

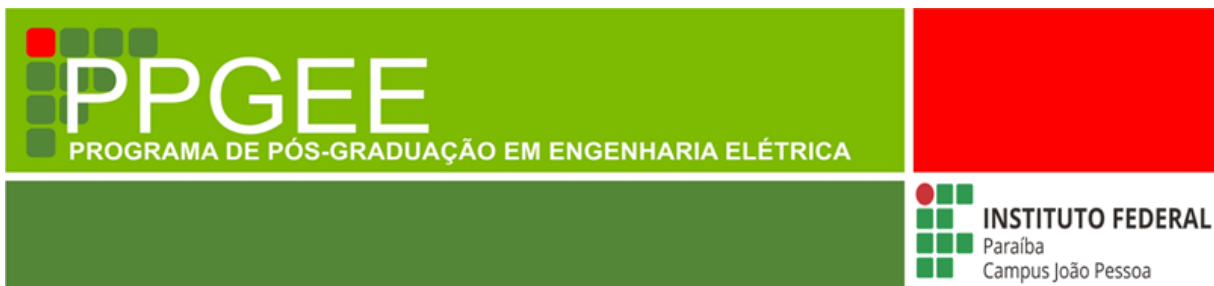


**Lucas Cardoso Dias**

**Detecção de patologias laríngeas por meio da  
análise de sinais de voz utilizando *Deep Neural  
Networks***

**João Pessoa - PB**

**Julho de 2020**



**Lucas Cardoso Dias**

**Detecção de patologias laríngeas por meio da  
análise de sinais de voz utilizando *Deep Neural  
Networks***

Dissertação apresentada à banca examinadora do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal da Paraíba, como requisito necessário à obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Processamento Digital de Sinais

Silvana Luciene do Nascimento Cunha Costa, Dra.  
Orientadora

Suzete Élide Nóbrega Correia, Dra.  
Coorientadora

João Pessoa - PB, Julho de 2020

Dados Internacionais de Catalogação na Publicação (CIP)  
Biblioteca Nilo Peçanha do IFPB, *campus* João Pessoa

D541d

Dias, Lucas Cardoso.

Detecção de patologias laríngeas por meio da análise de sinais de voz utilizando *Deep Neural Networks* / Lucas Cardoso Dias. – 2020.

82 f. : il.

Dissertação (Mestrado em Engenharia Elétrica) – Instituto Federal de Educação da Paraíba / Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), 2020.

Orientadora : D.ra Silvana Luciene do Nascimento.

Coorientadora : D.ra Suzete Élide Nóbrega Correia.

1. Saúde – patologias laríngeas. 2. Processamento digital de sinais - análise de sinais de voz. 3. Redes neurais artificiais. 4. Aprendizagem profunda. I. Título.

CDU 61:621.391(043)

Lucrecia Camilo de Lima  
Bibliotecária  
CRB 15/132

Lucas Cardoso Dias

**Detecção de patologias laríngeas por meio da análise de sinais de voz  
utilizando *Deep Neural Networks***

Dissertação apresentada à banca examinadora do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal da Paraíba, como requisito necessário à obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Dissertação aprovada pela banca examinadora em: 15/07/2020

BANCA EXAMINADORA



---

**Silvana Luciene do Nascimento Cunha Costa, Dra. – IFPB**  
**Orientadora**



---

**Suzete Élide Nóbrega Correia, Dra. – IFPB**  
**Coorientadora**



---

**Thaís Gaudencio do Rêgo, Dra. – UFPB**  
**Examinador Externo**



---

**Vinícius Jefferson Dias Vieira, Dr. – UFPB**  
**Examinador Externo**

João Pessoa - PB

Julho de 2020

## AGRADECIMENTOS

Primeiramente agradeço a Deus pelo dom da vida e pelas oportunidades que foram inseridas em minha jornada até o presente momento. Eternamente grato a minha família por todo apoio durante minha formação, em especial ao meu pai, Reginaldo Cardoso (*in memorian*), por ter sempre me incentivado a seguir em frente nos estudos.

Agradeço as professoras Silvana Luciene do Nascimento Cunha Costa e Suzete Élide Nóbrega Correia por todas as orientações/contribuições prestadas ao desenvolver desta pesquisa, assim como aos conselhos pessoais e profissionais.

Agradeço a todos os colegas do Programa de Pós-Graduação em Engenharia Elétrica do IFPB, com os quais compartilhei grandes momentos nesta jornada, em especial aos meus amigos Arthur Medeiros e Igor Forcelli (vulgo galego); aos professores da pós-graduação por todo o conhecimento compartilhado; a todas as pessoas que, de alguma forma, contribuíram para a minha formação acadêmica e que compartilham da minha alegria em concluir mais uma etapa.

Agradeço a Máisa Lauriane por todo o companheirismo durante esta grande etapa da minha vida. Por fim, agradeço ao Instituto Federal da Paraíba (IFPB) *Campus* João Pessoa, pela disponibilização de laboratórios e materiais necessários para a conclusão deste trabalho; à Pró-Reitoria de Inovação, Ensino e Pós-Graduação do IFPB/JP pelo apoio financeiro.

## RESUMO

A fala é o principal mecanismo natural de comunicação entre seres humanos. O sistema de formação e transmissão natural da voz, principal elemento da fala, é comprometido pelo surgimento de patologias laríngeas. Esta pesquisa trata da aplicação de classificadores baseados em redes neurais profundas (*Deep Neural Networks* - DNNs) na discriminação entre sinais de vozes saudáveis e de vozes afetadas pelas patologias laríngeas organofuncionais edema de *Reinke*, carcinoma, leocoplasia, pólipos e a paralisia das pregas vocais, de origem neurológica. A metodologia proposta é baseada na análise do comportamento dinâmico do sinal de voz avaliado, dispensando medidas ou aplicações de técnicas comumente usadas na extração de características. Foi investigado o uso de DNNs com 04, 05 e 06 camadas com 200 neurônios ocultos ativados pela função unidade linear retificada (*Rectified Linear Unit* - ReLU), um neurônio na camada de saída, ativado pela função sigmoide e uma camada de entrada que recebe os 400 dados que compõe cada segmento extraído do sinal de voz avaliado. No total, 07 algoritmos de aprendizagem, utilizando como função custo a entropia cruzada binária (*Binary Cross-entropy*), foram avaliados individualmente para o treinamento de cada DNN. Os sinais de voz utilizados nesta pesquisa foram extraídos da base de dados *Saarbruecken Voice Database* (SVD), desenvolvida na Alemanha. Da base, foram selecionados 640 sinais de voz da vogal sustentada /a/, sendo 320 sinais de vozes saudáveis e 320 afetados por patologias laríngeas. A discriminação foi realizada por classes, sendo: a classe saudável; a classe patologias, composta por todos os sinais patológicos selecionados da base SVD; a classe das vozes afetadas apenas por patologias laríngeas organofuncionais; e, por fim, a classe de sinais de voz afetados apenas por paralisia das pregas vocais, compondo a categoria de patologia laríngea de origem neurológica. Foram considerados 04 casos de classificação entre os sinais de voz selecionados, sendo eles: saudável × patologias, saudável × patologias organofuncionais, saudável × paralisia das pregas vocais e patologias organofuncionais × paralisia das pregas vocais. Para cada caso discriminativo, 28 classificadores foram implementados e avaliados por meio do  $F1_{score}$  e pelo coeficiente de correlação de Mathews (CCM) (aplicado apenas na discriminação entre as classes patológicas), e pelas métricas acurácia, sensibilidade e especificidade. Além disso, foram investigados os efeitos da inclusão de taxas de sobreposição (0%, 25%, 50% e 75%) aplicadas durante a extração dos segmentos. A técnica de validação cruzada *k-fold*, com  $k = 10$ , foi implementada nesta pesquisa para seleção dos conjuntos de dados de treino e teste. Os resultados indicam que o método proposto possui o seu melhor desempenho na discriminação entre vozes saudáveis e afetadas por paralisia das pregas vocais, com base na detecção de segmentos do sinal de voz sem taxa de sobreposição, utilizando o classificador com 4 camadas ocultas, treinado pelo algoritmo de aprendizagem Adadelta, no qual foram obtidos após a validação cruzada  $88,68 \pm 3,04\%$  para acurácia,  $92,04 \pm 5,82\%$  para sensibilidade,  $85,33 \pm 6,53\%$  para especificidade e  $F1_{score}$  igual 0,89. Conclui-se que é possível discriminar vozes saudáveis e afetadas por patologias laríngeas, com base na análise do comportamento dinâmico de segmentos do sinal de voz utilizando DNNs.

**Palavras-chave:** aprendizagem profunda, processamento digital de sinais de voz, patologias laríngeas, redes neurais artificiais.

# ABSTRACT

Speech is the main natural mechanism of communication between human beings. The natural voice formation and transmission system, the main element of speech, is compromised by the appearance of laryngeal pathologies. This study deals with the application of classifiers based on Deep Neural Networks (DNNs) in the discrimination between healthy voice signals and affected by laryngeal organofunctional pathologies Reinke's edema, carcinoma, leukoplakia, polyps and vocal cord paralysis, neurological origin. The proposed methodology is based on the analysis of the dynamic behavior of the evaluated voice signal, dispensing with measures or applications of techniques commonly used in the extraction of characteristics. The use of DNNs with 04, 05 and 06 layers with 200 hidden neurons activated by the Rectified Linear Unit (ReLU) function, a neuron in the output layer, activated by the sigmoid function, and an input layer that receives the 400 data were investigated. that makes up each segment extracted from the evaluated voice signal. In total, 07 learning algorithms, using Binary Cross-entropy as a cost function, were evaluated individually for the training of each DNN. The voice signals used in this research were extracted from the Saarbruecken Voice Database (SVD), developed in Germany. From the base, 640 voice signals of the sustained vowel /a/ were selected, 320 of which were signals of healthy voices and 320 were affected by language pathologies. The discrimination was carried out by classes, being the healthy one; the pathologies class, composed of all pathological signals selected from the SVD database; the class of voices affected only by organofunctional laryngeal pathologies; and, finally, the class of voice signals affected only by paralysis of the vocal folds, composing the category of laryngeal pathology of neurological origin four cases of classification were considered among the selected voice signals, namely: healthy  $\times$  pathological, healthy  $\times$  organofunctional pathologies, healthy  $\times$  vocal fold paralysis, and organofunctional pathologies  $\times$  paralysis of vocal folds. For each discriminative case, 28 classifiers were implemented and evaluated using the  $F1_{score}$  and the Mathews correlation coefficient (CCM) (applied only to the discrimination between pathological classes), and by the metrics accuracy, sensitivity and specificity. In addition, the effects of including overlapping rates (0%, 25%, 50% and 75%) applied during the extraction of the segments were investigated. The  $k$ -fold cross-validation technique, with  $k = 10$ , was implemented in this research to select the training and test data sets. The results indicate that the proposed method has its best performance in discriminating between healthy voices and those affected per vocal fold paralysis, based on the detection of segments of the voice signal without overlapping rate, using classifier with 4 hidden layers trained by the Adadelta learning algorithm, in which were obtained after cross-validation  $88.68 \pm 3.04\%$  for accuracy,  $92.04 \pm 5.82\%$  for sensitivity,  $85.33 \pm 6.53\%$  for specificity and  $F1_{score}$  equal 0.89. We conclude that it is possible to discriminate healthy voices affected by laryngeal pathologies based on the analysis of the dynamic behavior of segments of the voice signal using DNNs.

**Keywords:** *deep learning, digital processing of voice signals, pathologies laryngeal, artificial neural networks.*

## LISTA DE FIGURAS

|   |    |
|---|----|
| Figura 2.1 – Diagrama esquemático do mecanismo de produção da fala. . . . .   | 18 |
| Figura 2.2 – <b>(a)</b> - Pregas vocais (abdução); <b>(b)</b> - Pregas vocais (adução); <b>(c)</b> - Pulsos glotais.  | 19 |
| Figura 2.3 – Estrutura anatômica da laringe. . . . .  | 20 |
| Figura 2.4 – Edema de Reinke. . . . .   | 21 |
| Figura 2.5 – Pólipo unilateral. . . . .   | 22 |
| Figura 2.6 – Leocoplasia da prega vocal. . . . .  | 22 |
| Figura 2.7 – Carcinoma com característica escamosa presente nas pregas vocais. . . . .  | 23 |
| Figura 2.8 – Ilustração do nervo vago e seus ramos. . . . .   | 24 |
| Figura 2.9 – Paralisia nas pregas vocais : <b>(a)</b> - Unilateral; <b>(b)</b> - Bilateral . . . . .  | 24 |
| Figura 2.10–Comportamento dinâmico de sinais de vozes: <b>(a)</b> - sinal saudável; <b>(b)</b> - afetado por carcinoma; <b>(c)</b> - afetado por edema; <b>(d)</b> afetado por leocoplasia; <b>(e)</b> afetado por paralisia; <b>(f)</b> afetado por pólipos. . . . . | 26 |
| Figura 2.11–Exemplificação do processo de segmentação aplicada a sinais de voz. . . . .   | 29 |
| Figura 3.1 – Neurônio <i>perceptron</i> . . . . .   | 34 |
| Figura 3.2 – Estrutura de uma DNN. . . . .  | 36 |
| Figura 4.1 – Fluxograma representativo do sistema de classificação proposto. . . . .  | 45 |
| Figura 4.2 – <b>(a)</b> - Segmento original; <b>(b)</b> - Segmento escalonado. . . . .  | 48 |
| Figura 4.3 – Exemplo - Partição de dados <i>k - fold</i> . . . . .  | 49 |
| Figura 4.4 – Fluxograma da etapa de treinamento do classificador. . . . .   | 50 |
| Figura 4.5 – Fluxograma da etapa de teste com validação cruzada. . . . .  | 51 |
| Figura 4.6 – Resumo dos ensaios experimentais realizados nesta pesquisa. . . . .  | 54 |
| Figura 5.1 – Análise comparativa com trabalhos do estado da arte: $F1_{score}$ . . . . .  | 60 |
| Figura A.1 – Comparação entre as curvas de aprendizagem para o melhor classificador da discriminação: <b>(a)</b> saudável <i>versus</i> patológicos; <b>(b)</b> saudável <i>versus</i> patologias organofuncionais. . . . .   | 79 |
| Figura A.2 – Comparação entre as curvas de aprendizagem para o melhor classificador da discriminação: <b>(a)</b> saudável <i>versus</i> paralisia das pregas vocais; <b>(b)</b> patologias organofuncionais <i>versus</i> paralisia das pregas vocais. . . . .        | 79 |



## LISTA DE TABELAS

|  |    |
|--|----|
| Tabela 3.1 – Equações de atualização dos pesos sinápticos referente aos algoritmos das subsecções 3.4.1 a 3.4.5. . . . .   | 42 |
| Tabela 4.1 – Matriz de confusão em um teste de diagnóstico da presença ou ausência da patologia. . . . .   | 52 |
| Tabela 4.2 – Parâmetros iniciais dos otimizadores aplicados no cálculo dos pesos sinápticos das DNNs. . . . .  | 55 |
| Tabela 5.1 – Melhores resultados individuais da discriminação entre vozes saudáveis e patológicas utilizando DNNs de 4, 5 e 6 camadas para análise de segmentos com diferentes taxas de sobreposição. . . . .  | 57 |
| Tabela 5.2 – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes taxas de sobreposição. Caso saudável versus patologias. . . . .  | 57 |
| Tabela 5.3 – Resultados referentes aos melhores classificadores para o estudo de caso saudável <i>versus</i> patológico para cada arquitetura investigada. . . . .   | 58 |
| Tabela 5.4 – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes arquiteturas da Tabela 5.3. . . . .  | 58 |
| Tabela 5.5 – Análise comparativa com trabalhos do estado da arte. . . . .  | 58 |
| Tabela 5.6 – Matrizes de confusão referente aos classificadores da Tabela 5.3. . . . .   | 59 |
| Tabela 5.7 – Matrizes de confusão: Wu (2018) e Harár (2017). . . . .   | 59 |
| Tabela 5.8 – Melhores resultados individuais da discriminação entre vozes saudáveis e afetadas por patologias laríngeas organofuncionais utilizando DNNs de 4, 5 e 6 camadas para análise de segmentos com diferentes taxas de sobreposição. . . . . | 60 |
| Tabela 5.9 – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes taxas de sobreposição. Caso saudável versus patologias organofuncionais. . . . .                                 | 61 |
| Tabela 5.10–Resultados referentes aos melhores classificadores para o estudo de caso saudável <i>versus</i> patologias organofuncionais para cada arquitetura investigada. . . . .   | 62 |
| Tabela 5.11–Matrizes de confusão referente aos classificadores da Tabela 5.10 . . . . .  | 62 |
| Tabela 5.12–Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes arquiteturas da Tabela 5.10. . . . .  | 62 |
| Tabela 5.13–Melhores resultados individuais da discriminação entre vozes saudáveis e afetadas paralisia das pregas vocais utilizando DNNs de 4, 5 e 6 camadas para análise de segmentos com diferentes taxas de sobreposição. . . . .                | 63 |
| Tabela 5.14–Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes taxas de sobreposição. Caso saudável versus paralisia das pregas vocais. . . . .                                  | 63 |

|   |    |
|---|----|
| Tabela 5.15–Resultados referentes aos melhores classificadores para o estudo de caso saudável <i>versus</i> paralisia das pregas vocais. . . . .  | 64 |
| Tabela 5.16–Matrizes de confusão referente aos classificadores da Tabela 5.15 . . . . .   | 64 |
| Tabela 5.17–Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes arquiteturas da Tabela 5.15. . . .   | 64 |
| Tabela 5.18–Melhores resultados individuais da discriminação entre vozes afetadas por patologias organofuncionais e afetadas paralisia das pregas vocais utilizando DNNs de 4, 5 e 6 camadas. . . . .   | 65 |
| Tabela 5.19–Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes taxas de sobreposição. Caso patologias organofuncionais <i>versus</i> paralisia das pregas vocais. . . . . | 66 |
| Tabela 5.20–Resultados referentes aos melhores classificadores para o estudo de caso discriminativo entre patologias laríngeas organofuncionais e por paralisia das pregas vocais. . . . .  | 66 |
| Tabela 5.21–Matrizes de confusão referente aos classificadores da Tabela 5.20 . . . . .   | 66 |
| Tabela 5.22–Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes arquiteturas da Tabela 5.20. . . .   | 67 |

## LISTA DE SIGLAS

|         |  |
|---------|--|
| Adagrad | <i>Adaptive Gradient</i> (Gradiente adaptativo)  |
| Adam    | <i>Adaptive Momentum Estimation</i> (Estimação adaptativa do <i>momentum</i> )   |
| ANN     | <i>Artificial Neural Networks</i> (Redes Neurais Artificiais)  |
| ANOVA   | Análise de Variância   |
| BP      | <i>Backpropagation</i> (Retropropagação)   |
| CCM     | Coefficiente de Correlação de Matthews   |
| CNN     | <i>Convolutional Neural Networks</i> (Redes Neurais Convolucionais)  |
| CSV     | <i>Comma Separated Values</i> (Valores Separados por Vírgula)  |
| DNN     | <i>Deep Neural Network</i> (Redes Neurais Profundas)   |
| EKG     | Eletrocardiografia   |
| ELU     | <i>Exponential Linear Unit</i> (Unidade Exponencial Linear)  |
| $F_0$   | Frequência Fundamental   |
| FEMH    | <i>Far Eastern Memorial Hospital</i>   |
| FN      | Falso Negativo   |
| FP      | Falso Positivo   |
| GNE     | <i>Glottal Noise Excitation</i> (Excitação do Ruído Glótico)   |
| GMM     | <i>Gaussian Mixture Models</i> (Modelo de Misturas Gaussianas)   |
| HNR     | <i>Harmonic-Noise Ratio</i> (Relação Ruído-Harmônico)  |
| LIEV    | Laboratório Integrado de Estudos da Voz  |
| LMS     | <i>Least Mean Square</i> (Mínimos Quadrados Médios)  |
| LPC     | <i>Linear Predictive Coding</i> (Codificação por Predição Linear)  |
| MEEI    | <i>Massachusetts Eye and Ear Infirmary</i>   |
| MLPNN   | <i>Multilayer Perceptron Neural Networks</i> (Rede Neural Perceptron Multicamadas)                                       |
| Nadam   | <i>Nesterov-accelerated Adaptive Moment Estimation</i> (Estimativa de <i>momentum</i> adaptativo acelerado por Nesterov) |

|       |  |
|-------|--|
| RBFNN | <i>Radial Basis Function Neural Network</i> (Rede Neural de Função de Base Radial) |
| ReLU  | <i>Rectified Linear Unit</i> (Unidade Linear Retificada)                           |
| RP    | <i>Recurrence Plots</i> (Gráfico de Recorrência)                                   |
| Rprop | <i>Resilient Backpropagation</i> (Retropropagação Resiliente)                      |
| SGD   | <i>Stochastic Gradient Descent</i> (Gradiente Descendente Estocástico)             |
| SVD   | <i>Saarbruecken Voice Database</i>   |
| SVM   | <i>Support Vector Machine</i> (Máquina de Vetor de Suporte)                        |
| UFPB  | Universidade Federal da Paraíba  |
| VP    | Verdadeiro Positivo  |
| VN    | Verdadeiro Negativo  |

# SUMÁRIO

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>  | <b>13</b> |
| 1.1      | Formulação do problema   | 13        |
| 1.2      | Justificativa  | 15        |
| 1.3      | Objetivos  | 16        |
| 1.3.1    | Geral  | 16        |
| 1.3.2    | Específicos  | 16        |
| 1.4      | Organização do trabalho  | 16        |
| <b>2</b> | <b>SISTEMA DE PRODUÇÃO VOCAL E ESTADO DA ARTE</b>                                  | <b>17</b> |
| 2.1      | Anatomia e fisiologia do sistema de produção da fala                               | 17        |
| 2.2      | Laringe  | 18        |
| 2.3      | Patologias Laríngeas   | 20        |
| 2.3.1    | Edema de Reinke  | 21        |
| 2.3.2    | Pólipos nas pregas vocais  | 21        |
| 2.3.3    | Leocoplasia da prega vocal   | 22        |
| 2.3.4    | Carcinoma na laringe   | 23        |
| 2.3.5    | Paralisia das pregas vocais  | 23        |
| 2.4      | Análise acústica de sinais de voz  | 25        |
| 2.4.1    | Estado da arte   | 29        |
| <b>3</b> | <b>REDES NEURAIS ARTIFICIAIS</b>   | <b>33</b> |
| 3.1      | Neurônio Perceptron  | 33        |
| 3.2      | Função de ativação   | 35        |
| 3.3      | Rede Neural Perceptron de Múltiplas Camadas  | 36        |
| 3.4      | Regras de aprendizagem   | 37        |
| 3.4.1    | <i>Stochastic Gradient Descent</i> - SGD   | 38        |
| 3.4.2    | <i>Adaptive Gradient</i> - Adagrad   | 38        |
| 3.4.3    | Adadelta   | 39        |
| 3.4.4    | Algoritmo RMSprop  | 40        |
| 3.4.5    | <i>Adaptive Moment Estimation</i> - Adam   | 40        |
| 3.4.5.1  | Algoritmos Adamax e <i>Nesterov-accelerated Adaptive Moment Estimation</i> - Nadam | 41        |
| <b>4</b> | <b>MATERIAIS E MÉTODOS</b>   | <b>43</b> |
| 4.1      | Introdução   | 43        |
| 4.2      | Base de dados  | 44        |
| 4.3      | Metodologia  | 45        |
| 4.3.1    | Pré-processamento  | 46        |
| 4.3.1.1  | Filtragem  | 46        |
| 4.3.1.2  | Segmentação  | 47        |

|          |  |           |
|----------|--|-----------|
| 4.3.1.3  | Escalonamento dos dados . . . . .  | 47        |
| 4.3.2    | Arquitetura do Classificador . . . . .   | 48        |
| 4.3.3    | Treinamento e teste do classificador . . . . .   | 49        |
| 4.3.4    | Avaliação do classificador . . . . .   | 51        |
| 4.3.5    | Teste de significância . . . . .   | 54        |
| 4.3.6    | Implementação do classificador . . . . .   | 54        |
| <b>5</b> | <b>RESULTADOS E DISCUSSÃO . . . . .</b>  | <b>56</b> |
| 5.1      | Resultados da discriminação entre vozes saudáveis e patológicas . . . . .  | 56        |
| 5.1.1    | Comparação com trabalhos do estado da arte . . . . .   | 58        |
| 5.2      | Resultados da discriminação entre vozes saudáveis e afetadas por patologias organofuncionais . . . . .   | 60        |
| 5.3      | Resultados da discriminação entre vozes saudáveis e afetadas por paralisia da pregas vocais . . . . .  | 62        |
| 5.4      | Resultados da discriminação entre vozes afetadas por patologias organofuncionais e paralisia das pregas vocais . . . . .   | 65        |
| 5.5      | Discussão . . . . .  | 67        |
| <b>6</b> | <b>CONSIDERAÇÕES FINAIS . . . . .</b>  | <b>69</b> |
| 6.1      | Sugestões para trabalhos futuros . . . . .   | 70        |
|          | <b>REFERÊNCIAS . . . . .</b>   | <b>71</b> |
|          | <b>APÊNDICE A – CURVAS DE APRENDIZAGEM DURANTE O TREINAMENTO DOS MELHORES CLASSIFICADORES . . . . .</b>  | <b>79</b> |
|          | <b>ANEXO A – TABELA DE REFERÊNCIA PARA ESTIMAR O VALOR DE <math>F_{critico}</math> PARA O TESTE DE SIGNIFICÂNCIA ANOVA PARA UM INTERVALO DE CONFIANÇA DE 95% . . . . .</b> | <b>80</b> |

# 1 INTRODUÇÃO

Dentre os principais mecanismos naturais de comunicação utilizados pelo ser humano, a fala é o que mais apresenta informações socioemocionais do indivíduo (CIELO et al., 2015). A comunicação por meio da fala diferencia claramente o ser humano dos outros animais e, através dela, são transmitidas informações das emoções, da saúde e bem estar do indivíduo.

A ação dos diferentes grupos musculares presentes em todo o sistema de formação e transmissão natural da voz é comprometida pelo surgimento de alterações fisiológicas ocasionadas por patologias vocais capazes de impedir, comprometer ou dificultar a produção da voz (MENDOZA et al., 2003).

Dentre os tipos de disfunções provocadas por efeitos patológicos, as originadas de patologias laríngeas alteram o padrão da vibração das pregas vocais provocando modificações no timbre, na frequência, na amplitude e na composição dos harmônicos presentes na voz. Conseqüentemente, estas mudanças afetam a qualidade vocal do indivíduo, comprometendo tanto os aspectos comunicativos por meio da fala, como também a sua qualidade de vida devido aos desconfortos causados pela doença (QUEIROZ et al., 2018).

## 1.1 Formulação do problema

Existe uma grande variedade de doenças relacionadas à laringe que causam modificações na voz, podendo ser tanto de origem orgânica (organofuncionais), quanto de origem neurológica. As lesões laríngeas organofuncionais como os nódulos, os pólipos e edema de Reinke, são consideradas afecções benignas decorrentes principalmente do comportamento vocal inadequado e a hábitos sociais não saudáveis como o tabagismo e o alcoolismo (CIELO et al., 2011). As patologias laríngeas de origem neurológica, como a paralisia das pregas vocais, decorrem de lesões do nervo vago ou de seus ramos que interligam o tronco cerebral à laringe, podendo levar a alterações das funções que requerem o fechamento glótico (STEFFEN et al., 2004). No caso de paralisia, as principais causas incluem traumas cirúrgicos ou acidentais, doença cardiovascular e doenças neurológicas (STEMPLE, KLABEN, 2010).

A avaliação perceptivo-auditiva é considerada a técnica comumente utilizada pelos fonoaudiólogos para análise da qualidade vocal, a fim de verificar a presença de desordens na voz (MADAZIO, LEÃO, BEHLAU, 2011; MARYN, ROY, 2012). A análise de características, julgando a qualidade vocal de forma perceptiva e intuitiva, por ser uma avaliação subjetiva, está sujeita a erros e variações, tendo em vista que, para este tipo de investigação, fatores como a experiência do avaliador, o tipo de tarefa e do protocolo seguido para a avaliação influenciam no diagnóstico final (ZRAICK, WENDEL, SMITH-OLINDE, 2005).

Outras técnicas para diagnóstico de patologias utilizam-se de instrumentos e equipamentos biomédicos aplicados na inspeção da laringe, como por exemplo a videoestroboscopia e a

videolaringoscopia (ANDRADE et al., 2018). No entanto, tratam-se de exames considerados invasivos, podendo gerar desconforto, causando resistência de alguns pacientes para que o procedimento seja realizado até o final (COUTO, 2018).

A análise acústica tem sido sugerida como uma ferramenta auxiliar não-invasiva e de baixo custo aplicadas no auxílio ao diagnóstico de patologias (COSTA et al., 2013). Consiste em extrair características a partir do sinal de voz, necessitando apenas da voz gravada do paciente e de sistemas computacionais que irão realizar o tratamento e o processamento do sinal. Vários são os parâmetros acústicos estudados nessa análise a fim de se obter dados que sejam relevantes na identificação de patologias. As medidas comumente utilizadas na prática clínica são: a frequência fundamental (*pitch*), o *jitter*, o *shimmer*, a relação harmônica-ruído (*Harmonic-Noise Ratio* - HRN) e a excitação do ruído glotal (*Glottal Noise Excitation* - GNE) (MEDINA, SIMÕES-ZERANI, NEMR, 2015).

Sinais de vozes patológicas possuem, geralmente, uma característica ruidosa, apresentando dificuldades na obtenção da frequência fundamental. Desse modo, as medidas acústicas obtidas do *pitch*, como o *jitter* e o *shimmer*, podem não ser confiáveis para alguns sinais que apresentem uma patologia considerada severa (GODINO-LLORENTE, GOMEZ-VILDA, VELASCO, 2006; COSTA, 2008). Desse modo, outras medidas que independem do *pitch* têm sido investigadas em diversos trabalhos. Algumas pesquisas anteriores se baseiam no modelo linear de produção da fala (FANT, 1970), aplicando técnicas como a análise de coeficientes de predição linear (*Linear Predictive Coding* - LPC) e a análise cepstral, para obter parâmetros capazes de distinguir vozes saudáveis de patológicas (VIEIRA et al., 2013; GRZYWALSKI et al., 2018; RABEH, SALAH, ADNANE, 2018; WOLK, WOLK, 2019). Contudo, devido as várias não linearidades envolvidas na vibração das pregas vocais, estudos mais recentes utilizam-se de técnicas de análise dinâmica não-linear no intuito de investigar os aspectos da voz humana não explorados na abordagem linear (VIEIRA, 2014; SOUZA, 2015; PINHO, 2017; QUEIROZ et al., 2018).

