

Luana Rodrigues Barros

**Classificação de Patologias Laríngeas baseada em
Gráficos de Recorrência e Redes Neurais
Convolucionais**

João Pessoa – PB
Março de 2023

Luana Rodrigues Barros

Classificação de Patologias Laríngeas baseada em Gráficos de Recorrência e Redes Neurais Convolucionais

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, como requisito necessário à obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Telecomunicações

Linha de Pesquisa: Processamento de Sinais

Orientação: Suzete Élide Nobrega Correia, Doutora

Coorientação: Silvana Luciene do Nascimento Cunha Costa, Doutora

João Pessoa – PB, Março de 2023

© Luana Rodrigues Barros — luana.rodrigues@academico.ifpb.edu.br

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

B277c	<p>Barros, Luana Rodrigues.</p> <p>Classificação de patologias laríngeas baseada em gráficos de recorrência e redes neurais / Luana Rodrigues Barros. – 2023. 87 f. : il.</p> <p>Dissertação (Mestrado em Engenharia Elétrica) – Instituto Federal de Educação da Paraíba – IFPB / Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), 2023.</p> <p>Orientação : Profa. D.ra Suzete Élide Nóbrega Correia.</p> <p>Coorientação : Profa. D.ra Silvana Luciene do Nascimento Cunha Costa.</p> <p>1.Processamento de sinais - gráficos de recorrência. 2. Patologias laríngeas. 3. Sinais de voz. 4. Sistema dinâmico não linear. 5. MEEI. I Título.</p> <p style="text-align: right;">CDU 621.391(043)</p>
-------	--



MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA

PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU*
MESTRADO EM ENGENHARIA ELÉTRICA

LUANA RODRIGUES BARROS

**CLASSIFICAÇÃO DE PATOLOGIAS LARÍNGEAS BASEADA EM GRÁFICOS DE RECORRÊNCIA E REDES
NEURAIS CONVOLUCIONAIS**

Dissertação apresentada como requisito para obtenção do título de Mestre em Engenharia Elétrica, pelo Programa de Pós- Graduação em Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB - Campus João Pessoa.

Aprovado em 28 de março de 2023.

Membros da Banca Examinadora:

Dr^a. Suzete Élide Nóbrega Correia

IFPB - PPGE

Dr^a. Silvana Luciene do Nascimento Cunha Costa

IFPB - PPGEE

Dr. Carlos Danilo Miranda Regis

IFPB - PPGEE

Dr. Washington Cesar de Almeida Costa

IFPB

Dr^a. Guênia Mara Vieira Ladeira

UFU

João Pessoa/2023

Documento assinado eletronicamente por:

- **Suzete Elida Nobrega Correia**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 05/04/2023 00:50:44.
- **Carlos Danilo Miranda Regis**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 05/04/2023 08:51:18.
- **Washington Cesar de Almeida Costa**, PROF ENS BAS TEC TECNOLOGICO-SUBSTITUTO, em 06/04/2023 09:46:51.
- **Guênia Mara Vieira Ladeira**, PROFESSOR DE ENSINO SUPERIOR NA ÁREA DE ORIENTAÇÃO EDUCACIONAL, em 10/04/2023 13:47:08.
- **Silvana Luciene do Nascimento Cunha Costa**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 11/04/2023 11:42:31.

Este documento foi emitido pelo SUAP em 30/03/2023. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifpb.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código 411505
Verificador: 91e0debabb
Código de Autenticação:



Av. Primeiro de Maio, 720, Jaguaribe, JOÃO PESSOA / PB, CEP 58015-435
<http://ifpb.edu.br> - (83) 3612-1200

AGRADECIMENTOS

A minha gratidão é eterna à minha família. Meus pais Luzia Barros e Otacílio Barros, meu esposo Gabriel Soares, meu sogro Manoel Brasileiro, minha sogra Djanira Pereira, e minha cunhada Nicole Pereira.

Gabriel Soares, a nossa parceria de tantos anos se estende por inúmeros aspectos. Neste trabalho, fomos parceiros de pesquisa, e a sua contribuição foi imensurável. Desde a atualização deste template, não apenas para mim, mas como na revisão dos mínimos detalhes técnicos.

Esta pesquisa foi conduzida durante uma época de adversidades e transformações, em particular, ressalto o período da pandemia de COVID-19, que afligiu a todos. Suzete Correia e Silvana Costa, respectivamente orientadora e coorientadora desta pesquisa, não se ativeram apenas ao papel da orientação técnica, que por sinal foi executada com maestria. A sensibilidade, dedicação e amor com os quais fui conduzida durante este percurso nunca serão esquecidos. Foram luz quando o caminho estava escuro. Hoje, tento transbordar este apreço a todos que me cercam, como um reflexo do que me foi proporcionado.

Deixo aqui uma menção honrosa a Michel Dias, que foi conselheiro, amigo, chefe, entre tantos outros que aqui não caberia. Você me mostrou que eu tinha forças para trilhar todos os caminhos e me conduziu por desafios que nunca imaginei.

Não posso deixar de mencionar e agradecer àquela família que se formou ao longo da minha trajetória, Gabriel Daltro, Priscilla Sebadelhe, Sérgio Sales e Edson Bandeira. Obrigada por todo o apoio e por todo companheirismo durante tantos anos.

À companheira durante todo o mestrado, minha amiga Máisa Lauriane, deixo aqui todo o meu agradecimento por cada dúvida e cada desafio que enfrentamos juntas.

Agradeço aos demais professores do IFPB que tanto colaboraram para a minha formação, sendo pilares fundamentais para toda vida.

Agradeço ao IFPB pelo acolhimento, pelos recursos e por todas as oportunidades que pude desfrutar graças aos conhecimentos adquiridos neste espaço.

*Nada seria possível sem aqueles que deixaram fértil o solo.
E assim cultivei.*

Luana Barros

RESUMO

A laringe é uma das principais estruturas do sistema fonatório humano e pode ser afetada por patologias que ocasionam alterações nas características intrínsecas da voz do locutor. Com isso, a análise acústica tem se destacado como uma ferramenta na detecção de patologias laríngeas. Nesse contexto, os sinais de voz podem ser modelados como gráficos de recorrência (RPs), que permitem a análise de recorrências em forma de matrizes bidimensionais. Contudo, até o presente momento apenas RPs binários foram aplicados com essa finalidade. Este estudo apresenta sinais de vozes saudáveis e patológicas modelados como quatro tipos de RPs: RP binário, RP sem limiar (URP - *Unthresholded Recurrence Plot*), o ReLU-RP (*Rectified Linear Unit based Recurrence Plot*) e o RP do espectro da frequência (*FFT-RP - Fast Fourier Transform Recurrence Plot*). Foi definida uma rede neural convolucional (CNN - *Convolutional Neural Network*) para realizar a classificação individual dos quatro tipos de RPs. Foi utilizado o método de cross-validation ($k = 4$) para treinamento e avaliação de desempenho das CNNs. Os melhores resultados de classificação do sinal de áudio completo foram obtidos para o gráfico de recorrência do tipo ReLU-RP ($\varepsilon = 0,3$), considerando métricas de acurácia, precisão, sensibilidade, especificidade e F1-score.

Palavras-chave: Gráfico de Recorrência. Patologias Laríngeas. Sinais de Voz. Sistema Dinâmico não Linear. MEEI.

ABSTRACT

The larynx is one of the main structures of the human phonatory system and can be affected by pathologies which cause changes to the intrinsic voice characteristics of the speaking individual. Because of that, acoustic analysis can be highlighted as a larynx pathology detection tool. In this context, voice signals can be modelled as recurrence plots (RPs) which allow recurrence analysis in a bidimensional matrix format. However, until now only binary RPs were used for this purpose. This study presents healthy and pathological voice samples modelled as four recurrence plot types: binary RP, Unthresholded RP (URP), Rectified Linear Unit-based RP (ReLU-RP) and frequency domain RP (Fast Fourier Transform RP, FFT-RP). A Convolutional Neural Network (CNN) architecture was proposed and used for classification of healthy and pathological samples. Cross-validation method ($k = 4$) was used to train and to evaluate the neural network performance. The best results for whole audio signal classification were obtained using ReLU-RPs ($\varepsilon = 0.3$) considering metrics such as accuracy, precision, sensibility, specificity and F1-score.

Keywords: Recurrence Plot. Laryngeal Pathologies. Voice Signals. Non-Linear Dynamic System. MEEI.

LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema geral do aparelho de produção de voz.	20
Figura 2 – Exemplos de patologias laríngeas.	23
Figura 3 – Trajetória no espaço de fase bidimensional de um sistema dinâmico arbitrário.	24
Figura 4 – (a) Espaço de fase em um sistema dinâmico. b) Espaço de fase delimitado por um raio ε . (c) Gráfico de recorrência.	27
Figura 5 – Sinal constante.	30
Figura 6 – Sinal constante (FFT).	31
Figura 7 – Sinal senoidal 200 Hz e seus respectivos RPs-binários, URP e ReLU-RPs.	33
Figura 8 – Projeção bidimensional do atrator de um sinal senoidal de 200 Hz.	33
Figura 9 – FFT do sinal senoidal 200 Hz e suas respectivas variações de FFT-RPs.	34
Figura 10 – Projeção bidimensional do atrator do espectro da frequência um sinal senoidal de 200 Hz.	34
Figura 11 – Sinal senoidal 800 Hz e seus respectivos RPs-binários, URP e ReLU-RPs.	35
Figura 12 – Projeção bidimensional do atrator de um sinal senoidal de 800 Hz.	36
Figura 13 – FFT do sinal senoidal de 800 Hz e suas respectivas variações de FFT-RPs.	37
Figura 14 – Projeção bidimensional do atrator do espectro da frequência um sinal senoidal de 800 Hz.	37
Figura 15 – Sinal senoidal de 200 Hz somado a um sinal senoidal de 800 Hz e seus respectivos RPs-binários, URP e ReLU-RPs.	38
Figura 16 – Projeção bidimensional do atrator de um sinal senoidal de 200 Hz somado a um sinal senoidal de 800 Hz.	39
Figura 17 – Projeção bidimensional do atrator do espectro da frequência um sinal senoidal de 200 Hz somado a um sinal senoidal de 800 Hz.	39
Figura 18 – FFT do sinal senoidal de 200 Hz somado ao sinal senoidal de 800 Hz e suas respectivas variações de FFT-RPs.	40
Figura 19 – Ruído branco e seus respectivos RPs-binários, URP e ReLU-RPs.	41
Figura 20 – Projeção bidimensional do espaço de estados de um sinal do tipo ruído branco.	41
Figura 21 – FFT do sinal do tipo ruído branco e suas respectivas variações de FFT-RPs.	42
Figura 22 – Projeção bidimensional do espaço de estados do espectro da frequência de um sinal do tipo ruído branco.	42

Figura 23 – Sinal de voz saudável (identificado por AXH125 na base Kay Eleme- trics (1994)) e seus respectivos RPs-binários, URP e ReLU-RPs.	44
Figura 24 – Projeção bidimensional do atrator do sinal de voz saudável (identifi- cado por AXH125 na base Kay Elemetrics (1994)).	44
Figura 25 – Espectro de frequência de sinal de voz saudável (identificado por AXH125 na base Kay Elemetrics (1994)) e suas respectivas varia- ções de FFT-RPs.	45
Figura 26 – Projeção bidimensional do atrator do espectro da frequência do sinal de voz saudável (identificado por AXH125 na base Kay Elemetrics (1994)).	45
Figura 27 – Projeção bidimensional do atrator do sinal de voz afetada por nódulos bilaterais (identificado por ANB28AN na base Kay Elemetrics (1994)).	46
Figura 28 – Sinal de voz afetada por nódulos bilaterais (identificado por ANB28AN na base Kay Elemetrics (1994)) e seus respectivos RPs-binários, URP e ReLU-RPs.	47
Figura 29 – Espectro de frequência de sinal de voz afetada por nódulos bilaterais (identificado por ANB28AN na base Kay Elemetrics (1994)) e suas respectivas variações de FFT-RPs.	47
Figura 30 – Projeção bidimensional do atrator do espectro da frequência do sinal de voz afetada por nódulos bilaterais (identificado por ANB28AN na base Kay Elemetrics (1994)).	48
Figura 31 – Sinal de voz afetada por edemas bilaterais (identificado por ANA15 na base Kay Elemetrics (1994)) e seus respectivos RPs-binários, URP e ReLU-RPs.	49
Figura 32 – Projeção bidimensional do atrator do sinal de voz afetada por edemas bilaterais (identificado por ANA15 na base Kay Elemetrics (1994)).	49
Figura 33 – Espectro de frequência de sinal de voz afetada por edemas bilaterais (identificado por ANA15 na base Kay Elemetrics (1994)) e suas respectivas variações de FFT-RPs.	50
Figura 34 – Projeção bidimensional do atrator do espectro da frequência do sinal de voz afetada por edemas bilaterais (identificado por ANA15 na base Kay Elemetrics (1994)).	50
Figura 35 – Sinal de voz afetada paralisia unilateral nas pregas vocais (identifi- cado por ABB09AN na base Kay Elemetrics (1994)) e seus respecti- vos RPs-binários, URP e ReLU-RPs.	52
Figura 36 – Projeção bidimensional do atrator do sinal de voz afetada paralisia unilateral nas pregas vocais (identificado por ABB09AN na base Kay Elemetrics (1994)).	52

Figura 37 – Espectro de frequência de sinal de voz afetada paralisia unilateral nas pregas vocais (identificado por ABB09AN na base Kay Elemetrics (1994)) e suas respectivas variações de FFT-RPs.	53
Figura 38 – Projeção bidimensional do atrator do espectro da frequência do sinal de voz afetada paralisia unilateral nas pregas vocais (identificado por ABB09AN na base Kay Elemetrics (1994)).	53
Figura 39 – Elementos estruturais do neurônio artificial <i>perceptron</i>	55
Figura 40 – Principais funções de ativação.	56
Figura 41 – Gráfico de fluxo de sinal de uma rede neural de uma camada.	57
Figura 42 – Classificação de dados utilizando uma rede <i>perceptron</i> de uma camada.	57
Figura 43 – Classificação de dados utilizando uma MLP.	58
Figura 44 – Camadas convolucionais de uma CNN com <i>kernels</i> retangulares.	59
Figura 45 – Conexões entre camadas convolucionais.	59
Figura 46 – Camada do tipo <i>max pooling</i> com kernel de tamanho 2×2	60
Figura 47 – Dimensões das camadas convolucionais.	60
Figura 48 – Arquitetura típica de uma CNN.	61
Figura 49 – Método de validação cruzada <i>4-fold</i>	62
Figura 50 – Método de validação cruzada <i>4-fold</i>	63
Figura 51 – Diagrama de etapas.	68
Figura 52 – Diferentes RPs gerados a partir de um segmento do sinal de voz.	71
Figura 53 – Diferentes RPs gerados a partir da transformada de Fourier de um segmento do sinal de voz.	72
Figura 54 – Corte do gráfico de recorrência.	72
Figura 55 – Arquitetura da CNN para classificação de gráficos de recorrência.	73
Figura 56 – Relação entre predição da rede e classificação real.	75

LISTA DE TABELAS

Tabela 1 – Resultados para classificação dos segmentos de áudio usando RP-binários.	77
Tabela 2 – Resultados para classificação dos sinais de áudio completos usando RP-binários.	77
Tabela 3 – Resultados para classificação dos segmentos de áudio usando URPs. 77	
Tabela 4 – Resultados para classificação dos sinais de áudio completos usando URPs.	78
Tabela 5 – Resultados para classificação dos segmentos de áudio usando ReLU-RPs.	78
Tabela 6 – Resultados para classificação dos sinais de áudio completos usando ReLU-RPs.	78
Tabela 7 – Resultados para classificação dos segmentos de áudio usando FFT-RPs.	79
Tabela 8 – Resultados para classificação dos sinais de áudio completos usando FFT-RPs.	79
Tabela 9 – Melhores resultados para classificação dos segmentos de áudio. . .	79
Tabela 10 – Melhores resultados para classificação dos arquivos de áudio completos.	80

LISTA DE ABREVIATURAS E SIGLAS

AFNN	<i>Averaged False Nearest Neighbor</i>
CNN	<i>Convolutional Neural Networks</i>
DBN	<i>Deep Belief Network</i>
DET	Determinismo
ECG	Eletrocardiograma
EEG	Eletroencefalograma
ENTR	Entropia de Shannon
FFT-RP	<i>Fast Fourier Transform Recurrence Plot</i>
FN	<i>Falso Negativo</i>
FNN	<i>False Nearest Neighbor</i>
FP	<i>Falso Positivo</i>
GNE	<i>Glottal-to-Noise Excitation Ratio</i>
KNN	<i>K-nearest Neighbors</i>
LR	<i>Logistic Regression</i>
MEEI	Massachusetts Eye and Ear Infirmary
MLP	<i>Multiplayer Perceptron</i>
PPG	<i>Photoplethysmogram, fotopletismograma</i>
ReLU-RP	<i>Rectified Linear Unit based Recurrence Plot</i>
RP	<i>Recurrence Plot</i>
RQA	<i>Recurrence Quantification Analysis</i>
RR	<i>Recurrence Rate</i>
SVM	<i>Support Vector Machine</i>
TN	<i>True Negative, Verdadeiro Negativo</i>
TP	<i>True Positive, Verdadeiro Positivo</i>
URP	<i>Unthresholded Recurrence Plot</i>

LISTA DE SÍMBOLOS

\in	Pertence a
m	Quantidade de dimensões do espaço de estados
τ	Passo de reconstrução, defasagem utilizada para construção do espaço de estados
$x(t)$	Série temporal
T	Tamanho da série temporal $x(t)$
ε	Raio de vizinhança para determinação de recorrência entre estados
$\theta(\cdot)$	Função degrau unitário
$\ \cdot\ $	Norma euclidiana
$\vec{\xi}_i$	Estado i , ponto i no espaço de estados
N	Quantidade de estados $\vec{\xi}_i$
f	Frequência
f_s	Frequência de amostragem
\mathbb{R}	Conjunto dos números reais
\mathbb{R}_m	Espaço real m -dimensional
$\varphi(\cdot)$	Função de ativação do neurônio <i>perceptron</i>
$R_{i,j}^{m,\varepsilon}$	Ponto do gráfico de recorrência de coordenadas (i, j) do espaço de estados de dimensão m usando raio de vizinhança ε

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Objetivos	19
1.1.1	Geral	19
1.1.2	Específicos	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	Sistema de Produção de Voz	20
2.1.1	Patologias Laríngeas	21
2.2	Formação dos Gráficos de Recorrência	22
2.2.1	Espaço de Estados	22
2.2.2	Dimensão de Imersão	24
2.2.3	Passo de Reconstrução	25
2.2.4	RP Binários	26
2.2.5	RP sem limiar	27
2.2.6	ReLU-RP	28
2.2.7	FFT-RP	28
2.3	Análise da formação de estruturas nos gráficos de recorrência	29
2.3.1	Formação de RPs para um sinal constante	29
2.3.2	Formação de RPs para sinais senoidais	32
2.3.3	Formação de RPs para ruído branco	40
2.3.4	Gráficos de recorrência de sinais de voz	43
2.4	Redes Neurais Convolucionais	54
2.4.1	Neurônio artificial <i>perceptron</i>	54
2.4.2	Camada convolucional	58
2.4.3	Tipo de aprendizado	61
2.4.4	Validação cruzada	62
3	ESTADO DA ARTE	64
4	METODOLOGIA	68
4.1	Base de dados	69
4.2	Gráficos de Recorrência	70
4.3	Classificador	73
5	RESULTADOS	75
5.1	Análise dos resultados	75
5.2	Discussão dos resultados	80

6	CONSIDERAÇÕES FINAIS	82
	REFERÊNCIAS	83

1. INTRODUÇÃO

No corpo humano, a laringe e suas estruturas compõem o sistema fonatório, que consiste em uma das três partes principais do mecanismo de produção da voz. As demais partes são constituídas pelo sistema de pressão de ar (composto pelo diafragma, músculos do peito, costelas, músculos abdominais e pulmões) e pelo sistema articulatório (composto pelo trato vocal e suas estruturas) (MIRANDA, 2017). Contudo, há algumas patologias que podem afetar diretamente a laringe, ocasionando distúrbios vocais. Nesses casos, as pregas vocais (uma das principais estruturas da laringe) não vibram normalmente, afetando as características intrínsecas da voz do locutor. Essa condição também é denominada disfonia.

De acordo com Martins et al. (2016), cerca de 10% da população geral apresenta distúrbios vocais e, entre os profissionais que utilizam a voz, a proporção chega a 50%. Martins et al. (2016) ainda destacam que crianças e adultos são igualmente afetados, porém as causas são diferentes de acordo com as faixas etárias. Além disso, o comportamento vocal alterado pode colaborar para o aparecimento de algumas lesões tais como nódulos e pólipos vocais. (CIELO et al., 2011).

Nesse contexto, a identificação correta de uma patologia laríngea é essencial para a indicação do melhor tratamento. Tal diagnóstico é realizado de forma subjetiva por profissionais especialistas em patologias da fala. O estudo de Zraick, Wendel e Smith-Olinde (2005) investigou a avaliação perceptivo-auditivo por profissionais acerca da gravidade da disfonia em amostras de áudio de pacientes com vozes afetadas. Os resultados indicam que, mesmo considerando profissionais com mais de sete anos de experiência, a análise individual ainda está sujeita a variações. A pesquisa também corroborou com as evidências literárias de que a confiabilidade do diagnóstico está diretamente relacionada com a experiência do avaliador.

Diante desse cenário, nas últimas décadas, a análise acústica de sinais de voz tem se destacado como ferramenta para auxiliar no diagnóstico dos especialistas em patologias da fala. Por meio desse tipo de análise pode-se extrair atributos do sinal de voz que caracterizam, de forma singular, o sistema de produção vocal, possibilitando a discriminação das diferentes patologias que podem afetar a laringe. Determinados distúrbios vocais, por exemplo, podem alterar a frequência fundamental dos sons sonoros produzidos por um locutor.

O processo de produção da voz pode ser modelado por um sistema dinâmico não linear. Dentre os trabalhos relacionados à análise acústica com foco na detecção de patologias laríngeas, as aplicações baseadas na análise de recorrência de sistemas dinâmicos não lineares torna-se interessante em sinais que apresentam alto grau de

disfonia ou para sinais severamente afetados por patologias laríngeas, nos quais há dificuldade na obtenção de medidas baseadas no modelo linear de produção da fala, especificamente aquelas baseadas na frequência fundamental. A ideia de recorrência, proposta pelo físico e filósofo francês J. H. Poincaré, foi utilizada como base por Eckmann, Kamphorst e Ruelle (1987) para a criação dos gráficos de recorrência (RPs - *recurrence plots*). Os RPs são utilizados para representar a trajetória de sistemas dinâmicos no espaço de fases, permitindo a análise de recorrências em forma de matrizes bidimensionais. Matematicamente, a composição de um RP consiste em verificar se a distância entre cada dois estados de um sistema não linear é menor que um determinado limiar. O resultado é, então, convertido em uma matriz de valores, análoga a uma imagem binária.

A fim de quantificar o número e a duração das recorrências de um sistema dinâmico, apresentado por sua trajetória no espaço de estados, um conjunto de métricas foi proposto por Zbilut e Webber (1992, 1994). Posteriormente, novas métricas de complexidade foram acrescentadas por Marwan et al. (2002). Esse conjunto de métricas é normalmente referenciado como análise de quantificação de recorrência (RQA - *Recurrence Quantification Analysis*).

Apesar da RQA proporcionar métricas que permitem a análise de características variadas presentes nos RPs, ainda existem limitações. Como reforçado por Lee, Lee e Shin (2019), essas medidas não representam totalmente todas as mudanças ao longo do tempo ou tendências mostradas nos RPs. Com isso, os autores consideraram que as redes neurais convolucionais (CNN - *Convolutional Neural Networks*) poderiam colaborar para que essas restrições fossem suplantadas, uma vez que permitem a utilização de RPs diretamente como entrada. Dessa forma, o uso das CNNs eliminaria a necessidade de extração de informações por meio da RQA para classificação posterior.

Além dos gráficos de recorrência tradicionais (ou RPs binários), outros estudos trazem ainda a aplicação de variações dos RPs em sistemas dinâmicos derivados de variados tipos de sinais, como sinais de frequência cardíaca, eletrocardiograma (ECG), entre outros. Pode ser destacado o uso dos gráficos de recorrência sem limiar (URP - *Unthresholded Recurrence Plot*), utilizados por Hatami, Gavet e Debayle (2018) e Chen, Su e Yang (2020); o ReLU-RP (*Rectified Linear Unit based Recurrence Plot*), proposto e aplicado no estudo de Lee, Lee e Shin (2019); e o mais recente RP do espectro da frequência, proposto por Ladeira et al. (2020).

Considerando que não foram encontrados, até o presente momento desta pesquisa, estudos que apliquem variações do RP à análise e discriminação de sinais afetados por patologias laríngeas, este trabalho se propõe a investigar a eficácia dos URPs, ReLU-RPs e RPs do espectro da frequência (ou FFT-RP, como será designado)

nesse contexto. O método de classificação consiste na utilização de uma CNN. O estudo também deve incluir uma comparação com o uso de RPs tradicionais. A base de dados utilizada nessa pesquisa foi gravada pelo Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab.

1.1 OBJETIVOS

1.1.1 Geral

Classificar vozes afetadas por patologias laríngeas com base em diferentes tipos de gráficos de recorrência e redes neurais convolucionais.

1.1.2 Específicos

- Obter diferentes tipos de gráficos de recorrência (RP-binário, URP, ReLU-RP e FFT-RP) de sinais de voz saudáveis e afetados por patologia;
- Propor um modelo de CNN para classificação dos RPs;
- Realizar a classificação dos RPs obtidos da base de dados;
- Avaliar o desempenho do classificador aplicando os diferentes tipos de RPs.

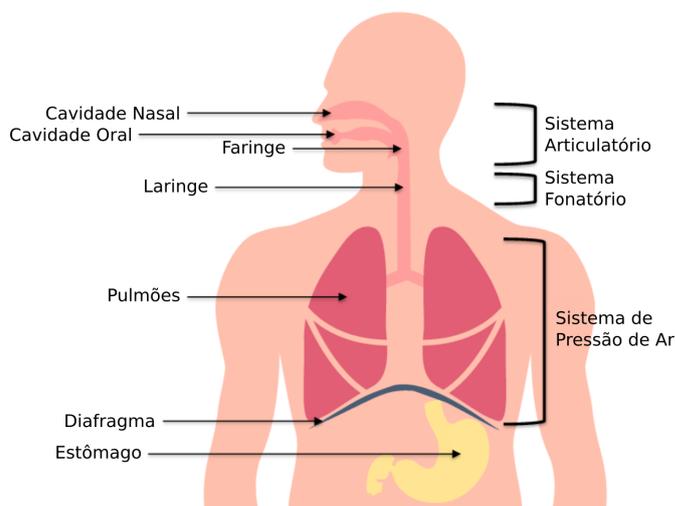
2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados conteúdos relevantes para o desenvolvimento do experimento proposto neste estudo. É caracterizado, de forma resumida, o sistema básico da produção de voz humana, bem como são descritas as patologias laríngeas consideradas nas bases de dados utilizadas neste trabalho. A produção dos gráficos de recorrência também é explorada. Além disso, são abordados os principais conceitos referentes às CNNs.

2.1 SISTEMA DE PRODUÇÃO DE VOZ

O sistema de produção de voz humana é dado por várias estruturas capazes de trabalhar em conjunto. O mecanismo de voz pode ser dividido em três partes principais: o sistema de pressão de ar (composto pelo diafragma, músculos do peito, costelas, músculos abdominais e pulmões), o sistema fonatório (composto pela laringe e suas estruturas) e o sistema articulatório (composto pelo trato vocal e suas estruturas) (MIRANDA, 2017). Na Figura 1 é ilustrada a localização dessas estruturas no corpo humano.

Figura 1 – Esquema geral do aparelho de produção de voz.



Fonte: Adaptado de Miranda (2017).

De acordo com Miranda (2017), para que o ar seja inalado para os pulmões, a caixa torácica precisa se expandir e o diafragma, que forma a base do tórax, precisa se achatar para baixo. Após esse evento, para que haja vocalização, é necessário que o ar armazenados nos pulmões seja liberado, passando pela traqueia e atingindo a laringe, também conhecida como caixa vocal. Dessa forma, a laringe consiste na passagem de ar da traqueia até a faringe. É também responsável pela produção de voz

e proteção das vias aéreas durante a deglutição. Na laringe, o principal componente vibratório são as pregas vocais que, de acordo com The Voice Foundation (2020), consistem em um tecido mole dobrável composto por uma capa (epitélio e lâmina própria superficial), ligamento vocal (lâmina própria intermediária e profunda) e corpo (músculo tireoaritenóideo). Quando o ar é expelido dos pulmões, as pregas vocais vibram, modulando o ar e conseqüentemente produzindo a fala. A abertura entre as pregas vocais é denominada glote, que abre durante a respiração e fecha no momento de produção de voz de deglutição. Na faringe, nas cavidades nasais e na cavidade oral, o som é amplificado e modificado. Essas estruturas compõem o que é chamado de trato vocal. A articulação que caracteriza as palavras é atribuída à língua, ao palato mole e aos lábios, localizados na cavidade oral do trato vocal.

