al. Classificação de patologias laríngeas por meio de características do espaço de fase reconstruído. *SBrT2016: Anais do*, v. 34, 2016.
- PINTO, J. A. et al. Lesões pré-malignas da laringe: revisão de literatura. *Rev. bras. cir. cabeça pescoço*, v. 41, n. 1, p. 42–7, 2012.
- QUATIERI, T. *Discrete-time Speech Signal Processing: Principles and Practice*. First. Upper Saddle River, NJ, USA: Prentice Hall Press, 2001. ISBN 9780132442138.
- RABINER, L.; JUANG, B.-H. *Fundamentals of Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993. ISBN 0-13-015157-2.
- REMACLE, A. et al. Description of patients consulting the voice clinic regarding gender, age, occupational status, and diagnosis. *European Archives of Oto-Rhino-Laryngology*, v. 274, n. 3, p. 1567–1576, Mar 2017.
- RIBEIRO, C. E. M. *Processamento Digital de Fala*. Lisboa: Instituto Superior de Engenharia de Lisboa, 2003.
- ROSENBLATT. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, p. 386–408, 1958.

SAEEDI, N. E.; ALMASGANJ, F. Wavelet adaptation for automatic voice disorders sorting. *Computers in biology and medicine*, Elsevier, v. 43, n. 6, p. 699–704, 2013.

SALEHI, P. Using patient's speech signal for vocal disorders detection based on lifting scheme. In: IEEE. *Knowledge-Based Engineering and Innovation (KBEI), 2015 2nd International Conference on*. [S.l.], 2015. p. 561–568.

SALMA, C.; ASMA, B.; AICHA, B. Organic voice pathology classification. In: IEEE. *Systems, Signals & Devices (SSD), 2017 14th International Multi-Conference on*. [S.l.], 2017. p. 203–206.

SALMA, C. et al. Recognition of pathological voices. In: IEEE. *Systems, Signals & Devices (SSD), 2014 11th International Multi-Conference on*. [S.l.], 2014. p. 1–6.

SANTOS, M. O. *Emprego do Estimador KSG na Classificação de Patologias Laríngeas por Intermédio da Análise Dinâmica Não-Linear*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2019. Doutorado em Engenharia Elétrica.

SARMENTO, P. *Leucoplasia da Laringe*. 2017. Disponível em <<https://repositoriodavoz.wordpress.com/2015/06/30/leucoplasia-da-laringe/>>. Acessado em 30-05-2017.

SAVI, M. *Dinâmica não-linear e caos*. [S.l.]: Editora E-papers, 2006.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, JSTOR, v. 52, n. 3/4, p. 591–611, 1965.

SHI, W.; DONGKAI, J.; KE, W. Application of modified wavelet features and multi-class sphere svm to pathological vocal detection. In: *2011 Seventh International Conference on Natural Computation*. [S.l.: s.n.], 2011. v. 1, p. 544–548.

SODRÉ, B.; ROSA, M.; DASSIE-LEITE, A. Evaluating the use of neural networks and acoustic measurements to identify laryngeal pathologies. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.: s.n.], 2017. p. 4590–4593.

SOM, P. M.; CURTIN, H. D. *Head and neck imaging*. [S.l.]: Elsevier Health Sciences, 2011.

SOTOMAYOR, C. A. M. *Realce de Voz Aplicado à Verificação Automática de Locutor*. Dissertação (Mestrado) — Instituto Militar de Engenharia, 2003. Dissertação de Mestrado.

SOUZA, E. G. *Caracterização de sistemas dinâmicos através de gráficos de recorrência*. Dissertação (Mestrado) — Universidade Federal do Paraná, 2008.

STUDENT. The probable error of a mean. *Biometrika*, JSTOR, p. 1–25, 1908.

TAKENS, F. Detecting strange attractors in turbulence. In: *Dynamical systems and turbulence, Warwick 1980*. [S.l.]: Springer, 1981. p. 366–381.

TITZE, I. Evidence of chaos in vocal fold vibration. *Vocal Fold Physiology, Frontiers in Basic Science*, Singular Pub., p. 143–188, 1993.

TITZE, I. *Workshop on Acoustic Voice Analysis: Summary Statement*. [S.l.]: National Center for Voice and Speech, 1995.

TUMA, J. et al. Configuração das pregas vestibulares em laringes de pacientes com nódulo vocal. *Revista Brasileira de Otorrinolaringologia*, v. 71, n. 5, p. 576–581, 2005.

ULOZA, V. et al. Categorizing normal and pathological voices: automated and perceptual categorization. *Journal of Voice*, Elsevier, v. 25, n. 6, p. 700–708, 2011.

VAICIUKYNAS, E. et al. Exploring similarity-based classification of larynx disorders from human voice. *Speech Communication*, v. 54, n. 5, p. 601 – 610, 2012.

VAPNIK, V. *Statistical learning theory*. 1998. [S.l.]: Wiley, New York, 1998. v. 3.

VERDOLINI, K. et al. *Guide to vocology*. Iowa City, IA: National Center for Voice and Speech, 1998.

VERDOLINI, K.; ROSEN, C.; BRANSKI, R. *Classification Manual for Voice Disorders-I*. [S.l.]: Taylor & Francis, 2014.

VOICE, F. Y. *Find Your Voice*. 2017. Disponível em <https://findyourvoiceaustralia.com/warming-up-your-voice/>. Acessado em 30-05-2017.

Apêndice A

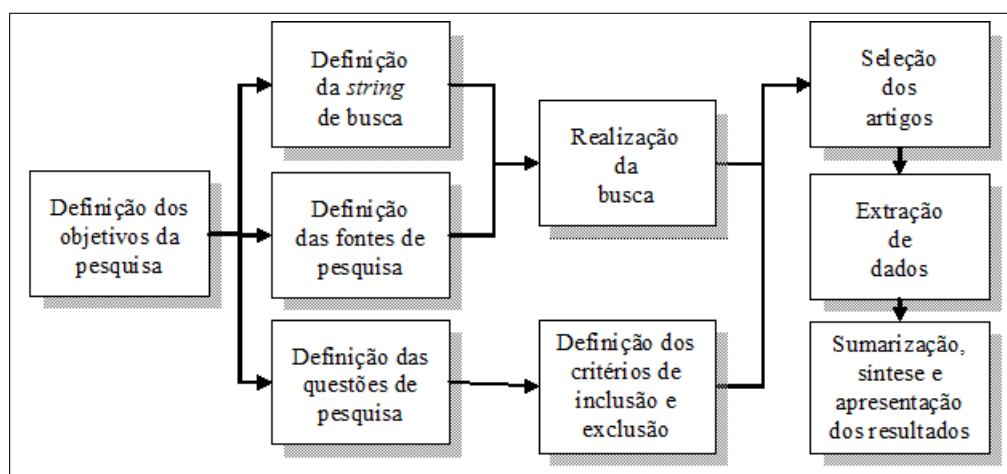
Metodologia da Revisão Sistemática

A estratégia de busca, para a revisão sistemática adotada, é definida por meio de um protocolo de pesquisa, conforme descrito a seguir.

1. Definição dos objetivos da pesquisa.
2. Definição das questões de pesquisa, da string de busca e das fontes de pesquisa.
3. Definição dos critérios de inclusão e exclusão de artigos a partir das questões de pesquisa.
4. Realização da pesquisa a partir nas fontes usando a string de busca. Os artigos são selecionados por meio dos critérios de inclusão.
5. Extração dos dados e sumarização, síntese e apresentação dos resultados do estudo.

As etapas da revisão sistemática são mostradas na Figura A.1, enquanto os artefatos gerados se encontram na Tabela A.1.

Figura A.1: Etapas da revisão sistemática.



Quadro A.1: Artefatos gerados pela revisão sistemática.

| Descrição | Artefato |
|-----------------------------|---|
| Objetivos de pesquisa | <p>Objetivo Geral</p> <ul style="list-style-type: none"> • Selecionar e revisar sistematicamente a literatura publicada entre 2011 e 2018 sobre análise acústica de sinais afetados por patologias nas pregas vocais, com enfoque nas pesquisas que visaram à classificação entre diferentes patologias. <p>Objetivos Específicos</p> <ul style="list-style-type: none"> • Verificar quais são as patologias das pregas vocais mais estudadas atualmente. • Identificar as bases de dados de sinais de vozes com distúrbios utilizadas. • Estudar técnicas para extração de características do sinal de voz afetada por patologias nas pregas vocais. • Identificar a utilização de técnicas para seleção de características. • Estudar técnicas para classificação de sinais de vozes afetadas por patologias nas pregas vocais. |
| Continua na página seguinte | |

Quadro A.1 – Continuação

| Descrição | Artefato |
|-----------------------------|--|
| Questões de pesquisa | <ul style="list-style-type: none"> • Quais são as patologias estudadas? • Quais são as bases de dados utilizadas? • Quais são as técnicas utilizadas para extração de características do sinal de voz? • Quais são as técnicas de seleção de características utilizadas? • Quais são as técnicas de classificação utilizadas? |
| String de busca | (pathology OR disease OR pathological OR disorder) AND (voice OR speech) AND (detection OR classification OR identification OR assessment) AND ("vocal fold" OR larynx) NOT image NOT emotion NOT Parkinson's) |
| Fontes de pesquisa | ACM, IEEE, Science Direct e Springer |
| Continua na página seguinte | |

Quadro A.1 – Continuação

| Descrição | Artefato |
|---|---|
| <p>Critérios de inclusão e exclusão das pesquisas</p> | <p>Critérios de inclusão</p> <ol style="list-style-type: none"> 1. Pesquisas publicadas de janeiro de 2011 até dezembro de 2018. 2. O artigo descreve pesquisas relacionadas à detecção automática de patologias nas pregas vocais por meio de sinais de áudio. 3. O artigo possui relevância com relação ao tema abordado. 4. O artigo apresenta uma solução para o tema abordado nesta revisão sistemática. 5. Os estudos respondem as questões de pesquisa. 6. Os artigos possuem resultados empíricos claros e completos. <p>Critérios de exclusão</p> <ol style="list-style-type: none"> 1. O artigo não possui relevância ao tema abordado. 2. O artigo não faz diferenciação entre patologias. 3. A patologia estudada não ocorre nas pregas vocais (Ex: Mal de Parkinson). |
| <p>Continua na página seguinte</p> | |

Quadro A.1 – Continuação

| Descrição | Artefato |
|---|---|
| Procedimento para seleção dos estudos primários | <ul style="list-style-type: none"> ● Passo 1: Identificar estudos relevantes com base no IEEE, ACM, Science Direct, e Springer. ● Passo 2: Excluir artigos duplicados. ● Passo 3: Excluir estudos com base apenas nos seus resumos (critérios de exclusão 1 e 3). ● Passo 4: Excluir estudos com base na sua introdução, análise experimental e conclusão (critério de exclusão 2). ● Passo 5: Obter os artigos primários e analisar os estudos. |

A partir de pesquisas realizadas em anos anteriores (COSTA, 2008; MARINUS, 2010; BRANDT, 2012; COSTA et al., 2012; ESPINOLA, 2014) e de pesquisas encontradas em uma busca preliminar, foram identificadas palavras-chave e sinônimos para a criação da string de busca, a qual passou por vários refinamentos até se chegar à string de busca descrita na Tabela A.1. Na expressão de busca utilizada, o conectivo lógico de conjunção (AND - e) permitiu restringir (especializar) o escopo da busca e o conectivo lógico de disjunção (OR - ou) foi utilizado para agrupar as palavras sinônimas.

As fontes de pesquisa foram escolhidas por sua abrangência em revistas e conferências de maior impacto na comunidade científica e por permitirem o uso de buscas com expressões lógicas. Além disso, foram selecionados apenas artigos no idioma inglês, pois os principais veículos da área objeto de estudo adotam a língua inglesa, sendo este também o idioma padrão nos engenhos de busca das fontes de pesquisa.

