

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE ELÉTRICA
CURSO DE ENGENHARIA ELÉTRICA

MURILLO AUGUSTO FERMINO

**CLASSIFICAÇÃO DE DISTÚRBIOS VOCAIS
UTILIZANDO REDES NEURAS ARTIFICIAIS**

TRABALHO DE CONCLUSÃO DE CURSO

CORNÉLIO PROCÓPIO
2017

MURILLO AUGUSTO FERMINO

**CLASSIFICAÇÃO DE DISTÚRBIOS VOCAIS
UTILIZANDO REDES NEURAS ARTIFICIAIS**

Trabalho de Conclusão de Curso de graduação do curso de Engenharia Elétrica da Universidade Tecnológica Federal do Paraná – UTFPR, como requisito parcial para a obtenção do título de Bacharel.

Orientadora: Prof. Dra. María Eugenia Dajer
Coorientador: Prof. Dr. Danilo Hernane Spatti

CORNÉLIO PROCÓPIO
2017



Universidade Tecnológica Federal do Paraná
Campus Cornélio Procópio
Departamento Acadêmico de Elétrica
Curso de Engenharia Elétrica



FOLHA DE APROVAÇÃO

Murillo Augusto Fermino

Classificação de distúrbios vocais utilizando redes neurais artificiais

Trabalho de conclusão de curso apresentado às 8:20hs do dia 20/06/2017 como requisito parcial para a obtenção do título de Engenheiro Eletricista no programa de Graduação em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná. O candidato foi arguido pela Banca Avaliadora composta pelos professores abaixo assinados. Após deliberação, a Banca Avaliadora considerou o trabalho aprovado.

Prof(a). Dr(a). María Eugenia Dajer - Presidente (Orientador)

Prof(a). Dr(a). Danilo Hernane Spatti - (Coorientador)

Prof(a). Dr(a). Bruno Catarino Bispo - (Membro)

Prof(a). Dr(a). Wagner Endo - (Membro)

*Com muito amor dedico este trabalho
aos meus familiares, a minha namorada
e a todas as pessoas mais que especiais
presentes em minha vida por seu amor, carinho
e paciência durante os anos de graduação.*

AGRADECIMENTOS

Gostaria de agradecer, primeiramente, a todos os meus professores que me guiaram até aqui, em especial ao Prof. Danilo Hernane Spatti e Prof. María Eugenia Dajer, que desde o início foram grandes conselheiros e hoje, no término desta jornada, tornaram-se verdadeiros amigos.

À minha família, em especial à minha mãe Neusa, meu pai Adilson e minha irmã Dayane pelo amor e apoio incondicional que sempre me concederam.

À minha namorada Beatriz pelo amor, carinho, companheirismo e compreensão nos períodos de ausência.

Aos amigos, em especial ao Higor e Andriéli, sempre presentes durante todos os anos de graduação, me apoiando e incentivando nos momentos que eu mais precisava.

À todas as outras pessoas que, de uma forma ou de outra, participaram de maneira especial para essa conquista, fazendo parte das boas recordações que guardarei para sempre em minha memória.

E, por fim, ao Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GPEM/CNPq) pela grande contribuição com os dados utilizados no desenvolvimento deste trabalho.

Muito obrigado!

*“It is paradoxical, yet true, to say that the more we know,
the more ignorant we become in the absolute sense,
for it is only through enlightenment that we become conscious of our limitations.
Precisely one of the most gratifying results of intellectual evolution
is the continuous opening up of new and greater prospects.”*

Nikola Tesla

RESUMO

FERMINO, Murillo Augusto. **Classificação de distúrbios vocais utilizando redes neurais artificiais**. 2017. 99 f. Trabalho de Conclusão de Curso (Graduação) – Engenharia Elétrica – Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2017.

A voz é um importante meio de comunicação, afetando o desempenho de profissionais em todas as áreas, exigindo extremo cuidado com a saúde vocal. No entanto, existem muitos desafios nos métodos de avaliação vocal, fornecendo oportunidades de estudo para abordagens não-invasivas. Em linha com esta exigência, o presente trabalho aplica redes neurais artificiais para classificação dos padrões relacionados a distúrbios vocais em sinais de fala. Propõe-se a utilização da transformada *wavelet packet* para extrair características do sinal em análise tempo-frequência, redes neurais artificiais para reconhecimento e classificação dos padrões apresentados por distúrbios vocais. Por fim, o classificador modelado se mostrou consistente nos testes desenvolvidos, provando-se promissor para uma futura aplicação como método não-invasivo que auxilie os profissionais da voz.

Palavras-chave: Distúrbios Vocais. Redes Neurais Artificiais. Transformada *Wavelet Packet*. Reconhecimento de Padrões.

ABSTRACT

FERMINO, Murillo Augusto. **Classification of voice disorders with artificial neural networks**. 2017. 99 f. Trabalho de Conclusão de Curso (Graduação) – Engenharia Elétrica – Federal Technology University - Paraná. Cornélio Procópio, 2017.

The voice is an important means of communication, affecting the performance of professionals in all areas, requiring extreme care with vocal health. However, there are many challenges in voice assessment methods, providing research opportunities for non-invasive approaches. In line with this requirement, this paper investigates the application of artificial neural network to classify patterns related to voice disorders in speech signals. It is proposed to use the wavelet packet transform to extract signal characteristics in time-frequency analysis, artificial neural networks for recognition and classification of the patterns presented by vocal disorders. Finally, the modeling intelligent classifier has been shown to be consistent in the developed tests, proving itself promising for a future application as a non-invasive method that assists voice professionals.

Keywords: Voice Disorder. Artificial Neural Networks. Wavelet Packet Transform. Pattern Recognition.

LISTA DE FIGURAS

Figura 1 – Pregas vocais durante um ciclo glótico normal.....	20
Figura 2 – Visualização esquemática do trato vocal.	21
Figura 3 – Representação da forma de onda vocal e seus intervalos.	23
Figura 4 – Sinal vocal durante elocução da frase “ <i>should we chase</i> ”.	23
Figura 5 – Estimativa do período de pitch para a forma de onda.	24
Figura 6 – Equipamento necessário para realização da videolaringoestroboscopia.....	25
Figura 7 – Quadros capturados identificando fases distintas do ciclo vocal.	26
Figura 8 – Edema de Reinke.....	29
Figura 9 – Exemplo de Filtro Passa-Baixa.	32
Figura 10 – Exemplo de Filtro Passa-Alta.	32
Figura 11 – Exemplo de Filtro Passa-Banda.	33
Figura 12 – Exemplo de Filtro Rejeita-Banda.....	33
Figura 13 – Exemplo de Ambiguidade na Amostragem.	34
Figura 14 – Exemplo de do processo de <i>upsampling</i>	36
Figura 15 – Exemplo de <i>downsampling</i>	36
Figura 16 – Transformada de Fourier no sinal de voz.....	37
Figura 17 – Transformada de Fourier de tempo curto.....	38
Figura 18 – Transformada <i>wavelet</i>	39
Figura 19 – Funcionamento da transformada <i>wavelet</i>	40
Figura 20 – <i>Downsampling</i> presente na decomposição.	40
Figura 21 – Decomposição da transformada <i>wavelet packet</i> em árvore binária.	41
Figura 22 – Representação do neurônio artificial.....	48
Figura 23 – Exemplo de rede <i>feedforward</i> de camada única.	51
Figura 24 – Exemplo de rede <i>feedforward</i> de camada múltiplas.	52
Figura 25 – Exemplo de rede recorrente.	52
Figura 26 – Fluxograma simplificado das etapas de um classificador de padrões.....	58
Figura 27 – Exemplo de sinal completo a ser janelado.	60
Figura 28 – Exemplo de sinal após procedimento de janelamento simples.	60
Figura 29 – Exemplo de sinal com janelamento cruzado.....	61
Figura 30 – Sinal completo para exemplificação.	62
Figura 31 – Sinal segmentado em 4096 pontos para exemplificação.....	63
Figura 32 – Exemplo de decomposição da TWP até terceiro nível.....	63
Figura 33 – Exemplo de folha de aproximação.....	64
Figura 34 – Exemplo de folha de detalhe.	64
Figura 35 – Fluxograma do processo de preparação dos dados.	66
Figura 35 – Fluxograma de extração de características.....	68
Figura 37 – Fluxograma do processo automatizado de treinamento das RNA.	71
Figura 38 – Fluxograma de análise estatística da robustez das RNA.....	72
Figura 39 – Testes de topologia (nível 2) – melhor de 10 treinamentos.	74
Figura 40 – Resultado de acertos das 7 melhores topologias (nível 2).	75
Figura 41 – Testes de topologia (nível 3) – melhor de 10 treinamentos.	76
Figura 42 – Resultado de acertos das 7 melhores topologias (nível 3).	77
Figura 45 – Histograma dos 60 treinamentos adicionais da melhor topologia.	79
Figura 46 – Imprecisões na validação da melhor topologia (5102 amostras).	81

Figura 47 – Resultados da validação da melhor topologia (5102 amostras).....	82
Figura 48 – Histograma de erro na validação da melhor topologia (5102 amostras).....	83
Figura 49 – Resposta para amostras desconhecidas – arquivo do tipo edema (I).	87
Figura 50 – Resposta para amostras desconhecidas – arquivo do tipo edema (II).....	88
Figura 51 – Resposta para amostras desconhecidas – arquivo do tipo edema (III).	89
Figura 52 – Resposta para amostras desconhecidas – arquivo do tipo saudável (I).....	90
Figura 53 – Resposta para amostras desconhecidas – arquivo do tipo saudável (II).	91
Figura 54 – Resposta para amostras desconhecidas – arquivo do tipo saudável (III).	92

LISTA DE QUADROS

Quadro 1 – Sistemas e estruturas participantes da produção vocal.	19
Quadro 2 – Pontos fortes em diversas metodologias inteligentes.	44
Quadro 3 – Potenciais áreas de aplicação das RNAs.	47
Quadro 4 – Principais funções de ativação.	50
Quadro 5 – Composição do banco de dados.	56
Quadro 6 – Características da RNA desenvolvida.	70
Quadro 7 – Variações de topologia.	70
Quadro 8 – Taxas de acertos (<i>underfitting</i> nível 3) – melhor de 3 treinamentos.	78
Quadro 9 – Critérios de pós-processamento.	80
Quadro 10 – Matriz de confusão.	84
Quadro 11 – Resultados – arquivo do tipo edema (I).	87
Quadro 12 – Resultados – arquivo do tipo edema (II).	88
Quadro 13 – Resultados – arquivo do tipo edema (III).	89
Quadro 14 – Resultados – arquivo do tipo saudável (I).	90
Quadro 15 – Resultados – arquivo do tipo saudável (II).	91
Quadro 16 – Resultados – arquivo do tipo saudável (III).	92

LISTA DE SIGLAS, ABREVIATURAS E ACRÔNOMOS

HC-FMUSP	Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo
GPEM/CNPq	Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico
RNA	Rede Neural Artificial
STFT	Short Time Fourier Transform
TWP	Transformada <i>Wavelet Packet</i>

SUMÁRIO

1	INTRODUÇÃO	13
1.1	PROPOSTA DO TRABALHO	15
1.2	PUBLICAÇÕES OBTIDAS	15
1.3	OBJETIVOS.....	15
1.3.1	OBJETIVO GERAL.....	15
1.3.2	OBJETIVOS ESPECÍFICOS	16
1.4	ORGANIZAÇÃO DO TRABALHO	16
2	ESTUDOS DA VOZ.....	18
2.1	CONCEITOS INICIAIS.....	18
2.2	PRODUÇÃO DA VOZ.....	19
2.3	PROPRIEDADES E FORMATO DA VOZ	22
2.4	MÉTODOS DE AVALIAÇÃO VOCAL.....	24
2.4.1	Métodos Invasivos.....	24
2.4.2	Métodos Não-Invasivos.....	26
2.5	DISTÚRBIOS VOCAIS.....	27
2.5.1	Edema de Reinke.....	28
2.6	CONSIDERAÇÕES	29
3	ESTUDOS DE SINAIS E SISTEMAS INTELIGENTES	30
3.1	PROCESSAMENTO DIGITAL DE SINAIS	30
3.1.1	Conceitos Iniciais	30
3.1.2	Filtros.....	31
3.1.3	Conceito de Amostragem	33
3.1.4	Transformações dos Sinais.....	37
3.1.5	Transformada de Fourier	37
3.1.6	Transformada <i>wavelet packet</i>	39
3.1.7	Parâmetros Relevantes	41
3.1.8	Considerações.....	42
3.2	SISTEMAS INTELIGENTES	43
3.2.1	Conceitos Iniciais	43
3.2.2	Redes Neurais Artificiais.....	44
3.2.3	Principais Características das RNA.....	45
3.2.4	Áreas de Aplicações	46
3.2.5	Estrutura Básica da RNA – Neurônio Artificial.....	47
3.2.6	Funções de Ativação	48
3.2.7	Arquiteturas de Redes Neurais	50
3.2.8	Treinamento e Aprendizado	53
3.2.9	Considerações.....	54
4	MATERIAIS E MÉTODOS.....	55
4.1	MATERIAIS	55
4.1.1	Banco de Dados.....	55

4.1.2	Ferramentas Utilizadas	56
4.2	MÉTODOS	56
4.2.1	Sistema Classificador de Padrões	57
4.2.2	Preparação dos Dados	59
4.2.3	Segmentação e Janelamento	59
4.2.4	Aplicação da <i>wavelet packet</i> e extração de características.....	61
4.2.5	Treinamento das Redes Neurais	65
4.2.6	Pós-Processamento e Obtenção de Resultados	65
4.3	PROCEDIMENTOS	66
4.3.1	Transformação do banco de dados em sinais janelados	66
4.3.2	Transformação das janelas em amostras	67
4.3.3	Rotina de treinamento automatizado	68
5	RESULTADOS E DISCUSSÕES	73
5.1	RESULTADOS DA VARIAÇÃO DE TOPOLOGIAS NO NÍVEL 2	73
5.2	RESULTADOS DA VARIAÇÃO DE TOPOLOGIAS NO NÍVEL 3	75
5.3	ANÁLISE DE <i>UNDERFITTING</i> NAS TOPOLOGIAS DO NÍVEL 3	77
5.4	VERIFICAÇÃO DA ROBUSTEZ DA MELHOR TOPOLOGIA	79
5.5	PÓS-PROCESSAMENTO DOS RESULTADOS OBTIDOS.....	80
5.6	MATRIZ DE CONFUSÃO.....	83
5.7	ANÁLISE DE RESPOSTA DA RNA PARA AMOSTRAS DESCONHECIDAS ..	86
5.7.1	Simulação de operação – arquivo do tipo edema (I)	86
5.7.2	Simulação de operação – arquivo do tipo edema (II).....	87
5.7.3	Simulação de operação – arquivo do tipo edema (III)	88
5.7.4	Simulação de operação – arquivo do tipo saudável (I)	89
5.7.5	Simulação de operação – arquivo do tipo saudável (II)	90
5.7.6	Simulação de operação – arquivo do tipo saudável (III).....	91
5.7.7	Simulação de operação – discussão dos resultados.....	92
6	CONSIDERAÇÕES FINAIS.....	94
	REFERÊNCIAS	95

1 INTRODUÇÃO

A produção da voz, juntamente com a face e as mãos, são as primeiras vias de expressão com o mundo ao redor, sinalizando a identidade do indivíduo, suas vontades e sentimentos. Em relações sociais do dia-a-dia usa-se a voz para ganhar aceitação ou para comunicar ideias. (TITZE, 2013, p. 9). A voz é uma habilidade única, desenvolvida com o passar do tempo, envolvendo o crescimento físico e amadurecimento emocional.

O uso inadequado da voz e suas consequências tem sido tema de estudos de profissionais. Segundo Pinho (1997, p.1), principalmente os profissionais da voz (professores, locutores, cantores, entre outros) utilizam atividades verbais inadequadas em sua profissão, podendo danificar os tecidos da laringe e produzir um distúrbio vocal resultante do abuso ou mau uso da voz. Como Köhle (2004, p. 33) discute, a comunicação humana é o foco da fonoaudiologia e, em conjunto com ela, a filosofia de prevenção tem sido disseminada e aplicada na prática, conscientizando a todos sobre a importância da saúde vocal.

Lima-Silva et al (2012, p. 391) e Macedo, Souza e Thomé (2014, p. 74 e 75), apresentam dados sobre os danos financeiros e sociais acarretados pelas disfunções vocais. De acordo com os autores, o distúrbio de voz é motivo de situações de afastamento e incapacidade para o desempenho de suas atividades, assim como os transtornos com readaptação e prejuízos para os indivíduos e instituições. É estimado, para o Brasil, um gasto anual de 200 milhões de reais destinados a lidar com os afastamentos de professores em função de problemas na voz. (ABORL-CCF, 2004, p.5).

Quanto a complexidade, a voz humana é gerada de interações em diversos órgãos e, para que ocorra a produção vocal, é necessário a coordenação de inúmeros subsistemas envolvidos. (SCALASSARA; DAJER; MACIEL; 2005, p. 1). De tal modo, a complexidade no estudo dos sinais vocais deriva dos obstáculos em encontrar abordagens opostas à métodos invasivos para a solução dos problemas como a tentativa de relacionar patologias com padrões de perturbação no sinal produzido pela voz.

Diversas doenças ligadas ao sistema de produção de voz, como o câncer de laringe ou pólipos, são responsáveis por afetar muito além da saúde do paciente, reduzindo a qualidade de vida e relações sociais. Sabe-se que os melhores tratamentos para as disfunções ocorrem quando detectados nos estágios mais precoces. Contudo, além de invasiva e desconfortantes, as técnicas convencionais fundamentadas na utilização de equipamentos de captura de imagens das pregas

vocais possuem custo elevado e demandam muito tempo para realização e obtenção de resultados. (GHASEMZADEH et al., 2015, p. 135).

Métodos não-invasivos, apesar de não participarem prioritariamente no procedimento de diagnóstico diferencial de distúrbios laríngeos, permitem a observação de elementos objetivos, como dados quantitativos. Com o procedimento adequado, tais medidas podem ser coletadas automaticamente por processo computacional, digitalizadas e analisadas criticamente por um profissional da área, oferecendo recurso para os obstáculos enfrentados pela fonoaudiologia. (PINHO; TSUJI; BOHADANA, 2006, p. 33 e 50).

Contudo, como é abordado por Behlau (2001, p.26), Godino-Llorente e Gomez-Vilda (2004, p. 380) e Arjmandi et al (2011, p. 275), existem inúmeros parâmetros acústicos introduzidos hoje para medir e classificar o grau de normalidade da voz (*pitch, jitter, shimmer*, amplitude, entre muitos outros), contribuindo na identificação de características patológicas. Porém, sabe-se que o maior obstáculo é encontrar uma forma de ajustar esses valores de forma que, em um caso generalizado, seja possível representar toda uma população e seus grupos de controle, independente de sexo ou idade.

Auxiliando na classificação de padrões, Prarraga (2002, p. 2) propõe a utilização de uma ferramenta consagrada nos últimos anos na área de processamento de sinais, a transformada *wavelet packet* (TWP). Desse modo, é possível realizar análise tempo-frequência, permitindo um detalhamento do sinal na extração de características e, conseqüentemente, melhora no desempenho do reconhecimento de padrões em desordem, vencendo as limitações no domínio do tempo, presente em outras transformadas. Nos trabalhos de Hilton (1997, p. 394), Ramirez-Villegas e Ramirez-Moreno (2012, p. 82) e Zhang et al (2015, p. 1795), fica claro que a transformada possui inúmeras aplicações e se tornou promissora para trabalhos envolvendo sinais biológicos.

Por sua vez, em relação ao aprendizado e automatização da identificação, tem-se como sugestão, inspirado em Salhi, Talbi e Cherif (2008, p.330) e Carvalho, Cavalcante e Cortez (2011, p. 371), um classificador fundamentado em sistemas inteligentes, mais especificamente, em redes neurais artificiais. Tal sistema apresenta variadas aplicações em extração de conhecimento mediante a variáveis que não são facilmente correlacionadas por métodos convencionais, tornando-se de grande valor para a área de sinais da fala.

1.1 PROPOSTA DO TRABALHO

Propõe-se então, como solução, a modelagem de um extrator de características do sinal da voz utilizando transformada *wavelet packet* em conjunto com redes neurais artificiais para a classificação dos padrões de distúrbio na voz.

Tal classificador torna-se, então, capaz de identificar de maneira automatizada a presença de uma patologia na produção vocal do indivíduo. Em outras palavras, a principal motivação do estudo é propor outro método para a avaliação da fala que permita dispor facilidades aos profissionais da área conjuntamente com os métodos já consolidados e que traga conforto ao indivíduo analisado.

1.2 PUBLICAÇÕES OBTIDAS

Durante a fase de desenvolvimento, os estudos realizados para a realização do trabalho apresentaram resultados promissores, sendo publicados nos seguintes congressos:

- FERMINO, Murillo A. et al **Classificação de Distúrbios Vocais Aplicando Redes Neurais Artificiais e Transformada Wavelet Packet**. 12th IEEE/IAS International Conference on Industry Applications – INDUSCON. Curitiba – PR, Brazil (2016).
- FERMINO, Murillo A. et al **Reconhecimento de Distúrbios Vocais Utilizando Redes Neurais Artificiais**. XXV Congresso Brasileiro de Engenharia Biomédica – CBEB. Foz do Iguaçu – PR, Brazil (2016).

1.3 OBJETIVOS

Nesta seção serão destacados os objetivos referentes a proposta do trabalho.

1.3.1 OBJETIVO GERAL

Desenvolver uma rotina de processamento automatizado que se inicia pelo carregamento do arquivo de voz do indivíduo, realize a extração de características de energia e entropia aplicando a transformada *wavelet packet* e finalize na classificação nos padrões de distúrbio encontrados utilizando redes neurais artificiais, tornando-se uma alternativa para

análises acústicas e auxilie um profissional da fonoaudiologia na detecção de patologias em procedimento não-invasivo.

1.3.2 OBJETIVOS ESPECÍFICOS

Delimitando-se o objetivo geral em estágios mais aprofundados e detalhados obtém-se os seguintes objetivos específicos:

- Preparação de um banco de dados com diferentes sinais de voz (saudáveis e com presença do Edema de Reinke) para o desenvolvimento do classificador inteligente;
- Propor uma abordagem de detalhamento do sinal vocal em análise tempo-frequência utilizando a transformada *wavelet packet*;
- Extrair características do sinal, ajustando e normalizando em amostras para as redes neurais artificiais;
- Automatizar o processo de treinamento e validação do classificador visando encontrar a melhor topologia para o problema proposto;
- Analisar e discutir os resultados obtidos nas informações de pós-processamento e respostas do classificador;

1.4 ORGANIZAÇÃO DO TRABALHO

No capítulo 2 serão apresentados alguns estudos básicos sobre a voz, realizando-se uma breve abordagem sobre os aspectos que a caracterizam e os desafios que proporciona à consolidação dos objetivos propostos. No capítulo 3 serão levantadas as fundamentações teóricas pertinentes para a compreensão dos estudos propostos, apresentando brevemente os conceitos básicos da área de processamento digital de sinais e de sistemas inteligentes. No capítulo 4 será apresentada a origem do banco de dados utilizado, bem como, os métodos aplicados para obtenção dos resultados. No capítulo 5 se encontram as descrições detalhadas de cada procedimento desenvolvido, interligando os conceitos teóricos explanados com os métodos descritos no capítulo anterior. No capítulo 6 serão reunidos todos os resultados obtidos e sua segmentação em diversas etapas para facilitar o entendimento das discussões propostas. No capítulo 7, por fim, serão feitas as considerações finais a respeito deste trabalho, elencando

quais os ganhos desta pesquisa para a área de reconhecimento de padrões na voz utilizando a modelagem fundamentada em redes neurais artificiais e transformada *wavelet packet*.

2 ESTUDOS DA VOZ

Levando-se em conta a multidisciplinaridade do trabalho proposto, faz-se necessário dedicar um capítulo para introduzir conceitos iniciais sobre os estudos da voz antes de se aprofundar nos tópicos de processamento digital de sinais e sistemas inteligentes para resolução dos problemas propostos.

Neste capítulo será abordado o que é a voz humana, quais são os mecanismos envolvidos na produção e como é caracterizado o sinal da fala. Tem-se também, com objetivo de justificar a proposta de estudo, a abordagem dos métodos tradicionais para discriminação entre distúrbios e quais são as características básicas da patologia escolhida.

2.1 CONCEITOS INICIAIS

O estudo da produção vocal é tópico essencial para diversas áreas profissionais, incluindo música, artes, comunicação, fonoaudiologia e medicina. Os aspectos envolvidos no tratamento de disfunções vocais são detalhados e complexos. Por exemplo, um elemento comum, a reabilitação, é muito mais do que simplesmente retornar a voz do paciente ao normal, mas sim um processo de fortalecimento da voz para aplicação em determinada função (comunicar, cantar, lecionar), pois uma voz em condições normais pode não ser suficiente para satisfazer os esforços requeridos para uso profissional. (TITZE, 2013, p. 15).

As técnicas de análise acústica são amplamente utilizadas tanto na clínica como na pesquisa, e têm demonstrado serem ferramentas auxiliares úteis e não invasivas. Contudo, essas abordagens enfrentam dificuldades em função da necessidade de extração de parâmetros acústicos e os métodos necessários ainda não apresentam a robustez suficiente. (DAJER, 2010, p. 20). Segundo Umaphy et al (2005, p. 421), apesar do grande espaço já conquistado, as ferramentas de análise de voz por meio acústico ainda carecem de respaldo quanto a sua utilização e confiabilidade na discriminação de vozes normais e patológicas.

Contudo, apesar dos obstáculos enfrentados, Pinho, Tsuji e Bohadana (2006, p.33) discutem em sua obra as principais motivações em utilizar métodos alternativos às análises perceptivas ou subjetivas. Apesar de seus interesses estarem fundamentados nos princípios da simplicidade e baixo custo, apresentam também outros incentivos de grande importância como:

- Apoio em observações subjetivas, por exemplo em métodos que utilizam imagens videolaringoestroboscópicas;
- Detecção de perturbação fonatória nos indivíduos presentes em uma população não-disfônica;
- Auxílio no monitoramento terapêutico;
- Facilidade de compreensão ao paciente, visto que a representação em imagens e gráficos ajuda nos esclarecimentos;
- Servir de embasamento na defesa relacionado a eficácia dos tratamentos convencionais realizados.

2.2 PRODUÇÃO DA VOZ

Sinais de fala são compostos por sequências de sons que, para meios de estudos, são tratados como produções físicas de símbolos discretos. Portanto, o som e suas variações servem como representações simbólicas das informações governadas pelas regras da linguagem. (RABINER; SCHAFER, 2012, p. 68).

A voz é um sistema complexo e que, para facilitar o entendimento da sua operação, pode ser segmentado em subsistemas mais simples. Contudo, para uma produção vocal adequada e satisfatória, é necessário que todos os subsistemas trabalhem em conjunto de maneira sincronizada. (BEHLAU, 2001, p. 26). No Quadro 1 está apresentado de maneira resumida os sistemas, as funções e quais as estruturas relacionadas dentro do contexto de produção da voz:

Quadro 1 – Sistemas e estruturas participantes da produção vocal.

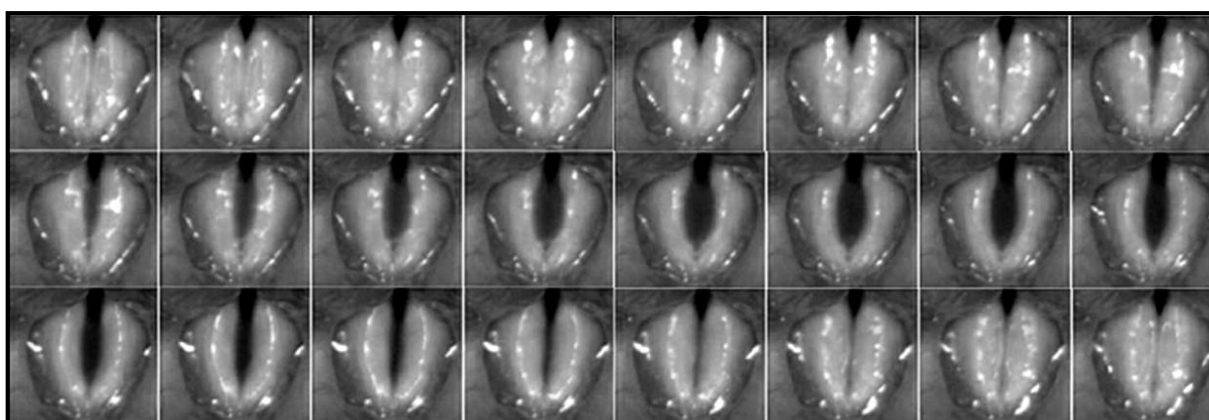
Sistema	Função	Estruturas Relacionadas
Respiratório	Fonte de Energia	Pulmões
Fonatório	Fonte Vibratória	Pregas vocais
Ressonância	Fonte de Ressonância	Cavidade oral e cavidade nasal
Articulatório	Articulação	Língua, lábios, mandíbula, palato, dentes
Nervoso	Coordenação	Córtex

Fonte: Adaptado de Behlau (2001, p. 26).

A principal fonte de produção da voz é baseada na variação do fluxo aéreo transglótico no momento da vibração das pregas vocais. Entretanto, o som pode ser produzido por diversas ou quase todas partes do trato vocal, tanto pela turbulência do ar numa região constrita quanto pela repentina liberação de pressão numa bolsa de ar. Deve-se considerar da mesma forma, que tecidos moles do trato vocal (lábios, palato, partes da língua) podem oscilar como as pregas e produzir som. (TITZE, 2013, p. 151).

Com o auxílio da Figura 1 é possível notar as oscilações das pregas vocais durante um ciclo glótico normal.

Figura 1 – Pregas vocais durante um ciclo glótico normal.



Fonte: Adaptado de Tsuji (2014, p. 114).