2.1.1 Patologias Laríngeas

Visto que na laringe ocorre a produção de som, as patologias laríngeas comumente ocasionam distúrbios vocais. Isso ocorre devido às pregas vocais não vibrarem normalmente, afetando as características intrínsecas da voz do locutor. Um desses distúrbios é a paralisia das pregas vocais, no qual as pregas vocais podem estar paralisadas ou parcialmente paralisadas (Figura 2b). Como apontado por Cedars-Sinai (2020), pode ser causada por uma infecção viral que afeta os nervos das pregas vocais, uma lesão em um nervo durante uma cirurgia, derrame ou câncer. Ainda de acordo com Cedars-Sinai (2020), se uma ou ambas as pregas vocais estiverem paralisadas em uma posição quase fechada, o locutor pode ter respiração ruidosa ou difícil. Se estiverem paralisadas em uma posição aberta, o locutor pode ter uma voz fraca e ofegante. No trabalho de Murry, Jr e Morris (1995) é feita uma análise detalhada acerca da frequência fundamental (F_0) em pacientes com paralisia das pregas vocais. Além disso, o trabalho de Patel e Parsram (2005) promove um estudo a cerca dos parâmetros acústicos que caracterizam vozes com este tipo de disfonia.

Outra patologia responsável por distúrbios vocais é o edema de Reinke, caracterizada como uma doença benigna das pregas vocais. É designada por Tavaluc e Tan-Geller (2019) como a degeneração polipóide difusa de toda a extensão de uma ou, mais comumente, de ambas as pregas vocais. Recebe a denominação de edema de Reinke devido ao inchaço que ocorre no espaço abaixo da camada epitelial, região conhecida como camada de tecido subepitelial de Reinke ou espaço de Reinke (YONEKAWA, 1988). Um exemplo é ilustrado na Figura 2c. Dentre os principais fatores que ocasionam o aparecimento do edema de Reinke, pode ser destacado o tabagismo. Há cerca de quatro décadas Yonekawa (1988) evidenciou, com base em estudos, que os fumantes constituíam mais de 90% com esta doença. Recentemente, Tavaluc e Tan-Geller (2019) reforçaram que a fumaça proveniente de cigarros ainda é uma das

principais causas da manifestação do edema. A disfonia ocasionada pela doença pode ser observada principalmente pela alteração da F_0 da voz do locutor. Em mulheres constata-se uma F_0 inferior a 130 Hz e em homens inferior a 100 Hz (TAVALUC; TANGELLER, 2019). Normalmente, os valores de F_0 variam entre 80 Hz e 150 Hz para vozes masculinas e entre 150 Hz e 250 Hz para vozes femininas (BEHLAU, 2001).

A laringe também pode ser afetada por nódulos das pregas vocais, que são resultantes de tensões perpendiculares cumulativas nas bordas livres da prega vocal (Figura 2d) e ocorrem com mais frequência em indivíduos com uma ocupação que exige o uso excessivo da voz (SALTÜRK et al., 2019). O problema pode ser agravado caso o abuso vocal persista, de forma que os nódulos fiquem com o aspecto cada vez mais sólido. Este tipo de patologia pode ainda ocorrer devido à alergias ou tabagismo, e o tratamento pode variar de acordo com a causa. Um dos tratamentos consiste na terapia de voz, que auxilia o paciente na manutenção da qualidade da voz, evitando abusos. Saltürk et al. (2019) avaliou a terapia de voz ressonante no tratamento de nódulos de prega vocal em mulheres e observou um aumento significativo da frequência fundamental após o tratamento.

De forma geral, pode haver diversas causas para os distúrbios vocais, como por exemplo crescimento e desenvolvimento do corpo, inflamações e inchaços nas pregas vocais, problemas nos nervos que controlam as pregas vocais, variação na produção de hormônios e abusos vocais. Os principais sintomas podem ser a produção de sons trêmulos, ásperos (roucos), instáveis, fracos, ofegantes, muito altos ou muito baixos, podendo haver inclusive variação de tom (CEDARS-SINAI, 2020). Esses sintomas afetam diretamente os parâmetros de prosódia como F_0 , durações segmentares e energia dos segmentos.

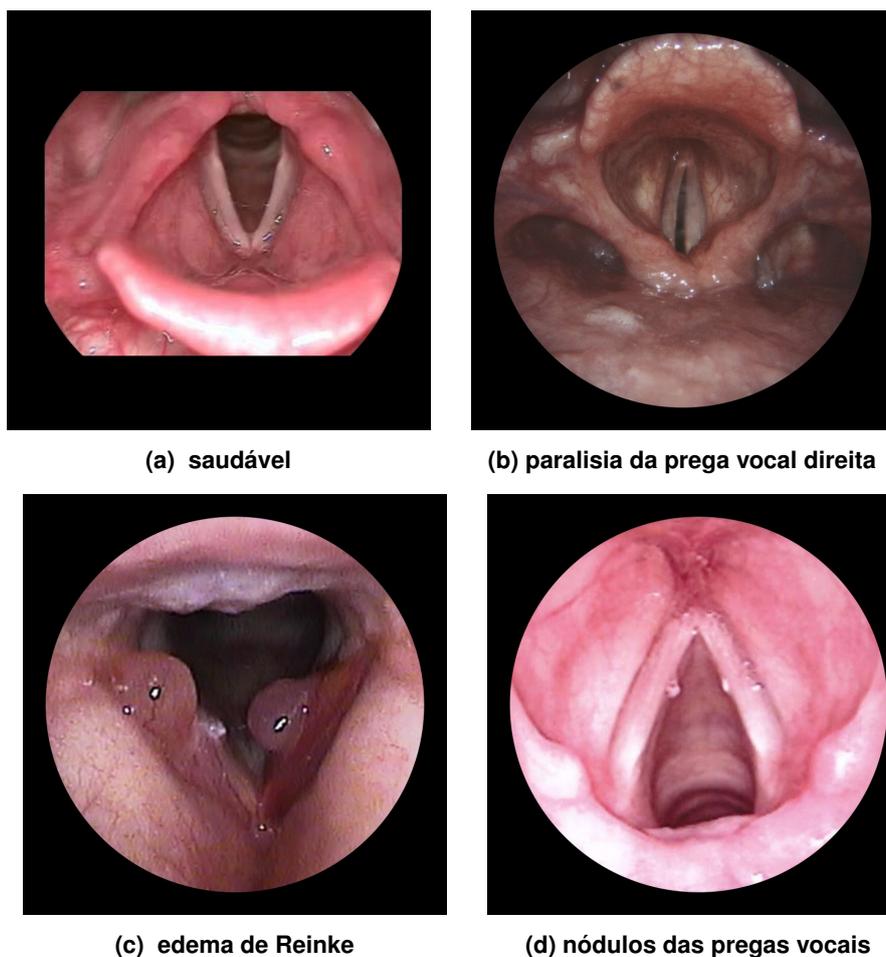
2.2 FORMAÇÃO DOS GRÁFICOS DE RECORRÊNCIA

Os gráficos de recorrência são representações da quantificação de recorrência que ocorrem em uma trajetória. Dessa forma, pode-se dizer que há recorrência quando uma trajetória se encontra na mesma vizinhança no espaço de estados em que já esteve anteriormente.

2.2.1 Espaço de Estados

Um modelo de espaço de estados descreve o comportamento de um sistema dinâmico como um conjunto de equações diferenciais ordinárias (EDO) de primeira ordem. Se um modelo dinâmico é descrito por uma EDO de ordem superior, usando o espaço de estados, o mesmo modelo pode ser descrito como um conjunto de EDOs de primeira ordem acopladas. Pode-se dizer que, o espaço de estados é o conjunto

Figura 2 – Exemplos de patologias laríngeas.



Fontes: Muhammad et al. (2011), Sasaki (2020), Weill Cornell Medicine (2021)

de todos os estados possíveis de um sistema dinâmico, onde cada estado do sistema corresponde a um único ponto no espaço de estados. Esse espaço de estados é frequentemente chamado de espaço de fase e a trajetória formada pelos estados é chamada de atrator (MONTEIRO, 2006).

Em sistemas tais como as séries temporais, há apenas uma variável de estado, $x(t)$, o que limita a análise. Uma técnica denominada reconstrução do espaço de fase permite que variáveis não observadas possam ser estudadas. Como descrito por Casdagli et al. (1991), a ideia básica por trás da reconstrução do espaço de estados é que o passado e o futuro de uma série temporal contêm informações sobre variáveis de estado não observadas que podem ser usadas para definir um estado no tempo presente. Dado que os atratores contêm propriedades geométricas e dinâmicas do espaço de fase original, estudar as propriedades dos atratores permite uma análise mais profunda acerca das propriedades do sistema.

Após observar as características dos sinais estudados no domínio do tempo e

verificar que há informações relevantes na evolução da série temporal, a reconstrução do espaço de estados é necessariamente o primeiro passo que deve ser dado para analisar uma única série temporal em termos de teoria de sistemas dinâmicos (CAS-DAGLI et al., 1991). Nesse sentido, o teorema de Takens (1981) possibilita que um espaço de estado multidimensional possa ser reconstruído a partir de uma única série temporal escalar. A Equação 1, definida e provada por Takens, encapsula em um vetor de defasagens, as informações passadas e futuras de uma série $x(t)$, onde N vetores $\vec{\xi}_i$ com m dimensões são constituídos.

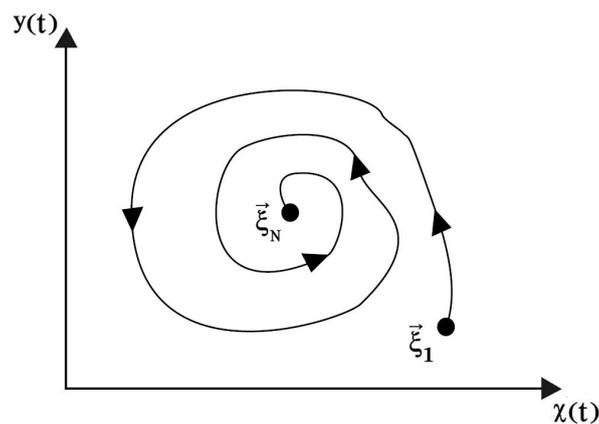
$$\vec{\xi}_i = \{x(t_i), x(t_i + \tau), \dots, x(t_i + (m - 1)\tau)\} \quad (1)$$

Cada vetor $\vec{\xi}_i$ representa um estado em que m é a dimensão de imersão, τ o passo de reconstrução ou tempo de defasagem, e i é o índice do vetor ($i = 1, 2, \dots, N$). Dado que a série temporal $x(t)$ possui tamanho T , a quantidade total N de estados $\vec{\xi}_i$, quando aplicado o teorema de Takens, pode ser dada pela Equação 2.

$$N = T - (m - 1)\tau \quad (2)$$

Um exemplo de um estado de fases bidimensional, representado pelas grandezas $x(t)$ e $y(t)$, pode ser observado na Figura 3, na qual é retratada a evolução temporal do estado inicial $\vec{\xi}_1$ até o estado final $\vec{\xi}_N$.

Figura 3 – Trajetória no espaço de fase bidimensional de um sistema dinâmico arbitrário.



Fonte: Vieira et al. (2018)

2.2.2 Dimensão de Imersão

Como apontado por Souza (2008), a preservação das estruturas topológicas da trajetória original de um sistema é garantida se, e somente se, $m \geq 2d + 1$, onde d é a dimensão do espaço de estados e m a dimensão de imersão. Contudo, nem sempre

se conhece a dimensão d , tornando impossível encontrar um valor adequado de m que valide essa relação. Nesse caso, pode-se recorrer à duas hipóteses: 1 - imersão com base em um diagnóstico conhecido e 2 - investigação de mudanças na vizinhança de pontos no espaço de fase (SOUZA, 2008). Na primeira, aumenta-se a dimensão m a fim de avaliar o comportamento do sistema. O valor escolhido deve corresponder ao menor m para o qual o comportamento do sistema permaneça invariante. Na segunda hipótese, avalia-se a vizinhança de cada ponto ao variar a dimensão m . Dessa forma, os vizinhos verdadeiros permanecerão vizinho mesmo com a variação do m . Portanto, o valor ótimo de m é aquele em que o número de falsos vizinho cai para zero. Essa técnica também é conhecida como falso vizinho mais próximo (FNN - *False Nearest Neighbor*).

Segundo Krakovská, Mezeiová e Budáčová (2015), o método FNN é um dos mais utilizados para encontrar a dimensão de imersão ótima, sendo introduzido primeiramente por Kennel, Brown e Abarbanel (1992). Contudo, o método de Kennel é bastante afetado pela escolha subjetiva de alguns parâmetros. A fim de contornar este problema, Cao (1997) introduziu uma modificação conhecida como média dos falsos vizinhos mais próximos (AFNN - *Averaged False Nearest Neighbors*) ou simplesmente método de Cao. Como explica Krakovská, Mezeiová e Budáčová (2015), em vez de testar se os vizinhos são falsos ou não, Cao calcula como, em média, as distâncias entre os vizinhos mais próximos mudam depois de passar da dimensão m para $m + 1$. Quando não ocorre mais mudança, pode-se dizer que o valor de m é ótimo, já que é possível assumir que a trajetória está totalmente desdobrada e que a adição de outras dimensões não altera a distância média entre os vizinhos mais próximos. A principal vantagem deste método é que o número de limites de parâmetros escolhidos subjetivamente é reduzido (KRAKOVSKÁ; MEZEIOVÁ; BUDÁČOVÁ, 2015).

2.2.3 Passo de Reconstrução

A informação mútua média provê uma medida útil para identificar componentes em amostragem multivariada que parecem estar relacionadas ou independentes, sendo uma importante medida estatística na análise de séries temporais experimentais (MOON; RAJAGOPALAN; LALL, 1995). Dessa forma, pode-se detectar a quantidade de informação presente em uma série temporal com defasagem τ que também está presente num instante de tempo t . Calculada com base na entropia de Shannon, a informação mútua média entre uma série temporal $x(t)$ e sua versão defasada $x(t + \tau)$, com b intervalos, pode ser dada por

$$I(\tau) = \sum_{i=1}^b \sum_{j=1}^b P_{i,j}(x(t), x(t+\tau)) \log_2 \left[\frac{P_{i,j}(x(t), x(t+\tau))}{P_i(x(t))P_j(x(t+\tau))} \right], \quad (3)$$

em que

- $P_i(x(t))$ é a probabilidade de $x(t)$ assumir um valor dentro do i -ésimo intervalo
- $P_j(x(t+\tau))$ é a probabilidade de $x(t+\tau)$ assumir um valor dentro do j -ésimo intervalo
- $P_{i,j}(x(t), x(t+\tau))$ é a probabilidade de que $x(t)$ esteja no i -ésimo intervalo e $(x(t+\tau))$ no j -ésimo intervalo simultaneamente

Dessa forma, o valor ótimo de τ será aquele para o qual a função de informação mútua média atinge seu primeiro mínimo local (FRASER; SWINNEY, 1986).

2.2.4 RP Binários

Os gráficos de recorrência originalmente apresentam-se como uma matriz bidimensional quadrada de ordem N , que permite visualizar o comportamento recorrente de sistemas dinâmicos, independentemente da sua dimensionalidade. Nesta matriz são marcados vários pontos, que podem ser de cor branca ou preta, na qual a recorrência de um estado no instante de tempo i em outro instante de tempo j é indicada por um ponto preto (ECKMANN; KAMPHORST; RUELLE, 1987).

Segundo Marwan (2003), um gráfico de recorrência pode ser definido matematicamente pela Equação 4:

$$R_{i,j}^{m,\varepsilon} = \theta(\varepsilon - \|\vec{\xi}_i - \vec{\xi}_j\|), \quad \vec{\xi}_i \in \mathbb{R}_m, \quad i, j = 1 \dots N \quad (4)$$

Na qual $\theta(\cdot)$ é a função degrau unitário, $\|\cdot\|$ é a distância euclidiana entre dois vetores, e N é o número de estados $\vec{\xi}_i$ formados pelo sistema, obtidos por meio do método dos atrasos temporais (TAKENS, 1981) que imerge a série temporal num espaço m -dimensional com tempo de defasagem τ . O raio de vizinhança, ε , é um parâmetro fundamental na construção do gráfico de recorrência. Ao calcular a distância entre os estados do sistema, o raio de vizinhança define quais pontos estão dentro ou fora dessa vizinhança (pontos recorrentes ou não recorrentes). Nesse contexto, observa-se que, as dimensões finais dos RPs correspondem a uma matriz quadrática de ordem $N \times N$, que obedece à Equação 2.

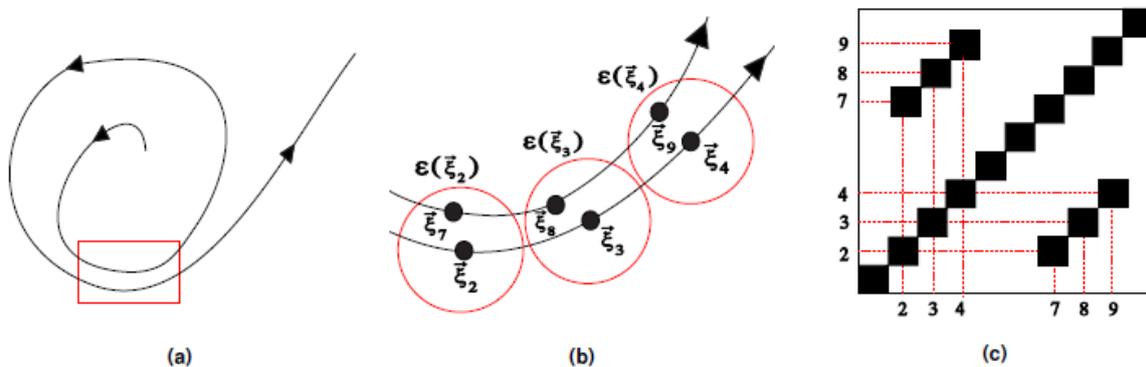
Uma das formas de definir um ε ótimo é fazer com que ele assuma um valor que mantenha uma determinada taxa de recorrência (RR, do inglês, *Recurrence Rate*).

A RR representa simplesmente a densidade de pontos de recorrência em um gráfico de recorrência, podendo ser representada matematicamente pela Equação 5:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}^{m,\varepsilon} \quad (5)$$

Observando ainda a função Equação 4, a função degrau unitário, $\theta(\cdot)$, define um limiar responsável por colocar os pontos recorrentes dentro do gráfico de recorrência, da seguinte forma: caso a operação contida pela função degrau obtenha um valor menor ou igual a 0, $R_{i,j}^{m,\varepsilon}$ assume o valor de 0 indicando que não há recorrência entre os estados. Caso a operação resulte em um valor maior que 0, $R_{i,j}^{m,\varepsilon}$ assume o valor de 1, indicando que há recorrência entre os estados (COSTA et al., 2015). Nesse caso, um ponto preto é marcado na coordenada (i, j) da matriz que representa o gráfico de recorrência. A formação de uma linha no eixo diagonal dos gráficos de recorrência é natural, e indica a recorrência dos estados consigo próprios. Na Figura 4 é apresentado o processo de geração de um gráfico de recorrência, a partir da observação da recorrência dos estados dentro de um raio ε em um sistema dinâmico.

Figura 4 – (a) Espaço de fase em um sistema dinâmico. b) Espaço de fase delimitado por um raio ε . (c) Gráfico de recorrência.



Fonte: Adaptado de Costa et al. (2015)

2.2.5 RP sem limiar

Os gráficos de recorrência sem limiar ou URPs (*unthresholded recurrence plot*) eliminam o uso da função degrau, apresentada na Equação 4, na filtragem da distância entre um estado e os demais. Portanto, um URP basicamente representa a distância entre um estado e todos os outros em um espaço de fases, sendo representado pela relação da Equação 6 (MCGUIRE; AZAR; SHELHAMER, 1997).

$$R_{i,j}^m = \|\vec{\xi}_i - \vec{\xi}_j\|, \vec{\xi}_i \in \mathbb{R}_m, i, j = 1 \dots N \quad (6)$$

Dessa forma, os valores resultantes da Equação 6 deixam de ser binários. Normalmente, mapas de cores são aplicados para uma melhor visualização das estruturas formadas.

2.2.6 ReLU-RP

Proposto por Lee, Lee e Shin (2019), o ReLU-RP (*Rectified Linear Unit based Recurrence Plot*) é um gráfico de recorrência que pode ser obtido ao filtrar um URP com a função ReLU (unidade linear retificada). Matematicamente, é representada pela relação da Equação 7.

$$R_{i,j}^{m,\varepsilon} = \begin{cases} D, & \text{se } D = \|\vec{\xi}_i - \vec{\xi}_j\| \geq \varepsilon \\ 0, & \text{caso contrário} \end{cases}, \quad \in \mathbb{R}_m, \quad i, j = 1 \dots N \quad (7)$$

Nesse caso, D é a distância filtrada pela função ReLU entre dois estados do espaço de fases. Lee, Lee e Shin (2019) apontam que se ε for muito grande, o RP resultante terá uma grande quantidade de pontos com valor zero, resultando na perda de muitas informações do URP. Contudo, um valor muito pequeno de ε tornaria o ReLU-RP muito similar ao URP. Portanto, o valor do limiar deve ser precisamente escolhido para que as informações do gráfico sejam significativas.

Lee, Lee e Shin (2019) ainda reforçam que, o ReLU-RP tem como foco o padrão formado pelas linhas verticais (ou horizontais), enfatizando o período de tempo em que um estado não muda significativamente ao longo do tempo. Por outro lado, o RP binário concentra-se no padrão formado pelas linhas diagonais e enfatiza a repetitividade de uma evolução de estado semelhante na dinâmica do sistema.

2.2.7 FFT-RP

O gráfico de recorrência do espectro de frequência foi proposto por Ladeira et al. (2020), e como sugerido pelo nome, é baseado no domínio da frequência, diferente dos RPs citados anteriormente. No presente trabalho, este tipo de RP será denominado FFT-RP (*Fast Fourier Transform Recurrence Plot*).

Inicialmente, a série temporal é transformada para o domínio da frequência. A partir do espectro obtido, os RPs podem ser calculados por meio da relação da Equação 8.

$$R_{i,j}^{m,\varepsilon} = \theta(\varepsilon - \|f_i - f_j\|), \quad f_i \in \mathbb{R}_m, \quad i, j = 1 \dots N \quad (8)$$

Ladeira et al. (2020) propuseram a geração de dois RPs para um mesmo sinal, considerando um valor baixo e um alto de ε (0,4 e 0,95 foram utilizados). Posterior-

mente os pontos das matrizes são subtraídos, originando uma matriz com pontos entre os dois limiares. Por fim, a matriz resultante consiste de pontos significativos do sinal. Isso ocorre porque um baixo valor de ε elimina ruído por meio da remoção dos pontos com amplitude muito baixa e um alto valor de ε elimina as amplitudes muito altas referentes à perturbação.

2.3 ANÁLISE DA FORMAÇÃO DE ESTRUTURAS NOS GRÁFICOS DE RECORRÊNCIA

Para uma melhor compreensão da formação de estruturas nos gráficos de recorrência, foram escolhidos alguns sinais comuns na literatura para análise. Além disso também foram demonstrados a formação dos gráficos de recorrência abordados neste estudo para os sinais de áudio de pacientes saudáveis, afetados por paralisia, edema de Reinke e nódulos nas pregas vocais.

2.3.1 Formação de RPs para um sinal constante

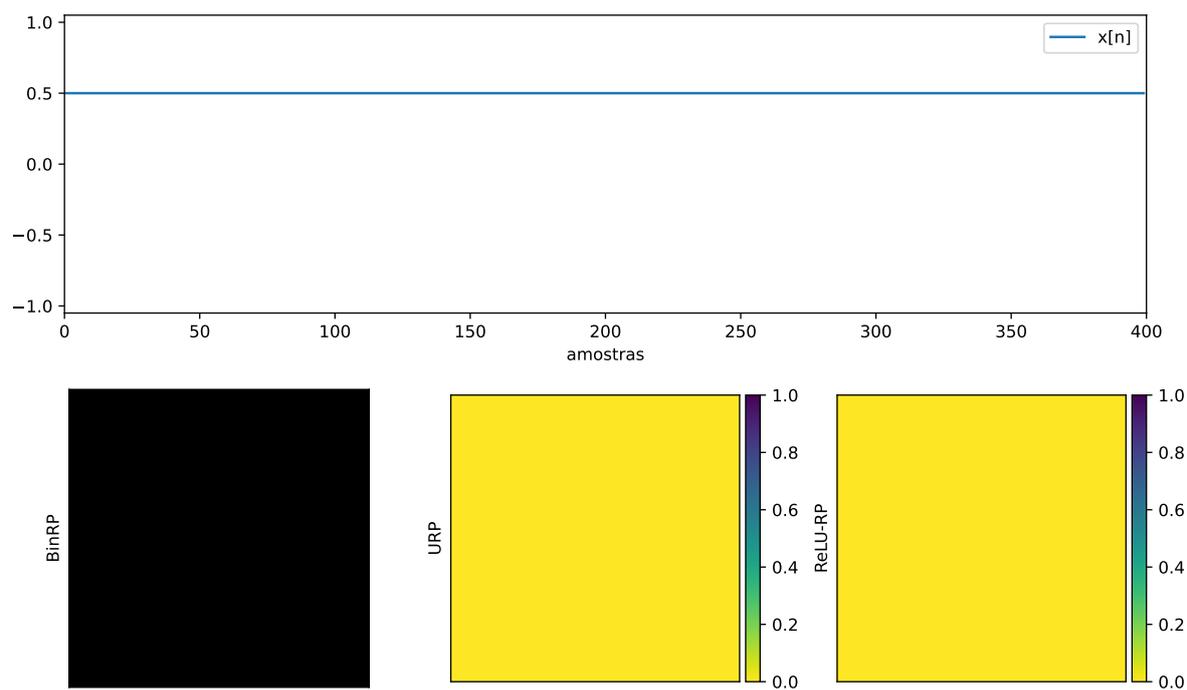
Na Figura 5 são apresentados os gráficos de recorrência para um sinal constante em 0,5 de 400 amostras e taxa de amostragem de 25.000 amostras/s. Um sinal constante, por natureza, tem recorrência com ele mesmo independente dos valores de τ , m e ε . Para gerar o RP-binário foram utilizados $\tau = 1$ e $m = 2$, valores mínimos para se obter um espaço de estados. Esses mesmos valores também foram usados para gerar o URP e ReLU-RP correspondentes, apresentados na sequência. O raio de vizinhança aplicado para o RP-binário foi de $\varepsilon = 1$. Para esse cenário, não faz sentido calcular um ε ótimo por meio da taxa de recorrência (RR), uma vez que a RR é sempre máxima (100%) independente do valor de ε escolhido. Dessa forma, o RP-binário do sinal de uma constante terá todos os pontos pretos.

Já para a geração do URP não se calcula o ε , dado que este tipo de gráfico representa uma matriz de distância entre um estado e os demais. Dessa forma, nenhuma estrutura é formada no gráfico de recorrência dado que a distância entre os estados em qualquer momento é nula. No mapa de cores escolhido para representar o URP e o Relu-RP, os tons amarelados representam as menores distâncias e os tons roxos representam as maiores distâncias.

De forma similar ao RP-binário, no ReLU-RP a distância entre um estado e os demais será nula para qualquer valor de ε .