Apêndice B

Descrição das Bases de Dados Utilizadas

Tabela B.1: Descrição das bases de dados MEEI e SVD.

| | Base MEEI | Base SVD |
|--|-----------|----------|
| Sessões de voz normal | 53 | 869 |
| Homens | 21 | 433 |
| Mulheres | 32 | 436 |
| Sujeitos com voz normal | 53 | 851 |
| Homens | 21 | 423 |
| Idade - Intervalo | 26-59 | 15-73 |
| Idade Média | 38,8 | 30,7 |
| Desvio Padrão | 8,3 | 11,5 |
| Mulheres | 32 | 428 |
| Idade - Intervalo | 22-52 | 9-84 |
| Idade Média | 34,2 | 25,1 |
| Desvio Padrão | 7,7 | 11 |
| Sessões com voz afetada por patologia | 657 | 1356 |
| Homens | 170 | 629 |
| Mulheres | 240 | 727 |
| Não identificado | 247 | 0 |
| Sujeitos com voz afetada por patologia | 527 | 1002 |

Continua na página seguinte

Tabela B.1 – Continuação

| | Base MEEI | Base SVD |
|--------------------------------------|------------------------|------------------------|
| Homens | 138 | 454 |
| Idade - Intervalo | 15-85 | 6-89 |
| Idade Média | 49 | 51,1 |
| Desvio Padrão | 17,5 | 16,3 |
| Mulheres | 195 | 548 |
| Idade - Intervalo | 11-93 | 9-94 |
| Idade Média | 46,6 | 47,8 |
| Desvio Padrão | 17,8 | 15,7 |
| Não identificado | 194 | 0 |
| Quantidade de patologias catalogadas | 152 | 71 |
| Sinais contidos em cada sessão | /a/ | /a/ - normal |
| | <i>Rainbow Passage</i> | /i/ - normal |
| | | /u/ - normal |
| | | /a/ - baixo |
| | | /i/ - baixo |
| | | /u/ - baixo |
| | | /a/ - alto |
| | | /i/ - alto |
| | | /u/ - alto |
| | | /a/ - baixo-alto-baixo |
| | | /i/ - baixo-alto-baixo |
| | | /u/ - baixo-alto-baixo |
| | | Sentença |
| Taxa de amostragem | 50000 Hz (Normal) | 50000 Hz |
| | 25000 Hz (Patologia) | |
| Resolução | 16 bits | 16 bits |
| Formato dos arquivos | .NSP | .WAV |

Continua na página seguinte

Tabela B.1 – Continuação

| | Base MEEI | Base SVD |
|-----------|-----------|----------|
| Microfone | | |

Tabela B.2: Descrição do subconjunto MEEI.

| | Homens | Mulheres | Total |
|------------------------|--------|----------|-------|
| Normal | 21 | 32 | 53 |
| Intervalo de Idade | 26-59 | 22-52 | 22-59 |
| Média de Idade | 38,8 | 34,2 | 36 |
| Desvio Padrão de Idade | 8,3 | 7,7 | 8,3 |
| Patologia | 70 | 88 | 158 |
| Intervalo de Idade | 15-77 | 17-85 | 15-85 |
| Média de Idade | 48,4 | 46,3 | 47,2 |
| Desvio Padrão de Idade | 17,1 | 17,1 | 17,1 |
| Paralisia | 33 | 33 | 66 |
| Intervalo de Idade | 15-77 | 19-80 | 15-80 |
| Média de Idade | 53,9 | 53,7 | 53,8 |
| Desvio Padrão de Idade | 18,7 | 16,4 | 17,4 |
| Edema | 10 | 28 | 38 |
| Intervalo de Idade | 23-63 | 17-85 | 17-85 |
| Média de Idade | 39 | 43,7 | 42,5 |
| Desvio Padrão de Idade | 13,5 | 14,8 | 14,4 |
| Queratose | 13 | 7 | 20 |
| Intervalo de Idade | 24-73 | 38-78 | 24-78 |
| Média de Idade | 49,2 | 49 | 49,1 |
| Desvio Padrão de Idade | 13,1 | 14,6 | 13,3 |
| Lesão benigna | 14 | 20 | 34 |
| Intervalo de Idade | 21-76 | 18-74 | 18-76 |

Continua na página seguinte

Tabela B.2 – Continuação

| | Homens | Mulheres | Total |
|------------------------|--------|----------|-------|
| Média de Idade | 41,3 | 37 | 38,8 |
| Desvio Padrão de Idade | 14,4 | 17,7 | 16,3 |
| Nódulo | 0 | 13 | 13 |
| Intervalo de Idade | 0 | 18-45 | 18-45 |
| Média de Idade | 0 | 27,8 | 27,8 |
| Desvio Padrão de Idade | 0 | 9,6 | 9,6 |
| Pólipo | 11 | 5 | 16 |
| Intervalo de Idade | 21-76 | 39-74 | 21-76 |
| Média de Idade | 39,2 | 60,8 | 45,9 |
| Desvio Padrão de Idade | 15,6 | 13,1 | 17,8 |
| Cisto | 3 | 2 | 5 |
| Intervalo de Idade | 47-52 | 26-48 | 26-52 |
| Média de Idade | 49 | 37 | 44,2 |
| Desvio Padrão de Idade | 2,6 | 15,6 | 10,4 |

Tabela B.3: Descrição do subconjunto SVD.

| | Homens | Mulheres | Total |
|------------------------|--------|----------|-------|
| Normal | 250 | 418 | 668 |
| Intervalo de Idade | 16-69 | 9-84 | 9-84 |
| Média de Idade | 30,8 | 24,9 | 27,1 |
| Desvio Padrão de Idade | 11,8 | 10,8 | 11,6 |
| Patologia | 97 | 152 | 249 |
| Intervalo de Idade | 9-81 | 19-79 | 9-81 |
| Média de Idade | 54,6 | 51,4 | 52,7 |
| Desvio Padrão de Idade | 14,5 | 14,2 | 14,4 |
| Paralisia | 42 | 88 | 130 |

Continua na página seguinte

Tabela B.3 – Continuação

| | Homens | Mulheres | Total |
|------------------------|--------|----------|-------|
| Intervalo de Idade | 14-81 | 21-79 | 14-81 |
| Média de Idade | 57,8 | 55,2 | 56,0 |
| Desvio Padrão de Idade | 15,7 | 13,2 | 14,0 |
| Edema | 7 | 33 | 40 |
| Intervalo de Idade | 52-66 | 31-77 | 31-77 |
| Média de Idade | 60,1 | 51,1 | 52,7 |
| Desvio Padrão de Idade | 5,1 | 11,1 | 10,9 |
| Queratose | 23 | 8 | 31 |
| Intervalo de Idade | 32-71 | 26-66 | 26-71 |
| Média de Idade | 52,6 | 45 | 50,6 |
| Desvio Padrão de Idade | 10,7 | 12,9 | 11,6 |
| Lesão benigna | 25 | 23 | 48 |
| Intervalo de Idade | 9-78 | 19-79 | 9-79 |
| Média de Idade | 49,5 | 39,8 | 44,9 |
| Desvio Padrão de Idade | 15,9 | 15,8 | 16,4 |
| Nódulo | 4 | 12 | 16 |
| Intervalo de Idade | 9-64 | 19-57 | 9-64 |
| Média de Idade | 40,3 | 31,3 | 33,5 |
| Desvio Padrão de Idade | 23,1 | 11,1 | 14,6 |
| Pólipo | 20 | 9 | 29 |
| Intervalo de Idade | 29-78 | 28-79 | 28-79 |
| Média de Idade | 50,7 | 50,9 | 50,7 |
| Desvio Padrão de Idade | 14,1 | 15,8 | 14,4 |
| Cisto | 1 | 2 | 3 |
| Intervalo de Idade | 64-64 | 31-52 | 31-64 |
| Média de Idade | 64 | 41,5 | 49 |
| Desvio Padrão de Idade | 0 | 14,8 | 16,7 |

Tabela B.4: Sinais de voz do subconjunto MEEI.

| Normal | Patologia | |
|-------------|-------------|-------------|
| AXH1NAL.NSP | AMC14AN.NSP | MCA07AN.NSP |
| BJB1NAL.NSP | CAC10AN.NSP | MCW21AN.NSP |
| BJV1NAL.NSP | CAK25AN.NSP | MEC06AN.NSP |
| CAD1NAL.NSP | CER16AN.NSP | MEC28AN.NSP |
| CEB1NAL.NSP | CLS31AN.NSP | MFC20AN.NSP |
| DAJ1NAL.NSP | CMA06AN.NSP | MNH14AN.NSP |
| DFP1NAL.NSP | CTB30AN.NSP | MPF25AN.NSP |
| DJG1NAL.NSP | DAP17AN.NSP | MPS09AN.NSP |
| DMA1NAL.NSP | DBF18AN.NSP | MRB30AN.NSP |
| DWS1NAL.NSP | DJF23AN.NSP | MRC20AN.NSP |
| EDC1NAL.NSP | DJP04AN.NSP | MXN24AN.NSP |
| EJC1NAL.NSP | DMC03AN.NSP | NFG08AN.NSP |
| FMB1NAL.NSP | DMG07AN.NSP | NJS06AN.NSP |
| GPC1NAL.NSP | DRC15AN.NSP | NLC08AN.NSP |
| GZZ1NAL.NSP | DSW14AN.NSP | NMB28AN.NSP |
| HBL1NAL.NSP | DVD19AN.NSP | PAT10AN.NSP |
| JAF1NAL.NSP | DXC22AN.NSP | PLW14AN.NSP |
| JAN1NAL.NSP | EAS11AN.NSP | PMF03AN.NSP |
| JAP1NAL.NSP | EEC04AN.NSP | RCC11AN.NSP |
| JEG1NAL.NSP | EED07AN.NSP | RHP12AN.NSP |
| JKR1NAL.NSP | EJH24AN.NSP | RJF14AN.NSP |
| JMC1NAL.NSP | EMP27AN.NSP | RJL28AN.NSP |
| JTH1NAL.NSP | EXE06AN.NSP | RJR15AN.NSP |
| JXC1NAL.NSP | EXI04AN.NSP | RJZ16AN.NSP |
| KAN1NAL.NSP | FMM21AN.NSP | RPJ15AN.NSP |
| LAD1NAL.NSP | FXC12AN.NSP | RTL17AN.NSP |
| LDP1NAL.NSP | GSB11AN.NSP | RWC23AN.NSP |

Continua na página seguinte

Tabela B.4 – Continuação

| Normal | Patologia | |
|-------------|-------------|-------------|
| LLA1NAL.NSP | GXT10AN.NSP | RWR16AN.NSP |
| LMV1NAL.NSP | HLM24AN.NSP | RXP02AN.NSP |
| LMW1NAL.NSP | JAP02AN.NSP | SAM25AN.NSP |
| MAM1NAL.NSP | JBS17AN.NSP | SCC15AN.NSP |
| MAS1NAL.NSP | JCC10AN.NSP | SCH15AN.NSP |
| MCB1NAL.NSP | JFG26AN.NSP | SEC02AN.NSP |
| MFM1NAL.NSP | JFN11AN.NSP | SHC07AN.NSP |
| MJU1NAL.NSP | JJD29AN.NSP | SJD28AN.NSP |
| MXB1NAL.NSP | JLS11AN.NSP | SLC23AN.NSP |
| MXZ1NAL.NSP | JMC18AN.NSP | SLM27AN.NSP |
| NJS1NAL.NSP | JPP27AN.NSP | SMA08AN.NSP |
| OVK1NAL.NSP | JXC21AN.NSP | SRB31AN.NSP |
| PBD1NAL.NSP | JXF11AN.NSP | SWS04AN.NSP |
| PCA1NAL.NSP | KAB03AN.NSP | TDH12AN.NSP |
| RHG1NAL.NSP | KAS09AN.NSP | TLS09AN.NSP |
| RHM1NAL.NSP | KCG25AN.NSP | TPS16AN.NSP |
| RJS1NAL.NSP | KJB19AN.NSP | VAW07AN.NSP |
| SCK1NAL.NSP | KLC06AN.NSP | WCB24AN.NSP |
| SCT1NAL.NSP | KLC09AN.NSP | WDK47AN.NSP |
| SEB1NAL.NSP | KMC27AN.NSP | WST20AN.NSP |
| SIS1NAL.NSP | KMS29AN.NSP | WTG07AN.NSP |
| SLC1NAL.NSP | LAD13AN.NSP | |
| SXV1NAL.NSP | LBA24AN.NSP | |
| TXN1NAL.NSP | LGM01AN.NSP | |
| VMC1NAL.NSP | LJS31AN.NSP | |
| WDK1NAL.NSP | LVD28AN.NSP | |

Tabela B.5: Sinais de voz do subconjunto SVD.