Apesar de diversos métodos serem apresentados na literatura, não existe uma definição da melhor técnica, característica ou parâmetro adequado para identificar a existência de patologias por meio da análise acústica do sinal de voz (TORRES, 2018). Além disso, alguns trabalhos investigam problemas quanto a normalização de medidas (FELIPE, GRILLO, GRECCI, 2006) e o potencial discriminativo que as mesmas trazem ao especialista durante a avaliação vocal realizada em laboratório (ULIZA et al., 2016; LOPES et al., 2017; KRIK et al., 2019). Diante do exposto, a definição da melhor técnica, característica ou parâmetro adequado para identificar a existência de patologias, por meio da análise acústica do sinal de voz, ainda é um campo em aberto na área científica.



## 1.2 Justificativa

Diversos pesquisadores têm dedicado esforços no desenvolvimento de técnicas eficientes para discriminar vozes patológicas usando análise acústica. Grande parte dos métodos propostos apresentados na literatura utilizam-se de uma avaliação multivariada de parâmetros extraídos do sinal de voz, sendo geralmente obtidos por análise dinâmica linear, sendo esta baseada no modelo linear de produção da fala (FANT, 1970); ou não linear com base na aplicação de técnicas originadas da teoria do caos (LETELLIER, AGUIRRE, 2009; COSTA, 2012). Desse modo, a maioria dos estudos buscam criar uma série de dados discriminativos que podem ser aplicados no treinamento de algoritmos inteligentes, como redes neurais artificiais (*Artificial Neural Networks* - ANNs), com o objetivo de desenvolver sistemas computacionais capazes de distinguir vozes saudáveis de patológicas.

Contudo, para que um algoritmo classificador consiga detectar a presença patológica, faz-se necessário que características obtidas representem significativamente as desordens provocadas por uma determinada patologia (COSTA, 2012). A escolha de características confiáveis que representem cada tipo de patologia pode se tornar uma tarefa difícil, uma vez que dependerá de vários fatores como o tipo e o grau da lesão, a severidade dos efeitos causados pela patologia na produção vocal e a quantidade de ruído presente no sinal de voz analisado.

Acredita-se que a voz pode apresentar fortes indicadores da presença de patologias que provocam problemas no processo natural da fonação humana. O especialista em análise perceptivo-auditiva escuta a voz dos seus pacientes com o intuito de detectar, pela escuta do som, a presença de desordens vocais. O seu cérebro treinado possui informações eminentes que possibilitam a tomada de decisão que indica se a voz possui ou não algum distúrbio.

A aprendizagem profunda (*Deep Learning*), por meio do uso de redes neurais profundas (*Deep Neural Networks* - DNNs), pode proporcionar o desenvolvimento de sistemas inteligentes aptos para realizar o reconhecimento de padrões característicos presentes em sinais complexos, como o da voz humana.

DNNs utilizam-se de múltiplas camadas ocultas de neurônios para criar uma representação distribuída de problemas complexos. O seu potencial despertou o interesse de pesquisadores que passaram a estudar o uso de DNNs no desenvolvimento de sistemas inteligentes capazes de auxiliar no diagnóstico médico. Diversos trabalhos abordam o uso de DNNs na detecção de patologias por meio da análise de bancos de dados com informações médicas complexas. Dentre as aplicações, tem-se a detecção de câncer (SHAH et al., 2017; QIU et al., 2018; LI et al., 2018; TURAN; BILGIN, 2019), afecções torácicas (BAR et al., 2015), doenças encefálicas (ALHUSSEIN, MUHAMMAD, HOSSAIN, 2019; SCHIRRMEISTER et al., 2017), distúrbios gastrointestinais (SRIVASTAVA et al., 2019), edema macular diabético (VAHADANE et al., 2018) e patologias vocais (HARAR et al., 2017; ALHUSSEIN; MUHAMMAD, 2018; CHUANG et al., 2018; FANG et al., 2018).

O uso de DNNs na detecção de patologias vocais pode se tornar uma ferramenta útil no auxílio de diagnóstico médico como técnica complementar. Apesar de não substituir os métodos baseados na inspeção direta das pregas vocais, a aplicação de sistemas inteligentes baseados em DNNs pode contribuir, significativamente, na redução da quantidade de exames invasivos a serem realizados pelo paciente. Além disso, a aplicação pode beneficiar positivamente o acompanhamento de tratamentos e terapias vocais.

## **1.3 Objetivos**

### **1.3.1 Geral**

Avaliação do método de discriminação entre vozes saudáveis e afetadas por patologias laríngeas com base na análise do comportamento dinâmico presente no próprio sinal de voz utilizando DNNs.

### **1.3.2 Específicos**

- Análise do uso de diferentes algoritmos de otimização para o treinamento das DNNs;
- Análise de significância dos efeitos ocasionados pelo aumento da quantidade de camadas ocultas das DNNs nos processo discriminativos.
- Validação de uma ou mais arquiteturas de DNNs aplicáveis na discriminação entre sinais de vozes saudáveis e afetados por patologias laríngeas;

## **1.4 Organização do trabalho**

No Capítulo 2 os principais temas referentes aos processo natural de produção da fala e dos efeitos ocasionados por patologias laríngeas são apresentados. Além disso, o Capítulo 2 traz uma abordagem sobre os diferentes métodos empregados na extração de medidas quantitativas utilizadas em análise acústica, bem como os resultados obtidos por pesquisadores presentes no estado da arte.

O Capítulo 3 aborda os principais tópicos sobre a estrutura de uma ANN e dos diferentes algoritmos que podem ser utilizados na etapa de treinamento destas ferramenta computacionais.

A metodologia empregada na pesquisa encontra-se no Capítulo 4. Resultados e discussão podem ser visualizados no Capítulo 5. Por fim, as considerações finais sobre a pesquisa realizadas estão presentes no Capítulo 6.

## **2 SISTEMA DE PRODUÇÃO VOCAL E ESTADO DA ARTE**

Uma compreensão do sistema de produção da fala humana é essencial no contexto desta pesquisa. Neste capítulo é realizada uma breve apresentação dos elementos anatômicos e fisiológicos do processo de produção da fala, das causas das desordens vocais e das patologias que afetam a saúde da laringe.

### **2.1 Anatomia e fisiologia do sistema de produção da fala**

O estudo da anatomia e fisiologia dos mecanismos de produção da fala é de fundamental importância para a compreensão do processo da fonação normal e patológica e, para reconhecer o impacto de lesões específicas na produção vocal. O sinal de voz oriundo do sistema natural de produção vocal humano é resultado da ação conjunta dos seguintes subsistemas (BEHLAU, 2001):

- i. Sistema Respiratório - Fonte de energia (pulmões);
- ii. Sistema Fonatório - Fonte vibratória (pregas vocais);
- iii. Sistema e Ressonância - Fonte de ressonância (cavidades oral e nasal);
- iv. Sistema Articulatório - Articuladores (língua, lábios, mandíbula, palato, dentes);
- v. Sistema Nervoso Central e Periférico - Coordenação dos demais sistemas (córtex, estruturas sub-corticais e nervos).

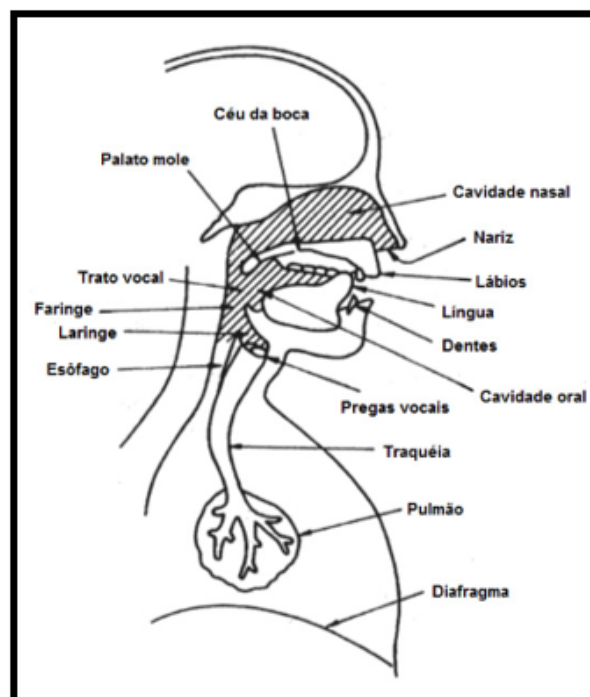
O sistema respiratório constitui a principal fonte de energia para produzir sons durante a fonação humana. A energia é proporcionada a partir do fluxo da corrente de ar que, a partir das perturbações ocorrentes até a saída do ar pela boca, proporciona que diferentes sons sejam produzidos. A fala é produzida a partir da liberação do ar dos pulmões para o trato vocal, formado basicamente por cavidades e órgãos articuladores, que começa na abertura da pregas vocais e termina nos lábios (RUFINER, 2005). A Figura 2.1 apresenta os elementos principais e secundários presentes no mecanismo de produção vocal.

Alterações estruturais decorrentes de patologias que afetam um ou mais órgãos que compõem o sistema de produção vocal ocasionam distúrbios na voz. Entre os aspectos relacionados aos distúrbios da voz estão as suas causas e consequências, no qual podem possuir diferentes origens. Dentre as possíveis causas podem ser elencadas (VIEIRA, 2014):

- Uso inadequado da voz: refere-se a comportamentos de produção vocal que não atingem um desempenho vocal eficiente. Desse modo, a voz passa a divergir do que seria considerado uma entonação normal.

- Abuso vocal: compreende um conjunto de comportamentos que podem ou não ocorrer de forma isolada, e que podem causar danos mais severos ao sistema de produção vocal do que o uso inadequado. Entre os comportamentos abusivos podem ser citados a intensidade excessiva por tempo prolongado da voz e o uso excessivo da voz em períodos de inchaço, inflamação ou patologias laríngeas sem o devido tratamento médico.
- Patologias da laringe: estão relacionadas a deformações anatômicas na região da laringe, que podem ter origem natural ou ocasionadas por maus hábitos vocais.

**Figura 2.1** – Diagrama esquemático do mecanismo de produção da fala.



Fonte: Costa (2012) (Adaptado).

Quando a voz está fora dos padrões de normalidade, diz-se que está perturbada ou disfônica, sendo este um termo mais geral para qualquer alteração na vocalização normal (DANIEL et al., 1994). Patologias da laringe afetam o padrão da vibração das pregas vocais o que conseqüentemente, ocasiona a produção de vozes desordenadas. Para compreender os efeitos que as afecções laríngeas provocam na produção vocal, é necessário o entendimento básico referente ao funcionamento deste órgão.

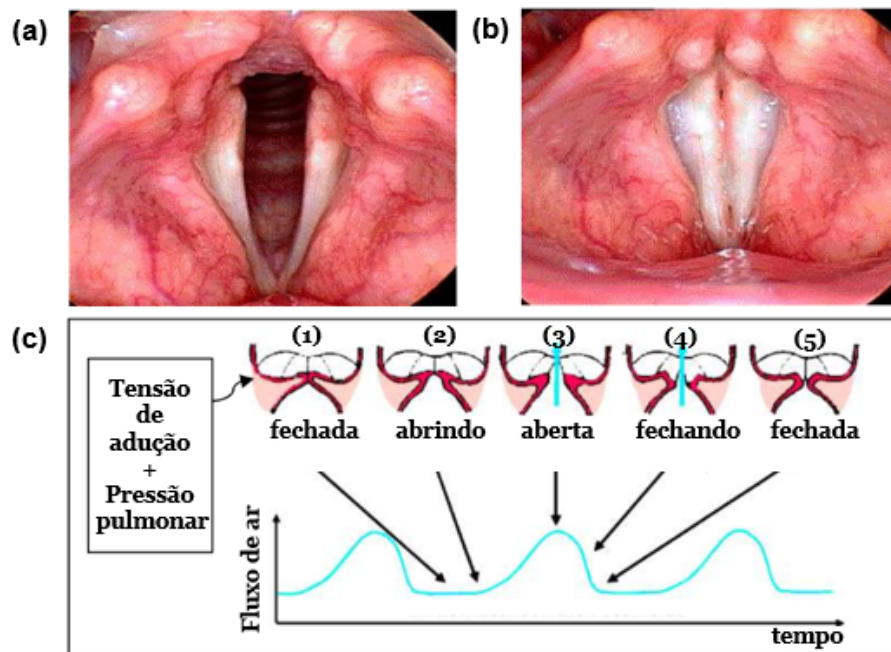
## 2.2 Laringe

A laringe é um órgão situado na garganta entre o plano superior da traqueia e o plano inferior da faringe. Dentre suas principais funções biológicas, a laringe exerce papel importante no processo respiração e fonação.

As funções básicas da laringe são: proteção das vias aéreas, respiração e fonação. A laringe protege as vias aéreas durante a deglutição funcionando como uma válvula, evitando que o alimento entre no sistema respiratório, coordena e otimiza as vias aéreas superiores durante a respiração e promove o controle da fonação por meio de mecanismos intrínsecos, em conjunto com a faringe e cavidades oral e nasal (ALMEIDA, 2010).

A fonação da laringe se realiza mediante um mecanismo que intervem das pregas vocais, das cartilagens e dos músculos laríngeos intrínsecos, dependendo também das características do fluxo de ar proveniente dos pulmões (RUFINER, 2005). O movimento de abertura e fechamento das pregas vocais, ocorrentes em intervalos semelhantes, provocam alterações no fluxo de ar que passa a ser transmitido por meio de pulsos sequenciados (pulsos glotais). Os mesmos seguem em direção às outras estruturas presentes no mecanismo de produção vocal, onde serão modulados e processados naturalmente até a formação final da fala. A Figura 2.2 ilustra, por meio de pulsos glotais, a transmissão do fluxo de ar durante o processo de abdução e adução das pregas vocais.

**Figura 2.2** – (a) - Pregas vocais (abdução); (b) - Pregas vocais (adução); (c) - Pulsos glotais.



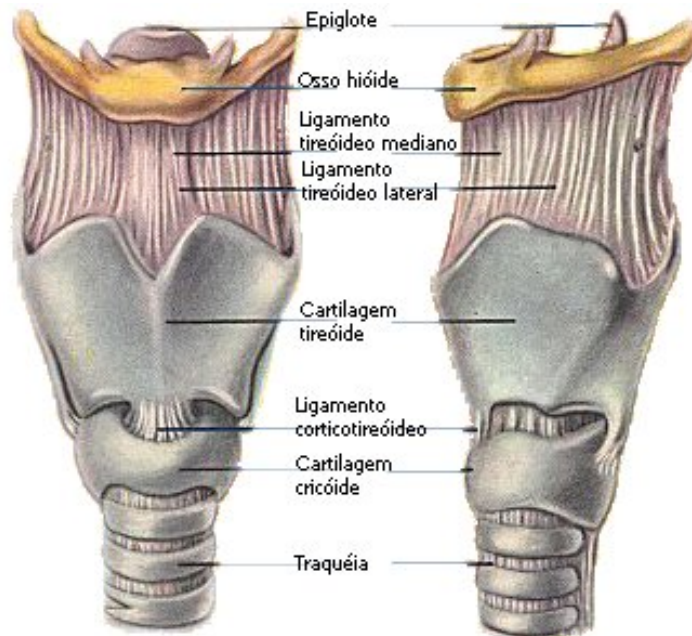
Fonte: Dias (2012) (Adaptado).

A frequência de vibração das pregas varia em torno de 113 Hz para os homens, 220 Hz para as mulheres (RUSSO e BEHLAU, 1993). Esta frequência de vibração se modifica, constantemente, devido a efeitos fisiológicos que podem ocorrer naturalmente durante a fonação, ou ocasionados por anomalias, como patologias na laringe.

A cartilagem tireoidea é a maior e mais importante cartilagem para o processo de fonação. Possuindo um formato semelhante a um escudo, esta cartilagem é formada por duas lâminas laterais (direita e esquerda) de formato quadrangular e dois pares de cornos posteriores. As lâminas possuem um ângulo de união, denominado de proeminência laríngea. Os ângulos variam de acordo com o sexo, em que entre homens observa-se ângulos em torno de 90° e em mulheres

em torno de  $120^\circ$ . Esta variação é responsável pela definição do tamanho das pregas vocais e contribui na definição da frequência vocal emitida (BEHLAU, 2001; COSTA, 2012). A Figura 2.3 ilustra a estrutura anatômica de uma laringe humana e seus elementos constituintes.

**Figura 2.3** – Estrutura anatômica da laringe.



Disponível em: <http://www.studiomel.com/20.html>

Diversos fatores podem contribuir com o surgimento de distúrbios no sistema vocal. Dentre eles, as patologias laríngeas são as que mais comprometem o funcionamento das pregas vocais (fonte vibratória). Uma abordagem sobre tais patologias é apresentada a seguir.

## 2.3 Patologias Laríngeas

As alterações vocais decorrentes de patologias laríngeas causam grande impacto no contexto social e profissional do ser humano, interferindo na qualidade de vida, sendo este um fator importante, pois compromete o bem estar físico e psicológico do indivíduo (SILVA, 2018).

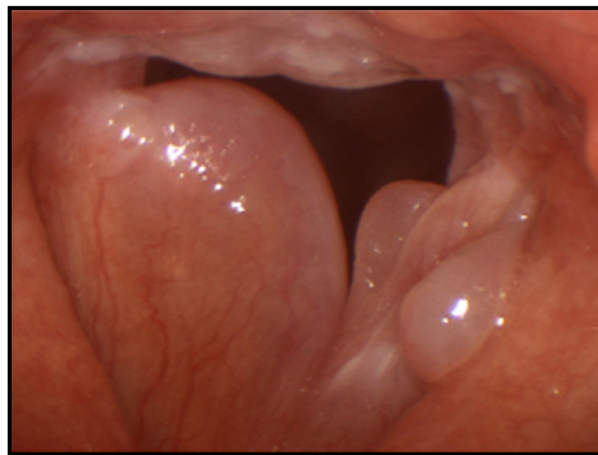
Patologias laríngeas provocam modificações orgânicas e funcionais das pregas vocais, comprometendo as camadas mucosas em massa e rigidez. Estas alterações afetam o padrão de vibração das pregas vocais resultando em distorções na intensidade da fonação (COSTA, 2008).

Em geral, as patologias que ocasionam distúrbios na laringe estão relacionadas a distúrbios organofuncionais ou neurológicos. Nesta pesquisa, serão abordadas as patologias de origem orgânica: edema de Reinke, pólipos, leucoplasia e carcinoma; além da patologia paralisia das pregas vocais de origem neurológica.

### 2.3.1 Edema de Reinke

O edema de Reinke (Figura 2.4) caracteriza-se por um processo inflamatório crônico que acomete a camada superficial da lâmina própria da prega vocal, sendo esta expandida por um muco espesso conferindo-lhe aspecto gelatinoso. Dentre os fatores que podem influenciar na aparição desta patologia, o tabagismo e o abuso vocal são, atualmente, os maiores propulsores desta doença. O edema de Reinke apresenta semelhanças histopatológicas com os nódulos vocais. Entretanto, essas lesões possuem diferenças endoscópicas e mecanismos fisiopatológicos distintos (MARTINS et al., 2009a; MARTINS et al., 2009b; CAVICHIOLO et al., 2019).

**Figura 2.4** – Edema de Reinke.



Disponível em: <http://voicemedicine.com/>.

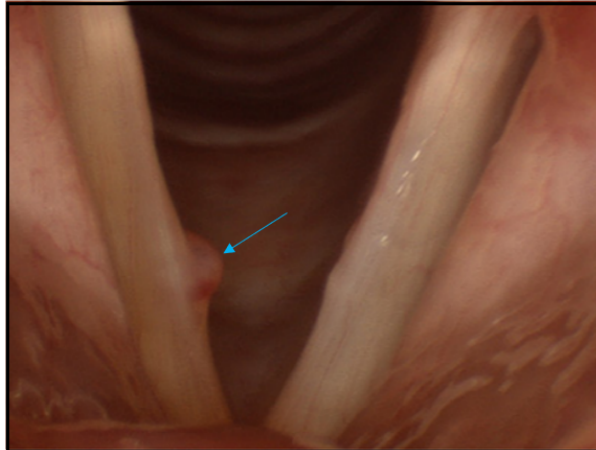
A rugosidade e a diminuição da frequência fundamental da vibração das pregas vocais também são sintomas recorrentes do edema de Reinke. Tais fatores provocam uma modificação na entonação da voz para um aspecto rouco e áspero. O diagnóstico pode ser feito através da laringoscopia indireta, por meio da videolaringoscopia, ou pode estar imediatamente evidente nos casos moderados ou avançados (ALMEIDA, 2010).

### 2.3.2 Pólipos nas pregas vocais

A literatura atual afirma que a principal etiologia dos pólipos nas pregas vocais é o fonotrauma súbito, que acarreta a formação de hematoma no espaço de Reinke podendo gerar um processo inflamatório, que estaria relacionado com a formação do pólipo vocal (ECKLEY et al., 2008).

Os pólipos nas pregas vocais (Figura 2.5) ocorrem, normalmente, na parte anterior ou média da prega vocal membranosa e constituem a patologia laríngea, que mais comumente exige remoção cirúrgica, sendo duas vezes mais comuns em homens do que em mulheres. Além disso, os pólipos vocais são predominantemente unilaterais, ou seja, surgem apenas em um lado das pregas vocais (BENJAMIN, 2000).

**Figura 2.5** – Pólipo unilateral.



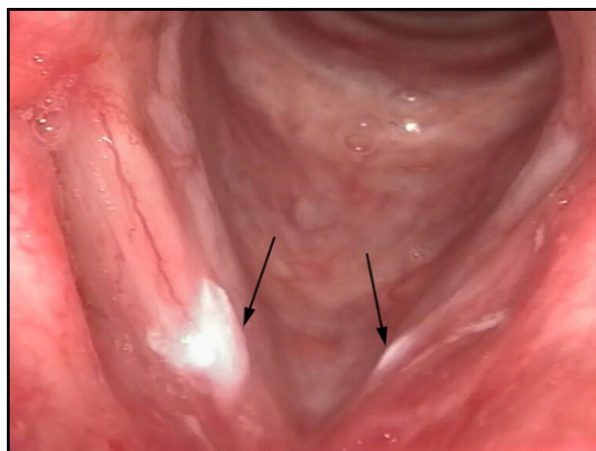
Fonte: Sulica (2013).

Dentre seus principais sintomas tem-se a rouquidão crônica e voz sussurrada, que tendem a se desenvolver ao longo de dias a semanas. Assim como outras patologias laríngeas organofuncionais, os pólipos vocais são comumente diagnosticados com base em exames de inspeção direta da laringe, como a laringoscopia e videolaringoscopia (CECATTO et al., 2002).

### 2.3.3 Leucoplasia da prega vocal

A leucoplasia da prega vocal trata-se de uma lesão formada por uma mancha branca espessa que se forma nas pregas vocais. Tal patologia é considerada uma lesão pré-maligna da laringe, ou seja, possui tendência a se transformar em carcinoma invasivo. O principal sintoma é a rouquidão, que tem caráter progressivo, podendo ocorrer ardor e sensação de corpo estranho na garganta, dor ao engolir e fadiga ao falar.

**Figura 2.6** – Leucoplasia da prega vocal.



Fonte: Chen et al. (2019).

A endoscopia com geração de imagens de banda estreita é uma das ferramentas modernas aplicadas para distinguir padrões benignos e malignos na leucoplasia das pregas vocais. Contudo,

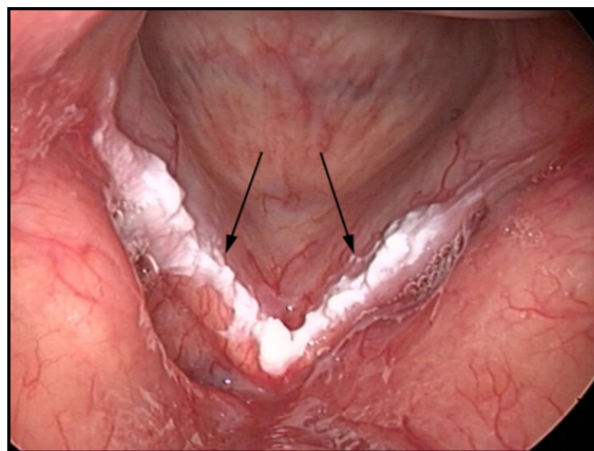


a laringoscopia normal sob luz branca é a ferramenta mais comum para a leucoplasia de prega vocal (CHEN et al., 2019).

### 2.3.4 Carcinoma na laringe

Mais frequente em homens, essa doença está intimamente ligada ao tabagismo e ao consumo de álcool, surgindo comumente entre os 40 e 70 anos de idade. Tendo como sintoma clássico a disfonia, o diagnóstico precoce do carcinoma na laringe é muito importante, uma vez que trata-se de um fator decisivo no êxito do tratamento (KUHL, 1982; CHEN et al., 2019).

**Figura 2.7** – Carcinoma com característica escamosa presente nas pregas vocais.



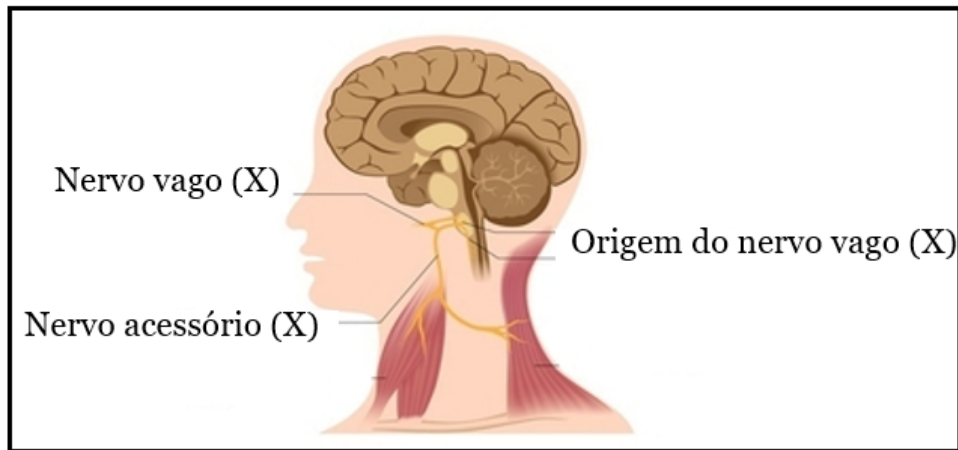
Fonte: Chen et al. (2019).

Muitas vezes, o tumor é tratado com métodos inadequados ou confundido com lesões benignas (edema, pólipos, nódulos) ou doenças comuns da laringe (KUHL, 1982). Os métodos clínicos mais utilizados para a avaliação da evolução do câncer de laringe, para definição da rapidez do crescimento e da presença ou não de metástases, são a graduação histológica e o estadiamento, importantes para a definição do prognóstico e do tratamento a ser realizado (MACIEL, LEITE e SOARES, 2013).

### 2.3.5 Paralisia das pregas vocais

A paralisia de prega vocal pode decorrer do surgimento de distúrbios cerebrais ou da lesão do nervo vago e de seus ramos, podendo levar a alterações das funções que requerem o fechamento glótico. O nervo vago (Figura 2.8) é o nervo craniano com o trajeto mais longo, estendendo-se do tronco cerebral ao abdome. Origina-se de quatro núcleos no bulbo, sendo que três deles convergem na cisterna basal para formar um único nervo, que emerge do crânio pelo forame jugular, percorre o pescoço e o tórax, terminando no abdome. Neste longo trajeto, o nervo vago dá origem a numerosos ramos que inervam a laringe e a faringe (STEFFEN et al., 2004; GARCIA et al., 2009)

**Figura 2.8** – Ilustração do nervo vago e seus ramos.



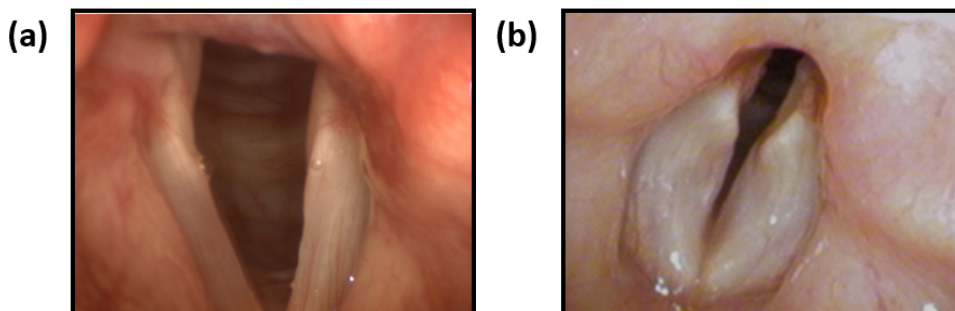
Disponível em: <https://www.tuasaude.com/nervo-vago/#articleSurvey> (Adaptado).

Os pacientes com paralisia das pregas vocais geralmente apresentam queixa de rouquidão. Outros sintomas mais graves incluem aspiração e pneumonias frequentes. A paralisia pode ocorrer de forma aguda ou crônica, podendo ser uni ou bilateral (SOUZA, 2015):

- Paralisia Unilateral: a prega vocal afetada não pode mover-se para a linha média e ajudar no fechamento da glote;
- Paralisia Bilateral: os dois lados tem espessura com aspecto reduzido e não conseguem mover-se totalmente para a linha média.

A paralisia unilateral (Figura 2.9 - **(a)**) não é uma afecção rara e pode ser resultado de lesão periférica que envolve o nervo vago acima da saída do nervo laríngeo superior. Paralisia bilateral (Figura 2.9 - **(b)**) de pregas vocais é uma apresentação rara, manifestando-se com disfonia, disfagia, e aspiração de alimentos e saliva para as vias aéreas inferiores. Algumas doenças cursam com essa manifestação, como patologias congênicas, neurológicas, neuromusculares, traumáticas e tumorais (DUARTE et al., 2006).

**Figura 2.9** – Paralisia nas pregas vocais : **(a)** - Unilateral; **(b)** - Bilateral



Disponível em: <http://voicemedicine.com/>.