Na Figura 6 observa-se a transformada de Fourier de um sinal constante em 0,5, cuja transformada é representada por uma constante em 0, dado que para esta análise desconsideramos a componente DC do espectro. Dessa forma, como o FFT-RP é

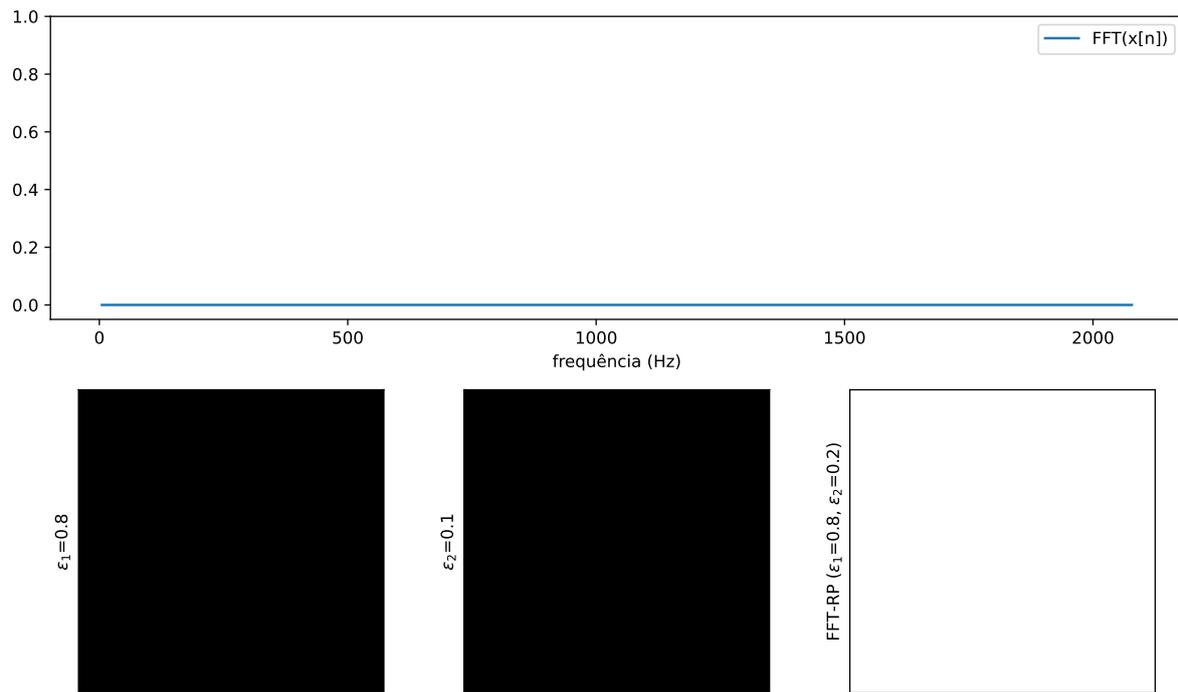
Figura 5 – Sinal constante.



Fonte: autoria própria.

formado pela subtração de um FFT-RP com ε mais alto e um FFT-RP com ε mais baixo, o gráfico final será branco, não havendo recorrência a ser destacada. Isso ocorre porque, para qualquer valor de ε escolhido, haverá recorrência entre um estado e os demais.

Figura 6 – Sinal constante (FFT).



Fonte: autoria própria.

2.3.2 Formação de RPs para sinais senoidais

Na Figura 7 é apresentado um sinal senoidal de 200 Hz, sendo 400 amostras com taxa de amostragem de 25.000 amostras/s. Como o sinal senoidal representa um sistema oscilante, diferentemente do sinal constante, haverá formação de estruturas nos RPs. Para a obtenção dos RPs-binário, URP e ReLU-RP foram aplicados o método de Cao (1997) para encontrar a dimensão de imersão m ótima e o método Fraser e Swinney (1986) para encontrar o passo de reconstrução τ ótimo.

O cálculo do ε ótimo para gerar o RP-binário foi feito com base na taxa de recorrência (RR). Contudo, a fim de demonstrar o efeito da variação do valor do ε , foram gerados dois RPs-binário, com dois valores de ε diferentes, um com base em uma taxa de recorrência RR=5% e outro com RR=20%. Em ambos, observa-se a formação de linhas diagonais totalmente preenchidas, que por sua vez, são paralelas à diagonal principal. Contudo, a quantidade de recorrências apresentadas no gráfico cuja RR=20% é maior que no gráfico cujo RR=5%. Isso porque é necessário um ε maior para que haja mais recorrência representada no RP-binário. A projeção bidimensional do atrator do sinal senoidal de 200 Hz apresentado na Figura 8 pode colaborar na compreensão da formação das estruturas, contudo, por ter apenas duas dimensões, pode não demonstrar a profundidade das dimensões de imersão para valores maiores de m .

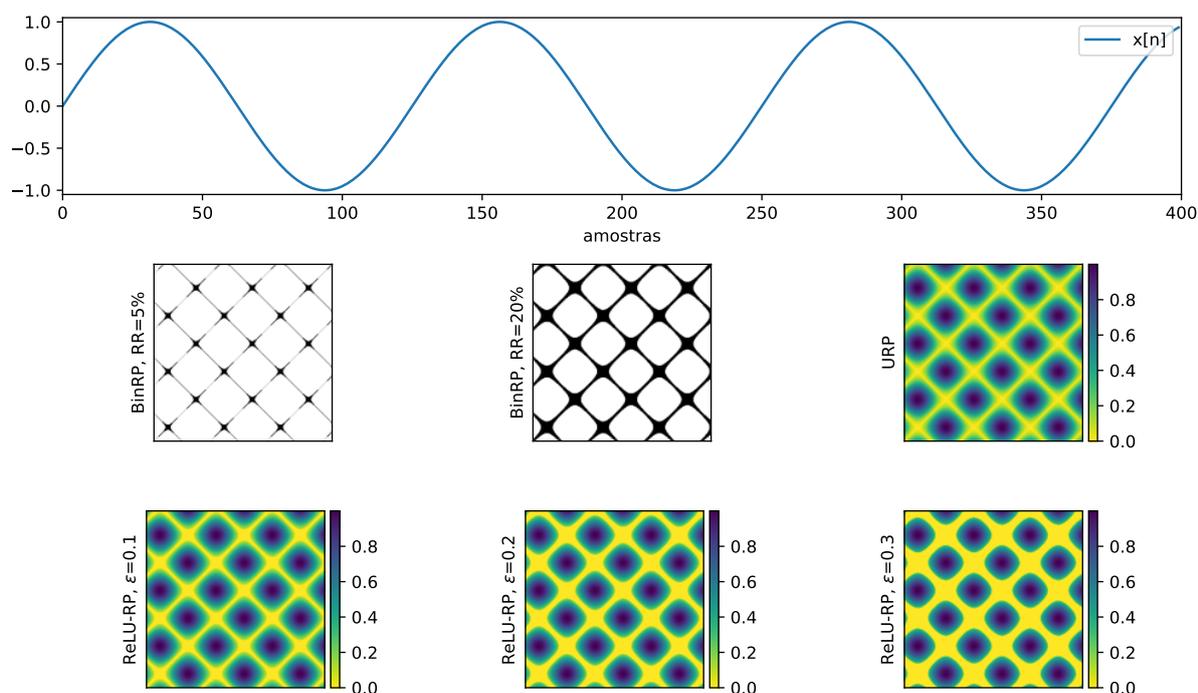
Diferentemente do URP do sinal constante, no URP de um sinal senoidal é possível ver a formação de estruturas bem nítidas. Observa-se que as diagonais principais e as diagonais paralelas a ela representam a distância mínima entre um estado e os demais.

O ReLU-RP, por sua vez, destaca as estruturas nas quais a recorrência é mínima, ou seja, as áreas do URP nas quais a distância entre os estados é maior que um limiar ε . Para que seja possível observar os efeitos da variação desse limiar, foram gerados ReLU-RPs com três valores de ε distintos. Para o ReLU-RP com $\varepsilon = 0,1$ observa-se uma suave atenuação dos pontos que representam distâncias pequenas entre os estados. Já para o ReLU-RP com $\varepsilon = 0,3$, observa-se uma forte atenuação das distâncias menores entre os estados do sistema se comparado ao URP.

Na Figura 9 é apresentado o espectro da frequência do sinal senoidal de 200 Hz. A metodologia para gerar o FFT-RP é similar ao método do RP-binário, contudo, utilizando o sinal no espectro da frequência. Também foram utilizados o método de Cao (1997) para encontrar a dimensão de imersão m ótima e o método Fraser e Swinney (1986) para encontrar o passo de reconstrução τ ótimo. Aplicando a metodologia de Ladeira et al. (2020), utiliza-se um valor de ε mais alto e um mais baixo,

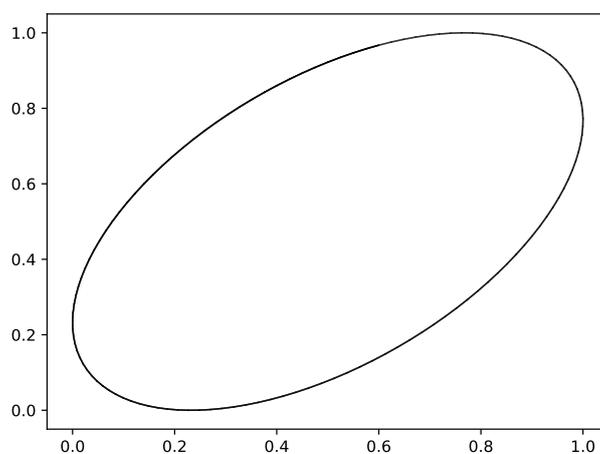
subtraindo os gráficos resultantes. Contudo, para um sinal composto por uma única harmônica, sem presença de ruído, qualquer valor de ϵ irá evidenciar a não recorrência dessa harmônica. Dessa forma, aplicando a subtração, o resultado final seria um gráfico sem recorrências. Na Figura 10 é apresentada a projeção bidimensional do atrator para o espectro da frequência de um sinal senoidal de 200 Hz.

Figura 7 – Sinal senoidal 200 Hz e seus respectivos RPs-binários, URP e ReLU-RPs.



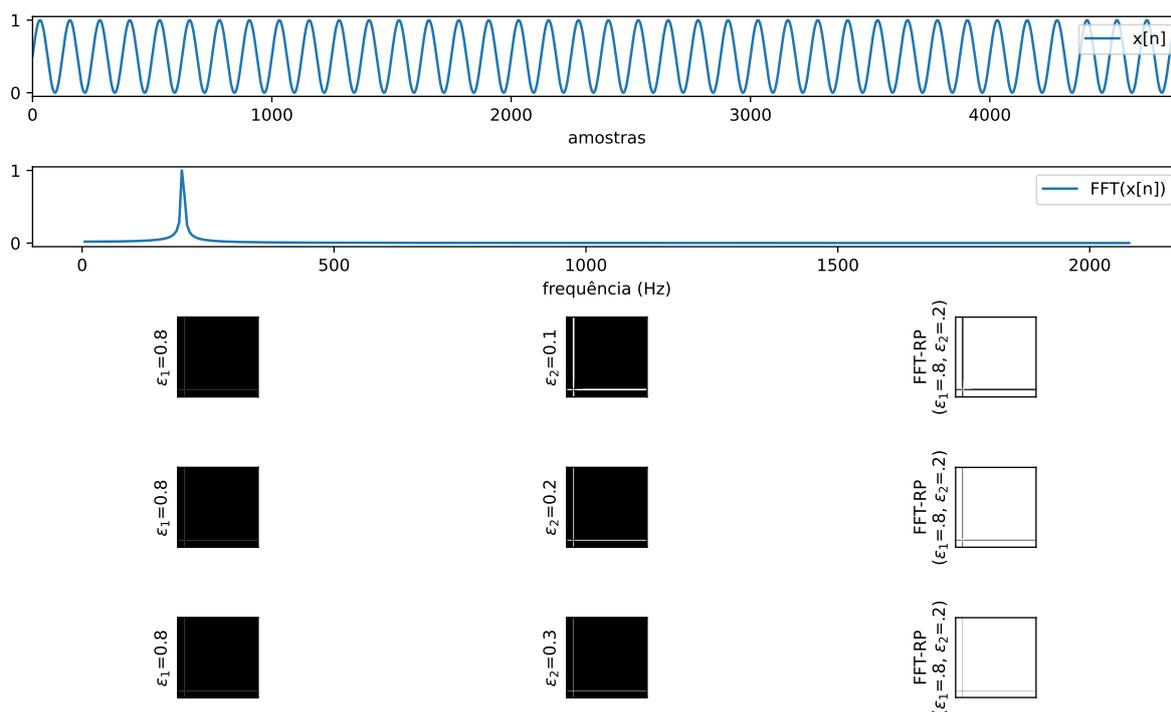
Fonte: autoria própria.

Figura 8 – Projeção bidimensional do atrator de um sinal senoidal de 200 Hz.



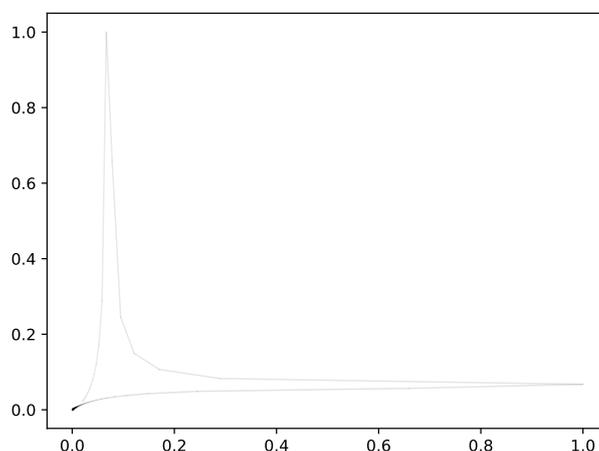
Fonte: autoria própria.

Figura 9 – FFT do sinal senoidal 200 Hz e suas respectivas variações de FFT-RPs.



Fonte: autoria própria.

Figura 10 – Projeção bidimensional do atrator do espectro da frequência um sinal senoidal de 200 Hz.



Fonte: autoria própria.

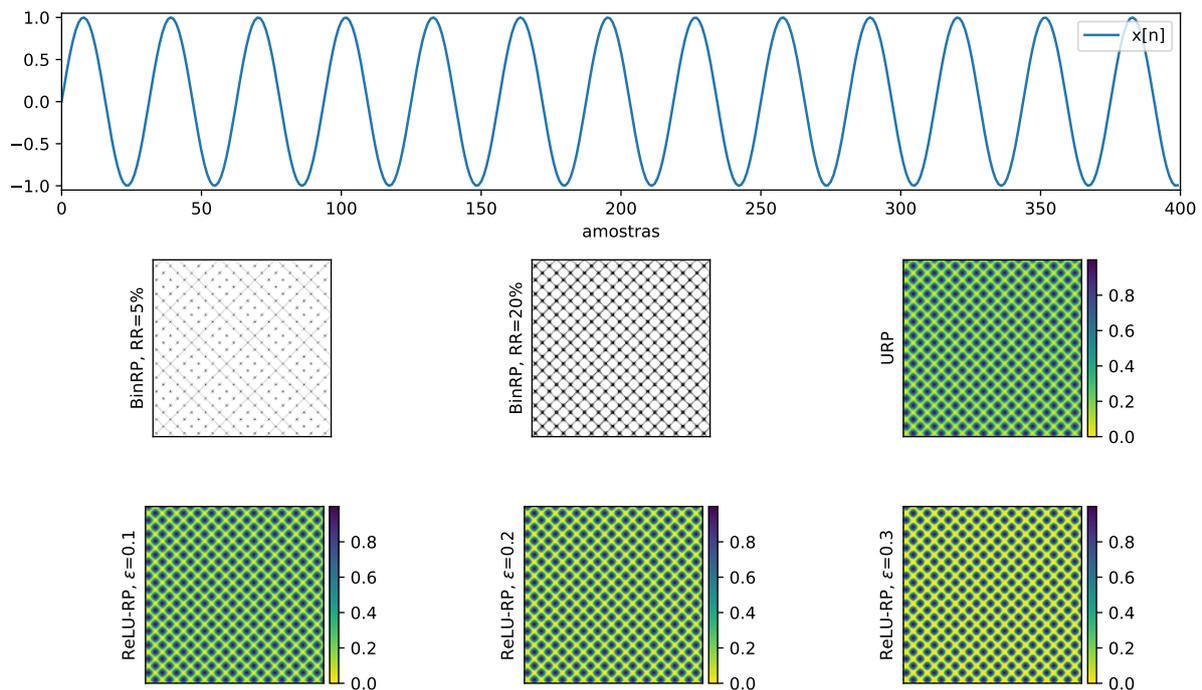
Em sinais periódicos é possível observar a relação entre taxa de amostragem, frequência do sinal e distância entre as linhas diagonais a partir da diagonal principal para cima (eixo vertical) ou para o lado (eixo horizontal) (FACCHINI; KANTZ, 2007). Essa relação matemática pode ser caracterizada pela Equação 9, onde D se refere à distância entre duas linhas diagonais, f_s se refere à taxa de amostragem do sinal

em Hertz, e f à frequência do sinal em Hertz. Para interpretação do valor de D obtido, a distância entre as linhas diagonais deve ser estimada em quantidade de estados necessários para a repetição de um ciclo.

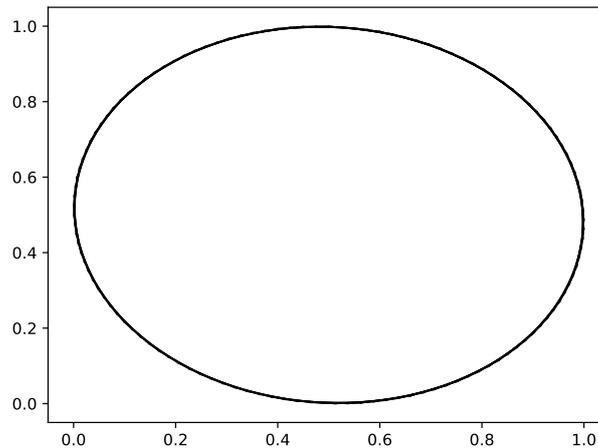
$$D = \frac{f_s}{f} \tag{9}$$

Para ilustrar o efeito da variação na estrutura de um gráfico de recorrência de acordo com a variação da frequência de um sinal periódico, pode-se comparar os RPs do sinal senoidal de 200 Hz apresentado na Figura 7, com o sinal senoidal de 800 Hz apresentado na Figura 11. Considerando que os dois sinais possuem uma taxa de amostragem igual a 25.000 amostras/s, para o sinal senoidal de 200 Hz D é igual a 120, e para o sinal senoidal de 800 Hz D é aproximadamente 31. A variação provocada nos RPs fica evidente quando observa-se a distância das linhas diagonais, que se aproximam para o sinal senoidal de 800 Hz e se distanciam para o sinal senoidal de 200 Hz. A projeção bidimensional do atrator do sinal senoidal de 800 Hz está apresentada na Figura 12.

Figura 11 – Sinal senoidal 800 Hz e seus respectivos RPs-binários, URP e ReLU-RPs.



Fonte: autoria própria.

Figura 12 – Projeção bidimensional do atrator de um sinal senoidal de 800 Hz.

Fonte: autoria própria.

Na Figura 13 é apresentado o espectro da frequência de um sinal senoidal de 800 Hz. De forma similar ao que ocorre no espectro da frequência do sinal senoidal de 200 Hz, ilustrado anteriormente, qualquer valor de ε evidenciará a não recorrência desta harmônica. Dessa forma, o resultado final, após a subtração dos dois gráficos, não apresenta recorrência. Na Figura 14 é apresentada a projeção bidimensional do atrator do espectro da frequência um sinal senoidal de 800 Hz.

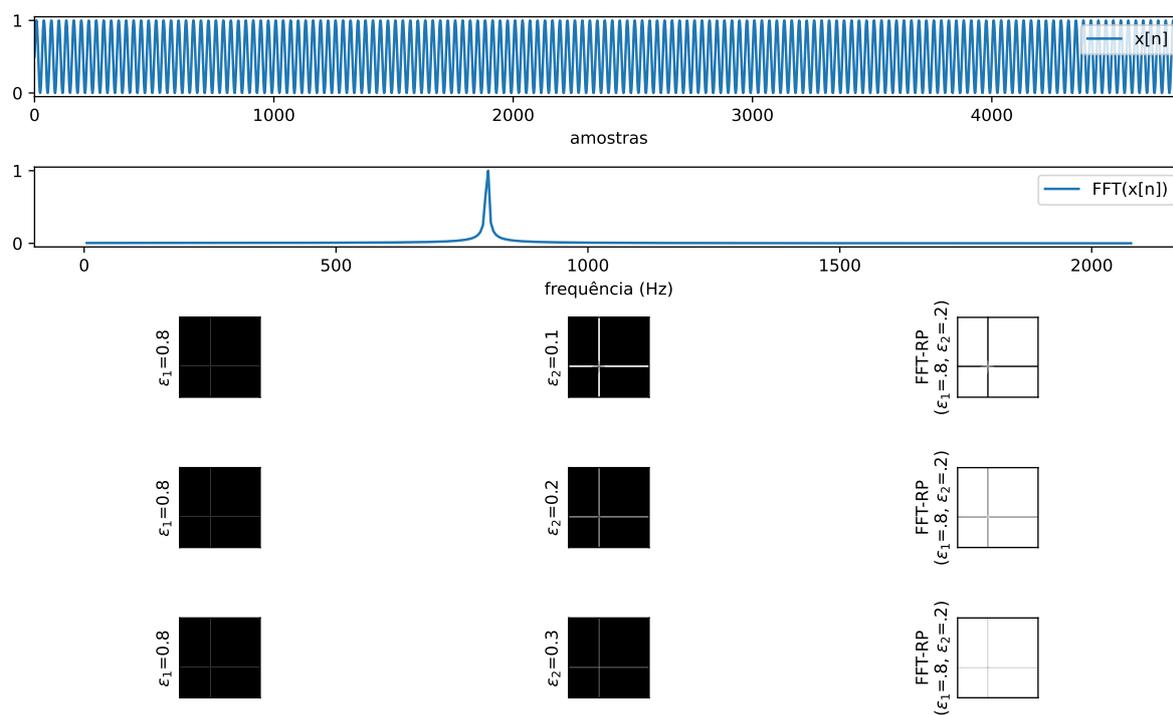
Analisando e comparando os gráficos de recorrência da Figura 13 e da Figura 9, é possível constatar que há uma relação direta entre a posição das linhas horizontais e verticais dos gráficos e a frequência das harmônicas. Nestes casos, para qualquer valor de ε (antes da subtração dos gráficos), observa-se que, quanto maior a frequência, mais para esquerda localiza-se a linha vertical, e mais para cima localiza-se a linha horizontal nos gráficos.

Na Figura 15 são exibidos os gráficos de recorrência de uma série temporal gerados a partir da soma de um sinal de 200 Hz e de 800 Hz, seguindo a Equação 10 (amostrado a uma frequência $f_s=25$ kHz).

$$x[n] = \sin(2\pi \cdot n \cdot 200/f_s) + 0,25 \cdot \sin(2\pi \cdot n \cdot 800/f_s) \quad (10)$$

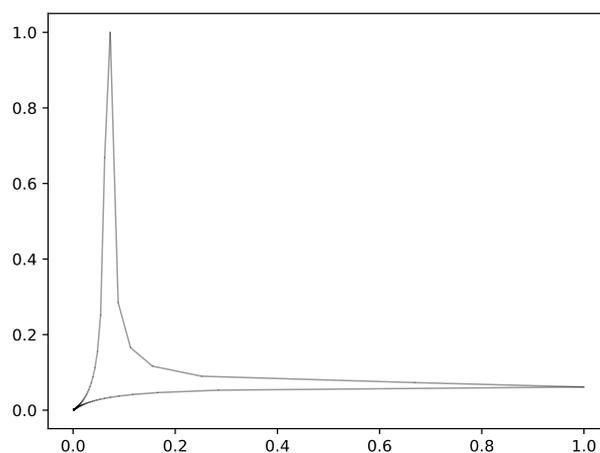
Observa-se que a relação apresentada pela Equação 9 se mantém. A distância entre duas linhas diagonais permanece evidenciada pela harmônica de menor frequência (200 Hz). As harmônicas de maior frequência, por sua vez, caracterizam as estruturas formadas ao longo das diagonais principais, que surgem bastante destacadas nos RPs binários. Na Figura 16 é apresentada a projeção bidimensional do atrator

Figura 13 – FFT do sinal senoidal de 800 Hz e suas respectivas variações de FFT-RPs.



Fonte: autoria própria.

Figura 14 – Projeção bidimensional do atrator do espectro da frequência um sinal senoidal de 800 Hz.

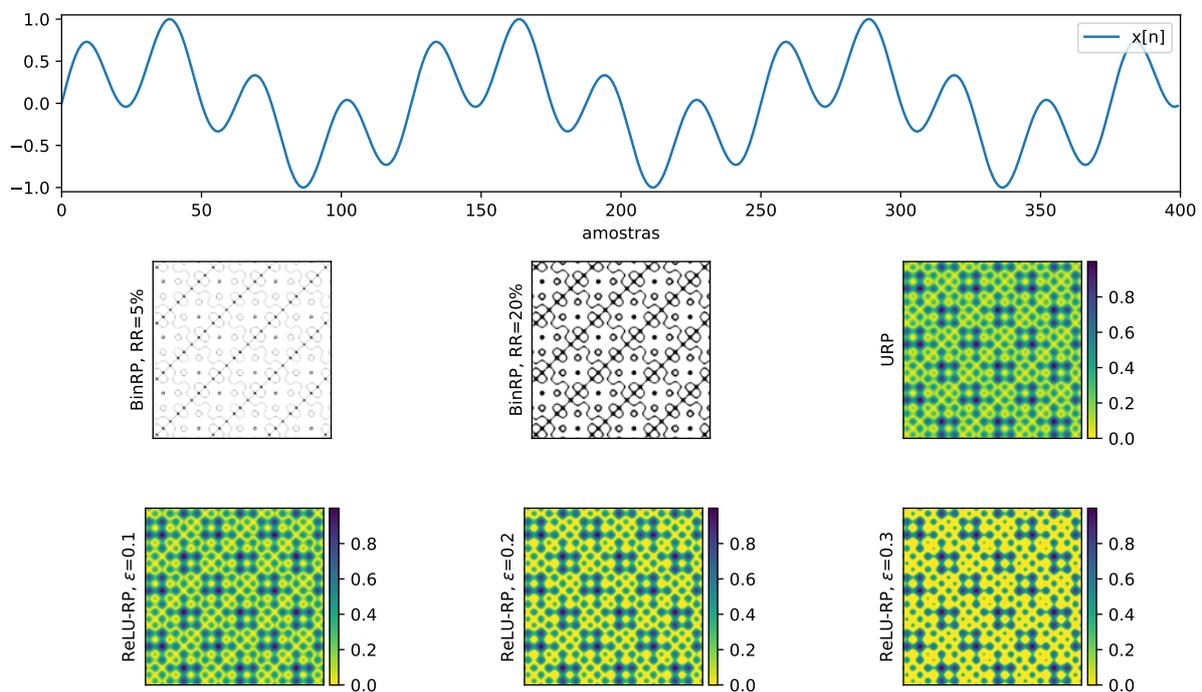


Fonte: autoria própria.

de um sinal senoidal de 200 Hz somado a um sinal senoidal de 800 Hz.

Já nos URPs é possível notar a variação da distância entre um estado e todos os outros. Por meio do mapa de cores aplicado nesses gráficos, destaca-se uma maior variação da distância entre os estados quando comparado ao URP da Figura 7 e Figura 11. Nos ReLU-RPs, à medida que se aumenta o valor de ε , aumenta-se o destaque para as regiões nas quais a recorrência é mínima, deixando evidente a contribuição das variações adicionadas pela harmônica de maior frequência. Essa variação pode ser facilmente percebida quando comparada aos ReLU-RPs da Figura 7.

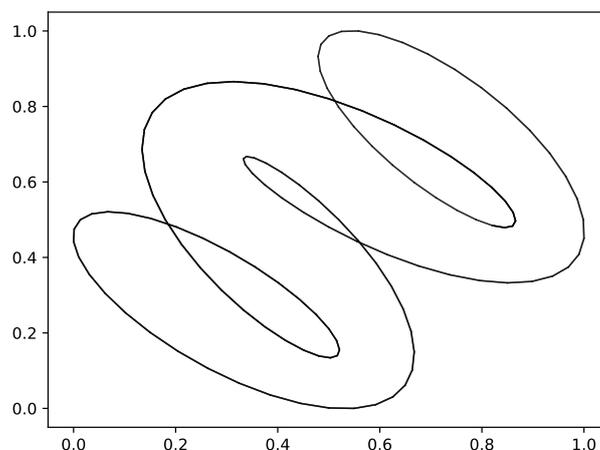
Figura 15 – Sinal senoidal de 200 Hz somado a um sinal senoidal de 800 Hz e seus respectivos RPs-binários, URP e ReLU-RPs.



Fonte: autoria própria.