| | Normal | | Patologia |
|------------|--------------|--------------|-------------|
| 1-a_n.wav | 987-a_n.wav | 1523-a_n.wav | 109-a_n.wav |
| 2-a_n.wav | 988-a_n.wav | 1524-a_n.wav | 119-a_n.wav |
| 3-a_n.wav | 989-a_n.wav | 1526-a_n.wav | 120-a_n.wav |
| 4-a_n.wav | 990-a_n.wav | 1527-a_n.wav | 128-a_n.wav |
| 5-a_n.wav | 991-a_n.wav | 1528-a_n.wav | 131-a_n.wav |
| 6-a_n.wav | 992-a_n.wav | 1530-a_n.wav | 146-a_n.wav |
| 7-a_n.wav | 993-a_n.wav | 1531-a_n.wav | 155-a_n.wav |
| 8-a_n.wav | 994-a_n.wav | 1532-a_n.wav | 356-a_n.wav |
| 9-a_n.wav | 995-a_n.wav | 1533-a_n.wav | 448-a_n.wav |
| 11-a_n.wav | 996-a_n.wav | 1534-a_n.wav | 450-a_n.wav |
| 12-a_n.wav | 997-a_n.wav | 1535-a_n.wav | 495-a_n.wav |
| 13-a_n.wav | 998-a_n.wav | 1536-a_n.wav | 497-a_n.wav |
| 14-a_n.wav | 999-a_n.wav | 1537-a_n.wav | 562-a_n.wav |
| 15-a_n.wav | 1000-a_n.wav | 1538-a_n.wav | 564-a_n.wav |
| 16-a_n.wav | 1002-a_n.wav | 1539-a_n.wav | 629-a_n.wav |
| 17-a_n.wav | 1003-a_n.wav | 1540-a_n.wav | 712-a_n.wav |
| 18-a_n.wav | 1004-a_n.wav | 1541-a_n.wav | 716-a_n.wav |
| 19-a_n.wav | 1005-a_n.wav | 1542-a_n.wav | 821-a_n.wav |
| 20-a_n.wav | 1006-a_n.wav | 1544-a_n.wav | 825-a_n.wav |
| 21-a_n.wav | 1007-a_n.wav | 1573-a_n.wav | 864-a_n.wav |
| 23-a_n.wav | 1008-a_n.wav | 1577-a_n.wav | 870-a_n.wav |
| 24-a_n.wav | 1009-a_n.wav | 1582-a_n.wav | 874-a_n.wav |
| 25-a_n.wav | 1010-a_n.wav | 1583-a_n.wav | 875-a_n.wav |
| 26-a_n.wav | 1011-a_n.wav | 1584-a_n.wav | 880-a_n.wav |
| 27-a_n.wav | 1012-a_n.wav | 1585-a_n.wav | 886-a_n.wav |
| 28-a_n.wav | 1013-a_n.wav | 1586-a_n.wav | 889-a_n.wav |
| 29-a_n.wav | 1014-a_n.wav | 1587-a_n.wav | 893-a_n.wav |

Continua na página seguinte

Tabela B.5 – Continuação

| Normal | | | Patologia |
|------------|--------------|--------------|--------------|
| 30-a_n.wav | 1015-a_n.wav | 1598-a_n.wav | 895-a_n.wav |
| 31-a_n.wav | 1016-a_n.wav | 1623-a_n.wav | 901-a_n.wav |
| 32-a_n.wav | 1017-a_n.wav | 1695-a_n.wav | 926-a_n.wav |
| 33-a_n.wav | 1018-a_n.wav | 1696-a_n.wav | 928-a_n.wav |
| 34-a_n.wav | 1019-a_n.wav | 1697-a_n.wav | 929-a_n.wav |
| 35-a_n.wav | 1020-a_n.wav | 1698-a_n.wav | 935-a_n.wav |
| 36-a_n.wav | 1021-a_n.wav | 1699-a_n.wav | 1044-a_n.wav |
| 37-a_n.wav | 1022-a_n.wav | 1700-a_n.wav | 1046-a_n.wav |
| 38-a_n.wav | 1023-a_n.wav | 1701-a_n.wav | 1049-a_n.wav |
| 39-a_n.wav | 1024-a_n.wav | 1702-a_n.wav | 1052-a_n.wav |
| 40-a_n.wav | 1025-a_n.wav | 1703-a_n.wav | 1054-a_n.wav |
| 41-a_n.wav | 1026-a_n.wav | 1704-a_n.wav | 1056-a_n.wav |
| 42-a_n.wav | 1027-a_n.wav | 1705-a_n.wav | 1166-a_n.wav |
| 43-a_n.wav | 1028-a_n.wav | 1706-a_n.wav | 1186-a_n.wav |
| 44-a_n.wav | 1029-a_n.wav | 1707-a_n.wav | 1188-a_n.wav |
| 45-a_n.wav | 1030-a_n.wav | 1708-a_n.wav | 1252-a_n.wav |
| 47-a_n.wav | 1031-a_n.wav | 1709-a_n.wav | 1253-a_n.wav |
| 48-a_n.wav | 1032-a_n.wav | 1710-a_n.wav | 1256-a_n.wav |
| 49-a_n.wav | 1033-a_n.wav | 1711-a_n.wav | 1263-a_n.wav |
| 50-a_n.wav | 1034-a_n.wav | 1712-a_n.wav | 1272-a_n.wav |
| 52-a_n.wav | 1035-a_n.wav | 1724-a_n.wav | 1274-a_n.wav |
| 53-a_n.wav | 1036-a_n.wav | 1725-a_n.wav | 1294-a_n.wav |
| 54-a_n.wav | 1059-a_n.wav | 1726-a_n.wav | 1317-a_n.wav |
| 55-a_n.wav | 1060-a_n.wav | 1727-a_n.wav | 1325-a_n.wav |
| 56-a_n.wav | 1061-a_n.wav | 1728-a_n.wav | 1334-a_n.wav |
| 57-a_n.wav | 1062-a_n.wav | 1729-a_n.wav | 1395-a_n.wav |
| 58-a_n.wav | 1063-a_n.wav | 1730-a_n.wav | 1405-a_n.wav |
| 59-a_n.wav | 1064-a_n.wav | 1731-a_n.wav | 1433-a_n.wav |

Continua na página seguinte

Tabela B.5 – Continuação

| Normal | | | Patologia |
|-------------|--------------|--------------|--------------|
| 60-a_n.wav | 1065-a_n.wav | 1732-a_n.wav | 1444-a_n.wav |
| 61-a_n.wav | 1066-a_n.wav | 1733-a_n.wav | 1453-a_n.wav |
| 62-a_n.wav | 1067-a_n.wav | 1734-a_n.wav | 1462-a_n.wav |
| 63-a_n.wav | 1068-a_n.wav | 1735-a_n.wav | 1464-a_n.wav |
| 67-a_n.wav | 1069-a_n.wav | 1736-a_n.wav | 1489-a_n.wav |
| 69-a_n.wav | 1070-a_n.wav | 1737-a_n.wav | 1525-a_n.wav |
| 72-a_n.wav | 1071-a_n.wav | 1738-a_n.wav | 1549-a_n.wav |
| 74-a_n.wav | 1072-a_n.wav | 1739-a_n.wav | 1563-a_n.wav |
| 80-a_n.wav | 1073-a_n.wav | 1740-a_n.wav | 1572-a_n.wav |
| 81-a_n.wav | 1074-a_n.wav | 1775-a_n.wav | 1576-a_n.wav |
| 82-a_n.wav | 1075-a_n.wav | 1815-a_n.wav | 1599-a_n.wav |
| 83-a_n.wav | 1076-a_n.wav | 1837-a_n.wav | 1607-a_n.wav |
| 85-a_n.wav | 1077-a_n.wav | 1838-a_n.wav | 1613-a_n.wav |
| 88-a_n.wav | 1078-a_n.wav | 1839-a_n.wav | 1616-a_n.wav |
| 89-a_n.wav | 1079-a_n.wav | 1840-a_n.wav | 1617-a_n.wav |
| 93-a_n.wav | 1080-a_n.wav | 1841-a_n.wav | 1620-a_n.wav |
| 94-a_n.wav | 1081-a_n.wav | 1842-a_n.wav | 1621-a_n.wav |
| 95-a_n.wav | 1082-a_n.wav | 1843-a_n.wav | 1624-a_n.wav |
| 97-a_n.wav | 1091-a_n.wav | 1844-a_n.wav | 1626-a_n.wav |
| 98-a_n.wav | 1092-a_n.wav | 1845-a_n.wav | 1631-a_n.wav |
| 102-a_n.wav | 1093-a_n.wav | 1846-a_n.wav | 1632-a_n.wav |
| 103-a_n.wav | 1094-a_n.wav | 1847-a_n.wav | 1639-a_n.wav |
| 104-a_n.wav | 1095-a_n.wav | 1848-a_n.wav | 1643-a_n.wav |
| 112-a_n.wav | 1096-a_n.wav | 1849-a_n.wav | 1656-a_n.wav |
| 113-a_n.wav | 1097-a_n.wav | 1850-a_n.wav | 1658-a_n.wav |
| 114-a_n.wav | 1098-a_n.wav | 1851-a_n.wav | 1662-a_n.wav |
| 115-a_n.wav | 1099-a_n.wav | 1852-a_n.wav | 1665-a_n.wav |
| 116-a_n.wav | 1100-a_n.wav | 1853-a_n.wav | 1667-a_n.wav |

Continua na página seguinte

Tabela B.5 – Continuação

| | Normal | | Patologia |
|-------------|--------------|--------------|--------------|
| 117-a_n.wav | 1101-a_n.wav | 1854-a_n.wav | 1669-a_n.wav |
| 121-a_n.wav | 1102-a_n.wav | 1855-a_n.wav | 1691-a_n.wav |
| 122-a_n.wav | 1103-a_n.wav | 1857-a_n.wav | 1719-a_n.wav |
| 123-a_n.wav | 1104-a_n.wav | 1858-a_n.wav | 1721-a_n.wav |
| 124-a_n.wav | 1105-a_n.wav | 1859-a_n.wav | 1745-a_n.wav |
| 125-a_n.wav | 1106-a_n.wav | 1860-a_n.wav | 1776-a_n.wav |
| 126-a_n.wav | 1107-a_n.wav | 1865-a_n.wav | 1783-a_n.wav |
| 132-a_n.wav | 1108-a_n.wav | 1870-a_n.wav | 1788-a_n.wav |
| 133-a_n.wav | 1109-a_n.wav | 1871-a_n.wav | 1800-a_n.wav |
| 134-a_n.wav | 1110-a_n.wav | 1872-a_n.wav | 1810-a_n.wav |
| 135-a_n.wav | 1111-a_n.wav | 1873-a_n.wav | 1818-a_n.wav |
| 136-a_n.wav | 1112-a_n.wav | 1874-a_n.wav | 1829-a_n.wav |
| 137-a_n.wav | 1121-a_n.wav | 1875-a_n.wav | 1831-a_n.wav |
| 145-a_n.wav | 1122-a_n.wav | 1876-a_n.wav | 1835-a_n.wav |
| 158-a_n.wav | 1123-a_n.wav | 1877-a_n.wav | 1836-a_n.wav |
| 675-a_n.wav | 1124-a_n.wav | 1878-a_n.wav | 1863-a_n.wav |
| 676-a_n.wav | 1125-a_n.wav | 1879-a_n.wav | 1866-a_n.wav |
| 677-a_n.wav | 1126-a_n.wav | 1880-a_n.wav | 1889-a_n.wav |
| 678-a_n.wav | 1127-a_n.wav | 1881-a_n.wav | 1913-a_n.wav |
| 679-a_n.wav | 1128-a_n.wav | 1882-a_n.wav | 1927-a_n.wav |
| 680-a_n.wav | 1129-a_n.wav | 1883-a_n.wav | 1931-a_n.wav |
| 681-a_n.wav | 1130-a_n.wav | 1884-a_n.wav | 1940-a_n.wav |
| 682-a_n.wav | 1131-a_n.wav | 1885-a_n.wav | 1957-a_n.wav |
| 683-a_n.wav | 1132-a_n.wav | 1886-a_n.wav | 1987-a_n.wav |
| 684-a_n.wav | 1133-a_n.wav | 1915-a_n.wav | 2014-a_n.wav |
| 685-a_n.wav | 1134-a_n.wav | 1916-a_n.wav | 2056-a_n.wav |
| 686-a_n.wav | 1135-a_n.wav | 1917-a_n.wav | 2075-a_n.wav |
| 687-a_n.wav | 1136-a_n.wav | 1918-a_n.wav | 2078-a_n.wav |

