

# APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS

Julio C. Uzinski  
Caio C. E. de Abreu  
Bruno R. de **Oliveira**  

---

organizadores



Pantanal Editora

2020

Julio Cezar Uzinski  
Caio Cesar Enside de Abreu  
Bruno Rodrigues de Oliveira  
(Organizadores)

**Aplicações de  
Inteligência Artificial e  
Ciência de Dados**



Pantanal Editora

2020

Copyright© Pantanal Editora  
Copyright do Texto© 2020 Os Autores  
Copyright da Edição© 2020 Pantanal Editora  
Editor Chefe: Prof. Dr. Alan Mario Zuffo  
Editores Executivos: Prof. Dr. Jorge González Aguilera  
Prof. Dr. Bruno Rodrigues de Oliveira

Diagramação: A editora  
Edição de Arte: A editora  
Revisão: Os autor(es), organizador(es) e a editora

#### Conselho Editorial

- Prof. Dr. Adailson Wagner Sousa de Vasconcelos – OAB/PB
- Profa. Msc. Adriana Flávia Neu – Mun. Faxinal Soturno e Tupanciretã
- Profa. Dra. Albys Ferrer Dubois – UO (Cuba)
- Prof. Dr. Antonio Gasparetto Júnior – IF SUDESTE MG
- Profa. Msc. Aris Verdecia Peña – Facultad de Medicina (Cuba)
- Profa. Arisleidis Chapman Verdecia – ISCM (Cuba)
- Prof. Dr. Bruno Gomes de Araújo - UEA
- Prof. Dr. Caio Cesar Enside de Abreu – UNEMAT
- Prof. Dr. Carlos Nick – UFV
- Prof. Dr. Claudio Silveira Maia – AJES
- Prof. Dr. Cleberton Correia Santos – UFGD
- Prof. Dr. Cristiano Pereira da Silva – UEMS
- Profa. Ma. Dayse Rodrigues dos Santos – IFPA
- Prof. Msc. David Chacon Alvarez – UNICENTRO
- Prof. Dr. Denis Silva Nogueira – IFMT
- Profa. Dra. Denise Silva Nogueira – UFMG
- Profa. Dra. Dennyura Oliveira Galvão – URCA
- Prof. Dr. Elias Rocha Gonçalves – ISEPAM-FAETEC
- Prof. Me. Ernane Rosa Martins – IFG
- Prof. Dr. Fábio Steiner – UEMS
- Prof. Dr. Gabriel Andres Tafur Gomez (Colômbia)
- Prof. Dr. Hebert Hernán Soto Gonzáles – UNAM (Peru)
- Prof. Dr. Hudson do Vale de Oliveira – IFRR
- Prof. Msc. Javier Revilla Armesto – UCG (México)
- Prof. Msc. João Camilo Sevilla – Mun. Rio de Janeiro
- Prof. Dr. José Luis Soto Gonzales – UNMSM (Peru)
- Prof. Dr. Julio Cezar Uzinski – UFMT
- Prof. Msc. Lucas R. Oliveira – Mun. de Chap. do Sul
- Prof. Dr. Leandro Argente-Martínez – ITSON (México)
- Profa. Msc. Lidiene Jaqueline de Souza Costa Marchesan – Consultório em Santa Maria
- Prof. Msc. Marcos Pisarski Júnior – UEG
- Prof. Dr. Mario Rodrigo Esparza Mantilla – UNAM (Peru)
- Profa. Msc. Mary Jose Almeida Pereira – SEDUC/PA
- Profa. Msc. Nila Luciana Vilhena Madureira – IFPA
- Profa. Dra. Patrícia Maurer
- Profa. Msc. Queila Pahim da Silva – IFB
- Prof. Dr. Rafael Chapman Auty – UO (Cuba)
- Prof. Dr. Rafael Felipe Ratke – UFMS
- Prof. Dr. Raphael Reis da Silva – UFPI

- Prof. Dr. Ricardo Alves de Araújo – UEMA
- Prof. Dr. Wéverson Lima Fonseca – UFPI
- Prof. Msc. Wesclen Vilar Nogueira – FURG
- Profa. Dra. Yilan Fung Boix – UO (Cuba)
- Prof. Dr. Willian Douglas Guilherme – UFT

#### Conselho Técnico Científico

- Esp. Joacir Mário Zuffo Júnior
- Esp. Maurício Amormino Júnior
- Esp. Tayronne de Almeida Rodrigues
- Esp. Camila Alves Pereira
- Lda. Rosalina Eufrausino Lustosa Zuffo