Dentre as diferentes avaliações clínicas empregadas no diagnóstico de patologias laríngeas, a análise acústica caracteriza-se por se tratar de uma metodologia não-invasiva, que pode contribuir com a redução da quantidade de exames invasivos realizados pelo paciente. Uma abordagem sobre análise acústica aplicada a sinais de voz será apresentada a seguir.

## **2.4 Análise acústica de sinais de voz**

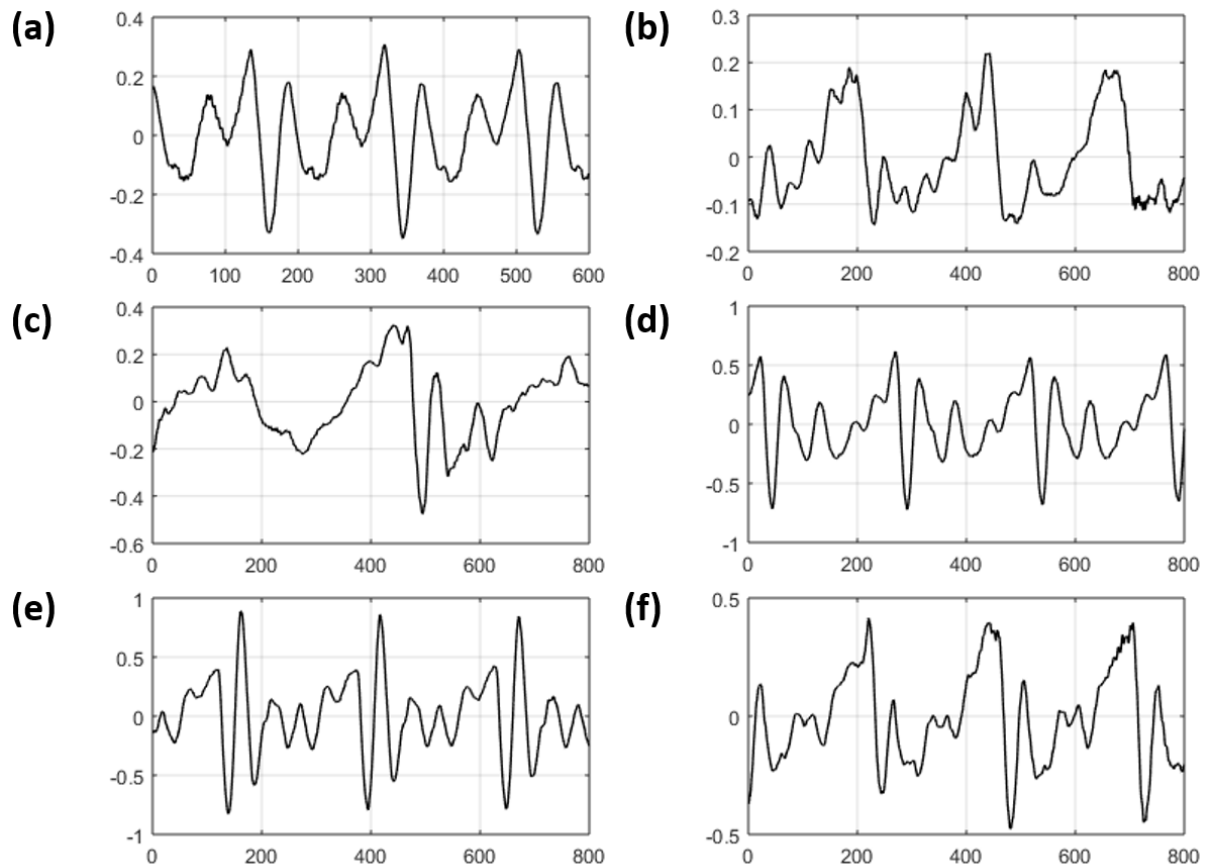
O sinal de voz é uma onda sonora emitida pelo processo natural da fonação humana que após ser digitalizada passa a ser representada por um conjunto de amostras sequenciadas em um intervalo finito de tempo, tornando-se uma série temporal discreta que traz consigo informações sobre o sistema de produção vocal (VIEIRA, 2014) (ASSEF et al., 2009).

Uma série temporal é qualquer conjunto de observações ordenadas no tempo (MORRETTIN; TOLOI, 2006). A análise de séries temporais possibilita a extração de informações relevantes sobre o comportamento dos dados observados. Dentre os principais objetivos que impulsionam o estudo do comportamento presente nestas amostras, podem ser citados (MORRETTIN; TOLOI, 2006):

- i. investigar o mecanismo gerador da série temporal;
- ii. fazer previsões de valores futuros;
- iii. descrever o comportamento da série;
- iv. procurar periodicidades relevantes nos dados;

Quando afetadas por patologias laríngeas, as pregas vocais passam a ter seu processo natural de vibração modificado resultando em alterações no sinal de voz emitido pelo indivíduo e, conseqüentemente, provocam alterações na dinâmica presente na sua respectiva série discreta, conforme ilustrado na Figura 2.10. Diante do sinal de voz digitalizado, a análise acústica realiza a extração de características do sinal de voz inerentes ao sistema de produção vocal, com o objetivo de obter valores numéricos capazes de quantificar as alterações vocais (COLTON; CASPER, LEONARD, 2011).

**Figura 2.10** – Comportamento dinâmico de sinais de vozes: **(a)** - sinal saudável; **(b)** - afetado por carcinoma; **(c)** - afetado por edema; **(d)** afetado por leucoplasia; **(e)** afetado por paralisia; **(f)** afetado por pólipos.



Fonte: Acervo Pessoal.

Tais características representativas do sinal de voz podem ser obtidas, tanto no domínio do tempo, quanto no domínio da frequência. Dentre elas podem ser citadas:

- **Frequência fundamental ( $f_0$ ):** é definida como o número de vibrações por segundo das pregas vocais.
- **Jitter:** é definido como o índice de perturbação da frequência fundamental em curto intervalo de tempo. Expressa o quanto o valor da frequência fundamental, em um período, é diferente do anterior ou de seu sucessor imediato. Altera-se, principalmente, com a falta de controle da vibração das pregas vocais (BEHLAU, 2013; CAVALCANTE, 2018).
- **Shimmer:** o *shimmer* é definido como o índice de perturbação da amplitude em curto intervalo de tempo. Indica a variação irregular da amplitude da onda sonora, de forma consecutiva, nos intervalos de observação do sinal de voz (BEHLAU, 2013; CAVALCANTE, 2018).
- **Autocorrelação:** trata-se da correlação cruzada de um sinal com ele próprio, a fim de encontrar padrões de repetição (LIU, WANG, 2002).

- **Espectrograma:** são gráficos que analisam dinamicamente a densidade espectral de energia de um sinal, sendo esta uma característica obtida no domínio da frequência. Mostra a intensidade por meio do escurecimento ou coloração do traçado, as faixas de frequência no eixo vertical e o tempo no eixo horizontal. Sua representação mostra estrias horizontais, denominadas harmônicos. O espectrograma demonstra visualmente as características acústicas da emissão, porém essas informações exigem interpretação por parte do avaliador (VALENTIM, CÔRTEZ, GAMA, 2010).
- **Transformada *wavelet*:** a transformada *wavelet* contínua permite uma análise dos sinais de voz por meio de escalogramas, uma representação tempo-frequência do sinal. Através da transformada *wavelet* discreta, características como energia, entropia e expoente de Hurst podem ser usadas para representar os sinais analisados em várias resoluções diferentes (SOUZA, 2015).
- **Coefficientes cepstrais:** com a análise cepstral permite-se trabalhar com o sinal da glote (excitação) separadamente das repercussões ressonantes do trato vocal, facilitando o entendimento das modificações que ocorrem nas pregas vocais. A aplicação dessa técnica identifica modelos no domínio da frequência para o sinal de voz. Os coeficientes cepstrais que compõem tais modelos fornecem uma boa medida das diferenças na envoltória espectral dos segmentos de voz em análise. Alguns dos tipos de coeficientes cepstrais apresentados na literatura são: coeficientes delta-cepstrais, coeficientes cepstrais ponderados, coeficientes delta cepstrais ponderados. Além disso, coeficientes podem ser extraídos a partir da denominada escala Mel-Cepstral. Na análise acústica, cada um destes coeficientes, bem como o conjunto de coeficientes, podem ser utilizados como características capazes de discriminar sinais de vozes saudáveis de patológicos (COSTA, 2008; KOOLAGUDI; RASTOGI; RAO, 2012).
- **Coefficientes de predição linear:** proposto originalmente por Fant na década de 1970, o modelo linear de produção da fala, também conhecido como teoria fonte-filtro, estuda a fonação dentro dos aspectos de sinais e sistemas, em que as pregas vocais são consideradas a fonte sonora e o trato vocal, o filtro, matematicamente representado por um preditor linear. O método de predição por codificação linear (*Linear Predictive Coding* - LPC), consiste em estimar os coeficientes que compõem a estrutura polinomial do filtro que modela o sistema dinâmico vocal. Para este caso, os coeficientes calculados podem ser utilizados como um conjunto de características extraídas do sinal de voz com a proposta de que os mesmos sejam capazes de servir para a análise de alterações vocais, assim como as medidas *jitter*, *shimmer*, frequência fundamental, entre outros (FANT, 1970; MAGI, 2009; DIAS, 2019).
- **Coefficientes com base na análise não linear:** devido aos aspectos não lineares presentes no sistema fonatório humano, medidas extraídas com base na análise dinâmica não linear

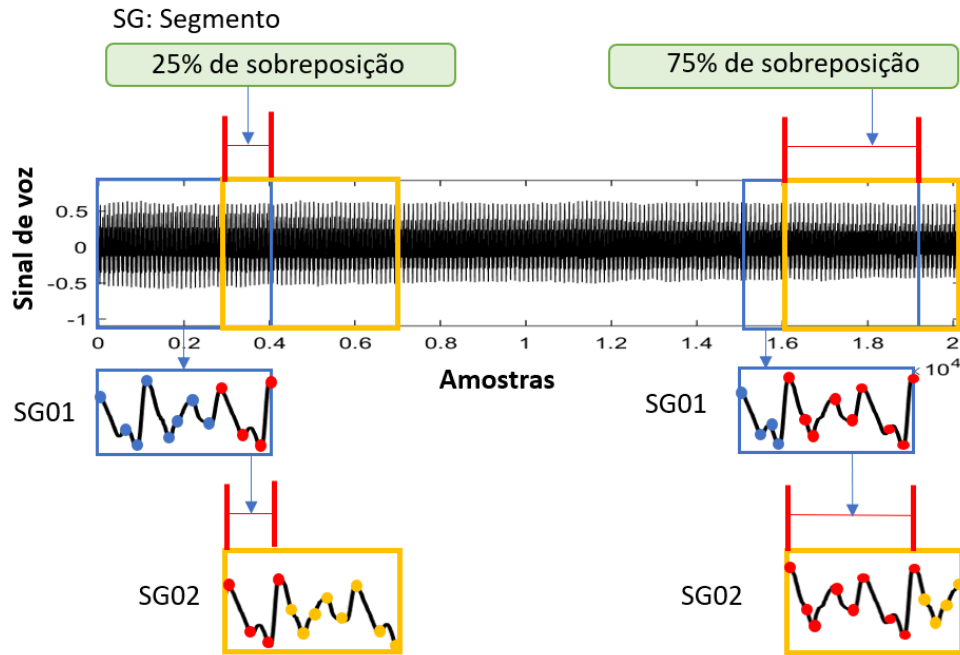
mostram-se também como uma alternativa interessante para a análise acústica de sinais de voz. A dimensão de correlação, a entropia de correlação e o maior expoente de Lyapunov, além da entropia de Shannon, a entropia de Tsallis e o expoente de Hurst são algumas das medidas quantitativas obtidas por meio da teoria do caos apresentadas na literatura, que podem ser aplicadas em uma análise discriminativa entre sinais de vozes saudáveis e patológicos (COSTA, 2012; COSTA et al., 2013).

- **Medidas de Recorrência:** recentemente, outro método baseado em dinâmica não linear de dados tem sido usado na análise acústica: os gráficos de recorrência (*Recurrence Plots* - RP). Recorrência é uma propriedade fundamental dos sistemas dinâmicos dissipativos e a principal vantagem dos gráficos de recorrência é a possibilidade do uso dessa técnica para análise de séries curtas e não-estacionárias. Por meio da análise de RPs, medidas quantitativas como a taxa de recorrência, determinismo, entropia, comprimento máximo das estruturas diagonais, tendência, liminaridade e comprimento máximo das estruturas verticais apresentam propriedades de potencial para realizar análises de alterações vocais (COSTA, 2012; VIEIRA, 2014).

As características podem ser extraídas unicamente para todo o sinal de voz analisado ou para segmentos (quadros) originados do sinal de voz. A segmentação (Figura 2.11) consiste em fracionar o sinal de voz em uma série de partições menores, sendo estas selecionadas por janelas ou quadros de duração previamente definidas. Uma série temporal é dita estacionária quando ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável (AGUIRRE, 2007). Um sinal de voz apresenta-se como uma série temporal não estacionária devido aos efeitos não lineares decorrentes dos elementos presentes no sistema humano de produção vocal (COSTA, 2012).

Para amenizar os efeitos decorrentes de tais fenômenos não lineares, os segmentos são escolhidos dentro dos limites de estacionariedade do sinal (16 ms - 32 ms) (RUFINER, 2005). É recomendado que os quadros possuam uma taxa de sobreposição entre os segmentos para evitar efeitos da descontinuidade entre os mesmos (FECHINE, 2000).

**Figura 2.11** – Exemplificação do processo de segmentação aplicada a sinais de voz.



Fonte: Acervo Pessoal.

### 2.4.1 Estado da arte

No Quadro 2.1 são apresentadas de forma sucinta as informações de pesquisas que utilizaram técnicas de extração de características aplicadas a sinais de voz junto a algoritmos classificadores a fim de discriminar vozes saudáveis de patológicas.

**Quadro 2.1** – Resumo de artigos que utilizaram extração de características e algoritmos classificadores aplicados na discriminação entre vozes saudáveis e patológicas.

| <b>Autoria</b>           | <b>Características Extraídas</b>  | <b>Metodologia</b>  | <b>Resultados</b>   |
|--------------------------|---|---|---|
| Costa (2008)             | Coeficientes LPC e cepstrais  | Base de dados: MEEI, sendo selecionados 53 sinais de vozes saudáveis, 44 afetadas por edema de <i>Reinke</i> , 23 afetadas por outras patologias (nódulos, pólipos e paralisia). Pré-classificação paramétrica e modelos de Markov escondidos de densidades discretas | Os melhores resultados foram obtidas na discriminação entre vozes saudáveis e afetadas por edema de <i>Reinke</i> com acurácia de 100%. Não utilizou validação cruzada. |
| Fonseca e Pereira (2009) | Aplicação de transformada Wavelets e extração da característica <i>jitter</i> para cada subbanda. | A base de dados foi construída na pesquisa, sendo composta por 30 sinais de vozes saudáveis, 25 afetadas por nódulos vocais e 16 afetadas por edema de <i>Reinke</i> . Classificação utilizando SVM.  | Os melhores resultados foram obtidas na discriminação entre vozes saudáveis e afetadas por nódulos vocais com acurácia de 90%. Não utilizou validação cruzada.          |

Continuação do Quadro 2.1.

| <b>Autoria</b>                  | <b>Características Extraídas</b>   | <b>Metodologia</b>   | <b>Resultados</b>   |
|---------------------------------|--|--|---|
| Costa (2012)                    | Dimensão de correlação; entropia de correlação; entropia de Shannon; entropia aproximada; entropia de Tsallis; maior expoente de Lyapunov; primeiro mínimo da função mútua; expoente de Hurst; determinismo; comprimento máximo das linhas diagonais (RP); entropia de Shannon da distribuição das linhas diagonais (RP); tendência (RP); laminaridade; tempo de permanência (RP); comprimento máximo das linhas verticais (RP); coeficientes LPC. | Base de dados: MEEI, sendo selecionados 53 sinais de vozes saudáveis, 44 afetados por edema, 55 afetados por paralisia e 18 afetados por nódulos vocais. Classificação com base em análise discriminante linear e quadrática.  | Os melhores resultados foram obtidas na discriminação entre vozes saudáveis e afetadas por nódulos vocais com acurácia em torno de 99%. Utilizou validação cruzada.                                     |
| Vieira <i>et al.</i> (2013)     | Coefficientes LPC  | Discriminação entre vozes saudáveis e patológicas. Base de dados: MEEI, sendo selecionados 53 sinais de vozes saudáveis e 114 patológicos. Classificador: <i>multilayer perceptron neural networks</i> (MLPNN).  | Foram obtidas taxas de acurácia em torno de 98% na discriminação entre vozes saudáveis e patológicas.   |
| Sellam e Jagadeesan (2014)      | <i>jitter</i> ; <i>shimmer</i> ; energia; frequência fundamental; formantes; resíduo médio quadrado do sinal.  | Sistema para classificar vozes infantis entre saudáveis e patológicas. Base de dados própria contendo 10 sinais de vozes infantis saudáveis e 10 patológicos. Análise comparativa entre dois classificadores: máquina de vetor de suporte ( <i>Support Vector Machine</i> - SVM) e redes neurais de função de base radial. | Melhores resultados: SVM - Acurácia de 83%; RBFNN - Acurácia de 91%. Não utilizou validação cruzada.  |
| Moon e Lee (2015)               | <i>Jitter</i> e <i>shimmer</i>   | Avaliação de análise acústica na detecção de vozes patológicas. Base de dados: MEEI, desta base foram utilizados 53 sinais de vozes saudáveis e 173 patológicas. Classificador: árvores de decisão.  | Melhor resultado: Acurácia - 83,11%. Não utilizou validação cruzada.  |
| Hammami, Salhi e Labidi (2016)  | frequência fundamental, LPC e frequência de formantes  | Discriminação entre vozes saudáveis e patológicas. Base de dados desenvolvida na pesquisa, sendo composta por 80 sinais saudáveis e 80 patológicos. Classificador: SVM.  | Melhores resultados: Acurácia - 86%.  |
| Belhaj, Bouzid e Ellouze (2016) | Frequência fundamental, quociente de abertura, coeficientes mel-cepstrais e energia.   | Análise acústica aplicada na discriminação entre vozes femininas saudáveis e patológicas. Base de dados: MEEI, no qual da base foram selecionados 53 sinais de vozes saudáveis e 173 patológicas. Classificador: SVM.  | Os melhores resultados na discriminação entre vozes saudáveis e patológicas foi de 99,10% de acurácia e para discriminar as vozes patológicas a acurácia foi de 79,77%. Não utilizou validação cruzada. |
| Souissi e Cherif (2016)         | Coefficientes mel-cepstrais e delta-cepstrais  | Realizou a discriminação entre vozes saudáveis e patológicas. Base de dados: SVD, sendo selecionados 50 sinais de vozes saudáveis e 70 patológicos. Classificador MLP.   | Melhores resultados para acurácia de 87,86%.  |



Continuação do Quadro 2.1.

| <b>Autoria</b>       | <b>Características Extraídas</b>   | <b>Metodologia</b>   | <b>Resultados</b>  |
|----------------------|--|--|--|
| AnkiShan (2017)      | <i>Jitter</i> , <i>shimmer</i> , maior expoente de Lyapunov, entropia, coeficientes de Skewnes e Kurtosis              | Avaliação de análise acústica na detecção de vozes patológicas. Base de dados desenvolvida na pesquisa, sendo constituída por 68 sinais de vozes saudáveis e 172 patológicos. Classificador: SVM.  | Melhor resultado: Acurácia - 93,45%. Não utilizou validação cruzada.   |
| Pinho (2017)         | Método da contagem de caixas, método da diferença, método da similaridade e o método da contagem de caixas ponderadas. | Analisar alterações causadas por patologias laríngeas e desvios vocais baseadas na dinâmica não linear. Base de dados: MEEI, com 53 sinais de vozes saudáveis e 114 patológicos; base de dados do LIEV contendo 30 sinais saudáveis e 90 desviados. Classificador MLP. | Melhores resultados de acurácia para base de dados do LIEV: Saudável $\times$ desvio soprosideade 87,33%. Foi utilizada validação cruzada.                                     |
| Fang et al. (2018)   | Coefficientes mel-cepstrais.   | Base de dados: MEEI, sendo selecionados 60 sinais de vozes saudáveis e 402 patológicas. Avaliou os classificadores: <i>Deep Neural Networks</i> - DNN, SVM, <i>Gaussian Mixture Models</i> - GMMs  | Melhores resultados obtidos para o classificador DNN de 3 camadas e 300 neurônios ocultos obtendo 99,32% de acurácia. Utilizou validação cruzada.                              |
| Chuang et al. (2018) | Coefficientes mel-cepstrais.   | A base de dados utilizada na pesquisa foi disponibilizada pelo <i>Far Eastern Memorial Hospital</i> (FEMH), na qual contém 50 sinais de vozes saudáveis e 150 sinais de vozes patológicas. Classificação utilizando DNNS   | O classificador DNN com 5 camadas ocultas com 200 neurônios demonstraram os melhores resultados em sua pesquisa, sendo obtidos 93,1% de sensibilidade e 46% de especificidade. |

Outras pesquisas estão relacionadas, assim como esta pesquisa, com a discriminação entre vozes saudáveis e patológicas sem uma ou mais características quantitativas extraídas do sinal de voz, optando pelo *deep learning*, sendo esta a principal diferença entre os trabalhos apresentados no Quadro 2.1 e o Quadro 2.2.

**Quadro 2.2** – Resumo de artigos que não utilizaram medidas quantitativas extraídas do sinal de voz na discriminação entre vozes saudáveis e patológicas.

| <b>Autoria</b>      | <b>Técnica aplicada ao sinal</b>   | <b>Metodologia</b>  | <b>Resultados</b>   |
|---------------------|--|---|---|
| Harar et al. (2017) | O método proposto pelo autor consistiu em particionar cada sinal de voz em quadros de 64ms com 30ms de sobreposição. Após o particionamento os sinais de vozes passaram a ser representados por uma matriz de dimensões $n$ segmentos $\times$ 3200 amostras do segmento (50000 Hz $\times$ 64ms = 3200 amostras). | Discriminação entre vozes saudáveis e patológicas. Base de dados: SVD, sendo selecionadas 687 sinais de vozes saudáveis e 1353 patológicas, dentre elas foram incluídos sinais afetados por carcinoma, edema, leucoplasia, pólipos e paralisia das pregas vocais. Utilizou uma rede neural convolucional ( <i>Convolutional Neural Networks</i> - CNN) para processar o sinal bidimensional obtido após aplicação da técnica proposta. O algoritmo de aprendizagem utilizado foi o Adam aplicado na minimização da função custo entropia cruzada binária (binary cross-entropy). O sistema foi validado com 15% da base de dados. | O melhor resultado foi para o classificador com 4 camadas convolucionais, 2 de <i>pooling</i> , e 3 camadas profundas, no qual obteve 71,36% de acurácia. |

## Continuação do Quadro 2.2.

| <b>Autoria</b>              | <b>Técnica aplicada ao sinal</b>  | <b>Metodologia</b>  | <b>Resultados</b>   |
|-----------------------------|---|---|---|
| Muhammad et al. (2018)      | Para cada sinal de voz a transformada rápida de Fourier foi aplicada e extraídos seus respectivos espectrogramas. | Discriminação entre vozes saudáveis e patológicas. Base de dados: SVD, sendo selecionados 687 sinais de vozes saudáveis e 853 patológicas, dentre eles foram incluídas as patologias laríngeas pólipos e paralisia das pregas vocais. Utilizou uma CNN com 4 camadas convolucionais, 1 de <i>pooling</i> e uma camada profunda com 4096 neurônios. Aplicou 70% da base para treino, 25% para validação e 5% para a validação. | Os autores apresentam como resultado uma acurácia - 98,5% para o conjunto de validação. Não apresenta a matriz de confusão.   |
| Wu et al. (2018)            | Para cada sinal de voz a transformada rápida de Fourier foi aplicada e extraídos seus respectivos espectrogramas. | Discriminação entre vozes saudáveis e patológicas. Base de dados: SVD, sendo selecionados 482 vozes patológicas e 482 saudáveis, dentre eles foram incluídas as patologias laríngeas edema, leoplasia, carcinoma, pólipos e paralisia das pregas vocais. Utilizou como classificador uma CNN com 3 camadas convolucionais e 8 profundas.  | Acurácia de 68% para a validação com 143 sinais de vozes.   |
| Alhussein e Muhammad (2019) | Para cada sinal de voz a transformada rápida de Fourier foi aplicada e extraídos seus respectivos espectrogramas. | Discriminação entre vozes saudáveis e patológicas. Base de dados: SVD, sendo selecionados 264 vozes patológicas e 686 saudáveis, dentre elas incluiu as patologias laríngeas pólipos e paralisia das pregas vocais. Também utilizou a base MEEI, porém não deixa claro a quantidade de sinais que aplicou no treino e teste. Utilizou como classificador CNNs.  | Os melhores resultados foram obtidos para o processo de classificação com 3 CNNs em paralelo. Para a base SVD, o melhor resultado foi uma acurácia de 75,4% para o teste com 30% de todo o conjunto de dados. |

Diante do exposto, diversos pesquisadores, assim como nesta pesquisa, fazem o uso de redes neurais como algoritmo classificador, a fim de realizar a discriminação de vozes saudáveis e patológicas, tanto pela utilização de características quantitativas, quanto com o uso de imagens de espectrograma. A seguir, são apresentados os principais conceitos sobre os classificadores baseados em redes neurais artificiais.

### 3 REDES NEURAIS ARTIFICIAIS

ANNs são estruturas de processamento da informação inspiradas pelo funcionamento do sistema nervoso dos animais (HAYKIN, 2001). Possuem processamento altamente paralelo e distribuído apresentando a capacidade de aprender e armazenar conhecimento experimental, sendo amplamente utilizadas na resolução de problemas de reconhecimento de padrões (ARAÚJO, FAGUNDES, BRANDÃO, 2012).

O desenvolvimento das ANNs pode ser creditado a diversos motivos, dentre os quais tem-se (TIMOSZCZUK, 2004):

- i. entender o funcionamento do sistema nervoso e, particularmente, o cérebro humano;
- ii. Simular aspectos de funcionamento do cérebro humano, para desenvolver métodos que possibilitem aos computadores desempenhar bem funções que são cotidianas para o cérebro, tais como reconhecimento de imagens, reconhecimento de voz, inteligência, aprendizado e compreensão.

Os benefícios das redes neurais se evidenciam em sua habilidade para executar computação distribuída e na sua generalização (BRAGA, 2006). De acordo com Haykin (2001), a generalização é a capacidade de apresentar saídas coerentes para entradas que não estavam presentes durante o processo de aprendizagem (treinamento). Essas duas características fazem com que as ANNs sejam capazes de solucionar problemas de maior grau de complexidade.

Outros aspectos relevantes sobre ANNs são (BRAGA, 2006):

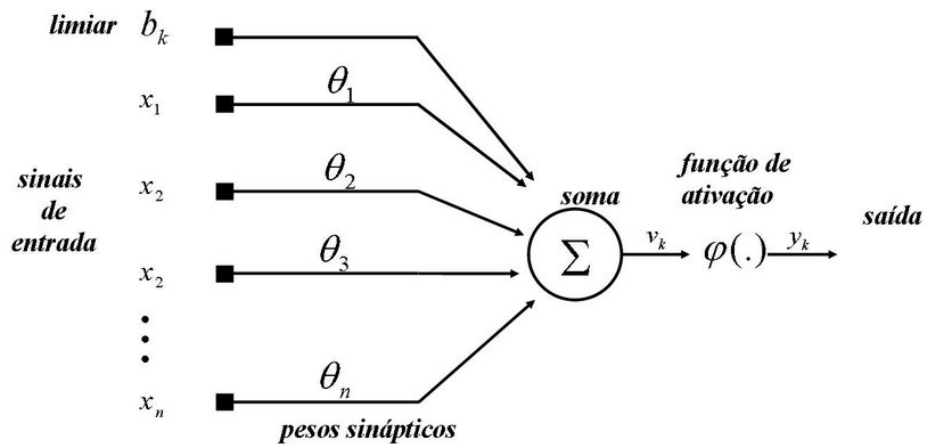
- i. possui a capacidade de trabalhar com problemas não lineares;
- ii. apresentam alto nível de adaptabilidade, ou seja, possuem a habilidade de se ajustar a novas informações.

ANNs são compostas por unidades menores chamadas de neurônios. Uma abordagem sobre a estrutura e funcionamento dos neurônios artificiais do tipo perceptron será abordada a seguir.

#### 3.1 Neurônio Perceptron

Os neurônios são unidades de processamento fundamentais de uma rede neural. O neurônio *perceptron* foi desenvolvido entre os anos de 1950 e 1960 por Frank Rosenblatt, inspirado em trabalhos anteriores realizados por Warren McCulloch e Walter Pitts (TIMOSZCZUK, 2004). O neurônio é constituído por um conjunto de conexões conhecidas como pesos sinápticos, um somador e uma função de ativação. As entradas do neurônio são ponderados pelos respectivos pesos e aplicadas ao somador. A Figura 3.1 ilustra o neurônio *perceptron* proposto por Rosenblatt (HUSH, HORNE, 1993).

**Figura 3.1** – Neurônio *perceptron*.



Fonte: Furtado, Campos Velho, Macau (2011).

O vetor  $x_n$  ( $x_1, x_2 \dots x_n$ ) é apresentado como a entrada do neurônio, o vetor  $\theta_n$  ( $\theta_1, \theta_2, \dots, \theta_n$ ) corresponde ao conjunto de pesos sinápticos que ponderam cada elemento presente em  $x_n$ , sendo este acrescido com o valor da polarização (bias)  $b$ . Posteriormente, são inseridos no combinador linear. O valor resultante da junção somatória  $v_k$  passa pela função de ativação  $\varphi(\cdot)$ , responsável por resultar um nível de atividade interna do neurônio. A resposta da função de ativação a este potencial interno corresponde a resposta final da saída  $y_n$  do neurônio.