Continua na página seguinte

Tabela B.5 – Continuação

| Normal | | | Patologia |
|-------------|--------------|--------------|--------------|
| 688-a_n.wav | 1137-a_n.wav | 1919-a_n.wav | 2079-a_n.wav |
| 689-a_n.wav | 1138-a_n.wav | 1920-a_n.wav | 2090-a_n.wav |
| 690-a_n.wav | 1139-a_n.wav | 1921-a_n.wav | 2096-a_n.wav |
| 691-a_n.wav | 1140-a_n.wav | 1922-a_n.wav | 2099-a_n.wav |
| 694-a_n.wav | 1141-a_n.wav | 1923-a_n.wav | 2100-a_n.wav |
| 695-a_n.wav | 1142-a_n.wav | 1924-a_n.wav | 2101-a_n.wav |
| 696-a_n.wav | 1143-a_n.wav | 1925-a_n.wav | 2112-a_n.wav |
| 697-a_n.wav | 1144-a_n.wav | 1926-a_n.wav | 2118-a_n.wav |
| 698-a_n.wav | 1145-a_n.wav | 1947-a_n.wav | 2144-a_n.wav |
| 699-a_n.wav | 1146-a_n.wav | 1995-a_n.wav | 2148-a_n.wav |
| 700-a_n.wav | 1147-a_n.wav | 1996-a_n.wav | 2152-a_n.wav |
| 701-a_n.wav | 1148-a_n.wav | 2018-a_n.wav | 2153-a_n.wav |
| 702-a_n.wav | 1149-a_n.wav | 2036-a_n.wav | 2157-a_n.wav |
| 703-a_n.wav | 1150-a_n.wav | 2037-a_n.wav | 2193-a_n.wav |
| 704-a_n.wav | 1151-a_n.wav | 2038-a_n.wav | 2211-a_n.wav |
| 705-a_n.wav | 1153-a_n.wav | 2039-a_n.wav | 2216-a_n.wav |
| 706-a_n.wav | 1155-a_n.wav | 2040-a_n.wav | 2218-a_n.wav |
| 707-a_n.wav | 1167-a_n.wav | 2041-a_n.wav | 2237-a_n.wav |
| 708-a_n.wav | 1168-a_n.wav | 2042-a_n.wav | 2242-a_n.wav |
| 709-a_n.wav | 1169-a_n.wav | 2043-a_n.wav | 2247-a_n.wav |
| 710-a_n.wav | 1170-a_n.wav | 2044-a_n.wav | 2274-a_n.wav |
| 711-a_n.wav | 1172-a_n.wav | 2045-a_n.wav | 2333-a_n.wav |
| 732-a_n.wav | 1173-a_n.wav | 2046-a_n.wav | 2341-a_n.wav |
| 733-a_n.wav | 1175-a_n.wav | 2047-a_n.wav | 2344-a_n.wav |
| 734-a_n.wav | 1176-a_n.wav | 2048-a_n.wav | 2345-a_n.wav |
| 735-a_n.wav | 1177-a_n.wav | 2049-a_n.wav | 2346-a_n.wav |
| 736-a_n.wav | 1178-a_n.wav | 2050-a_n.wav | 2370-a_n.wav |
| 737-a_n.wav | 1180-a_n.wav | 2051-a_n.wav | 2375-a_n.wav |

Continua na página seguinte

Tabela B.5 – Continuação

| | Normal | | Patologia |
|-------------|--------------|--------------|--------------|
| 738-a_n.wav | 1181-a_n.wav | 2052-a_n.wav | 2381-a_n.wav |
| 739-a_n.wav | 1183-a_n.wav | 2053-a_n.wav | 2411-a_n.wav |
| 740-a_n.wav | 1185-a_n.wav | 2054-a_n.wav | 2414-a_n.wav |
| 743-a_n.wav | 1207-a_n.wav | 2084-a_n.wav | 2426-a_n.wav |
| 744-a_n.wav | 1208-a_n.wav | 2108-a_n.wav | 2430-a_n.wav |
| 745-a_n.wav | 1209-a_n.wav | 2142-a_n.wav | 2443-a_n.wav |
| 746-a_n.wav | 1210-a_n.wav | 2162-a_n.wav | 2450-a_n.wav |
| 747-a_n.wav | 1211-a_n.wav | 2163-a_n.wav | 2460-a_n.wav |
| 782-a_n.wav | 1212-a_n.wav | 2164-a_n.wav | 2465-a_n.wav |
| 805-a_n.wav | 1214-a_n.wav | 2165-a_n.wav | 2466-a_n.wav |
| 806-a_n.wav | 1215-a_n.wav | 2166-a_n.wav | 2469-a_n.wav |
| 810-a_n.wav | 1216-a_n.wav | 2167-a_n.wav | 2476-a_n.wav |
| 811-a_n.wav | 1278-a_n.wav | 2168-a_n.wav | 2484-a_n.wav |
| 812-a_n.wav | 1308-a_n.wav | 2169-a_n.wav | 2501-a_n.wav |
| 813-a_n.wav | 1320-a_n.wav | 2170-a_n.wav | 2507-a_n.wav |
| 831-a_n.wav | 1336-a_n.wav | 2171-a_n.wav | 2508-a_n.wav |
| 834-a_n.wav | 1342-a_n.wav | 2172-a_n.wav | 2538-a_n.wav |
| 838-a_n.wav | 1343-a_n.wav | 2173-a_n.wav | 2557-a_n.wav |
| 841-a_n.wav | 1344-a_n.wav | 2174-a_n.wav | 2570-a_n.wav |
| 842-a_n.wav | 1345-a_n.wav | 2176-a_n.wav | 2572-a_n.wav |
| 843-a_n.wav | 1346-a_n.wav | 2177-a_n.wav | 2581-a_n.wav |
| 860-a_n.wav | 1347-a_n.wav | 2178-a_n.wav | 2587-a_n.wav |
| 861-a_n.wav | 1348-a_n.wav | 2179-a_n.wav | 2588-a_n.wav |
| 862-a_n.wav | 1349-a_n.wav | 2180-a_n.wav | 2591-a_n.wav |
| 865-a_n.wav | 1350-a_n.wav | 2181-a_n.wav | 2595-a_n.wav |
| 867-a_n.wav | 1351-a_n.wav | 2195-a_n.wav | |
| 941-a_n.wav | 1352-a_n.wav | 2196-a_n.wav | |
| 942-a_n.wav | 1353-a_n.wav | 2197-a_n.wav | |

Continua na página seguinte

Tabela B.5 – Continuação

| Normal | | | Patologia |
|-------------|--------------|--------------|-----------|
| 943-a_n.wav | 1354-a_n.wav | 2198-a_n.wav | |
| 944-a_n.wav | 1355-a_n.wav | 2199-a_n.wav | |
| 945-a_n.wav | 1356-a_n.wav | 2201-a_n.wav | |
| 946-a_n.wav | 1357-a_n.wav | 2202-a_n.wav | |
| 947-a_n.wav | 1358-a_n.wav | 2203-a_n.wav | |
| 948-a_n.wav | 1360-a_n.wav | 2204-a_n.wav | |
| 949-a_n.wav | 1361-a_n.wav | 2205-a_n.wav | |
| 950-a_n.wav | 1362-a_n.wav | 2206-a_n.wav | |
| 951-a_n.wav | 1363-a_n.wav | 2207-a_n.wav | |
| 952-a_n.wav | 1364-a_n.wav | 2208-a_n.wav | |
| 953-a_n.wav | 1365-a_n.wav | 2209-a_n.wav | |
| 954-a_n.wav | 1366-a_n.wav | 2210-a_n.wav | |
| 955-a_n.wav | 1367-a_n.wav | 2248-a_n.wav | |
| 956-a_n.wav | 1368-a_n.wav | 2249-a_n.wav | |
| 957-a_n.wav | 1369-a_n.wav | 2250-a_n.wav | |
| 958-a_n.wav | 1370-a_n.wav | 2251-a_n.wav | |
| 959-a_n.wav | 1371-a_n.wav | 2252-a_n.wav | |
| 960-a_n.wav | 1372-a_n.wav | 2253-a_n.wav | |
| 961-a_n.wav | 1373-a_n.wav | 2254-a_n.wav | |
| 962-a_n.wav | 1374-a_n.wav | 2255-a_n.wav | |
| 963-a_n.wav | 1375-a_n.wav | 2256-a_n.wav | |
| 964-a_n.wav | 1377-a_n.wav | 2257-a_n.wav | |
| 965-a_n.wav | 1427-a_n.wav | 2258-a_n.wav | |
| 967-a_n.wav | 1460-a_n.wav | 2259-a_n.wav | |
| 968-a_n.wav | 1504-a_n.wav | 2260-a_n.wav | |
| 969-a_n.wav | 1505-a_n.wav | 2261-a_n.wav | |
| 970-a_n.wav | 1506-a_n.wav | 2262-a_n.wav | |
| 971-a_n.wav | 1507-a_n.wav | 2263-a_n.wav | |

Continua na página seguinte

Tabela B.5 – Continuação

| Normal | | | Patologia |
|-------------|--------------|--------------|-----------|
| 972-a_n.wav | 1508-a_n.wav | 2264-a_n.wav | |
| 973-a_n.wav | 1509-a_n.wav | 2278-a_n.wav | |
| 974-a_n.wav | 1510-a_n.wav | 2279-a_n.wav | |
| 975-a_n.wav | 1511-a_n.wav | 2280-a_n.wav | |
| 976-a_n.wav | 1512-a_n.wav | 2281-a_n.wav | |
| 977-a_n.wav | 1513-a_n.wav | 2282-a_n.wav | |
| 978-a_n.wav | 1514-a_n.wav | 2283-a_n.wav | |
| 979-a_n.wav | 1515-a_n.wav | 2284-a_n.wav | |
| 980-a_n.wav | 1516-a_n.wav | 2285-a_n.wav | |
| 981-a_n.wav | 1517-a_n.wav | 2286-a_n.wav | |
| 982-a_n.wav | 1518-a_n.wav | 2462-a_n.wav | |
| 983-a_n.wav | 1519-a_n.wav | 2478-a_n.wav | |
| 984-a_n.wav | 1520-a_n.wav | 2532-a_n.wav | |
| 985-a_n.wav | 1521-a_n.wav | 2561-a_n.wav | |
| 986-a_n.wav | 1522-a_n.wav | 2576-a_n.wav | |

Apêndice C

Códigos-fonte

Código-Fonte C.1: Código-fonte da função que obtém o primeiro mínimo da curva de informação mútua média

```
function T = getTau(sinal)

T = 1;
Aux = mi(sinal, 40, 'silent');
Aux2 = reshape(Aux(1,1,:), 41, 1);
for m = 1 : 40
    if Aux2(m) <= Aux2(m+1)
        T = m-1; %o retorno e m vetor de indices 0 a 40, mas o matlab
                considera como sendo 1 a 41
        break;
    end
end
end
```

Código-Fonte C.2: Código-fonte da função que realiza o Time Stretching para o aumento de dados

```
function y = timeStretching(sinal, TS)
%TS = fator de time stretching
if TS == 1
    y = sinal;
else
    i = 1;
    while int32(i*TS) <= length(sinal)
```

```

        y(i) = sinal(int32(i*TS));
        i = i + 1;
    end
end

```

Código-Fonte C.3: Código-fonte da função que implementa o método da contagem de caixas

```

function IMG = criaImagemCaixas(Y, TAM, contaCaixa)
%Y = Espaço de fase reconstruído do sinal
%TAM = dimensão da caixa TAM x TAM
%contaCaixa = %caso contaCaixa seja falso, ele cria uma imagem "binária",
    ou preto (caixa não visitada) ou
%branco (caixa visitada)

plot(Y(:,1),Y(:,2));
axis off;

F = getframe(gcf);
I = rgb2gray(F.cdata);

if TAM > 1
    i=1;
    cont1 = 1;
    while i+TAM-1 <= size(I,1)
        j = 1;
        cont2 = 1;
        while j+TAM-1 <= size(I,2)
            %Os valores dos pixels da imagem gerada lá em cima é 240
            para
            %lugar não visitado e 29 para lugar visitado
            IMG(cont1, cont2) = sum(sum(I(i:i+TAM-1,j:j+TAM-1) < 200));
            cont2 = cont2 + 1;
            j = j + TAM;
        end
        cont1 = cont1 + 1;
        i = i + TAM;
    end
end