Com o intuito de exemplificar a eliminação da influência de amplitudes muito altas, três diferentes FFT-RPs foram gerados para o espectro desse sinal, sendo apresentados na Figura 18. Na Figura 17 é apresentada a projeção bidimensional do atrator do espectro da frequência de um sinal senoidal de 200 Hz somado a um sinal senoidal de 800 Hz. Dado que para a geração dos FFT-RPs é necessário subtrair um RP com ε maior de outro com ε menor, o RP com maior valor de ε foi mantido em $\varepsilon_1 = 0,8$ e os de menor valor foram variados entre $\varepsilon_2 = 0,1$, $\varepsilon_2 = 0,2$ e $\varepsilon_2 = 0,3$. Observa-se que, enquanto o RP gerado com $\varepsilon_1 = 0,8$ registra apenas a harmônica de menor frequência e maior amplitude, o RP gerado com $\varepsilon_2 = 0,1$ registra também a harmônica de maior frequência e menor amplitude. Com isso, o FFT-RP gerado registra apenas o ponto de menor amplitude do espectro. Para os RPs gerados com

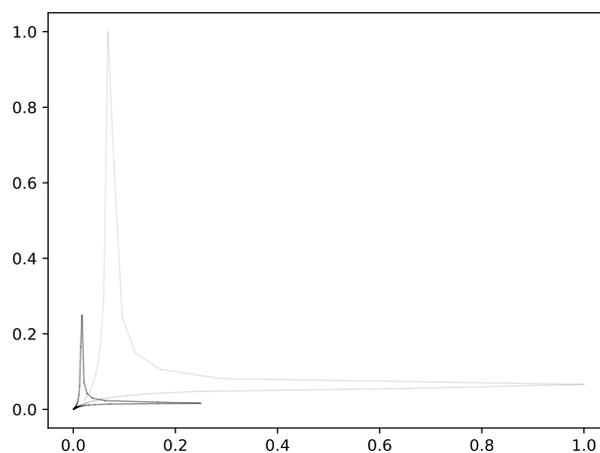
Figura 16 – Projeção bidimensional do atrator de um sinal senoidal de 200 Hz somado a um sinal senoidal de 800 Hz.



Fonte: autoria própria.

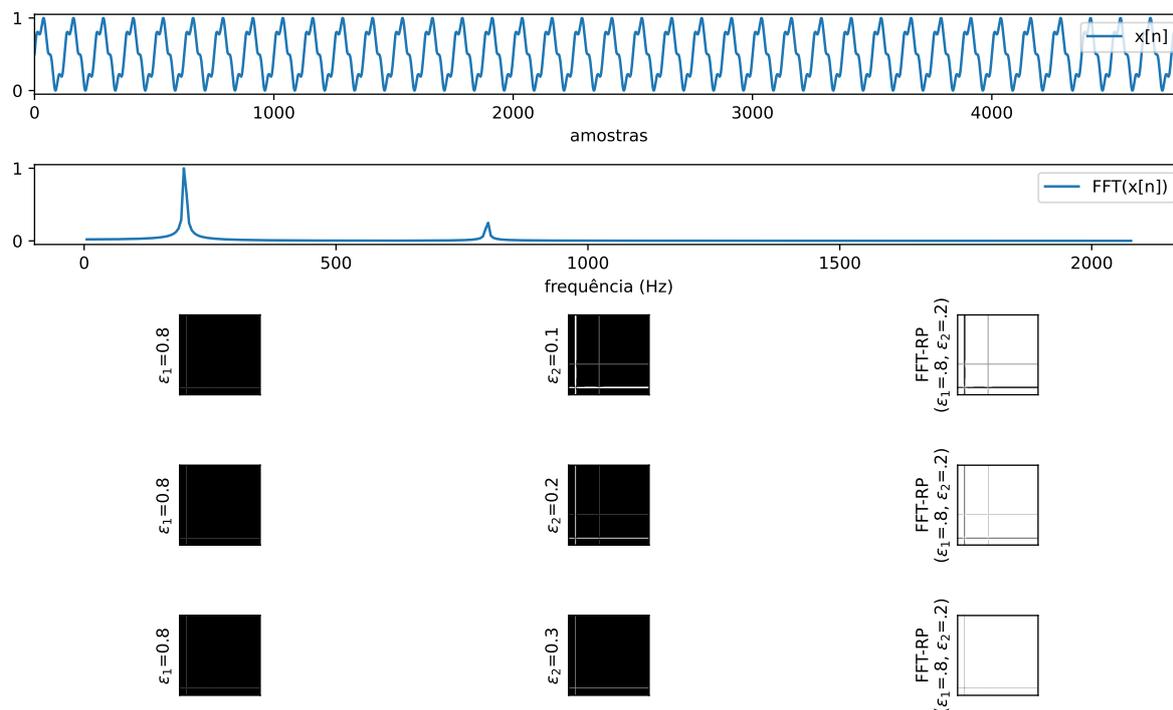
os demais valores de ε_2 , observa-se uma perda gradativa na informação referente à harmônica de 800 Hz. Isso ocorre porque um baixo valor de ε elimina ruído por meio da remoção dos pontos com amplitude muito baixa e um alto valor de ε elimina as amplitudes muito altas referentes à perturbação. Dessa forma, é importante evidenciar a importância de escolher corretamente os valores de ε de acordo com as informações de interesse, de forma que a matriz resultante consista de pontos significativos do sinal.

Figura 17 – Projeção bidimensional do atrator do espectro da frequência um sinal senoidal de 200 Hz somado a um sinal senoidal de 800 Hz.



Fonte: autoria própria.

Figura 18 – FFT do sinal senoidal de 200 Hz somado ao sinal senoidal de 800 Hz e suas respectivas variações de FFT-RPs.



Fonte: autoria própria.

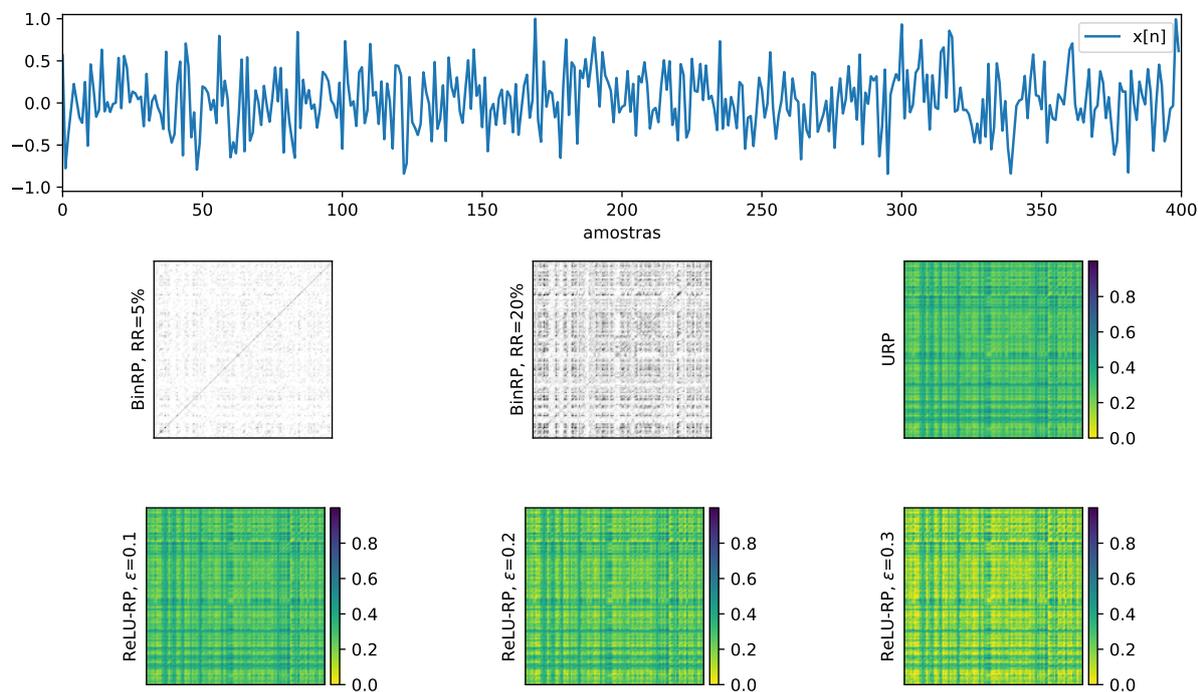
2.3.3 Formação de RPs para ruído branco

Os sinais podem ser afetados por diferentes tipos de ruídos, sendo o ruído branco um dos mais comuns. O ruído branco é caracterizado por conter todas as frequências no espectro do som audível em igual medida. A fim de compreender as estruturas formadas nos gráficos de recorrência por esse tipo de sinal, na Figura 19 é ilustrado um ruído branco com 400 amostras e taxa de amostragem de 25.000 amostras/s, seguido dos seus respectivos RPs-binário, URPs e ReLU-RPs. Nota-se, principalmente no RP-binário com RR=5%, a formação de uma única linha diagonal bem definida, sendo a diagonal principal, que indica a recorrência do estado com ele mesmo. As demais formações não denotam linhas diagonais em nenhum dos tipos de RPs gerados, formando apenas estruturas descontínuadas ao longo do gráfico. Na Figura 20 é apresentada a projeção bidimensional do espaço de estados de um sinal do tipo ruído branco.

Os FFT-RPs do ruído branco no domínio da frequência também denotam padrões desorganizados e são ilustrados na Figura 21. Ao contrário dos FFT-RPs dos sinais senoidais demonstrados anteriormente, o FFT-RP do sinal ruidoso é completamente preenchido por ter frequências espalhadas em todo espectro audível. Na Figura 22 é projeção bidimensional do espaço de estados do espectro da frequência

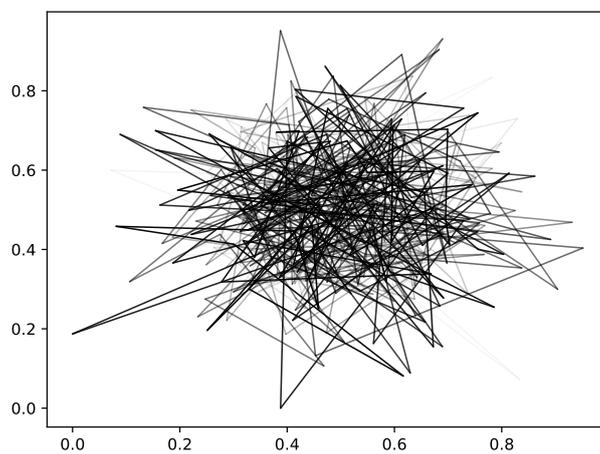
de um ruído branco.

Figura 19 – Ruído branco e seus respectivos RPs-binários, URP e ReLU-RPs.



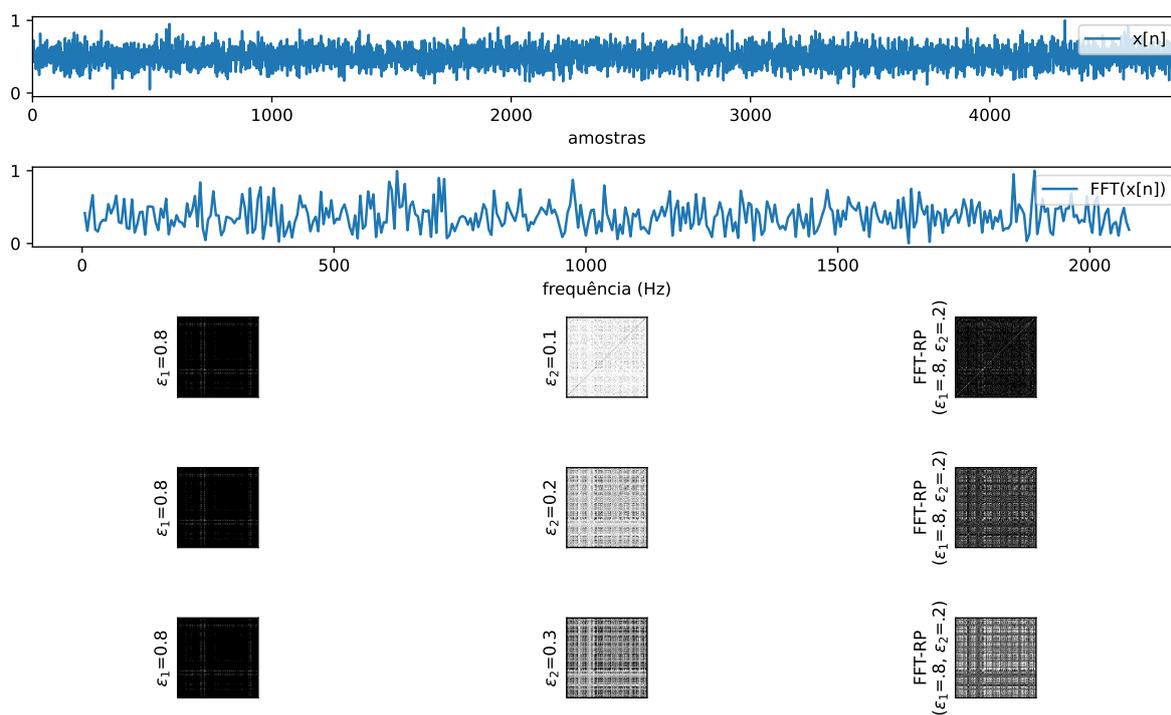
Fonte: autoria própria.

Figura 20 – Projeção bidimensional do espaço de estados de um sinal do tipo ruído branco.



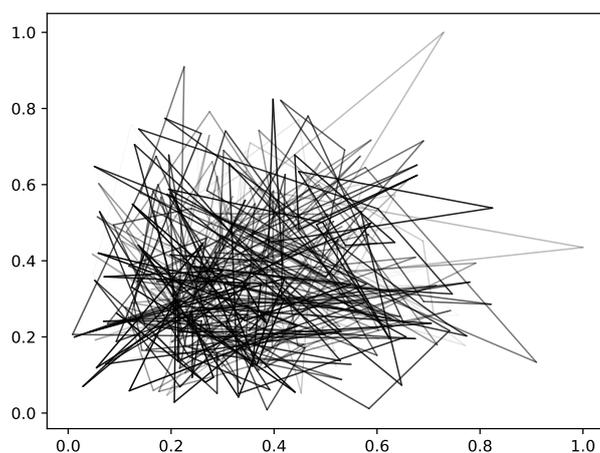
Fonte: autoria própria.

Figura 21 – FFT do sinal do tipo ruído branco e suas respectivas variações de FFT-RPs.



Fonte: autoria própria.

Figura 22 – Projeção bidimensional do espaço de estados do espectro da frequência de um sinal do tipo ruído branco.



Fonte: autoria própria.

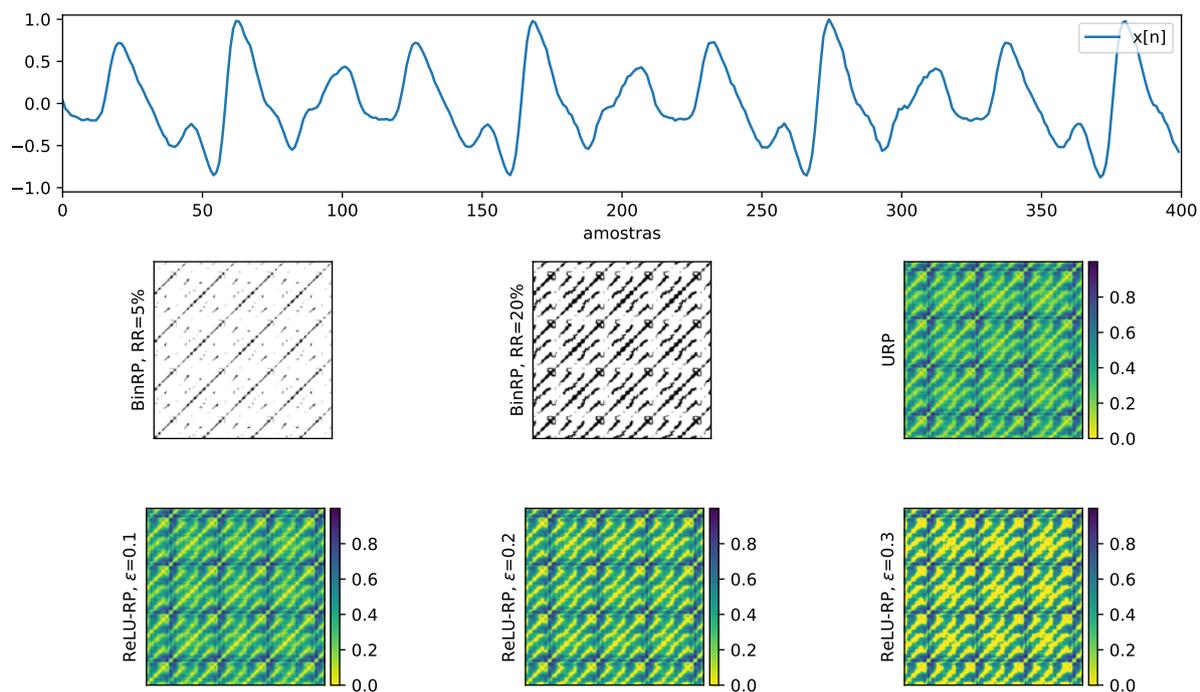
2.3.4 Gráficos de recorrência de sinais de voz

Neste estudo estão sendo utilizados sinais de voz gravados pelo Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab (KAY ELEMETRICS, 1994), consistindo em amostras correspondentes à vogal /a/ sustentada de 165 vezes de pessoas distintas, sendo 53 vozes normais, 51 com paralisia das pregas vocais, 43 com edema de Reinke e 18 com nódulos nas pregas vocais. Todos os sinais foram re-amostrados à taxa de 25.000 amostras/s, com resolução de 16 bits/amostra. Para gerar os RPs-binários, os URPs e os ReLU-RPs foram utilizados segmentos de 400 amostras. Já para gerar os FFT-RPs foram utilizados segmentos de 4800 amostras nos quais, após a obtenção do módulo da transformada de Fourier, foram ignoradas a componente DC e a parte espelhada do espectro. De um espectro até 12 kHz, foi considerada apenas a faixa entre 0 e 2 kHz. Ao final, de forma similar aos demais RPs, foram utilizadas 400 amostras, agora de pontos significativos da transformada de Fourier do sinal.

Na Figura 23 é apresentada a série temporal de um sinal de voz obtido de um paciente saudável. Observa-se que o sinal de voz apresenta certa regularidade dado a sua quase periodicidade. Nos RP-binários é possível notar a formação de diagonais principais com estruturas paralelas. Nos ReLU-RPs, mesmo para o maior valor de ε , observa-se uma distribuição equilibrada entre momentos de alta e baixa recorrência. Na Figura 24 é apresentada a projeção bidimensional do atrator de um sinal de voz saudável proveniente da base de sinais Kay Elemetrics (1994).

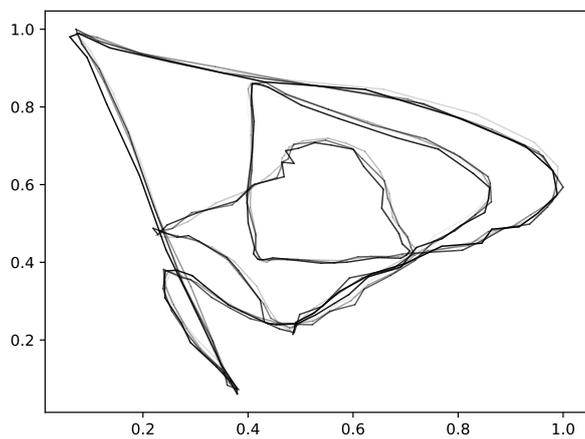
Na Figura 25 é apresentada a transformada de Fourier gerada a partir de um segmento de 4800 amostras obtidas a partir da mesma série temporal da Figura 23. Para um sinal de voz saudável, observa-se que as harmônicas com maior energia estão localizadas em 750 Hz, havendo outras formantes de menor amplitude que contribuem para o sinal. Os FFT-RPs gerados a partir desse espectro evidenciam a distribuição dessas harmônicas, havendo uma distância similar entre elas. Na Figura 26 é apresentada a projeção bidimensional do atrator do espectro da frequência de um sinal de voz saudável proveniente da base de sinais Kay Elemetrics (1994).

Figura 23 – Sinal de voz saudável (identificado por AXH125 na base Kay Elemetrics (1994)) e seus respectivos RPs-binários, URP e ReLU-RPs.



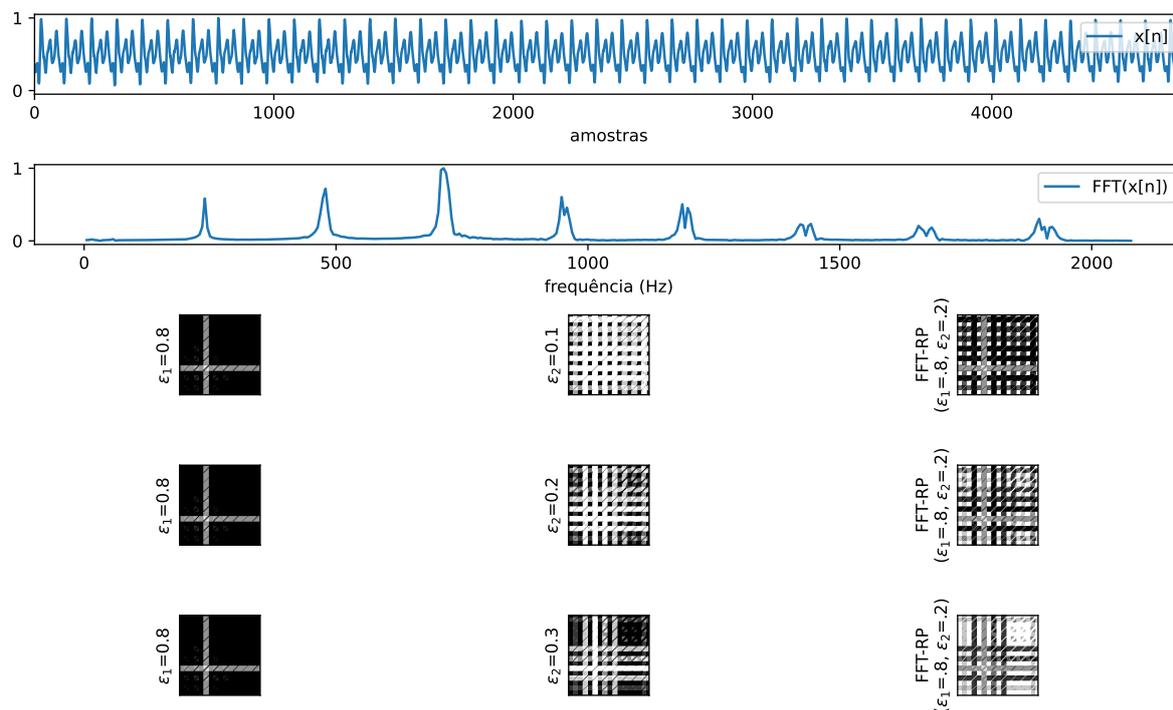
Fonte: autoria própria.

Figura 24 – Projeção bidimensional do atrator do sinal de voz saudável (identificado por AXH125 na base Kay Elemetrics (1994)).



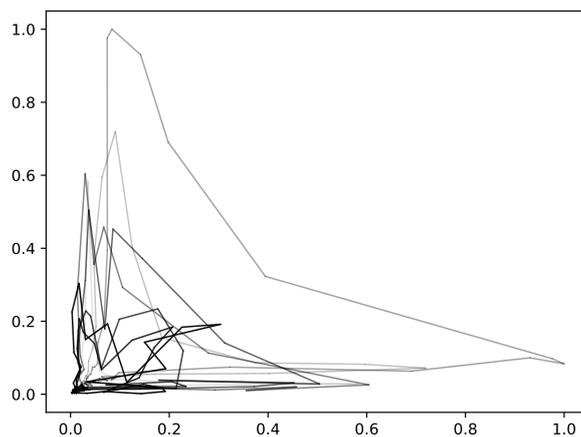
Fonte: autoria própria.

Figura 25 – Espectro de frequência de sinal de voz saudável (identificado por AXH125 na base Kay Elemetrics (1994)) e suas respectivas variações de FFT-RPs.



Fonte: autoria própria.

Figura 26 – Projeção bidimensional do atrator do espectro da frequência do sinal de voz saudável (identificado por AXH125 na base Kay Elemetrics (1994)).

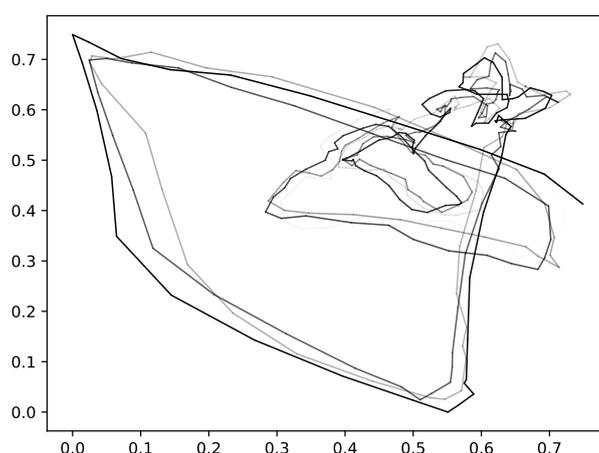


Fonte: autoria própria.

Na Figura 28 é apresentada a série temporal de um sinal de voz obtido de um paciente cuja voz é afetada pela presença de nódulos bilaterais na laringe. Comparada à voz saudável, a voz afetada por nódulos possui um período mais longo, impactando diretamente na formação das estruturas dos gráficos de recorrência. Nos RP-binários, nota-se que as diagonais principais possuem uma maior distância entre si, e que as estruturas formadas ao seu redor estão inteiramente conectadas a elas. Já nos ReLU-RPs, quanto maior o ε , maior a região de não recorrência do gráfico. Comparando o ReLU-RP com $\varepsilon = 0,3$ com o ReLU-RP com mesmo valor de ε da voz saudável, observa-se que a área de não recorrência é bem maior para o sinal de voz afetada pela presença de nódulos. Na Figura 26 é apresentada a projeção bidimensional do atrator do sinal do paciente cuja voz é afetada pela presença de nódulos bilaterais na laringe. Na Figura 24 é apresentada a projeção bidimensional do atrator de um sinal de voz afetada por nódulo proveniente da base de sinais Kay Elemetrics (1994).

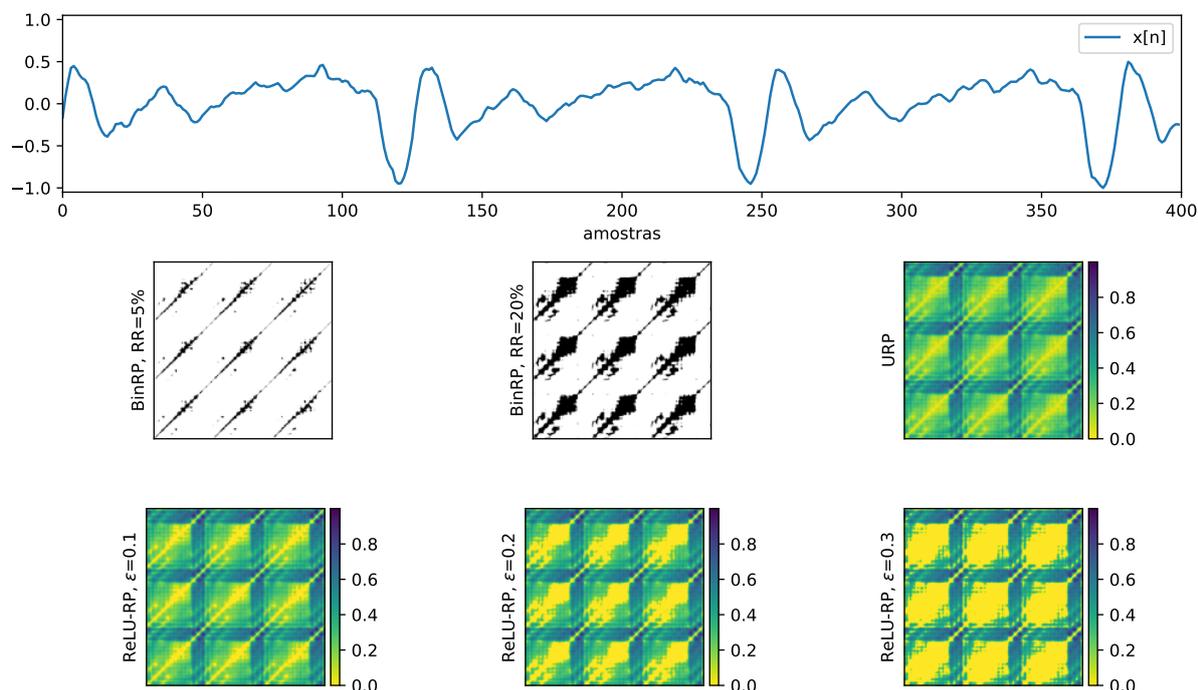
Na Figura 29 é apresentada a transformada de Fourier obtida do mesmo sinal de voz afetada pela presença de nódulos na laringe da Figura 28. Neste sinal, observa-se que, ao contrário da voz saudável, a distribuição das harmônicas é menos regular, além de uma maior variação da amplitude. Com isso, os FFT-RPs obtidos por meio do espectro deste sinal possuem uma distribuição menos regular, quando comparado aos FFT-RPs do sinal de voz saudável. Na Figura 30 é apresentada a projeção bidimensional do atrator do espectro da frequência de um sinal de voz afetada por nódulo da base de sinais Kay Elemetrics (1994).