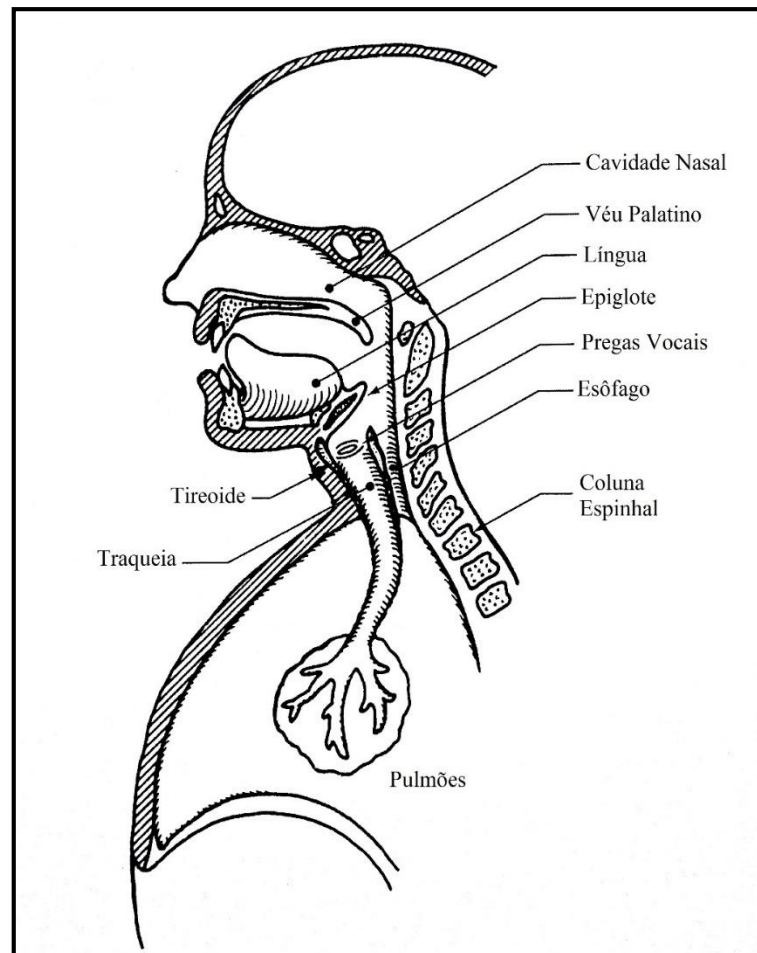
Os autores Rabiner e Schafer (2011, p. 71) descrevem o mecanismo de produção de sons vocais da seguinte forma:

- Primeiramente, o ar adentra os pulmões normalmente pela respiração (sem produção de sons na inspiração);
- Em seguida, o ar é expelido pelos pulmões passando pela traqueia, conduzindo as pregas vocais tensionadas a vibrarem de acordo com a Lei de Bernoulli, relacionando as variações de pressão com a abertura da glote;
- O ar, por sua vez, é interrompido pelo movimento de abertura a fechamento do orifício glótico, gerando pulsos periódicos;
- As frequências dos pulsos são moldadas quando atravessam as cavidades do trato vocal (faringe, boca e cavidades nasais);

- Finalmente, a posição dos articuladores (mandíbula, língua, véu palatino, lábios) determinam a modulação característica do som produzido.

Com auxílio da Figura 2 é possível visualizar o esquemático do trato vocal e seus elementos.

Figura 2 – Visualização esquemática do trato vocal.



Fonte: Adaptado de Rabiner e Schafer (2011, p. 71).

Dependendo de qual som será produzido pela excitação do aparelho vocal, três mecanismos básicos podem estar envolvidos, como trata Alcaim e Oliveira (2011, p. 25):

- **Sons sonoros (som de /u/ em uva):** o fluxo de ar saindo dos pulmões conduz a vibração das cordas vocais gerando pulsos sequenciais quase periódicos de pressão de ar, excitando o aparelho vocal.

- **Sons fricativos surdos (som de /s/ em sala):** composto por ruído contínuo de espectro amplo e uniforme, é criado na turbulência em constrictões estreitas situadas no trato vocal.
- **Sons oclusivos (som de /p/ em pato):** caracterizado por um súbito desprendimento de pressão após uma obstrução completa em algum ponto no aparelho vocal.

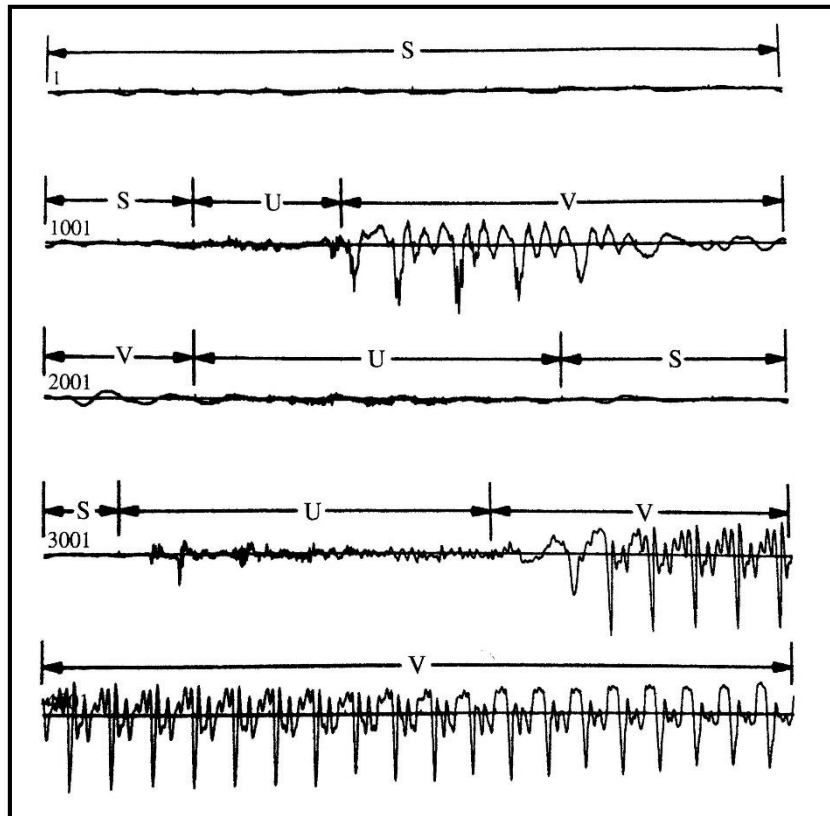
2.3 PROPRIEDADES E FORMATO DA VOZ

Rabiner e Schafer (2011, p. 75) descrevem algumas propriedades no sinal da voz, tais como:

- Voz é composta por uma série de sons em constante mudança;
- As propriedades do formato de onda estão fortemente relacionadas aos sons produzidos para transmitir a mensagem desejada;
- São também altamente dependentes do contexto onde os sons estão inseridos, ou seja, são alterados tanto pelos sons anteriores, quanto pelos seguintes. É chamada de coarticulação, sendo resultado dos mecanismos responsáveis antecipando as próximas articulações durante a produção do som atual;
- Sofrem alterações com as variações das pregas vocais, do mesmo modo com a posição, forma e tamanho dos articuladores (lábios, dentes, língua, mandíbula) em constante e lenta mudança com o passar do tempo.

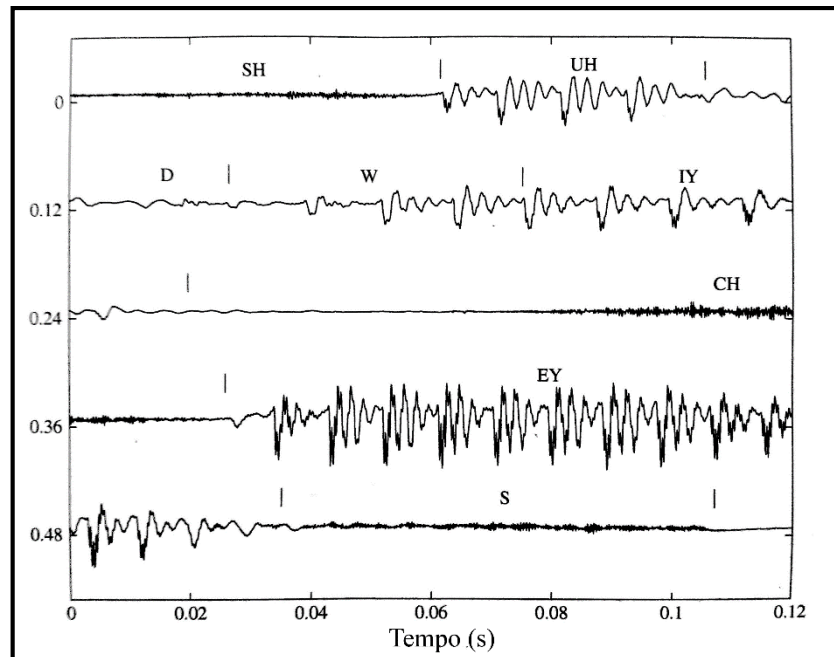
Dessa forma, percebe-se que a fala tem como origem um sinal complexo e altamente variante. Para exemplificar, pode-se observar na Figura 3 como são compostos os intervalos com vocalização (sons sonoros), sem vocalização (sons fricativos surdos) e também sinais de fundo. A letra “V” representa intervalos com vocalização, letra “U” os intervalos sem vocalização e “S” são sinais de fundo. Similarmente, na Figura 4 é possível visualizar as variações citadas durante a elocução da frase em inglês “*Should we chase*”.

Figura 3 – Representação da forma de onda vocal e seus intervalos.



Fonte: Adaptado de Rabiner e Schafer (2011, p. 76).

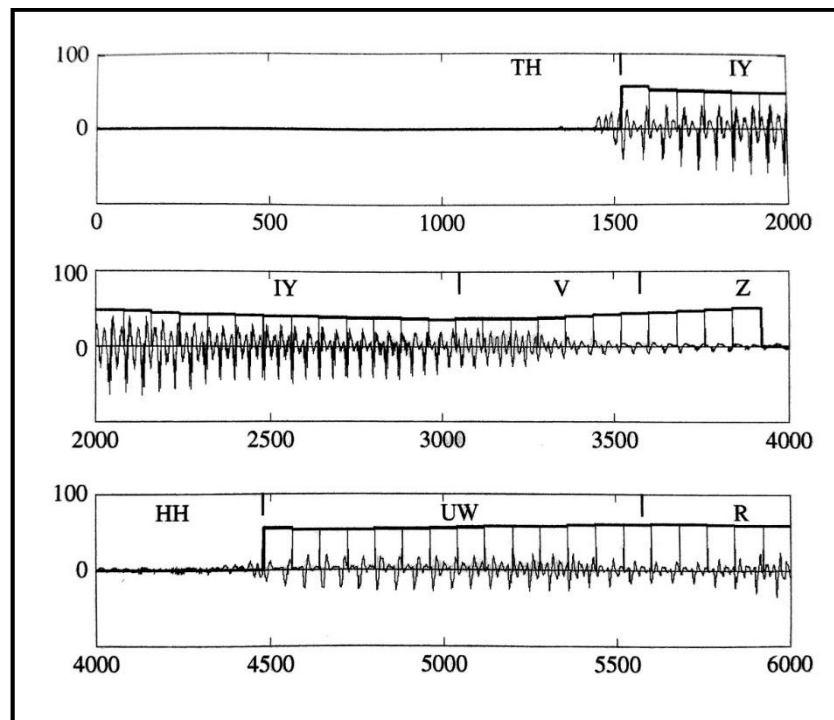
Figura 4 – Sinal vocal durante elocução da frase “*should we chase*”.



Fonte: Adaptado de Rabiner e Schafer (2011, p. 77).

Outra propriedade facilmente observada mediante a onda da fala é o tempo de periodicidade durante intervalos de vocalização. É definido pelo período local de repetição da forma de onda e seus valores medidos. Também chamado de período de *pitch* (RABINER; SCHAFER, 2012, p. 78). A Figura 5 é a estimativa do período de *pitch* observado nos primeiros fonemas da frase em inglês “*thieves who rob friends deserve jail*”:

Figura 5 – Estimativa do período de pitch para a forma de onda.



Fonte: Adaptado de Rabiner e Schafer (2011, p. 79).

2.4 MÉTODOS DE AVALIAÇÃO VOCAL

Esta seção tem o objetivo de apresentar os principais aspectos e diferenças entre os métodos invasivos e não-invasivos.

2.4.1 Métodos Invasivos

No quesito de avaliação vocal, os métodos invasivos baseiam-se na obtenção de imagens internas da laringe do paciente, permitindo avaliar as condições das pregas vocais. Ao observar a utilização da videolaringoscopia, por exemplo, percebe-se a maior vantagem dos métodos invasivos: a capacidade de visualizar por meio de uma câmera o real

comportamento das pregas vocais e das possíveis patologias que podem estar afetando a produção vocal do indivíduo.

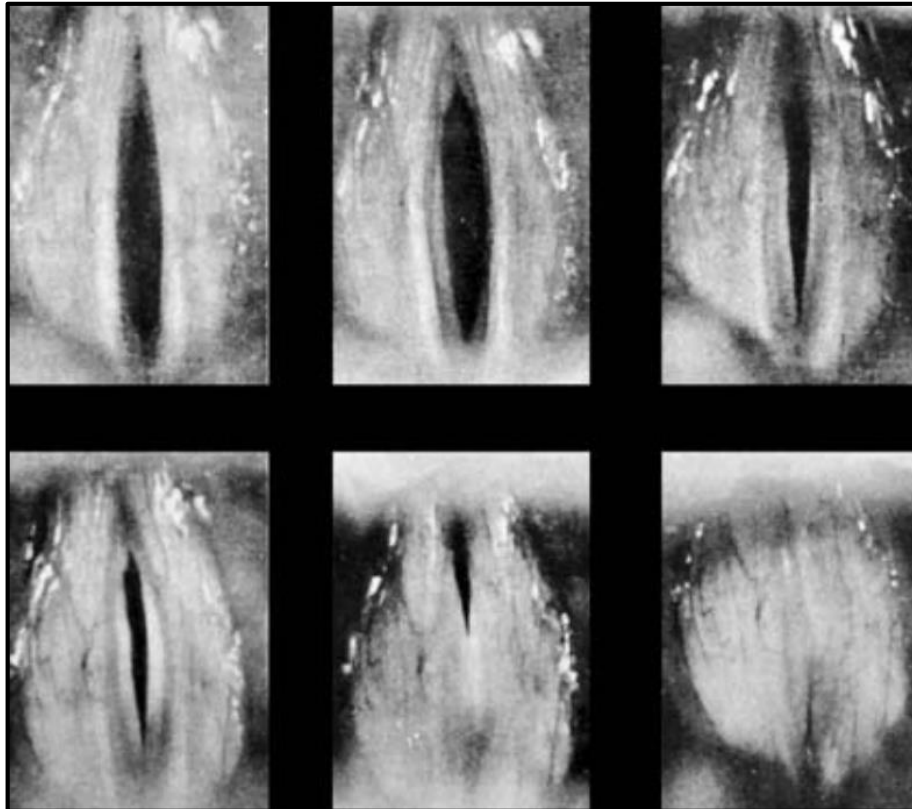
Como é explanado por González (2008, p. 424), o movimento vibratório das pregas vocais é muito rápido para a percepção do olho humano, tornando necessário a inserção de uma micro câmera com capacidade de captura em alta velocidade conjuntamente com um aparato de luz estroboscópica. Ajusta-se os pulsos da luz para uma frequência semelhante à frequência fundamental da voz analisada, possibilitando observar as características relevantes na avaliação vocal. Contudo, apesar de suas vantagens, é um exame de custo elevado onde há a necessidade de inserção do equipamento na orofaringe do paciente, causando desconforto e até mesmo dor. Pode-se observar, na Figura 6 tem-se os equipamentos necessários para a realização da videolaringostroboscopia e na Figura 7 as imagens resultantes do procedimento realizado.

Figura 6 – Equipamento necessário para realização da videolaringostroboscopia.



Fonte: Adaptado de González (2008, p. 425).

Figura 7 – Quadros capturados identificando fases distintas do ciclo vocal.



Fonte: Adaptado de González (2008, p. 425).

2.4.2 Métodos Não-Invasivos

Os métodos não-invasivos podem ser classificados em duas categorias básicas: os métodos de avaliação áudio-perceptual e avaliação acústica. Como é abordado por Freitas (2010, p. 10 e 44), é possível defini-los da seguinte maneira:

- **Avaliação áudio-perceptual:** é pressuposto que o profissional julgue auditivamente uma amostra vocal do indivíduo (com ou sem queixas na utilização da voz). Em geral, as amostras são compostas por vogais sustentadas e falas encadeadas. Ficam registradas em arquivos de áudio ou vídeo (preferencialmente) para que possa ser realizada a análise completa e sirva de comparativo em avaliações futuras.
- **Avaliação acústica:** permite quantificar, mensurar e descrever as características da voz. Propicia a obtenção de dados extraídos automaticamente por um processamento computacional adequado. Possibilita também uma integração entre as análises perceptivo-auditiva com o plano fisiológico, compondo uma visão mais detalhada sobre todo o processo de geração do sinal sonoro, suas modificações e disfonias. Para

tal, recorre a parâmetros acústicos que formam o sinal (periodicidade, amplitude, duração, composição espectral) caracterizando atributos físicos da voz no domínio do tempo, da frequência e também sua intensidade.

2.5 DISTÚRBIOS VOCAIS

A produção vocal normal requer refinado controle neuromuscular e integridade das estruturas laríngeas, e a maior parte dos casos de disfonias organofuncionais (decorrentes de mau uso ou abuso vocal) pode ser solucionada apenas com a fonoterapia. Muitas vezes, o quadro orgânico permanece após a reabilitação, mas ocorrem compensações funcionais, tais como: aumento do fluxo aéreo respiratório durante a fonação, maior eficiência da adução glótica e melhor projeção vocal, satisfazendo as necessidades de comunicação do paciente. (PINHO; TSUJI; BOHADANA, 2006, p. 69).

Por sua vez, Behlau (2001, p. 65) comenta sobre a dificuldade em definir as distinções entre uma voz normal e disfônica, sendo um dos temas mais subjetivos de sua área, podendo apresentar diferentes percepções sobre o grau de qualidade vocal do paciente. A presença de um distúrbio na produção vocal leva à redução da qualidade da voz, adquirindo aspectos como:

- **Rouquidão:** apresenta irregularidade de vibração das pregas vocais com frequência e intensidade reduzidas;
- **Aspereza:** é notável um aspecto desagradável de emissão e fechamentos vocais bruscos, tornando visível o esforço para falar;
- **Soprosidade:** presença de ar não sonorizado como ruído à fonação (chiado na voz);
- **Rugosidade:** voz com aspecto trêmulo como se houvesse acúmulo de mucosa na passagem do ar.

Sendo assim, Behlau (2001, p.65) descreve como objetivo do tratamento das disfonias vocais o retorno da produção vocal do paciente aos padrões de normalidade vocal. De forma simplificada, pode-se considerar os seguintes padrões vocais como referência para uma voz de boa qualidade:

- A produção vocal deve conter certa qualidade musical e ausência de ruído, se tornando agradável ao ouvir a fala do paciente;
- A frequência deve ser apropriada ao sexo e idade do falante;
- A intensidade deve ser adequada em condições de fala ordinárias (sem exercer esforço, a voz não deve soar nem tão fraca, nem tão intensa);
- A flexibilidade deve ser adequada de modo que o falante possa variar frequência e intensidade, permitindo utilizar ênfases nos significados e sutilezas da expressão oral.

2.5.1 Edema de Reinke

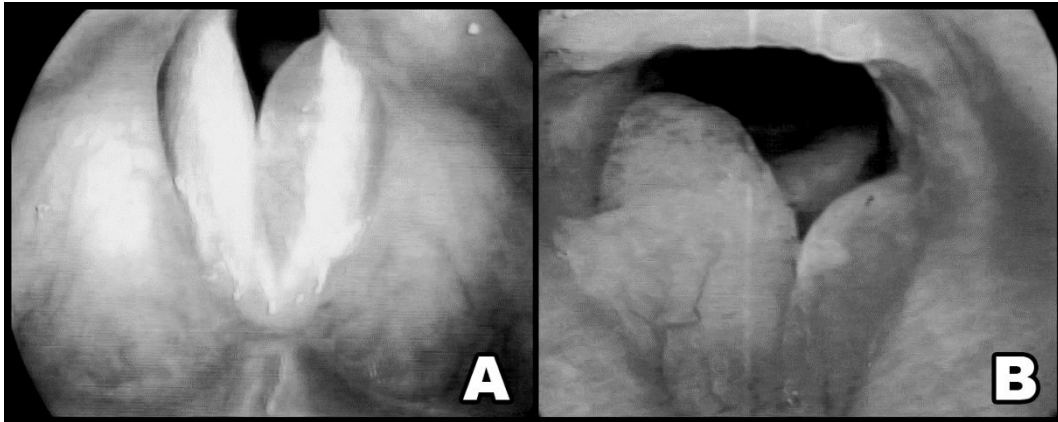
Segund Behlau (2001, p. 296), o Edema de Reinke é uma patologia classificada como disnofia organofuncional. É caracterizado pelo acúmulo de líquido no espaço de Reinke, dificultando o fluxo de ar e vibração das pregas vocais. Atribui aspecto de rouquidão e tonalidade mais grave à voz. Como é discutido em Lim (2014), o Edema de Reinke é um distúrbio vocal de grande ocorrência, tendo como causas principais o mau uso da voz (como o desgaste excessivo da voz em professores) e hábitos não saudáveis (tabagismo, por exemplo), incrementando a relevância do estudo proposto.

Por definição, o Edema de Reinke é uma afecção benigna, também conhecida como degeneração polipoide, abrangendo toda a porção membranosa das pregas vocais e envolvendo a camada superficial da lâmina própria (espaço de Reinke). É geralmente bilateral (ocorre em ambas pregas vocais) e assimétrico (proporções diferentes em cada prega vocal). Devido ao extenso edema presente na porção membranosa das pregas vocais, pode-se ter a presença de pequena fenda sempre aberta durante a fonação. (PINHO; TSUJI; BOHADANA, 2006, p. 72).

É possível visualizar na Figura 8 as imagens do Edema de Reinke (tanto em grau moderado quanto em grau severo):

Figura 8 – Edema de Reinke.

A) Bilateral, assimétrico, mais desenvolvido e translúcido à esquerda, paciente de 52 anos. B) Bilateral, em grau severo, em um paciente de 58 anos.



Fonte: Adaptado de Behlau (2001, p. 310).

2.6 CONSIDERAÇÕES

Considerando-se as informações discutidas ao lon, é possível compreender a importância exercida pela voz na vida das pessoas e como se torna essencial obter meios de recuperar uma fonação em desordem para seu estado normal. Entretanto, sabe-se que a produção vocal é resultado de um intrincado sistema com características de alta complexidade e suas peculiaridades influenciam nos padrões da voz.

Abordaram-se nesta seção alternativas aos métodos-invasivos, contudo deve-se levar em consideração seus pontos fracos. A análise áudio-perceptual pode sofrer variações entre avaliadores, imprecisões nas escalas aplicadas e até inconsistências nos arquivos apresentados aos especialistas. A análise acústica experimenta dificuldades na extração e correlação das características específicas de uma patologia e também na generalização de valores para as diversidades humanas, como sexo e idade.

3 ESTUDOS DE SINAIS E SISTEMAS INTELIGENTES

Feito a compreensão dos conceitos básicos da voz, tem-se o capítulo responsável pela introdução dos outros temas relacionados ao estudo proposto, revisando a teoria já consolidada na literatura. Primeiramente, na seção 4.1 será abordada as fundamentações da área de processamento digital de sinais, permitindo o entendimento dos conceitos base para a aplicação das ferramentas necessárias na realização do trabalho. Por fim, na seção 4.2 será apresentada a teoria de sistemas inteligentes e, de maneira detalhada, o funcionamento das redes neurais artificiais e suas estruturas essenciais.

3.1 PROCESSAMENTO DIGITAL DE SINAIS

O processamento digital de sinais (PDS) é um importante campo de estudo que surgiu dos avanços na teoria de comunicação e na tecnologia digital da computação aplicados atualmente.

Inicia-se esta seção discutindo brevemente os assuntos básicos como filtros e conceitos de amostragem para facilitar a compreensão dos tópicos mais avançados de transformação dos sinais, em específico a transformada *wavelet packet* que será aplicada na realização deste trabalho.

3.1.1 Conceitos Iniciais

A habilidade de comunicação é a chave para a vida em sociedade. Comunicação envolve o intercâmbio de informação, seja por curtas distâncias (diálogo entre duas pessoas, cara a cara) ou por longas distâncias (linha telefônica ou comunicação por satélite). A entidade responsável por transportar tais informações até seu destino é o sinal, assumindo diversas formas e carregando todo o tipo de dado. Por exemplo, uma onda acústica gerada pelo trato vocal humano carrega a fala, bem como, uma onda eletromagnética pode carregar som e imagem. (HAYES, 2009, p. 1).

3.1.2 Filtros

Uma das mais utilizadas operações em processamento de sinais complexos é a filtragem, sendo definida por componentes que permitem ou rejeitam a passagem de certas frequências. Considerando o escopo de estudo, faz-se interessante uma breve explicação do funcionamento da filtragem de sinais para facilitar o entendimento da separação de frequências de detalhe e aproximação que serão realizadas na aplicação da transformada *wavelet packet* no sinal da voz.

A implementação de tal componente no sistema é chamada de filtro e o intervalo de frequências permitidas pelo filtro é conhecido como banda passante. (MITRA; KUO, 2006, p. 4). Na maioria dos casos o princípio de operação dos filtros para sinais analógicos, por exemplo, pode ser definido pela integral da convolução, como pode ser observado na equação (1):

$$y(t) = \int_{-\infty}^{\infty} h(t - \tau).x(\tau) d\tau \quad (1)$$

Onde tem-se:

- $x(t)$ como sinal de entrada,
- $y(t)$ como sinal de saída caracterizado pela resposta de impulso $h(t)$.

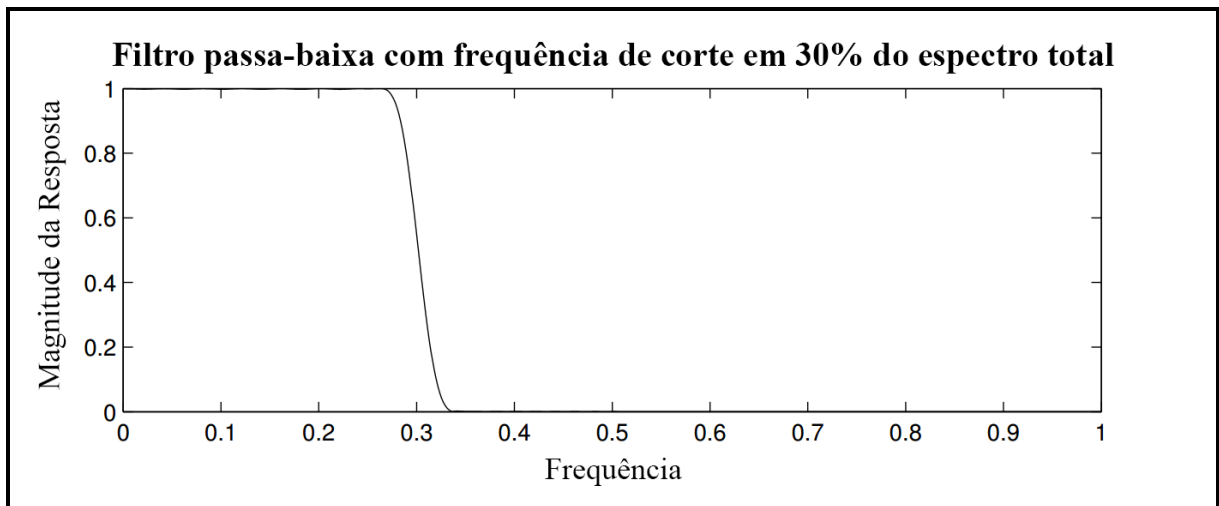
Por sua vez, Mitra e Kuo (2006, p. 4) discutem as distinções entre os variados tipos de filtros que podem ser aplicados na área de estudo de PDS. De tal modo, para uma dada frequência de corte, os filtros podem ser divididos nas seguintes categorias:

- **Filtro passa-baixa:** permite passagem de todas frequências inferiores à frequência de corte estipulada, bloqueando as demais frequências superiores;
- **Filtro passa-alta:** possui operação inversa ao filtro passa-baixa, bloqueando o espectro frequências inferiores à frequência de corte e possibilitando a passagem dos sinais de alta frequência;

- **Filtro passa-banda:** propicia a passagem de frequências entre duas frequências de corte, definindo uma faixa (ou banda) de frequências permitidas;
- **Filtro rejeita-banda:** de maneira similar ao filtro anterior, rejeita os valores de frequência contidos dentro da faixa definida por duas frequências de corte, possibilitando a passagem das restantes frequências contidas no espectro.

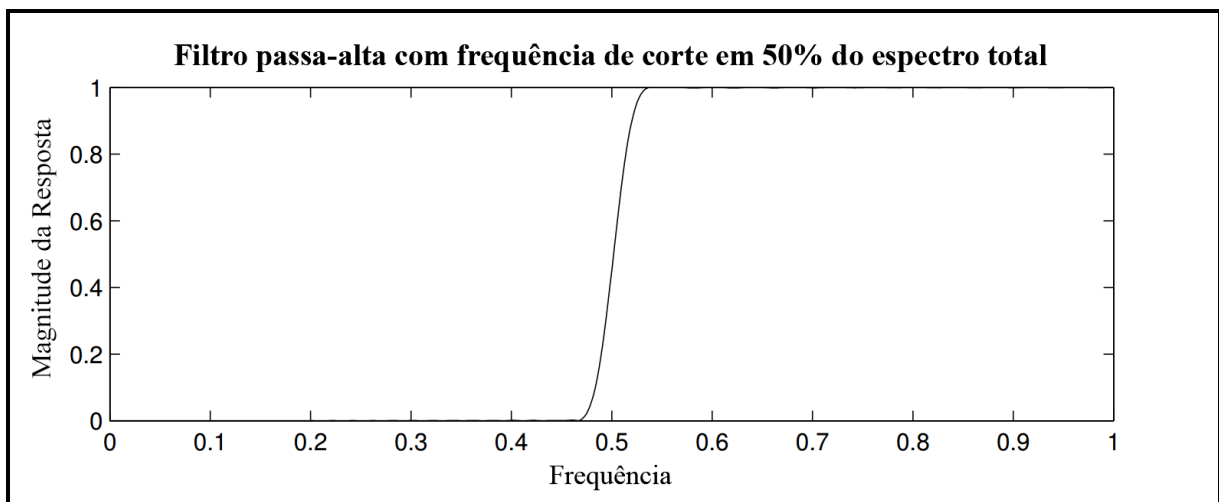
Por meio das Figuras 9, 10, 11 e 12 é possível facilitar o entendimento das nuances de cada um dos filtros citados anteriormente. Os eixos de frequência (eixo X) e magnitude da resposta do filtro (eixo Y) estão normalizados de 0 a 1 e as porcentagens representam as parcelas do espectro total de frequência que sofrerão corte no filtro:

Figura 9 – Exemplo de Filtro Passa-Baixa.