```

```
    IMG = uint8(IMG);  
    if ~contaCaixa  
        IMG(IMG > 0) = 255;  
    end  
else  
    IMG = uint8(I < 200);  
    IMG(IMG == 1) = uint8(255);  
end
```

Apêndice D

Resumo dos Resultados

Tabela D.1: Melhores Resultados da Classificação Paralisia *versus* Edema.

| Base de Dados | TS | Classificador | Caixa | Janela | Paralisia (%) | Edema (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 15 | 10 Ciclos | 83,53 ± 4,48 | 65,45 ± 3,80 | 79,11 ± 3,74 | 0,47 |
| MEEI | Sim | CNN | 15 | 10 Ciclos | 86,47 ± 5,83 | 63,64 ± 8,50 | 80,89 ± 3,08 | 0,49 |
| MEEI | Não | B32+SVM | 10 | 20 ms | 80,00 ± 5,08 | 67,27 ± 3,44 | 76,89 ± 2,48 | 0,44 |
| MEEI | Sim | B16+SVM | 15 | 10 Ciclos | 88,24 ± 3,77 | 72,73 ± 5,62 | 84,44 ± 2,77 | 0,59 |
| SVD | Não | CNN | 10 | 10 Ciclos | 61,40 ± 5,34 | 60,00 ± 9,50 | 61,27 ± 4,64 | 0,13 |
| SVD | Sim | CNN | 10 | 20 ms | 69,82 ± 11,02 | 65,00 ± 11,56 | 69,37 ± 9,05 | 0,22 |
| SVD | Não | B16+SVM | 15 | 10 Ciclos | 76,84 ± 3,54 | 65,00 ± 6,32 | 75,71 ± 2,90 | 0,28 |
| SVD | Sim | B32+SVM | 15 | 10 Ciclos | 78,95 ± 3,32 | 73,33 ± 5,94 | 78,41 ± 2,56 | 0,35 |

Tabela D.2: Melhores Resultados da Classificação Paralisia *versus* Queratose.

| Base de Dados | TS | Classificador | Caixa | Janela | Paralisia (%) | Queratose (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 15 | 10 Ciclos | 77,92 ± 5,52 | 77,14 ± 10,85 | 77,82 ± 4,23 | 0,40 |
| MEEI | Sim | CNN | 15 | 10 Ciclos | 79,17 ± 4,32 | 82,86 ± 7,79 | 79,64 ± 2,77 | 0,46 |
| MEEI | Não | B16+SVM | 10 | 10 Ciclos | 79,17 ± 3,34 | 80,00 ± 8,14 | 79,27 ± 2,95 | 0,44 |
| MEEI | Sim | B32+SVM | 10 | 10 Ciclos | 80,83 ± 4,46 | 94,29 ± 3,91 | 82,55 ± 4,07 | 0,55 |
| SVD | Não | CNN | 15 | 10 Ciclos | 78,03 ± 7,74 | 55,00 ± 7,13 | 76,71 ± 6,98 | 0,18 |
| SVD | Sim | CNN | 10 | 10 Ciclos | 77,27 ± 2,51 | 50,00 ± 4,42 | 75,71 ± 2,34 | 0,15 |
| SVD | Não | B32+SVM | 15 | 20 ms | 62,58 ± 4,96 | 77,50 ± 5,23 | 63,43 ± 4,51 | 0,19 |
| SVD | Sim | B16+SVM | 15 | 20 ms | 77,73 ± 7,45 | 62,50 ± 7,13 | 76,86 ± 6,76 | 0,22 |

Tabela D.3: Melhores Resultados da Classificação Paralisia *versus* Pólipo.

| Base de Dados | TS | Classificador | Caixa | Janela | Paralisia (%) | Pólipo (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 10 | 10 Ciclos | 83,85 ± 5,72 | 72,00 ± 11,37 | 82,81 ± 4,35 | 0,39 |
| MEEI | Sim | CNN | 15 | 10 Ciclos | 85,38 ± 5,70 | 72,00 ± 11,37 | 84,21 ± 5,42 | 0,41 |
| MEEI | Não | B32+SVM | 15 | 10 Ciclos | 86,54 ± 2,77 | 68,00 ± 11,40 | 84,91 ± 2,66 | 0,40 |
| MEEI | Sim | B32+SVM | 10 | 10 Ciclos | 85,38 ± 5,70 | 60,00 ± 20,00 | 83,16 ± 5,13 | 0,33 |
| SVD | Não | CNN | 10 | 10 Ciclos | 62,20 ± 7,08 | 68,89 ± 11,11 | 62,65 ± 5,93 | 0,16 |
| SVD | Sim | CNN | 15 | 10 Ciclos | 60,31 ± 6,46 | 64,44 ± 8,24 | 60,59 ± 6,06 | 0,12 |
| SVD | Não | B16+SVM | 15 | 10 Ciclos | 63,46 ± 10,46 | 73,33 ± 18,68 | 64,12 ± 8,67 | 0,19 |
| SVD | Sim | B16+SVM | 15 | 10 Ciclos | 75,28 ± 7,14 | 68,89 ± 12,04 | 74,85 ± 5,98 | 0,25 |

Tabela D.4: Melhores Resultados da Classificação Paralisia *versus* Nódulo.

| Base de Dados | TS | Classificador | Caixa | Janela | Paralisia (%) | Nódulo (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 10 | 20 ms | 93,96 ± 1,65 | 76,00 ± 11,00 | 92,41 ± 1,35 | 0,60 |
| MEEI | Sim | CNN | 10 | 10 Ciclos | 95,85 ± 3,24 | 80,00 ± 4,47 | 94,48 ± 2,87 | 0,69 |
| MEEI | Não | B32+SVM | 10 | 20 ms | 92,45 ± 1,26 | 76,00 ± 11,00 | 91,03 ± 5,34 | 0,56 |
| MEEI | Sim | B32+SVM | 10 | 20 ms | 93,58 ± 2,44 | 80,00 ± 5,48 | 92,41 ± 2,48 | 0,62 |
| SVD | Não | CNN | 10 | 10 Ciclos | 65,83 ± 4,65 | 84,00 ± 8,37 | 66,44 ± 4,41 | 0,19 |
| SVD | Sim | CNN | 15 | 20 ms | 68,75 ± 4,70 | 80,00 ± 7,07 | 69,13 ± 4,70 | 0,19 |
| SVD | Não | B16+SVM | 10 | 10 Ciclos | 67,78 ± 6,27 | 96,00 ± 4,47 | 68,72 ± 6,05 | 0,24 |
| SVD | Sim | B16+SVM | 10 | 10 Ciclos | 68,47 ± 7,21 | 96,00 ± 5,48 | 69,40 ± 6,89 | 0,25 |

Tabela D.5: Melhores Resultados da Classificação Paralisia *versus* Lesão Benigna.

| Base de Dados | TS | Classificador | Caixa | Janela | Paralisia (%) | Lesão (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 15 | 10 Ciclos | 77,71 ± 6,58 | 67,27 ± 9,96 | 75,22 ± 3,65 | 0,41 |
| MEEI | Sim | CNN | 10 | 10 Ciclos | 80,57 ± 9,23 | 56,36 ± 7,47 | 74,78 ± 7,55 | 0,35 |
| MEEI | Não | B32+SVM | 15 | 10 Ciclos | 80,00 ± 6,08 | 72,73 ± 6,62 | 78,26 ± 4,81 | 0,48 |
| MEEI | Sim | B16+SVM | 10 | 20 ms | 82,86 ± 7,95 | 61,82 ± 5,77 | 77,83 ± 6,08 | 0,43 |
| SVD | Não | CNN | 15 | 20 ms | 76,43 ± 9,50 | 51,43 ± 13,97 | 73,65 ± 6,91 | 0,20 |
| SVD | Sim | CNN | 15 | 10 Ciclos | 72,14 ± 7,86 | 55,71 ± 11,41 | 70,32 ± 6,19 | 0,19 |
| SVD | Não | B32+SVM | 10 | 10 Ciclos | 63,84 ± 2,09 | 68,57 ± 4,42 | 64,37 ± 1,93 | 0,21 |
| SVD | Sim | B16+SVM | 15 | 10 Ciclos | 68,45 ± 2,61 | 64,29 ± 3,57 | 67,99 ± 1,99 | 0,22 |

Tabela D.6: Melhores Resultados da Classificação Edema *versus* Queratose.

| Base de Dados | TS | Classificador | Caixa | Janela | Edema (%) | Queratose (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|--------------|---------------|--------------|------|
| MEEI | Não | CNN | 15 | 10 Ciclos | 83,20 ± 3,77 | 77,14 ± 8,14 | 81,88 ± 3,90 | 0,54 |
| MEEI | Sim | CNN | 10 | 10 Ciclos | 88,00 ± 5,33 | 74,29 ± 14,32 | 85,00 ± 4,37 | 0,59 |
| MEEI | Não | B32+SVM | 15 | 20 ms | 84,80 ± 3,32 | 77,14 ± 9,31 | 83,13 ± 3,15 | 0,57 |
| MEEI | Sim | B32+SVM | 10 | 20 ms | 88,00 ± 3,06 | 80,00 ± 4,59 | 86,25 ± 3,68 | 0,63 |
| SVD | Não | CNN | 15 | 20 ms | 56,67 ± 6,01 | 87,50 ± 6,25 | 63,16 ± 3,72 | 0,36 |
| SVD | Sim | CNN | 15 | 20 ms | 62,00 ± 7,23 | 77,50 ± 8,15 | 65,26 ± 5,21 | 0,32 |
| SVD | Não | B32+SVM | 10 | 20 ms | 73,33 ± 3,33 | 75,00 ± 4,42 | 73,68 ± 3,09 | 0,41 |
| SVD | Sim | B32+SVM | 10 | 20 ms | 73,33 ± 5,60 | 87,50 ± 8,15 | 76,32 ± 3,30 | 0,51 |

Tabela D.7: Melhores Resultados da Classificação Edema *versus* Pólipo.

| Base de Dados | TS | Classificador | Caixa | Janela | Edema (%) | Pólipo (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 15 | 10 Ciclos | 87,59 ± 6,00 | 84,00 ± 8,44 | 87,06 ± 4,16 | 0,60 |
| MEEI | Sim | CNN | 10 | 10 Ciclos | 77,93 ± 4,45 | 80,00 ± 10,00 | 78,24 ± 3,52 | 0,45 |
| MEEI | Não | B32+SVM | 15 | 10 Ciclos | 89,66 ± 5,67 | 84,00 ± 8,94 | 88,82 ± 5,34 | 0,64 |
| MEEI | Sim | B32+SVM | 10 | 10 Ciclos | 84,83 ± 4,82 | 76,00 ± 10,95 | 83,53 ± 4,72 | 0,50 |
| SVD | Não | CNN | 10 | 10 Ciclos | 73,60 ± 4,47 | 77,78 ± 11,79 | 74,71 ± 2,33 | 0,46 |
| SVD | Sim | CNN | 10 | 10 Ciclos | 65,60 ± 11,19 | 66,67 ± 8,78 | 65,88 ± 6,71 | 0,29 |
| SVD | Não | B32+SVM | 10 | 20 ms | 76,00 ± 4,56 | 82,22 ± 8,43 | 77,65 ± 1,32 | 0,53 |
| SVD | Sim | B32+SVM | 10 | 20 ms | 75,20 ± 4,77 | 80,00 ± 4,65 | 76,47 ± 3,75 | 0,50 |

Tabela D.8: Melhores Resultados da Classificação Edema *versus* Nódulo.

| Base de Dados | TS | Classificador | Caixa | Janela | Edema (%) | Nódulo (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|---------------|------|
| MEEI | Não | CNN | 15 | 10 Ciclos | 66,90 ± 5,43 | 75,00 ± 15,26 | 67,88 ± 4,17 | 0,28 |
| MEEI | Sim | CNN | 15 | 10 Ciclos | 72,41 ± 5,53 | 75,00 ± 10,49 | 72,73 ± 3,84 | 0,33 |
| MEEI | Não | B16+SVM | 10 | 10 Ciclos | 79,31 ± 6,57 | 55,00 ± 16,30 | 76,36 ± 4,20 | 0,26 |
| MEEI | Sim | B16+SVM | 10 | 10 Ciclos | 65,52 ± 13,05 | 75,00 ± 12,50 | 66,67 ± 10,48 | 0,27 |
| SVD | Não | CNN | 10 | 10 Ciclos | 63,90 ± 5,62 | 75,00 ± 15,31 | 64,89 ± 4,40 | 0,23 |
| SVD | Sim | CNN | 15 | 10 Ciclos | 68,29 ± 4,36 | 65,00 ± 11,18 | 68,00 ± 3,33 | 0,20 |
| SVD | Não | B32+SVM | 10 | 20 ms | 78,54 ± 2,91 | 60,00 ± 14,25 | 76,89 ± 1,69 | 0,25 |
| SVD | Sim | B32+SVM | 15 | 20 ms | 79,02 ± 6,84 | 67,50 ± 16,77 | 78,00 ± 5,53 | 0,31 |