Figura 27 – Projeção bidimensional do atrator do sinal de voz afetada por nódulos bilaterais (identificado por ANB28AN na base Kay Elemetrics (1994)).



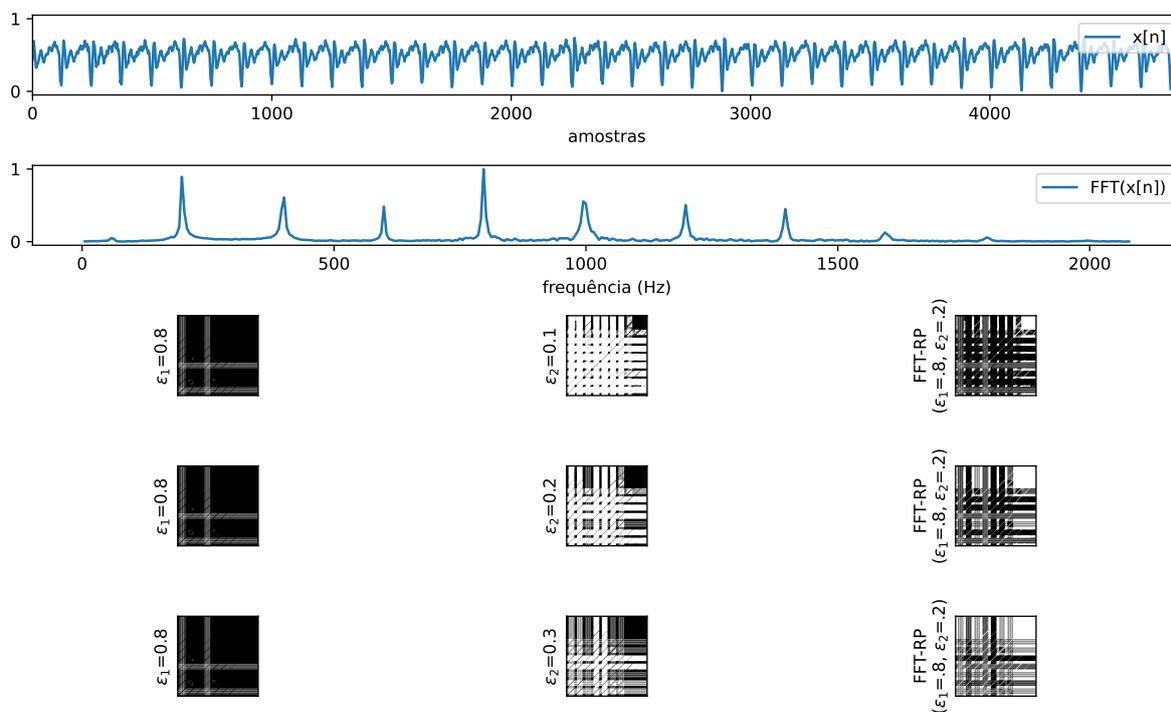
Fonte: autoria própria.

Figura 28 – Sinal de voz afetado por nódulos bilaterais (identificado por ANB28AN na base Kay Elemetrics (1994)) e seus respectivos RPs-binários, URP e ReLU-RPs.



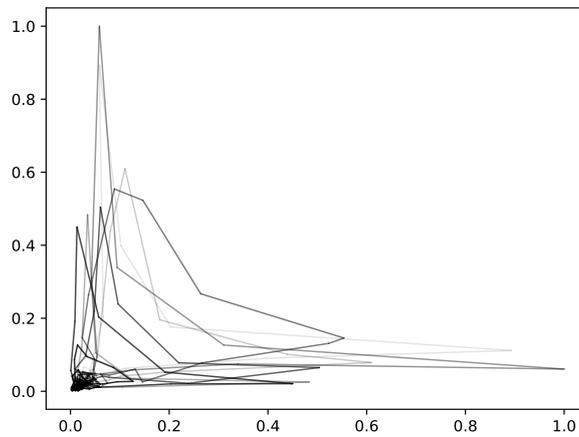
Fonte: autoria própria.

Figura 29 – Espectro de frequência de sinal de voz afetado por nódulos bilaterais (identificado por ANB28AN na base Kay Elemetrics (1994)) e suas respectivas variações de FFT-RPs.



Fonte: autoria própria.

Figura 30 – Projeção bidimensional do atrator do espectro da frequência do sinal de voz afetada por nódulos bilaterais (identificado por ANB28AN na base Kay Elemetrics (1994)).

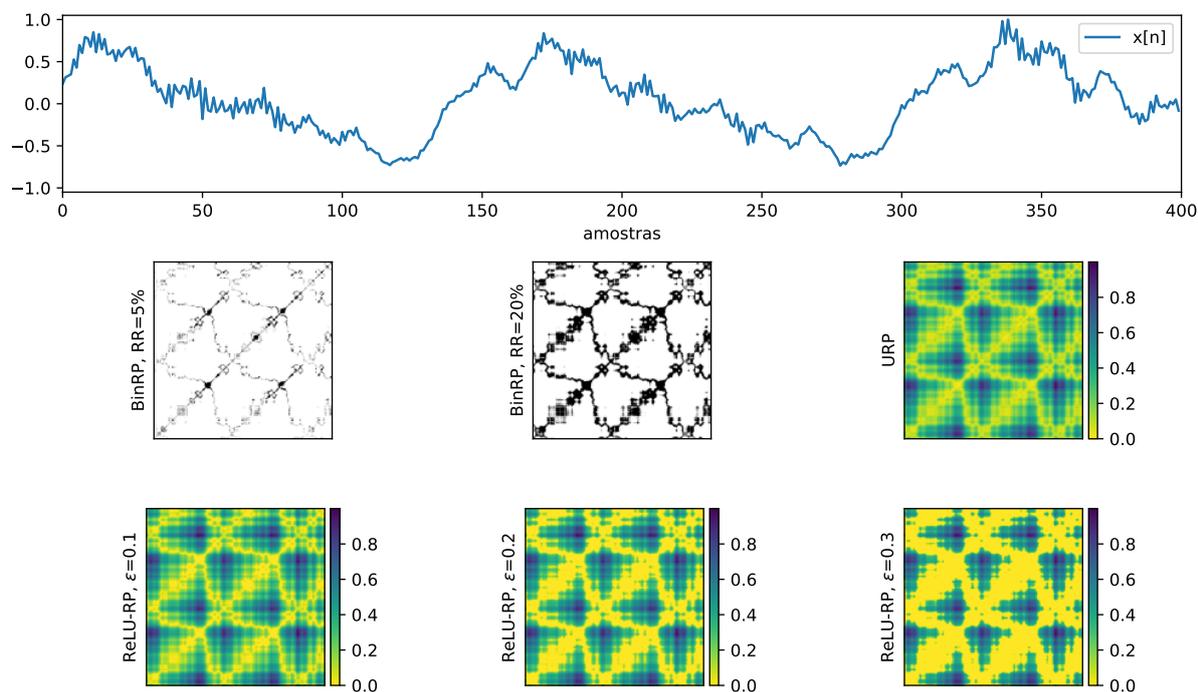


Fonte: autoria própria.

Na Figura 31 é apresentada a série temporal de um sinal de voz obtido de um paciente cuja voz é afetada pela presença de edemas bilaterais na laringe. É possível perceber que o período do sinal afetado por edema é ainda maior que o período do sinal afetado por nódulos. Sendo assim, a distância entre as diagonais principais é ainda maior. Nos RP-binários, nota-se ainda que, as estruturas formadas em torno das diagonais principais são tão extensas que conectam as diagonais. As estruturas formadas nos demais RPs também possuem formatos bastante distintos dos RPs de sinais de voz saudável e afetada por nódulo. Na Figura 32 é apresentada a projeção bidimensional do atrator de um sinal de voz afetada por edemas bilaterais proveniente da base de sinais Kay Elemetrics (1994).

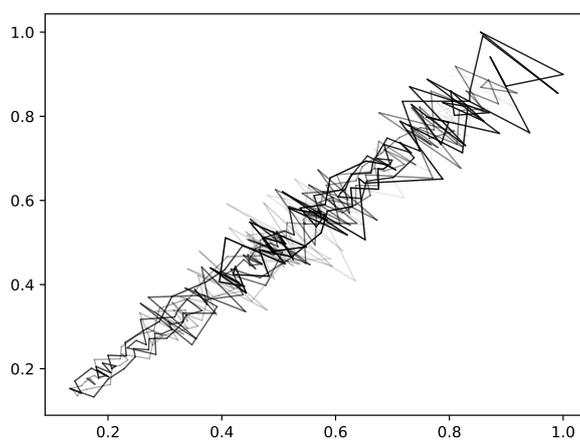
Diferente das demais patologias, vozes afetadas por edemas, possuem harmônicas de alta energia em baixas frequências de seus espectros. Na Figura 33 é apresentada a transformada de Fourier obtida do mesmo sinal de voz afetada pela presença de edema da Figura 31. Os FFT-RPs desse sinal apresentam os padrões mais irregulares. Os pontos de interesse representados nos gráficos concentram-se principalmente na lateral esquerda e parte inferior. Isso ocorre devido a remoção de pontos com amplitude muito baixa pelos RPs formados com um ε baixo. Dessa forma, restam no FFT-RPs apenas informações relevantes acerca do espectro do sinal. Na Figura 32 é apresentada a projeção bidimensional do atrator do espectro da frequência de um sinal de voz afetada por edemas bilaterais proveniente da base de sinais Kay Elemetrics (1994).

Figura 31 – Sinal de voz afetado por edemas bilaterais (identificado por ANA15 na base Kay Elemetrics (1994)) e seus respectivos RPs-binários, URP e ReLU-RPs.



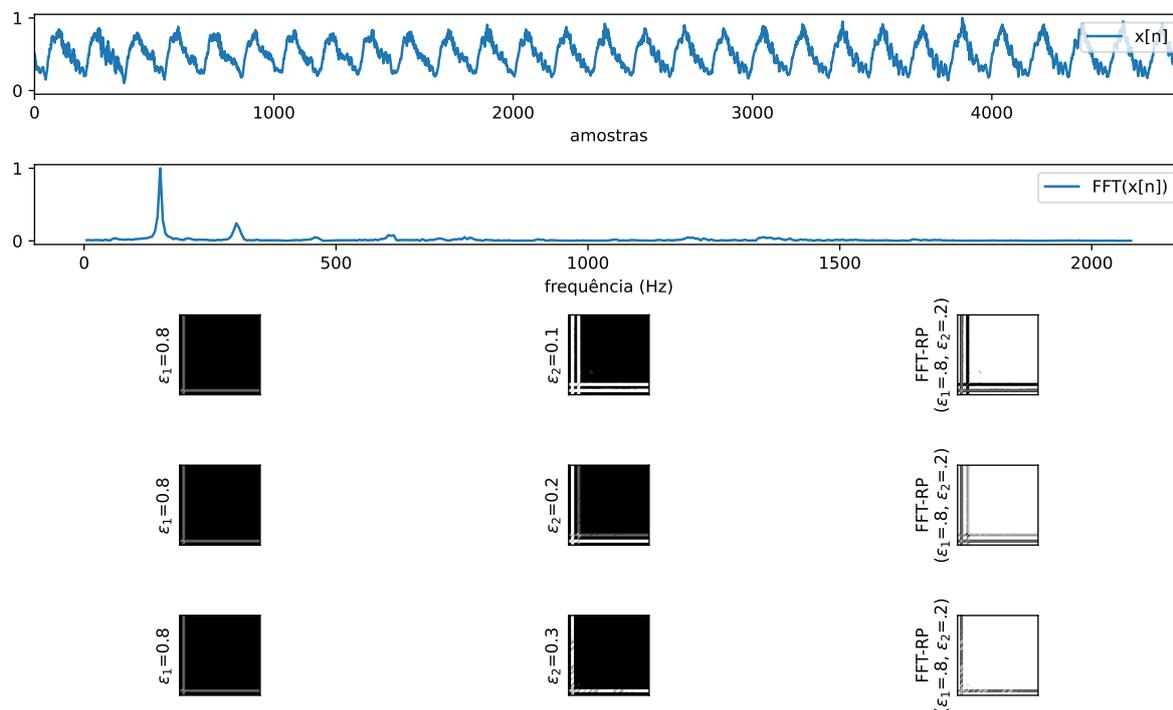
Fonte: autoria própria.

Figura 32 – Projeção bidimensional do atrator do sinal de voz afetado por edemas bilaterais (identificado por ANA15 na base Kay Elemetrics (1994)).



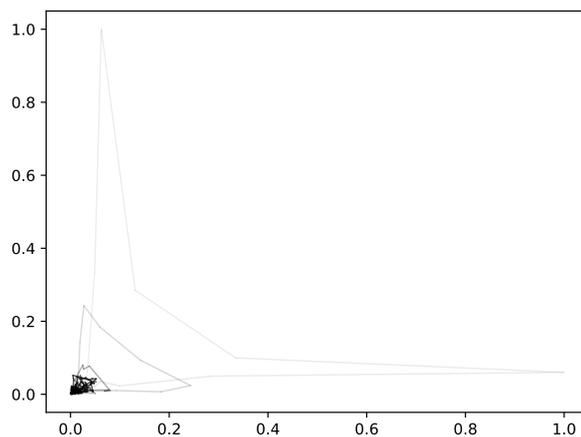
Fonte: autoria própria.

Figura 33 – Espectro de frequência de sinal de voz afetada por edemas bilaterais (identificado por ANA15 na base Kay Elemetrics (1994)) e suas respectivas variações de FFT-RPs.



Fonte: autoria própria.

Figura 34 – Projeção bidimensional do atrator do espectro da frequência do sinal de voz afetada por edemas bilaterais (identificado por ANA15 na base Kay Elemetrics (1994)).

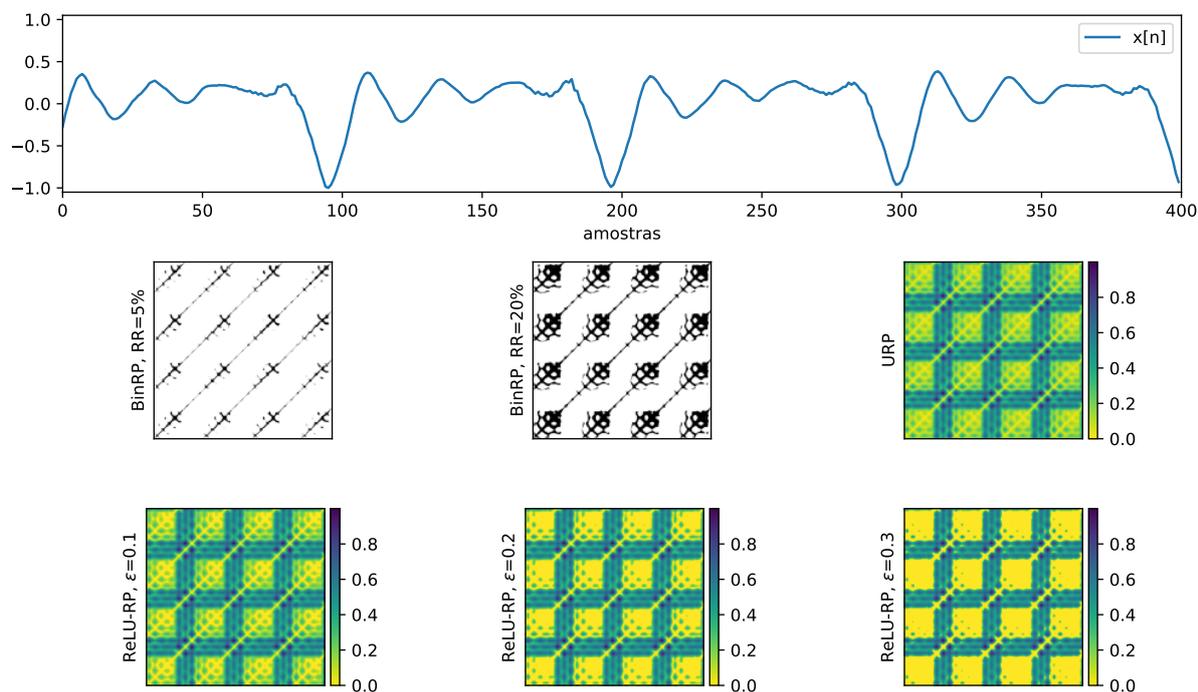


Fonte: autoria própria.

Na Figura 35 é apresentada a série temporal de um sinal de voz obtido de um paciente cuja voz é afetada por paralisia unilateral nas pregas vocais. Este sinal de voz é caracterizado por possuir períodos bem definidos, que chegam a se assemelhar ao sinal de voz saudável, porém há oscilações mais suaves dentro de cada período. Vale ressaltar que as alterações em sinais de vozes afetadas por paralisia podem variar de acordo com a gravidade da patologia. Nos RP-binários dessa série temporal, é possível observar que há estruturas bem definidas que acompanham as diagonais principais, sendo mais evidentes para uma $RR=20\%$. Os ReLU-RPs também capturam informações importantes para distinguir a patologia. Comparado aos ReLU-RPs de um sinal de voz saudável, este possui muito mais áreas de não recorrência evidenciadas. Figura 36 é apresentada a projeção bidimensional do atrator de um sinal de voz afetada por paralisia unilateral nas pregas vocais proveniente da base de sinais Kay Elemetrics (1994).

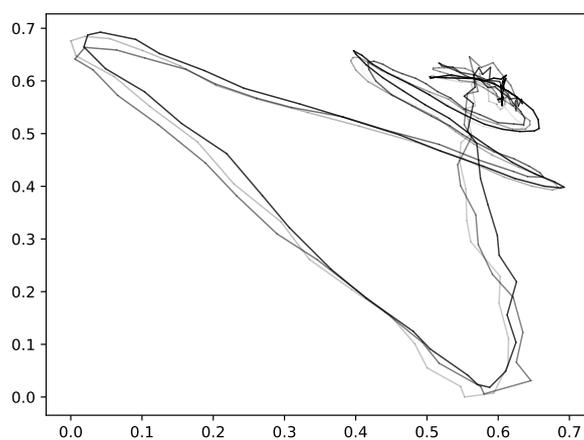
No espectro da frequência, outras características do sinal de voz afetada por paralisia podem ser analisadas. Na Figura 37 é apresentada a transformada de Fourier obtida do mesmo sinal de voz da Figura 35. Percebe-se um evanescimento da energia das harmônicas após 1 kHz. Esse padrão é realçado nos FFT-RPs que possuem um traço regular até pouco mais da metade do gráfico, ocultando os pontos com amplitude muito baixa em frequências mais altas. Na Figura 32 é apresentada a projeção bidimensional do atrator do espectro da frequência de um sinal de voz afetada por paralisia unilateral nas pregas vocais proveniente da base de sinais Kay Elemetrics (1994).

Figura 35 – Sinal de voz afetada paralisia unilateral nas pregas vocais (identificado por ABB09AN na base Kay Elemetrics (1994)) e seus respectivos RPs-binários, URP e ReLU-RPs.



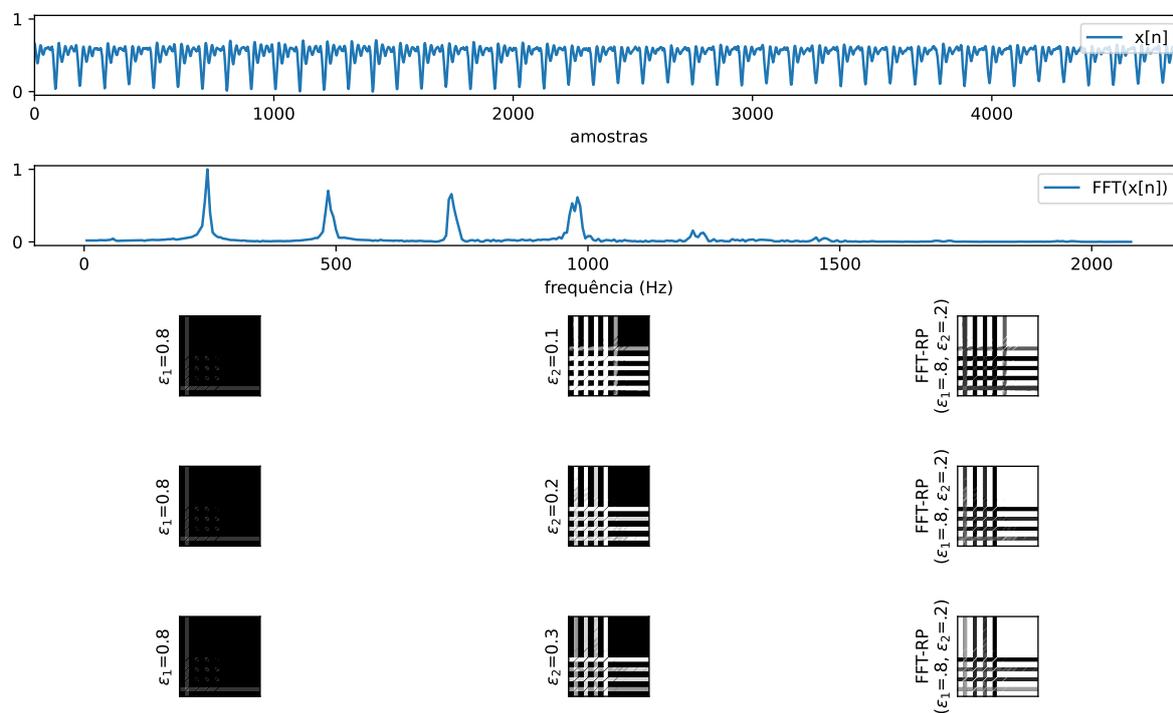
Fonte: autoria própria.

Figura 36 – Projeção bidimensional do atrator do sinal de voz afetada paralisia unilateral nas pregas vocais (identificado por ABB09AN na base Kay Elemetrics (1994)).



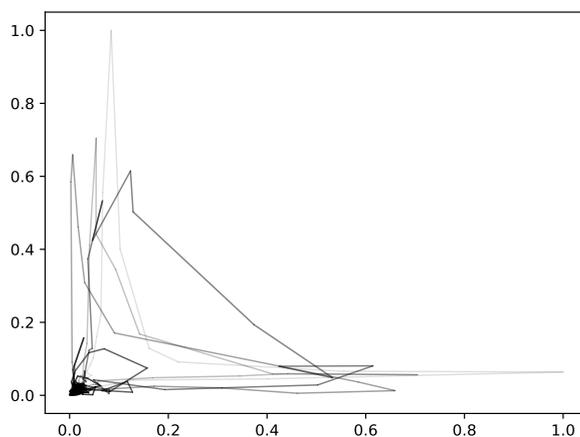
Fonte: autoria própria.

Figura 37 – Espectro de frequência de sinal de voz afetada paralisia unilateral nas pregas vocais (identificado por ABB09AN na base Kay Elemetrics (1994)) e suas respectivas variações de FFT-RPs.



Fonte: autoria própria.

Figura 38 – Projeção bidimensional do atrator do espectro da frequência do sinal de voz afetada paralisia unilateral nas pregas vocais (identificado por ABB09AN na base Kay Elemetrics (1994)).



Fonte: autoria própria.

2.4 REDES NEURAIS CONVOLUCIONAIS

Na última década, observou-se que vários estudos propuseram o uso de redes neurais convolucionais (CNNs) para análise de gráficos de recorrência como alternativa à extração manual de recursos, como é o caso das métricas de RQA (HSUEH et al., 2019; GARCIA-CEJA; UDDIN; TORRESEN, 2018; AFONSO et al., 2019). Esse tipo de abordagem tem se destacado pelo fato de ser possível extrair modelos de dados brutos, e nesse contexto, permitir que as variações presentes nos RPs sejam representadas de forma mais fidedigna. Abordagens baseadas em RQA são fundamentadas em fórmulas matematicamente bem definidas, que muitas vezes não acompanham totalmente as mudanças ou tendências apresentadas ao longo do tempo nos RPs (LEE; LEE; SHIN, 2019). O mesmo impasse ocorre quando os RPs são tratados como imagens de textura. Segundo Andrearczyk (2017) não há como generalizar de forma satisfatória os métodos clássicos de processamento de dados para texturas complexas, numerosas ou com alta variação intra-classe, como encontradas em vários problemas.

Nesse cenário, a aplicação das CNNs tem se destacado entre os algoritmos baseados em redes neurais artificiais (ANNs - Artificial Neural Networks). Estes algoritmos possuem estrutura similar à sistemas biológicos e foram desenvolvidos para aplicações de dados como imagens e vídeos. As CNNs eliminam a necessidade do processo de extração de características, de forma que as imagens (tais como RPs) podem ser utilizadas diretamente como entrada da rede.

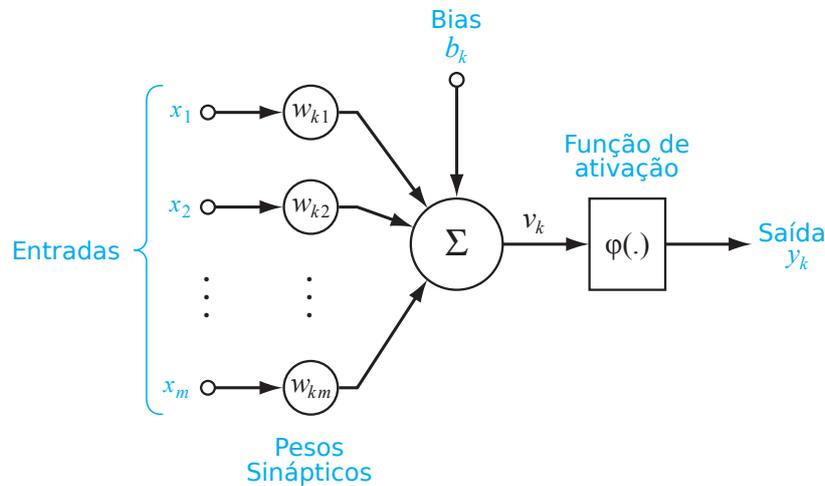
De acordo com Géron (2019), as CNNs emergiram de um estudo do córtex visual do cérebro, e tem sido usadas no reconhecimento de imagens desde a década de 1980. A relevância desse campo foi ampliada com o aumento do poder computacional, possibilitando o processamento de grandes quantidades de dados. A principal diferença entre CNNs e outros tipos de redes neurais consiste justamente na abstração da extração de características. Essas redes são capazes de extrair progressivamente as características dos dados por meio de múltiplos filtros, sendo consideradas ferramentas poderosas para a aprendizagem supervisionada. Após a camada convolucional, as CNNs se assemelham aos tipos mais tradicionais de redes neurais, sendo compostas de camadas de neurônios totalmente conectadas. A arquitetura mais simples e popular de neurônio é o *perceptron*. O funcionamento destes elementos deve ser entendido para uma melhor compreensão do funcionamento das CNNs.

2.4.1 Neurônio artificial *perceptron*

A arquitetura do tipo *perceptron*, proposta por Rosenblatt, com base no neurônio de McCulloch-Pitts e nas descobertas de Hebb, é uma das mais simples e possui

três elementos básicos: conjunto de sinapses, somador e função de ativação (HAYKIN, 2001; GÉRON, 2019). Na Figura 39 é apresentada a estrutura básica de um neurônio.

Figura 39 – Elementos estruturais do neurônio artificial *perceptron*.



Fonte: Adaptado de Haykin (2001).

Uma sinapse é caracterizada por um número real denominado peso sináptico. Cada sinapse de um conjunto está associada a uma das componentes do sinal de entrada e tem o objetivo de ponderá-la (HAYKIN, 2001). Ainda segundo Haykin (2001), o somador é responsável pela soma dos sinais de entrada ponderados pelas respectivas sinapses dos neurônios, sendo esta operação caracterizada como uma combinação linear. A função de ativação, por sua vez, recebe o resultado da combinação linear, e por fim, define a amplitude da saída do neurônio.

No modelo da Figura 39, um neurônio k possui m sinais de entrada x_j , conectados a seus respectivos pesos $w_{k,j}$. Um parâmetro denominado “bias”, representado por b_k , é responsável por ajustar o resultado da combinação linear, deslocando a função de decisão para esquerda ou direita. A Equação 11 demonstra a saída y_k do neurônio determinada pela função de ativação $\varphi(v_k)$.

$$y_k = \varphi(v_k) = \varphi\left(b_k + \sum_{j=1}^m x_j w_{k,j}\right) \quad (11)$$

Apesar de originalmente Rosenblatt ter utilizado a função degrau na ativação do *perceptron*, outras podem ser aplicadas. Na Figura 40 são apresentados os gráficos das funções de ativação mais utilizadas.