Matematicamente, a resposta de saída do neurônio *perceptron* para o instante  $i$ , baseado em uma quantidade  $N$  de entradas, pode ser representado pelas Equações 3.1-3.2.

$$v_k[i] = \sum_{n=1}^N \theta_n[i] \cdot x_n[i] + b \quad (3.1)$$

$$y_k[i] = \varphi(v_k[i]) \quad (3.2)$$

A grande diferença entre o *perceptron* e o neurônio desenvolvido por Warren McCulloch e Walter Pitts é a presença de um algoritmo de aprendizado para ajuste dos pesos sinápticos criado com base na Regra Delta (descida do gradiente) (LIPPMANN, 1987). Os pesos do neurônio são ajustados conforme a minimização do erro local  $e(i)$ , entre a resposta de saída do neurônio  $y_k(i)$  e um valor de referência  $d_k(i)$ . Além disso, o método de ajuste utiliza um parâmetro chamado de taxa de aprendizagem  $\eta$  que, para este caso, é atribuído um valor fixo entre 0 e 1 (HAYKIN, 2001). A atualização dos pesos é realizada pela Equação 3.3:

$$\theta_n[i + 1] = \theta_n[i] + \eta \cdot e[i] \cdot x_n[i] \quad (3.3)$$

## 3.2 Função de ativação

A função de ativação  $\varphi(\cdot)$  define o modelo para um neurônio. Sem o uso de  $\varphi(\cdot)$ , o neurônio torna-se um simples elemento de regressão linear. Desta forma, possuiria as mesmas limitações presentes em modelos lineares. As funções de ativação são essenciais para dar capacidade representativa ao neurônio, introduzindo um componente de não linearidade, sendo a não linearidade diferenciável em qualquer ponto. Algumas das funções mais empregadas na ativação de neurônios, bem como as suas derivadas  $\varphi'(\cdot)$ , são definidas pelas Equações 3.4-3.11:

1. Função Sigmóide:

$$\varphi(v_k) = \frac{1}{1 + e^{v_k}} \quad (3.4)$$

$$\varphi(v_k)' = \varphi(v_k)(1 - \varphi(v_k)) \quad (3.5)$$

2. Função Tangente Hiperbólica:

$$\varphi(v_k) = \frac{e^{v_k} - e^{-v_k}}{e^{v_k} + e^{-v_k}} \quad (3.6)$$

$$\varphi(v_k)' = 1 - \varphi(v_k)^2 \quad (3.7)$$

3. Função ReLU (*Rectified Linear Unit*):

$$\varphi(v_k) = \max(0, v_k) \quad (3.8)$$

$$\varphi(v_k)' = \begin{cases} 1, & \text{se } v_k > 0 \\ 0, & \text{c.c} \end{cases} \quad (3.9)$$

4. Função ELU (*Exponential Linear Unit*):

$$\varphi(v_k, \alpha) = \begin{cases} v_k, & \text{se } v_k > 0 \\ \alpha(e^{v_k} - 1), & \text{c.c} \end{cases} \quad (3.10)$$

$$\varphi(v_k, \alpha)' = \begin{cases} 1, & \text{se } v_k > 0 \\ \varphi(v_k, \alpha) + \alpha, & \text{c.c} \end{cases} \quad (3.11)$$

em que  $\alpha > 0$ .

Equações gaussianas também podem ser utilizadas como função de ativação em algumas topologias de ANNs, como ocorre nas redes neurais com funções de base radial (*Radial Basis Function Neural Network* - RBFNN) (LU, WANG, 2010).

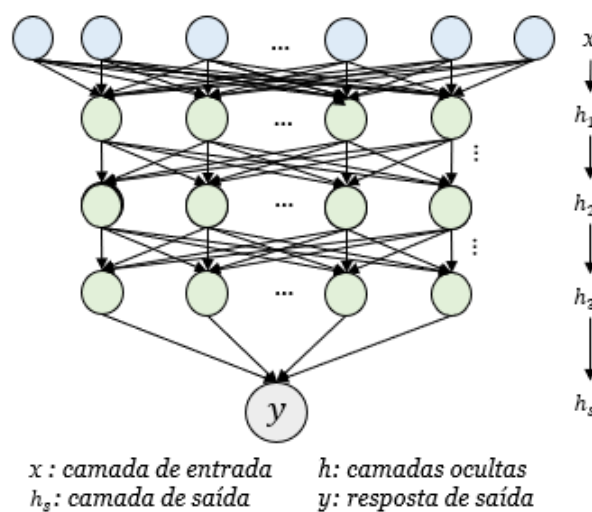
### 3.3 Rede Neural Perceptron de Múltiplas Camadas

O neurônio *perceptron* apresenta desempenho limitado na resolução de problemas mais complexos, como na classificação de padrões cujo os dados não apresentam uma distribuição linear. A forma de solucionar este problema é usar múltiplos *perceptrons* distribuídos em camadas totalmente conectadas. As redes neurais *perceptron* de múltiplas camadas (*Multilayer Perceptron Neural Network* - MLP) representam uma generalização do *perceptron* de camada única proposto por Rosenblatt. Redes MLP podem tratar dados que não são linearmente separáveis, ou seja, apresentam um poder computacional muito maior do que o uso de um único neurônio aplicado na solução de problemas não lineares (NIED, 2007).

Uma rede MLP é subdividida em camadas: camada de entrada, uma ou mais camadas ocultas e camada de saída. A camada oculta possui neurônios ocultos que intervem de maneira útil entre a camada neural de entrada e a de saída, tornando a rede capaz de realizar estatísticas de ordem elevada (BARAVIEIRA, 2016).

A camada de entrada recebe o vetor de dados  $x_n$  que serão apresentados para treinamento da rede. Cada um dos valores  $x_n$  de entrada está conectado a todos os pesos sinápticos presentes na primeira camada oculta. Uma rede neural profunda (*Deep Neural Network* - DNN) utiliza-se de múltiplas camadas compostas por neurônios ocultos totalmente conectados. O número mínimo de camadas, no qual uma DNN é considerada profunda, trata-se de um conceito controverso. Contudo, diversos pesquisadores adotam o conceito de que ANNs com mais de 03 camadas caracterizam-se como uma DNN. A Figura 3.2 ilustra a estrutura de uma DNN organizada em camadas.

Figura 3.2 – Estrutura de uma DNN.



Fonte: Acervo Pessoal.

A utilização de redes MLP, de camada única ou profundas, somente foi viabilizada com o desenvolvimento de algoritmos capazes de treinar os múltiplos neurônios presentes na(s) camada(s) oculta(s). A regra de aprendizado é o que determina o processo de atualização de todos

os pesos sinápticos durante a etapa de treinamento da rede. Desse modo, a regra de aprendizado é uma das características mais importantes de uma ANN.

### 3.4 Regras de aprendizagem

O algoritmo de retropropagação (*Backpropagation* - BP) desenvolvido por Rumelhart (1986) tornou-se efetivamente conhecido e alavancou o estudo sobre ANNs para um patamar de maior interesse entre os pesquisadores e desenvolvedores. Baseado na regra de aprendizagem por correção do erro, o algoritmo BP pode ser visto com a generalização do algoritmo Mínimo Quadrado Médio (*Least Mean Square* - LMS), também conhecido como Regra Delta (NIED, 2007).

O algoritmo BP consiste na propagação do erro da saída da rede para a sua entrada, utilizando a técnica do gradiente descendente para efetuar a correção dos pesos em função da derivada do erro com relação aos pesos anteriores (RUMELHART, MCCCELLAND, 1986).

Contudo, o algoritmo BP apresenta convergência lenta, além de estar sujeito a problemas ocorrentes por mínimos locais, onde os gradientes são nulos (LEE, DELBRUCK, PFEIFFER, 2016). Diante do exposto, outras regras de aprendizagem foram desenvolvidas no decorrer dos primeiros anos após o surgimento do BP, ambas tendo como objetivo suprir ou minimizar as deficiências ocorrentes no processo de treinamento utilizando o BP. Dentre estes algoritmos, podem ser citados (NIED, 2007): QuickProp (FAHLMAN et al., 1988), RProp (RIEDMILLER, BRAUN, 1993), ajuste da taxa de aprendizado (SILVA, ALMEIDA, 1990), e o algoritmo do gradiente conjugado (BRENT, 1991).

Atualmente, o método do gradiente descendente ainda é amplamente utilizado no treinamento de ANNs. O algoritmo BP e seus sucessores não apresentam uma grande eficiência no processo de aprendizagem de ANNs durante o processamento de banco de dados mais complexos (como áudios, imagens e vídeos), onde faz-se necessário, na maioria dos casos, o uso de redes profundas. A necessidade de algoritmos aplicados na otimização do processo de aprendizagem motivou pesquisadores a desenvolverem técnicas que apresentassem um maior rendimento durante o treinamento de DNNs, quando comparados ao desempenhos das redes treinadas pelo algoritmos considerados tradicionais.

Os estudos em otimização contribuíram significativamente no desenvolvimento de métodos e algoritmos aplicáveis em diversas áreas da computação, em busca de solucionar problemas lógicos complexos. De acordo com o estudo apresentado por Ruder (2016), dentre os algoritmos de otimização comumente utilizados em aplicações envolvendo *Deep Learning* implementadas em trabalhos realizados pela comunidade científica são: gradiente descendente estocástico (*Stochastic Gradient Descent* - SGD), RMSprop, Adagrad, Adadelata, Adam, Adamax e Nadam.

### 3.4.1 *Stochastic Gradient Descent - SGD*

Em termos de ANNs, o gradiente descendente é uma maneira de minimizar uma função custo  $J(\theta)$  parametrizada pelos pesos  $\theta$  de um modelo neural, sendo estes atualizados na direção oposta do gradiente da função custo  $\nabla_{\theta}J(\theta)$ . A atualização de  $\theta$  pelo gradiente descendente é dada pela Equação 3.12.

$$\theta = \theta + \eta \nabla_{\theta}J(\theta; X; Y), \quad (3.12)$$

em que  $X$  trata-se do conjunto de dados de entrada inseridos na rede e  $Y$  o conjunto de respostas corretas associadas ao conjunto  $X$ .

Os parâmetros  $\theta$  são computados após a apresentação de todas as amostras do conjunto  $X$  e  $Y$  e seguem na direção do gradiente com a taxa de aprendizado ( $\eta$ ) determinando o tamanho de uma atualização realizada. O método converge eficientemente para problemas de otimização convexa. No entanto, sofre com problemas de mínimos locais (valores sub-ótimos) para erros com superfícies não convexas (RUDER, 2016).

Por sua vez, a otimização por meio do algoritmo SGD realiza atualizações com base em mini-lotes. Desse modo, determina-se a quantidade de amostras que serão utilizadas para efetuar a atualização das sinapses (Equação 3.13). Esta metodologia proporciona cálculos frequentes dos pesos sinápticos provocando uma variação estocástica na descida do gradiente, a fim de minimizar a função custo ao decorrer da sequência de iterações realizadas.

$$\theta = \theta + \eta \nabla_{\theta}J(\theta; x^{[i:i+n]}; y^{[i:i+n]}) \quad (3.13)$$

### 3.4.2 *Adaptive Gradient - Adagrad*

O método de otimização Adagrad trata-se de um algoritmo baseado em gradiente descendente que utiliza uma taxa de aprendizagem adaptativa aplicada no cálculo de cada peso sináptico. Dessa forma, atualizações maiores são realizadas a medida que o gradiente se encontra mais distante do seu ponto mínimo global e ajustes mais suaves aplicados quando a descida do gradiente se encontra próximo ao mínimo global (RUDER, 2016).

Sendo  $\theta$  o conjunto de pesos sinápticos  $\{\theta^i, \theta^{i+1}, \dots, \theta^N\}$ , o cálculo do gradiente para  $\theta^i$  referente a época de treinamento  $t$ , pode ser obtido por meio da Equação 3.14.

$$g_t^i = \nabla_{\theta_t}J(\theta_t^i), \quad (3.14)$$

em que  $g_t^i$  refere-se ao gradiente calculado para o peso sináptico  $\theta^i$ ;  $i$ , o peso sináptico a ser atualizado;  $N$ , o total de pesos sinápticos do neurônio. A taxa de aprendizagem adaptativa  $\kappa_t^i$  é



definida pela Equação 3.15.

$$\kappa_t^i = \frac{\eta}{\sqrt{\alpha_t^i + \varepsilon}}, \quad (3.15)$$

sendo  $\eta$ , o valor da taxa de aprendizagem inicial;  $\varepsilon$ , uma contante para evitar uma divisão por 0, no qual o valor  $\varepsilon$  pode ser definido igual a 1e-8; e  $\alpha_t^i$ , trata-se de um coeficiente que leva em consideração o valor de  $g_t^i$  e seu antecessores, o mesmo é expresso pela Equação 3.16. Por fim, a atualização das sinapses é efetuada por meio da Equação 3.17.

$$\alpha_t^i = \sum_{j=1}^t (g_j^i)^2 \quad (3.16)$$

$$\theta_{t+1}^i = \theta_t^i - \kappa_t^i \cdot g_t^i \quad (3.17)$$

### 3.4.3 Adadelta

O algoritmo Adadelta propõe uma nova metodologia aplicada na atualização da taxa de aprendizagem adaptativa  $\kappa_t$ . Para tal, a otimização pelo método Adadelta utiliza-se de duas variáveis determinantes na computação dos pesos sinápticos, sendo elas o fator de relação ao gradiente quadrado ( $E[g^2]_t$ ) e o fator de relação à atualização sináptica ao quadrado ( $E[\Delta\theta^2]_t$ ), sendo este obtidos, respectivamente, pela Equação 3.18 e Equação 3.19.

$$E[g^2]_t^i = \gamma \cdot E[g^2]_{t-1}^i + (1 - \gamma) \cdot (g_t^i)^2, \quad (3.18)$$

$$E[\Delta\theta^2]_t^i = \gamma \cdot E[\Delta\theta^2]_{t-1}^i + (1 - \gamma) \cdot (\Delta\theta_t^i)^2, \quad (3.19)$$

em que  $\gamma$  refere-se a uma constante similar ao fator *momentum* aplicado em outros algoritmos como, por exemplo, o gradiente descendente conjugado. Segundo Dozat (2016), o valor comumente atribuído a  $\gamma$  é 0,9. Por sua vez, os pesos sinápticos com base no algoritmo Adadelta passa a ser realizada por meio da Equação 3.20.

$$\theta_{t+1}^i = \theta_t^i + \Delta\theta_t^i, \quad (3.20)$$

em que  $\Delta\theta_t^i$  representa a taxa de atualização dos pesos sinápticos definida pela Equação 3.21.

$$\Delta\theta_t^i = -\frac{\sqrt{E[\Delta\theta^2]_{t-1}^i + \varepsilon}}{\sqrt{E[g^2]_t^i + \varepsilon}} \cdot g_t^i \quad (3.21)$$

### 3.4.4 Algoritmo RMSprop

O algoritmo RMSprop apresenta características semelhantes ao Adadelta, tendo como proposta uma redução exponencial da taxa de aprendizagem, com base no fator de relação ao gradiente quadrado (Equação 3.22) durante a atualização dos pesos sinápticos (Equação 3.23). De acordo com Ruder (2016), o coeficiente  $\eta$  pode ser definido como 0,001.

$$E[g^2]_t^i = 0.9 \cdot E[g^2]_{t-1}^i + 0.1 \cdot (g_t^i)^2 \quad (3.22)$$

$$\theta_{t+1}^i = \theta_t^i - \frac{\eta}{\sqrt{E[g^2]_t^i + \varepsilon}} \cdot g_t^i \quad (3.23)$$

### 3.4.5 Adaptive Moment Estimation - Adam

O algoritmo Adam calcula momentos de aprendizagem de forma adaptativa para cada parâmetro  $\theta_t$  presente na rede. Além disso, o Adam mantém uma média exponencial decadente dos gradientes passados. Algumas características presentes no Adam também ocorrem em outros otimizadores como o RMSprop (MUKKAMALA, HEIN, 2017), Adagrad e o Adadelta (ZEILER, 2012). Durante seu treinamento são calculadas as variáveis  $m_t$  (Equação 3.24) e  $v_t$  (Equação 3.25). A primeira é um fator de correlação médio entre os gradientes antecedentes; a segunda, trata-se de um coeficiente denominado de variância descentralizada (KINGMA, BA, 2015; RUDER, 2016).

$$m_t^i = \beta_1 \cdot m_{t-1}^i + (1 - \beta_1) \cdot g_t^i, \quad (3.24)$$

$$v_t^i = \beta_2 \cdot v_{t-1}^i + (1 - \beta_2) \cdot (g_t^i)^2, \quad (3.25)$$

sendo  $\beta_1$  e  $\beta_2$  as taxas de decaimento, as quais são valores fixos que podem ser definidos entre 0 e 1. Em posse dos valores referentes aos termos  $m_t$  e  $v_t$ , faz-se necessário calcular o coeficientes de estimativa  $\hat{m}_t$  (Equação 3.26) e  $\hat{v}_t$  (Equação 3.27).

$$\hat{m}_t^i = \frac{m_t^i}{1 - \beta_1^t} \quad (3.26)$$

$$\hat{v}_t^i = \frac{v_t^i}{1 - \beta_2^t} \quad (3.27)$$

Por fim, a atualização dos pesos sinápticos leva em consideração os coeficientes de estimativa dos momentos de aprendizagem  $\hat{m}_t$  e  $\hat{v}_t$ , conforme definido pela Equação 3.28.

$$\theta_{t+1}^i = \theta_t^i - \frac{\eta}{\sqrt{\hat{v}_t^i + \varepsilon}} \cdot \hat{m}_t^i \quad (3.28)$$

### 3.4.5.1 Algoritmos Adamax e *Nesterov-accelerated Adaptive Moment Estimation* - Nadam

O algoritmo de aprendizagem Adam trata-se de uma metodologia de otimização proposta para cálculo dos pesos sinápticos, aplicado em redes neurais, que possui vantagens quando comparados ao tradicional *Backpropagation* e derivados, por apresentar em sua arquitetura matemática fatores de correção eficientes, a ponto de suprir problemas como mínimo local. No entanto, algumas alterações são propostas a fim aumentar a eficiência do algoritmo Adam e desta forma resultando em algoritmos derivados como o Adamax e o Nadam.

O algoritmo Adamax apresenta uma modificação no cálculo do coeficiente  $v_t$  apresentado na Equação 3.25. Os autores Kingma e Ba (2015) propõe que os parâmetros  $\beta_2$  e  $g_t^2$  sejam substituídos por  $\beta_2^p$  e  $g_t^p$ , conforme definido na Equação 3.29.

$$v_t^i = \beta_2^p \cdot v_{t-1}^i + (1 - \beta_2^p) \cdot |g_t^i|^p, \quad (3.29)$$

em que  $p$  trata-se de um coeficiente de valor positivo que irá amplificar tanto a taxa de decaimento  $\beta_2^p$  quanto o gradiente  $g_t$ . Com base na computação de  $v_t$ , efetuado por meio da Equação 3.29, calcula-se o valor máximo entre o resultado do produto entre  $v_t$ , sob a taxa de decaimento  $\beta_2$ , e o módulo do gradiente  $g_t$ , dando origem ao coeficiente  $u_t$  (Equação 3.30), responsável por maximizar o decaimento da taxa de aprendizagem em função do *momentum*  $v_t$  e o gradiente  $g_t$ . Com a adição de  $u_t$ , a atualização dos pesos sinápticos realizada pelo algoritmo Adamax passa a ser definida pela Equação 3.31.

$$u_t^i = \max(\beta_2 \cdot v_{t-1}^i, |g_t^i|) \quad (3.30)$$

$$\theta_{t+1}^i = \theta_t^i - \frac{\eta}{u_t^i} \cdot \hat{m}_t^i \quad (3.31)$$

O algoritmo Nadam é o resultado da combinação entre o *Nesterov Accelerated Gradient* (NAG) e o Adam, proposto com base na análise numérica apresentada em Dozat (2016). O algoritmo NAG otimiza a aprendizagem pelo gradiente inserindo o fator  $v_t$  (Equação 3.32), que irá auxiliar na movimentação da descida do gradiente em busca do mínimo global. A ideia principal desta otimização é estabelecer um termo de previsão, a fim de estimar a variância entre os parâmetros  $\theta_t$  da iteração atual com os da iteração futura. Com base na variação dos valores de  $v_t$ , os parâmetros  $\theta_t$  serão atualizados (Equação 3.33) em busca dos coeficientes que proporcionem a minimização de  $J(\theta_t)$ .

$$v_t^i = \gamma v_t^{i-1} + \eta \nabla_{\theta_t} J(\theta_t^i - \gamma v_t^{i-1}) \quad (3.32)$$

$$\theta_{t+1}^i = \theta_t^i - v_t^i \quad (3.33)$$

De acordo com Dozat (2016), a atualização dos pesos sinápticos por meio da combinação do algoritmo NAG e Adam é realizada por meio da Equação 3.34.

$$\theta_{t+1}^i = \theta_t^i - \frac{\eta}{\sqrt{\hat{v}_t^i} + \epsilon} \cdot (\beta_1 \cdot \hat{m}_t^i + \frac{(1 - \beta_1) \cdot g_t^i}{1 - \beta_1^t}) \quad (3.34)$$

As equações de atualização dos pesos sinápticos referentes a todos os métodos apresentados são resumidas na Tabela 3.1, as mesmas devem ser associadas ao conteúdo exposto nas subsecções 3.4.1 a 3.4.5.

**Tabela 3.1** – Equações de atualização dos pesos sinápticos referente aos algoritmos das subsecções 3.4.1 a 3.4.5.

| <b>Algoritmos</b>   | <b>Atualização sináptica</b>   |
|---|--|
| Gradiente Descendente   | $\theta_{t+1} = \theta_t + \eta \nabla_{\theta_t} J(\theta_t; X; Y)$   |
| SGD   | $\theta_{t+1} = \theta_t + \eta \nabla_{\theta_t} J(\theta_t; x^{[i:i+n]}; y^{[i:i+n]})$   |
| <b>Métodos com redução exponencial da taxa de aprendizagem</b>                                |  |
| $\theta_t$ o conjunto de pesos sinápticos $\{\theta_t^i, \theta_t^{i+1}, \dots, \theta_t^N\}$ |  |
| Adagrad   | $\theta_{t+1}^i = \theta_t^i - \kappa_t^i \cdot g_t^i$   |
| Adadelta  | $\theta_{t+1}^i = \theta_t^i + -\frac{\sqrt{E[\Delta\theta^2]_{t-1}^i + \epsilon}}{\sqrt{E[g^2]_t^i + \epsilon}} \cdot g_t^i$                                  |
| RMSprop   | $\theta_{t+1}^i = \theta_t^i - \frac{\eta}{\sqrt{E[g^2]_t^i + \epsilon}} \cdot g_t^i$  |
| Adam  | $\theta_{t+1}^i = \theta_t^i - \frac{\eta}{\sqrt{\hat{v}_t^i + \epsilon}} \cdot \hat{m}_t^i$   |
| Adamax  | $\theta_{t+1}^i = \theta_t^i - \frac{\eta}{u_t^i} \cdot \hat{m}_t^i$   |
| Nadam   | $\theta_{t+1}^i = \theta_t^i - \frac{\eta}{\sqrt{\hat{v}_t^i + \epsilon}} \cdot (\beta_1 \cdot \hat{m}_t^i + \frac{(1 - \beta_1) \cdot g_t^i}{1 - \beta_1^t})$ |

Fonte: Acervo Pessoal.

## 4 MATERIAIS E MÉTODOS

Neste capítulo são descritos os materiais utilizados para o desenvolvimento desta pesquisa, bem como a metodologia empregada nos experimentos computacionais.

### 4.1 Introdução

A caracterização acústica do sinal de voz é uma etapa fundamental para o processo de detecção de uma determinada patologia, ou para discriminação entre patologias (COSTA, 2008). Nesta pesquisa a caracterização é realizada por meio da análise do comportamento dinâmico da forma de onda do sinal de voz gravado e digitalizado. Uma vez digitalizado, o sinal de voz torna-se uma série temporal que traz consigo informações importantes sobre o sistema de produção vocal (VIEIRA, 2014). Desse modo, tem-se como objetivo detectar, com o uso de DNN's como algoritmo classificador, padrões comportamentais presentes no sinal de voz que caracterizem o perfil referente ao de uma voz saudável ou de uma voz patológica. Assim, espera-se, por meio do reconhecimento de padrões e aprendizagem de máquina, efetuar a detecção das classes investigadas.

De uma maneira geral, o método proposto consiste em inserir um sinal de voz digitalizado no classificador que irá analisar o comportamento dinâmico presente na sua série temporal de dados e detectar a qual classe o sinal de voz pertence. A confiabilidade e a eficiência do processo proposto dependerá muito da capacidade que o algoritmo classificador terá de conseguir aprender os padrões de comportamento de cada amostra do sinal de voz apresentado durante as etapas de treinamento, além de conseguir generalizar todos aspectos de identificação adquiridos durante o treino na etapa de validação.

Um sinal de voz digitalizado com tempo de duração entre 1 a 3 segundos (tempos mínimos de duração comumente empregados) apresenta-se como uma série temporal com um grande número de dados. Essa quantidade de instâncias inseridas de uma única vez no classificador pode tornar o processo de aprendizagem de máquina ainda mais difícil, uma vez que processar esse grande volume de dados irá aumentar a complexidade computacional tanto do treinamento, quanto na generalização do classificador. Por exemplo, um sinal de voz digital resultante da gravação de um indivíduo emitindo o som da vogal sustentada /a/, no período de 1 segundo pode conter, em média, de 16000 a 75000 amostras em sua série temporal (para taxas de amostragem de 16000 amostras/s e 75000 amostras/s, respectivamente), valores estes que podem aumentar ou diminuir conforme a taxa de amostragem por segundo do dispositivo utilizado para efetuar a captura do sinal.

Subdividir o sinal de voz em partições com o mesmo tamanho para realizar uma análise segmental apresenta-se como uma alternativa atrativa para o método proposto, tendo em vista que, para este fim, segmentar o sinal de voz traz consigo algumas vantagens. Dentre elas, temos:

- Diminuição considerável da complexidade computacional durante o processamento dos dados nas etapas de treinamento do classificador, uma vez que a análise não parte da utilização de todo o sinal de voz de uma única vez, mas da análise de segmentos com dimensões pré-definidas;
- A segmentação favorece na ampliação do número de exemplos que serão apresentados ao classificador, o que possibilita realizar uma análise do comportamento dinâmico do sinal de voz, em diferentes momentos decorrentes na série temporal;
- Favorece na redução da complexidade computacional durante a etapa de teste e validação do classificador, uma vez que o menor número de dados em análise pode proporcionar uma diminuição significativa da complexidade que o classificador terá ao detectar a classe que o sinal de voz pertence.

Sendo assim, nesta pesquisa, o sinal de voz digitalizado passa a ser fracionado em uma sequência de segmentos de mesmo comprimento. Cada um destes segmentos será inserido no classificador (um por vez) para que o mesmo defina a qual classe cada um dos segmentos apresentados pertencem. Ao fim da etapa de classificação dos segmentos realiza-se a computação da classe predominante a fim de definir em qual categoria o sinal de voz como um todo será classificado. O fluxograma apresentado na Figura 4.1 ilustra a sequência das etapas do processo proposto para efetuar a detecção de patologias.

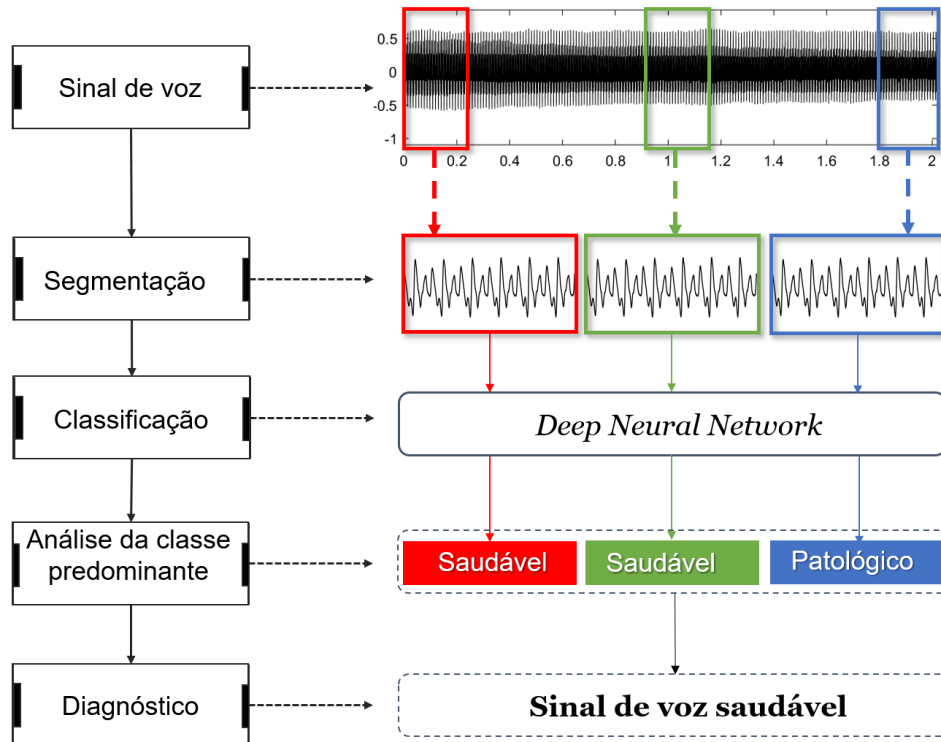
## 4.2 Base de dados

A base de dados aplicada neste trabalho trata-se da *Saarbruecken Voice Database (SVD)*, desenvolvida na Alemanha e distribuída por meio de um repositório digital aberto, na qual possui tanto sinais de vozes gravados, quanto sinais de eletroglotografia (EGG) capturados em laboratório. Na base SVD estão presentes 2043 sinais de vozes, apresentando 687 saudáveis e 1356 patológicos, distribuídos em 71 categorias diferentes de desordem vocais, sendo 57% dos dados referentes a vozes de indivíduos do sexo feminino. Da base, foram selecionados 640 sinais de voz, sendo 320 sinais de vozes saudáveis (180 femininos e 140 masculinos), 63 afetados por edema de *Reinke* (56 femininos e 7 masculinos), 21 afetados por carcinoma (1 feminino e 20 masculinos), 41 afetados por leocoplasia (17 femininos e 27 masculinos), 45 afetados por pólipos vocais (19 femininos e 26 masculinos) e 150 afetados por paralisia das pregas vocais (90 femininos e 60 masculinos).