Tabela D.9: Melhores Resultados da Classificação Edema *versus* Lesão Benigna.

| Base de Dados | TS | Classificador | Caixa | Janela | Edema (%) | Lesão (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 15 | 20 ms | 65,45 ± 5,93 | 70,00 ± 11,73 | 67,62 ± 3,91 | 0,35 |
| MEEI | Sim | CNN | 15 | 20 ms | 60,00 ± 6,10 | 62,00 ± 11,94 | 60,95 ± 6,61 | 0,22 |
| MEEI | Não | B32+SVM | 10 | 10 Ciclos | 67,27 ± 5,85 | 76,00 ± 12,13 | 71,43 ± 3,39 | 0,43 |
| MEEI | Sim | B32+SVM | 15 | 10 Ciclos | 69,09 ± 7,10 | 80,00 ± 11,32 | 74,29 ± 3,16 | 0,49 |
| SVD | Não | CNN | 10 | 10 Ciclos | 64,62 ± 11,08 | 74,12 ± 7,08 | 70,00 ± 1,67 | 0,39 |
| SVD | Sim | CNN | 10 | 10 Ciclos | 67,69 ± 10,39 | 77,65 ± 7,38 | 73,33 ± 4,47 | 0,46 |
| SVD | Não | B32+SVM | 15 | 10 Ciclos | 64,62 ± 7,89 | 74,12 ± 7,82 | 70,00 ± 2,62 | 0,39 |
| SVD | Sim | B16+SVM | 15 | 10 Ciclos | 50,77 ± 7,50 | 85,88 ± 5,34 | 70,67 ± 3,61 | 0,40 |

Tabela D.10: Melhores Resultados da Classificação Queratose *versus* Pólipo.

| Base de Dados | TS | Classificador | Caixa | Janela | Queratose (%) | Pólipo (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 15 | 10 Ciclos | 70,91 ± 9,85 | 84,00 ± 11,00 | 75,00 ± 6,65 | 0,51 |
| MEEI | Sim | CNN | 15 | 10 Ciclos | 81,82 ± 7,19 | 76,00 ± 13,00 | 80,00 ± 6,00 | 0,56 |
| MEEI | Não | B32+SVM | 15 | 10 Ciclos | 81,82 ± 5,57 | 84,00 ± 11,00 | 82,50 ± 6,25 | 0,63 |
| MEEI | Sim | B16+SVM | 15 | 10 Ciclos | 85,45 ± 2,49 | 80,00 ± 15,00 | 83,75 ± 4,64 | 0,64 |
| SVD | Não | CNN | 10 | 10 Ciclos | 62,50 ± 9,88 | 67,14 ± 10,29 | 65,45 ± 7,11 | 0,29 |
| SVD | Sim | CNN | 15 | 10 Ciclos | 55,00 ± 8,39 | 54,29 ± 11,46 | 54,55 ± 5,33 | 0,09 |
| SVD | Não | B16+SVM | 10 | 10 Ciclos | 75,00 ± 7,65 | 64,29 ± 8,14 | 68,18 ± 5,43 | 0,38 |
| SVD | Sim | B16+SVM | 10 | 10 Ciclos | 67,50 ± 16,30 | 57,14 ± 6,68 | 60,91 ± 2,59 | 0,24 |

Tabela D.11: Melhores Resultados da Classificação Queratose *versus* Nódulo.

| Base de Dados | TS | Classificador | Caixa | Janela | Queratose (%) | Nódulo (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|---------------|--------------|------|
| MEEI | Não | CNN | 15 | 10 Ciclos | 80,00 ± 7,60 | 100,00 ± 0,00 | 85,88 ± 5,26 | 0,74 |
| MEEI | Sim | CNN | 10 | 10 Ciclos | 83,33 ± 3,49 | 100,00 ± 0,00 | 88,24 ± 2,53 | 0,77 |
| MEEI | Não | B32+SVM | 10 | 10 Ciclos | 78,33 ± 7,00 | 100,00 ± 0,00 | 84,71 ± 2,90 | 0,72 |
| MEEI | Sim | B16+SVM | 10 | 10 Ciclos | 86,67 ± 5,43 | 100,00 ± 0,00 | 90,59 ± 3,83 | 0,81 |
| SVD | Não | CNN | 15 | 10 Ciclos | 75,00 ± 8,84 | 84,00 ± 8,37 | 76,80 ± 5,55 | 0,49 |
| SVD | Sim | CNN | 15 | 10 Ciclos | 69,00 ± 7,79 | 92,00 ± 5,48 | 73,60 ± 6,10 | 0,49 |
| SVD | Não | B16+SVM | 10 | 20 ms | 91,00 ± 3,26 | 80,00 ± 14,14 | 88,80 ± 2,97 | 0,67 |
| SVD | Sim | B32+SVM | 10 | 20 ms | 91,00 ± 7,58 | 88,00 ± 7,07 | 90,40 ± 6,54 | 0,73 |

Tabela D.12: Melhores Resultados da Classificação Queratose *versus* Lesão Benigna.

| Base de Dados | TS | Classificador | Caixa | Janela | Queratose (%) | Lesão (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|--------------|--------------|------|
| MEEI | Não | CNN | 10 | 10 Ciclos | 90,00 ± 4,59 | 70,43 ± 3,92 | 74,48 ± 2,29 | 0,50 |
| MEEI | Sim | CNN | 15 | 10 Ciclos | 83,33 ± 8,25 | 77,39 ± 6,36 | 78,62 ± 3,54 | 0,52 |
| MEEI | Não | B32+SVM | 15 | 10 Ciclos | 86,67 ± 5,30 | 73,91 ± 4,99 | 76,55 ± 3,20 | 0,50 |
| MEEI | Sim | B32+SVM | 10 | 10 Ciclos | 93,33 ± 3,80 | 78,26 ± 6,43 | 81,38 ± 3,83 | 0,60 |
| SVD | Não | CNN | 15 | 20 ms | 67,50 ± 7,13 | 61,18 ± 2,63 | 62,38 ± 1,77 | 0,23 |
| SVD | Sim | CNN | 10 | 10 Ciclos | 70,00 ± 11,35 | 61,76 ± 6,74 | 63,33 ± 3,82 | 0,25 |
| SVD | Não | B32+SVM | 15 | 20 ms | 72,50 ± 6,92 | 58,82 ± 3,61 | 61,43 ± 1,97 | 0,25 |
| SVD | Sim | B32+SVM | 10 | 20 ms | 75,00 ± 7,65 | 54,71 ± 3,22 | 58,57 ± 1,81 | 0,23 |

Tabela D.13: Melhores Resultados da Classificação Nódulo *versus* Pólipo.

| Base de Dados | TS | Classificador | Caixa | Janela | Nódulo (%) | Pólipo (%) | Total (%) | MCC |
|---------------|-----|---------------|-------|-----------|---------------|--------------|--------------|------|
| MEEI | Não | CNN | 10 | 20 ms | 100,00 ± 0,00 | 93,33 ± 4,57 | 96,36 ± 2,51 | 0,93 |
| MEEI | Sim | CNN | 10 | 10 Ciclos | 92,00 ± 5,48 | 93,33 ± 4,56 | 92,73 ± 3,80 | 0,85 |
| MEEI | Não | B32+SVM | 10 | 10 Ciclos | 100,00 ± 0,00 | 96,67 ± 1,87 | 98,18 ± 0,57 | 0,96 |
| MEEI | Sim | B32+SVM | 10 | 20 ms | 88,00 ± 8,94 | 86,67 ± 3,73 | 87,23 ± 3,80 | 0,74 |
| SVD | Não | CNN | 15 | 20 ms | 80,00 ± 67,89 | 68,00 ± 0,89 | 69,66 ± 8,99 | 0,34 |
| SVD | Sim | CNN | 10 | 10 Ciclos | 80,00 ± 49,53 | 60,80 ± 1,67 | 63,45 ± 6,52 | 0,28 |
| SVD | Não | B32+SVM | 10 | 20 ms | 82,50 ± 0,00 | 79,20 ± 7,40 | 79,66 ± 6,38 | 0,47 |
| SVD | Sim | B32+SVM | 10 | 20 ms | 85,00 ± 6,85 | 74,40 ± 7,01 | 75,86 ± 5,31 | 0,43 |

Apêndice E

Implementações de outros Trabalhos

Código-Fonte E.1: Código-fonte da implementação do trabalho de Sodre, Rosa e Dassie-Leite (2017).

```
%Evaluating the use of neural networks and acoustic measurements to  
    identify laryngeal pathologies
```

```
S = svdMDVP; %script que carrega todos os valores do descritor MDVP da  
    base SVD. Para usar a base MEEI
```

```
    %basta utilizar meeimDVP
```

```
base = 'SVD'; %nome da base utilizada. Para mudar para SVD basta  
    modificar esse valor
```

```
classe1Aux(1) = {'paralisiaE'};  
classe1Aux(2) = {'paralisiaQ'};  
classe1Aux(3) = {'paralisiaN'};  
classe1Aux(4) = {'paralisiaO'};  
classe1Aux(5) = {'paralisiaL'};  
classe1Aux(6) = {'edemaQ'};  
classe1Aux(7) = {'edemaN'};  
classe1Aux(8) = {'edemaO'};  
classe1Aux(9) = {'edemaL'};  
classe1Aux(10) = {'queratoseN'};  
classe1Aux(11) = {'queratoseO'};  
classe1Aux(12) = {'queratoseL'};
```

```

classe1Aux(13) = {'noduloO'};

classe2Aux(1) = {'edemaP'};
classe2Aux(2) = {'queratoseP'};
classe2Aux(3) = {'noduloP'};
classe2Aux(4) = {'polipoP'};
classe2Aux(5) = {'lesaoP'};
classe2Aux(6) = {'queratoseE'};
classe2Aux(7) = {'noduloE'};
classe2Aux(8) = {'polipoE'};
classe2Aux(9) = {'lesaoE'};
classe2Aux(10) = {'noduloQ'};
classe2Aux(11) = {'polipoQ'};
classe2Aux(12) = {'lesaoQ'};
classe2Aux(13) = {'polipoN'};
resultado = [];
for c1 = 1:length(classe1Aux)
    classe1 = classe1Aux{c1};
    classe2 = classe2Aux{c1};
    cT1 = [];
    files = importdata(['files/' base '/' classe1 '.txt']);
    for i=1:length(files)
        for j=1:length(S)
            if S(j).nome == files{i}
                aux = [S(j).Fo S(j).To S(j).Fhi S(j).Flo S(j).STD S(j).PFR S(
                    j).Tsam S(j).Jita S(j).Jitt S(j).RAP S(j).PPQ S(j).vFo S(j)
                    ).ShdB S(j).Shim S(j).APQ S(j).vAm S(j).NHR S(j).VTI S(j).
                    SPI];
                cT1 = [cT1; aux];
                break;
            end
        end
    end
end

cT2 = [];
files = importdata(['files/' base '/' classe2 '.txt']);
for i=1:length(files)

```

```

for j=1:length(S)
    if S(j).nome == files{i}
        aux = [S(j).Fo S(j).To S(j).Fhi S(j).Flo S(j).STD S(j).PFR S(
            j).Tsam S(j).Jita S(j).Jitt S(j).RAP S(j).PPQ S(j).vFo S(j)
            ).ShdB S(j).Shim S(j).APQ S(j).vAm S(j).NHR S(j).VTI S(j).
            SPI];
        cT2 = [cT2; aux];
        break;
    end
end
end

cT = [cT1;cT2];
cT = zscore(cT);
cT1 = cT(1:size(cT1,1),:);
cT2 = cT(size(cT1,1)+1:end,:);

C1 = 0;
C2 = 0;
C3 = 0;

for i=1:10
    shuf1 = randperm(size(cT1,1));
    shuf2 = randperm(size(cT2,1));

    train1 = cT1(shuf1(1:int16(length(shuf1)*0.6)),:);
    train2 = cT2(shuf2(1:int16(length(shuf2)*0.6)),:);

    val1 = cT1(shuf1(size(train1,1)+1:size(train1,1)+int16(length(shuf1)
        *0.2)),:);
    val2 = cT2(shuf2(size(train2,1)+1:size(train2,1)+int16(length(shuf2)
        *0.2)),:);

    test1 = cT1(shuf1(size(train1,1)+size(val1,1)+1:end),:);
    test2 = cT2(shuf2(size(train2,1)+size(val2,1)+1:end),:);

    p = pca([train1;train2]);