#### Ficha Catalográfica

<b>Dados Internacionais de Catalogação na Publicação (CIP)</b> (eDOC BRASIL, Belo Horizonte/MG)	
A642	<p>Aplicações de inteligência artificial e ciência de dados [recurso eletrônico] / Organizadores Julio Cezar Uzinski, Caio Cesar Enside de Abreu, Bruno Rodrigues de Oliveira. – Nova Xavantina, MT: Pantanal, 2020. 100p.</p> <p>Formato: PDF            Requisitos de sistema: Adobe Acrobat Reader            Modo de acesso: World Wide Web            ISBN 978-65-88319-26-0            DOI <a href="https://doi.org/10.46420/9786588319260">https://doi.org/10.46420/9786588319260</a></p> <p>1. Inteligência artificial. 2. Automação. 3. Processamento de dados. I. Uzinski, Julio Cezar. II. Abreu, Caio Cesar Enside de. III. Oliveira, Bruno Rodrigues de.</p> <p style="text-align: right;">CDD 001.535</p>
<b>Elaborado por Maurício Amormino Júnior – CRB6/2422</b>	

O conteúdo dos livros e capítulos, seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva do(s) autor (es). O download da obra é permitido e o compartilhamento desde que sejam citadas as referências dos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

#### **Pantanal Editora**

Rua Abaete, 83, Sala B, Centro. CEP: 78690-000.  
 Nova Xavantina – Mato Grosso – Brasil.  
 Telefone (66) 99682-4165 (Whatsapp).  
<https://www.editorapantanal.com.br>  
[contato@editorapantanal.com.br](mailto:contato@editorapantanal.com.br)

## APRESENTAÇÃO

Ciência de dados (*Data science*) e Inteligência artificial (*Artificial intelligence*) podem ser consideradas como algumas das áreas da ciência mais importantes de nosso tempo. Ademais, não estão apenas no mundo acadêmico ou no futuro como no filme *The terminator*, estão ao nosso redor nesse exato momento em situações simples, por exemplo, são empregadas para sugerir qual o próximo emoji que você enviará em uma conversa no seu aplicativo de bate papo favorito. Ciência de dados se refere a um campo de estudo muito amplo que se vale de métodos científicos para se obter informações e percepções de conjuntos de dados, enquanto que a inteligência artificial refere-se ao que chamamos de inteligência das máquinas. As aplicações dessas ferramentas estão relacionadas às tecnologias que nos envolvem em nosso cotidiano, e.g., as redes sociais ou comércio eletrônico, bem como aplicações que escapam ao senso comum, como em cibersegurança ou astrofísica.

Tais abordagens, às vezes, são utilizadas juntas, de forma intercalada ou separadas. Uma pesquisa poderá contar com ambas as ferramentas ou apenas com uma delas. Nesse livro, o leitor encontrará aplicações diversas dessas ferramentas em diferentes perspectivas. Em seis capítulos são apresentados resultados de pesquisas de dezesseis pesquisadores sobre temas diversos que dão uma demonstração do poder da Inteligência Artificial e da Ciência de Dados.

O primeiro capítulo traz uma comparação dos resultados da Transformada Wavelet (WT) e do Filtro Savitzky-Golay (SG) na estimação da tendência em séries temporais. Os resultados são obtidos a partir de uma série temporal artificial e uma série de dados reais, i.e, a cotação da moeda americana (Dólar US\$) frente à moeda brasileira (Real R\$).

Considerando que a principal abordagem em Aprendizado de Máquinas é a construção de algoritmos que, por meio das experiências e aprendizado, possam ser melhorados automaticamente, o segundo capítulo desse livro traz uma avaliação de um algoritmo usando Redes Neurais Artificiais LSTM. Essa abordagem é justificada pela necessidade de reconhecimento de padrões de comportamento de presença de elementos em um ambiente para realizar a predição de ações no local por meio dos dados coletados fornecendo uma base para o desenvolvimento de um sistema doméstico inteligente.

Já no terceiro capítulo, uma metodologia para descrição de Casos de Uso (Uc) no desenvolvimento de sistemas de informação é apresentada. A motivação para a criação desta metodologia é que muitos UC's são descritos com pouca riqueza de detalhes, o que prejudica a qualidade dos sistemas de informação.

O quarto capítulo traz uma aplicação bastante diversa de todas as outras: um estudo dos aspectos e causas da obesidade por meio do algoritmo de agrupamento Farthest First para agrupar

pessoas com informações demográficas, socioeconômicas, de biotipo e de conjunto de ações observáveis recorrentes.