Todos os sinais selecionados são gravações da vogal sustentada /a/. Os sinais possuem originalmente uma taxa de amostragem de 50000 amostras por segundo, mas foram re-amostradas nesta pesquisa para 25000 amostras por segundo, para diminuir a dimensão dos segmentos que serão analisados, a fim de reduzir a complexidade computacional realizado pelo

classificador. Todas as gravações possuem uma resolução de 16 bits/amostra com tempo de duração de 1 a 3 segundos gravados em formato '.wav'.

**Figura 4.1** – Fluxograma representativo do sistema de classificação proposto.



Fonte: Acervo Pessoal.

### 4.3 Metodologia

Os estudos realizados neste trabalho tem como objetivo a avaliação de sinais de vozes afetadas por patologias laríngeas, sendo consideradas as classes patológicas edema de *Reinke*, carcinoma, leucoplasia, pólipos e paralisia das pregas vocais. Tais patologias foram escolhidas por apresentarem a maior quantidade de sinais de vozes presentes na base SVD, referentes a classe investigada, no caso patologias laríngeas.

Ao decorrer do estudo são desenvolvidos 04 casos de classificação utilizando DNNs. Para cada caso, 02 classes são avaliadas por vez em um processo discriminativo. Cada classe constitui em um grupo de sinais de voz, selecionados da base SVD. Tais classes empregadas nas etapas de classificação foram:

- classe patologias, grupo composto por todos os sinais de voz afetados pelas patologias laríngeas selecionados da base SVD;
- patologias organofuncionais, no qual possui todos sinais de voz afetados pelas patologias laríngeas edema, pólipos, leucoplasia e carcinoma;

- paralisia das pregas vocais, sendo esta a única patologia laríngea de origem neurológica selecionada da base de SVD, por esse motivo constitui um grupo único;
- classe saudável, composta pelos sinais de vozes sem distúrbios provocados por patologias laríngeas.

Os processos discriminativos realizados foram: saudável *versus* patologias, saudável *versus* patologias organofuncionais, saudável *versus* paralisia das pregas vocais, patologias organofuncionais *versus* paralisia das pregas vocais. Para os processos discriminativos saudável *versus* patologias organofuncionais e saudável *versus* paralisia das pregas vocais a quantidade de sinais saudáveis foram equiparados aos patológicos, sendo estes selecionados aleatoriamente. No entanto, na discriminação patologias organofuncionais *versus* paralisia das pregas vocais, não foi possível igualar o tamanho de sinais das classes, pois a quantidade de vozes afetadas por paralisia de pacientes do sexo masculino, presentes na base SVD, é inferior a quantidade de vozes masculinas referentes a classe de patologias organofuncionais. Desse modo, para não descartar sinais da classe patologias organofuncionais, a quantidade de sinais da mesma foi mantida.

Os aspectos metodológicos aplicados no desenvolvimento de cada classificador são estruturados em 3 etapas principais: pré-processamento, treinamento e validação, sendo cada uma delas descritas a seguir.

### 4.3.1 Pré-processamento

O pré-processamento é um conjunto de atividades que envolvem preparação, organização e estruturação dos dados. Trata-se de uma etapa fundamental que precede a realização do processo de classificação (SILVA, 2018). Apresenta-se como uma etapa importante, pois será determinante para aprimorar a qualidade final dos dados que serão analisados. O pré-processamento dos dados empregados neste trabalho pode ser subdividido em três partes: filtragem, segmentação e normalização dos dados.

#### 4.3.1.1 Filtragem

Inicialmente, cada sinal de voz selecionado da base de dados é filtrado por um filtro linear de suavização do tipo média-móvel. O filtro foi aplicado para reduzir possíveis ruídos presentes nos sinais da base. Este filtro opera calculando a média de vários pontos do sinal de entrada para produzir cada ponto no sinal de saída, conforme definido na Equação 4.1 (LI et al., 2012).

$$\hat{s}[n] = \frac{1}{M} \sum_{j=0}^{M-1} s[n+j], \quad (4.1)$$



sendo  $\hat{s}[n]$  a amostra filtrada do sinal original e  $s[n + j]$ , as amostras levadas em consideração para calcular  $\hat{s}[n]$ , no qual a quantidade destas amostras é definida de acordo com o valor de  $M$ . Neste trabalho foi estabelecido, com base em ensaios experimentais, o valor de  $M$  igual a 5.

#### 4.3.1.2 Segmentação

Nesta etapa, cada sinal de voz foi fracionado em segmentos estacionários de 16 ms, valor este dentro da faixa de tempo em que um sinal de voz é considerado estacionário, resultando em partições de 400 amostras, sendo todos os segmentos extraídos sob janelamento retangular. Neste trabalho foram investigados os efeitos que as taxas de sobreposição aplicadas durante a segmentação causam nos resultados de treino e validação do classificador. Desse modo, 4 taxas de sobreposição foram avaliadas, sendo elas: 0%, 25%, 50% e 75%. Assim, durante a etapa de implementação do sistema de detecção de patologias laríngeas pelo método proposto, 04 experimentos foram realizados com os classificadores baseados em DNN's: o primeiro consistiu em realizar o particionamento do sinal de voz em segmentos de 16 ms com 0% de taxa de sobreposição e inseri-los no classificador para as etapas de treino e validação; no segundo experimento, a segmentação foi realizada com a taxa de sobreposição de 25%; no terceiro e no quarto experimentos foram investigadas a segmentação do sinal de voz, com as taxas de sobreposição de 50% e 75%. Os desempenhos obtidos, para cada conjunto de dados, foram avaliados individualmente.

#### 4.3.1.3 Escalonamento dos dados

O escalonamento de um conjunto de dados é um requisito comum para muitos estimadores de aprendizado de máquina. De acordo com Raschka (2015) e Gron (2017), algoritmos classificadores e preditores podem ter seu desempenho comprometido ao processar dados com altos níveis de dispersão ou sob a presença de *outliers*, caso este ocorrente na dinâmica aleatória presente em sinais de voz.

Para efeitos de escalonamento dos dados fez-se, neste trabalho, o uso da distribuição normal padrão. Antes de ser inserido no classificador, cada segmento passa pela distribuição normal padrão. Desse modo, um novo conjunto de valores é construído com base no *score-z* calculado pela Equação 4.2. O *score-z* é uma medida que indica o número de desvios padrão em que um valor se encontra a partir da média (LARSON, FARBER, 2004).

$$z[n] = \frac{s[n] - \mu}{\sigma}, \quad (4.2)$$

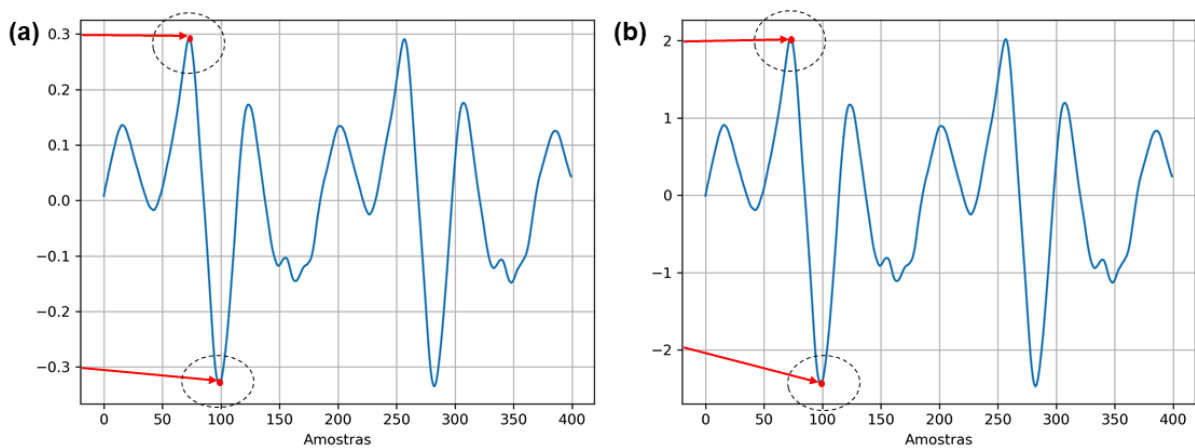
sendo  $\mu$  (Equação 4.3) a média dos valores de  $s$  do segmento analisado e  $\sigma$ , o desvio padrão (Equação 4.4).

$$\mu = \frac{1}{N} \sum_{n=1}^N s[n] \quad (4.3)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N (s[n] - \mu)^2}, \quad (4.4)$$

em que  $N$  é o total de amostras presentes no segmento. A Figura 4.2 apresenta o segmento do sinal de voz antes e após o escalonamento das amostras. Diante do exposto, é possível perceber que o comportamento dinâmico do sinal escalonado é semelhante ao do sinal original. No entanto, percebe-se uma mudança nas amplitudes do sinal, no qual o mesmo apresenta-se distribuído sob valores escalonados correspondentes ao *escore-z*, calculado por meio das equações que definem a distribuição normal padrão.

**Figura 4.2** – (a) - Segmento original; (b) - Segmento escalonado.



Fonte: Acervo Pessoal.

### 4.3.2 Arquitetura do Classificador

DNNs foram utilizadas para realizar a detecção de patologias laríngeas com base no método proposto. Nesta pesquisa, foram implementadas 03 arquiteturas de DNNs, no qual a diferença entre elas está relacionada com a quantidade de camadas ocultas, sendo analisados classificadores com 04, 05 e 06 camadas ocultas. A motivação de alterar a quantidade de camadas ocultas está relacionada com o interesse de avaliar os efeitos que a profundidade da DNN causam na eficiência do classificador.

Durante as etapas de treino e validação, o sinal de voz é fracionado em segmentos estacionários de 16 ms, possuindo no total 400 amostras. A camada de entrada da DNN recebe as 400 amostras (instâncias) referentes ao segmento em análise. Neste trabalho, cada camada oculta dos classificadores é composta por 200 neurônios, valor definido com base na Equação 4.5, empregada nos trabalhos de Costa (2008), Vieira (2013) e Souza (2015). De acordo Panchal F. & Panchal M. (2014) *apud* Abreu (2019), não há uma equação exata que dará o número correto de neurônios na camada oculta para se ter uma quantidade menor de erro. O valor obtido por meio da Equação 4.5 serve apenas como um valor de referência para a quantidade de neurônios da camada oculta, uma vez que o mesmo pode ser alterado de acordo com o projeto da rede

neural implementada. Para este trabalho, o valor obtido com base na Equação 4.5 foi mantido para todos os experimentos computacionais realizados. Os neurônios da camada de entrada, bem como os demais neurônios presentes nas camadas ocultas, foram ativados pela função ReLU (Equação 3.8). Por sua vez, a camada de saída dos classificadores é composta por 01 neurônio, sendo este ativado pela função sigmoide (Equação 3.4).

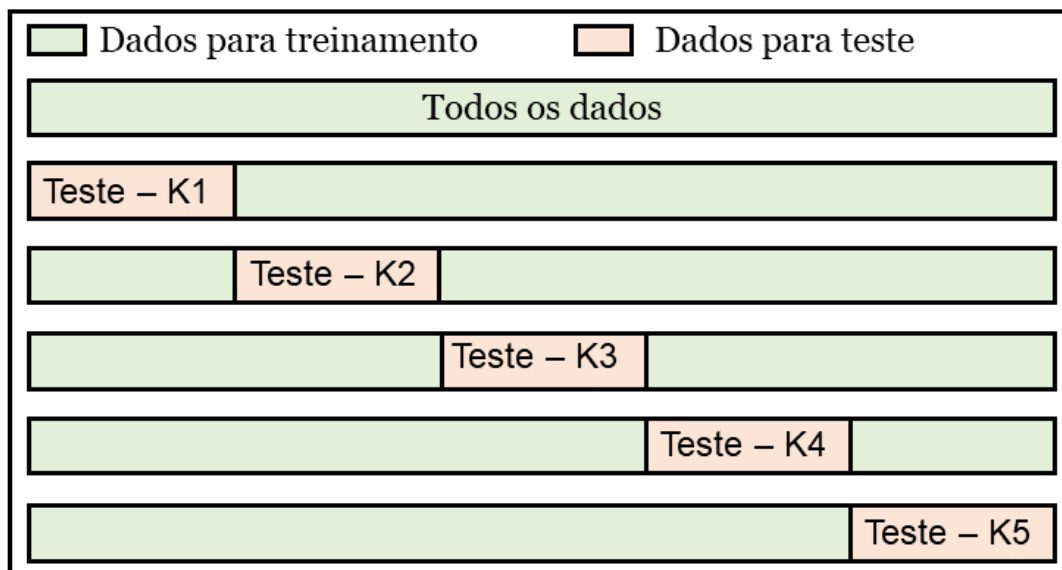
$$N_i = \frac{N_e + N_s}{2}, \quad (4.5)$$

em que  $N_i$  refere-se a quantidade neurônios da camada oculta;  $N_e$ , a quantidade de neurônios na camada de entrada; e  $N_s$ , a quantidade neurônios da camada de saída.

### 4.3.3 Treinamento e teste do classificador

Dois conjuntos de dados foram criados, sendo um para ser utilizado na etapa de treinamento e outro na de teste. Para definir o conjunto de treino e teste a técnica de validação cruzada (*Cross-Validation*) *k-fold* (DUCHESNE; REMILLARD, 2005). Desse modo, são criadas *k* partições aleatórias de dados para definir os conjuntos de treinamento e teste. Estas partições possuem aproximadamente a mesma quantidade de dados. Destes subconjuntos, apenas um é utilizado para teste e o restante para treino. Sendo assim, o classificador é treinado e testado *k* vezes com diferentes partições de dados e a cada uma destas interações são obtidas as medidas de desempenho do classificador sobre os erros e acertos da predição. Ao final dos testes são calculadas a média e o desvio padrão dos resultados das métricas referentes a cada iteração. A Figura 4.3 ilustra a seleção dos dados de treino e teste para uma validação cruzada do tipo *k-fold* para um valor de  $k = 5$ .

**Figura 4.3** – Exemplo - Partição de dados *k-fold*.



Fonte: Acervo Pessoal.

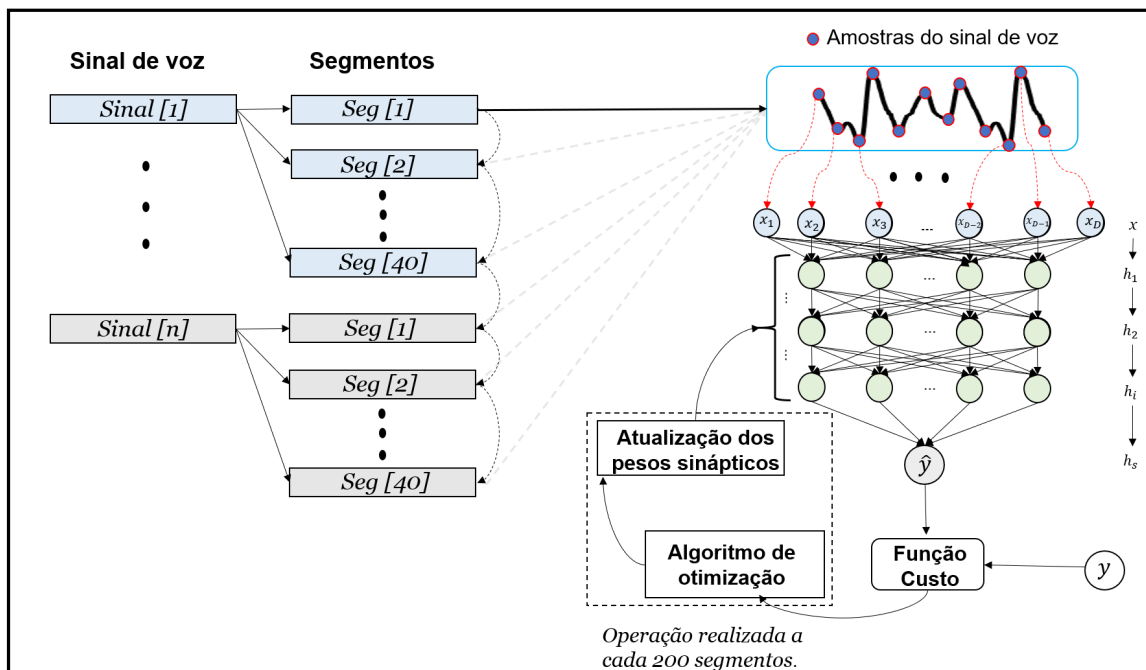
Neste trabalho é empregada a técnica de validação cruzada *k - fold* com o valor de *k* igual a 10. O valor de *k = 10* devido ao que foi observado nos trabalhos sobre aprendizagem de máquina encontrados na literatura (KOHAVI, 1995) (WITTEN; FRANK; HALL, 2011).

Inicialmente, cada sinal de voz selecionado para o conjunto de dados para treino foi particionado em 40 segmentos consecutivos (a partir do início da gravação) e foram inseridos individualmente na DNN. Por se tratar de um processo de aprendizagem supervisionado, as classes correspondentes de cada segmento são apresentadas a DNN para possibilitar o cálculo da função custo. A regra de aprendizagem, empregada neste trabalho, realiza a computação dos pesos sinápticos de cada neurônio com base na minimização da função custo entropia cruzada binária (*binary cross entropy*), sendo esta definida pela Equação 4.6. Durante o treinamento, a atualização das sinapses ocorre a cada 200 segmentos. O fluxograma apresentado na Figura 4.4 ilustra o processo de treinamento do classificador.

$$Custo = -\frac{1}{N} \sum_{n=1}^N y[n] \cdot \log(\hat{y}[n]) + (1 - y[n]) \cdot \log(1 - \hat{y}[n]), \quad (4.6)$$

em que  $y[n]$  refere-se a classe real do segmento analisado;  $\hat{y}[n]$ , a classe estimada pela DNN;  $N$ , a quantidade total de segmentos analisados.

Figura 4.4 – Fluxograma da etapa de treinamento do classificador.



Fonte: Acervo Pessoal.

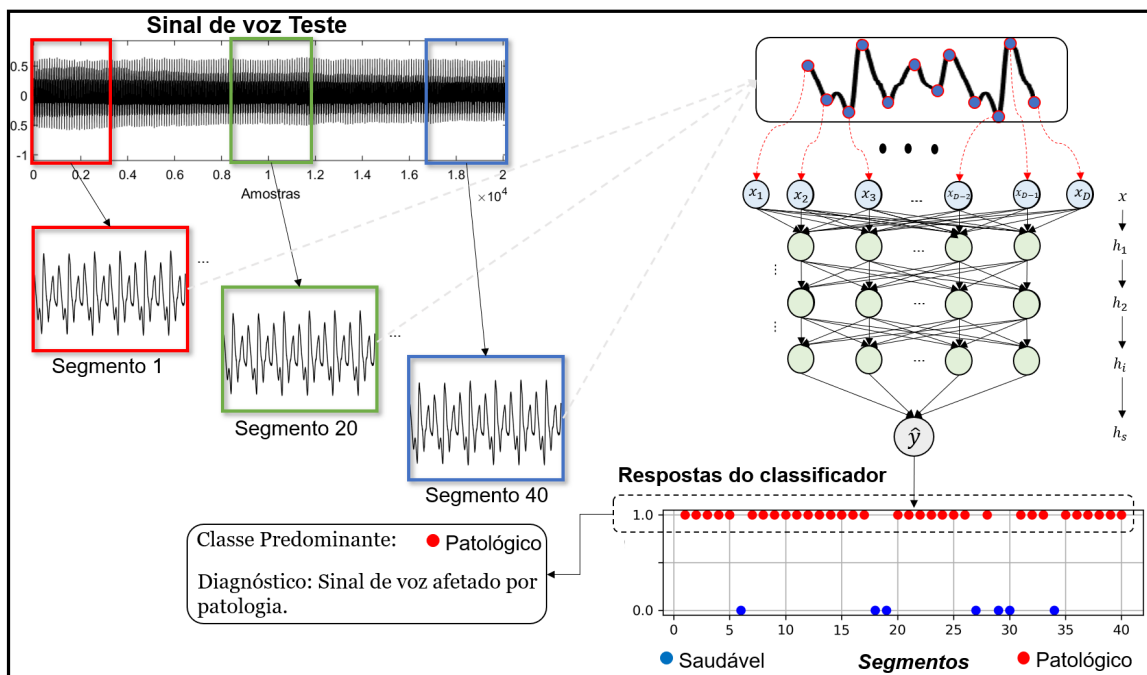
Foram investigados 07 otimizadores aplicados no treinamento dos classificadores, sendo eles: SGD, RMSprop, Adagrad, Adadelta, Adam, Adamax, e Nadam. Tais algoritmos foram selecionados com base na revisão bibliográfica realizada em Dozat (2016), no qual apresenta

os algoritmos de otimização baseados em gradiente descendente comumente empregados em treinamento de redes neurais artificiais pela comunidade científica.

A implementação de diferentes algoritmos de aprendizagem nesta pesquisa tem como objetivo investigar qual deles proporcionará os melhores classificadores para cada arquitetura das DNNs aplicadas nos processos discriminativos entre vozes. A avaliação das diferentes arquiteturas dos classificadores, treinados pelos seus respectivos algoritmos de aprendizagem, foram realizadas com base nas métricas da seção 4.3.4 calculadas após o teste com validação cruzada.

A finalização do treinamento do classificador marca o início da etapa de teste. Para tal, cada sinal de voz presente no conjunto de dados de teste foi particionado em 40 segmentos consecutivos, partindo do início da gravação. Cada um dos segmentos foi inserido individualmente no classificador treinado e o mesmo categoriza o segmento avaliado a uma das classes. Após os 40 segmentos do sinal de voz selecionado passarem pelo processo de classificação, avalia-se a classe predominante no qual a maioria dos segmentos foram classificados. O diagnóstico final será definido com base na determinação da classe predominante dos segmentos correspondentes ao sinal de voz analisado, conforme ilustrado na Figura 4.5.

**Figura 4.5** – Fluxograma da etapa de teste com validação cruzada.



Fonte: Acervo Pessoal.

#### 4.3.4 Avaliação do classificador

As métricas acurácia, sensibilidade e especificidade foram utilizadas para mensurar o desempenho do classificador empregado nesta pesquisa durante a aplicação da validação  $k$ -fold. Tais medidas são obtidas por meio da análise da matriz de confusão, sendo esta matriz uma

ferramenta importante que determina se o valor classificado corresponde ou não ao valor real, de tal forma que todos os casos em cada classe são aferidos e os mesmos são exibidos na matriz. A Tabela 4.1 apresenta a estrutura de uma matriz de confusão que fornece as possíveis situações encontradas em um teste de classificação entre duas categorias.

**Tabela 4.1** – Matriz de confusão em um teste de diagnóstico da presença ou ausência da patologia.

| Resultado | Patologia                |                          |
|-----------|--------------------------|--------------------------|
|           | Presente                 | Ausente                  |
| Positivo  | Verdadeiro Positivo (VP) | Falso Positivo (FP)      |
| Negativo  | Falso Negativo (FN)      | Verdadeiro Negativo (VN) |

Fonte: Costa (2012) (Adaptado).

A acurácia quantifica a taxa de acerto obtida durante a classificação na etapa de teste, desse modo mensurando a proporção de predições corretas (KOHAVI, PROVOST, 1998). A acurácia baseada em matriz de confusão é definida pela Equação 4.7.

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.7)$$

A sensibilidade mensura a capacidade do classificador identificar de modo correto os sinais afetados por patologias (VIEIRA, 2014; COSTA, 2012). A sensibilidade baseada em matriz de confusão é definida pela Equação 4.8.

$$Sensibilidade = \frac{VP}{VP + FN} \quad (4.8)$$

A especificidade mensura a capacidade que o classificador possui em excluir corretamente aqueles sinais que não são afetados por patologias (VIEIRA, 2014; COSTA, 2012). A especificidade baseada em matriz de confusão é definida pela Equação 4.9.

$$Especificidade = \frac{VN}{VN + FP} \quad (4.9)$$

Uma vez finalizada a etapa de validação cruzada, faz-se necessário resumir a qualidade da tabela de matriz de confusão em um único valor numérico passível de ser comparado (KOHAVI, PROVOST, 1998). Para tal, utilizou-se o  $F1_{score}$  e o coeficiente de correlação de Matthews (CCM) para criar um *ranking* dos melhores classificadores desenvolvidos para cada caso investigado.

Atualmente, a medida  $F1_{score}$  (Equação 4.11) é amplamente utilizada na maioria das áreas de aplicação do aprendizado de máquina, sendo representado pela média ponderada entre a precisão (razão entre as observações positivas previstas corretamente e o total de observações positivas previstas - Equação 4.10) e o *recall* (o mesmo que sensibilidade - Equação 4.8). O  $F1_{score}$  é definido pela Equação 4.11. Recomenda-se o seu uso para mensurar o desempenho de processos de classificação cujos dados retirados sejam balanceados. Nesta pesquisa, o  $F1_{score}$  foi empregado como índice de desempenho nos processo discriminativos saudável *versus* patologias,

saudável *versus* patologias organofuncionais e saudável *versus* paralisia das pregas vocais, uma vez que em todos estes casos as classes estão balanceadas. O  $F1_{score}$  retorna valores entre 0 e 1 (CHICCO, JURMAN, 2020).

$$precisão = \frac{VP}{VP + FP}, \quad (4.10)$$

$$F1_{score} = 2 \cdot \frac{precisão \cdot recall}{precisão + recall} \quad (4.11)$$

O CCM (CCM - Equação 4.12) é um coeficiente de correlação entre o valor observado e o valor previsto que avalia todos os casos (VP, FP, FN e VN) obtidos em cada matriz de confusão da validação cruzada. O CCM retorna um valor entre -1 e +1, no qual, quanto mais próximo de +1 melhor o desempenho do classificador. Entretanto, valores próximos a 0 representam um classificador incapaz de predição diferente de aleatória. Resultados próximos a -1 indicam que o classificador pode ser melhorado através da inversão da resposta desejada do teste. Recomenda-se seu uso para avaliar desempenhos de processos de classificação que apresentam uma quantidade de dados desbalanceada, apresentando-se nestes casos, segundo Chicco & Jurman (2020), como um índice considerado mais confiável do que o  $F1_{score}$ . Desse modo, empregou-se o índice CCM para mensurar o desempenho do processo discriminativo patologias organofuncionais *versus* paralisia das pregas vocais (MATTHEWS, 1975; SOUZA, 2016; CHICCO, JURMAN, 2020).

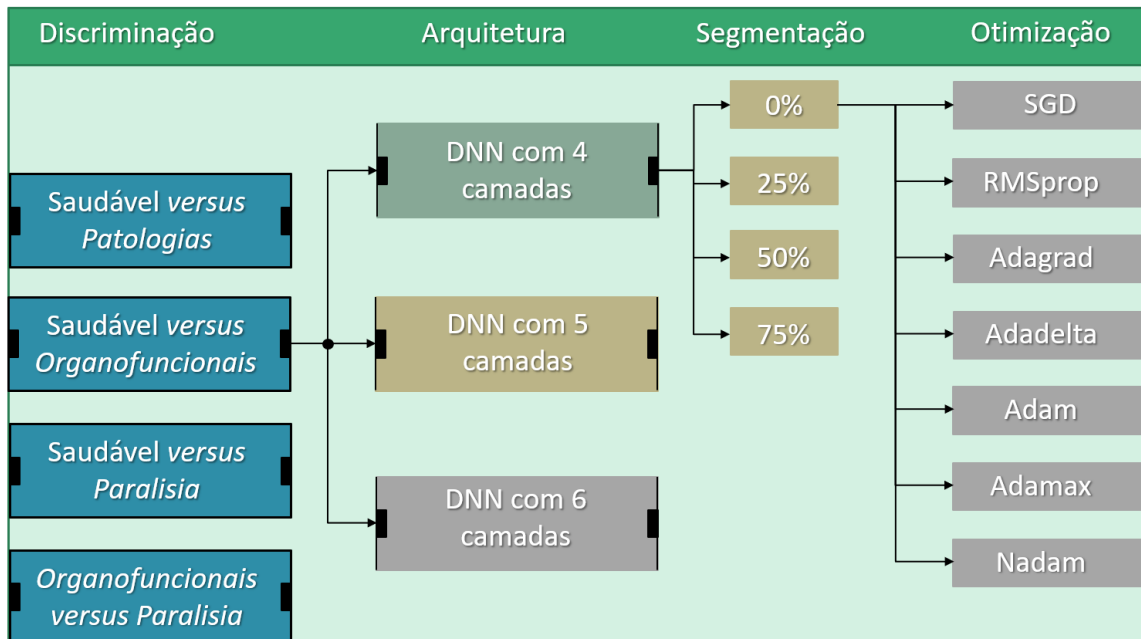
$$CCM = \frac{VP \cdot VN - FP \cdot FN}{\sqrt{(VP + FP) \cdot (VP + FN) \cdot (VN + FP) \cdot (VN + FN)}} \quad (4.12)$$

O fluxograma presente na Figura 4.6 resume a sequência de experimentos realizados, utilizando o método proposto nas etapas discriminativas.

Conforme ilustrado na Figura 4.6, para cada processo discriminativo foram avaliados 03 arquiteturas de classificadores. Cada classificador foi submetido individualmente ao método proposto, com base na análise dos segmentos sob 04 diferentes taxas de sobreposição. Por fim, para cada etapa de classificação, avaliou-se o desempenho de classificadores treinados por 07 algoritmos de otimização distintos.

Para todos os classificadores foram mensurados os valores para o  $F1_{score}$ , com exceção da discriminação patologias organofuncionais *versus* paralisia das pregas vocais, no qual empregou-se o CCM. Por arquitetura, foram selecionados os melhores classificadores, treinados pelos seus respectivos algoritmos de otimização sob as diferentes taxas de sobreposição. Como critério de seleção desta etapa, empregou-se a análise do maior valor de  $F1_{score}$  e CCM. Os casos em que os valores de  $F1_{score}$  ou CCM foram iguais, selecionou-se aqueles que apresentaram menores valores de desvio padrão entre as métricas acurácia, sensibilidade e especificidade.

**Figura 4.6** – Resumo dos ensaios experimentais realizados nesta pesquisa.



Fonte: Acervo Pessoal.

#### 4.3.5 Teste de significância

A fim de avaliar se os efeitos da variação da taxa de sobreposição e das arquiteturas de DNNs, treinadas pelos otimizadores, apresentam contribuições estatisticamente significativas, aplicou-se o teste de significância para analisar os desempenhos dos classificadores com base. Para tal, utilizou-se o teste de significância da Análise de Variância (ANOVA).