```

```

Train = [train1*p(:,1:6);train2*p(:,1:6)];
Train = Train';
val = [val1*p(:,1:6);val2*p(:,1:6)];
val = val';
test = [test1*p(:,1:6);test2*p(:,1:6)];
test = test';

net = feedforwardnet(20, 'trainlm');
net.divideFcn = 'divideind';
net.divideParam.trainInd = [1:size(Train,2)];
net.divideParam.valInd = [size(Train,2)+1:size(Train,2)+size(val,2)];
net.divideParam.testInd = [size(Train,2)+size(val,2)+1:size(Train,2)+
    size(val,2)+size(test,2)];

net.layers{1}.transferFcn = 'tansig';
net.layers{2}.transferFcn = 'purelin';

X = [Train val test];
Yaux = [zeros(1,size(train1,1))+1 zeros(1,size(train2,1))-1 zeros(1,
    size(val1,1))+1 zeros(1,size(val2,1))-1 zeros(1,size(test1,1))+1
    zeros(1,size(test2,1))-1];
Y = [Yaux;Yaux*(-1)];
[net, tr] = train(net,X,Y);
y1 = net((test1*p(:,1:6))');
y2 = net((test2*p(:,1:6))');
[~,pos1] = max(y1);
[~,pos2] = max(y2);

SE1 = sum(pos1==1)/size(test1,1);
ES1 = sum(pos2==2)/size(test2,1);
ACC1 = (sum(pos1==1)+sum(pos2==2))/size(test,2);
C1 = C1+SE1;
C2 = C2+ES1;
C3 = C3+ACC1;

end
C1 = C1/10;

```

```
C2 = C2/10;
C3 = C3/10;
resultado = [resultado; C1*100 C2*100 C3*100];
end
```

Código-Fonte E.2: Código-fonte da implementação do trabalho de Dahmani e Guerti (2017).

```
%Vocal folds pathologies classification using Naïve Bayes Networks
```

```
S = svdMDVP;
```

```
diretorio = 'bases/SVD/Patologia/';
```

```
base = 'SVD';
```

```
classe1Aux(1) = {'paralisiaE'};
```

```
classe1Aux(2) = {'paralisiaQ'};
```

```
classe1Aux(3) = {'paralisiaN'};
```

```
classe1Aux(4) = {'paralisiaO'};
```

```
classe1Aux(5) = {'paralisiaL'};
```

```
classe1Aux(6) = {'edemaQ'};
```

```
classe1Aux(7) = {'edemaN'};
```

```
classe1Aux(8) = {'edemaO'};
```

```
classe1Aux(9) = {'edemaL'};
```

```
classe1Aux(10) = {'queratoseN'};
```

```
classe1Aux(11) = {'queratoseO'};
```

```
classe1Aux(12) = {'queratoseL'};
```

```
classe1Aux(13) = {'noduloO'};
```

```
classe2Aux(1) = {'edemaP'};
```

```
classe2Aux(2) = {'queratoseP'};
```

```
classe2Aux(3) = {'noduloP'};
```

```
classe2Aux(4) = {'polipoP'};
```

```
classe2Aux(5) = {'lesaoP'};
```

```
classe2Aux(6) = {'queratoseE'};
```

```
classe2Aux(7) = {'noduloE'};
```

```
classe2Aux(8) = {'polipoE'};
```

```
classe2Aux(9) = {'lesaoE'};
```

```
classe2Aux(10) = {'noduloQ'};
```

```
classe2Aux(11) = {'polipoQ'};
```

```
classe2Aux(12) = {'lesaoQ'};
classe2Aux(13) = {'polipoN'};
resultado = [];
for c1 = 1:length(classe1Aux)
    classe1 = classe1Aux{c1};
    classe2 = classe2Aux{c1};
    cT1 = [];
    files = importdata(['files/' base '/' classe1 '.txt']);
    for i=1:length(files)
        [sinal,fs] = audioread(['diretorio files{i} '.wav']);
        if fs == 50000
            sinal = resample(sinal, 1, 2);
            fs = fs/2;
        end
        sinal=filter([1 -0.95],1,sinal);
        e = melcepst(sinal, fs, 'M', 12, floor(3*log(fs)), floor(0.025*fs),
            floor(0.025*fs/2));
        e = mean(e);
        for j=1:length(S)
            if S(j).nome == files{i}
                aux = [S(j).Fo S(j).Jitt S(j).Shim];
                e = [e aux];
                break;
            end
        end
        cT1 = [cT1; e];
    end

cT2 = [];
files = importdata(['files/' base '/' classe2 '.txt']);
for i=1:length(files)
    [sinal,fs] = audioread(['diretorio files{i} '.wav']);
    if fs == 50000
        sinal = resample(sinal, 1, 2);
        fs = fs/2;
    end
    sinal=filter([1 -0.95],1,sinal);
```

```

e = melcepst(sinal, fs, 'M', 12, floor(3*log(fs)), floor(0.025*fs),
    floor(0.025*fs/2));
e = mean(e);
for j=1:length(S)
    if S(j).nome == files{i}
        aux = [S(j).Fo S(j).Jitt S(j).Shim];
        e = [e aux];
        break;
    end
end
cT2 = [cT2; e];
end

C1 = 0;
C2 = 0;
C3 = 0;

for i=1:10
    shuf1 = randperm(size(cT1,1));
    shuf2 = randperm(size(cT2,1));

    minimo = min(length(shuf1), length(shuf2));

    train1 = cT1(shuf1(1:int16(minimo*0.75)),:);
    train2 = cT2(shuf2(1:int16(minimo*0.75)),:);

    test1 = cT1(shuf1(size(train1,1)+1:end),:);
    test2 = cT2(shuf2(size(train2,1)+1:end),:);

    Train = [train1;train2];

    labels = [zeros(size(train1,1),1); zeros(size(train2,1),1)+1];
    modelo = fitcnb(Train, categorical(labels));

    predictedLabels1 = predict(modelo, test1);
    predictedLabels2 = predict(modelo, test2);

```

```

SE1 = sum(predictedLabels1==categorical(0))/size(test1,1);
ES1 = sum(predictedLabels2==categorical(1))/size(test2,1);
ACC1 = (sum(predictedLabels1==categorical(0))+sum(predictedLabels2==
    categorical(1)))/(size(test1,1)+size(test2,1));
C1 = C1+SE1;
C2 = C2+ES1;
C3 = C3+ACC1;
end
C1 = C1/10;
C2 = C2/10;
C3 = C3/10;

resultado = [resultado; C1*100 C2*100 C3*100];

end

```

Código-Fonte E.3: Código-fonte da implementação do trabalho de Fang et al. (2018).

%Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach

```

diretorio = 'bases/SVD/Patologia/';
base = 'SVD';

classe1Aux(1) = {'paralisiaE'};
classe1Aux(2) = {'paralisiaQ'};
classe1Aux(3) = {'paralisiaN'};
classe1Aux(4) = {'paralisiaO'};
classe1Aux(5) = {'paralisiaL'};
classe1Aux(6) = {'edemaQ'};
classe1Aux(7) = {'edemaN'};
classe1Aux(8) = {'edemaO'};
classe1Aux(9) = {'edemaL'};
classe1Aux(10) = {'queratoseN'};
classe1Aux(11) = {'queratoseO'};
classe1Aux(12) = {'queratoseL'};
classe1Aux(13) = {'noduloO'};

```

```

classe2Aux(1) = {'edemaP'};
classe2Aux(2) = {'queratoseP'};
classe2Aux(3) = {'noduloP'};
classe2Aux(4) = {'polipoP'};
classe2Aux(5) = {'lesaoP'};
classe2Aux(6) = {'queratoseE'};
classe2Aux(7) = {'noduloE'};
classe2Aux(8) = {'polipoE'};
classe2Aux(9) = {'lesaoE'};
classe2Aux(10) = {'noduloQ'};
classe2Aux(11) = {'polipoQ'};
classe2Aux(12) = {'lesaoQ'};
classe2Aux(13) = {'polipoN'};
resultado = [];

for c1 = 1:length(classe1Aux)
    classe1 = classe1Aux{c1};
    classe2 = classe2Aux{c1};
    files = importdata(['files/' base '/' classe1 '.txt']);
    cT1 = cell(length(files),1);
    for i=1:length(files)
        [sinal,fs] = audioread(['diretorio files{i} '.wav']);
        if fs == 50000
            sinal = resample(sinal, 1, 2);
            fs = fs/2;
        end
        sinal=filter([1 -0.95],1,sinal);
        e = melcepst(sinal, fs, 'M', 13, floor(3*log(fs)), floor(0.016*fs),
            floor(0.016*fs/2));
        cT1(i) = {e};
    end

files = importdata(['files/' base '/' classe2 '.txt']);
cT2 = cell(length(files),1);
for i=1:length(files)
    [sinal,fs] = audioread(['diretorio files{i} '.wav']);
    if fs == 50000

```

```

        sinal = resample(sinal , 1, 2);
        fs = fs / 2;
    end
    sinal=filter([1 -0.95],1,sinal);
    e = melcepst(sinal , fs , 'M' , 13, floor(3*log(fs)) , floor(0.016*fs) ,
        floor(0.016*fs/2));
    cT2(i) = {e};
end

C1 = 0;
C2 = 0;
C3 = 0;

minimo = min(length(cT1) , length(cT2));
[train1aux , val1aux , test1aux ] = criakfoldVal2(length(cT1) , 5, 6, floor(
    minimo*0.8));
[train2aux , val2aux , test2aux ] = criakfoldVal2(length(cT2) , 5, 6, floor(
    minimo*0.8));
shuf1 = randperm(size(cT1,1));
shuf2 = randperm(size(cT2,1));

for i=1:5
    train1 = cT1(shuf1(train1aux{i}));
    train2 = cT2(shuf2(train2aux{i}));

    val1 = cT1(shuf1(val1aux{i}));
    val2 = cT2(shuf2(val2aux{i}));

    test1 = cT1(shuf1(test1aux{i}));
    test2 = cT2(shuf2(test2aux{i}));

    Train = [];
    label = [];
    for j=1:length(train1)
        Train = [Train; train1{j}];
        label = [label; zeros(size(train1{j},1),1)];
    end
end

```

```
for j=1:length(train2)
    Train = [Train; train2{j}];
    label = [label; zeros(size(train2{j},1),1)+1];
end

Val = [];
labelval = [];
for j=1:length(val1)
    Val = [Val; val1{j}];
    labelval = [labelval; zeros(size(val1{j},1),1)];
end
for j=1:length(val2)
    Val = [Val; val2{j}];
    labelval = [labelval; zeros(size(val2{j},1),1)+1];
end

Train = reshape(Train', [1 13 1 size(Train,1)]);
Val = reshape(Val', [1 13 1 size(Val,1)]);
layers = [
    imageInputLayer([1 13 1])

    fullyConnectedLayer(300)
    fullyConnectedLayer(300)
    fullyConnectedLayer(300)
    fullyConnectedLayer(2)
    softmaxLayer
    classificationLayer];

miniBatchSize = 128;
validationFrequency = floor(size(Train,4)/miniBatchSize/2);
options = trainingOptions('sgdm',...
    'MaxEpochs',100, ...
    'MiniBatchSize',miniBatchSize,...
    'ValidationData',{Val,categorical(labelval)},...
    'ValidationFrequency',validationFrequency,...
    'Verbose',true,...
    'ExecutionEnvironment','cpu',...