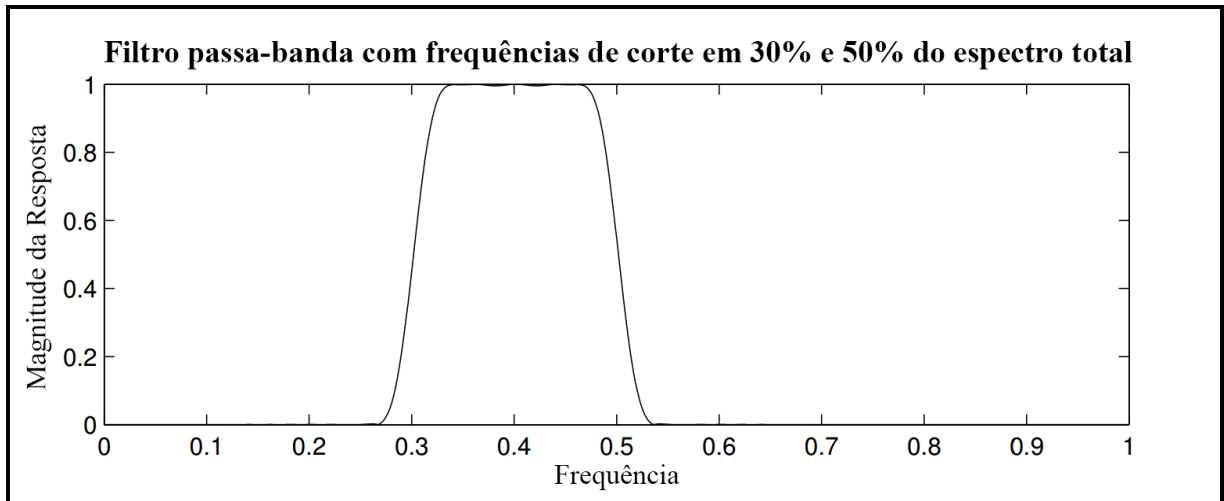
Fonte: Adaptado de Weeks (2010, p. 105).

Figura 10 – Exemplo de Filtro Passa-Alta.



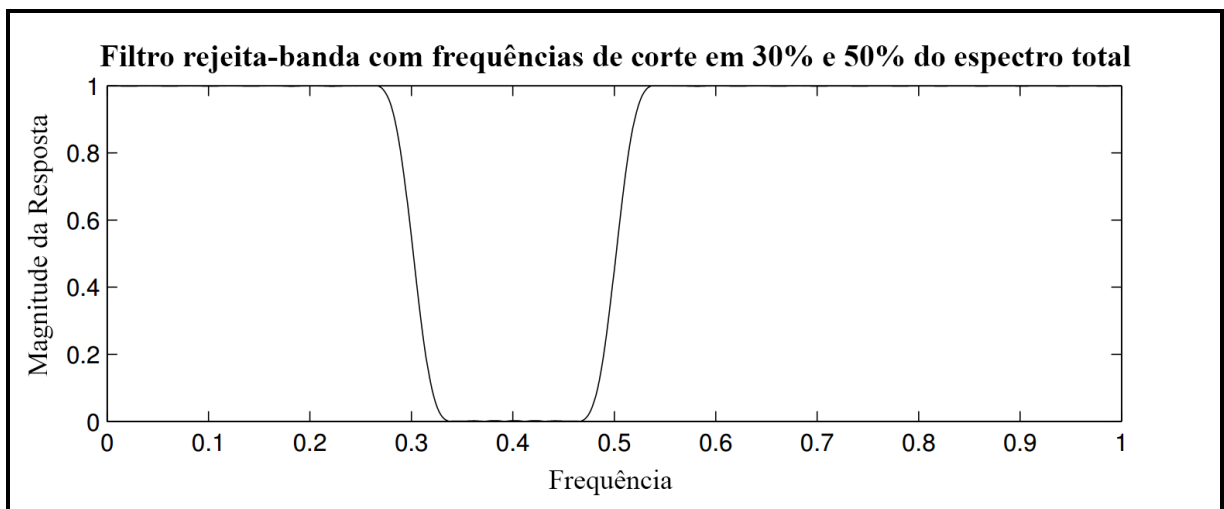
Fonte: Adaptado de Weeks (2010, p. 105).

Figura 11 – Exemplo de Filtro Passa-Banda.



Fonte: Adaptado de Weeks (2010, p. 108).

Figura 12 – Exemplo de Filtro Rejeita-Banda.



Fonte: Adaptado de Weeks (2010, p. 108).

3.1.3 Conceito de Amostragem

Tem-se conhecimento que o mundo real é analógico e os sinais relevantes ao estudo das ciências são tipicamente contínuos. Contudo, computadores são dispositivos digitais capazes de armazenar apenas informações discretas, finitas e limitadas em precisão e magnitude. Logo, é necessária uma aproximação do sinal antes de armazená-lo e manipulá-lo.

O conceito de amostragem (*sampling*) é dado pelo processo de extrair um sinal digital a partir de algo analógico. Quando um sinal é amostrado, tem-se a gravação em intervalo de tempo especificado do valor correspondente ao sinal. (WEEKS, 2010, p. 159).

A explicação do conceito de amostragem tem o propósito de exemplificar os cuidados necessários com a preparação dos arquivos de áudio utilizados no banco de dados. É importante que seja garantido uma uniformidade na amostragem dos sinais, mesmo na presença de diferentes frequências de amostragem (22kHz ou 44kHz, por exemplo).

Como é definido por Manolakis e Ingle (2011, p. 293), o modo mais comum de amostragem é a uniforme (ou periódica), delimitado pela obtenção de uma sequência de amostras a partir de um sinal contínuo no tempo, extraíndo seus valores a partir de pontos uniformemente separados no tempo. Por meio da equação (2) pode-se detalhar os conceitos:

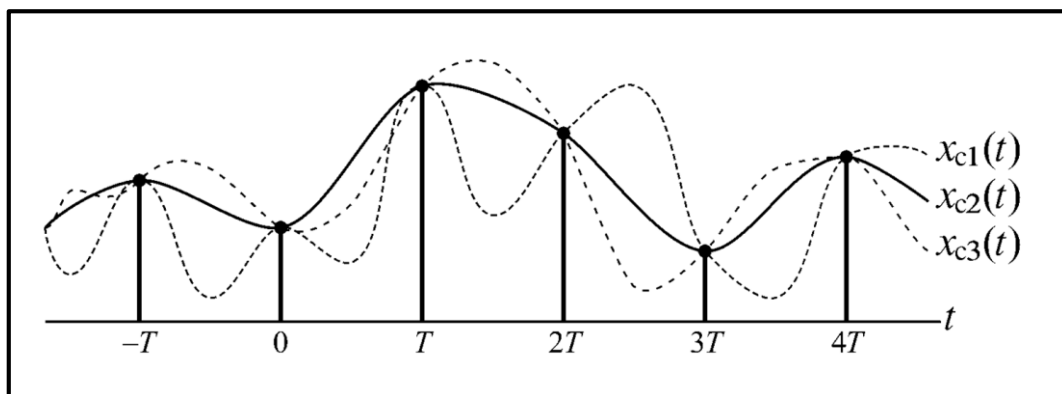
$$F_s = \frac{1}{T} \quad (2)$$

Onde tem-se que,

- F_s é a frequência de amostragem (quando apresentadas em ciclos por segundos ou Hertz (Hz)) ou, também, taxa de amostragem (quando expressa em termos de amostras por segundo);
- T é o período de amostragem, estabelecendo o valor de tempo fixo entre cada ponto a ser extraído.

Tais parâmetros, presentes na equação (2), possuem grande relevância no procedimento de conversão do sinal analógico para digital, pois, dependendo da frequência de amostragem, ambiguidades podem ocorrer para sinais distintos. Pode-se perceber na Figura 13 os obstáculos em representar as características do sinal entre os pontos de amostragem. Para três sinais de características distintas (x_{c1} , x_{c2} e x_{c3}) tem-se os mesmos valores nas amostras obtidas, indicando sinais idênticos na amostragem.

Figura 13 – Exemplo de Ambiguidade na Amostragem.



Fonte: Adaptado de Manolakis e Ingle (2011, p. 293).

Antes de discutir o problema apresentado previamente, é necessário definir conceitos pertinentes. Weeks (2010, p. 160) aborda o intervalo de frequências que compõe um sinal, denominando-o largura de banda (*bandwidth*). Pode-se delimitá-lo também por intermédio da frequência mais alta presente, indicando o interesse em analisar todas os valores entre 0 Hz e o máximo.

Como pode ser lido em Weeks (2010, p. 160), a equação (3), também conhecida como critério de Nyquist, relaciona uma frequência de amostragem ideal para que se possa reconstruir o sinal apropriadamente no futuro:

$$F_s \geq 2B \quad (3)$$

Sendo composta por,

- F_s , frequência de amostragem (*sampling frequency*);
- B , largura da banda (*bandwidth*).

Considerando o critério de Nyquist, pode-se encontrar duas situações indesejadas:

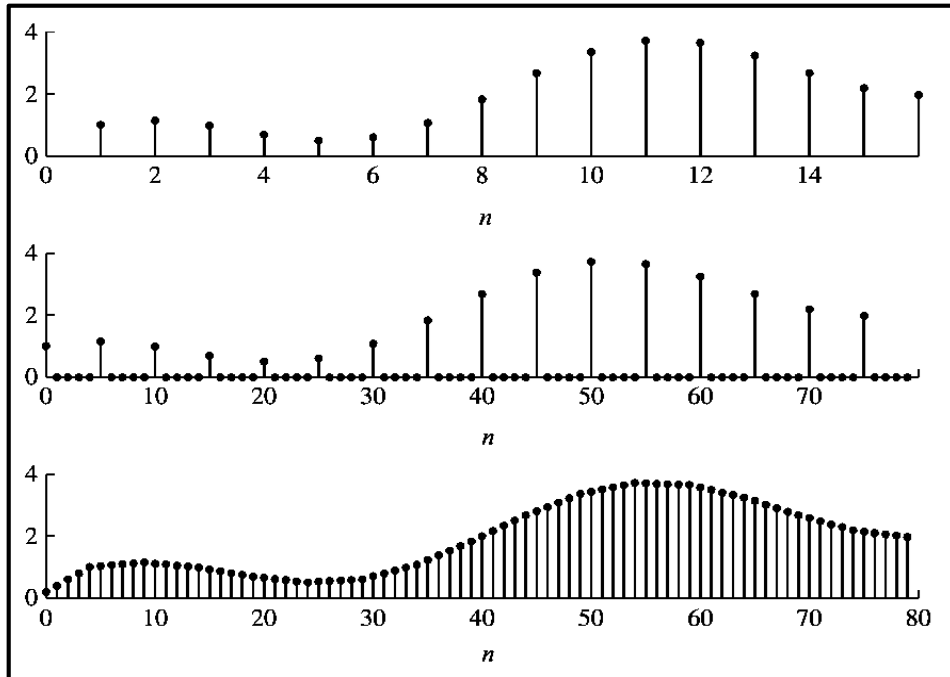
- ***Oversampling***: definido por uma frequência de amostragem superior ao dobro da largura de banda, tendo um número de amostras superior ao necessário, criando um número de dados excessivo;
- ***Undersampling***: definido por uma frequência de amostragem inferior ao dobro da largura de banda, tendo um número de amostras insuficientes para a reconstrução apropriada do sinal;

Por fim, o último conceito relevante dentro do tópico de amostragem é a conversão da taxa de amostragem. Tal conceito permite a correção do banco de dados em uma única taxa de amostragem. Manolakis e Ingle (2011, p. 715 e 706-707) abordam os métodos utilizados para alterar a taxa de amostragem de um sinal. São classificados da seguinte maneira:

- ***Upsampling***: procedimento onde, para um fator L de *upsampling*, altera-se o número de amostras inserindo $L-1$ pontos entre consecutivas amostras, conceituando-o como um método de preservação do sinal. Pode ser observado na Figura 14 um exemplo de *upsampling* aplicado a um sinal. Primeiramente tem-se a representação do sinal

original, em seguida a aplicação do fator “L” igual a 5 (que define a proporção de amostras incrementadas) e, por fim, o sinal resultante da aplicação.

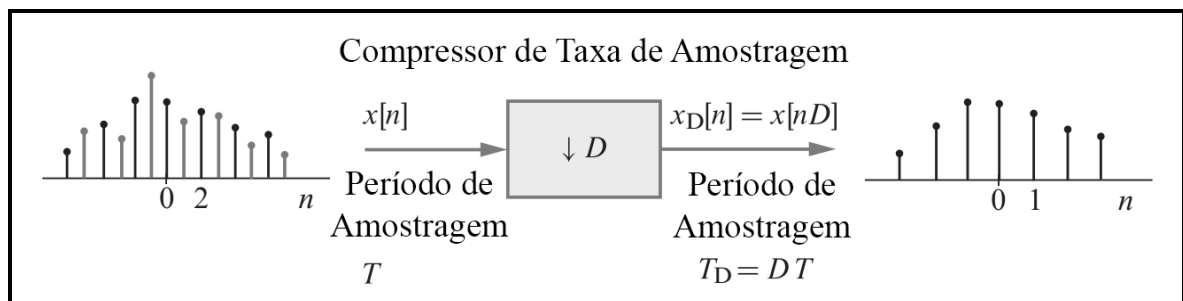
Figura 14 – Exemplo de do processo de *upsampling*.



Fonte: Adaptado de Manolakis e Ingle (2011, p. 724).

- **Downsampling:** situação onde ocorre a redução do número de amostras por um fator D. Seu funcionamento pode ser simplificado em reter apenas uma amostra em um segmento de D pontos. Em outras palavras, o fator “D” define a razão de amostras entre entrada e saída. Tem operação representada pela Figura 15:

Figura 15 – Exemplo de *downsampling*.



Fonte: Adaptado de Manolakis e Ingle (2011, p. 707).

3.1.4 Transformações dos Sinais

O objetivo da análise de sinais é extrair informações pertinentes utilizando alguma forma de transformação do sinal. Logo, tais métodos são essenciais para aplicações em áreas de estimação de parâmetros, codificação e reconhecimento de padrões. (RIOUL; VETTERLI, 1991, p. 15).

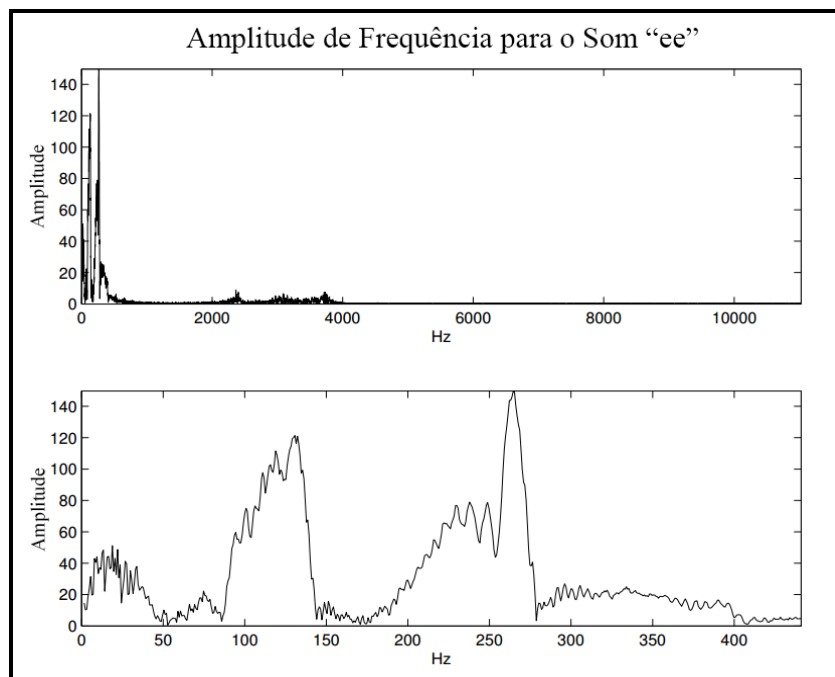
Rabiner e Schafer (2011, p. 22) comentam sobre a facilidade que representações de sinais e sistemas no domínio da frequência trazem para o contexto de análises em PDS, citando como exemplo a transformada de Fourier e a transformada Z.

3.1.5 Transformada de Fourier

É utilizada em diversos ramos da ciência (medicina, ótica, física, engenharia elétrica) como uma poderosa ferramenta de análise. Essencialmente, permite observar as informações da frequência no lugar do tempo, possibilitando concluir resultados mais intuitivos em certas aplicações, principalmente com sinais estacionários. (WEEKS, 2010, p. 187).

É possível, então, observar na Figura 16 a aplicação da transformada de Fourier na vocalização do som “ee” (proveniente do inglês).

Figura 16 – Transformada de Fourier no sinal de voz.



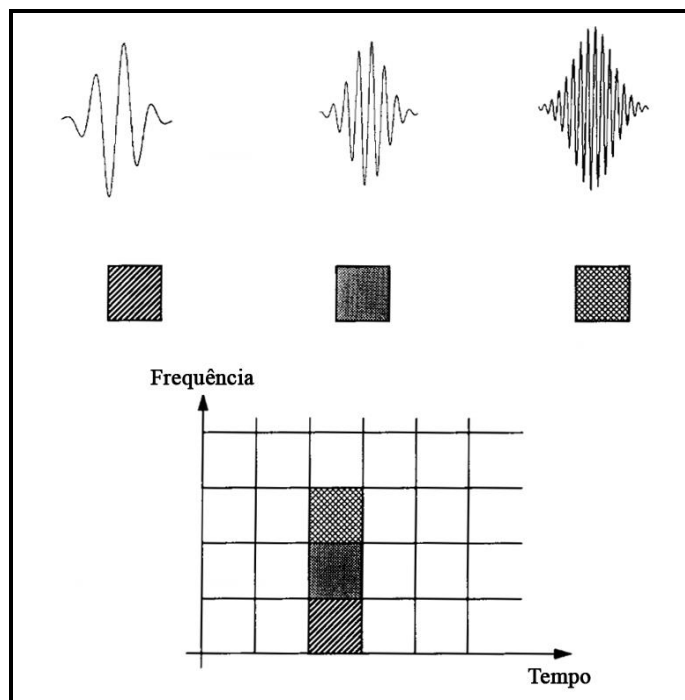
Fonte: Adaptado de Weeks (2010, p. 188).

Contudo, Rioul e Vetterli (1991, p. 15 e 16) discutem sobre as limitações da transformada de Fourier para casos não-estacionários. Em sinais que apresentam variações de frequência ao longo do intervalo analisado, pode-se perder informações relevantes para a caracterização da patologia estudada, pois aplicando a transformada de Fourier as frequências são linearizadas em um espectro médio ao longo da duração da amostra.

Faz-se necessário, então, adaptações para uma abordagem bidimensional, considerando a frequência e o tempo. O primeiro método sugerido é a transformada de Fourier de tempo curto ou, do inglês, *short time Fourier transform* (STFT). Baseia-se na segmentação do sinal em janelas contendo informações do tempo e frequência.

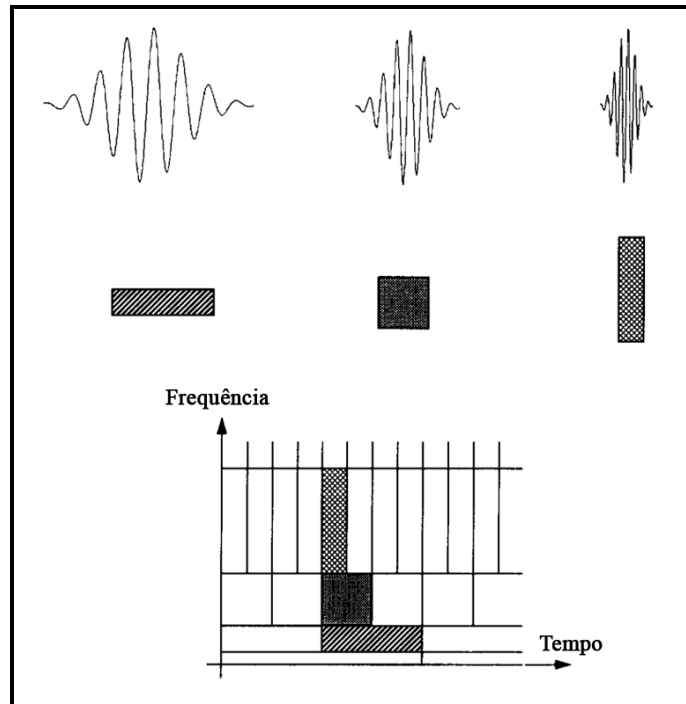
Por sua vez, a STFT ainda apresenta particularidades para a análise de sinais não-estacionários. Vetterli e Herley (1992, p. 2209) discorrem sobre as dificuldades encontradas em relação à resolução na aplicação da STFT e a comparam com outra metodologia, a transformada *wavelet*. Em comparação com a resolução fixa da STFT, a transformada *wavelet* permite um ajuste na precisão de tempo e frequência, tendo maiores resoluções de tempo nos segmentos de altas frequências e maiores resoluções de frequência nos segmentos de baixa resolução de tempo. As Figuras 17 e 18 representam a comparação entre as duas abordagens de transformação do sinal, onde tem-se no segundo caso a representação do grau de flexibilidade na análise tempo-frequência:

Figura 17 – Transformada de Fourier de tempo curto.



Fonte: Adaptado de Vetterli e Herley (1992, p. 2209).

Figura 18 – Transformada *wavelet*.

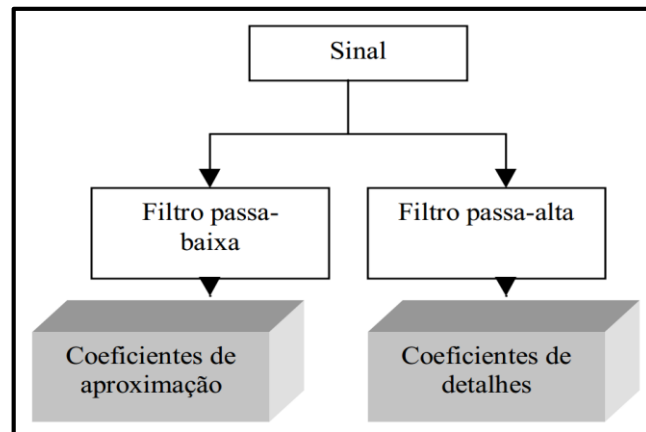


Fonte: Adaptado de Vetterli e Herley (1992, p. 2210).

3.1.6 Transformada *wavelet packet*

Aprofundando-se no conceito de transformação dos sinais, Prarraga (2002, p. 51) demonstra que o funcionamento da transformada *wavelet* pode ser simplificado por filtros passa-baixa e passa-alta. A função *wavelet* está relacionada com o filtro passa-alta, responsável por produzir os coeficientes de detalhes da decomposição. Por sua vez, a função de escalonamento está associada ao filtro passa-baixa, produzindo os coeficientes de aproximação. Pode-se observar visualmente o processo na Figura 19:

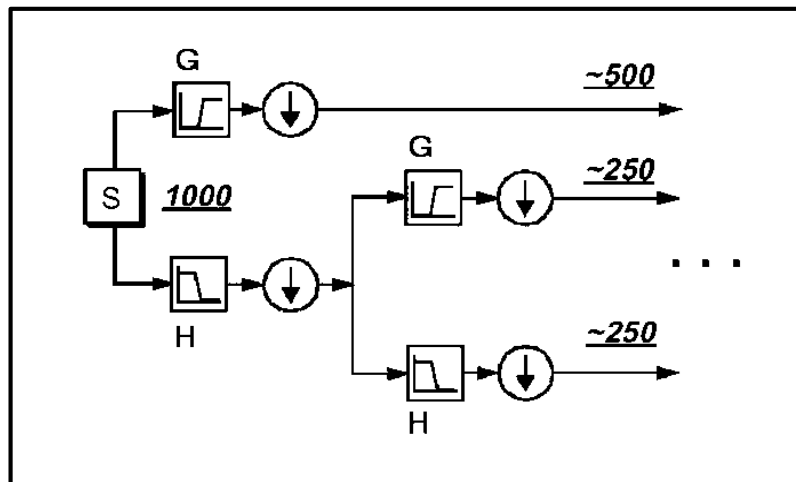
Figura 19 – Funcionamento da transformada *wavelet*.



Fonte: Adaptado de Prarraga (2002, p. 51).

Conforme é dada sequência ao processo de decomposição, é realizado o procedimento de *downsampling* por fator 2, como é exemplificado na Figura 20:

Figura 20 – *Downsampling* presente na decomposição.

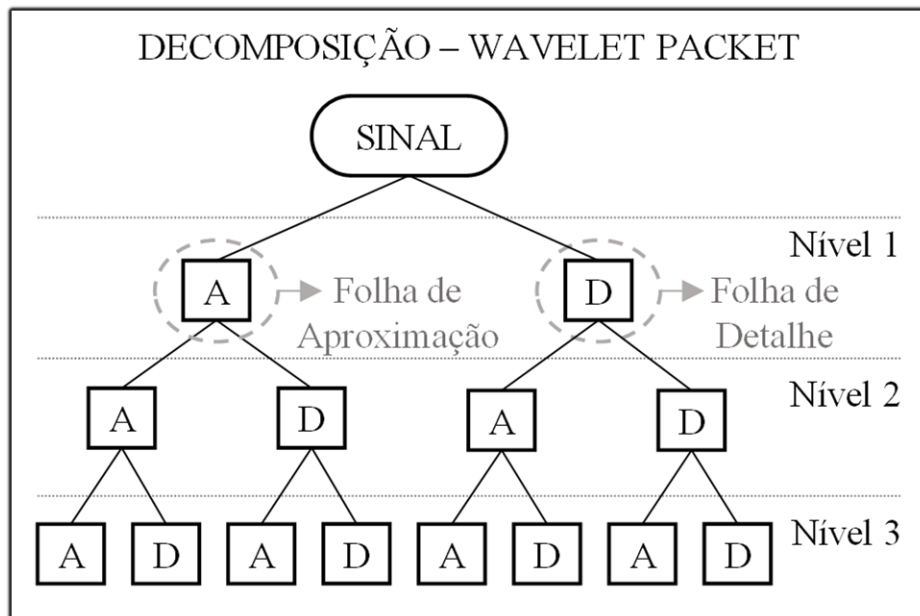


Fonte: Adaptado de Prarraga (2002, p. 53).

Graps (1995, p. 56) apresenta a transformada *wavelet packet* (TWP) como uma variação mais versátil da transformada *wavelet*. É composta pela combinação de várias transformadas *wavelet* e os coeficientes decorrentes de cada decomposição são calculados por algoritmos recursivos, onde o resultado da computação de cada coeficiente se torna a raiz para novos cálculos (decompondo-os em árvore binária).

Dado a característica de decomposição binária, os coeficientes de detalhe e aproximação também são chamados de “folhas” da *wavelet packet*. Permite-se observar melhor o conceito por meio da Figura 21:

Figura 21 – Decomposição da transformada *wavelet packet* em árvore binária.



Fonte: Autoria própria.

Devido sua versatilidade em análises de tempo e frequência, optou-se por utilizar a TWP como extrator de características do sinal para o classificador inteligente, ou seja, o método responsável por preparar o sinal para a identificação de padrões no sinal de voz com distúrbio.

3.1.7 Parâmetros Relevantes

Após a explicação do extrator de características, deve-se conceituar quais características relevantes a serem extraídas de cada um dos coeficientes da decomposição *wavelet packet* para o nível escolhido.

Nos testes empíricos preliminares realizados, dentre as estatísticas disponíveis para as análises das folhas *wavelets* (nível de energia, potência, entropia de Shannon, entropia de energia logarítmica, *threshold*, entre outras) os melhores resultados relacionados ao Edema de Reinke foram obtidos utilizando o nível de energia e entropia de Shannon. Uma possível resposta para tal desempenho pode estar relacionada a capacidade do nível de energia indicar a

distribuição do sinal pelas faixas de frequência (ressaltando irregularidades na periodicidade da voz) e a entropia avalia a quantidade de informação produzida por um processo (influenciada por comportamento em desordem). Tais características são apresentadas na sequência.

Hsu (1995, p. 5) define o valor da energia para um sinal discreto por meio da equação (4):

$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad (4)$$

Onde tem-se que:

- $x[n]$ é o conjunto de pontos no segmento a ser analisado;
- E é o valor da energia obtido.

Quanto a entropia de Shannon, Lima, Assis e De Souza (2011, p. 2) apresentam o conceito como uma relação da quantidade de informação em uma mensagem utilizando medida estatística. Dito isso, uma variável aleatória X com distribuição discreta de probabilidade $p_i = p_1, p_2, \dots, p_k$ associada aos possíveis resultados de eventos correspondentes tem a entropia definida pela equação (5):

$$H(X) = \sum_{i=1}^k p_i \cdot \ln(p_i) \quad (5)$$

Onde tem-se que,

- $H(X)$ é a entropia de Shannon relacionada a variável X ;
- p_i é a probabilidade do i -ésimo termo da distribuição discreta de probabilidade.

3.1.8 Considerações

As informações apresentadas ao longo da seção permitiram aprofundar o entendimento da área de processamento digital de sinais, contribuindo para o amadurecimento da proposta do trabalho. Abordou-se os conceitos fundamentais de transformação dos sinais, tanto por meio da utilização dos filtros quanto pela aplicação de transformadas. Introduziu-se, também,

procedimentos em relação a obtenção de um sinal digital a partir do meio analógico, assim como parâmetros relevantes para a análise sugerida nos objetivos de estudo.

De tal modo, inserido no contexto de classificação de padrões, desenvolveu-se as aplicações e limitações da transformação do sinal, desde a abordagem mais simples de Fourier até atingir os fundamentos de resolução em tempo e frequência presentes na transformada *wavelet packet*, sendo escolhida por suas versatilidades e capacidades de análise no domínio do tempo, sem a necessidade de linearização das características escolhidas (nível de energia e entropia de Shannon) no domínio da frequência.