No quinto capítulo é desenvolvida uma aplicação utilizando regressão logística como um mapeamento do espaço de atributos não-linearmente separável baseado em um tipo especial de funções denominadas de Golden Wavelets. Nesse texto, os autores buscam tratar a matemática envolvida no método de aprendizado de máquina denominado de Regressão Logística e propor a utilização de funções wavelets para mapeamento do espaço de atributos.

Para encerrar esse e-book com chave de ouro, o sexto capítulo, apresenta uma classificação de cenas acústicas utilizando a Transformada Wavelet e Aprendizagem de Máquina. Os autores dedicam a redação à apresentação dos principais métodos de Classificação de Cenas Acústicas (CCA) e os principais atributos utilizados para esta classificação, exemplos de aplicações, e uma investigação adicional centrada na utilização da Transformada Wavelet Contínua (TWC).

Caríssimo(a) leitor(a), convidamos você à leitura de Aplicações de Inteligência Artificial e Ciência de Dados, e esperamos que a mesma seja um deleite ilimitado e que contribua de alguma forma com suas aspirações.

**Julio Cezar Uzinski**

## SUMÁRIO

<b>Apresentação</b> .....	4
<b>Capítulo I</b> .....	7
Obtenção de Tendências em Série Temporal empregando a Transformada Wavelet e o filtro Savitzky-Golay.....	7
<b>Capítulo II</b> .....	24
Aplicação de Redes Neurais Artificiais LSTM em domótica inteligente .....	24
<b>Capítulo III</b> .....	32
Uma Metodologia Simples para Descrição de Casos de Uso no Desenvolvimento de Sistemas de Informação.....	32
<b>Capítulo IV</b> .....	49
Aspectos e causas da obesidade: uma análise utilizando algoritmo de agrupamento <i>Farthest First</i> .....	49
<b>Capítulo V</b> .....	59
Mapeamento de espaço de atributos não-linearmente separável baseado na Golden wavelet: aplicação na regressão logística .....	59
<b>Capítulo VI</b> .....	85
Classificação de Cenas Acústicas Utilizando a Transformada Wavelet Contínua e Aprendizagem de Máquina .....	85
<b>Índice Remissivo</b> .....	99

# Classificação de Cenas Acústicas Utilizando a Transformada Wavelet Contínua e Aprendizagem de Máquina

Recebido em: 18/09/2020

Aceito em: 01/10/2020

 10.46420/9786588319260cap6

Caio Cesar Enside de Abreu<sup>1\*</sup> 

Edilaine Gomes da Silva<sup>2</sup> 

Lucas Angelo Mattesco<sup>3</sup> 

Bruno Rodrigues de Oliveira<sup>4</sup> 

Carlos Alex Sander J. Gulo<sup>5</sup> 

Lucas Kriesel Sperotto<sup>6</sup>

## INTRODUÇÃO

De acordo com Halliday et al. (2002), o som é a propagação de uma frente de compressão mecânica ou onda mecânica; o som é uma onda longitudinal, que se propaga de forma circuncêntrica, apenas em meios materiais, como os meios sólidos, líquidos ou gasosos, não podendo propagar-se no vácuo. Os sons audíveis pelo ouvido humano têm uma frequência entre 20Hz e 20.000Hz, portanto, os sons que estão abaixo ou acima desta faixa de frequência são denominados, respectivamente, de infrassom e ultrassom.

O som que percebemos ao caminhar por um parque, uma rua movimentada ou ao andar de ônibus, por exemplo, contém várias informações sobre o ambiente e os eventos que estão ocorrendo naquele ambiente. Os seres humanos utilizam estas informações para compreenderem o ambiente onde estão, como por exemplo, uma pessoa que está se preparando para sair de casa, ouvindo o que está

---

<sup>1,2,3</sup> Universidade do Estado de Mato Grosso – UNEMAT, Rua Santa Rita, 128, Alto Araguaia-MT. CEP 78780-000.

E-mail: <sup>1</sup>caioenside@unemat.br, <sup>2</sup>edilaine.gomes@unemat.br, <sup>3</sup>lucas.mattesco@unemat.br

<sup>4</sup> Pantanal Editora. E-mail: bruno@editorapantanal.com.br

<sup>5,6</sup> Grupo de Pesquisa PIXEL - Ciência da Computação – UNEMAT. Rua Santa Rita, 128, Centro - Alto Araguaia-MT.