- Função de Unidade Linear Retificada (ReLU):

$$\varphi(v) = \max(0, v) \quad (12)$$

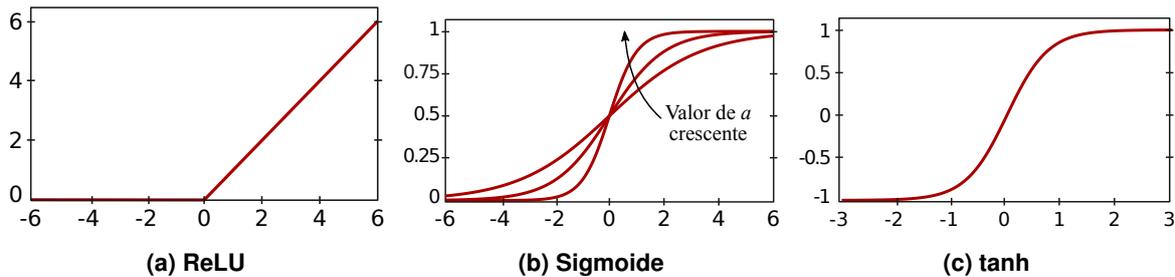
- Função Sigmoide:

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad (13)$$

- Função Tangente Hiperbólica:

$$\varphi(v) = \tanh(v) \quad (14)$$

Figura 40 – Principais funções de ativação.



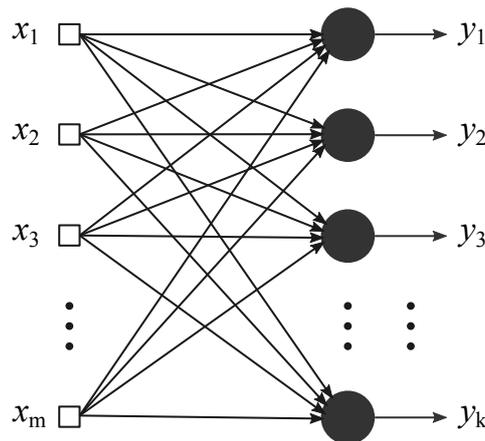
Fonte: autoria própria.

Faceli et al. (2019) relatam que os neurônios são dispostos em uma ou mais camadas interligadas por muitas conexões, geralmente unidirecionais. Durante um processo de aprendizagem, o conhecimento da rede neural é assimilado por meio do ajuste dos pesos w_k , de acordo com os dados de treinamento de entrada da rede. O processo de aprendizagem consiste, portanto, em ajustar cada peso sináptico da rede para que ela produza uma saída desejada para uma dada entrada. Assim, uma CNN é caracterizada por sua estrutura (quantidade de neurônios e camadas) e pelo seu algoritmo de aprendizagem.

Na forma mais simples de uma rede em camadas, temos uma camada de entrada de nós de fonte que se projeta sobre uma camada de saída de neurônios. Na Figura 41 é apresentada, utilizando a representação grafo arquitetural, uma rede neural artificial *perceptron* de camada única, na qual a representação do neurônio foi abstraída. Segundo Haykin (2001), a representação grafo arquitetural é caracterizada por três elementos:

1. **Nó de fonte:** Fornece sinal de entrada para o grafo;
2. **Nó computacional:** Representa um neurônio artificial;
3. **Elo de comunicação:** Conectam nós de fonte a nós computacionais e nós computacionais entre si. Eles não carregam pesos, apenas fornecem direções ao fluxo de sinal do grafo.

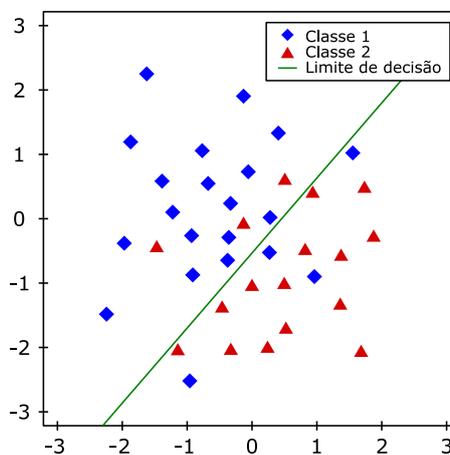
Figura 41 – Gráfico de fluxo de sinal de uma rede neural de uma camada.



Fonte: Adaptado de Braga, Ferreira e Ludermir (2007).

As redes neurais de uma camada apenas, no entanto, são bastante limitadas devido ao fato de que só são capazes de classificar dados linearmente separáveis. Na Figura 42 é apresentado um espaço bidimensional, no qual o hiperplano é representado por uma reta que separa duas classes. Devido a essa limitação, redes neurais *perceptron* de uma única camada possuem pouca aplicação em problemas do cotidiano.

Figura 42 – Classificação de dados utilizando uma rede *perceptron* de uma camada.

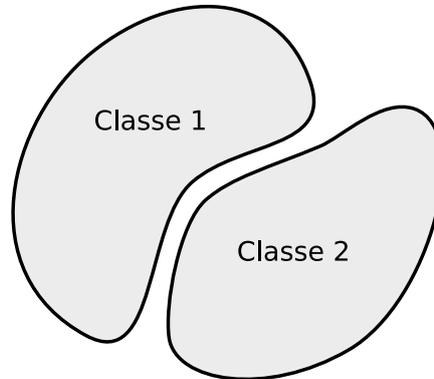


Fonte: autoria própria.

Redes neurais *perceptron* de múltiplas camadas (MLPs - *Multiplayer Perceptron*) consistem, tradicionalmente, de um conjunto de unidades sensoriais (nós de fonte), uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída de nós computacionais (neurônios), de forma que o sinal de entrada se propaga para frente através da rede, camada por camada (HAYKIN, 2001). MLPs lidam de maneira bastante satisfatória com problemas não lineares que representam a grande maioria

dos problemas reais, podendo ser aplicadas em problemas de classificação como, por exemplo, mostrado na Figura 43.

Figura 43 – Classificação de dados utilizando uma MLP.



Fonte: Adaptado de Haykin (2001).

As redes neurais de múltiplas camadas podem ou não ser totalmente conectadas. São ditas totalmente conectadas se todos os nós existentes em uma camada estão totalmente conectados todos os nós da camada seguinte. Quando não há a presença de alguns dos elos de comunicação, a rede pode ser qualificada como parcialmente conectada.

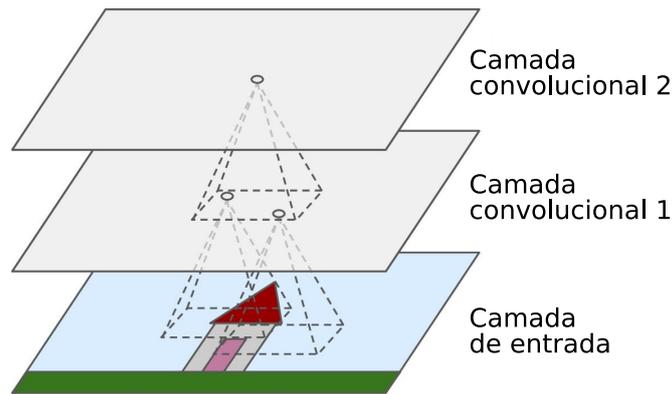
2.4.2 Camada convolucional

Na CNN, a camada convolucional pode ser considerada a mais importante. Nesta camada, os neurônios são representados por *kernels* (bancos de filtros) com campos receptivos locais em imagens de entrada, permitindo que a correlação de valores de *pixels* seja explorada (ANDREARCZYK, 2017). Os neurônios localizados na primeira camada convolucional não estabelecem conexão com todos os *pixels* da camada de entrada, apenas com *pixels* em seu campo receptivo (GÉRON, 2019), como ilustrado na Figura 44. De forma similar, os neurônios da próxima camada só se conectarão com os neurônios correspondentes à área do *kernel* seguinte.

Segundo Géron (2019), em um *kernel*, os pesos podem ser representados como uma pequena imagem do tamanho do campo receptivo. Na Figura 45 é ilustrado um neurônio na posição i, j conectado a saída de outra camada, com um campo receptivo de tamanho 3×3 . A operação linear que caracteriza a saída do neurônio na posição i, j é denominada “convolução”.

A convolução pode ser destacada como a operação mais importante da camada convolucional, e é representada pelo símbolo $*$. Levando em consideração o exemplo

Figura 44 – Camadas convolucionais de uma CNN com *kernels* retangulares.

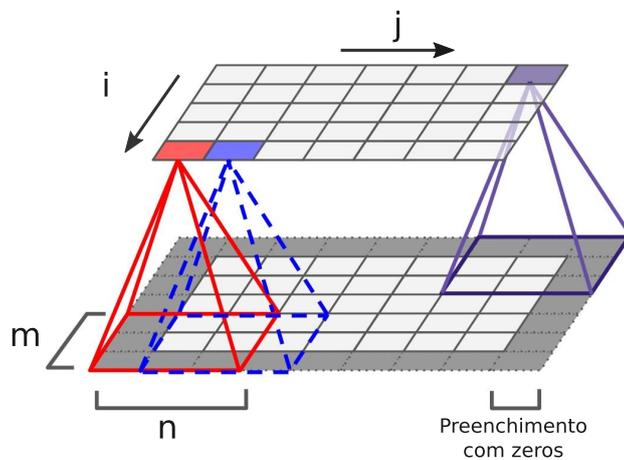


Fonte: Adaptado de Géron (2019).

da Figura 45, essa operação pode ser descrita pela Equação 15, na qual I se refere à imagem de entrada e K ao *kernel* de tamanho 3×3 . A matriz resultante dessa operação é conhecida como mapa de características.

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \quad (15)$$

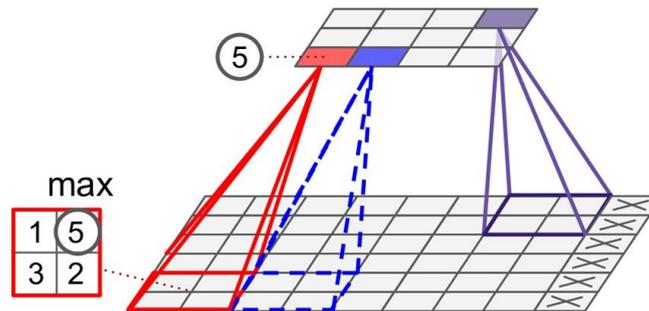
Figura 45 – Conexões entre camadas convolucionais.



Fonte: Adaptado de Géron (2019).

Após a camada convolucional, uma camada denominada *pooling* é utilizada com o intuito de reduzir a imagem de entrada. Essa operação também ajuda a minimizar o uso de memória, custo computacional e quantidade de parâmetros. O funcionamento de uma camada *pooling* é similar ao de uma camada convolucional. Contudo, funções são utilizadas dentro de um campo receptivo para determinar a saída, como máxima e mínima. Um exemplo pode ser encontrado na imagem da Figura 46.

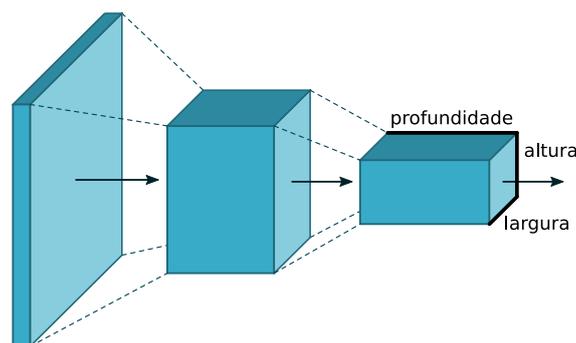
Figura 46 – Camada do tipo *max pooling* com kernel de tamanho 2×2 .



Fonte: Adaptado de Géron (2019).

Buduma e Locascio (2017) apontam que a camada convolucional é necessária para processar um volume tridimensional de informação para produzir um novo volume tridimensional de informação. Isso ocorre pois a estrutura da camada convolucional é definida por um conjunto de *kernels* com os quais os dados de entrada são convoluídos, sendo a quantidade e o tamanho dos filtros os principais parâmetros da camada (ANDREARCZYK, 2017). Conseqüentemente, a saída de cada camada é composta por mapas de características, que posteriormente são convoluídos com o conjunto de *kernels* da próxima camada. Para CNNs dedicadas à classificação de imagens, a entrada e saída das camadas convolucionais são o produto da altura, largura e profundidade (BUDUMA; LOCASCIO, 2017), como ilustrado na Figura 47.

Figura 47 – Dimensões das camadas convolucionais.



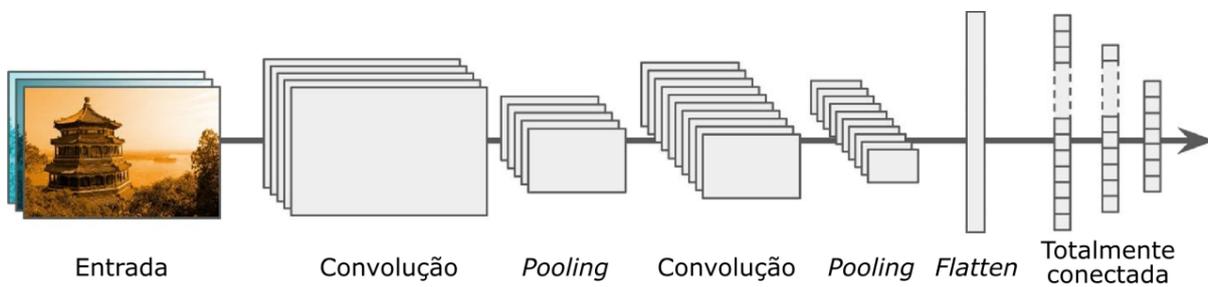
Fonte: Adaptado de Buduma e Locascio (2017).

Os mapas de características formados após o processamento dos dados pelas camadas convolucionais e de *pooling* geram uma estrutura multi-dimensional conhecida como “tensor”. Para que os dados possam ser utilizados posteriormente por um conjunto de camadas totalmente conectadas, é necessário converter essa estrutura

em um elemento unidimensional. Esta tarefa é executada por uma camada *flatten*.

No conjunto de camadas totalmente conectadas, localizado após a camada *flatten*, cada neurônio recebe como entrada todas as saídas da camada anterior. Um procedimento conhecido por *backward propagation* é utilizado para ajustar os pesos de cada neurônio a cada nova entrada durante o treinamento. Um dos métodos mais comuns empregados para otimizar o conjunto de pesos é a descida do gradiente. Na sequência, todo o processo é similar ao funcionamento de um algoritmo MLP. As saídas dessas camadas definem a qual classe pertence a imagem de entrada, permitindo a predição da classe de novas imagens. Na Figura 48 é apresentado um exemplo genérico da arquitetura de uma CNN.

Figura 48 – Arquitetura típica de uma CNN.



Fonte: Adaptado de Géron (2019).

2.4.3 Tipo de aprendizado

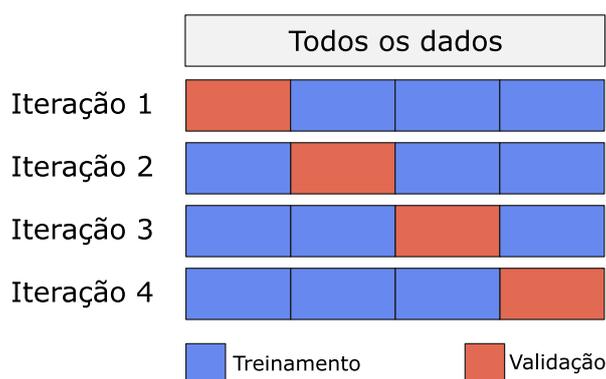
O tipo de treinamento de uma CNN também pode ser subdividido em dois tipos: aprendizado supervisionado e aprendizado não supervisionado. As redes que utilizam aprendizado supervisionado necessitam que os dados de treinamento estejam previamente rotulados de acordo com a sua classe. Esse método de treinamento é análogo aos papéis desempenhados por um professor e um aluno, em uma situação na qual o professor apresenta exemplos para que o aluno assimile as informações. Dessa forma, o algoritmo será capaz de realizar uma correção dos pesos a cada nova entrada, generalizando o conhecimento e conseqüentemente executando predições mais acuradas. No aprendizado não supervisionado não é necessário que os dados estejam rotulados. Este tipo de rede é bastante utilizada em situações nas quais tenta-se descobrir a distribuição de probabilidades que forma um conjunto de dados, ou mesmo agrupá-los em classes (GOODFELLOW; BENGIO; COURVILLE, 2016). Sistemas que funcionam em tempo real geralmente empregam este tipo de aprendizagem.

2.4.4 Validação cruzada

A validação cruzada consiste em uma técnica que avalia se a rede proverá uma boa generalização independente da divisão do conjunto de dados. Segundo Goodfellow, Bengio e Courville (2016), um pequeno conjunto de testes implica em incerteza estatística em torno do erro médio estimado de teste, dificultando a afirmação de que o algoritmo *A* funciona melhor que o algoritmo *B* na tarefa especificada. Dessa forma, o processo de validação cruzada propõe a repetição do treinamento utilizando diferentes subconjuntos de treinamento e teste, separados de forma aleatória.

De acordo com Goodfellow, Bengio e Courville (2016), a forma mais comum é baseada na técnica de *k-fold*, na qual uma partição do conjunto de dados é formada dividindo-a em *k* subconjuntos sem sobreposição. As métricas utilizadas na avaliação do desempenho da rede são calculadas a partir da média do desempenho de cada treinamento. Na Figura 49 a técnica de *k-fold* é exemplificada considerando $k = 4$. Nela, os conjuntos de treinamento e validação foram compostos de quatro maneiras diferentes.

Figura 49 – Método de validação cruzada 4-fold.



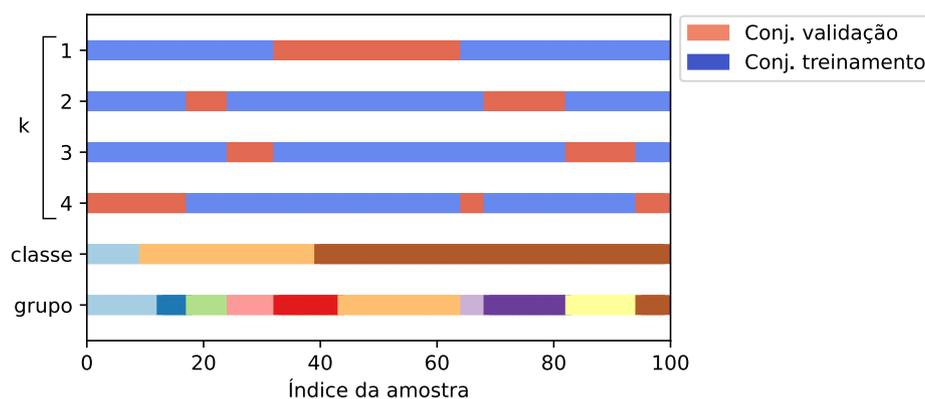
Fonte: autoria própria.

Contudo, há cenários mais complexos nos quais uma ou mais classes são compostas por diferentes grupos (por exemplo, a classe de patologias composta por amostras de diferentes patologias). Nesses casos, há a necessidade de garantir que haja a mesma proporção de amostras de cada grupo que compõe uma classe para que o processo de aprendizado não seja enviesado.

A fim de garantir essa proporcionalidade entre as classes e grupos, pode ser utilizada a técnica de *stratified group k-fold*, uma derivação do método comum *k-fold*. Nesse procedimento, tenta-se preservar a distribuição das classes em cada divisão, fazendo com que cada divisão *k* possua quantidades proporcionais de amostras de cada grupo tanto no conjunto de validação quanto no conjunto de treinamento. Este tipo de abordagem também é implementada a fim de garantir que conjuntos de dados

desequilibrados não produzam resultados de treinamento distorcidos, como é o caso do banco de dados utilizado neste estudo. Na Figura 50 a técnica de *stratified group k-fold* é exemplificada considerando $k = 4$. De forma similar ao exemplo anterior, o conjunto de testes foi fixado, enquanto os conjuntos de treinamento e validação foram compostos de quatro maneiras diferentes. Contudo, agora é garantido que a formação de cada conjunto manterá a proporção de amostras de acordo com cada grupo.

Figura 50 – Método de validação cruzada 4-fold.



Fonte: Adaptado de scikit-learn.org (2022)

Neste estudo, consideramos duas classes de sinais de áudio: saudáveis e patológicos. A classe saudável é composta por apenas um grupo de pessoas, que são as consideradas saudáveis. Já a classe de patológicos é composta por três grupos de pessoas, afetadas pelas seguintes patologias: paralisia, edema de Reinke e nódulos nas pregas vocais.

3. ESTADO DA ARTE

Em vista da quantidade de informações proporcionada pelos gráficos de recorrência, estudos de diversas áreas introduziram a RQA como alternativa para análise de dados baseados em sistemas dinâmicos. Lewenstein, Jamrózy e Leyko (2016) propuseram o uso de RPs e da RQA para analisar aquisições de sinais de batimentos cardíacos a fim de avaliar características correlacionadas com a síndrome avançada de falha crônica do coração. Em um estudo mais recente, os RPs foram aplicados por Godoy e Gregório (2019) na análise qualitativa e quantitativa da variabilidade da frequência cardíaca. Os padrões definidos nos RPs permitiram a verificação de homeostase adequada em diferentes graus de comprometimento.

Na última década, a utilização de RPs e RQA também tem se destacado na detecção de patologias laríngeas por meio da análise da fala. A voz pode ser caracterizada como um sistema dinâmico não linear e sofre alterações quando a laringe é afetada. Nesse contexto, pode ser citado o trabalho de Vieira et al. (2018), que apresenta a detecção e discriminação de patologias laríngeas e a avaliação do distúrbio de voz. Os autores utilizaram 15 medidas de quantificação de recorrência e aplicaram uma análise discriminante para realizar a classificação dos dados. Pode-se destacar a acurácia obtida na classificação de amostras saudáveis e patológicas, que alcançou 92,48% com desvio padrão de 6,49%. Em outro estudo mais recente, Lopes, Vieira e Behlau (2020) comparam a eficiência da quantificação de recorrência com medidas tradicionais, cepstrais e não lineares para a discriminação de sinais vocais com e sem distúrbio de voz. O desempenho do classificador também foi avaliado ao combinar tais medidas acústicas. Os autores observaram que, ao combinar diferentes tipos de medidas, apenas uma combinação obteve os melhores resultados (86,43% de acurácia, 90,83% de sensibilidade e 80,83% de especificidade), consistindo na combinação da relação de excitação glótica-ruído (GNE - *glottal-to-noise excitation ratio*), uma medida acústica e da entropia de Shannon (ENTR), uma métrica pertencente à RQA.

Considerando o recente avanço das redes neurais convolucionais, em alguns estudos têm-se eliminado a obtenção de métricas por meio do RQA, e optado por utilizar os RPs diretamente como entrada das redes. Nesse contexto, nota-se a contribuição de Hsueh et al. (2019), que utilizaram uma CNN para classificar RPs provenientes de sinais de corrente coletados de motores de indução trifásicos. Comparado a outros estudos que também visam o monitoramento das condições de máquinas de rotação e utilizam outros métodos de classificação, a abordagem proposta apresentou maior acurácia, alcançando 99,81%.

Em outro estudo, Garcia-Ceja, Uddin e Torresen (2018) propuseram a classificação de atividades físicas e de locomoção (como *cooper*, caminhar ou sentar-se). Dados brutos do acelerômetro de um smartphone foram modelados como gráficos de recorrência, que posteriormente foram utilizados como entrada de uma CNN. Para fins de comparação, uma Deep Belief Network (DBN) foi treinada diretamente com os dados dos três eixos do acelerômetro e suas respectivas magnitudes. A CNN obteve uma acurácia de 94,2% enquanto a DBN alcançou 83%, mostrando que análise de RPs pela CNN é promissora.

Esse tipo de abordagem também se mostrou eficiente na identificação de patologias. Afonso et al. (2019) utilizaram RPs obtidos a partir de dados da dinâmica da escrita como entradas de uma CNN. Os dados foram adquiridos de indivíduos submetidos a tarefas que medem habilidades relacionadas à escrita. O método foi proposto com a finalidade de auxiliar na identificação da doença de Parkinson. Os resultados apresentaram uma precisão média de 87%, superior à abordagem anterior proposta pelos autores.

Ainda na área da saúde, em outro estudo recente, elaborado por Zhao et al. (2019), foram aplicados RPs e CNNs com o propósito de estruturar um modelo novo, consistente, robusto e eficaz para a detecção de hipóxia fetal (diminuição da assimilação de oxigênio pelo feto). Os RPs foram gerados a partir de sinais de frequência cardíaca de fetos. O modelo obteve uma acurácia de 98,69%, identificando corretamente 99,29% dos casos de hipóxia fetal.

Alguns autores optaram ainda por testar modelos baseados em diferentes tipos de RPs. Um deles é o URP, que não emprega um limiar de distância entre os estados para obtenção da matriz de valores. Consequentemente, tal matriz é composta de valores não binários. Nesse sentido, pode ser citado o estudo de Hatami, Gavet e Debayle (2018), no qual URPs foram utilizados como entrada de uma CNN com a finalidade de investigar a eficácia desta abordagem como solução para a classificação de séries temporais. Na validação do método, foram utilizados sinais provenientes da base de dados do conjunto UCR de séries temporais. Segundo os autores, o motivo da utilização de URPs foi evitar a perda de informações por binarização da matriz. Resultados indicaram que o método proposto foi mais eficiente quando comparado a outras abordagens de classificação.

Uma aplicação mais específica dos URPs pode ser vista no trabalho de Chen, Su e Yang (2020). RPs e URPs foram utilizadas com CNN para a detecção de anomalias no processo de soldagem instantânea, comumente utilizado na fabricação de correntes de âncora na indústria naval. Além de serem utilizados individualmente como entradas da rede, os RPs e URPs foram combinados para o treinamento de uma mesma CNN.

Para fins de comparação, as métricas da RQA também foram extraídas e classificadas por uma SVM (*Support Vector Machine*). Os resultados apontaram que a abordagem que utilizou apenas RPs foi capaz de detectar as transições dinâmicas melhor do que URPs. De acordo com Chen, Su e Yang (2020), isso ocorre porque os URPs levam em consideração apenas as informações de distância entre dois estados dinâmicos, enquanto os RPs contêm os comportamentos de recorrência das transições de estado no espaço de fases. Apesar disso, todas as abordagens com base no uso de CNN obtiveram boa acurácia, com o máximo de 96%. É importante destacar que os autores validaram o método desenvolvido tanto utilizando sinais de corrente elétrica quanto para os dados de posição do eletrodo.

Outro tipo de RP foi proposto no estudo de Lee, Lee e Shin (2019), denominado ReLU-RP. Neste trabalho, os autores geraram Bin-RPs (RPs), Cont-RPs (mesmo que URPs) e ReLU-RPs a partir de sinais de ECG e fotopleletismograma (PPG - photoplethysmogram), que consiste em uma técnica óptica para detecção de alterações volumétricas no sangue na circulação periférica. No estudo, foram avaliados o desempenho individual das variações de RPs na classificação de sonolência de motoristas por meio de uma CNN. O ReLU-RP consiste basicamente em um URP filtrado por uma função ReLU. Segundo Lee, Lee e Shin (2019), os ReLU-RPs enfatizam o padrão de linhas verticais (ou horizontais) do Cont-RP, o que indica que um estado não muda ou muda muito lentamente durante um certo tempo. Para comprovar a eficácia da utilização de CNNs, as métricas de RQA foram extraídas e classificadas por quatro modelos: SVM, LR (*logistic regression*), KNN (*K-nearest neighbors*) e *random forest*. De forma geral, os modelos de CNN utilizando Bin-RPs, Cont-RPs e ReLU-RPs obtiveram melhores resultados do que a classificação das medidas da RQA. Entretanto, o modelo que utilizou os ReLU-RPs resultou na melhor performance considerando acurácia, precisão, recall e F-score. O desempenho foi similar tanto para o uso de sinais de ECG quanto para PPG. Contudo o uso de ECG se mostrou superior.

Recentemente, foi proposta por Ladeira et al. (2020) uma nova técnica de análise de recorrência do espectro de frequência. Os autores consideram que recorrências também podem aparecer em domínios diferentes do tempo, sendo um deles a frequência. No trabalho, essa abordagem é investigada por meio da análise de sinais eletroencefalográficos, considerando dois estados: olhos abertos e olhos fechados. Primeiramente, os sinais são convertidos para o domínio da frequência aplicando a transformada rápida de Fourier. Em seguida, os novos RPs são gerados de maneira similar aos RPs tradicionais. Uma nova métrica foi concebida para analisar regiões de recorrência na direção vertical pois, de acordo com Ladeira et al. (2020), as matrizes de recorrência dos sinais de frequência apresentam maiores diferenças de recorrência nas linhas verticais/horizontais do que nas diagonais. Dessa forma, pode-se verificar a

concentrações de recorrência na frequência (R_{cf}) e a concentrações de recorrência nas bandas de frequências ($R_{c(i)}$). Para validação do estudo, foram extraídos das matrizes de recorrência dos sinais de frequência quantificadores já conhecidos: determinismo (DET), comprimentos de linhas diagonais (L) e ENTR. Além disso, foram extraídos os novos quantificadores criados, R_{cf} e $R_{c(0,25)}$. Notou-se que, na análise de recorrência do espectro de frequência, o intervalo de confiança da média (CI) dos resultados dos quantificadores não resulta em sobreposição e cria dois intervalos diferentes para cada estado (olhos abertos ou fechados). Ainda foi observado que há uma maior possibilidade de uma média igual para os resultados de recorrência da série temporal do que na banda de frequência. Esses pontos indicam que a abordagem pode ser promissora em situações nas quais a recorrência da série temporal se aproxime para classes diferentes. Outro ponto positivo destacado pelos autores é que este tipo de RP permite que os sinais adquiridos por longos intervalos de tempo sejam analisados em uma única etapa, dentro do espectro de frequência. Por fim, Ladeira et al. (2020) ressaltam que foi possível obter características relevantes dos sinais de EEG tanto por meio dos quantificadores previamente estabelecidos (DET, L e ENTR) quanto por meio dos novos quantificadores propostos R_{cf} e $R_{c(0,25)}$.