3.2 SISTEMAS INTELIGENTES

Os sistemas inteligentes são definidos por Jang, Sun e Mizutani (1997, p. 1) como uma abordagem inovadora para explorar a inexatidão e incerteza visando elaborar inteligência computacional. Sabe-se que os problemas encontrados no dia a dia requerem metodologias com características peculiares combinadas que permitem o desenvolvimento de ferramentas inteligentes com atributos da experiência humana ou biológica.

Inicia-se esta seção discutindo brevemente os conceitos básicos das redes neurais artificiais, suas principais características e aplicações e como sua estrutura e seu treinamento são capazes de consolidar o objetivo do trabalho.

3.2.1 Conceitos Iniciais

No contexto de inteligência computacional, existem diversas ramificações em tópicos de estudo e aplicação tendo como exemplo as redes neurais artificiais, sistemas *fuzzy*, computação evolutiva, inteligência coletiva, sistemas imunológicos artificiais e agentes inteligentes. (DA SILVA; SPATTI; FLAUZINO, 2010, p. 24).

Cada uma das áreas apresenta elementos distintos que destacam pontos fortes de aplicação. Como pode ser observado no Quadro 2, as metodologias convencionais possuem destaques para diferentes finalidades de aplicação.

Quadro 2 – Pontos fortes em diversas metodologias inteligentes.

MÉTODOS	PONTOS FORTES
Redes neurais artificiais	Aprendizado e adaptação
Teoria de lógica <i>fuzzy</i>	Representação de conhecimento com regras linguísticas
Algoritmos genéticos e <i>simulated annealing</i> (recozimento simulado)	Busca aleatória sistemática
Inteligência computacional convencional	Manipulação simbólica

Fonte: Adaptado de Jang, Sun e Mizutani (1997, p. 2).

3.2.2 Redes Neurais Artificiais

O termo rede neural artificial (RNA) originou-se das tentativas matemáticas de representar como a informação é processada em sistemas biológicos e vem sendo utilizado até hoje para cobrir diversos modelos diferentes especializados em soluções específicas de problemas. (BISHOP, 2007, p. 226).

As redes neurais artificiais são uma representação de como aproveitar aspectos do raciocínio humano para resolução de problemas do mundo real. Portanto, devem ser capazes de extrair conhecimento de determinada área ou problema, se adaptarem para atingir melhor desempenho dentro do contexto inserido mediante ao treinamento e conseguirem atingir o processo de generalização para estimar uma solução ou resultado, mesmo em situações distintas ao que foi presenciado durante o processo de treinamento.

Contudo, para modelagens complexas em classificação de padrões, por exemplo, o problema da generalização é definido pela dificuldade em separar as amostras de forma eficiente e correta. É improvável que para boa generalização a fronteira de decisão esteja apenas refinada aos eventos do treinamento ao invés das características presentes na real modelagem do problema a ser resolvido. (DUDA; HART; STORK, 2012, p. 5)

Como abordado por Wilamowski (2009, p. 60), deve-se escolher o menor número de neurônios artificiais (estrutura básica da modelagem de uma RNA) possível, pois uma rede com quantidade reduzida de neurônios artificiais e interconexões geralmente apresenta a melhor generalização, mesmo que seja necessário sacrificar uma pequena parcela de acertos na etapa de treinamento.

3.2.3 Principais Características das RNA

Como é discutido por Haykin (1997, p. 24) e Da Silva, Spatti e Flauzino (2010, p. 24), o poder computacional da RNA deriva de uma estrutura massivamente paralela e distribuída capaz de aprender e generalizar. Portanto, tal processamento de informação possibilita o desenvolvimento de soluções de alta complexidade e larga escala. Logo, é importante ressaltar algumas das características principais de uma rede neural artificial:

- **Não-linearidade:** é propriedade de fundamental importância, principalmente em situações onde o mecanismo físico responsável pelas amostras de entrada é não-linear.
- **Capacidade de aprendizado:** aplicando diferentes técnicas de treinamento, a rede consegue absorver a relação entre as variáveis de entrada que compõe o modelo e a saída desejada, permitindo a extração de contextos e interpretações sobre o problema.
- **Habilidade de generalização:** como já discutido anteriormente, seguido de um processo de treinamento bem executado, a rede deve ser capaz de se adaptar ao universo exposto e generalizar o conhecimento adquirido, encontrando soluções que não eram previamente de fácil obtenção.
- **Resposta probatória:** na operação como classificador de padrões, permite ser projetada para desenvolver respostas sobre a categoria da amostra como também a confiança no resultado apresentado, permitindo o descarte de estimações ambíguas e, conseqüentemente, melhorando a confiabilidade da rede.
- **Organização de dados:** por meio da percepção de características desconhecidas dos dados apresentados, a rede é capaz de realizar a sua organização interna visando possibilitar o agrupamento de padrões que muitas vezes não são facilmente observados por métodos convencionais.
- **Tolerância a falhas:** apresentando elevado nível de interconexões entre os neurônios artificiais, paralelizando o processamento semelhante ao biológico, a rede neural torna-se um sistema mais tolerante a falhas quando parte de sua estrutura interna é sensivelmente corrompida.

- **Armazenamento distribuído:** como abordado no tópico anterior, a estrutura composta por diversas conexões entre subestruturas modulares (neurônios artificiais), o armazenamento se torna distribuído e incrementa a confiabilidade no processo caso eventuais neurônios deixem de funcionar.
- **Facilidade de prototipagem:** considerando que o processo de treinamento pode ser realizado antes da prototipagem para um *hardware*, a implementação da maioria das arquiteturas é um processo relativamente simples e de pouco esforço computacional, uma vez que os seus resultados são normalmente obtidos por algumas operações matemáticas elementares.

3.2.4 Áreas de Aplicações

As redes neurais artificiais são ferramentas poderosas, sendo utilizadas em diversos problemas presentes nos estudos de engenharia e ciências. Como sugerido por Hagan (2014, p. 1-5 a 1-8), as potenciais áreas de aplicação estão definidas no Quadro 3:

Quadro 3 – Potenciais áreas de aplicação das RNAs.

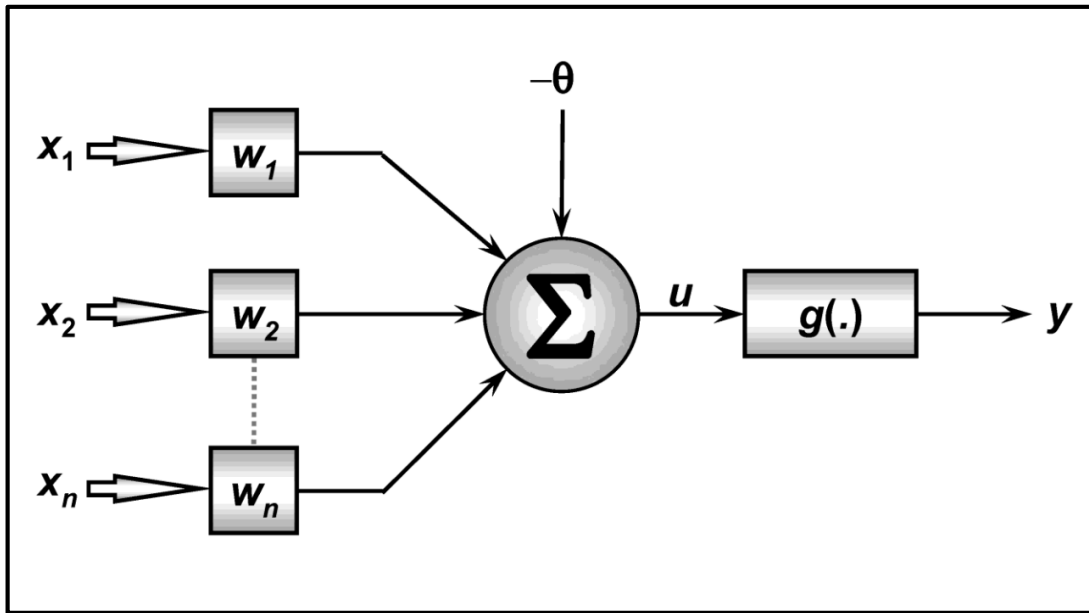
ÁREA	APLICAÇÃO
Aeroespacial	Sistemas de controle de aeronaves, simulação de falhas e componentes, desempenho de piloto automático
Automotivo	Controle de injeção de combustível, sistemas de frenagem automática, detecção de falhas e sensores
Bancária	Leitura de documentos, avaliação de crédito, previsão de caixa, reconhecimento de assinatura, risco de crédito
Defesa Militar	Controle de direção de armas, identificação de objetos, sonares, radares, processamento de sinais e imagens,
Fala	Reconhecimento de fala, compressão de som, classificação de vogais, sintetização de voz,
Médica	Análise de células cancerígenas, análise de sinais biológicos, projeto de próteses, otimização de tempo de transplante
Produção	Projeto e análise de produtos, controle de qualidade e processos, planejamento e gestão, sistemas de inspeção
Robótica	Controle de trajetória, manipulação de controladores e sensores, sistemas de visão, veículos autônomos
Telecomunicações	Compressão de imagens e dados, serviços automatizados de informação, traduções em tempo real

Fonte: Adaptado de Hagan (2014, p. 1-5 a 1-8).

3.2.5 Estrutura Básica da RNA – Neurônio Artificial

Antes de se aprofundar nas arquiteturas mais complexas, deve-se apresentar a estrutura básica de uma RNA: o neurônio artificial. São representados na Figura 22 os elementos que compõe o neurônio artificial.

Figura 22 – Representação do neurônio artificial.



Fonte: Spatti (2011, p. 28).

Detalhadamente, Spatti (2011, p.28) caracteriza que o funcionamento consiste da soma linear ponderada por meio dos pesos sinápticos (w_i) referentes a cada valor de entrada (x_i). Essa soma deve superar uma barreira potencial definida pelo valor do limiar de ativação (θ) para que ocorra a ativação do neurônio e o valor da soma seja repassado a função de ativação (representado por $g(\cdot)$), responsável por confinar o valor de saída da maneira desejada.

O neurônio artificial define também as arquiteturas mais simples de RNA (*perceptron* simples ou *adaline*), composta por apenas um único neurônio artificial.

3.2.6 Funções de Ativação

O funcionamento das RNAs mais frequentes é definido pelo conceito de *feedforward* (em tradução livre, “alimentação para frente”) representando que o processamento dentro da estrutura segue um fluxo da entrada para saída, sem laços ou realimentações.

Bishop (1995, p. 117) ressalta a importância dessa característica, pois permite representar funções não-lineares de várias variáveis em termos de composições de funções não-lineares de única variável, chamada de função de ativação.









Existem diversas funções de ativação, cada qual com sua específica aplicabilidade. Fausett (1993, p. 17-19), Kriesel (2007, p. 42-44) e Hagan (2014, p. 2-3 a 2-6) discutem as funções mais comuns:

- **Função degrau (ou Heaviside):** se comporta de maneira discreta, limitando a saída entre 0 e 1 (ou entre -1 e 1, para degrau bipolar).
- **Função rampa simétrica:** apresenta valores definidos pelos mesmos potenciais de ativação ou, caso exceda o domínio da função, geralmente limitada entre -1 e 1. Obtém valores nem sempre saturados, permitindo extrair o grau de confiança no resultado.
- **Função logística (ou Fermi):** representa comportamento semelhante a função rampa, mas pode ser utilizada em algoritmos de requerem diferenciação. É confinada em valores de 0 a 1.
- **Função tangente hiperbólica:** similar a função logística, contudo é definida para o intervalo de -1 a 1.

Tanto a função logística quanto a tangente hiperbólica podem ter seus coeficientes ajustados, permitindo um grau de rápida saturação com comportamento semelhante à função degrau.

Por intermédio do Quadro 4 possibilita-se compreender outras principais funções de ativação, bem como, seu comportamento em relação a entrada e saída (sendo “ Y ” o valor de saída e “ u ” o valor recebido pela função de ativação) e a simbologia que representa seu princípio de operação:

Quadro 4 – Principais funções de ativação.

NOME	RELAÇÃO ENTRADA (u) E SAÍDA (Y)	SÍMBOLO
Degrau	$u < 0 \quad Y = 0$ $u \geq 0 \quad Y = 1$	
Degrau bipolar	$u < 0 \quad Y = -1$ $u \geq 0 \quad Y = +1$	
Linear	$Y = u$	
Linear simétrica	$u < 0 \quad Y = 0$ $0 \leq u \leq 1 \quad Y = u$ $u > 1 \quad Y = 1$	
Rampa simétrica	$u < -1 \quad Y = -1$ $-1 \leq u \leq 1 \quad Y = u$ $u > 1 \quad Y = 1$	
Logística	$Y = \frac{1}{1 + e^{-u}}$	
Tangente hiperbólica	$Y = \frac{e^u - e^{-u}}{e^u + e^{-u}}$	
Linear positiva	$u < 0 \quad Y = 0$ $0 \leq u \quad Y = u$	
Competição	$Y = 1$ neurônio com maior u $Y = 0$ outros neurônios	-----

Fonte: Adaptado de Hagan (2014, p. 1-5 a 1-8).

3.2.7 Arquiteturas de Redes Neurais

Os autores Da Silva, Spatti e Flauzino (2010, p. 46) definem as redes neurais em três estruturas fundamentais:

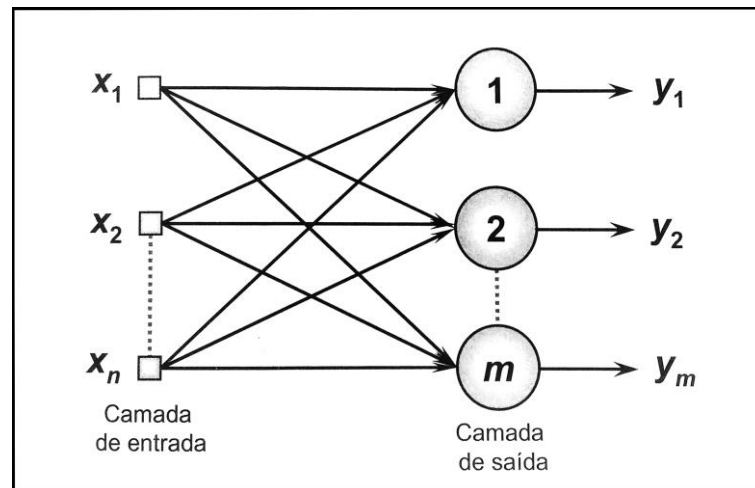
- **Camada de Entrada:** responsável pela introdução dos dados ao sistema (medidas, sinais, características, padrões). Medidas geralmente normalizadas nas faixas de variação para obter-se os melhores resultados em relação a função de ativação.
- **Camadas Escondidas (ou intermediárias, ocultas):** compostas pelos neurônios encarregados de extrair o conhecimento do problema proposto. Maior parte do processamento acontece nas camadas escondidas.

- **Camada de Saída:** estrutura incumbida de gerar e apresentar os resultados do processo das camadas anteriores.

Quanto a organização das camadas, Haykin (1997, p. 43) separa as redes neurais de maneira generalizada em três categorias mais comuns:

- **Redes *feedforward* de camada única:** composta por apenas uma camada de entrada e uma camada de saída, onde as informações seguem apenas um sentido. É denominada camada única por não levar em consideração a camada de entrada, uma vez que nenhum processamento é realizado por ela. Exemplo de redes: *Perceptron* simples e *Adaline*. Pode ser observado na Figura 23 o arranjo de camada única.

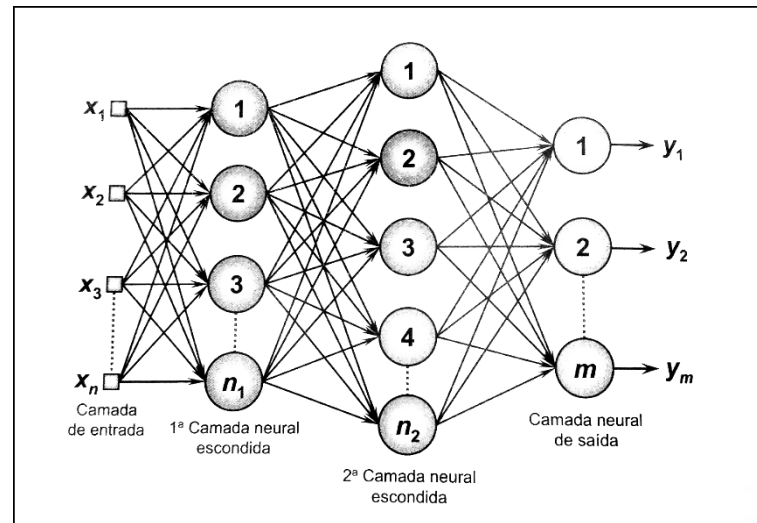
Figura 23 – Exemplo de rede *feedforward* de camada única.



Fonte: Da Silva, Spatti e Flauzino (2010, p. 47).

- **Redes *feedforward* de múltiplas camadas:** se diferencia da primeira categoria por possuir uma ou mais camadas escondidas. A adição de camadas possibilita a extração de características de ordens elevadas, se tornando particularmente interessante para situações de muitas entradas. Exemplo de redes: *Perceptron* Multicamadas (PMC) e redes de base radial. Por sua vez, é demonstrado na Figura 24 a organização das conexões sinápticas em múltiplas camadas:

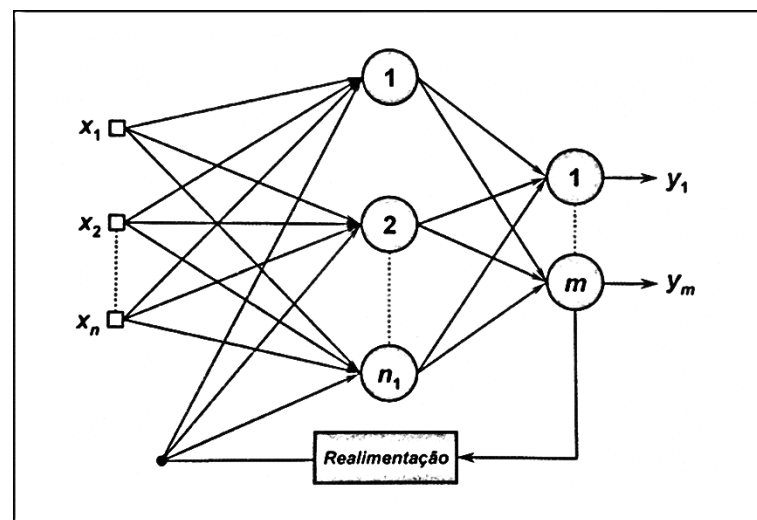
Figura 24 – Exemplo de rede *feedforward* de camada múltiplas.



Fonte: Da Silva, Spatti e Flauzino (2010, p. 48).

- Redes Recorrentes:** devem ser compostas por pelo menos um laço de realimentação, onde o resultado da camada de saída também realimenta os neurônios da camada de entrada, impactando profundamente a capacidade de aprendizado e desempenho da rede. Permite processamento dinâmico de informações e utilização em sistemas variantes em relação ao tempo. Exemplo de redes: Hopfield e PMC com realimentação. É possível definir sua organização com o auxílio da Figura 25, uma rede PMC com laço de realimentação.

Figura 25 – Exemplo de rede recorrente.



Fonte: Da Silva, Spatti e Flauzino (2010, p. 48).

3.2.8 Treinamento e Aprendizado

Sabe-se que a propriedade mais importante das redes neurais é a habilidade de aprender por meio dos dados e com isso melhorar seu desempenho. Isso é feito através de um processo iterativo de ajustes aplicado a seus pesos, o treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas.

Haykin (1997, p.85) aborda, de forma generalizada, dois tipos básicos de treinamento para a RNA:

- **Treinamento Supervisionado:** é análogo a um treinamento com professor que sabe as respostas certas enquanto a rede, por sua vez, não as conhece. Os erros entre as respostas da rede comparadas com as do “professor” são empregados nos ajustes dos parâmetros livres com o objetivo de aproximar o comportamento da RNA para uma simulação mais próxima possível do “professor”.
- **Treinamento Não-supervisionado:** intuitivamente, no método não-supervisionado não há respostas desejadas para a camada de saída, tornando responsabilidade da própria rede se ajustar e organizar os padrões de forma a reconhecer suas particularidades e os agrupá-los (princípio da clusterização).

Por fim, Kriesel (2007, p. 62) ressalta, similarmente, dois tipos diferentes de aprendizado para uma rede neural:

- **Off-line:** grande quantidade de padrões são mostrados para a rede simultaneamente, para depois calcular os erros e ajustar os parâmetros livres da rede (peso sinápticos e limiar de ativação). Também é conhecido por *batch training procedures* (procedimento de treinamento em lote), pois todos os valores são corrigidos na mesma etapa.
- **On-line:** inversamente, nesse caso é realizado ajustes a cada entrada exposta a rede, ajustando os valores para cada erro de amostra. É utilizado para aplicações em ambiente muito dinâmicos, que necessitam de rapidez para manter a relevância das amostras e requerem a passagem de muitas entradas para a rede.

3.2.9 Considerações

Como é desenvolvido por Hornik, Stinchcombe e White (1989, p. 359), fundamentado no Teorema da Superposição de Kolmogorov, estabeleceu-se que uma rede *feedforward* de múltiplas camadas é capaz de aproximar qualquer função mensurável em qualquer nível de precisão desejada, em outras palavras, um aproximador universal de funções. Tal afirmação reforça a hipótese de que a PMC seja suficientemente capaz de lidar com a complexidade do problema proposto. É observado, também, por Bishop (2007, p. 226) que a rede neural artificial que apresenta maior valor prático para resolver problemas de classificação de padrões é a *perceptron* multicamadas.

Dito isso, qualquer falta de sucesso na aplicação é derivada de um treinamento inadequado, número insuficiente de neurônios ou camadas escondidas ou também da falta de relação determinística entre entradas e saídas.

Logo, por intermédio dos fatos e informações expostas, conclui-se que a PMC exhibe as melhores características para resolver o objetivo proposto de classificação de padrões de alta complexidade e não-linearidade presentes no sinal da voz. Define-se similarmente a utilização de treinamento supervisionado e aprendizado *off-line* uma vez que são conhecidas as saídas desejadas na etapa de treinamento e o ambiente não é dinâmico a ponto de justificar um aprendizado *on-line*. E, por fim, será adotado a função de ativação de rampa simétrica para a camada de saída, permitindo posteriormente identificação de grau de confiança no contexto dos resultados obtidos.

4 MATERIAIS E MÉTODOS

4.1 MATERIAIS

Serão discutidos nesta seção os materiais e ferramentas necessárias para consolidar os objetivos propostos no trabalho.

4.1.1 Banco de Dados

Em relação aos sinais vocais, o banco de dados é composto por vozes gravadas em ambiente controlado, com equipamentos adequados e realizados por profissionais da área de fonoaudiologia. O material foi cedido pelo Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GPEM/CNPq), contendo amostras saudáveis e com distúrbio de Edema de Reinke. O grupo de amostras contém pacientes de diversas idades, (sexo masculinos e femininos) coletadas ao longo dos anos de atuação do grupo.

Dos pacientes selecionados nas amostras foram extraídos a produção vocal acústica da vogal sustentada /a/ emitidas em *pitch*¹ e *loudness*² confortáveis. A vogal sustentada /a/ apresenta uma maior abertura glótica em relação as outras vogais, permitindo melhor identificação da presença de irregularidades na produção vocal.

Tais gravações foram coletadas nas dependências do Ambulatório de Voz do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP). Os registros da emissão vocal foram realizados por fonoaudiólogos experientes voz em sala acusticamente tratada. Para a captura, gravação e armazenamento dos sinais de voz foram utilizados: o software *Audacity* em computador Pentium II acoplado a uma placa de som e um microfone *headset* unidirecional da marca AKG– C520 L. O microfone foi posicionado à distância de quatro centímetros da boca, formando com esta um ângulo de 45° à 90°.

¹ “*Pitch*, do ponto de vista perceptual, corresponde ao número de ciclos de vibração da pregas vocais por segundo (frequência fundamental).” (PINHO; TSUJI; BOHADANA, 2006, p. 16).

² “*Loudness*, ou volume sonoro, é a referência perceptiva da quantidade de som feita através do sistema auditivo (que inclui o cérebro).” (TITZE, 2013, p. 282).

Os sinais acústicos foram armazenados em extensão de arquivo .WAV à taxa de 44100 amostras por segundos. Contudo, para certas gravações mais antigas, o laboratório dependia de uma placa de aquisição de modelo defasado, permitindo apenas 22050 amostras por segundo.

A composição geral dos arquivos presentes no banco de dados está representada no Quadro 5:

Quadro 5 – Composição do banco de dados.

Característica	Quantidade
Saudável	85
Edema de Reinke	45
Total	130

Fonte: Autoria própria.

Vale ressaltar que cada arquivo de áudio é referente à indivíduos diferentes e os ajustes que deverão ser realizados para tratar as discrepâncias entre taxas de amostragem e número de arquivos por padrões serão discutidos detalhadamente posteriormente nos procedimentos.

4.1.2 Ferramentas Utilizadas

Todas as etapas práticas do trabalho serão construídas dentro do ambiente computacional do *software* MATLAB[®], permitindo facilidade para implementar os códigos, capacidade de automatizar os procedimentos e agilidade caso seja necessário processamento paralelizado onde há grande esforço computacional.

Será utilizado também as funções já implementadas nos pacotes instalados em conjunto com o MATLAB[®], por exemplo, para aplicação de transformadas e extração de parâmetros por meio da *toolbox* de processamento de sinais e treinamento do sistema inteligente presente na *toolbox* de redes neurais.

4.2 MÉTODOS

Serão abordados, então, os métodos que fundamentam e estruturam a proposta deste trabalho. Entretanto, primeiramente, é essencial definir o projeto de um classificador de padrões

antes de delinear os métodos que serão aplicados. Feito isso, em seguida, será explicado como esses métodos se relacionam com a estruturação do objetivo a ser alcançado.

4.2.1 Sistema Classificador de Padrões

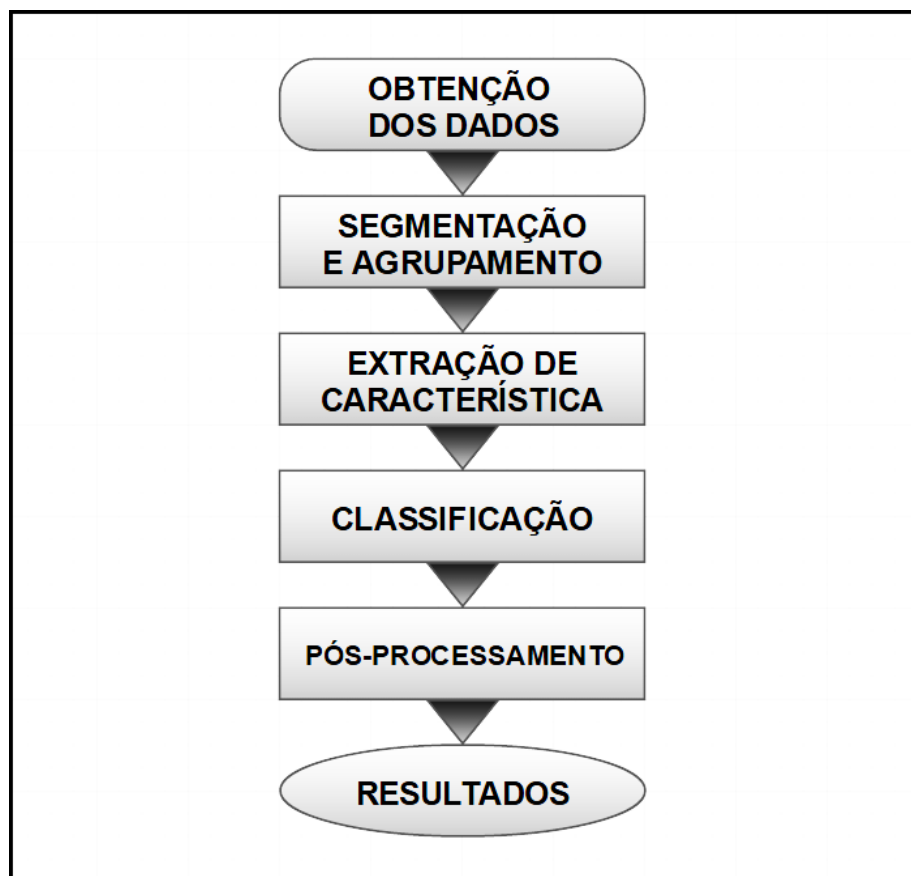
Segundo Duda, Hart e Stork (2012, p. 9), um sistema classificador de padrões pode ser dividido em etapas fundamentais: obtenção dos dados, segmentação e agrupamento, extração de característica, classificação e pós-processamento. Seus detalhes são descritos em sequência:

- **Obtenção dos dados:** etapa responsável pela coleta das informações que alimentarão o sistema. É geralmente composta por transdutores como microfones, câmeras ou sensores e também é limitado pelas características de construção dos mesmos como resolução, sensibilidade, distorção ou ruído.
- **Segmentação e agrupamento:** como consequência da etapa anterior, o dado obtido sob condição de não perder suas características relevantes à modelagem do problema consegue ser segmentado em porções menores com o intuito de detalhar o sinal e permitir aprofundar-se nas características e incrementar o número de amostras para o treinamento.
- **Extração de característica:** em essência, tem como função principal distinguir tal objeto de forma a ser reconhecido por meio de valores mensuráveis como participante de uma dada série de padrões. Em conjunto com o crescimento da complexidade do problema, esse procedimento enfrenta obstáculos definidos por pequenas variações de valores que, supostamente, deveriam ser semelhantes dentro da mesma categoria e o surgimento de valores similares em múltiplas classes.
- **Classificação:** como os autores ressaltam, a diferenciação entre extrair características e de fato classificá-las é arbitrária. Dado um classificador ideal, o próprio procedimento de separação das características possibilita a visualização do resultado de um classificador. Contudo, para sistemas mais elaborados é necessário dividir o processo com o propósito de se ter uma etapa destinada unicamente a entender os agrupamentos e identificá-los nas categorias propostas.