O ANOVA é um teste estatístico amplamente difundido entre os analistas e visa fundamentalmente verificar se existe uma diferença significativa entre as médias oriundas de 03 ou mais grupos diferentes a serem analisados (SILVA FILHO, 2012). Conforme apresentado em Levin (1978), o ANOVA produz um índice chamado de estatística  $F_{ANOVA}$  (ou razão  $F_{ANOVA}$ ) resultante da relação estatística das médias dos valores entre e dentro dos grupos.

Os grupos apresentam variações estatisticamente significativas quando o valor obtido para o  $F_{ANOVA}$  for positivo e maior do que o  $F_{critico}$ , estimado para um determinado intervalo de confiança. O valor referente ao  $F_{critico}$  pode ser obtido com base na análise da Tabela VI presente em Levin (1978), apresentado no Anexo A.

Nesta pesquisa, a ferramenta de Análise de Dados do *Microsoft Excel* foi empregada para calcular os valores estimados do  $F_{ANOVA}$ , obtidos para um intervalo de confiança de 95%.

#### 4.3.6 Implementação do classificador

Os dados referentes de cada segmento dos sinais de voz foram salvos em um arquivo de extensão CSV (*Comma Separated Values*) e importados no ambiente de programação do *software Spyder* utilizando a biblioteca *Pandas*. Por meio de rotinas elaboradas em *Python* em



conjunto a biblioteca *TensorFlow* foram projetadas DNNs para a etapa de classificação dos sinais de vozes selecionados da base SVD.

O *TensorFlow* é uma biblioteca de computação numérica *open source* desenvolvida pelo Google em 2015, tornando-se rapidamente a mais famosa e recomendada para *machine learning e deep learning* (GRON, 2017). A biblioteca disponibiliza várias ferramentas para o desenvolvimento e implementação de algoritmos aplicados em inteligência computacional e artificial.

O *Keras* trata-se de uma biblioteca para linguagem *Python* útil para projetos que envolvem o uso do *TensorFlow*. Algumas rotinas necessárias para trabalhos com o *TensorFlow* são longas e repetitivas, proporcionando códigos extensos e de maior complexidade. Além disso, o processamento de dados com *TensorFlow*, bem como a implementação de alguns classificadores mais robustos podem por vezes ser de difícil manipulação. O *Keras* oferece uma série de rotinas pré-definidas e parametrizáveis que facilitam o uso do *TensorFlow*. Todas as ferramentas disponibilizadas pelo *Keras* funcionam apenas como uma interface simplificada para o uso *TensorFlow*, uma vez que toda a computação numérica é realizada pela biblioteca desenvolvida pelo Google.

Uma vez realizadas as etapas de pré-processamento e partição, os dados são preparados para serem inseridos no classificador neural. Quando importados no ambiente de programação, os dados encontram-se no formato de uma matriz. Rotinas preparadas em *Python* acessam cada uma destas linhas e inserem na camada de entrada do classificador. Por sua vez, o classificador constitui de uma DNN implementada utilizando *Keras e TensorFlow*. A combinação destas rotinas finalizam a etapa de implementação dos classificadores baseados em DNNs utilizados nesta pesquisa. Os parâmetros iniciais utilizados no treinamento das DNNs para cada algoritmo de aprendizagem são apresentados na Tabela 4.2, nos quais foram definidos com base em ensaios experimentais. Os resultados dos classificadores são apresentados na seção seguinte.

**Tabela 4.2** – Parâmetros iniciais dos otimizadores aplicados no cálculo dos pesos sinápticos das DNNs.

| Otimizador/Parâmetros | $\eta$ | <i>mini-batch</i> | $\varepsilon$ | $\alpha$ | $\gamma$ | $\beta_1$ | $\beta_2$ |
|-----------------------|--------|-------------------|---------------|----------|----------|-----------|-----------|
| SGD                   | 0,01   | 200               | -             | -        | -        | -         | -         |
| RMSprop               | 0,001  | 200               | 1,00e-06      | 0,9      | -        | -         | -         |
| Adadelta              | 0,001  | 200               | 1,00e-07      | 0,95     | -        | -         | -         |
| Adagrad               | 0,001  | 200               | 1,00e-07      | -        | 0,1      | -         | -         |
| Adam                  | 0,001  | 200               | 1,00e-07      | -        | -        | 0,9       | 0,99      |
| Adamax                | 0,001  | 200               | 1,00e-07      | -        | -        | 0,9       | 0,99      |
| Nadam                 | 0,001  | 200               | 1,00e-07      | -        | -        | 0,9       | 0,99      |

$\eta$  - Taxa de aprendizagem;  $\varepsilon$  - Constante para evitar divisão por zero;  $\alpha$  - Gradiente acumulado;  $\gamma$  - Fator *momentum*;  $\beta_1, \beta_2$  - Fator de amortecimento 1 e 2.

Fonte: Acervo Pessoal.

## 5 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados os resultados obtidos do processo de discriminação entre sinais de vozes saudáveis e afetados pelas patologias laríngeas edema, pólipos, leucoplasia, carcinoma e paralisia das pregas vocais utilizando DNNs.

Os primeiros resultados apresentados são referentes ao processo de classificação entre vozes saudáveis e patológicas, na qual este tipo de classificação em um ambiente clínico é chamada de triagem. Em sequência são abordados os resultados das outras etapas de classificação, sendo elas: saudável  $\times$  patologias organofuncionais, saudável  $\times$  paralisia das pregas vocais, e patologias organofuncionais  $\times$  paralisia das pregas vocais.

A avaliação de desempenho na classificação dos sinais entre os grupos é realizada considerando-se a análise em curtos intervalos de tempo por meio de segmentos do sinal de voz. São comparados os efeitos da variação da taxa de sobreposição a fim de identificar qual método fornece as melhores taxas de classificação entre os diferentes grupos de sinais de voz selecionados.

Utilizou-se o índice  $F1_{score}$  para avaliar o desempenho geral de cada classificador e selecionar o que apresenta o melhor rendimento geral para seu respectivo caso avaliado. Por sua vez, as métricas acurácia, sensibilidade e especificidade foram aplicadas para para cada classificador durante o processo de classificação do sinal de voz à sua correta categoria.

Neste capítulo serão discutidos apenas os melhores resultados individuais de cada experimento realizado. As curvas de aprendizagem dos melhores classificadores foram adicionadas no Apêndice A. Para cada classificador, foram estabelecidas, com base em ensaios experimentais, 40 épocas de treinamento. A validação cruzada implementada utilizando o método  $k$ -fold para  $k = 10$ .

### 5.1 Resultados da discriminação entre vozes saudáveis e patológicas

Na Tabela 5.1 são apresentados os melhores resultados de cada DNN investigada na discriminação entre vozes saudáveis e patológicas, com base na detecção de segmentos da voz sob as taxas de sobreposição de 0% a 75%.

De acordo com os dados presentes na Tabela 5.1, tem-se que para a discriminação entre sinais de vozes saudáveis e patológicas, com base na detecção de segmentos com taxa de sobreposição de 0% e 50%, os classificadores compostos por 04 camadas ocultas treinados, respectivamente, pelos algoritmos Nadam e Adamax, apresentam os melhores desempenhos com base no índice  $F1_{score}$  quando comparado as demais DNNs.

**Tabela 5.1** – Melhores resultados individuais da discriminação entre vozes saudáveis e patológicas utilizando DNNs de 4, 5 e 6 camadas para análise de segmentos com diferentes taxas de sobreposição.

| Camadas  | Otimizador    | TS         | F1          | Acurácia           | Sensibilidade      | Especificidade     |
|----------|---------------|------------|-------------|--------------------|--------------------|--------------------|
| <b>4</b> | <b>Nadam</b>  | <b>0%</b>  | <b>0,85</b> | <b>85,55±3,39%</b> | <b>86,87±5,36%</b> | <b>84,39±5,76%</b> |
| 4        | Adamax        | 25%        | 0,83        | 83,34±2,27%        | 84,69±5,49%        | 82,03±6,32%        |
| 4        | Adamax        | 50%        | 0,84        | 84,02±2,85%        | 86,25±5,27%        | 81,89±6,3%         |
| 4        | Nadam         | 75%        | 0,84        | 84,27±4,57%        | 85,00±5,56%        | 83,43±6,76%        |
| <b>5</b> | <b>Nadam</b>  | <b>0%</b>  | <b>0,86</b> | <b>85,69±5,06%</b> | <b>85,94±8,18%</b> | <b>85,55±5,38%</b> |
| 5        | Adadelta      | 25%        | 0,85        | 85,22±3,47%        | 87,81±5,84%        | 82,57±7,88%        |
| 5        | Adam          | 50%        | 0,83        | 84,77±2,14%        | 88,44±4,85%        | 81,22±6,24%        |
| 5        | Adadelta      | 75%        | 0,85        | 83,97±5,48%        | 85,31±6,71%        | 82,52±9,19%        |
| 6        | Adamax        | 0%         | 0,84        | 84,75±3,91%        | 87,81±7,96%        | 81,76±5,41%        |
| <b>6</b> | <b>Adamax</b> | <b>25%</b> | <b>0,85</b> | <b>85,49±2,82%</b> | <b>86,87±6,67%</b> | <b>84,12±6,75%</b> |
| 6        | Adagrad       | 50%        | 0,84        | 84,48±2,32%        | 86,56±3,71%        | 82,51±6,37%        |
| 6        | Adamax        | 75%        | 0,83        | 84,56±4,98%        | 85,94±4,47%        | 83,15±9,11%        |

F1 -  $F1_{score}$ ; TS - Taxa de sobreposição.

Fonte: Acervo Pessoal.

Com relação ao processo discriminativo saudável *versus* patologias com base na detecção de segmentos sob 25% de taxa de sobreposição, observa-se valores de  $F1_{score}$  igual 0,86 para o classificador de 06 camadas, sendo este o maior valor obtido para esta análise. Por sua vez, para a discriminação saudável *versus* patológico com base na detecção de segmentos sob 75% de taxa de sobreposição, o maior valor de  $F1_{score}$  foi alcançado pelo classificador de 04 camadas.

Conforme apresentado na Tabela 5.1, os resultados obtidos para o índice  $F1_{score}$ , bem como para as demais métricas, referentes a discriminação entre vozes saudáveis e patológicas, utilizando o método proposto, são muito próximos. Um teste de significância, com base na análise ANOVA, foi efetuado, a fim de definir se a variação da taxa de sobreposição proporcionou contribuições estatisticamente consideráveis com base nos valores de acurácia, sensibilidade e especificidade, mensurados para cada arquitetura, durante o teste com validação cruzada. Os resultados obtidos para o teste de significância são apresentados na Tabela 5.2.

**Tabela 5.2** – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes taxas de sobreposição. Caso saudável versus patologias.

| DNN       | $F_{ANOVA}$ | $F_{critico}$ |
|-----------|-------------|---------------|
| 4 Camadas | 0,71        | 2,68          |
| 5 Camadas | 0,39        | 2,68          |
| 6 Camadas | 0,17        | 2,68          |

Fonte: Acervo Pessoal.

Diante dos resultados expostos na Tabela 5.2, observa-se que a variação de taxas de sobreposição, aplicadas durante a etapa de segmentação do sinal de voz, não apresenta contribuições estatisticamente significativas com relação a metodologia discriminativa proposta, pois o valor obtido para o  $F_{ANOVA}$ , presente na Tabela 5.2, é inferior ao  $F_{critico}$ . Desse modo, os

melhores classificadores para cada arquitetura foram selecionados com base no maior valor de  $F1_{score}$ , conforme apresentado na Tabela 5.3.

**Tabela 5.3** – Resultados referentes aos melhores classificadores para o estudo de caso saudável *versus* patológico para cada arquitetura investigada.

| Camadas | Otimizador | TS  | F1   | Acurácia    | Sensibilidade | Especificidade |
|---------|------------|-----|------|-------------|---------------|----------------|
| 4       | Nadam      | 0%  | 0,85 | 85,55±3,39% | 86,87±5,36%   | 84,39±5,76%    |
| 5       | Nadam      | 0%  | 0,86 | 85,69±5,06% | 85,94±8,18%   | 85,55±5,38%    |
| 6       | Adamax     | 25% | 0,85 | 85,49±2,82% | 86,87±6,67%   | 84,12±6,75%    |

F1 -  $F1_{score}$ ; TS - Taxa de sobreposição.

Fonte: Acervo Pessoal.

O teste de significância ANOVA foi aplicado a fim de avaliar se a diferença entre as arquiteturas dos classificadores da Tabela 5.3 apresentaram variações estatisticamente significativas. O resultado para o teste de significância é apresentado na Tabela 5.5.

**Tabela 5.4** – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes arquiteturas da Tabela 5.3.

| Arquiteturas | $F_{ANOVA}$ | $F_{critico}$ |
|--------------|-------------|---------------|
| 04 - 05 - 06 | 0,01        | 3,10          |

Fonte: Acervo Pessoal.

Diante do exposto na Tabela 5.4, o desempenho das 03 arquiteturas de classificadores validados, presentes na Tabela 5.3, são estatisticamente equiparáveis, uma vez que não apresentam variações significativas entre seus resultados mensurados pelas métricas.

### 5.1.1 Comparação com trabalhos do estado da arte

A Tabela 5.5 apresenta uma análise comparativa entre os resultados obtidos pelos classificadores propostos, neste trabalho, com resultados das principais pesquisas presentes no estado da arte que utilizaram dados da base SVD e *deep learning*.

**Tabela 5.5** – Análise comparativa com trabalhos do estado da arte.

| Referência    | Classificador | Acurácia    | Sensibilidade | Especificidade |
|---------------|---------------|-------------|---------------|----------------|
| Esta Pesquisa | 4 camadas     | 85,55±3,39% | 86,87±5,36%   | 84,39±5,76%    |
| Esta Pesquisa | 5 camadas     | 85,69±5,06% | 85,94±8,18%   | 85,55±5,38%    |
| Esta Pesquisa | 6 camadas     | 85,49±2,82% | 86,87±6,67%   | 84,12±6,75%    |
| Wu (2018)     | CNN           | 68%         | 77%           | 60%            |
| Harár (2017)  | CNN           | 71,36%      | 65,04%        | 77,67%         |

Fonte: Acervo Pessoal.

Observa-se na Tabela 5.5, os classificadores desenvolvidos neste trabalho apresentam resultados de acurácia, sensibilidade e especificidade satisfatórios, quando comparados às principais pesquisas encontradas no estado da arte. Não foram encontrados trabalhos que utilizaram o mesmo método de detecção de patologias proposto nesta pesquisa.

Conforme abordado na seção 4.3.4, o índice  $F1_{score}$  avalia os casos presentes na matriz de confusão composta pelo total de casos positivos e negativos predeterminados pelo classificador, após o teste por meio da validação cruzada. As matrizes de confusão, obtidas para o processo discriminativo saudável *versus* patologias, para os classificadores presentes na Tabela 5.3 são apresentadas na Tabela 5.6.

Com base na matrizes de confusão da Tabela 5.6, observa-se que o classificador de 05 camadas detectou mais sinais de vozes afetados por patologias laríngeas, quando comparado as demais DNNs da Tabela 5.3. No entanto, detectou menos sinais de vozes saudáveis.

Nos trabalhos de Wu (2018) e Harár (2017) nenhuma técnica de validação cruzada foi aplicada, sendo utilizado um conjunto fixo de amostras para validação do classificador proposto, no qual estes conjuntos de validação são compostos, respectivamente, por 145 (83 saudáveis e 62 patológicos) e 206 (103 saudáveis e 103 patológicos) sinais de vozes. A Tabela 5.7 apresenta a matriz de confusão dos classificadores propostos por Wu (2018) e Harár (2017), na qual pode ser comparada com as matrizes presentes na Tabela 5.6.

**Tabela 5.6** – Matrizes de confusão referente aos classificadores da Tabela 5.3.

| <b>Matriz Final</b> | DNN - 4 camadas  |         | DNN - 5 camadas  |         | DNN - 6 camadas  |         |
|---------------------|------------------|---------|------------------|---------|------------------|---------|
| Resultado           | <b>Patologia</b> |         | <b>Patologia</b> |         | <b>Patologia</b> |         |
|                     | Presente         | Ausente | Presente         | Ausente | Presente         | Ausente |
| <b>Positivo</b>     | 270              | 50      | 276              | 44      | 270              | 50      |
| <b>Negativo</b>     | 42               | 278     | 45               | 275     | 42               | 278     |

Fonte: Acervo Pessoal.

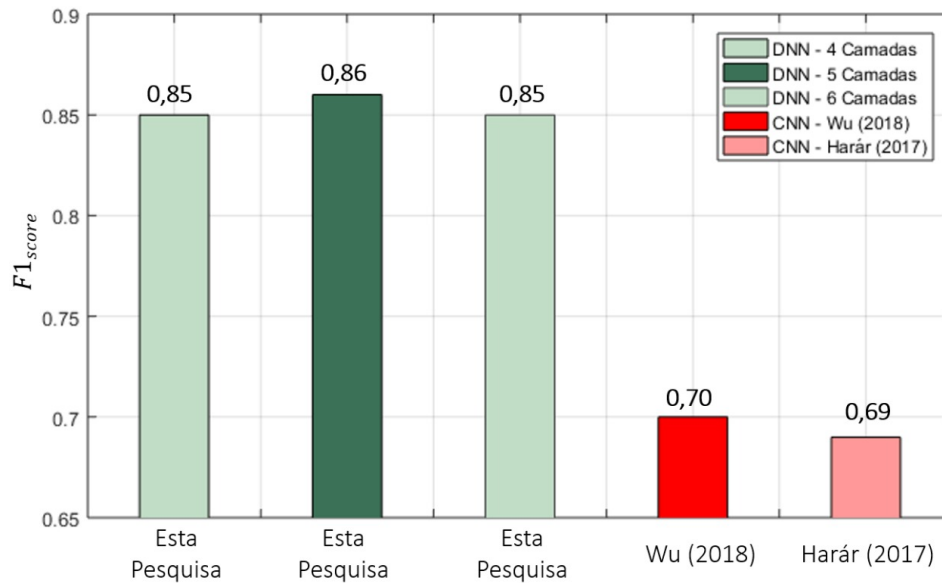
**Tabela 5.7** – Matrizes de confusão: Wu (2018) e Harár (2017).

| <b>Matriz Final</b> | Wu (2018)        |         | Harár (2017)     |         |
|---------------------|------------------|---------|------------------|---------|
| Resultado           | <b>Patologia</b> |         | <b>Patologia</b> |         |
|                     | Presente         | Ausente | Presente         | Ausente |
| <b>Positivo</b>     | 53               | 30      | 67               | 36      |
| <b>Negativo</b>     | 16               | 46      | 23               | 80      |

Fonte: Acervo Pessoal.

Com base da matriz de confusão, obtida após a validação dos seus respectivos classificadores, disponibilizada pelos autores, os valores de índice  $F1_{score}$  podem ser calculados para ambos os trabalhos e comparados com os desempenhos dos classificadores da Tabela 5.3 pelo método proposto nesta pesquisa, conforme apresentado no gráfico da Figura 5.1.

Diante dos resultados expostos na Tabela 5.5 e na Figura 5.1, tem-se que, quando comparado aos métodos usando *deep learning* apresentados por Wu (2018) e Harár (2017) para discriminação saudável *versus* patológico com sinais de vozes da base SVD, os melhores classificadores desta pesquisa (Tabela 5.3) apresentam desempenho superior, tanto com relação as métricas acurácia, sensibilidade e especificidade, quanto na análise geral da matriz de confusão, após validação cruzada utilizando o índice  $F1_{score}$ .

**Figura 5.1** – Análise comparativa com trabalhos do estado da arte:  $F1_{score}$ .

Fonte: Acervo Pessoal.

## 5.2 Resultados da discriminação entre vozes saudáveis e afetadas por patologias organofuncionais

A Tabela 5.8 apresenta os melhores resultados para cada classificador investigado na discriminação entre vozes saudáveis e afetadas por patologias laríngeas organofuncionais, com base na detecção de segmentos da voz sob as taxas de sobreposição de 0% a 75%.

**Tabela 5.8** – Melhores resultados individuais da discriminação entre vozes saudáveis e afetadas por patologias laríngeas organofuncionais utilizando DNNs de 4, 5 e 6 camadas para análise de segmentos com diferentes taxas de sobreposição.

| Camadas  | Otimizador      | TS         | F1          | Acurácia           | Sensibilidade      | Especificidade      |
|----------|-----------------|------------|-------------|--------------------|--------------------|---------------------|
| <b>4</b> | <b>Nadam</b>    | <b>0%</b>  | <b>0,83</b> | <b>81,65±4,63%</b> | <b>79,59±5,72%</b> | <b>83,53±12,0%</b>  |
| 4        | Adamax          | 25%        | 0,82        | 811,3±5,65%        | 79,38±6,19%        | 82,11±9,13%         |
| 4        | Adam            | 50%        | 0,81        | 79,28±5,84%        | 77,26±7,95%        | 83,53±7,35%         |
| 4        | Nadam           | 75%        | 0,81        | 80,37±7,79%        | 76,72±11,2%        | 84,12±9,13%         |
| 5        | Adam            | 0%         | 0,81        | 80,43±5,58%        | 78,26±6,56%        | 82,34±10,52%        |
| <b>5</b> | <b>Adadelta</b> | <b>25%</b> | <b>0,83</b> | <b>81,34±3,44%</b> | <b>77,75±8,63%</b> | <b>84,71±8,80%</b>  |
| 5        | RMSprop         | 50%        | 0,81        | 79,99±5,80%        | 78,15±9,75%        | 81,76±5,55%         |
| 5        | Adamax          | 75%        | 0,82        | 80,97±6,56%        | 77,67±10,85%       | 84,12±6,99%         |
| <b>6</b> | <b>Adadelta</b> | <b>0%</b>  | <b>0,83</b> | <b>81,97±6,42%</b> | <b>81,31±7,67%</b> | <b>82,35±13,42%</b> |
| 6        | Adamax          | 25%        | 0,82        | 80,77±4,58%        | 76,81±6,97         | 84,71±7,98          |
| 6        | Nadam           | 50%        | 0,79        | 78,71±5,90%        | 77,87±7,15%        | 79,41±6,58%         |
| 6        | RMSprop         | 75%        | 0,82        | 80,11±6,89%        | 75,24±12,09%       | 84,71±8,40%         |

F1 -  $F1_{score}$ ; TS - Taxa de sobreposição.

Fonte: Acervo Pessoal.

Como pode ser visto na Tabela 5.8, a discriminação entre sinais de vozes saudáveis *versus*

patologias organofuncionais, com base na detecção de segmentos com taxa de sobreposição de 50%, dois classificadores apresentam  $F1_{score}$  iguais a 0,81, sendo eles o de 04 e o de 05 camadas treinados, respectivamente, pelos otimizadores Adam e RMSprop.

Para o processo discriminativo com base na detecção de segmentos com taxa de sobreposição de 0%, os classificadores com 04 e 06 camadas, treinados utilizando os algoritmos Nadam e Adadelta, obtiveram valores de  $F1_{score}$  iguais a 0,83. Tais valores são os maiores mensurados entre os classificadores presentes na Tabela 5.8.

Na discriminação entre vozes saudáveis *versus* patologias organofuncionais, com base na detecção de segmentos com taxa de sobreposição de 75%, o maior valor de  $F1_{score}$  foi obtido pelo classificador de 06 camadas treinado pelo algoritmo Adadelta. Por sua vez, no processo discriminativo com base nos segmentos sob 25% de taxa de sobreposição, dois classificadores obtiveram resultados iguais a 0,82 para o índice  $F1_{score}$ , sendo eles o classificador de 04 camadas e o de 06 camadas, ambos treinados pelo otimizador Adamax.

O teste de significância ANOVA foi aplicado a fim de avaliar se a variação da taxa de sobreposição aplicada aos segmentos apresentou significância estatística na discriminação entre vozes saudáveis *versus* vozes afetadas por patologias organofuncionais, utilizando as diferentes arquiteturas de DNNs. Para tal, avaliou-se a significância entre os resultados referentes as métricas acurácia, sensibilidade e especificidade mensurados ao decorrer da validação cruzada *k - fold* de cada experimento. Os resultados para o teste ANOVA são apresentados na Tabela 5.9.

**Tabela 5.9** – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes taxas de sobreposição. Caso saudável *versus* patologias organofuncionais.

| DNN       | $F_{ANOVA}$ | $F_{crítico}$ |
|-----------|-------------|---------------|
| 4 Camadas | 0,71        | 2,68          |
| 5 Camadas | 0,39        | 2,68          |
| 6 Camadas | 0,17        | 2,68          |

Fonte: Acervo Pessoal.

Com base nos resultados expostos na na Tabela 5.9, observa-se que a variação da taxa de sobreposição aplicada aos segmentos não apresentou contribuições significativas no processo discriminativo, uma vez que o valor calculado para  $F_{ANOVA}$  é positivo e inferior ao  $F_{crítico}$ . Diante do exposto, os melhores classificadores para cada arquitetura foram selecionados com base no maior valor de  $F1_{score}$ . Na Tabela 5.10, o resumo dos melhores resultados para cada arquitetura das DNNs, investigadas no processo discriminativo saudáveis *versus* patologias organofuncionais, são apresentados. A matriz de confusão final após a validação cruzada pode ser vista na Tabela 5.11.

Diante do exposto nas Tabelas 5.10 e 5.11, observa-se a maior eficiência do classificador de 06 camadas em detectar corretamente sinais de vozes afetados por patologias organofuncionais. Contudo, o resultado obtido para a quantidade de vozes saudáveis corretamente identificadas pelo classificador de 06 camadas mostra-se inferior aos demais presentes na Tabela 5.10.

**Tabela 5.10** – Resultados referentes aos melhores classificadores para o estudo de caso saudável *versus* patologias organofuncionais para cada arquitetura investigada.

| Camadas | Algoritmo | TS  | F1   | Acurácia    | Sensibilidade | Especificidade |
|---------|-----------|-----|------|-------------|---------------|----------------|
| 4       | Nadam     | 0%  | 0,83 | 81,65±4,63% | 83,53±12,0%   | 79,59±5,72%    |
| 5       | Adadelta  | 25% | 0,83 | 81,34±3,44% | 84,71±8,80%   | 77,75±8,63%    |
| 6       | Adadelta  | 0%  | 0,83 | 81,97±6,42% | 82,35±13,42%  | 81,31±7,67%    |

F1 -  $F1_{score}$ ; TS - Taxa de sobreposição.

Fonte: Acervo Pessoal.

**Tabela 5.11** – Matrizes de confusão referente aos classificadores da Tabela 5.10

| Matriz Final<br>Resultado | DNN - 4 camadas |         | DNN - 5 camadas |         | DNN - 6 camadas |         |
|---------------------------|-----------------|---------|-----------------|---------|-----------------|---------|
|                           | Patologia       |         | Patologia       |         | Patologia       |         |
|                           | Presente        | Ausente | Presente        | Ausente | Presente        | Ausente |
| <b>Positivo</b>           | 138             | 32      | 135             | 35      | 141             | 29      |
| <b>Negativo</b>           | 28              | 142     | 26              | 144     | 30              | 140     |

Fonte: Acervo Pessoal.

O teste de significância ANOVA foi aplicado para avaliar se existe diferenças relevantes entre os classificadores da Tabela 5.10, o resultado é exposto na Tabela 5.12.

**Tabela 5.12** – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes arquiteturas da Tabela 5.10.

| Arquiteturas | $F_{ANOVA}$ | $F_{critico}$ |
|--------------|-------------|---------------|
| 04 - 05 - 06 | 0,25        | 3,10          |

Fonte: Acervo Pessoal.

O valor obtido para o  $F_{ANOVA}$ , presente na Tabela 5.12, é inferior ao  $F_{critico}$ . Desse modo, o teste indica que, estatisticamente, não existe diferenças significantes entre os desempenhos mensurados entre as arquiteturas dos classificadores da Tabela 5.10, quando aplicados no processo discriminativo saudável *versus* patologias organofuncionais, utilizando o método proposto.

### 5.3 Resultados da discriminação entre vozes saudáveis e afetadas por paralisia da pregas vocais

Os melhores resultados para cada classificador investigado na discriminação entre vozes saudáveis e afetadas por paralisia das pregas vocais, com base na detecção de segmentos da voz sob as taxas de sobreposição de 0% a 75%, são apresentados na Tabela 5.13.

De acordo com os dados presentes na Tabela 5.13, tem-se que para a discriminação entre sinais de vozes saudáveis e afetadas por paralisia as pregas vocais, com base na detecção de segmentos com taxa de sobreposição de 0% e 25%, os classificadores compostos por 04 camadas ocultas treinados pelo algoritmo Adadelta, apresentam os melhores desempenhos com base no índice  $F1_{score}$  quando comparado as demais DNNs.



Para o processo discriminativo entre vozes com base na detecção de segmentos sob taxa de sobreposição de 50% e 75% os melhores resultados foram obtidos, respectivamente, pelo classificador de 06 camadas treinado pelo algoritmo Nadam e pelo de 05 camadas treinado pelo algoritmo Adagrad.

Aplicou-se o teste de significância ANOVA foi com o objetivo de avaliar se a variação da taxa de sobreposição realizada durante a segmentação proporcionou efeitos estatisticamente considerações no processo discriminativo saudável *versus* paralisia das pregas vocais, mediante o uso das diferentes arquiteturas de DNNs empregadas no método proposto. Os resultados do teste de significância são apresentado na Tabela 5.14.