```

```
'Shuffle', 'every-epoch', ...
'ValidationPatience', 10, ...
'LearnRateSchedule', 'piecewise');
net = trainNetwork(Train, categorical(label), layers, options);

SE1 = 0;
for j=1:length(test1)
    t = test1{j};
    t = reshape(t', [1 13 1 size(t,1)]);
    pL = classify(net, t);
    if sum(pL == categorical(0)) > sum(pL == categorical(1))
        SE1 = SE1+1;
    end
end

ES1 = 0;
for j=1:length(test2)
    t = test2{j};
    t = reshape(t', [1 13 1 size(t,1)]);
    pL = classify(net, t);
    if sum(pL == categorical(1)) > sum(pL == categorical(0))
        ES1 = ES1+1;
    end
end

ACC1 = (SE1+ES1)/(length(test1)+length(test2));
SE1 = SE1/length(test1);
ES1 = ES1/length(test2);

C1 = C1+SE1;
C2 = C2+ES1;
C3 = C3+ACC1;

end

C1 = C1/5;
C2 = C2/5;
C3 = C3/5;

resultado = [resultado; C1*100 C2*100 C3*100];
end
```

Código-Fonte E.4: Código-fonte da implementação do trabalho de Ali et al. (2018).

```
%An intelligent healthcare system for detection and classification to
discriminate vocal fold disorders
diretorio = 'bases/MEEIwav/PATHOL/AH/';
base = 'MEEI';

K = 3;

classe1Aux(1) = {'paralisiaE'};
classe1Aux(2) = {'paralisiaQ'};
classe1Aux(3) = {'paralisiaN'};
classe1Aux(4) = {'paralisiaO'};
classe1Aux(5) = {'paralisiaL'};
classe1Aux(6) = {'edemaQ'};
classe1Aux(7) = {'edemaN'};
classe1Aux(8) = {'edemaO'};
classe1Aux(9) = {'edemaL'};
classe1Aux(10) = {'queratoseN'};
classe1Aux(11) = {'queratoseO'};
classe1Aux(12) = {'queratoseL'};
classe1Aux(13) = {'noduloO'};

classe2Aux(1) = {'edemaP'};
classe2Aux(2) = {'queratoseP'};
classe2Aux(3) = {'noduloP'};
classe2Aux(4) = {'polipoP'};
classe2Aux(5) = {'lesaoP'};
classe2Aux(6) = {'queratoseE'};
classe2Aux(7) = {'noduloE'};
classe2Aux(8) = {'polipoE'};
classe2Aux(9) = {'lesaoE'};
classe2Aux(10) = {'noduloQ'};
classe2Aux(11) = {'polipoQ'};
classe2Aux(12) = {'lesaoQ'};
classe2Aux(13) = {'polipoN'};
resultado = [];
```

```
for c1 = 1:length(classe1Aux)
    disp(c1);
    classe1 = classe1Aux{c1};
    classe2 = classe2Aux{c1};
    files = importdata(['files/' base '/' classe1 '.txt']);
    cT1 = cell(length(files),1);
    for i=1:length(files)
        [sinal,fs] = audioread([diretorio files{i} '.wav']);
        if fs == 50000
            sinal = resample(sinal, 1, 2);
            fs = fs/2;
        end
        e=getFCB(sinal, fs, 256);
        cT1(i) = {e'};
    end

    files = importdata(['files/' base '/' classe2 '.txt']);
    cT2 = cell(length(files),1);
    for i=1:length(files)
        [sinal,fs] = audioread([diretorio files{i} '.wav']);
        if fs == 50000
            sinal = resample(sinal, 1, 2);
            fs = fs/2;
        end
        e=getFCB(sinal, fs, 256);
        cT2(i) = {e'};
    end

    [train1, test1] = criakfold(length(cT1), K);
    [train2, test2] = criakfold(length(cT2), K);

    shuf1 = randperm(length(cT1));
    shuf2 = randperm(length(cT2));
    r = [0 0 0 0];
    for j=1:K
        disp(j);
        t1 = [];
```

```
for k=1:length(train1{j})
    t1 = [t1; cT1{shuf1(train1{j}(k))}];
end

t2 = [];
for k=1:length(train2{j})
    t2 = [t2; cT2{shuf2(train2{j}(k))}];
end

mix1 = calcula_gmm(t1, 32);
mix2 = calcula_gmm(t2, 32);

r1 = [0 0];
for k=1:length(test1{j})
    te = cT1{shuf1(test1{j}(k))};
    p1 = calcula_probabilidade(te', mix1);
    p2 = calcula_probabilidade(te', mix2);
    if p1 > p2
        r1(1) = r1(1) + 1;
    else
        r1(2) = r1(2) + 1;
    end
end

r2 = [0 0];
for k=1:length(test2{j})
    te = cT2{shuf2(test2{j}(k))};
    p1 = calcula_probabilidade(te', mix1);
    p2 = calcula_probabilidade(te', mix2);
    if p1 > p2
        r2(1) = r2(1) + 1;
    else
        r2(2) = r2(2) + 1;
    end
end

r = r + [r1 r2];
end

resultado = [resultado; r];
```

end

Apêndice F

Comparação com Resultados Obtidos na Literatura

Tabela F.1: Resumo dos resultados de trabalhos publicados de 2011 até 2018 ordenados por classificação e por acurácia. N.I. significa "Não Informado".

| Posição | Trabalho | Base | Classe 1 | Classe 2 | Quant. Sinais | Classe 1 (%) | Classe 2 (%) | Total (%) |
|---------|---|---------|-----------|--------------|---------------|---------------|---------------|---------------|
| 1° | Al-Nasheri et al. (2018) | MEEI | Paralisia | Pólipo+Cisto | N.I. | N.I. | N.I. | 98,04 |
| 2° | Al-Nasheri et al. (2017) | MEEI | Paralisia | Pólipo+Cisto | 71 - 20+10 | N.I. | N.I. | 97,29 |
| 3° | Al-Nasheri et al. (2018) | SVD | Paralisia | Pólipo+Cisto | N.I. | N.I. | N.I. | 94,83 |
| 4° | Al-Nasheri et al. (2017) | SVD | Paralisia | Pólipo+Cisto | 212 - 45+6 | N.I. | N.I. | 92,26 |
| 5° | Al-Nasheri et al. (2018) | Própria | Paralisia | Pólipo+Cisto | N.I. | N.I. | N.I. | 91,37 |
| 6° | Al-Nasheri et al. (2017) | Própria | Paralisia | Pólipo+Cisto | 56 - 46+25 | N.I. | N.I. | 90,14 |
| 7° | Al-nasheri et al. (2017) | SVD | Paralisia | Pólipo+Cisto | 195 - 43+6 | N.I. | N.I. | 79,17 |
| 8° | Pesquisa | MEEI | Paralisia | Lesão | 67 - 43 | 80,00 ± 6,08 | 72,73 ± 6,62 | 78,26 ± 4,81 |
| 9° | Lopes et al. (2017) | Própria | Paralisia | Cisto | 10 - 34 | 40,00 ± 14,54 | 87,50 ± 5,16 | 72,5 ± 5,88 |
| 10° | Mehdizadehfard, Almasganj e Torabinezhad (2017) | Própria | Paralisia | Cisto | 5 - 10 | N.I. | N.I. | 70,0 |
| 11° | Pesquisa | SVD | Paralisia | Lesão | 155 - 53 | 68,45 ± 2,61 | 64,29 ± 3,57 | 67,99 ± 1,99 |
| 12° | Al-nasheri et al. (2017) | MEEI | Paralisia | Pólipo+Cisto | 66 - 19+10 | N.I. | N.I. | 65,56 |
| 13° | Al-nasheri et al. (2017) | Própria | Paralisia | Pólipo+Cisto | 32 - 30+13 | N.I. | N.I. | 57,14 |
| 1° | Ali et al. (2018) | MEEI | Queratose | Pólipo | 23 - 17 | 99,42 | 98,63 | 99,08 ± 1,1 |
| 2° | Markaki e Stylianou (2011) | MEEI | Queratose | Pólipo | 26 - 20 | N.I. | N.I. | 84,77 ± 1,42 |
| 3° | Pesquisa | MEEI | Queratose | Pólipo | 23 - 17 | 85,45 ± 2,49 | 80,00 ± 15,00 | 83,75 ± 4,64 |
| 4° | Saeedi e Almasganj (2013) | MEEI | Queratose | Pólipo | 26 - 20 | N.I. | N.I. | 80,00 |
| 5° | Pesquisa | SVD | Queratose | Pólipo | 31 - 37 | 75,00 ± 7,65 | 64,29 ± 8,14 | 68,18 ± 5,43 |
| 1° | Pesquisa | MEEI | Nódulo | Pólipo | 18 - 19 | 100,00 ± 0,00 | 96,67 ± 1,87 | 98,18 ± 0,57 |
| 2° | Ali et al. (2018) | MEEI | Nódulo | Pólipo | 19 - 19 | 96,49 | 97,02 | 96,75 ± 1,2 |
| 3° | Markaki e Stylianou (2011) | MEEI | Nódulo | Pólipo | 20 - 20 | N.I. | N.I. | 91,66 ± 1,14 |
| 4° | Pesquisa | SVD | Nódulo | Pólipo | 16 - 37 | 82,50 ± 0,00 | 79,20 ± 7,40 | 79,66 ± 6,38 |
| 1° | Ali et al. (2018) | MEEI | Queratose | Nódulo | 26 - 20 | 97,44 | 96,67 | 97,10 ± 1,2 |
| 2° | Pesquisa | MEEI | Queratose | Nódulo | 26 - 16 | 86,67 ± 5,43 | 100,00 ± 0,00 | 90,59 ± 3,83 |
| 3° | Pesquisa | SVD | Queratose | Nódulo | 32 - 17 | 91,00 ± 7,58 | 88,00 ± 7,07 | 90,40 ± 6,54 |
| 4° | Markaki e Stylianou (2011) | MEEI | Queratose | Nódulo | 26 - 20 | N.I. | N.I. | 89,11 ± 1,33 |
| 1° | Saeedi e Almasganj (2013) | MEEI | Paralisia | Nódulo | 67 - 19 | N.I. | N.I. | 100,00 |
| 2° | Pesquisa | MEEI | Paralisia | Nódulo | 67 - 19 | 95,85 ± 3,24 | 80,00 ± 4,47 | 94,48 ± 2,87 |
| 3° | Costa (2012) | MEEI | Paralisia | Nódulo | 55 - 18 | 75,00 ± 42,49 | 96,67 ± 10,54 | 88,00 ± 17,35 |
| 4° | Lopes et al. (2017) | Própria | Paralisia | Nódulo | 10 - 93 | 25,00 ± 13,45 | 88,57 ± 5,57 | 76,79 ± 5,68 |
| 5° | Pesquisa | SVD | Paralisia | Nódulo | 155 - 17 | 68,47 ± 7,21 | 96,00 ± 5,48 | 69,40 ± 6,89 |
| 1° | Pesquisa | MEEI | Edema | Queratose | 43 - 25 | 88,00 ± 3,06 | 80,00 ± 4,59 | 86,25 ± 3,68 |
| 2° | Pesquisa | SVD | Edema | Queratose | 52 - 32 | 73,33 ± 5,60 | 87,50 ± 8,15 | 76,32 ± 3,30 |
| 3° | Saeedi e Almasganj (2013) | MEEI | Edema | Queratose | 44 - 26 | N.I. | N.I. | 75,00 |
| 1° | Saeedi e Almasganj (2013) | MEEI | Paralisia | Edema | 67 - 44 | N.I. | N.I. | 100,00 |
| 2° | Pesquisa | MEEI | Paralisia | Edema | 66 - 43 | 88,24 ± 3,77 | 72,73 ± 5,62 | 84,44 ± 2,77 |
| 3° | Costa (2012) | MEEI | Paralisia | Edema | 55 - 45 | 86,67 ± 17,21 | 81,67 ± 25,40 | 84,33 ± 13,52 |
| 4° | Pesquisa | SVD | Paralisia | Edema | 147 - 48 | 78,95 ± 3,32 | 73,33 ± 5,94 | 78,41 ± 2,56 |
| 5° | Brandt (2012) | MEEI | Paralisia | Edema | 46 - 43 | 79,1 | 53,2 | 65,6 |
| 1° | Carvalho, Cavalcante e Cortez (2011) | Própria | Edema | Nódulo | 15 - 15 | 91,19 | 90,37 | 90,78 |
| 2° | Costa (2012) | MEEI | Edema | Nódulo | 45 - 18 | 55,00 ± 43,78 | 93,33 ± 14,50 | 82,50 ± 9,50 |
| 3° | Pesquisa | SVD | Edema | Nódulo | 52 - 16 | 79,02 ± 6,84 | 67,50 ± 16,77 | 78,00 ± 5,53 |
| 4° | Pesquisa | MEEI | Edema | Nódulo | 41 - 16 | 72,41 ± 5,53 | 75,00 ± 10,49 | 72,73 ± 3,84 |