- **Pós-Processamento:** utiliza-se o período de pós-processamento para observar contextos mais abstratos como a relação das características extraídas em função das classes desejadas, eficiência do sistema, taxas de acertos, grau de confiabilidade nos resultados e como esses fatores podem influenciar a viabilidade da ferramenta na prática.
- **Resultados:** por fim, o último procedimento é responsável por lidar com os resultados obtidos na classificação e absorver os conceitos obtidos anteriormente.

Recorrendo a Figura 26, pode-se notar de forma simplificada o fluxograma das etapas de um sistema classificador de padrões:

Figura 26 – Fluxograma simplificado das etapas de um classificador de padrões.



Fonte: Autoria própria.

4.2.2 Preparação dos Dados

Partindo do princípio que o banco de dados já foi cedido conforme comentado no tópico de materiais, seguindo todos os cuidados necessários na aquisição e digitalização dos arquivos, a etapa de coleta de dados não é incluída no contexto do trabalho. Portanto, inicia-se os métodos pela preparação dos dados obtidos.

A primeira metodologia é composta pelo agrupamento dos arquivos de áudio por suas características, possibilitando a separação em dois grupos: voz saudável e Edema de Reinke.

Tem-se também nesta etapa a análise da taxa de amostragem de cada arquivo. Primeiramente, é coletado a menor taxa de amostragem do banco de dados e em seguida realiza-se a reamostragem dos arquivos para a taxa de amostragem coletada (22050Hz). Este procedimento tem o intuito de garantir uniformidade no banco de dados sem a inserção de informação sintética nos dados.

4.2.3 Segmentação e Janelamento

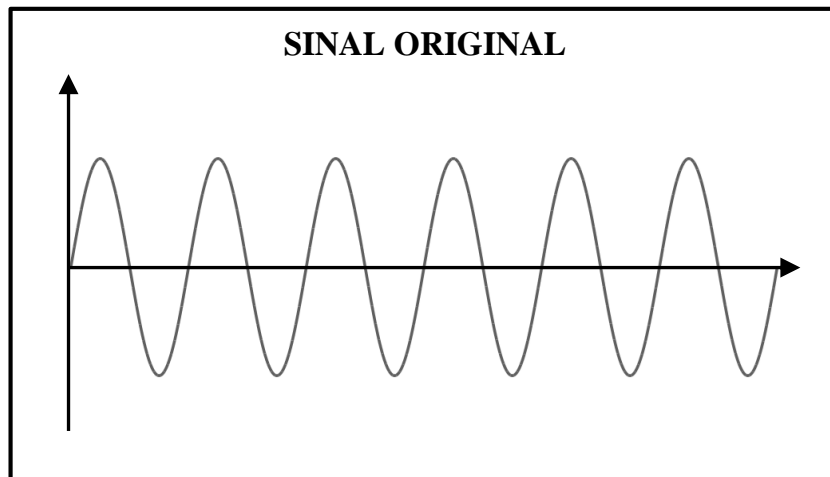
O método de segmentação e janelamento do sinal é essencial para alcançar os objetivos desejados. Será responsável tanto para agilizar o processo (reduzindo o tempo necessário para a aplicação da TWP e a extração das características), quanto para produzir uma quantidade mais elevada de amostras para o treinamento, aumentando o desempenho e validação das RNA.

Para a explicação do processo de janelamento, supõe-se um sinal original qualquer de 100 pontos, demonstrado na Figura 27.

O janelamento simples é análogo ao conceito de segmentação, onde tem-se um sinal completo e a cada janelamento é dado um passo correspondente ao próprio comprimento da janela (não há sobreposição entre os sinais em cada janela), em outras palavras, o sinal é simplesmente segmentado em porções de tamanho desejado. Pode-se então, dividi-lo em 2 partes de 50 pontos, como é demonstrado de maneira simplificada na Figura 28.

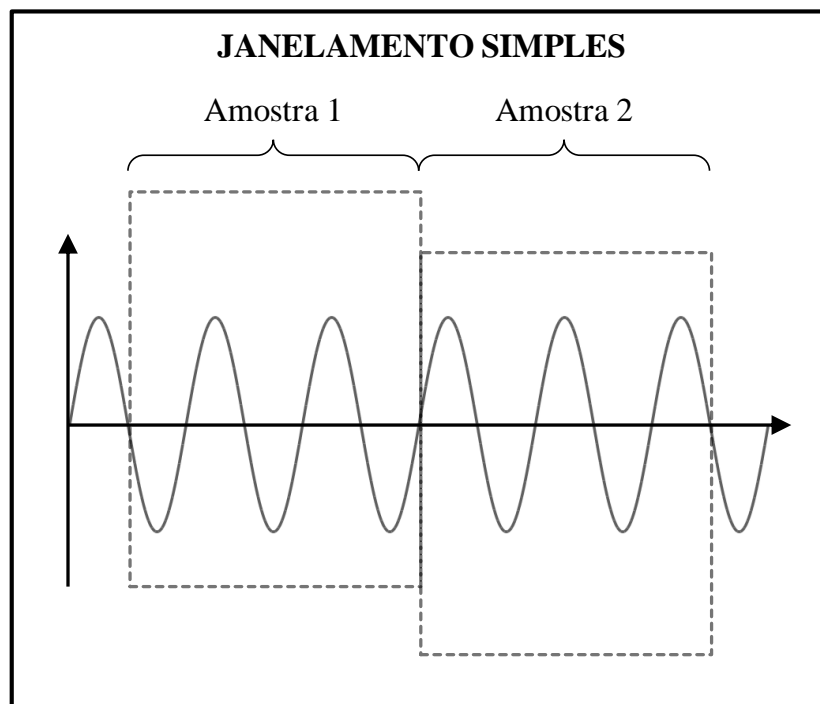
Contudo, no janelamento cruzado, após retirar a primeira parte do sinal (correspondente aos primeiros 50 pontos), é dado um valor de passo menor que o próprio tamanho da janela (no exemplo, 50% de superposição ou passo referente a 25 pontos), percorrendo um espaço menor e acarretando a segmentação em janelas cruzadas. O processo de janelas cruzadas é representado na Figura 29.

Figura 27 – Exemplo de sinal completo a ser janelado.



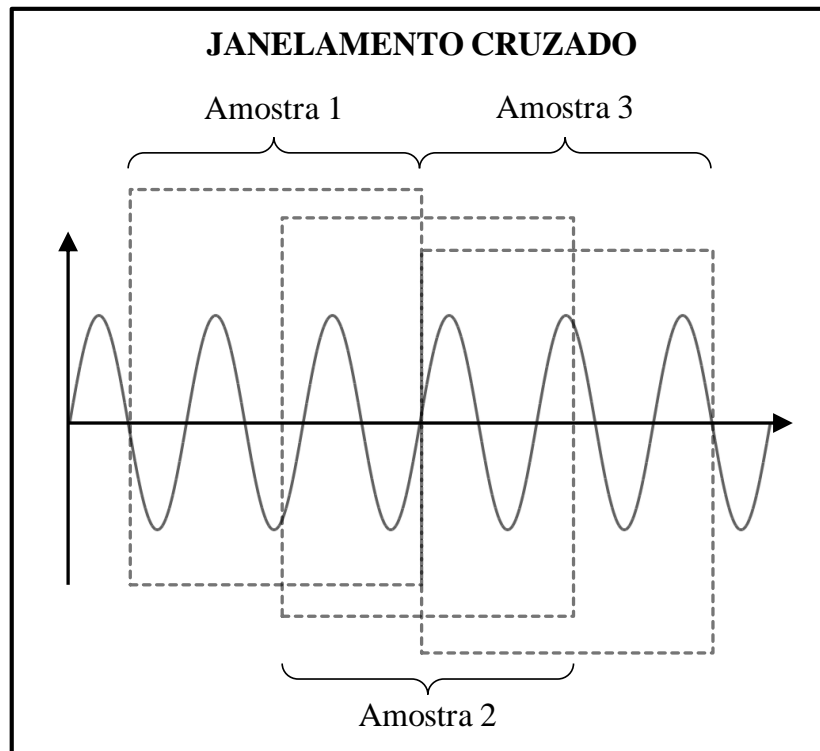
Fonte: Autoria própria.

Figura 28 – Exemplo de sinal após procedimento de janelamento simples.



Fonte: Autoria própria.

Figura 29 – Exemplo de sinal com janelamento cruzado.



Fonte: Autoria própria.

Para a construção do banco de amostras do estudo proposto optou-se pelo janelamento cruzado com o objetivo de aumentar o número de amostras e incrementar a diversidade de informações nas janelas, garantindo um melhor treinamento da RNA.

4.2.4 Aplicação da *wavelet packet* e extração de características

Conforme já abordado anteriormente, esta etapa dos métodos é responsável por detalhar o sinal por meio de uma análise tempo-frequência em função da TWP. Para realizar a esta aplicação é necessário, primeiramente, escolher até qual nível será feita a decomposição da *wavelet packet*.

Nota-se que, conforme eleva-se o nível de resolução dos coeficientes (ou folhas), é dobrado o número de coeficientes (tem-se o número de coeficientes realizando o cálculo: $2^{\text{nível}}$), aumentando consideravelmente o conjunto de entradas para as RNAs e incrementando fortemente o esforço computacional e complexidade do sistema.

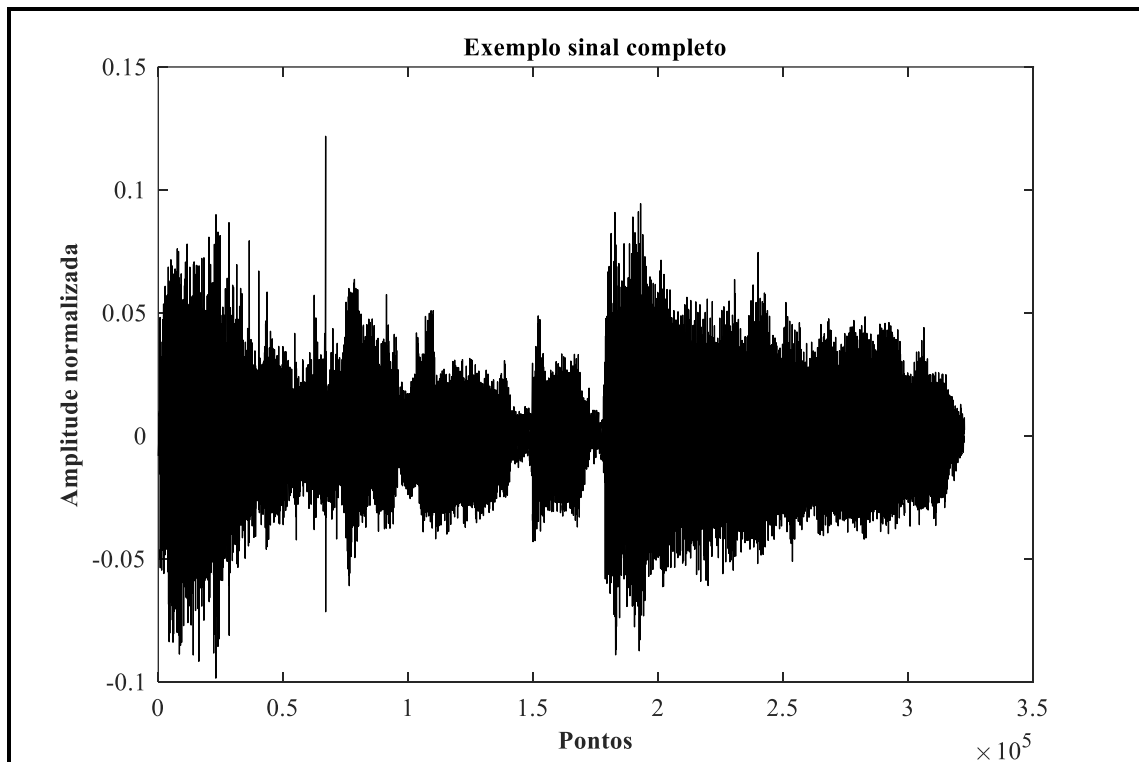
No escopo do estudo proposto e considerando-se limitados recursos computacionais, determinou-se análises no segundo e terceiro nível. Será comparado qual apresentará melhor

capacidade de extração de informações relevantes para generalização da RNA na etapa de treinamento.

Feito isso, tem-se o passo de extração das características. Optou-se por trabalhar com todos os coeficientes do mesmo nível e calculando o nível de energia e o valor da entropia de Shannon para cada coeficiente. Tais valores obtidos para cada janela de sinal formarão o conjunto de entradas da RNA.

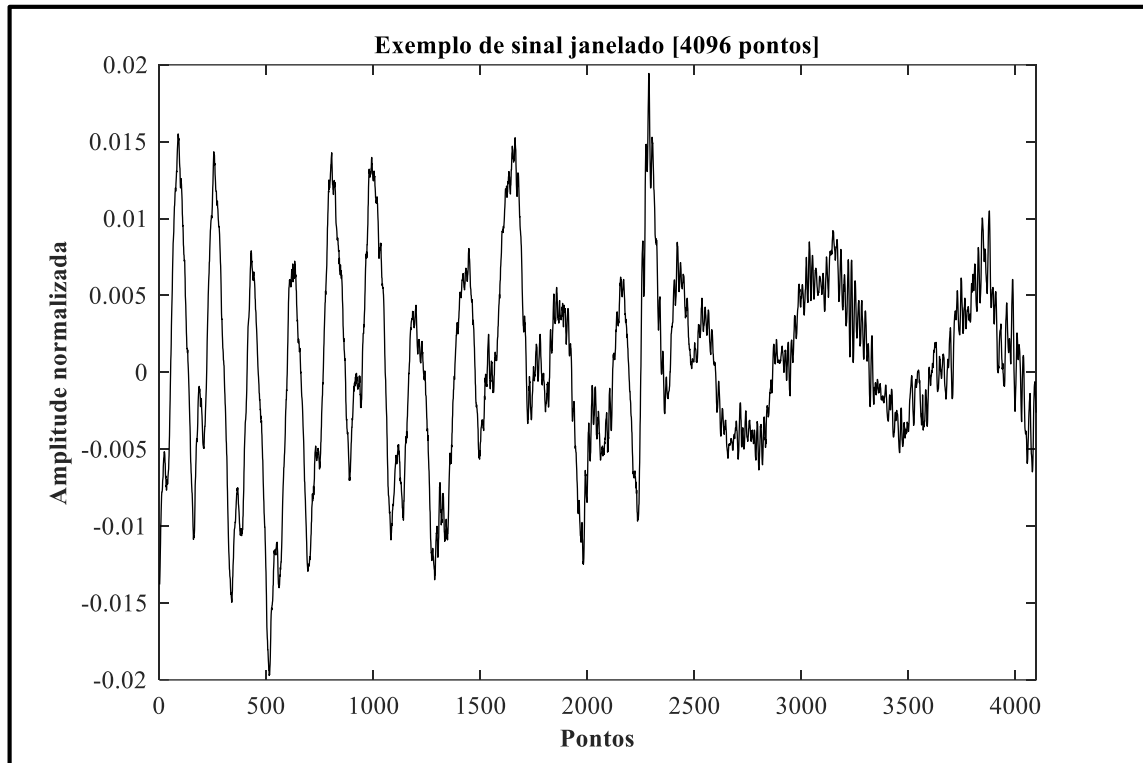
Mediante as Figuras 30, 31 e 32, é possível visualizar um exemplo de arquivo de áudio completo, uma janela qualquer do sinal de 4096 pontos e também a decomposição da wavelet packet dentro da interface computacional do *software* MATLAB[®] utilizando as funções implementadas em sua *toolbox*.

Figura 30 – Sinal completo para exemplificação.



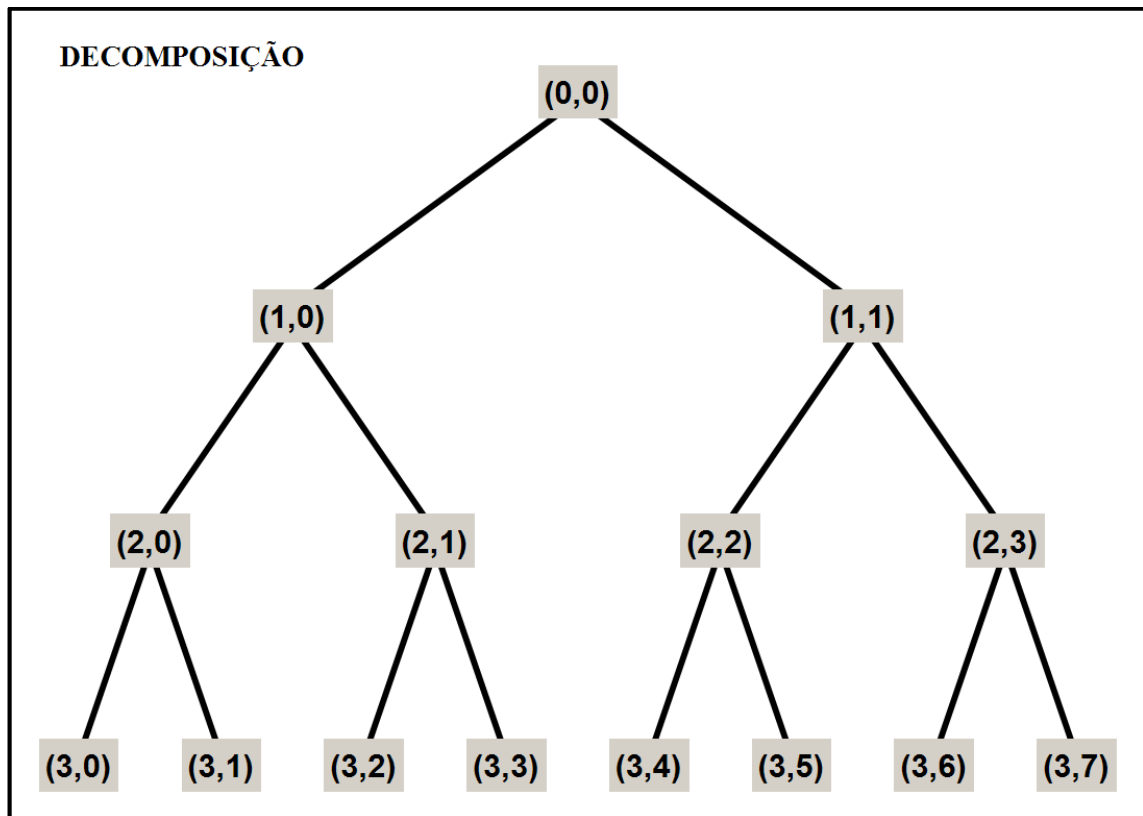
Fonte: Autoria própria.

Figura 31 – Sinal segmentado em 4096 pontos para exemplificação.



Fonte: Autoria própria.

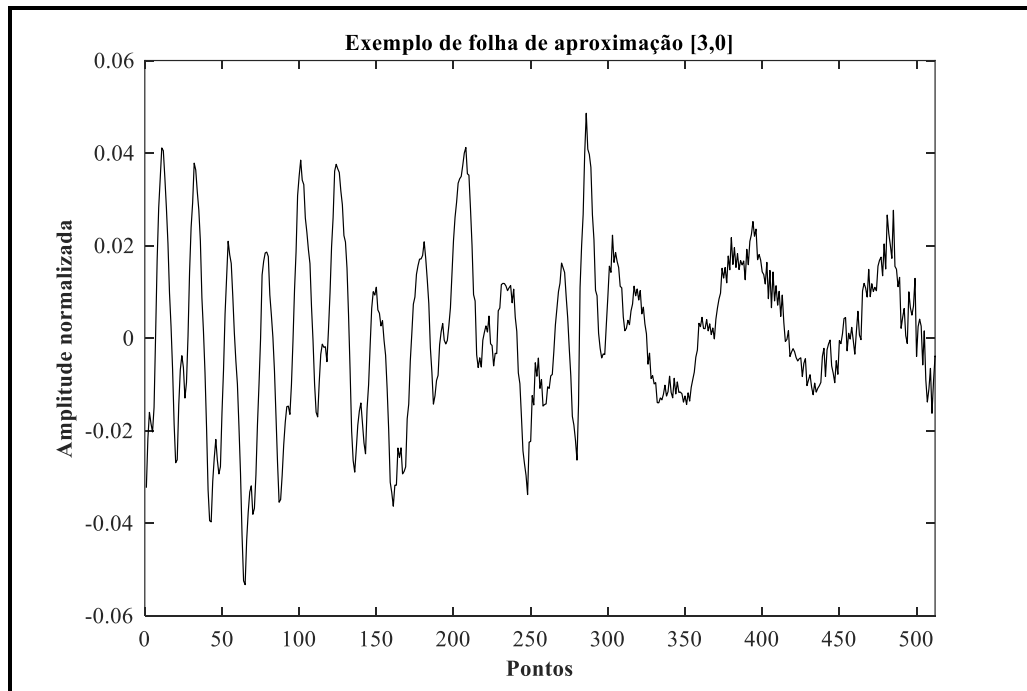
Figura 32 – Exemplo de decomposição da TWP até terceiro nível.



Fonte: Autoria própria.

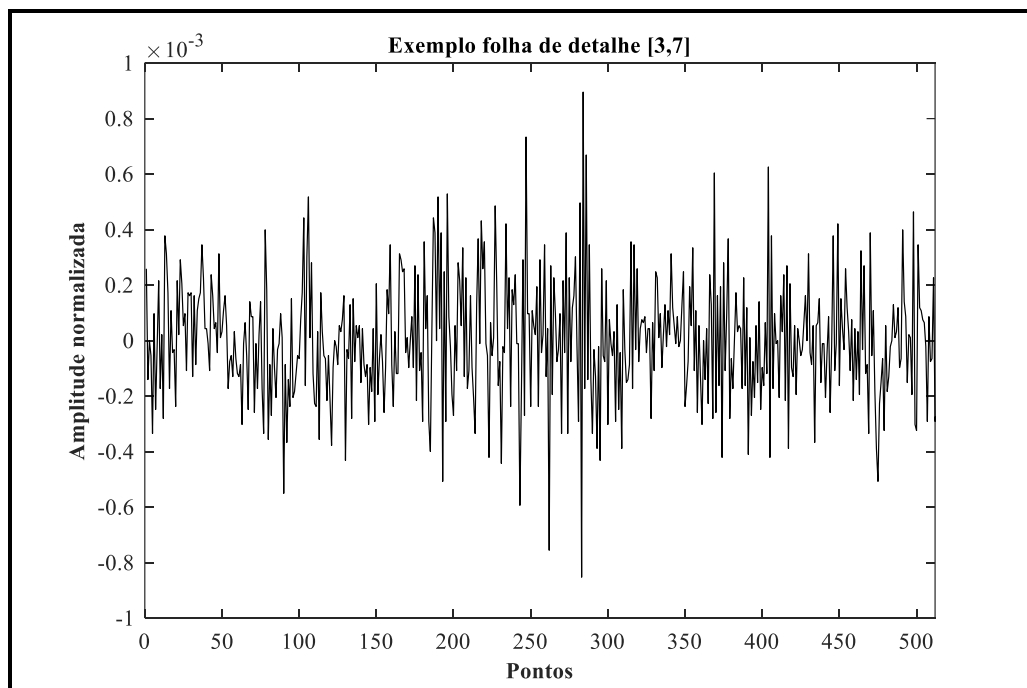
Por sua vez, por meio das Figuras 32 e 33, é possível notar as diferenças entre uma folha de aproximação (apresenta as baixas frequências e característica de onda semelhante ao sinal original) e uma folha de detalhe (apresenta as componentes de alta frequência, sem semelhança visual com o formato original do sinal).

Figura 33 – Exemplo de folha de aproximação.



Fonte: Autoria própria.

Figura 34 – Exemplo de folha de detalhe.



Fonte: Autoria própria.

4.2.5 Treinamento das Redes Neurais

O treinamento das RNA visa obter a topologia que melhor se adapte e generalize as condições delimitadas para o trabalho. Contudo, a determinação do arranjo das camadas e do número de neurônios que as compõe consiste em testes empíricos, buscando meios que validem a eficiência de uma topologia em relação a outras já testadas.

Em virtude do extenso banco de entradas e o grande esforço computacional exigido para o treinamento de uma RNA, é de extrema importância que se elabore um algoritmo automatizado capaz de incluir todas as variações de nível de resolução discutidas combinadas com as possíveis topologias a serem adotadas. Similarmente, é necessário incluir no processo automatizado a coleta de informações que comprovem a eficácia de cada rede testada (taxa de acertos, precisão, entre outros) para, enfim, definir consistentemente a melhor configuração do classificador.

4.2.6 Pós-Processamento e Obtenção de Resultados

Como abordado anteriormente, o método de pós-processamento é capaz de obter informações relevantes de difícil observação natural, tornando-se um grande incremento aos resultados convencionais da discriminação entre patologias.

Essa situação dentro do escopo do trabalho permite aferir, por exemplo, a porcentagem de amostras inclusas em cada classe para determinado arquivo de áudio. Outro exemplo é a confiabilidade de cada resultado devido a gradação que a função de ativação (rampa simétrica) dispõe antes de chegar em sua região de saturação. Da mesma forma, tal método propicia a observação e avaliação da relação entre os elementos utilizados nas entradas com os tipos de saída do sistema inteligente.

Conjuntamente, essas informações são fundamentais para avaliar a eficácia do desenvolvimento do trabalho, conjecturando sobre sua funcionalidade e robustez tanto para as amostras cuidadosamente coletadas e digitalizadas quanto para o sistema em funcionamento real, sem preocupação com pequenos ruídos ou perturbações.

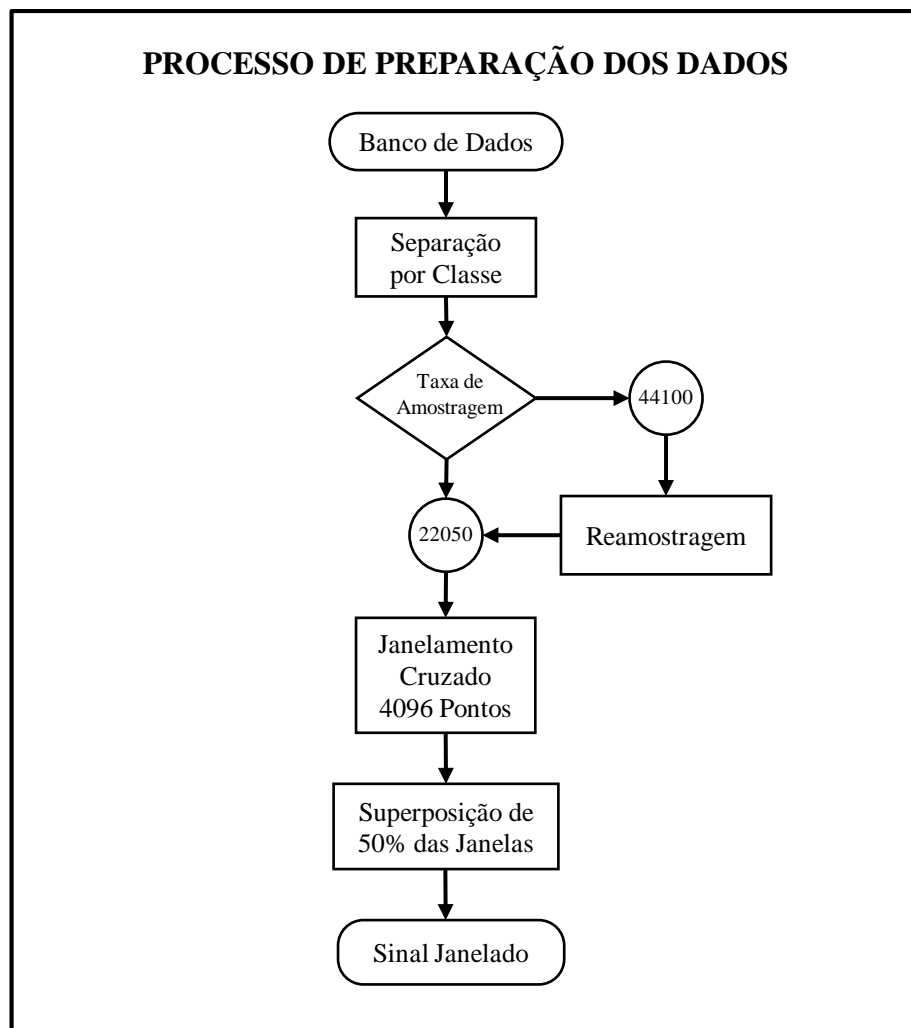
4.3 PROCEDIMENTOS

Essa seção tem como objetivo interligar todos os conceitos e métodos apresentados, auxiliando o entendimento por meio de processos individualizados e representações gráficas do fluxo de ideias que os compõe.

4.3.1 Transformação do banco de dados em sinais janelados

Processo responsável por carregar o banco de dados no MATLAB® e dentro de uma rotina verificar a necessidade de aplicar reamostragem para igualar as taxas de amostragens. Feito isso, são decididos o método de segmentação cruzado e o número de 4096 pontos por janela, totalizando um conjunto de 17004 janelas. Tal processo é exemplificado por meio da Figura 35.

Figura 35 – Fluxograma do processo de preparação dos dados.



Fonte: Autoria própria.