**Tabela 5.13** – Melhores resultados individuais da discriminação entre vozes saudáveis e afetadas paralisia das pregas vocais utilizando DNNs de 4, 5 e 6 camadas para análise de segmentos com diferentes taxas de sobreposição.

| Camadas  | Otimizador      | TS         | F1          | Acurácia           | Sensibilidade      | Especificidade     |
|----------|-----------------|------------|-------------|--------------------|--------------------|--------------------|
| <b>4</b> | <b>Adadelta</b> | <b>0%</b>  | <b>0,89</b> | <b>88,67±3,04%</b> | <b>92,04±5,82%</b> | <b>85,33±6,53%</b> |
| 4        | Adadelta        | 25%        | 0,88        | 88,02±5,78%        | 89,37±7,95%        | 86,67±5,96%        |
| 4        | Adagrad         | 50%        | 0,87        | 87,02±5,64%        | 89,47±8,52%        | 84,67±6,70%        |
| 4        | RMSprop         | 75%        | 0,87        | 87,35±5,3%         | 88,08±9,33%        | 86,68±4,22%        |
| 5        | Nadam           | 0%         | 0,87        | 87,68±4,92%        | 90,04±7,40%        | 85,33±7,18%        |
| 5        | Adam            | 25%        | 0,87        | 87,02±5,83%        | 88,75±9,90%        | 85,33±7,18%        |
| 5        | Adam            | 50%        | 0,87        | 87,33±5,93%        | 88,66±9,91%        | 86,01±5,54%        |
| <b>5</b> | <b>Adagrad</b>  | <b>75%</b> | <b>0,88</b> | <b>88,02±5,38%</b> | <b>91,42±7,28%</b> | <b>84,67±6,70%</b> |
| 6        | Adadelta        | 0%         | 0,87        | 87,68±4,92%        | 90,08±8,00%        | 85,33±6,53%        |
| 6        | Nadam           | 25%        | 0,87        | 87,66±4,49%        | 89,33±8,00%        | 86,00±5,54%        |
| <b>6</b> | <b>Nadam</b>    | <b>50%</b> | <b>0,89</b> | <b>88,68±5,58%</b> | <b>92,08±7,12%</b> | <b>85,33±7,18%</b> |
| 6        | Adadelta        | 75%        | 0,87        | 87,03±4,98%        | 89,46±7,86%        | 84,67±5,21%        |

F1 -  $F1_{score}$ ; TS - Taxa de sobreposição.

Fonte: Acervo Pessoal.

**Tabela 5.14** – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes taxas de sobreposição. Caso saudável *versus* paralisia das pregas vocais.

| DNN       | $F_{ANOVA}$ | $F_{critico}$ |
|-----------|-------------|---------------|
| 4 Camadas | 0,34        | 2,68          |
| 5 Camadas | 0,10        | 2,68          |
| 6 camadas | 0,29        | 2,68          |

Fonte: Acervo Pessoal.

Com base nos resultados presentes na Tabela 5.14, tem-se que a variação da taxa de sobreposição não proporcionou, estatisticamente, contribuições significativas entre os resultados mensurados para acurácia, sensibilidade e especificidade obtidos para as diferentes arquiteturas de classificadores avaliados, tendo em vista que o valor de  $F_{ANOVA}$  calculado está entre 0 e o  $F_{critico}$ .

Desse modo, os melhores classificadores, para cada arquitetura de DNN investigadas no processo discriminativo saudável *versus* paralisia das pregas vocais, foram selecionados com base no maior valor mensurado para o  $F1_{score}$ , conforme apresentado na Tabela 5.15. A matriz de confusão final após a validação cruzada pode ser vista na Tabela 5.16.

O teste de significância ANOVA foi aplicado para avaliar se existiu, estatisticamente, contribuições no desempenho mensurados pelas métricas acurácia, sensibilidade e especificidade diante da variação da quantidade camadas presentes em cada classificador da Tabela 5.15, o resultado apresenta-se na Tabela 5.17.

**Tabela 5.15** – Resultados referentes aos melhores classificadores para o estudo de caso saudável *versus* paralisia das pregas vocais.

| Camadas | Algoritmo | TS  | F1   | Acurácia    | Sensibilidade | Especificidade |
|---------|-----------|-----|------|-------------|---------------|----------------|
| 4       | Adadelta  | 0%  | 0,89 | 88,68±3,04% | 92,04±5,82%   | 85,33±6,53%    |
| 5       | Adagrad   | 75% | 0,88 | 88,02±5,38% | 91,42±7,28%   | 84,67±6,7%     |
| 6       | Nadam     | 50% | 0,89 | 88,69±5,58% | 92,08±7,12%   | 85,34±7,18%    |

F1 -  $F1_{score}$ ; TS - Taxa de sobreposição.

Fonte: Acervo Pessoal.

**Tabela 5.16** – Matrizes de confusão referente aos classificadores da Tabela 5.15

| Matriz Final    | DNN - 4 camadas |         | DNN - 5 camadas |         | DNN - 6 camadas |         |
|-----------------|-----------------|---------|-----------------|---------|-----------------|---------|
|                 | Patologia       |         | Patologia       |         | Patologia       |         |
| Resultado       | Presente        | Ausente | Presente        | Ausente | Presente        | Ausente |
| <b>Positivo</b> | 138             | 12      | 137             | 13      | 138             | 12      |
| <b>Negativo</b> | 22              | 128     | 23              | 127     | 22              | 128     |

Fonte: Acervo Pessoal.

**Tabela 5.17** – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes arquiteturas da Tabela 5.15.

| Arquiteturas | $F_{ANOVA}$ | $F_{critico}$ |
|--------------|-------------|---------------|
| 04 - 05 - 06 | 0,06        | 3,10          |

Fonte: Acervo Pessoal.

De acordo com a Tabela 5.17, os classificadores expostos na Tabela 5.13 são, estatisticamente, equiparáveis, uma vez que o valor obtido para  $F_{ANOVA}$  encontra-se entre 0 e o  $F_{critico}$ . Embora os classificadores de 04 e 06 camadas tenham apresentado o mesmo valor para o  $F1_{score}$ , os resultados para as métricas acurácia, sensibilidade e especificidade são melhores na arquitetura de 04 camadas, levando em consideração o desvio padrão. Tal análise sugere que a mesma possuiu melhor desempenho durante a validação cruzada.

## 5.4 Resultados da discriminação entre vozes afetadas por patologias organofuncionais e paralisia das pregas vocais

A Tabela 5.18 apresenta os melhores resultados para cada classificador investigado na discriminação entre vozes afetadas por patologias laríngeas organofuncionais e por paralisia das pregas vocais, com base na detecção de segmentos da voz sob as taxas de sobreposição de 0% a 75%. Para este experimento a matriz de confusão se remete aos casos de verdadeiros e falsos diagnósticos de patologias organofuncionais ou de paralisia das pregas vocais. As métricas sensibilidade e especificidade não foram aplicadas, uma vez que estas métricas são referidas em casos de avaliação de classificadores aplicados na triagem de pacientes. No caso discriminativo entre doenças, avalia-se apenas a acurácia.

**Tabela 5.18** – Melhores resultados individuais da discriminação entre vozes afetadas por patologias organofuncionais e afetadas paralisia das pregas vocais utilizando DNNs de 4, 5 e 6 camadas.

| Camadas  | Algoritmo      | TS        | CCM         | Acurácia            |
|----------|----------------|-----------|-------------|---------------------|
| <b>4</b> | <b>RMSprop</b> | <b>0%</b> | <b>0,59</b> | <b>80,29±17,21%</b> |
| 4        | Adam           | 25%       | 0,32        | 66,57±12,04%        |
| 4        | sgd            | 50%       | 0,35        | 67,84±8,62%         |
| 4        | Adagrad        | 75%       | 0,34        | 67,34±10,29%        |
| <b>5</b> | <b>RMSprop</b> | <b>0%</b> | <b>0,62</b> | <b>81,48±14,88%</b> |
| 5        | Adam           | 25%       | 0,32        | 66,39±10,04%        |
| 5        | Adam           | 50%       | 0,34        | 67,38±8,34%         |
| 5        | Adagrad        | 75%       | 0,33        | 66,68±8,63%         |
| <b>6</b> | <b>RMSprop</b> | <b>0%</b> | <b>0,59</b> | <b>80,25±16,43%</b> |
| 6        | Adamax         | 25%       | 0,37        | 68,58±8,71%         |
| 6        | Adagrad        | 50%       | 0,34        | 67,33±8,89%         |
| 6        | Nadam          | 75%       | 0,34        | 67,43±10,45%        |

CCM - Coeficiente de Correlação de Matthews; TS - Taxa de sobreposição.

Fonte: Acervo Pessoal.

Como pode ser visto na Tabela 5.18, a discriminação patologias organofuncionais *versus* paralisia das pregas vocais, com base na detecção de segmentos com taxa de sobreposição de 50% e 75%, os classificadores compostos por 04 camadas ocultas treinados, respectivamente, pelos algoritmos SGD e Adagrad, apresentaram os melhores desempenhos com base no índice CCM quando comparado as demais DNNs para este mesmo processo discriminativo.

Na classificação com base nos segmentos de 25% o melhor resultado foi obtido pela DNN de 06 camadas treinada pelo algoritmo Adamax. O maior índice CCM foi mensurado para o classificador de 05 camadas treinado pelo algoritmo RMSprop, obtendo valor mensurado igual a 0,62, aplicado no processo discriminativo com base na detecção dos segmentos sob taxa de 0% de sobreposição.

O teste de significância ANOVA, foi efetuado, a fim de definir se a variação da taxa de sobreposição proporcionou contribuições estatisticamente consideráveis com base nos valores de

acurácia, sensibilidade especificidade, mensurados para cada arquitetura, durante o teste com validação cruzada  $k$ -fold. Os resultados obtidos para o teste de significância são apresentados na Tabela 5.19.

Os valores referentes ao  $F_{ANOVA}$  presentes na Tabela 5.19 são superiores ao  $F_{critico}$ . Os resultados indicam que a comparação das métricas mensuradas entre os casos analisados apresentam diferenças estatísticas. Observa-se, com base na Tabela 5.18, que os melhores desempenhos foram mensurados para os classificadores cuja análise foi baseada na detecção de segmentos sem taxa de sobreposição. A variação da taxa de sobreposição provoca desempenhos inferiores na discriminação entre patologias utilizando o método proposto. Desse modo, os classificadores que atingiram os maiores valores de CCM apresentam-se estatisticamente superiores aos demais.

**Tabela 5.19** – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes taxas de sobreposição. Caso patologias organofuncionais versus paralisia das pregas vocais.

| DNN       | $F_{ANOVA}$ | $F_{critico}$ |
|-----------|-------------|---------------|
| 4 Camadas | 4,19        | 2,68          |
| 5 Camadas | 8,12        | 2,68          |
| 6 Camadas | 5,14        | 2,68          |

Fonte: Acervo Pessoal.

O resumo dos melhores resultados para cada arquitetura das DNNs investigadas no processo discriminativo patologias organofuncionais *versus* paralisia das pregas vocais é apresentado na Tabela 5.20. A matriz de confusão final, após a validação cruzada, pode ser vista na Tabela 5.21.

**Tabela 5.20** – Resultados referentes aos melhores classificadores para o estudo de caso discriminativo entre patologias laríngeas organofuncionais e por paralisia das pregas vocais.

| Camadas  | Algoritmo      | TS        | CCM         | Acurácia            |
|----------|----------------|-----------|-------------|---------------------|
| 4        | RMSprop        | 0%        | 0,59        | 80,29±17,21%        |
| <b>5</b> | <b>RMSprop</b> | <b>0%</b> | <b>0,62</b> | <b>81,48±14,88%</b> |
| 6        | RMSprop        | 0%        | 0,59        | 80,25±16,43%        |

CCM - Coeficiente de Correlação de Matthews; TS - Taxa de sobreposição.

Fonte: Acervo Pessoal.

**Tabela 5.21** – Matrizes de confusão referente aos classificadores da Tabela 5.20

OGF: Organofuncional; PRL: Paralisia

| Matriz Final<br>Resultado | DNN - 4 camadas |     | DNN - 5 camadas |     | DNN - 6 camadas |     |
|---------------------------|-----------------|-----|-----------------|-----|-----------------|-----|
|                           | Diagnóstico     |     | Diagnóstico     |     | Diagnóstico     |     |
|                           | OGF             | PRL | OGF             | PRL | OGF             | PRL |
| Organofuncional           | 142             | 28  | 137             | 33  | 137             | 33  |
| Paralisia                 | 35              | 115 | 26              | 124 | 30              | 120 |

Fonte: Acervo Pessoal.

Diante do exposto nas Tabelas 5.20 e 5.21, observa-se que o classificador de 05 camadas (CCM = 0,62) apresentou o melhor desempenho no diagnóstico da patologia paralisia das pregas vocais quando comparado aos demais classificadores. Com relação a detecção das patologias organofuncionais, o melhor resultado foi obtido pelo classificador de 04 camadas (CCM = 0,59).

O teste de significância ANOVA (Tabela 5.22) foi aplicado para avaliar se o aumento do número de camadas contribuiu, estatisticamente, com o desempenho entre os melhores classificadores para cada arquitetura utilizada na discriminação entre patologias organofuncionais *versus* paralisia das pregas vocais pelo método proposto.

**Tabela 5.22** – Resultados referentes aos teste de significância ANOVA com base na análise dos valores das métricas para as diferentes arquiteturas da Tabela 5.20.

| Arquiteturas | $F_{ANOVA}$ | $F_{critico}$ |
|--------------|-------------|---------------|
| 04 - 05 - 06 | 0,05        | 3,10          |

Fonte: Acervo Pessoal.

Com base na Tabela 5.22, os classificadores expostos na Tabela 5.20 são, estatisticamente, equiparáveis, uma vez que o valor obtido para  $F_{ANOVA}$  encontra-se entre 0 e o  $F_{critico}$ .

## 5.5 Discussão

Os resultados obtidos indicam que cada estudo de caso analisado apresenta suas particularidades, no que diz respeito as arquiteturas dos classificadores implementados, a fim de discriminar diferentes tipos de sinais de vozes com base na série temporal dinâmica de segmentos estacionários com e sem taxas de sobreposição. O primeiro grande fator a ser observado trata-se da aplicação de diferentes algoritmos de aprendizagem utilizados no processo de treinamento dos classificadores.

O algoritmo Nadam proporcionou a dois classificadores os melhores desempenhos. Tais resultados foram obtidos para diferentes arquiteturas utilizadas na discriminação saudável *versus* patologias, saudável *versus* patologias organofuncionais e saudável *versus* paralisia das pregas vocais. Por sua vez, o algoritmo RMSprop destacou-se no treinamento do classificador de 5 camadas aplicado na discriminação entre patologias. De acordo com Dozat (2016), os otimizadores com variáveis adaptativas, como o RMSprop e Nadam, destacam-se na literatura como algoritmos que possuem muito potencial em aprender e generalizar dados com características dispersas ou com aspectos pseudo-aleatórios, caso esse semelhante ao comportamento dinâmico de um sinal de voz.

As DNNs de 04, 05 e 06 camadas foram submetidas a processos discriminativos entre sinais de vozes com base na detecção de segmentos com taxas de sobreposição diferentes, sendo elas 0%, 25%, 50% e 75%. Em vários casos, foram obtidos valores de índice de  $F1_{score}$  e CCM distintos, uma vez que a mudança da taxa de sobreposição aplicada aos segmentos provocaram

variações no desempenho dos classificadores. No entanto, analisando os resultados das Tabelas 5.1, 5.4, 5.5, percebe-se que a diferença entre os valores calculados para o índice  $F_{ANOVA}$  do melhor e do pior classificador de mesma quantidade de camadas é pequena para os processos discriminativos: saudável *versus* patologias, saudável *versus* patologias organofuncionais e saudável *versus* paralisia das pregas vocais.

Com base no teste de significância ANOVA, conclui-se que os melhores classificadores baseados na análise de segmentos com diferentes taxas de sobreposição não apresentam variações estatisticamente consideráveis entre seus resultados mensurados pelas métricas acurácia, sensibilidade e especificidade. O mesmo ocorre diante dos melhores classificadores selecionados por arquitetura. Tal análise sugere que, para a metodologia proposta, os melhores classificadores, resumidos nas Tabelas 5.3, 5.11, 5.16, conseguem realizar seus respectivos processos discriminativos com desempenhos semelhantes.

Nesse caso, em termos de implementação computacional, os classificadores com arquitetura de 04 camadas destaca-se entre os demais por possuir um menor custo computacional, uma vez que possui menos camadas e neurônios ocultos quando comparado aos demais investigados nesta pesquisa.

Com relação aos processo discriminativo entre patologias organofuncionais *versus* paralisia das pregas vocais, a variação da taxa de sobreposição dos segmentos não favoreceu ao processo de classificação realizado pelas DNNs. Isso resultou em classificadores com baixo potencial de acerto, quando comparado aos classificadores que discriminaram as patologias com base na detecção de segmentos sem taxa de sobreposição, conforme apresentado na Tabela 5.18. Para o caso da discriminação entre patologias, todos os classificadores que obtiveram maiores índices de CCM foram os baseados na detecção de segmentos sob 0% de taxa de sobreposição.

Os classificadores com maiores valores de  $F1_{score}$  obtiveram os melhores desempenhos na métrica sensibilidade nos processos discriminativos entre sinais de vozes saudável *versus* patologias, saudável *versus* patologias organofuncionais e saudável *versus* paralisia das pregas vocais. Desse modo, tais classificadores obtiveram o melhor rendimento em detectar corretamente um sinal de voz afetado por alguma das patologias. O método proposto aplicado na discriminação saudável *versus* patologias, apresenta desempenho satisfatório, tendo em vista os resultados superiores ao de outros trabalhos recentes presentes na literatura. O maior índice de  $F1_{score}$  foi alcançado pelo classificador de 04 camadas treinado pelo algoritmo Adadelta aplicado na discriminação saudável *versus* paralisia das pregas vocais, conforme apresentado na Tabela 5.16.

Por sua vez, a discriminação entre patologias mostrou-se um processo complexo para o método proposto, tendo em vista que, neste caso de classificação, foram obtidos resultados inferiores e com altos valores de desvio padrão, apresentando uma variação entre os resultados durante a validação cruzada. Desta forma, o método proposto não apresentou desempenho satisfatório para este processo discriminativo, uma vez que apresenta rendimento inferior quando comparado as demais classificações investigadas nesta pesquisa.

## 6 CONSIDERAÇÕES FINAIS

Diante dos resultados obtidos, pode-se ser considerar que o método proposto utilizando DNNs na discriminação entre vozes saudáveis e patológicas da base SVD, referentes a indivíduos do sexo feminino e masculino, tem possibilidade de contribuir na triagem realizada por profissionais da saúde. Além disso, o método proposto apresenta diferenças significativas com grande parte dos trabalhos apresentados na literatura, devido a possibilidade de discriminar sinais de voz sem extrair medidas objetivas, sejam elas lineares ou não lineares.

No que se diz respeito a validação de um classificador baseado na análise dinâmica do sinal de voz junto a DNNs, tem-se que a arquitetura de 05 camadas com 200 neurônios ocultos apresentou o melhor desempenho para a atividade de triagem, com base no maior valor de  $F1_{score}$ . Por sua vez, a rede de 04 camadas com 200 neurônios ocultos obteve os melhores resultados para a correta classificação de vozes saudáveis e afetadas por paralisia, ambas utilizando o método de classificação baseados na detecção de segmentos sem taxa de sobreposição. Os resultados mostram que a análise do comportamento dinâmico do próprio sinal de voz pode ser uma ferramenta útil para discriminar sinais de vozes saudáveis de patológicos.

De uma maneira geral, as principais contribuições científicas desta pesquisa foram:

- Apresentação e análise comparativa entre diferentes algoritmos de otimização utilizados no treinamento de DNNs aplicadas em diferentes etapas de classificação envolvendo vozes saudáveis e patológicas, no qual não foram encontrados trabalhos relacionados a este tipo específico de investigação aplicados diretamente a análise vocal.
- Comparação entre diferentes arquiteturas de DNNs, bem como a análise do impacto ocasionado pela profundidade de uma DNN, a fim de avaliar seus efeitos sob a correta classificação de sinais de voz.
- Investigação da possibilidade de conseguir discriminar sinais de vozes saudáveis de patológicos, analisando apenas o comportamento da dinâmica do sinal de voz, sem a necessidade de extração de características, processo pré-discriminatório ou do uso de algoritmos responsáveis por testar uma série de características a fim de encontrar um conjunto de dados que apresentem melhores resultados.
- Proporcionar técnicas que podem se tornar ferramentas de apoio ao diagnóstico e tratamento de patologias laríngeas para profissionais da área da voz.
- Contribuição para o estudo de pesquisadores da área de *deep learning* e análise acústica aplicada a sinais de voz.

## 6.1 Sugestões para trabalhos futuros

- Avaliar novas metodologias com base na análise dinâmica do sinal de voz para melhorar a sua eficiência com relação a detecção de patologias laríngeas organofuncionais;
- Aperfeiçoar a estratégia de classificação, como por exemplo, utilizar classificadores DNNs combinados com outros algoritmos de classificação e variar o tamanho dos segmentos;
- Avaliar o processo de classificação para uma análise discriminativa com base no comportamento dinâmico da resposta em frequência do sinal de voz;
- Explorar o uso de outras arquiteturas de DNN aplicadas no processo de discriminação entre vozes saudáveis e patológicas, bem como a discriminação entre patologias;
- Utilizar arquiteturas prontas apresentadas na literatura, como a ResNET;
- Explorar o uso de parametrizadores automáticos como o *grid search*.



## REFERÊNCIAS

- ABREU, S. R. **Avaliação da transformada de Hilbert-Huang na detecção de desvios vocais**. Dissertação (Mestrado) — Instituto Federal da Paraíba - IFPB, João Pessoa, PB, março. 2019. 103f.
- AGUIRRE, L. **Introdução à Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais**. Editora UFMG, 2007. ISBN 9788570415844. Disponível em: <<https://books.google.com.br/books?id=f9IwE7Ph0fYC>>.
- AGUIRRE, L.; LETELLIER, C. Modeling nonlinear dynamics and chaos: A review. **Mathematical Problems in Engineering**, 06 2009.
- ALHUSSEIN, M.; MUHAMMAD, G. Voice pathology detection using deep learning on mobile healthcare framework. **IEEE Access**, v. 6, p. 41034–41041, 2018. ISSN 2169-3536.
- ALHUSSEIN, M.; MUHAMMAD, G. Automatic voice pathology monitoring using parallel deep models for smart healthcare. **IEEE Access**, v. 7, p. 46474–46479, 2019. ISSN 2169-3536.
- ALHUSSEIN, M.; MUHAMMAD, G.; HOSSAIN, M. S. Eeg pathology detection based on deep learning. **IEEE Access**, v. 7, p. 27781–27788, 2019. ISSN 2169-3536.
- ALMEIDA, N. C. de. **Sistema Inteligente para Diagnóstico de Patologias na Laringe utilizando Máquinas de Vetor de Suporte**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte - UFRN, Natal, RN, Julho. 2010. 119f.
- ANDRADE, R. G. A. da C. et al. Dificuldade na laringoscopia e na intubação orotraqueal: estudo observacional. **Revista Brasileira de Anestesiologia**, v. 68, n. 2, p. 168–173, 2018.
- ANKISHAN, H. A new approach for the acoustic analysis of the speech pathology. In: **2017 International Conference on Engineering and Technology (ICET)**. [S.l.: s.n.], 2017. p. 1–5.
- ARAUJO, T. S. D.; FAGUNDES, F.; BRANDAO, P. Aplicação de redes neurais artificiais para o diagnóstico de patologias traumato-ortopédicas dos membros inferiores. In: . [S.l.: s.n.], 2012.
- ASSEF, A. A. et al. Sistema para geração, aquisição e processamento de sinais de ultra-som. **Sba: Controle & Automação Sociedade Brasileira de Automatica**, scielo, v. 20, p. 145 – 155, 06 2009. ISSN 0103-1759.
- BAR, Y. et al. Chest pathology detection using deep learning with non-medical training. In: **2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)**. [S.l.: s.n.], 2015. p. 294–297. ISSN 1945-7928.
- BARAVIEIRA, P. B. **Aplicação de uma rede neural artificial para a avaliação da rugosidade e soprovidade vocal**. Tese (Doutorado) — Universidade de São Paulo, São Carlos, SP, 2016. 100f.
- BEHLAU, M. **Voz: o livro do especialista -**. Revinter, 2001. ISBN 9788573095258. Disponível em: <<https://books.google.com.br/books?id=vI8tAAAACAAJ>>.
- BELHAJ, A.; BOUZID, A.; ELLOUZE, N. Normal and organic pathology classification of female voices using svm classifier. **International Journal of Applied Engineering Research**, v. 11, n. 15, p. 8702–8707, 2016.

BENJAMIN, B. **Cirurgia Endolaríngea**. 1. ed. [S.l.]: Editora Revinter, 2000.

BRAGA, P. de L. **Reconhecimento de voz dependente de locutor utilizando Redes Neurais Artificiais**. 2006. Monografia (Bacharel em Engenharia da Computação), Universidade de Pernambuco, Recife-PE, Brazil.

BRENT, R. P. Fast training algorithms for multilayer neural nets. **IEEE Trans. on Neural Networks**, v. 2, n. 3, p. 346–354, 1991.

CAVICHIOLO, J. B. et al. Relação entre avaliação vocal, acústica e qualidade de vida em voz de mulheres com diferentes graus de edema de reinke. **Distúrb Comun, São Paulo**, v. 1, n. 31, p. 147–159, 03 2019.

CECATTO, S. B. et al. Paralisia de pregas vocais: aspectos clinicos e cirúrgicos. **Revista Brasileira de Otorrinolaringologia**, v. 68, p. 534 – 538, 08 2002. ISSN 0034-7299. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-72992002000400013&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-72992002000400013&nrm=iso)>.

CHEN, M. et al. A morphological classification for vocal fold leukoplakia. **Brazilian Journal of Otorhinolaryngology**, sciELO, v. 85, p. 588 – 596, 2019. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1808-86942019000500588&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1808-86942019000500588&nrm=iso)>.

CHICCO, D.; JURMAN, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. **BMC genomics**, v. 21, n. 1, p. 1–13, 2020.

CHUANG, Z. et al. Dnn-based approach to detect and classify pathological voice. In: **2018 IEEE International Conference on Big Data (Big Data)**. [S.l.: s.n.], 2018. p. 5238–5241.

CIELO, C. A. et al. Lesões organofuncionais do tipo nódulos, pólipos e edema de *Reinke*. **Revista CEFAC**, v. 13, n. 4, p. 735–748, 2011.

CIELO, C. A. et al. Qualidade de vida em voz, avaliação perceptivoauditiva e análise acústica da voz de professoras com queixas vocais. **Audiology - Communication Research**, v. 20, n. 2, p. 130–140, 2015.

COMPARISON of the predicted and observed secondary structure of T4 phage lysozyme. **Biochimica et Biophysica Acta (BBA) - Protein Structure**, v. 405, n. 2, p. 442 – 451, 1975. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0005279575901099>>.

COSTA, S. L. do N. C. **Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas**. Tese (Doutorado) — Universidade Federal de Campina Grande, Campina Grande, PB, 2008. 161f.

COSTA, W. C. de A. **Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas**. Tese (Doutorado) — Universidade Federal de Campina Grande, Campina Grande, PB, 2012. 176f.

COSTA, W. C. de A. et al. Análise dinâmica do sistema de produção vocal na presença de patologias da laringe. **Proceeding Series of the Brazilian Society of Applied and Computational Mathematics**, v. 1, n. 1, p. 1–6, 2013.

- COUTO, T. B. **Mudando a visão: videolaringoscopia como opção para intubações no departamento de emergência pediátrica.** Tese (Doutorado) — Universidade de São Paulo, São Paulo, SP, 2018. 81f.
- DE, P. P. H. U. **Avaliação de Desordens Vocais por Meio de Características do Espaço de Fase Reconstruído.** Dissertação (Mestrado em Engenharia Elétrica), João Pessoa-PB, 2017.
- DIAS, L. C.; CORREIA, S. E. N.; COSTA, S. L. do N. C. Análise de estimadores recursivos aplicados no cálculo de coeficientes lpc de sinais de voz com patologias laríngeas. **Anais da XIV Conferência Brasileira de Dinâmica, Controle e Aplicações**, 2019.
- DIAS, S. O. **Estimation of the glottal pulse from speech or singing voice.** Dissertação (Mestrado) — Universidade do Porto, Porto, PT, jul. 2012.
- DOZAT, T. Incorporating nesterov momentum into adam. In: . [S.l.: s.n.], 2016.
- DUARTE, B. B. et al. Paralisia bilateral em abdução de pregas vocais como manifestação de câncer de esôfago: Relato de caso e revisão de literatura. **Arq. Int. Otorrinolaringol**, v. 10, n. 4, p. 327–330, Dezembro 2006.
- DUCHESNE, P.; REMILLARD, B. **Statistical modeling and analysis for complex data problems.** [S.l.]: Springer Science & Business Media, 2005.
- ECKLEY, C. A. et al. Incidência de alterações estruturais das pregas vocais associadas ao pólipos de prega vocal. **Revista Brasileira de Otorrinolaringologia**, scielo, v. 74, p. 508 – 511, 08 2008. ISSN 0034-7299. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-72992008000400005&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-72992008000400005&nrm=iso)>.
- FAHLMAN, S. E. An empirical study of learning speed in back-propagation networks. In: . [S.l.: s.n.], 1988.
- FANG, S.-H. et al. Detection of pathological voice using cepstrum vectors: A deep learning approach. **Journal of Voice**, 2018. ISSN 0892-1997. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S089219971730509X>>.
- FANT, G. **Acoustic: Theory of speech production.** [S.l.]: Walter de Gruyter, Paris, 1970.
- FECHINE, J. M. **Reconhecimento Automático de Identidade Vocal Utilizando Modelagem híbrida: Paramétrica e Estatística.** Tese (Doutorado) — Universidade de Campina Grande, Campina Grande, PB, 2000. 237f.
- FELIPPE, A. C. N. de; GRILLO, M. H. M. M.; GRECBI, T. H. Normatização de medidas acústicas para vozes normais. **REVISTA BRASILEIRA DE OTORRINOLARINGOLOGIA**, v. 5, n. 72, p. 659–664, 2006.
- FILHO, A. S. da S. Estudo sobre a aplicação da análise de variância. **Revista Pensar Gestão e Administração**, v. 2, 2012.
- FONSECA, E.; PEREIRA, J. Normal versus pathological voice signals. **Engineering in Medicine and Biology Magazine, IEEE**, v. 28, p. 44 – 48, 11 2009.
- FURTADO, H.; VELHO, H. C.; MACAU, E. Assimilação de dados com redes neurais artificiais em equações diferenciais. In: . [S.l.: s.n.], 2011.