CEP 78.780-000. E-mail: <sup>5</sup>sander@unemat.br, <sup>6</sup>sperotto@unemat.br

\* Autor correspondente: caioenside@unemat.br

acontecendo do lado de fora pode colocar um casaco (se ouvir sons de ventos), utilizar um guarda-chuva (sons da chuva) ou também pode deduzir como está a situação do trânsito em sua rua. A combinação de vários tipos de sons em um determinado ambiente caracteriza uma cena acústica (Barchiesi et al., 2015; Stowell et al., 2015).

Segundo Chachada e Kuo (2013), a análise de áudio realizada de forma automática tem sido empregada em diversas áreas, tais como: robótica, inteligência artificial, dispositivos móveis, tecnologia assistiva, dentre outras. O estudo e o aperfeiçoamento de técnicas de análise de áudio podem fornecer, juntamente com sua posterior utilização em dispositivos tecnológicos, o desenvolvimento de dispositivos inteligentes e a criação de sistemas de tomada de decisão cada vez mais precisos. Dessa forma, a análise não somente dos sons individuais, mas da cena acústica que eles formam, pode trazer melhorias antes impensáveis para o ser humano.

Segundo Virtanen et al. (2018), a habilidade da audição é vista como algo importante para o ser humano pois, além de fazer parte do sistema sensorial, é por meio dela que se tornou possível o desenvolvimento da linguagem falada e escrita que conhecemos, sem as quais não seria possível ter evoluído tanto em termos tecnológicos. O desafio enfrentado na área de Classificação de Cenas Acústicas (CCA) é o de construir algoritmos e métodos computacionais, capazes de reconhecer os diversos sons pertencentes à um determinado ambiente ou localização.

O desenvolvimento de métodos de processamento de sinais capazes de extrair automaticamente informações do som, apresentam grande potencial em diversas aplicações do dia a dia, dentre aquelas mencionadas anteriormente. Porém, ainda é um grande desafio encontrar formas eficazes de realizar o reconhecimento de cenas acústicas, seja pela busca de novos descritores capazes de extrair informações relevantes do sinal de áudio, ou pela melhoria constante nos algoritmos de aprendizagem de máquina. Isto se deve ao fato de que em qualquer ambiente real, o arquivo de áudio capturado contém sons gerados por fontes e intensidades aleatórias, e que normalmente são sobrepostos de formas distintas.

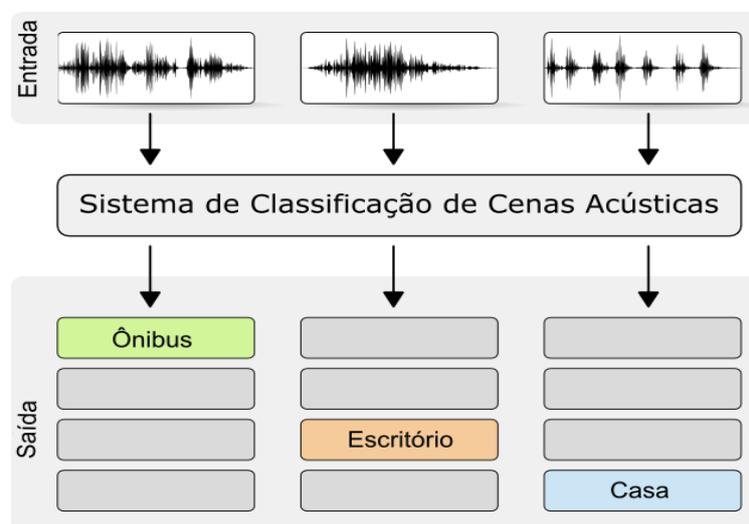
Neste capítulo, serão apresentados os principais métodos de CCA e os principais atributos utilizados para esta classificação, assim como exemplos de aplicações. Uma investigação adicional centrada na utilização da Transformada Wavelet Contínua (TWC) será apresentada. Para isso, simulações serão conduzidas utilizando um algoritmo padrão de *Machine Learning*, o tradicional Máquina de Vetores de Suporte (SVM), em conjunto com um critério de decisão tradicional, baseado no voto majoritário. Uma base de dados será fixada e apenas os descritores de entrada para o classificador serão alternados. O objetivo é fazer uma comparação justa dos descritores avaliados. Por fim, pretende-se discutir se a TWC compete, em relação a performance, com os descritores tradicionalmente empregado nesse problema, denominados de Coeficientes Cepstrais de Frequência Mel (MFCCs) e Mel-

Espectrograma, que são amplamente utilizados para análise de áudio e voz (Barchiesi et al., 2015; Mesaros et al., 2018).