4.3.2 Transformação das janelas em amostras

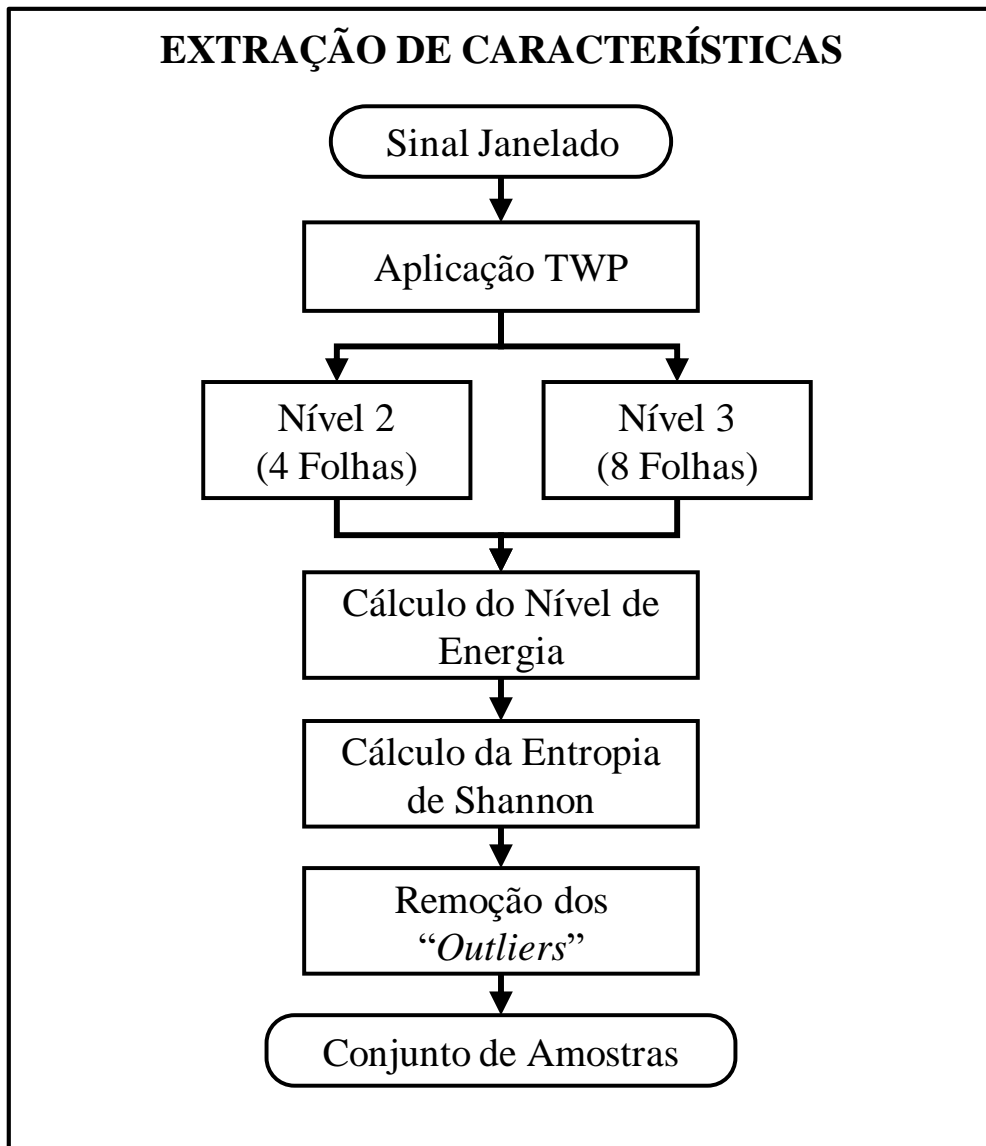
Este é o procedimento encarregado de receber as janelas do processo anterior e conseguir condicioná-las em entradas para as redes neurais. Composto basicamente pela aplicação da TWP por meio da *toolbox* do MATLAB[®], escolha do nível de resolução da decomposição e a extração das características utilizadas.

Com o intuito de garantir o melhor desempenho na adaptação das RNA, aplica-se nessa etapa uma análise estatísticas nos valores extraídos dos coeficientes, eliminando indivíduos discrepantes inclusos na população de amostras (*outliers*). Para ser considerado *outlier* dentro da metodologia proposta neste trabalho o valor deve estar três desvios-padrão distantes da média. Da mesma forma, é fundamental realizar a normalização dos valores em função de seus máximos e mínimos (valor máximo se torna 1, valor mínimo se torna 0 e os outros são ajustados proporcionalmente dentro desse intervalo) para que não ocorra problemas com as funções e limiares de ativação.

Será abordado o estudo separado de extração para o nível 2 e nível 3 de resolução na aplicação da TWP utilizando a família Daubechies 1. Tal proposta tem o intuito de determinar qual nível apresenta melhores condições para generalização da RNA. No primeiro caso, tem-se 4 folhas, obtendo 4 valores de energia e 4 valores de entropia, totalizando 8 entradas por janela na alimentação da RNA. Por sua vez, no segundo caso, tem-se 8 folhas, totalizando 16 entradas por janela para o classificador. Por meio da Figura 36, pode-se esclarecer a lógica das ideias envolvidas nesse processo.

Como critério de padronização e organização, adota-se o termo janela para referenciar os segmentos de áudio da voz e amostras os valores já extraídos das janelas, ou seja, as entradas prontas para serem apresentadas às RNA.

Figura 35 – Fluxograma de extração de características.



Fonte: Autoria própria.

4.3.3 Rotina de treinamento automatizado

Após garantir o sucesso dos procedimentos anteriores e ter mãos um grande conjunto de entradas, o objetivo seguinte é aplicar o método de treinamento das RNA. Deve-se atentar para cuidados antes de se iniciar o processo:

- I. **Embaralhar o conjunto de entradas:** visto que um único arquivo garante a geração de um extenso grupo de janelas (média de 130 janelas por arquivo), mesclar as entradas garante que todos os arquivos participaram tanto do treinamento quanto da validação.
- II. **Equilibrar o número de amostras:** como o banco de dados não é homogêneo em relação a duração da gravação por arquivo e da quantidade de arquivos por categoria

de classificação, é necessário equilibrar na etapa de treinamento o número de amostras de forma que a RNA não fique condicionada para um único tipo de classe. Portanto, foi registrado qual a menor quantidade de amostras por categoria e foi removido os excedentes das outras categorias, garantindo o equilíbrio (8502 amostras saudáveis e 8502 amostras com Edema de Reinke).

- III. **Separar as entradas em dois grupos:** o grupo de padrões para treinamentos e o grupo para validações. Considerando a grande quantidade de entradas que serão criadas, será utilizado a proporção de 70% do total para treinamento (cerca de 11900 amostras) e os 30% restantes para validação (cerca de 5100 amostras).
- IV. **Fixar os valores máximos e mínimos:** inserindo todas as amostras que contenham um valor máximo ou mínimo das características analisadas dentro do total de treinamento certifica-se que a rede será ajustada nos limites do universo das amostras, aperfeiçoando a generalização na etapa de validação.

Partindo dos princípios acima, pode-se iniciar a rotina automatizada de treinamento com a principal função testar o extenso número de possíveis configurações para o sistema. Foi realizado no mínimo dez treinamentos antes de alterar a topologia, pois considerando a inicialização aleatória dos pesos sinápticos e a existência de mínimos locais na função erro a ser minimizada pelo algoritmo de treinamento, efetuar vários treinamentos na mesma topologia evita descartar uma configuração promissora devido ao confinamento em um mínimo local da função.

É essencial o armazenamento de gráficos, histogramas e dados referentes a cada repetição, possibilitando comparar o melhor desempenho no final do processo. As informações detalhas da RNA aplicada no estudo estão descritas no Quadro 6. Por sua vez, o Quadro 7 contém as informações das variações de topologia que serão realizadas, alcançando um total de 600 treinamentos para cada conjunto (nível 2 e nível 3). Por fim, a Figura 37 permite compreender o processo automatizado responsável por determinar a melhor topologia.

Quadro 6 – Características da RNA desenvolvida.

Arquitetura da rede	Perceptron Multicamadas
Tipo de treinamento	Supervisionado – <i>Off-line</i>
Algoritmo de treinamento	Levenberg-Marquardt (<i>backpropagation</i>)
Taxa de aprendizagem	10^{-2}
Critério de parada (Erro)	10^{-4}
Épocas	2000
Proporção Treinamento / Validação	70% / 30%
Número total de Amostras	17004

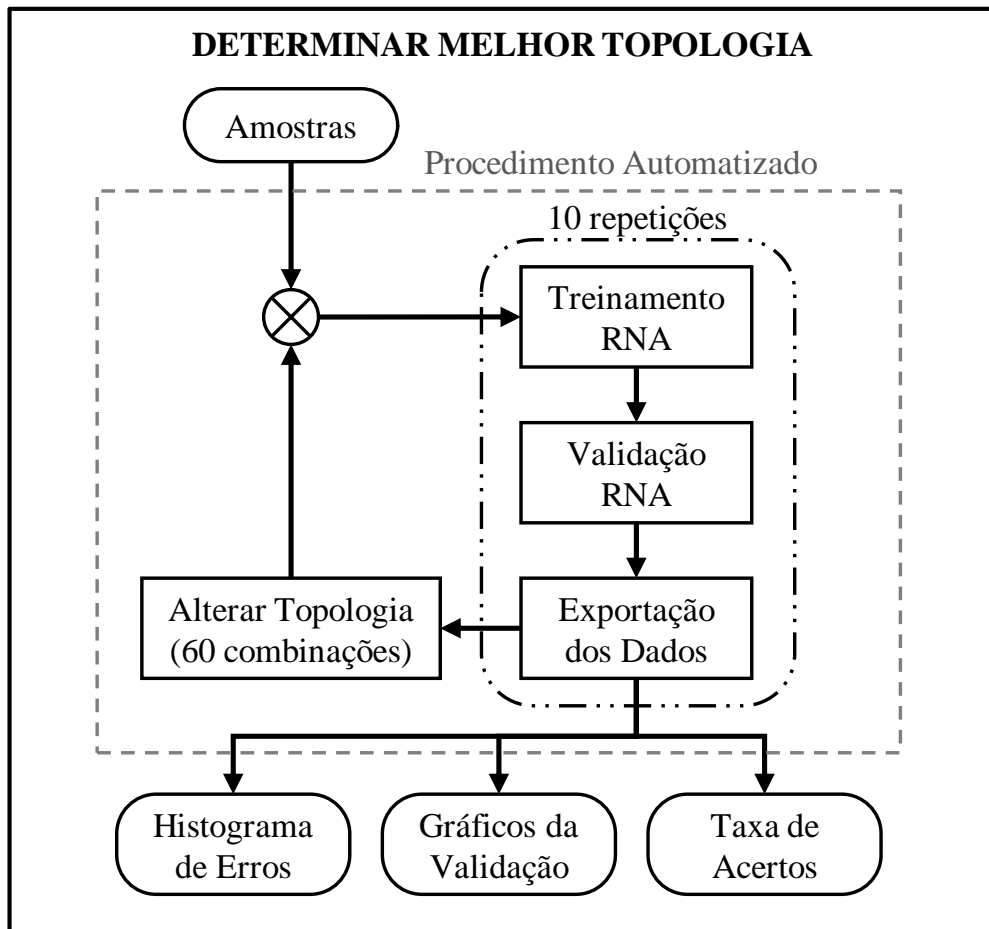
Fonte: Autoria própria.

Quadro 7 – Variações de topologia.

Neurônios (primeira camada)	11, 13, 15, 17, 19, 21
Neurônios (segunda camada)	3, 5, 7, 9, 11, 13, 15, 17, 19, 21
Variações de topologias	60 combinações
Treinamentos por topologia	10 treinamentos
Total de treinamentos	600 treinamentos

Fonte: Autoria própria.

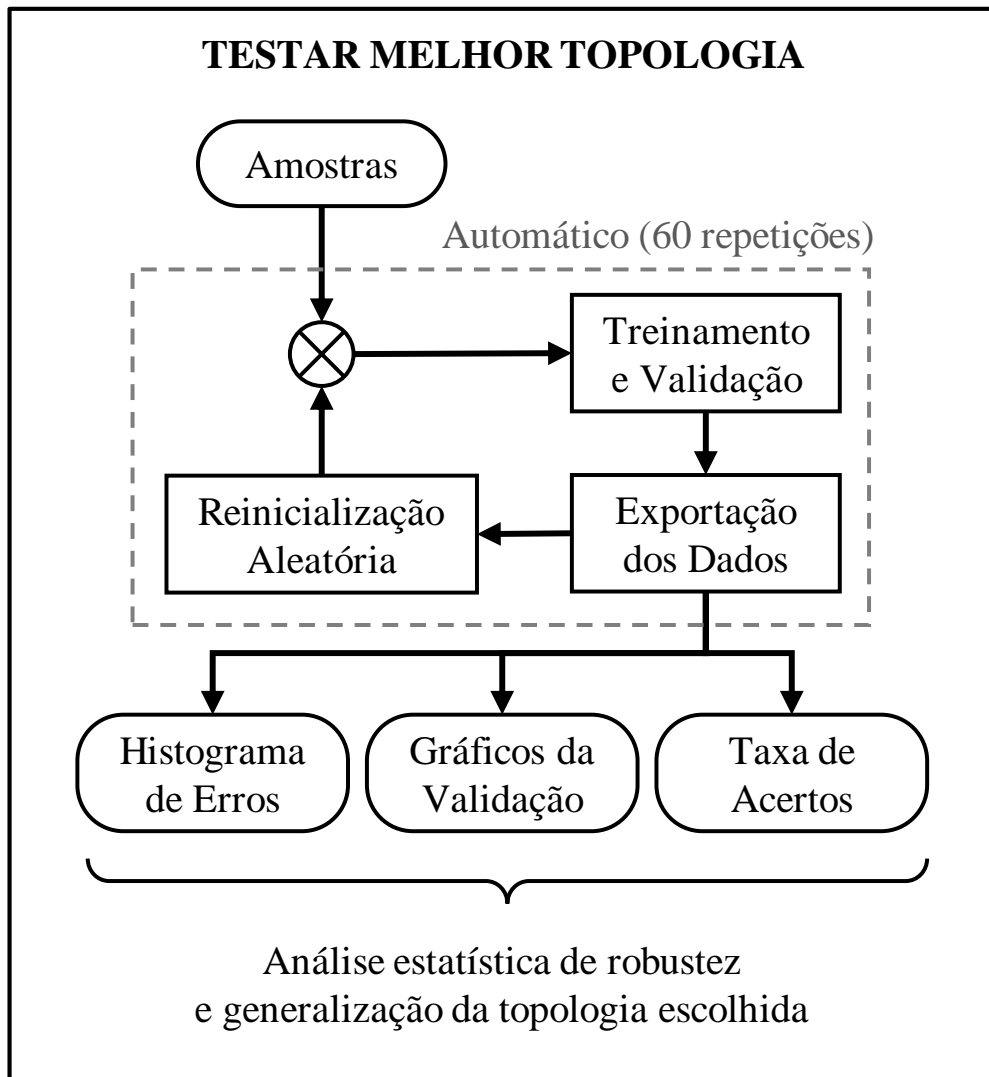
Figura 37 – Fluxograma do processo automatizado de treinamento das RNA.



Fonte: Autoria própria.

Com o objetivo de avaliar a robustez e capacidade de generalização, serão realizadas mais 60 repetições sobre a topologia com melhor desempenho, reiniciando o processo de treinamento e validação com os parâmetros livres da rede gerados aleatoriamente, implicando uma confiabilidade estatística em sua eficiência. Tal processo é representado pela Figura 38.

Figura 38 – Fluxograma de análise estatística da robustez das RNA.



Fonte: Autoria própria.

5 RESULTADOS E DISCUSSÕES

Este capítulo inicia-se pela discussão da capacidade de generalização das RNA mediante a variação no nível de resolução da decomposição da TWP, bem como, a escolha da melhor topologia.

Em sequência, será abordado alguns resultados para os testes de *underfitting*³ elaborados em topologias do terceiro nível de resolução, conjuntamente com os testes de robustez da topologia com os melhores resultados.

Serão analisados também os critérios de pós-processamento nos valores coletados e a discussão do significado dos resultados na validação. Feito isso, será construída a matriz de confusão (ou matriz de erro), envolvendo de maneira mais clara algumas estatísticas relacionadas ao comportamento da classificação realizada.

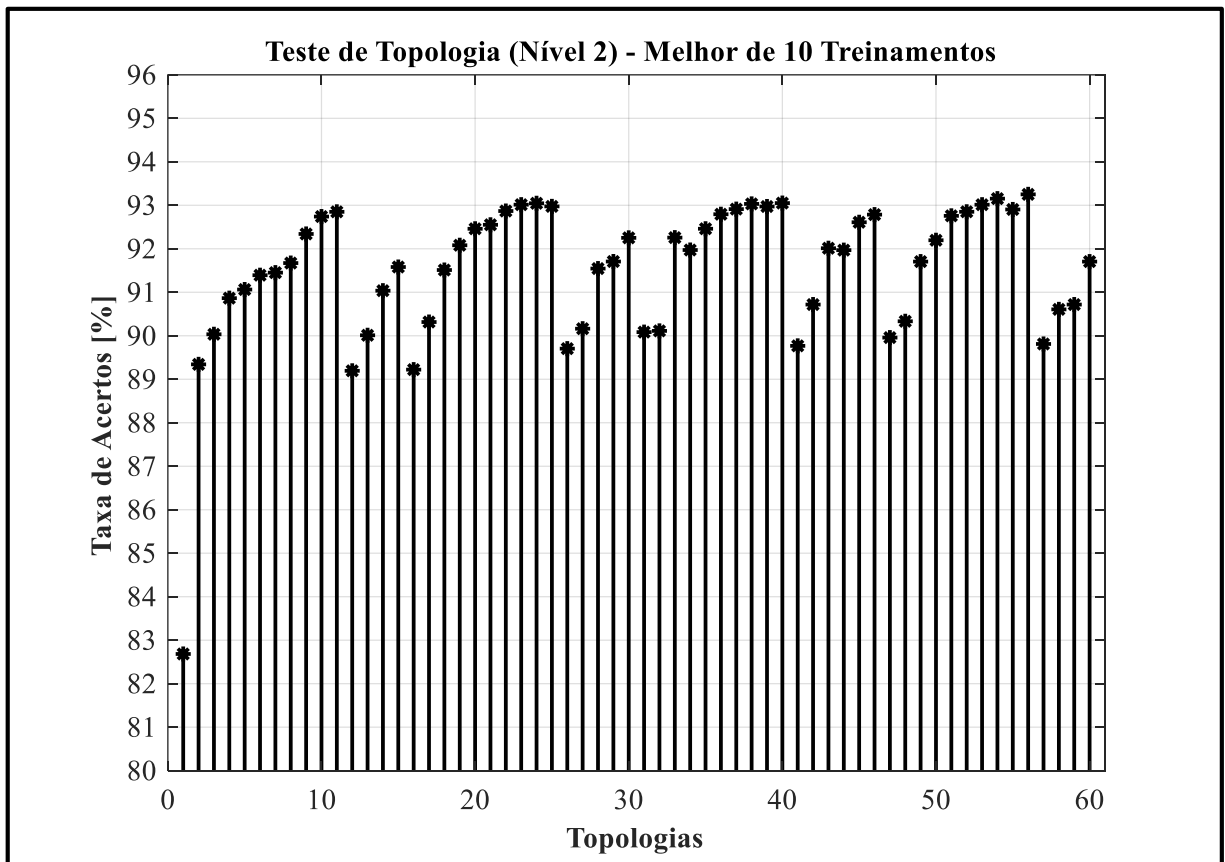
Por fim, será demonstrada a resposta da RNA para amostras desconhecidas tanto do processo de treinamento quanto ao processo de validação. Desse modo é possível realizar uma tentativa de simulação da operação do classificador em um ambiente real dos profissionais da VOZ.

5.1 RESULTADOS DA VARIAÇÃO DE TOPOLOGIAS NO NÍVEL 2

A primeira parcela de resultados obtidos é referente aos testes realizados no segundo nível de resolução, visando determinar a melhor topologia. Realizaram-se 10 treinamentos para cada uma das 60 variações de topologia, registrando-se apenas o treinamento que apresentou melhor taxa de acertos na etapa da validação. Tais resultados são representados graficamente por intermédio da Figura 39.

³ Testes de *underfitting*: testes de topologias com elevado número de neurônios artificiais para garantir que possíveis problemas de generalização não estão relacionados com a escolha de poucos neurônios por camada.

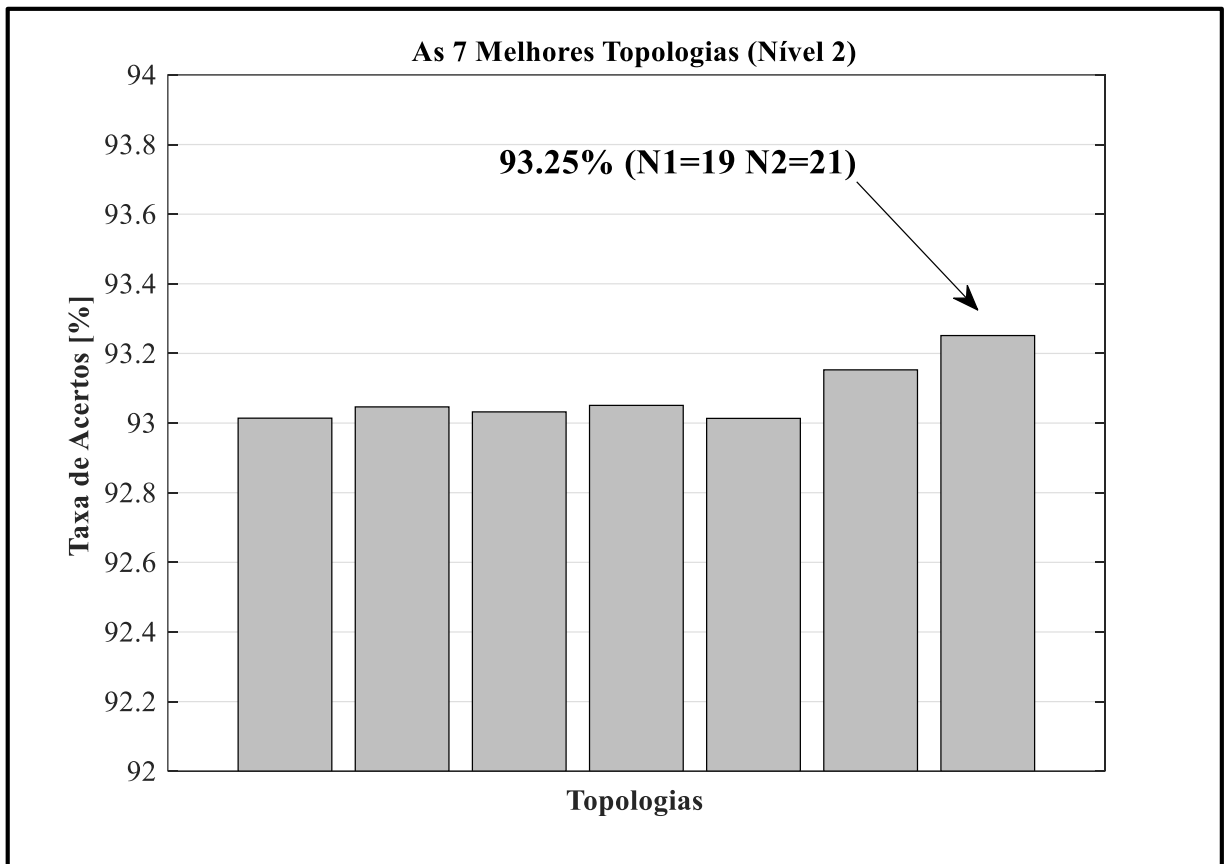
Figura 39 – Testes de topologia (nível 2) – melhor de 10 treinamentos.



Fonte: Autoria própria.

Com o intuito de facilitar a análise, filtrou-se o resultado das 7 melhores topologias em outra representação gráfica, podendo ser observado com o auxílio da Figura 40. É possível notar que, ao final desta primeira etapa de testes, tem-se como melhor topologia 19 neurônios artificiais na primeira camada escondida e 21 neurônios na segunda camada escondida, apresentando uma eficiência de 93,25% de acertos na etapa de validação.

Figura 40 – Resultado de acertos das 7 melhores topologias (nível 2).

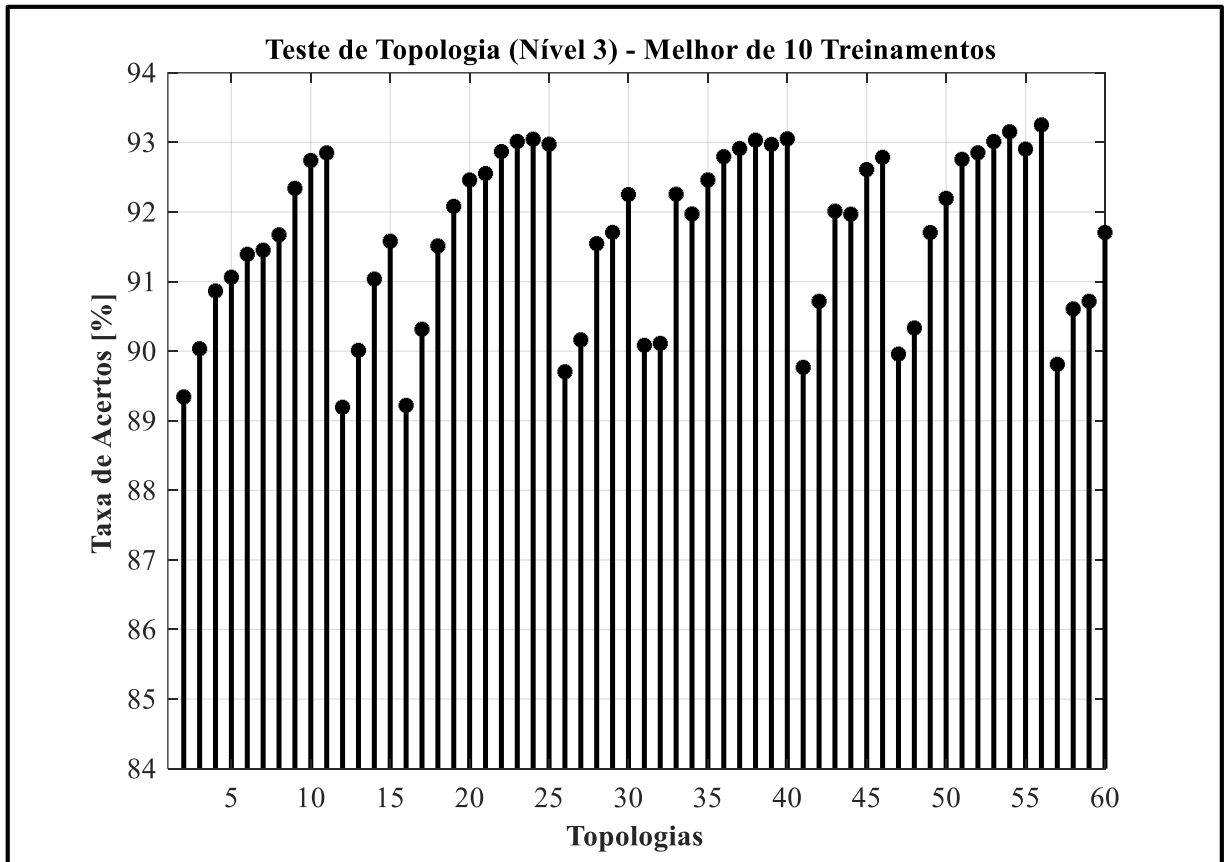


Fonte: Autoria própria.

5.2 RESULTADOS DA VARIAÇÃO DE TOPOLOGIAS NO NÍVEL 3

A segunda parcela de resultados consiste nos testes realizados com o conjunto de amostras referente ao terceiro nível de resolução. Seguindo o mesmo padrão da primeira etapa, realizaram-se 10 treinamentos para cada uma das 60 variações de topologia, registrando-se apenas o treinamento que apresentou melhor taxa de acertos na etapa da validação. Tais resultados são representados graficamente na Figura 41.

Figura 41 – Testes de topologia (nível 3) – melhor de 10 treinamentos.

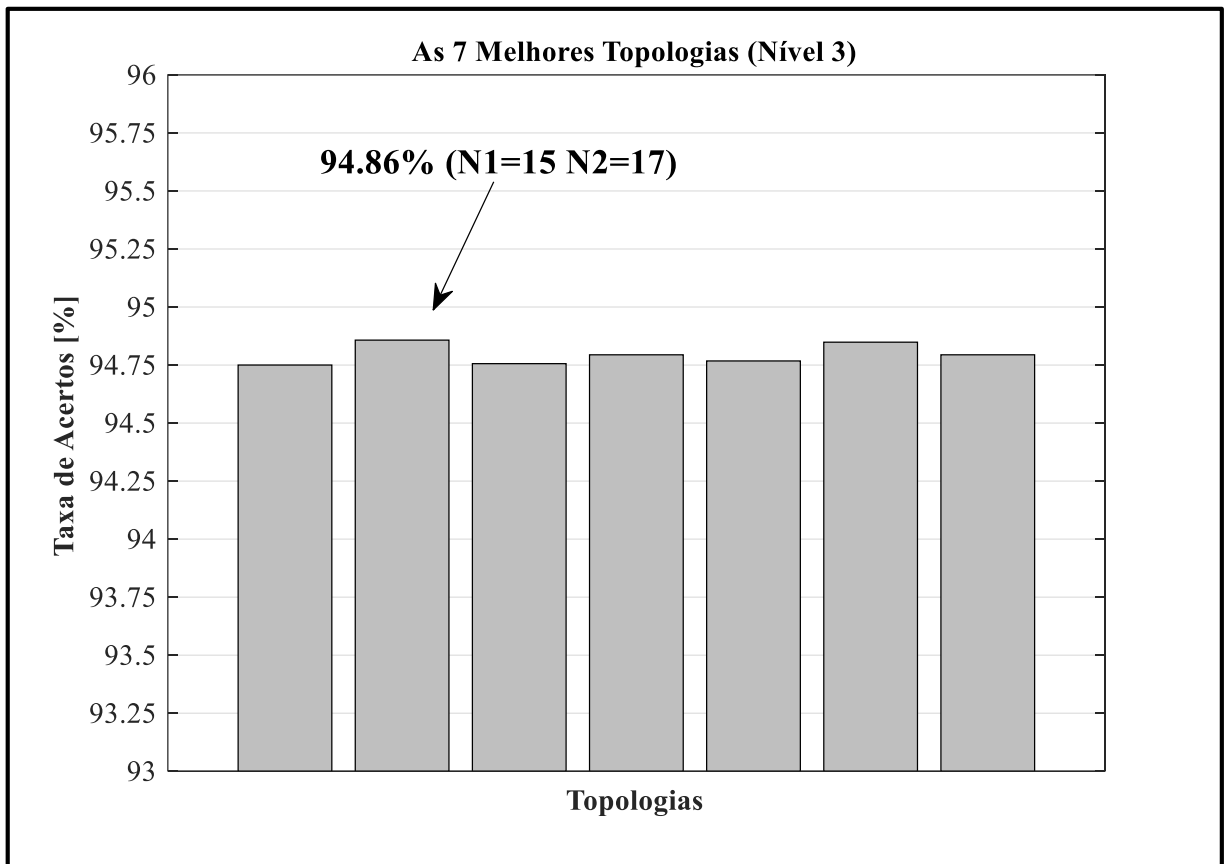


Fonte: Autoria própria.

Novamente, selecionou-se o resultado das 7 melhores topologias em outro arranjo gráfico, podendo ser observado com o auxílio da Figura 42.