- GARCIA, M. d. M. et al. Avaliação por imagem da paralisia de pregas vocais. **Radiologia Brasileira**, scielo, v. 42, p. 321 – 326, 10 2009.
- GODINO-LLORENTE, J.; GOMEZ-VILDA, P.; VELASCO, M. B. Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. **IEEE Transactions on Biomedical Engineering**, v. 10, n. 53, p. 1943–1953, 2006.
- GRON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 1st. ed. [S.l.]: O’Reilly Media, Inc., 2017. ISBN 1491962291, 9781491962299.
- GRZYWALSKI, T. et al. Parameterization of sequence of mfccs for dnn-based voice disorder detection. **IEEE Transactions on Biomedical Engineering**, v. 9, n. 11, 2018.
- HAMMAMI, I.; SALHI, L.; LABIDI, S. Pathological voices detection using support vector machine. In: **2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)**. [S.l.: s.n.], 2016. p. 662–666. ISSN null.
- HARAR, P. et al. Voice pathology detection using deep learning: a preliminary study. In: **2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)**. [S.l.: s.n.], 2017. p. 1–4.
- HAYKIN, S. **Redes Neurais - 2ed**. Bookman, 2001. ISBN 9788573077186. Disponível em: <<https://books.google.com.br/books?id=1Bp0X5qfyjUC>>.
- HUSH, D. T.; HORNE, B. G. Progress in supervised neural networks: what’s new since lippmann? **IEEE Signal Processing Magazine**, v. 10, n. 1, p. 8–39, 1 1993.
- I.C.P.RUSSO; BEHLAU, M. **Percepção da fala: análise acústica do português brasileiro**. [S.l.]: Editora Lovise, 1993.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **CoRR**, 2015.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI’95), p. 1137–1143. ISBN 1-55860-363-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=1643031.1643047>>.
- KOHAVI, R.; PROVOST, F. Glossary of terms. **Machine Learning**, v. 30, n. 2-3, p. 271–274, 1998. Disponível em: <<http://www.springerlink.com/content/x105525142v51t20/>>.
- KOOLAGUDI, S. G.; RASTOGI, D.; RAO, K. S. Identification of language using mel-frequency cepstral coefficients (mfcc). **Procedia Engineering**, v. 38, p. 3391 – 3398, 2012.
- KRIK, V. M. et al. Análise acústica da voz: comparação entre dois tipos de microfones. **Audiology - Communication Research**, scielo, v. 24, 00 2019.
- KUHL, I. **Manual Prático de Laringologia**. 1. ed. [S.l.]: Editora da Universidade – UFRGS, Porto Alegre, 1982.
- LARSON, R.; FARBER, B. **Estatística Aplicada**. 1. ed. [S.l.]: Pearson Education do Brasil, 2004.

- LEE, J. H.; DELBRUCK, T.; PFEIFFER, M. Training deep spiking neural networks using backpropagation. **Frontiers in Neuroscience**, v. 10, 08 2016.
- LEVIN, J. **Estatística aplicada a ciencias humanas**. Harbra, 1978. ISBN 9788529402079. Disponível em: <<https://books.google.com.br/books?id=zkdwugEACAAJ>>.
- LI, Y. et al. Deep learning based gastric cancer identification. In: **2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)**. [S.l.: s.n.], 2018. p. 182–185. ISSN 1945-8452.
- LIPPMANN, R. P. An introduction to computing with neural nets. **IEEE ASSP Magazine**, v. 4, n. 2, p. 4–22, 4 1987.
- LIU, X. W. e B. Wei e D. A autocorrelação geral e sua aplicação. In: SOCIETY, I. C. (Ed.). **Anais do Simpósio Internacional 2002 Mundos cibernéticos: teoria e práticas. CW2002**. Los Alamitos, CA, EUA: [s.n.], 2002. p. 0267. Disponível em: <<https://doi.ieeeecomputersociety.org/10.1109/CW.2002.1180889>>.
- LOPES, L. W. et al. Relationship between acoustic measurements and self-evaluation in patients with voice disorders. **Vf Journal of Voice**, v. 31, n. 1, p. 119 – 129, 2017.
- LUO, P. et al. A moving average filter based method of performance improvement for ultraviolet communication system. In: . [S.l.: s.n.], 2012.
- MACIEL, C. T. V. et al. Análise da qualidade de vida dos pacientes com câncer de laringe em hospital de referência na região sudeste do brasil. **Revista CEFAC**, v. 15, p. 932 – 940, 2013. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1516-18462013000400022&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-18462013000400022&nrm=iso)>.
- MADAZIO, G.; LEAO, S.; BEHLAU, M. The phonatory deviation diagram: A novel objective measurement of vocal function. **Folia Phoniatica et Logopaedica**, v. 3, n. 6, p. 305–311, 2011.
- MAGI, C. et al. Stabilised weighted linear prediction. **Speech Communication**, v. 51, n. 5, p. 401 – 411, 2009. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167639308001799>>.
- MARTINS, R. H. G. et al. Edema de reinke: estudo da imunoexpressão da fibronectina, da laminina e do colágeno iv em 60 casos por meio de técnicas imunoistoquímica. **Braz. j. otorhinolaryngol.**, v. 75, n. 6, p. 821–825, 2009.
- MARTINS, R. H. G. et al. Reinke's edema precancerous lesion? histological and electron microscopic aspects. **J. Voice**, v. 23, n. 6, p. 721–725, 2009.
- MARYN, Y.; ROY, N. N. sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. **Jornal da Sociedade Brasileira de Fonoaudiologia**, v. 2, n. 24, p. 107–112, 2012.
- MEDINA, V.; SIMOES-ZERANI, M.; NEMR, N. K. Análise vocal acústica: efeito do treinamento auditivo-visual para graduandos de fonoaudiologia. **Audiology - Communication Research**, v. 2, n. 20, p. 123–129, 2015.
- MENDOZA, M. E. et al. Acoustic andperceptual indicators of normal and pathological voice. **Folia Phoniatica et Logopaedica**, v. 55, n. 2, p. 102–114, 2003.

MOON, J.; LEE, J. Development of medical/electrical convergence software for classification between normal and pathological voices. **Journal of Digital Convergence**, v. 13, p. 187–192, 12 2015.

MORETTIN, P. A.; TOLOI, C. M. de C. **Análise de séries temporais**. [S.l.]: Edgard Blucher, 2006.

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. ISSN 1558-1896.

MUKKAMALA, M. C.; HEIN, M. Variants of RMSProp and Adagrad with logarithmic regret bounds. In: PRECUP, D.; TEH, Y. W. (Ed.). **Proceedings of the 34th International Conference on Machine Learning**. International Convention Centre, Sydney, Australia: PMLR, 2017. (Proceedings of Machine Learning Research, v. 70), p. 2545–2553. Disponível em: <<http://proceedings.mlr.press/v70/mukkamala17a.html>>.

NIED, A. **Treinamento de redes neurais artificiais baseado em sistemas de estrutura variável com taxa de aprendizado adaptativa**. Tese (Doutorado) — Universidade Federal de Minas Gerais, Belo Horizonte - MG, Brasil, 2007. 128f.

PANCHAL, F. S.; PANCHAL, M. Review on methods of selecting number of hidden nodes in artificial neural network. **International Journal of Computer Science and Mobile Computing**, v. 3, n. 11, p. 455–464, 2014.

PINHO, P. H. U. de. **Avaliação de distúrbios vocais por meio de características do espaço de fase reconstruído**. Dissertação (Mestrado) — Instituto Federal da Paraíba, João Pessoa, PB, abril. 2017. 139.

QIU, J. X. et al. Deep learning for automated extraction of primary sites from cancer pathology reports. **IEEE Journal of Biomedical and Health Informatics**, v. 22, n. 1, p. 244–251, Jan 2018. ISSN 2168-2194.

QUEIROZ, G. K. L. P. de et al. Classificação de disfonias por meio da análise de medidas não lineares e de quantificação de recorrência. **Revista Principia**, v. 38, n. 1, p. 148–160, 2018.

RABEH, H.; SALAH, H.; ADNANE, C. Voice pathology recognition and classification using noise related features. **International Journal of Advanced Computer Science and Applications**, v. 9, n. 11, p. 82 – 87, 2018.

RASCHKA, S. **Python Machine Learning**. [S.l.]: Packt Publishing, 2015. ISBN 1783555130, 9781783555130.

RIEDMILLER, M.; BRAUN, H. A direct adaptive method for faster backpropagation learning: The rprop algorithm. **Proceedings of the Int. Conf. on Neural Networks, San Francisco, CA**, p. 586–591, 1993.

RUDER, S. An overview of gradient descent optimization algorithms. **CoRR**, abs/1609.04747, 2016.

RUFINER, H. L. **Análisis y representación de la voz mediante técnicas no convencionales**. Tese (Doutorado) — Universidade Federal de Buenos Aires, Buenos Aires, ARG, 2005.

- RUMELHART, D. E.; MCCLELLAND, J. L. **Parallel distributed processing**. [S.l.]: San Diego, MIT Press, 1986.
- SCHIRRMESTER, R. et al. Deep learning with convolutional neural networks for decoding and visualization of eeg pathology. In: **2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)**. [S.l.: s.n.], 2017. p. 1–7. ISSN 2473-716X.
- SELLAM, V.; JAGADEESAN, J. Classification of normal and pathological voice using svm and rbfn. **Journal of Signal and Information Processing**, v. 05, p. 1–7, 01 2014.
- SHAH, M. et al. Deep learning assessment of tumor proliferation in breast cancer histological images. In: **2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**. [S.l.: s.n.], 2017. p. 600–603.
- SILVA, D. A. da. **Aplicação de técnicas de pré-processamento e agrupamento na base de dados de benefícios previdenciários do Ministério Público do Trabalho**. 2018. 67f. Trabalho de Conclusão de Curso, Universidade Federal de Uberlândia, MG, 2018.
- SILVA, F. M.; ALMEIDA, L. B. Speeding up backpropagation, in: eckmiller (ed.). **Advanced Neural Computers**, Amsterdam: Elsevier North Holland, p. 151–158, 1990.
- SILVA, S. S. L. da. Principais patologias laríngeas em professores. **Revista Distúrbios da Comunicação**, v. 30, n. 4, p. 767–775, 2018.
- SOUISSI, N.; CHERIF, A. Speech recognition system based on short-term cepstral parameters, feature reduction method and artificial neural networks. In: **2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)**. [S.l.: s.n.], 2016. p. 667–671. ISSN null.
- SOUSA, M. de. **Descritores Morfológicos para Identificação Automática de Padrões Epileptiformes em Sinais de EEG**. Dissertação (Mestrado) — Universidade Federal de Santa Catarina, Florianópolis, SC, 2016. 118p.
- SOUZA, T. A. de. **Aplicação da Transformada Wavelet em Análise de Texturas de Gráficos de Recorrência para Detecção de Patologias Laríngeas**. Dissertação (Mestrado) — Instituto Federal da Paraíba, João Pessoa, PB, abril. 2015. 81p.
- SRIVASTAVA, A. et al. Deep learning for detecting diseases in gastrointestinal biopsy images. In: **2019 Systems and Information Engineering Design Symposium (SIEDS)**. [S.l.: s.n.], 2019. p. 1–4.
- STEFFEN, L. M. et al. Paralisia unilateral de prega vocal: associação e correlação entre tempos máximos de fonação, posição e ângulo de afastamento. **Revista Brasileira de Otorrinolaringologia**, scielo, v. 70, p. 450 – 455, 08 2004.
- STEMPLE, J. C.; KLABEN, B. **Clinical Voice Pathology: Theory and management**. [S.l.]: 4th ed, Plural Publishing, 2010.
- TIMOSZCZUK, A. P. **Reconhecimento automático do locutor com redes neurais pulsadas**. Tese (Doutorado) — Universidade de São Paulo, São Paulo, SP, 2004. 237f.
- TORRES, W. de L. **Detecção de Desvios Vocais Utilizando Modelos Auto Regressivos e o Algoritmo KNN**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte - UFRN, Natal, RN, janeiro. 2018. 80p.

TURAN, S.; BILGIN, G. Semantic nuclei segmentation with deep learning on breast pathology images. In: **2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT)**. [S.l.: s.n.], 2019. p. 1–4.

ULIZA, V. et al. Validation of the acoustic voice quality index in the lithuanian language. **Vf Journal of Voice**, v. 31, n. 2, p. 257 – 268, 2016.

VAHADANE, A. et al. Detection of diabetic macular edema in optical coherence tomography scans using patch based deep learning. In: **2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)**. [S.l.: s.n.], 2018. p. 1427–1430. ISSN 1945-8452.

VALENTIM, A. F.; CÔRTEZ, M. G.; GAMA, A. C. C. Análise espectrográfica da voz: efeito do treinamento visual na confiabilidade da avaliação. **Revista da Sociedade Brasileira de Fonoaudiologia**, v. 15, p. 335 – 342, 2010. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1516-80342010000300005&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-80342010000300005&nrm=iso)>.

VIEIRA, V. J. D. **Avaliação de Distúrbios da Voz por meio de Análise de Quantificação de Recorrência**. Dissertação (Mestrado) — Instituto Federal da Paraíba - IFPB, João Pessoa, PB, fevereiro. 2014. 218p.

VIEIRA, V. J. D. et al. Avaliação de desempenho na classificação de patologias laríngeas por análise lpc de sinais de voz e redes neurais mlp. **Anais do XIII Congresso Brasileiro de Inteligência Computacional**, 2013.

WITTEN, I.; FRANK, E.; HALL, M. **Data Mining: Practical machine learning tools and techniques**. [S.l.]: Morgan Kaufmann, 2011.

WOLK t. K.; WOLK, A. Early and remote detection of possible heartbeat problems with convolutional neural networks and multipart interactive training. **IEEE Access**, p. 1–1, 2019. ISSN 2169-3536.

WU, H. et al. A deep learning method for pathological voice detection using convolutional deep belief networks. In: . [S.l.: s.n.], 2018. p. 446–450.

ZEILER, M. D. Adadelta: An adaptive learning rate method. v. 1212, 12 2012.

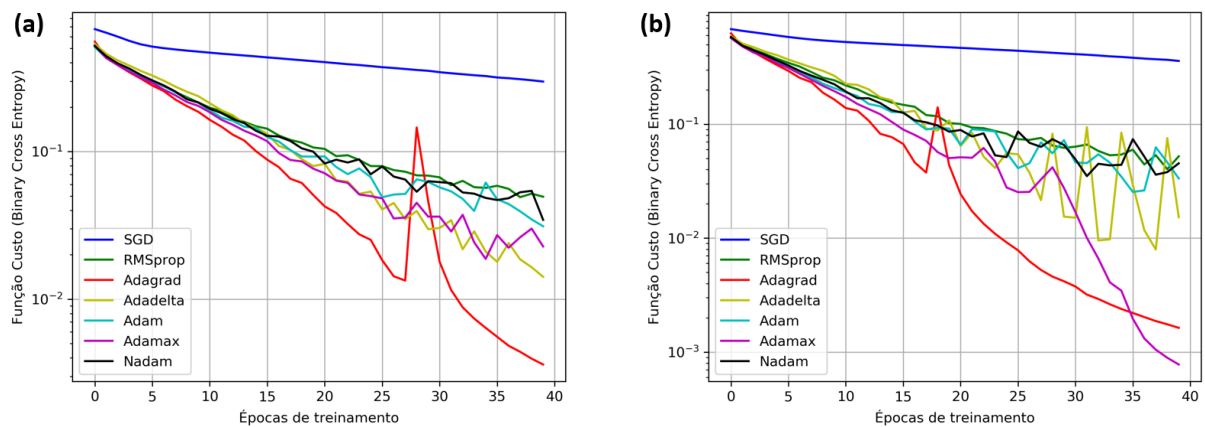
ZRAICK, R. I.; WENDEL, K.; SMITH-OLINDE, L. The effect of speaking task on perceptual judgment of the severity of dysphonic voice. **Journal of Voice**, v. 19, n. 4, p. 574–581, 2005.



## APÊNDICE A – CURVAS DE APRENDIZAGEM DURANTE O TREINAMENTO DOS MELHORES CLASSIFICADORES

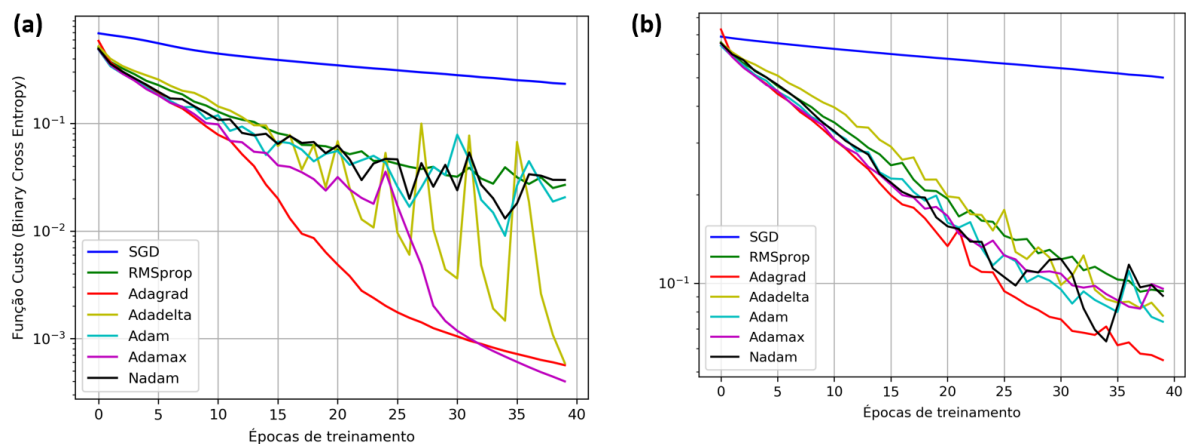
Os gráficos presentes nas Figuras A.1 e A.2 apresentam a comparação entre os algoritmos de aprendizagem durante o treinamento dos melhores classificadores. As curvas são relacionadas a aprendizagem com os dados referentes ao *fold* 1 de treinamento de cada processo discriminativo. O comportamento das curvas de aprendizagem são semelhantes entre os *folders* de cada processo discriminativo.

**Figura A.1** – Comparação entre as curvas de aprendizagem para o melhor classificador da discriminação: (a) saudável *versus* patológicos; (b) saudável *versus* patologias organofuncionais.



Fonte: Acervo Pessoal.

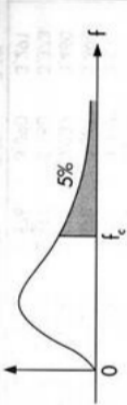
**Figura A.2** – Comparação entre as curvas de aprendizagem para o melhor classificador da discriminação: (a) saudável *versus* paralisia das pregas vocais; (b) patologias organofuncionais *versus* paralisia das pregas vocais.



Fonte: Acervo Pessoal.

# ANEXO A – TABELA DE REFERÊNCIA PARA ESTIMAR O VALOR DE $F_{critico}$ PARA O TESTE DE SIGNIFICÂNCIA ANOVA PARA UM INTERVALO DE CONFIANÇA DE 95%

**Tabela VI – Distribuição F**  
Corpo da tabela dá os valores  $f_c$  tais que  $P(F > f_c) = 0,05$ .



| Graus de liberdade do denominador de $F: v_2$ | Grau de liberdade do numerador de $F: v_1$ |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | ∞     |       |
|---|--|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|   | 1  | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 12    | 14    | 15    | 16    | 18    | 20    | 24    | 30    | 40    | 60    |       | 120   |
| 1   | 161,4                                      | 199,5 | 215,7 | 224,6 | 230,2 | 234,0 | 236,8 | 238,9 | 240,5 | 241,9 | 243,9 | 245,4 | 245,9 | 246,5 | 247,3 | 248,0 | 249,1 | 250,1 | 251,1 | 252,2 | 253,3 | 254,3 |
| 2   | 18,51                                      | 19,00 | 19,16 | 19,25 | 19,30 | 19,33 | 19,35 | 19,37 | 19,38 | 19,40 | 19,41 | 19,42 | 19,43 | 19,43 | 19,44 | 19,45 | 19,45 | 19,46 | 19,47 | 19,48 | 19,49 | 19,50 |
| 3   | 10,13                                      | 9,55  | 9,28  | 9,12  | 9,01  | 8,94  | 8,89  | 8,85  | 8,81  | 8,79  | 8,74  | 8,72  | 8,70  | 8,69  | 8,67  | 8,66  | 8,64  | 8,62  | 8,59  | 8,57  | 8,55  | 8,53  |
| 4   | 7,71                                       | 6,94  | 6,59  | 6,39  | 6,26  | 6,16  | 6,09  | 6,04  | 6,00  | 5,96  | 5,91  | 5,87  | 5,86  | 5,84  | 5,82  | 5,80  | 5,77  | 5,75  | 5,72  | 5,69  | 5,66  | 5,63  |
| 5   | 6,61                                       | 5,79  | 5,41  | 5,19  | 5,05  | 4,95  | 4,88  | 4,82  | 4,77  | 4,74  | 4,68  | 4,64  | 4,62  | 4,60  | 4,58  | 4,56  | 4,53  | 4,50  | 4,46  | 4,43  | 4,40  | 4,36  |
| 6   | 5,99                                       | 5,14  | 4,76  | 4,53  | 4,39  | 4,28  | 4,21  | 4,15  | 4,10  | 4,06  | 4,00  | 3,96  | 3,94  | 3,92  | 3,90  | 3,87  | 3,84  | 3,81  | 3,77  | 3,74  | 3,70  | 3,67  |
| 7   | 5,59                                       | 4,74  | 4,35  | 4,12  | 3,97  | 3,87  | 3,79  | 3,73  | 3,68  | 3,64  | 3,57  | 3,53  | 3,51  | 3,49  | 3,47  | 3,44  | 3,41  | 3,38  | 3,34  | 3,30  | 3,27  | 3,23  |
| 8   | 5,32                                       | 4,46  | 4,07  | 3,84  | 3,69  | 3,58  | 3,50  | 3,44  | 3,39  | 3,35  | 3,28  | 3,24  | 3,22  | 3,20  | 3,17  | 3,15  | 3,12  | 3,08  | 3,04  | 3,01  | 2,97  | 2,93  |
| 9   | 5,12                                       | 4,26  | 3,86  | 3,63  | 3,48  | 3,37  | 3,29  | 3,23  | 3,18  | 3,14  | 3,07  | 3,03  | 3,01  | 2,99  | 2,96  | 2,94  | 2,90  | 2,86  | 2,83  | 2,79  | 2,75  | 2,71  |
| 10  | 4,96                                       | 4,10  | 3,71  | 3,48  | 3,33  | 3,22  | 3,14  | 3,07  | 3,02  | 2,98  | 2,91  | 2,87  | 2,85  | 2,83  | 2,80  | 2,77  | 2,74  | 2,70  | 2,66  | 2,62  | 2,58  | 2,54  |
| 11  | 4,84                                       | 3,98  | 3,59  | 3,36  | 3,20  | 3,09  | 3,01  | 2,95  | 2,90  | 2,85  | 2,79  | 2,74  | 2,72  | 2,70  | 2,67  | 2,65  | 2,61  | 2,57  | 2,53  | 2,49  | 2,45  | 2,40  |
| 12  | 4,75                                       | 3,89  | 3,49  | 3,26  | 3,11  | 3,00  | 2,91  | 2,85  | 2,80  | 2,75  | 2,69  | 2,64  | 2,62  | 2,60  | 2,57  | 2,54  | 2,51  | 2,47  | 2,43  | 2,38  | 2,34  | 2,30  |
| 13  | 4,67                                       | 3,81  | 3,41  | 3,18  | 3,03  | 2,92  | 2,83  | 2,77  | 2,71  | 2,67  | 2,60  | 2,55  | 2,53  | 2,52  | 2,48  | 2,46  | 2,42  | 2,38  | 2,34  | 2,30  | 2,25  | 2,21  |
| 14  | 4,60                                       | 3,74  | 3,34  | 3,11  | 2,96  | 2,85  | 2,76  | 2,70  | 2,65  | 2,60  | 2,54  | 2,48  | 2,46  | 2,44  | 2,41  | 2,39  | 2,35  | 2,31  | 2,27  | 2,22  | 2,18  | 2,13  |
| 15  | 4,54                                       | 3,68  | 3,29  | 3,06  | 2,90  | 2,79  | 2,71  | 2,64  | 2,59  | 2,54  | 2,48  | 2,42  | 2,40  | 2,39  | 2,35  | 2,33  | 2,29  | 2,25  | 2,20  | 2,16  | 2,11  | 2,07  |
| 16  | 4,49                                       | 3,63  | 3,24  | 3,01  | 2,85  | 2,74  | 2,66  | 2,59  | 2,54  | 2,49  | 2,42  | 2,37  | 2,35  | 2,33  | 2,30  | 2,28  | 2,24  | 2,19  | 2,15  | 2,11  | 2,06  | 2,01  |
| 17  | 4,45                                       | 3,59  | 3,20  | 2,96  | 2,81  | 2,70  | 2,61  | 2,55  | 2,49  | 2,45  | 2,38  | 2,34  | 2,31  | 2,29  | 2,26  | 2,23  | 2,19  | 2,15  | 2,10  | 2,06  | 2,01  | 1,96  |
| 18  | 4,41                                       | 3,55  | 3,16  | 2,93  | 2,77  | 2,66  | 2,58  | 2,51  | 2,46  | 2,41  | 2,34  | 2,29  | 2,27  | 2,25  | 2,22  | 2,19  | 2,15  | 2,11  | 2,06  | 2,02  | 1,97  | 1,92  |
| 19  | 4,38                                       | 3,52  | 3,13  | 2,90  | 2,74  | 2,63  | 2,54  | 2,48  | 2,42  | 2,38  | 2,31  | 2,26  | 2,23  | 2,22  | 2,18  | 2,16  | 2,11  | 2,07  | 2,03  | 1,98  | 1,93  | 1,88  |
| 20  | 4,35                                       | 3,49  | 3,10  | 2,87  | 2,71  | 2,60  | 2,51  | 2,45  | 2,39  | 2,35  | 2,28  | 2,22  | 2,20  | 2,18  | 2,15  | 2,12  | 2,08  | 2,04  | 1,99  | 1,95  | 1,90  | 1,84  |
| 21  | 4,32                                       | 3,47  | 3,07  | 2,84  | 2,68  | 2,57  | 2,49  | 2,42  | 2,37  | 2,32  | 2,25  | 2,20  | 2,18  | 2,16  | 2,12  | 2,10  | 2,05  | 2,01  | 1,96  | 1,92  | 1,87  | 1,81  |
| 22  | 4,30                                       | 3,44  | 3,05  | 2,82  | 2,66  | 2,55  | 2,46  | 2,40  | 2,34  | 2,30  | 2,23  | 2,17  | 2,15  | 2,13  | 2,10  | 2,07  | 2,03  | 1,98  | 1,94  | 1,89  | 1,84  | 1,78  |
| 23  | 4,28                                       | 3,42  | 3,03  | 2,80  | 2,64  | 2,53  | 2,44  | 2,37  | 2,32  | 2,27  | 2,20  | 2,15  | 2,13  | 2,11  | 2,08  | 2,05  | 2,01  | 1,96  | 1,91  | 1,86  | 1,81  | 1,76  |
| 24  | 4,26                                       | 3,40  | 3,01  | 2,78  | 2,62  | 2,51  | 2,42  | 2,36  | 2,30  | 2,25  | 2,18  | 2,13  | 2,11  | 2,09  | 2,05  | 2,03  | 1,98  | 1,94  | 1,89  | 1,84  | 1,79  | 1,73  |
| 25  | 4,24                                       | 3,39  | 2,99  | 2,76  | 2,60  | 2,49  | 2,40  | 2,34  | 2,28  | 2,24  | 2,16  | 2,11  | 2,09  | 2,07  | 2,04  | 2,01  | 1,96  | 1,92  | 1,87  | 1,82  | 1,77  | 1,71  |
| 26  | 4,23                                       | 3,37  | 2,98  | 2,74  | 2,59  | 2,47  | 2,39  | 2,32  | 2,27  | 2,22  | 2,15  | 2,09  | 2,07  | 2,05  | 2,02  | 1,99  | 1,95  | 1,90  | 1,85  | 1,80  | 1,75  | 1,69  |
| 27  | 4,21                                       | 3,35  | 2,96  | 2,73  | 2,57  | 2,46  | 2,37  | 2,31  | 2,25  | 2,20  | 2,13  | 2,08  | 2,06  | 2,04  | 2,00  | 1,97  | 1,93  | 1,88  | 1,84  | 1,79  | 1,73  | 1,67  |
| 28  | 4,20                                       | 3,34  | 2,95  | 2,71  | 2,56  | 2,45  | 2,36  | 2,29  | 2,24  | 2,19  | 2,12  | 2,06  | 2,04  | 2,02  | 1,99  | 1,96  | 1,91  | 1,87  | 1,82  | 1,77  | 1,71  | 1,65  |
| 29  | 4,18                                       | 3,33  | 2,93  | 2,70  | 2,55  | 2,43  | 2,35  | 2,28  | 2,22  | 2,18  | 2,10  | 2,05  | 2,03  | 2,01  | 1,97  | 1,94  | 1,90  | 1,85  | 1,81  | 1,75  | 1,70  | 1,64  |
| 30  | 4,17                                       | 3,32  | 2,92  | 2,69  | 2,53  | 2,42  | 2,33  | 2,27  | 2,21  | 2,16  | 2,09  | 2,04  | 2,01  | 1,99  | 1,96  | 1,93  | 1,89  | 1,84  | 1,79  | 1,74  | 1,68  | 1,62  |
| 40  | 4,08                                       | 3,23  | 2,84  | 2,61  | 2,45  | 2,34  | 2,25  | 2,18  | 2,12  | 2,08  | 2,00  | 1,95  | 1,92  | 1,90  | 1,87  | 1,84  | 1,79  | 1,74  | 1,69  | 1,64  | 1,58  | 1,51  |
| 60  | 4,00                                       | 3,15  | 2,76  | 2,53  | 2,37  | 2,25  | 2,17  | 2,10  | 2,04  | 1,99  | 1,92  | 1,86  | 1,84  | 1,81  | 1,78  | 1,75  | 1,70  | 1,65  | 1,59  | 1,53  | 1,47  | 1,39  |
| 120   | 3,92                                       | 3,07  | 2,68  | 2,45  | 2,29  | 2,17  | 2,09  | 2,02  | 1,96  | 1,91  | 1,83  | 1,77  | 1,75  | 1,72  | 1,69  | 1,66  | 1,61  | 1,55  | 1,50  | 1,43  | 1,35  | 1,25  |
| ∞   | 3,84                                       | 3,00  | 2,60  | 2,37  | 2,21  | 2,10  | 2,01  | 1,94  | 1,88  | 1,83  | 1,75  | 1,69  | 1,67  | 1,63  | 1,60  | 1,57  | 1,52  | 1,46  | 1,39  | 1,32  | 1,22  | 1,00  |