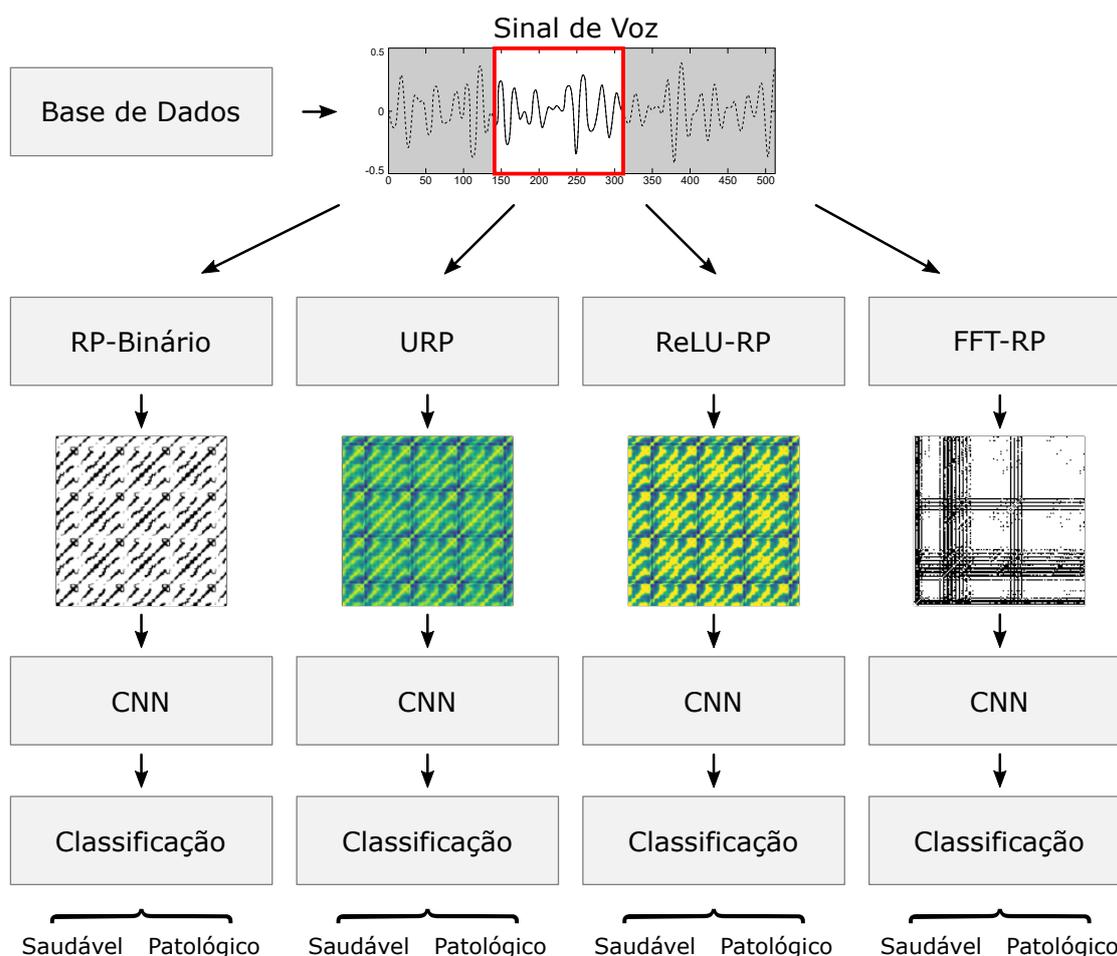
Considerando a análise realizada acerca dos trabalhos relacionados à RPs e suas aplicações, publicados e identificados até o presente momento, observou-se que não há pesquisas que avaliam o emprego de URPs, ReLU-RPs ou RPs do espectro de frequência aplicados à detecção de patologias laringeas por meio de sinais de voz. Além disso, em estudo prévio desenvolvido por Barros et al. (2020), foi observado que ao utilizar uma CNN para distinguir RPs de sinais de vozes afetados por patologia obteve-se uma acurácia média de 98,2%. O estudo também contribui com um novo método de *data augmentation* desenvolvido especificamente para RPs. Os resultados foram comparados com o método proposto por Souza et al. (2015), que utilizou a mesma base de dados e outro método de classificação. A acurácia média obtida por Barros et al. (2020) foi 10% maior para discriminação de vozes saudáveis e vozes afetadas por patologia.

4. METODOLOGIA

Este capítulo detalha a metodologia que foi aplicada para a validação desta pesquisa. É descrita a origem e composição do banco de dados que foi utilizado, a geração dos tipos RPs, e a classificação que foi empregada.

No diagrama ilustrado na Figura 51 são especificadas as etapas do método proposto. Inicialmente, os sinais de voz selecionados, provenientes da base de dados MEEI, são seccionados em fragmentos com mesma quantidade de pontos, sem sobreposição. Dessa forma, é possível ampliar a quantidade de amostras disponíveis. A partir de cada fragmento, gráficos de recorrência poderão ser gerados tais como RP binário, URP, ReLU-RP e FFT-RP. O principal propósito de obter uma variedade de RPs é avaliar qual deles caracteriza de forma mais eficiente as variações entre as vozes saudáveis e vozes afetadas por patologia.

Figura 51 – Diagrama de etapas.



Fonte: autoria própria.

Posteriormente, uma arquitetura de CNN é definida para realizar a classificação

individual dos tipos de RPs utilizados. A etapa de validação cruzada garantirá que a rede possui uma boa generalização. Por fim, a CNN proposta é utilizada para classificar os RPs pertencentes às amostras saudáveis e amostras afetadas por patologia.

4.1 BASE DE DADOS

A base de sinais de voz utilizada nessa pesquisa foi gravada pelo Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab (KAY ELEMETRICS, 1994), sendo uma das mais populares na área de patologias de voz. As amostras correspondem à vogal /a/ sustentada, com 53 vozes normais e 657 patológicas, e à primeira parte de um texto de domínio público intitulado *Rainbow Passage*¹ (em tradução livre, Passagem do Arco-íris), com 53 vozes normais e 662 patológicas. As gravações foram feitas em ambiente controlado e, no total, somam 1400 sinais de voz. Originalmente, os sinais de vozes normais possuem uma frequência de amostragem de 50 kHz, enquanto que as amostras patológicas variam entre 25 kHz e 50 kHz. De acordo com informações disponibilizadas pela Kay Elemetrics (1994), ferramentas como estroboscopia, exame físico do pescoço e boca e medidas aerodinâmicas acústicas foram utilizadas para avaliar a condição das vozes.

Para este estudo, todos os sinais foram re-amostrados a taxa de 25.000 amostras por segundo, com resolução de 16 bits por amostra. Da base de dados, foram selecionadas apenas amostras correspondentes à vogal /a/. Todos os 53 arquivos de vozes saudáveis foram utilizados, além de 112 sinais de vozes afetados unicamente por uma patologia cada, dos quais 51 apresentam paralisia, 43 edema de Reinke e 18 nódulos nas pregas vocais.

É importante destacar que, apesar das gravações terem sido realizadas em ambiente controlado, uma das limitações da base de dados MEEI consiste no fato de que as amostras patológicas e as amostras normais foram registradas em dois ambientes distintos (MUHAMMAD et al., 2017). Dessa forma, sistemas propostos para analisar a base podem ser influenciados pelas características ambientais. Outras desvantagens estão associadas às diferentes frequências de amostragem utilizadas nas gravações e à pequena quantidade de sinais saudáveis. Contudo, apesar das limitações, pode-se considerar que a MEEI é uma base consolidada pelos inúmeros estudos que a utilizam.

¹ Uma cópia do texto pode ser encontrada em <<https://www.dialectsarchive.com/the-rainbow-passage>> (acesso em 20 de agosto de 2021).

4.2 GRÁFICOS DE RECORRÊNCIA

A fim de realizar uma análise inicial acerca dos padrões formados nos diferentes tipos de RPs, foram gerados RPs-binários, URPs, ReLU-RPs e FFT-RPs a partir dos sinais de voz da base de dados MEEI. Todos os gráficos de recorrência utilizados neste estudo foram obtidos com o auxílio da biblioteca NeuroKit2 (MAKOWSKI et al., 2021), desenvolvida em Python.

Para gerar os quatro tipos de RPs abordados neste estudo foram considerados um total de 165 arquivos de áudio da base de dados produzida pela Kay Elemetrics (1994). Desses, 53 são arquivos de vozes saudáveis, 51 de vozes afetadas por paralisia, 43 de vozes afetadas por edema de Reinke e 18 de vozes afetadas por nódulos nas pregas vocais. Contudo, desse total, apenas 53 consistem em sinais de vozes saudáveis enquanto que 112 representam sinais de vozes afetadas por patologias. Como este estudo tem como objetivo avaliar qual tipo de gráfico melhor distingue sinais saudáveis de sinais patológicos, a base ficaria desbalanceada. Sendo assim, uma subamostragem foi feita no conjunto de sinais patológicos para que não afetasse o desempenho da rede perante a quantidade reduzida de sinais saudáveis. Nesse contexto, para compor a base final, foram selecionados aleatoriamente 30 sinais de vozes afetadas por paralisia, 30 de vozes afetadas por edema de Reinke e foram mantidos os 18 sinais de vozes afetadas por nódulos nas pregas vocais.

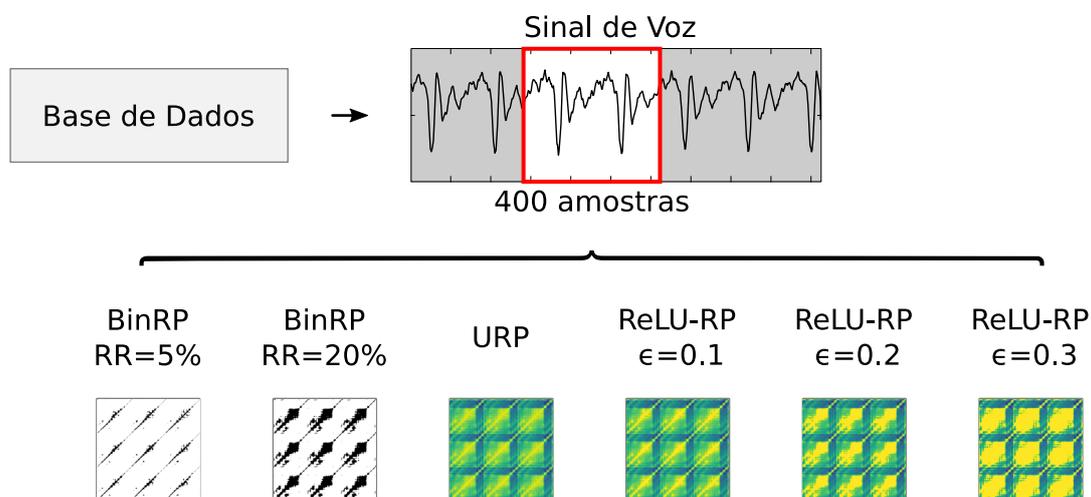
Com todos os sinais re-amostrados a uma taxa de 25.000 amostras por segundo, foram considerados apenas as primeiras 24.000 amostras de cada sinal, dado que alguns arquivos possuíam apenas 1 segundo de áudio. Para a geração dos RPs-binários, URPs, ReLU-RPs, as 24.000 amostras de cada sinal foram segmentadas em partes menores de 400 amostras sem sobreposição. Com isso, para cada arquivo de voz obteve-se um total de 60 segmentos.

Na obtenção dos RPs-binário, URPs e ReLU-RPs foi aplicado o método de Cao (1997) para encontrar a dimensão de imersão m ótima e o método Fraser e Swinney (1986) para encontrar o passo de reconstrução τ ótimo. O cálculo do ε ótimo para gerar o RP-binário foi feito com base na taxa de recorrência (RR). Como os gráficos de recorrência foram analisados pela CNN, optou-se por gerar dois RPs-binários para cada segmento. Um com uma RR aproximada de 5%, baseada no trabalho de Vieira et al. (2018) e outra mantendo uma maior quantidade de informações no RP, com uma RR aproximada de 20%. Já para os URPs, que são representados como uma matriz de distância entre um estado e todos os outros, há apenas uma possibilidade de gráfico. No caso dos ReLU-RPs, D é a distância filtrada pela função ReLU entre dois estados do espaço de fases. Para que o valor da distância D seja mantido no gráfico, é necessário escolher um ε ótimo que, não seja muito grande, evitando perda

de informações importantes do URP, nem muito pequeno, para que não se torne muito similar ao URP. Portanto, o valor do limiar deve ser precisamente escolhido para que as informações do gráfico sejam significativas. Dessa forma, para cada segmento foram gerados três ReLU-RPs, considerando $\epsilon = 0,1$, $\epsilon = 0,2$ e $\epsilon = 0,3$, sendo possível, por meio dos resultados obtidos com a CNN, analisar qual deles guarda uma quantidade ótima de informações.

Na Figura 52 é ilustrado quantos gráficos foram gerados por segmento. Cada conjunto de gráficos gerados foi utilizado individualmente para treinar uma CNN com a mesma arquitetura.

Figura 52 – Diferentes RPs gerados a partir de um segmento do sinal de voz.



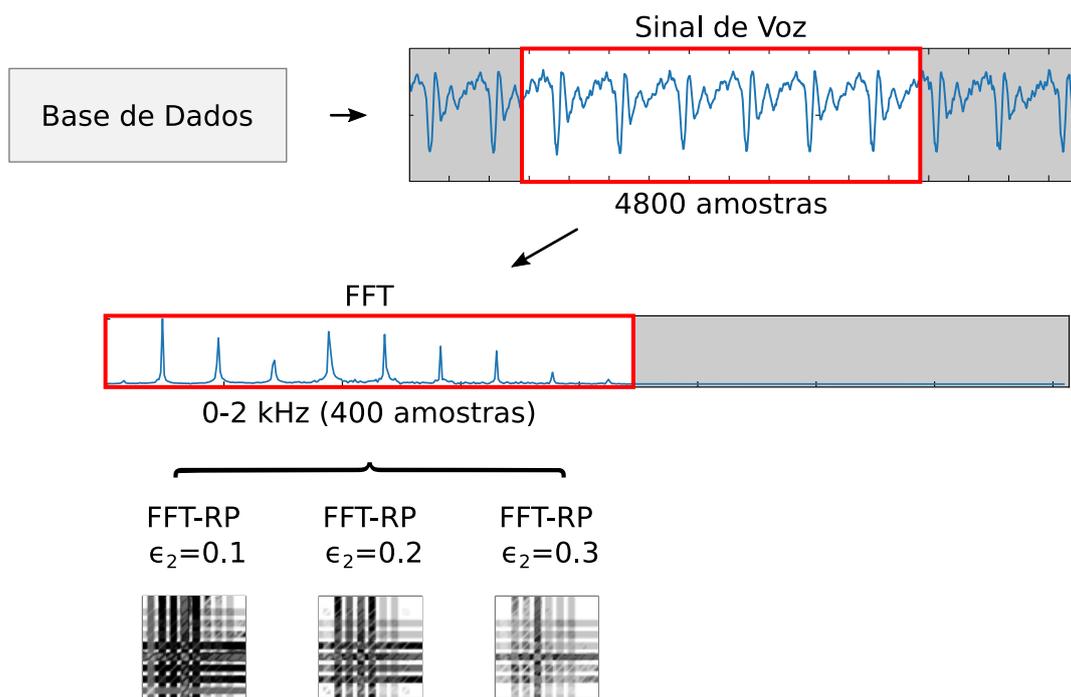
Fonte: autoria própria.

Para a geração dos FFT-RPs também foram considerados apenas as primeiras 24.000 amostras de cada sinal. Contudo, cada sinal foi segmentado em partes maiores, correspondentes a 4.800 amostras. Isso porque, ao fazer a FFT do segmento, que também retorna 4800 amostras, metade corresponde ao espelhamento do espectro, que é desconsiderado, junto com a componente DC. Além disso, o intervalo do espectro dos sinais de voz no qual está contida as informações mais significativas é entre 0 e 2 kHz. Sendo assim, foram selecionadas as primeiras 400 amostras do espectro, que correspondem a uma frequência de 0 e 2083 Hz. Com isso, a mesma quantidade final de amostras é utilizada para gerar tanto o FFT-RP, quanto os demais gráficos de recorrência abordados neste trabalho.

Na Figura 53 é ilustrado quantos gráficos foram gerados por segmento. Cada conjunto de gráficos gerados foi utilizado individualmente para treinar uma CNN com a mesma arquitetura.

As dimensões finais de todos os tipos de RPs, apesar de suas particularidades, estão diretamente relacionadas aos valores obtidos de m e de τ . Dessa forma, cada

Figura 53 – Diferentes RPs gerados a partir da transformada de Fourier de um segmento do sinal de voz.



Fonte: autoria própria.

segmento pode ter tamanhos diferentes de acordo com os valores obtidos por meios dos métodos de Cao (1997) e de Fraser e Swinney (1986), conforme demonstrado na Equação 2. Contudo, as matrizes utilizadas como entrada da CNN devem ter o mesmo tamanho. A fim de não utilizar nenhuma técnica de redimensionamento nos gráficos de recorrência, alterando assim as suas características fundamentais, optou-se por deixar todos os RPs com um tamanho final de 310×310 . Apesar da perda de informações, as principais características relacionadas ao período e ao espectro do sinal conseguem ser mantidas na matriz final. Na Figura 54 é ilustrado como foi feito o corte nos RPs.

Figura 54 – Corte do gráfico de recorrência.



Fonte: autoria própria.

Em caso de segmentos cujo N obtivesse um valor inferior à 310, este segmento é desconsiderado e não é analisado pela rede. Dessa forma, foram desconsiderados

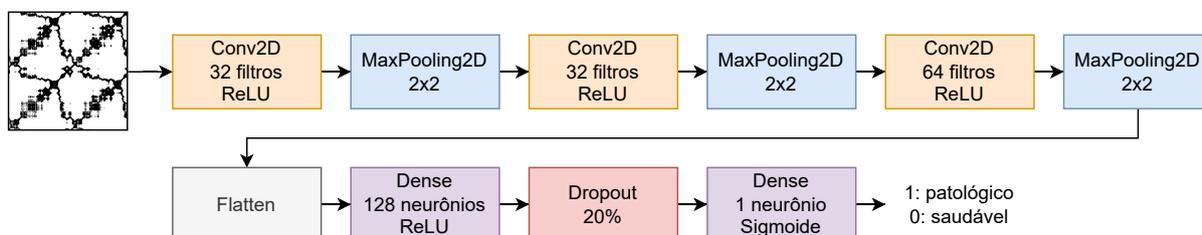
apenas os segmentos referentes aos seguintes sinais da base Kay Elemetrics (1994): Edema (segmento 16 do sinal RYL28) e Paralisia (segmento 37 do sinal SCH15AN, segmento 56 do sinal AJM05AN e segmento 15 do sinal RPJ15AN). Esses segmentos foram desconsiderados apenas nos conjuntos dos gráficos que utilizam o sinal no domínio do tempo (RPs-binários, URPs e ReLU-RPs). Para os FFT-RPs não ocorreram perdas de segmentos.

4.3 CLASSIFICADOR

Em trabalho anterior, Barros et al. (2020) abordaram os RPs binários como imagens de texturas, devido aos padrões e estruturas formados ao longo da matriz. Por isso, a CNN desenvolvida no estudo teve sua organização baseada na pesquisa de Andrearczyk (2017), que avaliou as melhores arquiteturas de CNN para aplicação em imagens de textura.

Na arquitetura proposta por Barros et al. (2020), a rede consistiu em 3 camadas convolucionais seguidas de operadores *pooling*, uma camada *flatten* e duas camadas totalmente conectadas. As duas primeiras camadas convolucionais foram compostas por 32 filtros de dimensões 3×3, e a terceira por 64 filtros de dimensões 3×3. Neste trabalho, foram mantidas todas as configurações antes da camada *flatten*. Contudo, observou-se melhores resultados ao utilizar apenas uma camada densa (totalmente conectada) com 128 neurônios, seguida de uma camada de *dropout* de 20% antes da última camada. Na camada com 128 neurônios foi utilizada a função de ativação ReLU, com a principal finalidade de evitar um problema conhecido como fuga do gradiente. Essa função também possibilita um treinamento mais rápido, sem comprometer a performance da generalização. Na última camada totalmente conectada, composta de apenas um único neurônio, foi utilizada a função de ativação Sigmoide, dado que a classificação da rede é binária. Na Figura 55 é ilustrada com detalhes a arquitetura utilizada no treinamento de todos os conjuntos de RPs produzidos neste estudo.

Figura 55 – Arquitetura da CNN para classificação de gráficos de recorrência.



Fonte: autoria própria.

Para esta mesma arquitetura, foram testadas taxas de aprendizagem de 10^{-3} e 10^{-4} . Os treinamentos com taxa de aprendizagem de 10^{-4} obtiveram os melhores

resultados, sendo este o valor considerado em todos os treinamentos.

Os resultados da classificação de cada segmento, por sua vez, foram utilizados para inferir a classificação de um sinal de voz completo. Para os RPs-binários, URPs e ReLU-RPs, foram classificados 60 segmentos, exceto para os segmentos desconsiderados por $N < 310$. Caso a maioria dos segmentos fossem classificados como saudáveis, o sinal completo pode ser considerado saudável. Caso a maioria dos segmentos fossem classificados como patológicos, o sinal completo pode ser considerado patológico. Em caso de empate, o sinal é considerado patológico.

A mesma regra se aplica à classificação dos FFT-RPs, contudo, como a quantidade de segmentos classificados é 5 por sinal de voz, dessa forma, não é possível ocorrer empate.

A fim de garantir que a rede proverá uma boa generalização independente da divisão do conjunto de dados, foi utilizada técnica de validação cruzada *stratified group k-fold*, com $k = 5$.

5. RESULTADOS

Este capítulo apresenta os resultados obtidos nos treinamentos da CNN utilizando individualmente os conjuntos de gráficos de recorrência gerados. Foram detalhados os valores obtidos de acurácia, precisão, sensibilidade, especificidade e F1-score tanto para a classificação individual dos segmentos como para a classificação dos arquivos de áudio completos.

5.1 ANÁLISE DOS RESULTADOS

Os valores obtidos durante o processo de predição da CNN fornecem uma comparação entre a classificação atribuída pela rede e a classificação real das amostras. As quatro saídas obtidas do classificador são especificadas na Figura 56.

Figura 56 – Relação entre predição da rede e classificação real.

		Previsto	
		patológico	saudável
Verdadeiro	patológico	TP	FN
	saudável	FP	TN

Fonte: Próprio autor.

Verdadeiro Positivo (TP): Amostras corretamente classificadas como patológicas

Falso Positivo (FP): Amostras erroneamente classificadas como patológicas

Verdadeiro Negativo (TN): Amostras corretamente classificadas como saudáveis

Falso Negativo (FN): Amostras erroneamente classificadas como saudáveis

Para analisar os resultados obtidos nesse estudo foram utilizadas as métricas listadas a seguir, calculadas com base nos parâmetros TP, FP, TN e FN, contribuindo para a avaliação do desempenho da rede.

Acurácia: Indica a habilidade da rede de diagnosticar corretamente as amostras. A expressão matemática correspondente é definida na Equação 16.

$$\text{Acurácia} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (16)$$

Precisão: Indica a habilidade da rede de classificar as amostras positivas dentre todos os segmentos classificados como positivos. Ou seja, dentro de tudo o que foi classificado como positivo, o que é de fato positivo. A expressão matemática correspondente é definida na Equação 17.

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (17)$$

Sensibilidade: Mede a habilidade da rede de classificar corretamente as amostras patológicas, isto é, qual percentual de amostras patológicas que foram classificadas como tal pela rede. Matematicamente, essa métrica pode ser definida pela Equação 18.

$$\text{Sensibilidade} = \frac{TP}{TP + FN} \quad (18)$$

Especificidade: Mede a habilidade da rede de determinar corretamente as amostras saudáveis, isto é, qual percentual de amostras saudáveis que foram classificadas como tal pela rede. Essa métrica pode ser obtida por meio da Equação 19.

$$\text{Especificidade} = \frac{TN}{TN + FP} \quad (19)$$

F1-score: É definida como a média harmônica de precisão e sensibilidade. Essa métrica pode ser obtida por meio da Equação 20.

$$\text{F1-score} = \frac{2 \cdot \text{precisão} \cdot \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}} \quad (20)$$

Para a identificação dos melhores resultados foram consideradas primeiramente a acurácia, seguida da sensibilidade. Na Tabela 1 são apresentados os valores obtidos na classificação dos RP-binários gerados a partir de cada segmento. Observa-se que, para a maioria das métricas, os RP-binários com RR=5% obtém melhores resultados. A acurácia indica que o conjunto com RR=5% teve uma melhor capacidade de classificar corretamente as amostras. Por outro lado, a sensibilidade, que mede a capacidade da rede de classificar corretamente as patologias, foi melhor para o conjunto gerado com RR=20%. No geral, a performance dos dois conjuntos de gráficos é bastante similar e caracterizam os sinais de voz de forma semelhante, mas para quesito de comparação com os demais tipos de RPs, considera-se que o RP-binários com RR=5% possuem o melhor desempenho, com base na acurácia.

Como explicado na seção 4.3, os resultados da classificação de cada segmento, por sua vez, foram utilizados para inferir a classificação de um sinal de voz completo.

Tabela 1 – Resultados para classificação dos segmentos de áudio usando RP-binários.

	acurácia	precisão	sensibilidade	especificidade	F1-score
RP-binário (RR=5%)	0,864 ± 0,031	0,892 ± 0,054	0,924 ± 0,041	0,770 ± 0,128	0,882 ± 0,028
RP-binário (RR=20%)	0,858 ± 0,018	0,868 ± 0,047	0,934 ± 0,021	0,739 ± 0,067	0,880 ± 0,016

Caso a maioria dos segmentos sejam classificados como saudáveis, o sinal completo pode ser considerado saudável. Caso a maioria dos segmentos sejam classificados como patológicos, o sinal completo pode ser considerado patológico. Em caso de empate, o sinal é considerado patológico. Na Tabela 2, são apresentados os resultados da classificação dos arquivos de áudio completos, baseado na classificação majoritária dos segmentos, como indicado. Observa-se que, ainda assim, os RP-binários gerados com RR=5% e com RR=20% obtiveram resultados bastante similares. Contudo, como a principal métrica a ser considerada é a acurácia, nota-se que, apesar do empate, o desvio padrão para o conjunto gerado com base em uma RR=20% é menor. Além disso, a sensibilidade para este conjunto também possui um valor superior com um desvio padrão menor, sendo este o melhor conjunto para a classificação dos áudios completos.

Tabela 2 – Resultados para classificação dos sinais de áudio completos usando RP-binários.

	acurácia	precisão	sensibilidade	especificidade	F1-score
RP-binário (RR=5%)	0,894 ± 0,054	0,905 ± 0,093	0,936 ± 0,040	0,829 ± 0,176	0,916 ± 0,038
RP-binário (RR=20%)	0,894 ± 0,042	0,875 ± 0,047	0,961 ± 0,032	0,793 ± 0,088	0,915 ± 0,032

Na Tabela 3 são exibidos os resultados obtidos com o conjunto de URPs gerados a partir dos segmentos dos áudios. Observa-se que os valores de todas as métricas foram superiores aos obtidos com ambos conjuntos de RPs-binários, detalhados na Tabela 1.

Tabela 3 – Resultados para classificação dos segmentos de áudio usando URPs.

	acurácia	precisão	sensibilidade	especificidade	F1-score
URP	0,925 ± 0,046	0,961 ± 0,056	0,957 ± 0,023	0,922 ± 0,107	0,932 ± 0,030

Na Tabela 4 são exibidos os resultados da classificação dos arquivos de áudio completos, baseado na classificação majoritária dos segmentos caracterizados como URPs. A classificação é consistente com a classificação dos segmentos, indicando uma alta relação entre a categorização dos segmentos e dos áudios como um todo. A acurácia para esse conjunto é 4.5% maior que ambos conjuntos de RP-binários citados anteriormente.