Contudo, com as análises das duas etapas de testes é possível notar um ganho de acertos ao utilizar o conjunto de amostras do terceiro nível, indicando que a topologia de 15 neurônios artificiais na primeira camada escondida e 17 neurônios na segunda camada, atingiu um grau de generalização superior, apresentando uma eficiência de 94,86% de acertos na etapa de validação.

Figura 42 – Resultado de acertos das 7 melhores topologias (nível 3).



Fonte: Autoria própria.

5.3 ANÁLISE DE *UNDERFITTING* NAS TOPOLOGIAS DO NÍVEL 3

Tendo em mãos as observações referentes à melhor topologia até o momento, é necessário realizar uma verificação fundamentada na hipótese de erros presentes na validação serem decorrentes de *underfitting* nas camadas escondidas. Em outras palavras, é preciso garantir que os erros não sejam consequência da escolha de poucos neurônios nas camadas, se tornando uma modelagem incapaz de atingir um maior nível de generalização.

Para tal verificação, elaborou-se uma nova rotina automatizada de testes envolvendo variações de topologias com elevados números de neurônios, servindo como critério de comparação para os resultados da etapa anterior.

Contudo, levando-se em conta o grande acréscimo de esforço computacional, reduziu-se para 3 o número de treinamentos dentro da mesma topologia. Portanto, a nova rotina consiste de mais 84 treinamentos (28 combinações de topologias x 3 treinamentos por topologia). Tais variações podem ser observadas com o auxílio do Quadro 8.

Quadro 8 – Taxas de acertos (*underfitting* nível 3) – melhor de 3 treinamentos.

Taxas de Acertos (<i>Underfitting</i> Nível 3) – Melhor de 3 Treinamentos		
Neurônios 1ª Camada	Neurônios 2ª Camada	Taxa de Acertos [%]
20	20	93.08250
20	25	92.17958
20	30	93.41083
20	35	92.83625
25	20	93.10986
25	25	92.45319
25	30	92.72681
25	35	92.75417
30	20	93.02778
30	25	93.00042
30	30	92.42583
30	35	93.00042
35	20	93.02778
35	25	92.50792
35	30	92.86361
35	35	92.91833
45	20	94.05958
45	25	94.44571
45	30	94.05958
45	35	94.31056
55	20	94.56154
55	25	94.75461
55	30	93.94375
55	35	94.63877
65	20	93.36456
65	25	94.56154
65	30	94.15612
65	35	94.61946

Fonte: Autoria própria.

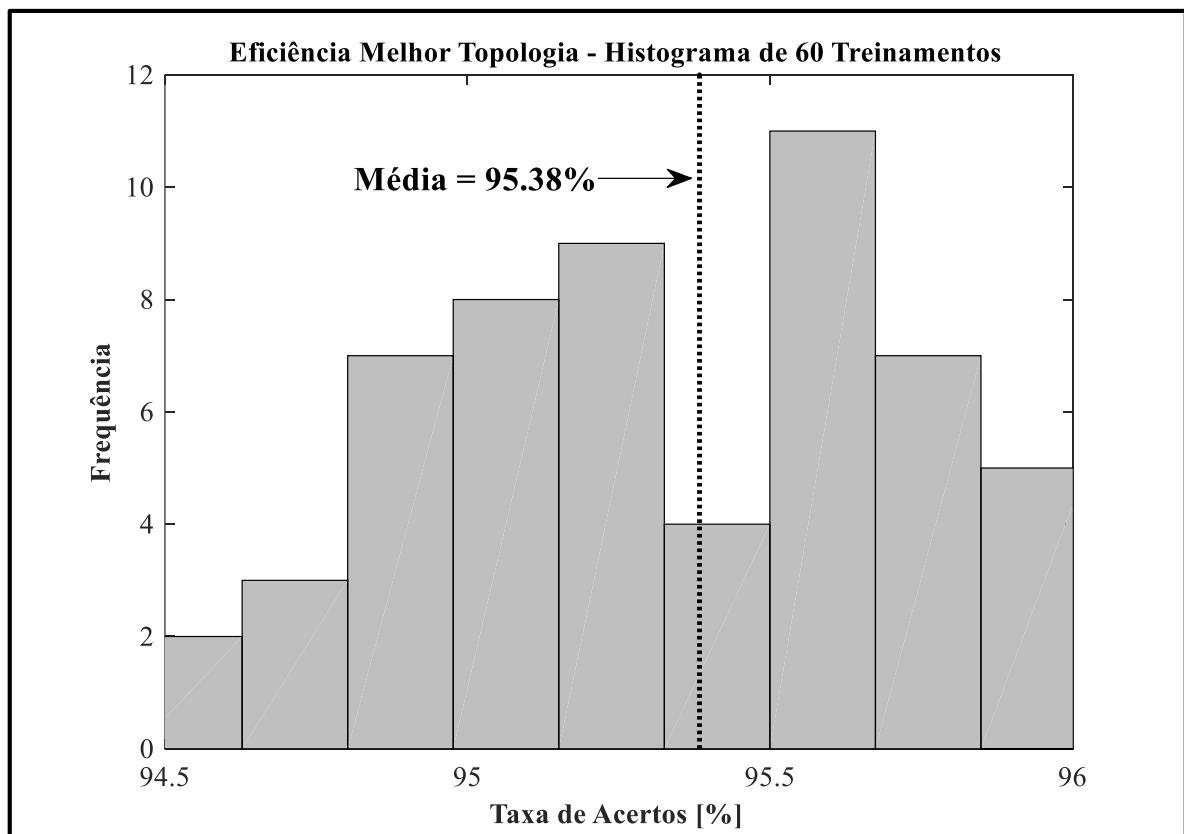
Feito isso, pode-se observar que mesmo elevando a composição de neurônios das camadas não houve uma melhoria alguma na taxa de acertos na validação. Portanto, é possível afirmar que os erros encontrados não são originados pelo fator de *underfitting*.

Nota-se, também, que a topologia com 55 neurônios na primeira camada e 25 neurônios na segunda camada atingiu resultados próximos à melhor topologia registrada anteriormente. Contudo, este resultado é descartado das análises nas próximas etapas devido ao extremo incremento do esforço computacional para se trabalhar com tal topologia.

5.4 VERIFICAÇÃO DA ROBUSTEZ DA MELHOR TOPOLOGIA

Nesta etapa inicia-se a quarta rotina automatizada de treinamento. Dessa vez com o intuito de verificar a consistência da taxa de acertos encontrada na melhor topologia. Portanto, foi realizado mais 60 treinamentos mantendo fixa a configuração de neurônios do melhor resultado ($N1=15$ e $N2=17$), tendo como variação apenas os parâmetros livres (pesos sinápticos e limiares de ativação) que são iniciados aleatoriamente em cada treinamento. Tal resultado pode ser observado por meio da Figura 45.

Figura 45 – Histograma dos 60 treinamentos adicionais da melhor topologia.



Fonte: Autoria própria.

Analisando a Figura 45, pode-se afirmar que os resultados se mantiveram consistentes durante a rotina automatizada de treinamentos, obtendo uma média de 95,38% de taxa de acertos. Logo, nota-se que a capacidade de generalização da RNA para as amostras foi robusta e confiável.

É importante ressaltar que, mesmo obtendo uma média de 95,38% de taxa de acertos, o melhor resultado de treinamento obtido apresentou 96,20% de taxa de acertos. Desta forma, foi registrado as configurações de parâmetros livres para que possa ser utilizado como referência nos demais testes nas seções seguintes, em outras palavras, é adotado como o melhor treinamento da melhor topologia.

5.5 PÓS-PROCESSAMENTO DOS RESULTADOS OBTIDOS

Em sequência, o próximo passo da análise dos resultados é realizar o pós-processamento das classificações obtidas na configuração da melhor topologia definida anteriormente (N1=15 e N2= 17).

O pós-processamento consiste em transformar a resposta da validação em valores discretos que definam as classes desejadas. Desse modo, simplifica-se a discussão dos resultados e permite, também, estruturar uma margem de confiabilidade.

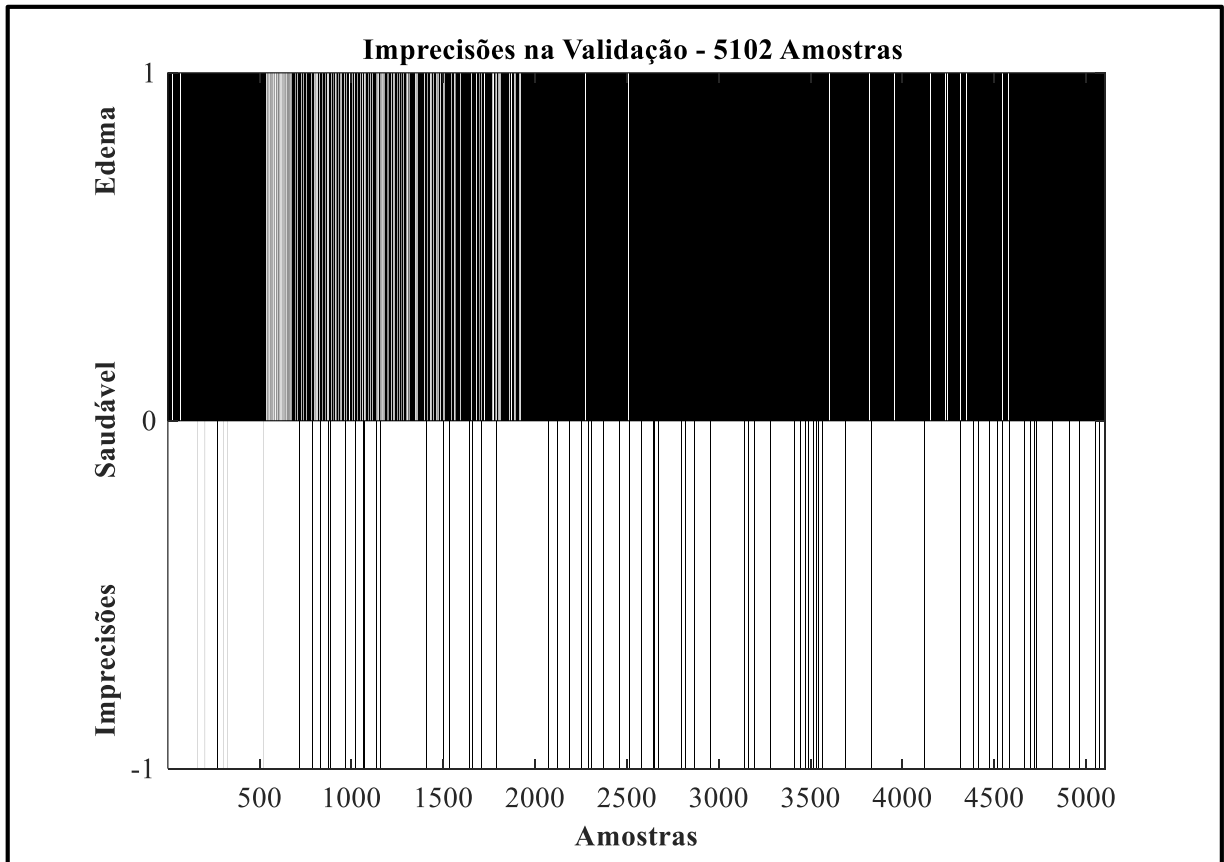
Os critérios adotados para o processamento estão representados no Quadro 9, caracterizando uma margem de confiabilidade de 95% na classificação e tratando qualquer valor encontrado no intervalo entre as margens como uma imprecisão. Pode-se observar a representação das imprecisões encontradas por meio da Figura 46.

Quadro 9 – Critérios de pós-processamento.

Valor da saída	Valor adotado	Significado
$0,00 \leq X \leq 0,05$	0	Saudável
$0,05 < X < 0,95$	-1	Imprecisão
$0,95 \leq X \leq 1,00$	1	Edema

Fonte: A autoria própria.

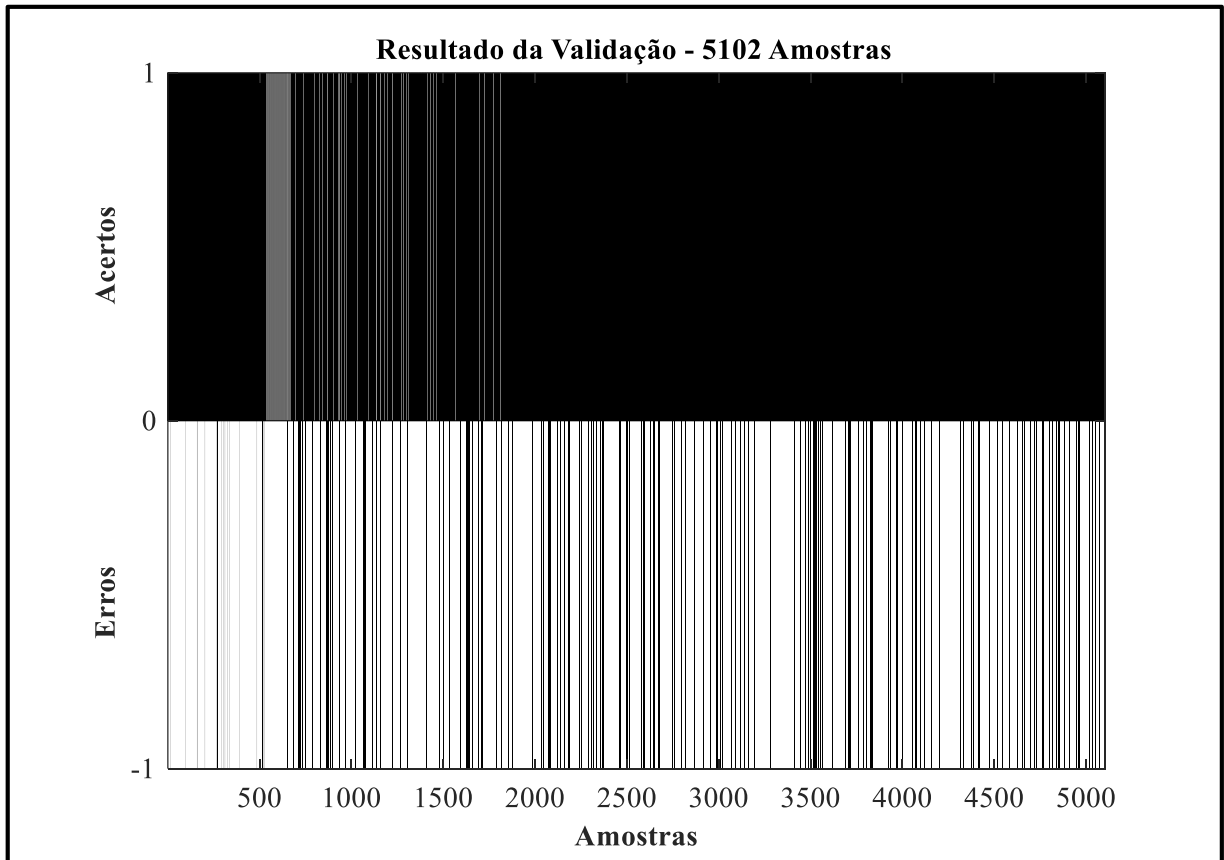
Figura 46 – Imprecisões na validação da melhor topologia (5102 amostras).



Fonte: Autoria própria.

Por meio do resultado do pós-processamento, pode-se também representar de maneira visual o desempenho da RNA para todo o conjunto de amostras de validação. Tal representação é fundamentada na comparação dos valores obtidos na classificação com os valores reais das entradas (a partir deste ponto todos os casos de imprecisão são tratados como erros cometidos pela rede). Dessa forma, os acertos foram representados com o valor 1 e os erros com o valor -1, facilmente observados na Figura 47.

Figura 47 – Resultados da validação da melhor topologia (5102 amostras).



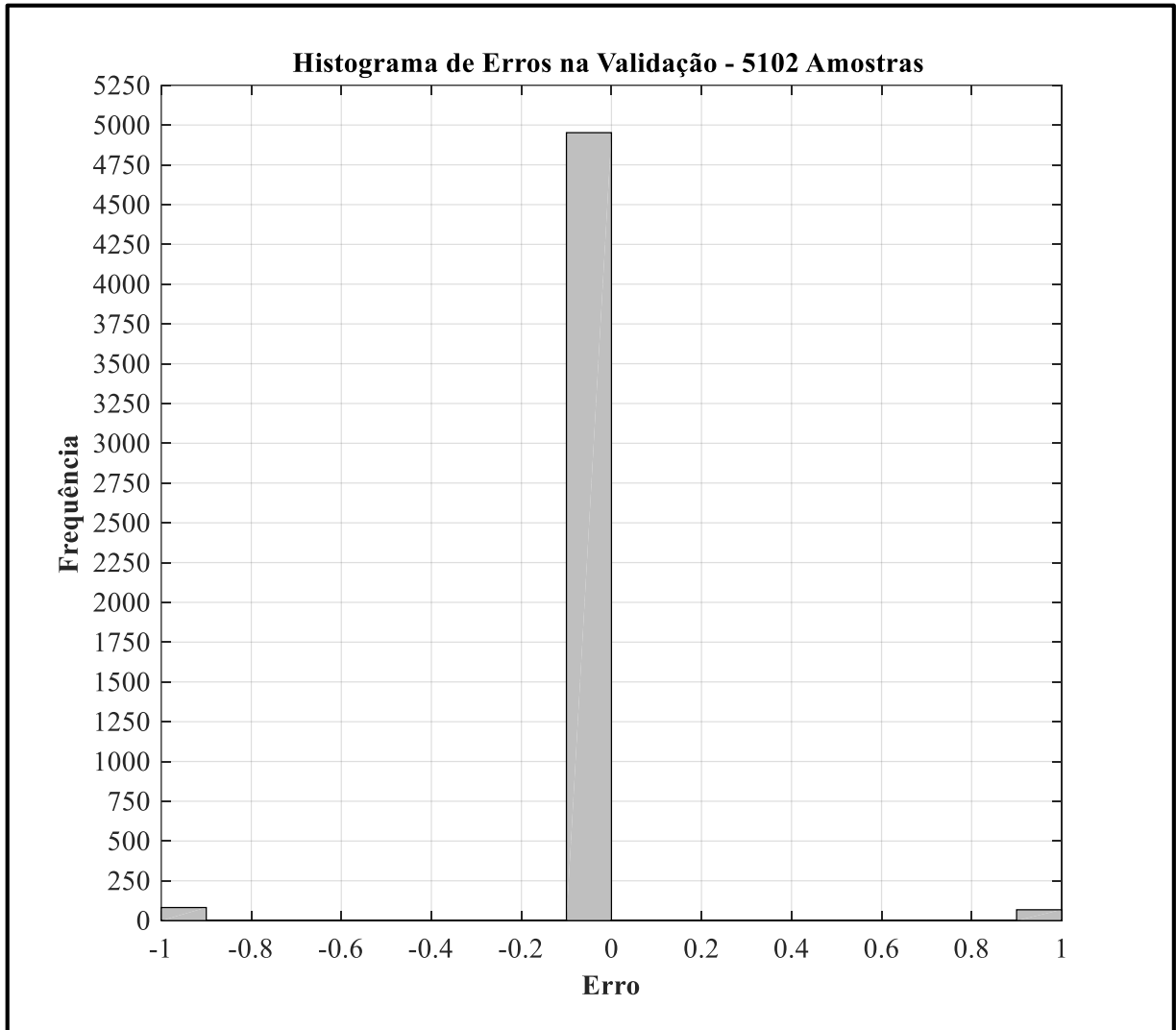
Fonte: Autoria própria.

Por fim, tem-se a representação dos tipos de erros cometidos pela rede. Tal observação é estruturada na subtração dos valores obtidos no pós-processamento pelos valores desejados. Partindo do princípio que tanto os valores desejados como os valores pós-processados têm suas classes definidas por valores discretos (0 para a classe saudável e 1 para a classe edema), tem-se as seguintes combinações de resultados:

- **Valor do erro igual a 0:** resultantes das subtrações de 1 por 1 ou de 0 por 0, indicando que os valores obtidos e desejado são idênticos, ou seja, é confirmado um caso de acerto na classificação;
- **Valor do erro igual a 1:** resultantes da subtração de 0 por 1, indicando amostras saudáveis classificadas como edema, ou seja, é confirmado um caso de erro na classificação;
- **Valor do erro igual a -1:** resultantes da subtração de 1 por 0, indicando amostras de edema classificadas como saudáveis, ou seja, é confirmado um caso de erro na classificação.

Dessa forma, o resultado da análise dos tipos de erros cometidos pela rede seguindo os conceitos explicados pode ser visualizado na Figura 48.

Figura 48 – Histograma de erro na validação da melhor topologia (5102 amostras).



Fonte: Autoria própria.

5.6 MATRIZ DE CONFUSÃO

Esta seção é elaborada com o objetivo de reunir todas as análises estatísticas realizadas anteriormente. Portanto, construiu-se uma matriz de confusão (ou matriz de erro) contendo todas as informações de maneira estruturada. Tal representação pode ser visualizada por intermédio do Quadro 10.

Quadro 10 – Matriz de confusão.

		Condição Classificada (CC)		
		Classificado Saudável (CS)	Classificado Edema (CE)	Imprecisão (IP)
População Total (PT)		2491	2536	75
Condição Verdadeira (CV)	Real Saudável (RS)	Verdadeiro Saudável (VS)	Falso Edema (FE)	Taxa de Falso Saudável (TFS)
	Real Edema (RE)	Falso Saudável (FS)	Verdadeiro Edema (VE)	Taxa de Falso Edema (TFE)
Acurácia (AC)		Taxa de Verdadeiro Saudável (TVS)	Taxa de Verdadeiro Edema (TVE)	

Fonte: Autoria própria.

Onde tem-se que,

- **Condição classificada (CC):** representa as respostas obtidas na etapa de validação da melhor RNA;
- **Classificado saudável (CS):** representa o total de classificações do tipo saudável;
- **Classificado edema (CE):** representa o total de classificações do tipo edema;
- **Condição verdadeira (CV):** representa os valores reais desejados na etapa de validação da melhor RNA;
- **Real saudável (RS):** representa o total de valores reais do tipo saudável;
- **Real edema (RE):** representa o total de valores reais do tipo edema;
- **População total (PT):** representa o valor total de amostras utilizadas na validação da rede;
- **Verdadeiro saudável (VS):** representa os valores reais saudáveis que foram classificados como saudável;

- **Verdadeiro edema (VE):** representa os valores reais de edema que foram classificados como edema;
- **Falso saudável (FS):** representa os valores reais de edema que foram classificados como saudáveis;
- **Falso edema (FE):** representa os valores reais saudáveis que foram classificados como edema;
- **Acurácia (AC):** representa o total de acertos na classificação em relação a população total. A acurácia é obtida por meio da Equação (6):

$$AC = \frac{VS + VE}{PT} \times 100\% \quad (6)$$

- **Taxa de verdadeiro saudável (TVS):** representa a razão de classificações do tipo verdadeiro saudável em relação a todas as classificações do tipo saudável, ou seja, a taxa de acertos na classificação do tipo saudável. A taxa é obtida por meio da Equação (7):

$$TVS = \frac{VS}{CS} \times 100\% \quad (7)$$

- **Taxa de verdadeiro edema (TVE):** representa a razão de classificações do tipo verdadeiro edema em relação a todas as classificações do tipo edema, ou seja, a taxa de acertos na classificação do tipo edema. A taxa é obtida por meio da Equação (8):

$$TVE = \frac{VE}{CE} \times 100\% \quad (8)$$

- **Taxa de falso saudável (TFS):** representa o total de classificações do tipo falso saudável em relação a todas as classificações do tipo saudável, ou seja, a taxa de erros na classificação do tipo saudável. A taxa é obtida por meio da Equação (9):

$$TFS = \frac{FS}{CS} \times 100\% \quad (9)$$

- **Taxa de falso edema (TFE):** representa o total de estimações do tipo falso edema em relação a todas as classificações do tipo edema, ou seja, a taxa de erros na classificação do tipo edema. A taxa é obtida por meio da Equação (10):

$$TFE = \frac{FE}{CE} \times 100\% \quad (10)$$

- **Imprecisões (IP):** representa os valores de estimação superiores a 0,05 e inferiores a 0,95 que foram considerados imprecisos nos critérios de pós-processamento. O valor é obtido por meio da Equação (11):

$$IP = PT - VS - VE - FS - FE \quad (11)$$

5.7 ANÁLISE DE RESPOSTA DA RNA PARA AMOSTRAS DESCONHECIDAS

Por fim, a última etapa da análise dos resultados tem como objetivo a tentativa de simular o comportamento do classificador desenvolvido em um ambiente real de operação. Portanto, foram selecionados três arquivos de cada classe para não participarem de nenhum processo de treinamento ou validação até o momento, ou seja, arquivos que serão tratados como amostras totalmente desconhecidas e externas ao banco de dados das RNA.

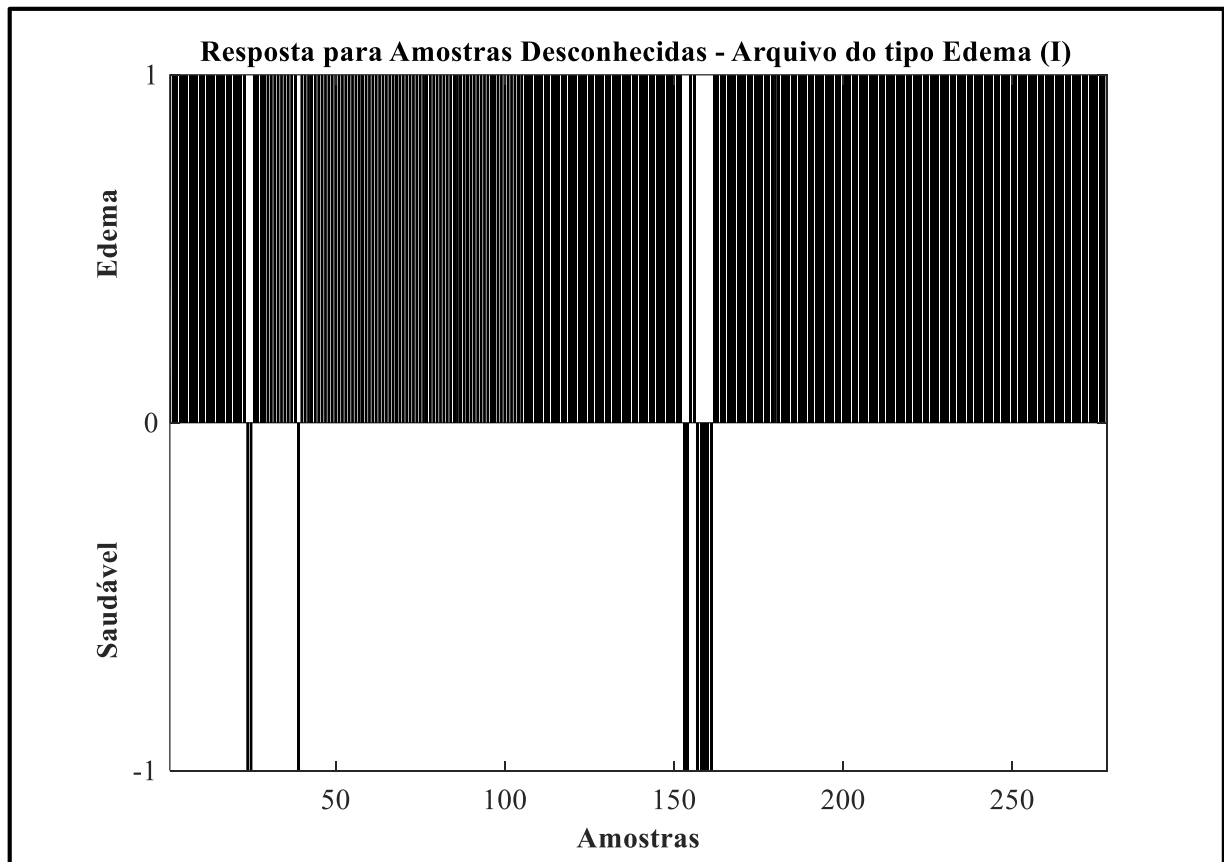
Contudo, para garantir consistência das respostas da simulação, os arquivos passaram pelo mesmo processo de preparação da taxa de amostragem, janelamento, extração de características e parâmetros de normalização.

Os resultados obtidos serão apresentados nas seções seguintes.

5.7.1 Simulação de operação – arquivo do tipo edema (I)

Os resultados referentes a resposta do classificador para a simulação do primeiro arquivo do tipo edema estão representados na Figura 49 e no Quadro 11.

Figura 49 – Resposta para amostras desconhecidas – arquivo do tipo edema (I).



Fonte: Autoria própria.

Quadro 11 – Resultados – arquivo do tipo edema (I).