Na Tabela 5 são apresentados os resultados obtidos com os conjuntos de ReLU-RPs gerados a partir dos segmentos dos áudios. Os ReLU-RPs destacam as estruturas nas quais a recorrência é mínima, ou seja, as áreas do URP nas quais a

Tabela 4 – Resultados para classificação dos sinais de áudio completos usando URPs.

	acurácia	precisão	sensibilidade	especificidade	F1-score
URP	0,939 ± 0,052	0,967 ± 0,067	0,935 ± 0,042	0,940 ± 0,120	0,949 ± 0,040

distância entre os estados é maior que um limiar ε . Portanto, o valor do limiar deve ser precisamente escolhido para que as informações do gráfico sejam significativas. Observa-se que, o conjunto de ReLU-RPs gerados com um $\varepsilon = 0,1$ obtiveram a melhor classificação dos segmentos. Apesar da sutil diferença de resultados obtidos com os diferentes valores de ε , com exceção da especificidade, todas as outras métricas apontam que aumentar o valor de ε tende a descartar informações relevantes para a caracterização dos gráficos. Já em comparação com os resultados obtidos com os URPs (Tabela 3), que consistem em um ReLU-RP com $\varepsilon = 0$, os conjunto de ReLU-RPs com $\varepsilon = 0,1$ obtiveram uma performance superior em todas as métricas.

Tabela 5 – Resultados para classificação dos segmentos de áudio usando ReLU-RPs.

	acurácia	precisão	sensibilidade	especificidade	F1-score
ReLU-RP ($\varepsilon = 0,1$)	0,929 ± 0,041	0,973 ± 0,046	0,958 ± 0,021	0,872 ± 0,061	0,930 ± 0,028
ReLU-RP ($\varepsilon = 0,2$)	0,918 ± 0,045	0,954 ± 0,053	0,955 ± 0,028	0,937 ± 0,043	0,933 ± 0,037
ReLU-RP ($\varepsilon = 0,3$)	0,920 ± 0,049	0,953 ± 0,065	0,953 ± 0,024	0,926 ± 0,105	0,936 ± 0,036

Na Tabela 6 são exibidos os resultados da classificação dos arquivos de áudio completos, baseado na classificação majoritária dos segmentos caracterizados como ReLU-RPs. Constata-se que a variação entre as classificações é bastante sutil, assim como para a classificação dos segmentos, o que justifica os resultados obtidos quando utilizados os ReLU-RPs com $\varepsilon = 0,3$. Contudo, é possível destacar que, os resultados obtidos pelos ReLU-RPs é significativamente superior aos obtidos pelos URPs.

Tabela 6 – Resultados para classificação dos sinais de áudio completos usando ReLU-RPs.

	acurácia	precisão	sensibilidade	especificidade	F1-score
ReLU-RP ($\varepsilon = 0,1$)	0,954 ± 0,057	0,967 ± 0,067	0,961 ± 0,032	0,940 ± 0,120	0,963 ± 0,043
ReLU-RP ($\varepsilon = 0,2$)	0,939 ± 0,067	0,973 ± 0,053	0,923 ± 0,061	0,960 ± 0,080	0,948 ± 0,056
ReLU-RP ($\varepsilon = 0,3$)	0,962 ± 0,042	0,977 ± 0,047	0,961 ± 0,032	0,960 ± 0,080	0,968 ± 0,033

Na Tabela 7 são apresentados os resultados obtidos com os conjuntos de FFT-RPs gerados a partir dos segmentos dos áudios. Para todos os conjuntos de FFT-RPs gerados, foi mantido o valor superior do raio em $\varepsilon_1 = 0,8$ e os valores inferiores foram variados entre $\varepsilon_2 = 0,1$, $\varepsilon_2 = 0,2$ e $\varepsilon_2 = 0,3$. Comparado aos demais tipos de RPs, o FFT-RP obteve performance inferior na caracterização dos sinais, considerando principalmente a acurácia. Além disso, obteve um desempenho similar para todas variações propostas.

Tabela 7 – Resultados para classificação dos segmentos de áudio usando FFT-RPs.

	acurácia	precisão	sensibilidade	especificidade	F1-score
FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,1$)	0,800 \pm 0,026	0,850 \pm 0,058	0,903 \pm 0,073	0,741 \pm 0,128	0,852 \pm 0,032
FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,2$)	0,818 \pm 0,033	0,867 \pm 0,038	0,938 \pm 0,053	0,739 \pm 0,028	0,867 \pm 0,029
FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,3$)	0,806 \pm 0,041	0,876 \pm 0,026	0,912 \pm 0,060	0,840 \pm 0,052	0,839 \pm 0,053

Analisando a Tabela 8, nota-se que, para os FFT-RPs, a acurácia obtida por meio da inferência de resultado utilizando a classificação dos segmentos aumentou significativamente. Esse comportamento não foi observada para os demais gráficos.

Tabela 8 – Resultados para classificação dos sinais de áudio completos usando FFT-RPs.

	acurácia	precisão	sensibilidade	especificidade	F1-score
FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,1$)	0,878 \pm 0,044	0,887 \pm 0,074	0,922 \pm 0,048	0,811 \pm 0,132	0,901 \pm 0,032
FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,2$)	0,885 \pm 0,043	0,878 \pm 0,012	0,934 \pm 0,073	0,811 \pm 0,009	0,904 \pm 0,042
FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,3$)	0,862 \pm 0,047	0,917 \pm 0,020	0,844 \pm 0,092	0,885 \pm 0,043	0,876 \pm 0,054

Na Tabela 9 é exibido um resumo dos melhores resultados obtidos por tipo de RP, considerando a acurácia da sua melhor variação para a classificação de segmentos. Nesse sentido, podemos avaliar que os ReLU-RPs apresentaram uma melhor capacidade de distinguir corretamente os segmentos saudáveis dos patológicas. Os ReLU-RPs também se destacam por apresentarem a menor dispersão, baseado na precisão, que indica a proporção de diagnósticos positivos que estão corretos, ou seja, a proporção de segmentos que a rede diagnosticou como patológicos que são de fato patológicos. Para esse mesmo conjunto de gráficos, a sensibilidade indica que a proporção de segmentos patológicos que foram diagnosticadas como patológicos de fato, ou seja, a proporção de amostras patológicas que foram diagnosticadas como tal. Já o conjunto de RPs que melhor diagnosticou os casos verdadeiramente negativos, ou seja, segmentos saudáveis, foram os URPs, como apontado pela especificidade. Os URPs também obtiveram um desempenho levemente superior ao dos ReLU-RPs no F1-score, que combina a precisão e a sensibilidade.

Tabela 9 – Melhores resultados para classificação dos segmentos de áudio.

	acurácia	precisão	sensibilidade	especificidade	F1-score
RP-binário (RR=5%)	0,864 \pm 0,031	0,892 \pm 0,054	0,924 \pm 0,041	0,770 \pm 0,128	0,882 \pm 0,028
URP	0,925 \pm 0,046	0,961 \pm 0,056	0,957 \pm 0,023	0,922 \pm 0,107	0,932 \pm 0,030
ReLU-RP ($\varepsilon = 0,1$)	0,929 \pm 0,041	0,973 \pm 0,046	0,958 \pm 0,021	0,872 \pm 0,061	0,930 \pm 0,028
FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,2$)	0,818 \pm 0,033	0,867 \pm 0,038	0,938 \pm 0,053	0,739 \pm 0,028	0,867 \pm 0,029

Analisando a classificação dos áudios como um todo, realizada por meio da inferência dos resultados dos segmentos, há pequenas variações de desempenho entre o URP e o ReLU-RP. Contudo, para a análise final dos sinais completos, o ReLU-RP obteve o melhor desempenho, podendo ser considerado o conjunto que melhor caracterizou os sinais de vozes, sendo capaz de distinguir sinais saudáveis

de patológicas com a melhor eficácia. Ao contrário dos demais tipos de RPs, o RP-binário foi o único que apresentou um melhor resultado na classificação individual dos segmentos com um parâmetro (RR=5%), mas apresentou um resultado melhor na classificação geral dos áudios com outro parâmetro (RR=20%).

Tabela 10 – Melhores resultados para classificação dos arquivos de áudio completos.

	acurácia	precisão	sensibilidade	especificidade	F1-score
RP-binário (RR=20%)	0,894 ± 0,042	0,875 ± 0,047	0,961 ± 0,032	0,793 ± 0,088	0,915 ± 0,032
URP	0,939 ± 0,052	0,967 ± 0,067	0,935 ± 0,042	0,940 ± 0,120	0,949 ± 0,040
ReLU-RP ($\varepsilon = 0,3$)	0,962 ± 0,042	0,977 ± 0,047	0,961 ± 0,032	0,960 ± 0,080	0,968 ± 0,033
FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,2$)	0,885 ± 0,043	0,878 ± 0,012	0,934 ± 0,073	0,811 ± 0,009	0,904 ± 0,042

5.2 DISCUSSÃO DOS RESULTADOS

De acordo com os resultados analisados na seção anterior, baseando-se primeiramente na acurácia, o ReLU-RP com $\varepsilon = 0,3$ ficou em primeiro lugar na classificação dos áudios completos, obtendo uma acurácia média de 96,2%. Em segundo lugar ficou o URP com uma acurácia média de 93,9%, seguido do RP-binário com uma RR aproximada de 20% e do FFT-RP ($\varepsilon_1 = 0,8, \varepsilon_2 = 0,2$), com respectivamente 89,4% e 88,5% de acurácia.

No estudo anterior de Barros et al. (2020), que utilizou uma abordagem diferente para a distinção de áudios patológicos de áudios saudáveis aplicada a mesma base de dados, a acurácia média obtida foi de 98,3%. Comparado ao melhor resultado obtido neste trabalho, que foi para a base de ReLU-RPs com $\varepsilon = 0,3$, este resultado é cerca de 2,1% acima do obtido com os ReLU-RPs. Contudo, este estudo leva em consideração cerca de 60 segmentos por arquivo de áudio para a classificação individual dos ReLU-RPs, sendo estes segmentos utilizados para a inferência do resultado geral de um áudio completo, o que denota uma confiabilidade superior do que a apresentada no estudo de Barros et al. (2020), cujo melhor resultado foi obtido com base na classificação de 3 segmentos por arquivo de áudios, sendo os mesmos utilizados para a inferência da classificação do áudio completo. Já quando comparada a sensibilidade dos dois trabalhos, os ReLUs-RPs mostraram uma capacidade superior na distinção de amostras que de fato são patológicas.

O presente estudo também traz resultados superiores aos obtidos no estudo de Souza et al. (2015), que também foi utilizado para comparação no trabalho de Barros et al. (2020) e utiliza uma abordagem baseada na análise de textura para a classificação dos áudios da mesma base de dados. Souza et al. (2015) apresentaram uma acurácia média de 86,3%, uma sensibilidade média de 66,52% e uma especificidade média de 89,83%. Este estudo apresenta valores superiores para todas essas métricas, tendo um aumento de respectivamente 9,9%, 29,6% e 6,2%.

Outro estudo que também propôs um método baseado no uso de gráficos de recorrência, contudo com uma metodologia baseada nas métricas de RQA, foi proposto por Vieira et al. (2018). Utilizando a mesma base de dados considerada neste trabalho, os autores obtiveram uma acurácia de 92,48%, sensibilidade de 93,26% e especificidade de 90,67%. Para os ReLU-RPs com $\varepsilon = 0,3$ foram obtidos um acréscimo nessas métricas de 3,7%, 2,8% e 5,33% respectivamente. Contudo, pode-se analisar que, o trabalho de Vieira et al. (2018) propôs um método mais eficiente para analisar os RPs-binários, que obteve valores superiores de acurácia e especificidade se comparado aos resultados obtidos nesse estudo para os RP-binários com RR=20%. Ainda assim, a sensibilidade foi superior nesse trabalho, atingindo uma taxa de 96,1%.

Em outro estudo mais recente, Lopes, Vieira e Behlau (2020) comparam a eficiência da quantificação de recorrência com medidas tradicionais, cepstrais e não lineares para a discriminação de sinais vocais com e sem distúrbio de voz, contudo utilizando uma base de dados diferente. O desempenho do classificador também foi avaliado ao combinar tais medidas acústicas. Os autores observaram que, ao combinar diferentes tipos de medidas, apenas uma combinação obteve os melhores resultados (86,43% de acurácia, 90,83% de sensibilidade e 80,83% de especificidade), consistindo na combinação da GNE e da ENTR. Apesar de uma combinação direta não poder ser estabelecida devido às diferentes bases de dados utilizados pelos estudos, pode-se sugerir que um melhor resultado poderia ser obtido pela mesma base utilizando o método proposto nesse trabalho.

6. CONSIDERAÇÕES FINAIS

Este trabalho se propôs a investigar a eficácia dos RP-binários, URPs, ReLU-RPs e FFT-RP para a discriminação de sinais de voz afetados por patologias laríngeas de sinais saudáveis. A base de dados utilizada nessa pesquisa foi gravada pelo Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab. Uma CNN de mesma arquitetura foi utilizada para treinar e classificar individualmente os gráficos de recorrência dos quatro tipos indicados nesse estudo, que foram geradas a partir de dos sinais da base Kay Elemetrics (1994). Além disso, para os RPs-binário, ReLU-RPs e FFT-RPs foram geradas mais de uma base de dados para avaliar como a alteração de parâmetros importantes para a geração desses gráficos de recorrência afetariam a caracterização dos sinais de voz.

Nesse cenário, este trabalho contribuiu de forma ímpar para a literatura relacionada à análise de gráficos de recorrência para caracterização e análise de sinais de voz e discriminação de patologias. No presente estudo, foram empregados, além dos RPs-binários, os URPs, ReLU-RPs e FFT-RPs, que nunca haviam sido utilizados para caracterização desse tipo de sinal. Dessa forma, foi possível identificar que os ReLU-RPs tem uma capacidade superior para a discriminação de patologias quando analisados por uma CNN. Na classificação dos áudios completos, todas as métricas discutidas nesse trabalho obtiveram valores iguais ou superiores à 96%, sendo elas acurácia, precisão, sensibilidade, especificidade e F1-score. Isso também indica que a rede tem uma capacidade equilibrada tanto para a análise de amostras classificadas como patológicas que são de fato patológicas, como para amostras classificadas como saudáveis que são de fato saudáveis.

Por fim, pode-se concluir que este trabalho obteve êxito no uso de outros tipos de gráficos de recorrência que não os RPs-binários, comumente usados em trabalhos que envolvem a análise de sinais com relativa periodicidade. Dessa forma, este trabalho pode ser utilizado como base para que trabalhos envolvendo tais sinais provenientes de diferentes fontes possam investigar a eficiência de outros tipos de gráficos em suas análises.

REFERÊNCIAS

- AFONSO, Luis CS; ROSA, Gustavo H; PEREIRA, Clayton R; WEBER, Silke AT; HOOK, Christian; ALBUQUERQUE, Victor Hugo C; PAPA, João P. A recurrence plot-based approach for parkinson's disease identification. **Future Generation Computer Systems**, Elsevier, v. 94, p. 282–292, 2019.
- ANDREARCZYK, Vincent. **Deep learning for texture and dynamic texture analysis**. Tese (Doutorado) — Dublin City University, 2017.
- BARROS, Luana Rodrigues; SOARES, Gabriel Gutierrez Pereira; DUARTE, Gabriel Daltro; CORREIA, Suzete Élide Nóbrega; COSTA, Silvana Luciene do Nascimento Cunha. Classification of recurrence plots of voice signals using convolutional neural networks. **XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais**, SBRT, 2020.
- BEHLAU, MARA SUZANA. **Voz O Livro do Especialista Rio de Janeiro: Revinter**. 2001.
- BRAGA, Antônio de Pádua; FERREIRA, André Carlos Ponce de Leon; LUDERMIR, Teresa Bernarda. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2007.
- BUDUMA, Nikhil; LOCASCIO, Nicholas. **Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms**. Sebastopol, Canada: O'Reilly Media, Inc., 2017.
- CAO, Liangyue. Practical method for determining the minimum embedding dimension of a scalar time series. **Physica D: Nonlinear Phenomena**, Elsevier, v. 110, n. 1-2, p. 43–50, 1997.
- CASDAGLI, Martin; EUBANK, Stephen; FARMER, J Doyne; GIBSON, John. State space reconstruction in the presence of noise. **Physica D: Nonlinear Phenomena**, Elsevier, v. 51, n. 1-3, p. 52–98, 1991.
- CEDARS-SINAI. **Voice Disorders**. 2020. Disponível em: <<https://www.cedars-sinai.org/health-library/diseases-and-conditions/v/voice-disorders.html>>. Acesso em: 2020-12-04.
- CHEN, Yun; SU, Shijie; YANG, Hui. Convolutional neural network analysis of recurrence plots for anomaly detection. **International Journal of Bifurcation and Chaos**, World Scientific, v. 30, n. 01, p. 2050002, 2020.
- CIELO, Carla Aparecida; FINGER, Leila Susana; ROSA, Juliana de Caldas; BRANCALIONI, Ana Rita. Lesões organofuncionais do tipo nódulos, pólipos e edema de reinke. **Revista CEFAC**, SciELO Brasil, v. 13, p. 735–748, 2011.
- COSTA, Silvana L. do N. C.; CORREIA, Suzete E. N.; COSTA, Washington C. de A.; VIEIRA, Vinícius J. D. Análise de sinais por meio de gráficos de recorrência. In: CIDRIM, L.; COSTA, S. C. **Tecnologias da Informação e da Comunicação (TIC) Aplicadas às Ciências da Linguagem**. Curitiba: CRV, 2015. cap. 6, p. 83–97.

ECKMANN, J.-P; KAMPHORST, S. Oliffson; RUELLE, D. Recurrence plots of dynamical systems. **Europhysics Letters (EPL)**, IOP Publishing, v. 4, n. 9, p. 973–977, nov 1987. Disponível em: <<https://doi.org/10.1209/0295-5075/4/9/004>>.

FACCHINI, A; KANTZ, H. Curved structures in recurrence plots: The role of the sampling time. **Physical Review E**, APS, v. 75, n. 3, p. 036215, 2007.

FACELI, Katti; LORENA, Ana Carolina; GAMA, João; CARVALHO, André Carlos Ponce de Leon et al. **Inteligência Artificial: Uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2019.

FRASER, Andrew M; SWINNEY, Harry L. Independent coordinates for strange attractors from mutual information. **Physical review A**, APS, v. 33, n. 2, p. 1134, 1986.

GARCIA-CEJA, Enrique; UDDIN, Md Zia; TORRESEN, Jim. Classification of recurrence plots' distance matrices with a convolutional neural network for activity recognition. **Procedia computer science**, Elsevier, v. 130, p. 157–163, 2018.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems**. 2. ed. Sebastopol, Canada: O'Reilly Media, Inc., 2019.

GODOY, Moacir Fernandes de; GREGÓRIO, Michele Lima. Relevância diagnóstica dos gráficos de recorrência na caracterização de saúde, doença ou morte, em humanos. **Journal of Human Growth and Development**, v. 29, n. 1, p. 39–47, 2019.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. MIT Press, 2016. Disponível em: <<http://www.deeplearningbook.org>>.

HATAMI, Nima; GAVET, Yann; DEBAYLE, Johan. Classification of time-series images using deep convolutional neural networks. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Tenth international conference on machine vision (ICMV 2017)**. Vienna, Austria, 2018. v. 10696, p. 106960Y.

HAYKIN, Simon. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.

HSUEH, Yumin; ITTANGIHALA, Veerasha Ramesha; WU, Wei-Bin; CHANG, Hong-Chan; KUO, Cheng-Chien. Condition monitor system for rotation machine by CNN with recurrence plot. **Energies**, Multidisciplinary Digital Publishing Institute, v. 12, n. 17, p. 3221, 2019.

KAY ELEMETRICS. **Disordered Voice Database, model 4337, 3a ed.** 1994. 1 CD.

KENNEL, Matthew B; BROWN, Reggie; ABARBANEL, Henry DI. Determining embedding dimension for phase-space reconstruction using a geometrical construction. **Physical review A**, APS, v. 45, n. 6, p. 3403, 1992.

KRAKOVSKÁ, Anna; MEZEIOVÁ, Kristína; BUDÁČOVÁ, Hana. Use of false nearest neighbours for selecting variables and embedding parameters for state space reconstruction. **Journal of Complex Systems**, Hindawi, v. 2015, 2015.

- LADEIRA, Guênia; MARWAN, Norbert; DESTRO-FILHO, João-Batista; RAMOS, Camila Davi; LIMA, Gabriela. Frequency spectrum recurrence analysis. **Scientific Reports**, Nature Publishing Group, v. 10, n. 1, p. 1–9, 2020.
- LEE, Hyeonjeong; LEE, Jaewon; SHIN, Miyoung. Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots. **Electronics**, Multidisciplinary Digital Publishing Institute, v. 8, n. 2, p. 192, 2019.
- LEWENSTEIN, K; JAMROŻY, M; LEYKO, T. The use of recurrence plots and beat recordings in chronic heart failure detection. **Bulletin of the Polish Academy of Sciences. Technical Sciences**, v. 64, n. 2, p. 339–345, 2016.
- LOPES, Leonardo; VIEIRA, Vinícius; BEHLAU, Mara. Performance of different acoustic measures to discriminate individuals with and without voice disorders. **Journal of Voice**, Elsevier, 2020.
- MAKOWSKI, Dominique; PHAM, Tam; LAU, Zen J.; BRAMMER, Jan C.; LESPINASSE, François; PHAM, Hung; SCHÖLZEL, Christopher; CHEN, S. H. Annabel. NeuroKit2: A python toolbox for neurophysiological signal processing. **Behavior Research Methods**, Springer Science and Business Media LLC, v. 53, n. 4, p. 1689–1696, feb 2021. Disponível em: <<https://doi.org/10.3758/s13428-020-01516-y>>.
- MARTINS, Regina Helena Garcia; AMARAL, Henrique Abrantes do; TAVARES, Elaine Lara Mendes; MARTINS, Maira Garcia; GONÇALVES, Tatiana Maria; DIAS, Norimar Hernandes. Voice disorders: etiology and diagnosis. **Journal of Voice**, Elsevier, v. 30, n. 6, p. 761–e1, 2016.
- MARWAN, Norbert. **Encounters with neighbours: current developments of concepts based on recurrence plots and their applications**. Tese (Doutorado) — Universität Potsdam, set. 2003.
- MARWAN, Norbert; WESSEL, Niels; MEYERFELDT, Udo; SCHIRDEWAN, Alexander; KURTHS, Jürgen. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. **Physical Review E**, American Physical Society, v. 66, p. 026702, Aug 2002. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevE.66.026702>>.
- MCGUIRE, Gary; AZAR, Nabeel B.; SHELHAMER, Mark. Recurrence matrices and the preservation of dynamical properties. **Physics Letters A**, v. 237, n. 1, p. 43–47, 1997. ISSN 0375-9601. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S037596019700697X>>.
- MIRANDA, Gustavo Xavier Andrade. **Analyzing of the vocal fold dynamics using laryngeal videos**. Tese (Doutorado) — Escuela Técnica Superior de Ingenieros de Telecomunicación - Universidad Politécnica de Madrid, 2017. Disponível em: <<https://oa.upm.es/47122/>>. Acesso em: 2020-12-03.
- MONTEIRO, Luiz Henrique Alves. **Sistemas dinâmicos**. [S.l.]: Editora Livraria da Física, 2006.

- MOON, Young-Il; RAJAGOPALAN, Balaji; LALL, Upmanu. Estimation of mutual information using kernel density estimators. **Physical Review E**, APS, v. 52, n. 3, p. 2318, 1995.
- MUHAMMAD, Ghulam; ALSULAIMAN, Mansour; ALI, Zulfiqar; MESALLAM, Tamer A.; FARAHAT, Mohamed; MALKI, Khalid H.; AL-NASHERI, Ahmed; BENCHERIF, Mohamed A. Voice pathology detection using interlaced derivative pattern on glottal source excitation. **Biomedical Signal Processing and Control**, v. 31, p. 156–164, 2017. ISSN 1746-8094. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1746809416300970>>.
- MUHAMMAD, Ghulam; MESALLAM, Tamer A; MALKI, Khalid H; FARAHAT, Mohamed; ALSULAIMAN, Mansour; BUKHARI, Manal. Formant analysis in dysphonic patients and automatic arabic digit speech recognition. **BioMedical Engineering Online**, BioMed Central, v. 10, n. 1, p. 1–12, 2011.
- MURRY, Thomas; JR, William S Brown; MORRIS, Richard J. Patterns of fundamental frequency for three types of voice samples. **Journal of Voice**, Elsevier, v. 9, n. 3, p. 282–289, 1995.
- PATEL, Rita; PARSRAM, KS. Acoustic analysis of subjects with vocal cord paralysis. **Indian Journal of Otolaryngology and Head and Neck Surgery**, Springer, v. 57, n. 1, p. 48–51, 2005.
- SALTÜRK, Ziya; ÖZDEMİR, Erdi; SARI, Hüseyin; KETEN, Seda; KUMRAL, Tolgar Lütfi; BERKİTEN, Guler; TUTAR, Belgin; UYAR, Yavuz. Assessment of resonant voice therapy in the treatment of vocal fold nodules. **Journal of Voice**, Elsevier, v. 33, n. 5, p. 810–e1, 2019.
- SASAKI, Clarence T. **Overview of Laryngeal Disorders**. MSD Manuals, 2020. Disponível em: <<https://www.msdmanuals.com/professional/ear,-nose,-and-throat-disorders/laryngeal-disorders/overview-of-laryngeal-disorders>>. Acesso em: 2021-07-26.
- SCIKIT-LEARN.ORG. **Cross-validation: evaluating estimator performance**. 2022. Disponível em: <https://scikit-learn.org/stable/modules/cross_validation.html>. Acesso em: 2022-10-22.
- SOUZA, Everton Granemann. **Caracterização de sistemas dinâmicos através de gráficos de recorrência**. Tese (Doutorado) — Universidade Federal do Paraná, 2008. Disponível em: <<https://acervodigital.ufpr.br/handle/1884/17165>>. Acesso em: 2022-08-07.
- SOUZA, Taciana A; SOUZA, Micael A; WASHINGTON, C de A; COSTA, Silvana C; CORREIA, Suzete EN; VIEIRA, Vinícius JD. Feature selection based on binary particle swarm optimization and neural networks for pathological voice detection. In: IEEE. **2015 Latin America Congress on Computational Intelligence (LA-CCI)**. [S.l.], 2015. p. 1–6.
- TAKENS, Floris. Detecting strange attractors in turbulence. In: SPRINGER-VERLAG. **Dynamical Systems and Turbulence, Lecture Notes in Mathematics, vol. 898**. [S.l.], 1981. p. 366–381.

TAVALUC, Raluca; TAN-GELLER, Melin. Reinke's edema. **Otolaryngologic Clinics of North America**, Elsevier, v. 52, n. 4, p. 627–635, 2019.

THE VOICE FOUNDATION. **Understanding How Voice is Produced**. Philadelphia, PA, USA: The Voice Foundation, 2020. Disponível em: <<https://voicefoundation.org/health-science/voice-disorders/anatomy-physiology-of-voice-production/understanding-voice-production/>>. Acesso em: 2020-12-01.

VIEIRA, Vinícius JD; COSTA, Silvana C; CORREIA, Suzete LN; LOPES, Leonardo W; COSTA, Washington C de A; ASSIS, Francisco M de. Exploiting nonlinearity of the speech production system for voice disorder assessment by recurrence quantification analysis. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, AIP Publishing LLC, v. 28, n. 8, p. 085709, 2018.

WEBBER, C. L.; ZBILUT, J. P. Dynamical assessment of physiological systems and states using recurrence plot strategies. **Journal of Applied Physiology**, v. 76, n. 2, p. 965–973, 1994. PMID: 8175612.

WEILL CORNELL MEDICINE. **Reinke's Edema**. New York, NY, USA: Sean Parker Institute for the Voice, 2021. Disponível em: <<https://voice.weill.cornell.edu/voice-disorders/reinkes-edema>>. Acesso em: 2021-07-26.

YONEKAWA, Hiroko. A clinical study of reinke's edema. **Auris Nasus Larynx**, Elsevier, v. 15, n. 1, p. 57–78, 1988.

ZBILUT, Joseph P.; WEBBER, Charles L. Embeddings and delays as derived from quantification of recurrence plots. **Physics Letters A**, v. 171, n. 3, p. 199–203, 1992. ISSN 0375-9601. Disponível em: <<https://www.sciencedirect.com/science/article/pii/037596019290426M>>.

ZHAO, Zhidong; ZHANG, Yang; COMERT, Zafer; DENG, Yanjun. Computer-aided diagnosis system of fetal hypoxia incorporating recurrence plot with convolutional neural network. **Frontiers in physiology**, Frontiers, v. 10, p. 255, 2019.

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