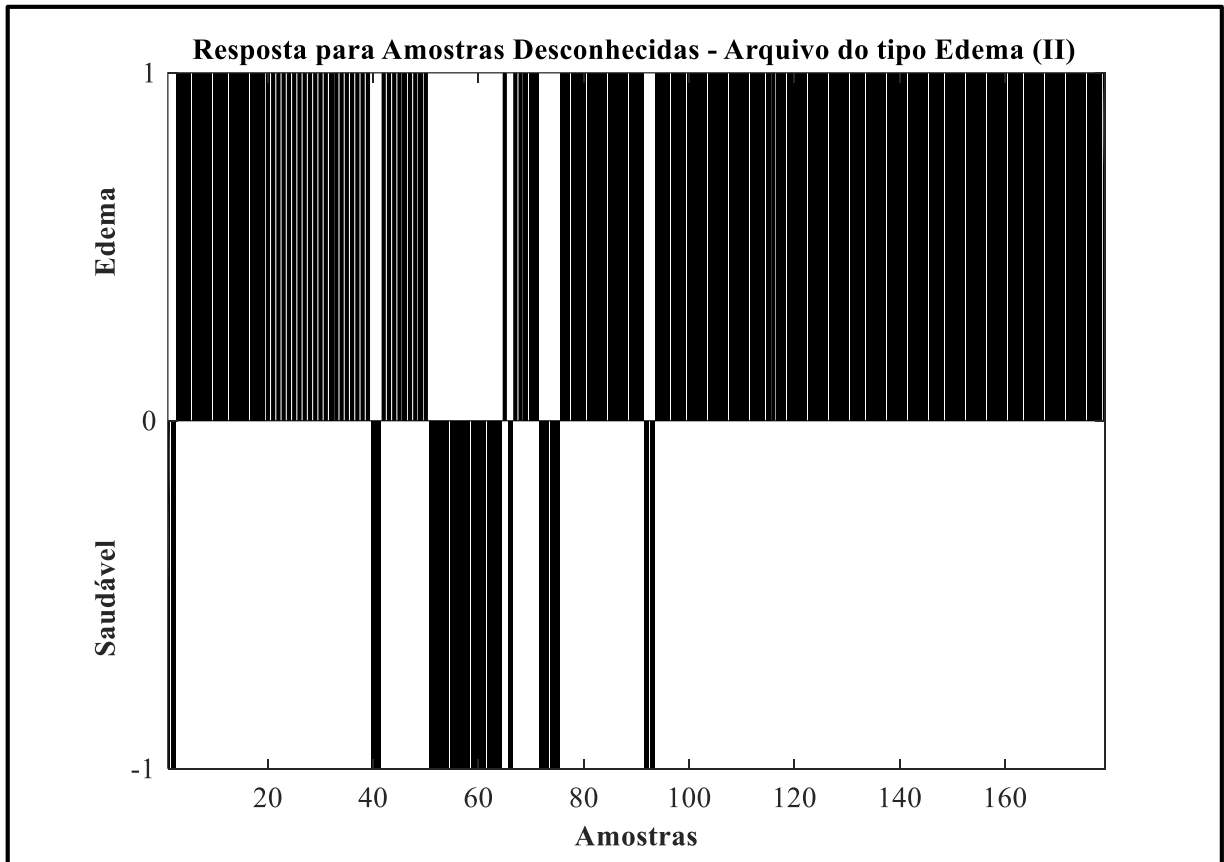
Resultados – arquivo do tipo edema (I)	
Total de amostras	278
Amostras do tipo edema	268
Amostras do tipo saudáveis	10
Amostras indeterminadas	0
Porcentagem de amostras com edema	96,40%

Fonte: Autoria própria.

5.7.2 Simulação de operação – arquivo do tipo edema (II)

Os resultados referentes a resposta do classificador para a simulação do segundo arquivo do tipo edema estão representados na Figura 50 e no Quadro 12.

Figura 50 – Resposta para amostras desconhecidas – arquivo do tipo edema (II).



Fonte: Autoria própria.

Quadro 12 – Resultados – arquivo do tipo edema (II).

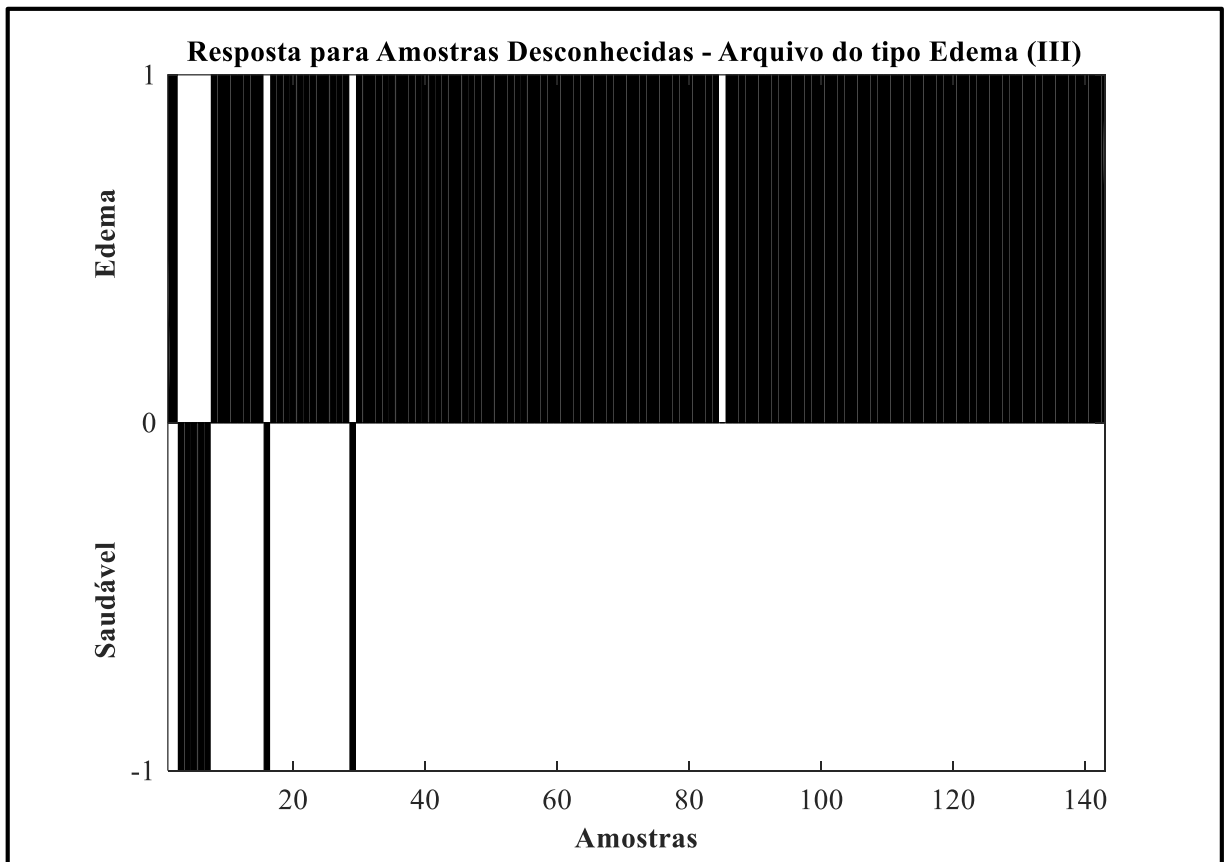
Resultados – arquivo do tipo edema (II)	
Total de amostras	179
Amostras do tipo edema	154
Amostras do tipo saudáveis	25
Amostras indeterminadas	0
Porcentagem de amostras com edema	86,03%

Fonte: Autoria própria.

5.7.3 Simulação de operação – arquivo do tipo edema (III)

Os resultados referentes a resposta do classificador para a simulação do terceiro arquivo do tipo edema estão representados na Figura 51 e no Quadro 13.

Figura 51 – Resposta para amostras desconhecidas – arquivo do tipo edema (III).



Fonte: Autoria própria.

Quadro 13 – Resultados – arquivo do tipo edema (III).

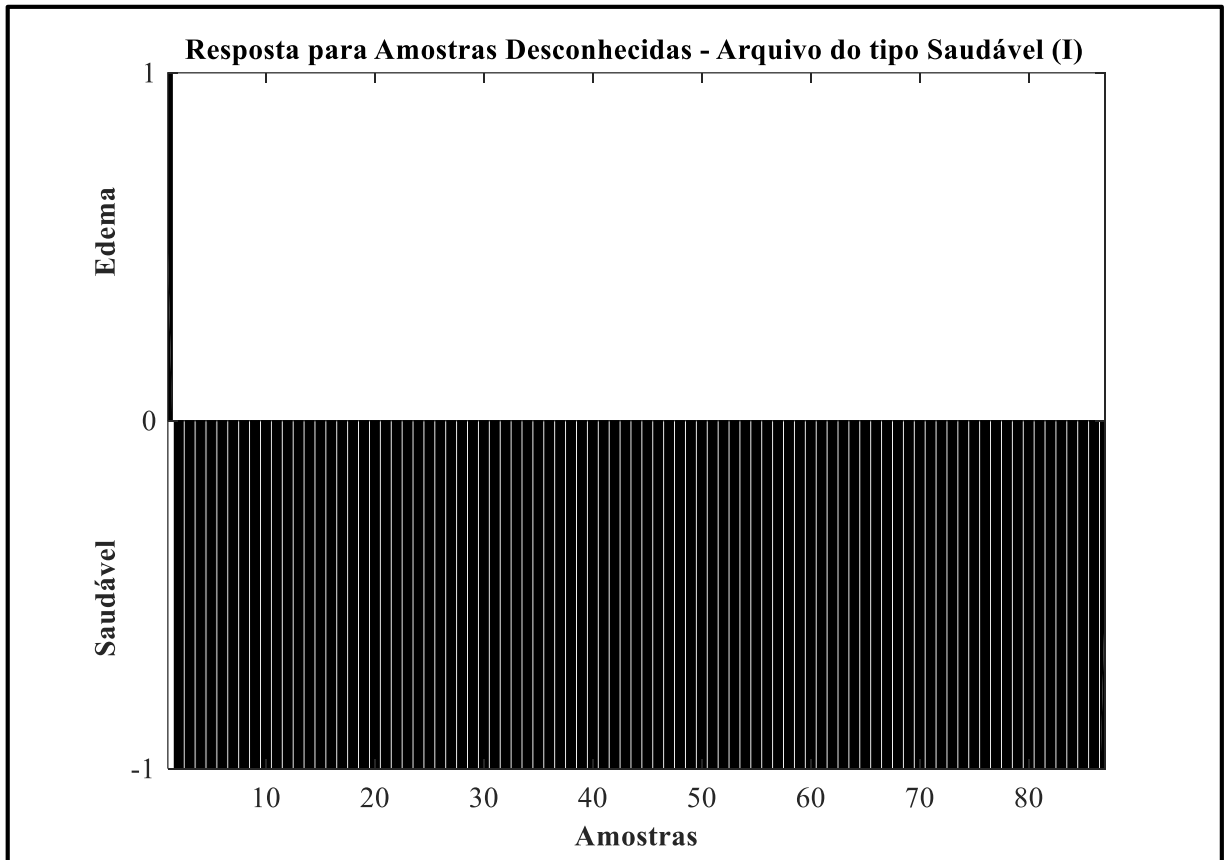
Resultados – arquivo do tipo edema (III)	
Total de amostras	143
Amostras do tipo edema	135
Amostras do tipo saudáveis	7
Amostras indeterminadas	1
Porcentagem de amostras com edema	94,41%

Fonte: Autoria própria.

5.7.4 Simulação de operação – arquivo do tipo saudável (I)

Os resultados referentes à resposta do classificador para a simulação do primeiro arquivo do tipo saudável estão representados na Figura 52 e no Quadro 14.

Figura 52 – Resposta para amostras desconhecidas – arquivo do tipo saudável (I).



Fonte: Autoria própria.

Quadro 14 – Resultados – arquivo do tipo saudável (I).

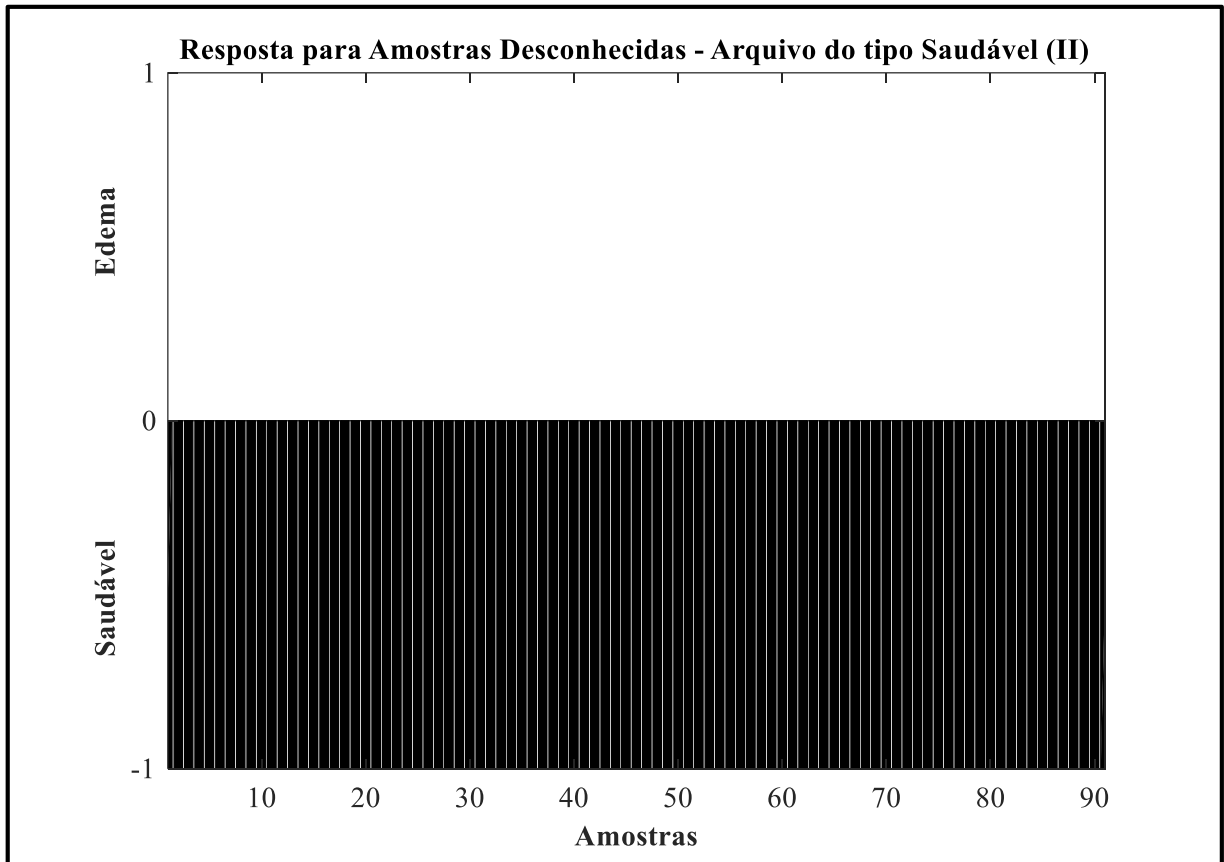
Resultados – arquivo do tipo saudável (I)	
Total de amostras	87
Amostras do tipo edema	1
Amostras do tipo saudáveis	86
Amostras indeterminadas	0
Porcentagem de amostras com edema	1,15%

Fonte: Autoria própria.

5.7.5 Simulação de operação – arquivo do tipo saudável (II)

Os resultados referentes à resposta do classificador para a simulação do segundo arquivo do tipo saudável estão representados na Figura 53 e no Quadro 15.

Figura 53 – Resposta para amostras desconhecidas – arquivo do tipo saudável (II).



Fonte: Autoria própria.

Quadro 15 – Resultados – arquivo do tipo saudável (II).

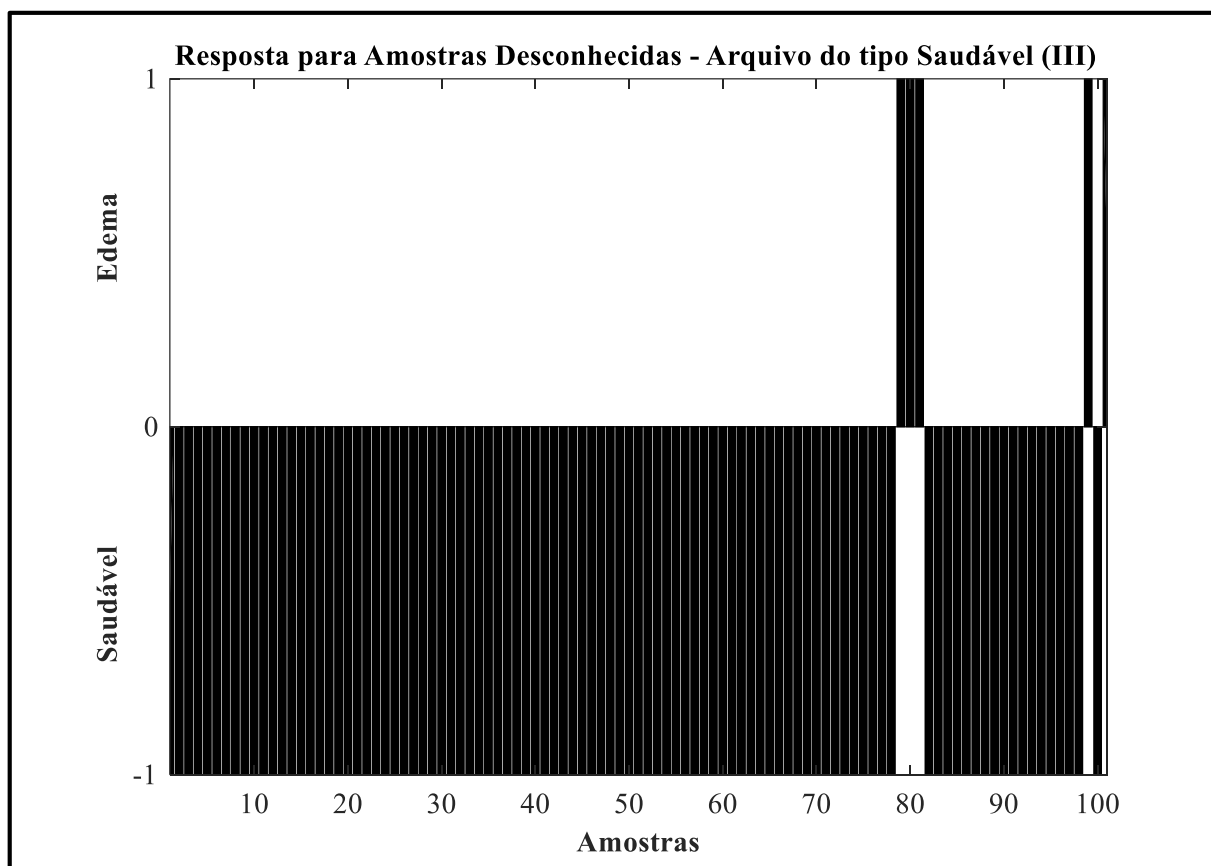
Resultados – arquivo do tipo saudável (II)	
Total de amostras	91
Amostras do tipo edema	0
Amostras do tipo saudáveis	91
Amostras indeterminadas	0
Porcentagem de amostras com edema	0,0%

Fonte: Autoria própria.

5.7.6 Simulação de operação – arquivo do tipo saudável (III)

Os resultados referentes à resposta do classificador para a simulação do terceiro arquivo do tipo saudável estão representados na Figura 54 e no Quadro 16.

Figura 54 – Resposta para amostras desconhecidas – arquivo do tipo saudável (III).



Fonte: Autoria própria.

Quadro 16 – Resultados – arquivo do tipo saudável (III).

Resultados – arquivo do tipo saudável (III)	
Total de amostras	101
Amostras do tipo edema	5
Amostras do tipo saudáveis	96
Amostras indeterminadas	0
Porcentagem de amostras com edema	4,95%

Fonte: Autoria própria.

5.7.7 Simulação de operação – discussão dos resultados

Após a realização dos testes de resposta para amostras de simulação, nota-se que o classificador apresentou índices de acerto consistentes com os resultados da validação. Contudo, a realização de apenas 6 testes (cerca de 5% da população total estudada) não apresenta forte embasamento estatístico para realmente comprovar que a RNA conseguiu adquirir aprendizado em relação as características da patologia e desenvolver um elevado grau

de generalização para operação com arquivos desconhecidos. Porém, como os arquivos utilizados nesta etapa foram totalmente desconsiderados nos procedimentos anteriores, optou-se pela remoção de apenas 6 arquivos para não ferir a quantidade e a diversidade de amostras para o treinamento das RNA.

Deve-se ressaltar, também, que os arquivos apresentados não sofreram nenhum tipo de tratamento em relação ao conteúdo dos áudios visando simular as condições reais, podendo conter trechos não vocalizados como sons de respiração, ruídos ou pequenas perturbações na gravação.

6 CONSIDERAÇÕES FINAIS

Ultimamente, com a evolução dos recursos computacionais, as aplicações de redes neurais artificiais já estão presentes em inúmeros setores da sociedade, permitindo o processamento de informações de maneiras que ultrapassam as limitações humanas. Contudo, ainda existem áreas com grande potencial para se beneficiarem com as modelagens inteligentes. Levando-se em conta, principalmente, os estudos biológicos, tais contribuições consolidam avanços nas pesquisas, enfrentando os obstáculos impostos pela natureza e melhorando indiretamente a qualidade de vida das pessoas.

O sinal da voz possui características peculiares que representam grandes desafios para seu estudo e processamento. Logo, tem-se a possibilidade de desenvolvimento de pesquisas focando soluções robustas que permitam a extração de características e reconhecimento de padrões. Foi proposto, então, a aplicação de redes neurais artificiais em conjunto com a transformada *wavelet packet* para a identificação do edema de Reinke em sinais de voz. Tem-se como maior motivação para o desenvolvimento da proposta a possibilidade de auxiliar na identificação de patologias em um indivíduo sem causar dor e desconforto.

Por meio dos resultados obtidos, comprova-se a eficácia do método proposto, atingindo elevado nível de eficiência e mantendo taxas de acertos acima dos 95% mesmo ao trabalhar com mais de 17000 amostras. É de grande valia destacar que o desempenho da rede englobou, também, baixas taxas de omissão de distúrbio e falso positivos na classificação, aumentando a confiabilidade nos resultados. Por fim, nota-se que os resultados se mantiveram constantes nas simulações com amostras desconhecidas, indicando qualidade no treinamento e alta capacidade de generalização.

Deve-se ressaltar que o estudo elaborado não descarta a importância e aplicação dos métodos convencionais de diagnóstico. É proposto apenas um método para auxiliar os profissionais das áreas médicas e uma alternativa as análises acústicas que já possuem aplicação consolidadas hoje em dia.

Para trabalhos futuros, pode-se destacar a investigação de outros parâmetros a serem estudados nos coeficientes da transformada, bem como, diferentes vogais, patologias e arquiteturas de RNA. Desse modo, é possível incrementar o desempenho do classificador desenvolvido e até mesmo adicionar outros tipos de distúrbios em sua classificação.

REFERÊNCIAS

- ALCAIM, Abraham; OLIVEIRA, Carlos A. S.; **Fundamentos do Processamento de Sinais de Voz e Imagem**. Rio de Janeiro, RJ: Editora Interciência: PUC-Rio, 2011.
- ARJMANDI, Meisam K. et al. **Identification of voice disorders using long-time features and support vector machine with different feature reduction methods**. *Journal of Voice*, vol 25, issue 6, p. 275- 289, 2011.
- ABORL-CCF – Associação Brasileira de Otorrinolaringologia e Cirurgia Cérvico-Facil. **Consenso nacional sobre voz profissional: voz e trabalho, uma questão de saúde e direito do trabalhador**. Rio de Janeiro, 2004.
- BEHLAU, Mara. **A voz. O livro do especialista. Vol. I**. Rio de Janeiro, RJ: Revinter, 2001.
- BISHOP, Christopher M. **Neural networks for pattern recognition**. Oxford, UK: Oxford university press, 1995.
- BISHOP, Christopher M. **Pattern Recognition and Machine Learning**. New York, US: Springer, 2007.
- CARVALHO, Raphael T.S.; CAVALCANTE, Charles C.; CORTEZ, Paulo C. **Wavelet transform and artificial neural networks applied to voice disorders identification**. In *Nature and Biologically Inspired Computing (NaBIC)*, 2011 Third World Congress on: p. 371-376. IEEE, 2011.
- DA SILVA, Ivan N.; SPATTI, Danilo. H.; FLAUZINO, Rogério. A. **Redes Neurais Artificiais para engenharia e ciências aplicadas**. São Paulo, SP: Artliber, 2010.
- DAJER, Maria E. **Análise de sinais de voz por padrões visuais de dinâmica vocal**. 154 f. Tese. Universidade de São Paulo, São Carlos, 2010.
- DUDA, Richard O.; HART, Peter E.; STORK, David G. **Pattern Classification**. Second Edition. New Jersey, US: John Wiley & Sons, 2012.
- FAUSETT, Laurene V. **Fundamentals of Neural Networks: Architectures, Algorithms And Applications**. New York, NY: Pearson, 1993.

FREITAS, Susana V. **Correlação entre avaliação acústica e perceptual na caracterização de vozes patológicas**. 132 f. Relatório do Estado-da-Arte. Faculdade de Engenharia Universidade do Porto, Porto, 2010.

GHASEMZADEH, Hamzeh et al. **Detection of vocal disorders based on phase space parameters and Lyapunov spectrum**. Biomedical Signal Processing and Control, 22, p. 135-145, 2015.

GODINO-LLORENTE, Juan. I.; GOMEZ-VILDA, Pedro. **Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors**. Biomedical Engineering, IEEE Transactions on 51, no. 2: p. 380-384, 2004.

GONZÁLEZ, Isabel V. **Videolaringoestroboscopia: una técnica para visualizar las cuerdas vocales**. Estudios de fonética experimental, no. 17: p. 418-432, 2008.

GRAPS, Amara. **An introduction to wavelets**. Computational Science & Engineering, IEEE 2(2): p. 50-61, 1995.

HAGAN, Martin T. **Neural Network Design**. Second Edition. Boston, MA: Martin Hagan, 2014.

HAYES, Monson H. **Statistical digital signal processing and modeling**. New Jersey, US: John Wiley & Sons, 2009.

HAYKIN, Simon S. **Neural Networks: A Comprehensive Foundation**. New Jersey, US: Prentice Hall, 1999.

HILTON, Michael L. **Wavelet and wavelet packet compression of electrocardiograms**. Biomedical Engineering, IEEE Transactions on, 44(5), p. 394-402, 1997.

HORNIK, Kurt; STINCHCOMBE, Maxwell; WHITE, Halbert. **Multilayer feedforward networks are universal approximators**. Neural networks, Elsevier, vol. 2, issue 5, p. 359-366, 1989.

HSU, Hwei P. **Signals and systems. Schaum outline series**. New York, US: McGraw-Hill, 1995.

JANG, Jyh S. R.; SUN, Chuen T.; MIZUTANI, Eiji. **Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence**. New Jersey, US: Prentice Hall, 1997.

KÖHLE, Juliana et al. **Ação de proteção de saúde vocal: perfil da população e correlação entre auto-avaliação vocal, queixas e avaliação fonoaudiológica perceptivo-auditiva e acústica**. *Distúrbios da Comunicação*. ISSN 2176-2724 16.3, 2004.

KRIESEL, David. **A Brief Introduction to Neural Networks**. 2007. Disponível em <http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf > Acesso em: 30 abri. 2016.

LIM S., et al. **The Incidence of Premalignant and Malignant Disease in Reinke's Edema**. *Otolaryngology - Head and Neck Surgery* 150.3: p.434-436, 2014.

LIMA, Christiane F.L.; ASSIS, Francisco M.; DE SOUZA, Cleonilson .P. **A comparative study of use of Shannon, Rényi and Tsallis entropy for attribute selecting in network intrusion detection**. In *Measurements and Networking Proceedings (M&N)*, 2011 IEEE International Workshop on: p. 77-82, 2011.

LIMA-SILVA, Maria F. B. de, et al. **Distúrbio de voz em professores: autorreferência, avaliação perceptiva da voz e das pregas vocais**. *Revista da Sociedade Brasileira de Fonoaudiologia* 17.4: p. 391-397, 2012.

MACEDO, Cristiane S.; SOUZA, Carla L. de; THOMÉ, Célia. **Readaptação de professores por disfonia na rede municipal de ensino de Salvador**. *Revista Baiana de Saúde Pública* 32.1: p.72, 2014.

MANOLAKIS, Dimitris G.; INGLE, Vinay K. **Applied digital signal processing: theory and practice**. Cambridge, UK: Cambridge University Press, 2011.

MITRA, Sanjit K.; KUO, Yonghong. **Digital signal processing: a computer-based approach**. Vol. 2. New York, US: McGraw-Hill, 2006

PINHO, Silvia M. R. **Manual de higiene vocal para profissionais da voz**. Carapicuíba, SP: Pró-Fono, 1997.

PINHO, Silvia M. R.; TSUJI, Domingos H.; BOHADANA, Saramira. C. **Fundamentos em Laringologia e Voz**. Rio de Janeiro, RJ: Revinter, 2006.

PRARRAGA, Adriane. **Aplicação da transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas**. 163 f. Dissertação. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

RABINER, Lawrence R.; SCHAFER, Ronald W. **Theory and Applications of Digital Speech Processing**. First Edition. New Jersey, US: John Wiley & Sons Inc., 2012.

RAMIREZ-VILLEGAS, Juan F.; RAMIREZ-MORENO, David F. **Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions**. *Neurocomputing* 77.1: p. 82-100, 2012.

RIOUL, Oliver.; VETTERLI, Martin. **Wavelets and signal processing**. *IEEE signal processing magazine*, 8, p.14-38, 1991.

ROSA, Marcelo de O. **Análise acústica da voz para pré-diagnóstico de patologias da laringe**. 219 f. Dissertação. Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos, 1998.

SALHI, Lotfi; TALBI, Mourad; CHERIF, Adnane. **Voice disorders identification using hybrid approach: Wavelet analysis and multilayer neural networks**. *World Academy of Science, Engineering and Technology* 45.21: p. 330-339, 2008.

SCALASSARA, Paulo R.; DAJER, Maria E.; MACIEL, C.D. **Application of autoregressive decomposition and pole tracking to pathological voice signals**. *Multimedia, Seventh IEEE International Symposium on*. IEEE, 2005.

SPATTI, Danilo H. **Automatização de Processos de Detecção de Faltas em Linhas de Distribuição Utilizando Sistemas Especialistas Híbridos**. 208 f. Tese. Universidade de São Paulo, São Carlos, 2011.

TITZE, I. R. **Princípios da Produção Vocal**. Salt Lake City, UT: National Center for Voice & Speech, 2013.

TSUJI, Domingos H. et al. **Improvement of vocal pathologies diagnosis using high-speed videolaryngoscopy**. *International archives of otorhinolaryngology*, 18(3): p. 294-302, 2014.

UMAPATHY, Karthikeyan, et al. **Discrimination of pathological voices using a time-frequency approach**. *Biomedical Engineering, IEEE Transactions on* 52.3: p. 421-430, 2005.

VETTERLI, Martin; HERLEY, Cormac. **Wavelets and filter banks: Theory and design.** Signal Processing, IEEE Transactions on 40.9: 2207-2232, 1992.

WEEKS, Michael. **Digital Signal Processing Using MATLAB & Wavelets.** Massachusetts, US: Jones & Bartlett Learning, 2010.

WILAMOWSKI, Bogdan M. **Neural network architectures and learning algorithms.** Industrial Electronics Magazine, IEEE, vol. 3, issue 4, p. 56-63, 2009.

ZHANG, Yudong, et al. **Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEP-SVM).** Entropy 17.4: p. 1795-1813, 2015.