## CLASSIFICAÇÃO AUTOMÁTICA DE CENAS ACÚSTICAS

A cena acústica, de acordo com Lignelli (2011), é o som que é formado quando vários sons, de diversas fontes, são combinados, formando o áudio que é ouvido num ambiente como um todo. Por exemplo, a cena acústica de uma grande avenida pode ser formada pelo som emitido por pessoas conversando, vendedores gritando, carros passando e/ou buzinando, sirenes de ambulâncias, entre outros. Assim, o ser humano consegue compreender qual é o ambiente, somente por ouvir a cena acústica.

Em Nguyen e Pernkopf (2018), a CCA é definida como o reconhecimento de um determinado ambiente, tomando como base uma gravação de áudio, e assumindo que a cena acústica é uma representação efetiva de um local ou situação. O objetivo da classificação automática de cenas acústicas é identificar o ambiente onde em que o um arquivo de áudio foi gravado, utilizando como base classes predefinidas de áudio que caracterizam o cenário original, por exemplo, “parque”, “rua de pedestres”, “estação de metrô”; ou para indicar que a cena é de um ambiente diferente ou desconhecido (Mesaros et al., 2018). A Figura 1 apresenta a ideia geral do problema.



**Figura 1.** Esquema da classificação de cenas acústicas. Fonte: Adaptado de <http://dcase.community/challenge2019/task-acoustic-scene-classification>.

## ESTADO DA ARTE

De acordo com Barchiesi et al. (2015), o primeiro método computacional que aborda especificamente o tema CCA foi publicado em 1997, em Sawhney e Maes (1997). Apesar de já haver um tempo considerável desde então, a área de pesquisa só ganhou força e notoriedade a partir de 2013, com o surgimento de desafios internacionais consecutivos. Denominado *IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events* (DCASE), este desafio internacional, que é apoiado pela *IEEE Signal Processing Society*, proporcionou um avanço significativo em um curto espaço de tempo, e tem conseguido agregar pesquisadores ao redor do mundo, entusiastas e representantes da indústria. A primeira edição do desafio ocorreu em 2013. Os organizadores desenvolveram uma base de dados sobre a qual os participantes desenvolveram e testaram seus algoritmos. Os participantes foram encorajados a submeter um artigo acompanhando o código fonte dos sistemas propostos, para que todos os detalhes das propostas pudessem ficar disponíveis para a comunidade. Após o recebimento das propostas, os organizadores testaram todos os sistemas sobre uma base de dados oculta.

Os resultados, juntamente com todo o material, estão disponíveis no *website* do evento. Além disso, os resultados, incluindo o estado da arte da área até aquela data, foram publicados em Barchiesi et al. (2015) e Stowell et al. (2015). A segunda edição do DCASE ocorreu em 2016 e, a partir daquele ano, o desafio passou a ser acompanhado por um *workshop*. Todas as edições do DCASE podem ser encontradas no endereço <http://dcase.community/events>.

A fim de caracterizar todos os sistemas submetidos às últimas edições do DCASE, a Tabela 1 apresenta uma visão geral dos atributos e dos algoritmos de aprendizagem de máquina utilizados.

**Tabela 1.** Visão geral dos atributos e algoritmos de aprendizagem de máquina utilizados em todos os desafios DCASE realizados entre 2013 e 2019, contabilizando 324 sistemas verificados.

Atributos	Edição DCASE				
	2013	2016	2017	2018	2019
Espalhamento espectral profundo	-	-	-	4	3
Filtro Gammatone + log-mel-espectrograma + transformada Q-Constante (CQT)	-	-	-	-	2
Log-mel-espectrograma + outros	-	-	25	32	72
Log-mel-espectrograma	-	-	9	5	-
MFCCs + outros	7	17	13	5	5
Espectro de potência ponderado perceptualmente	-	-	-	3	4
Forma de onda bruta	-	-	-	3	2
Forma de onda bruta + log-mel-espectrograma	-	-	-	-	7
Forma de onda bruta + espectrograma + i-vetor	-	-	-	4	-
Espectrograma + outros	-	5	11	-	1
Vários	8	27	38	4	2
Wavelet + outros	3	-	2	-	1
Algoritmos de aprendizagem de máquina					
Rede Neural Convolutacional	-	10	49	32	76
Rede Neural Profunda	-	7	4	-	1

<i>Ensemble</i>	-	4	2	4	3
SVM e Modelo de Mistura de Gaussianas	1	3	4	-	5
Rede Neural de Regressão Generalizada	-	1	2	-	-
I-vetor + outros	1	2	2	1	-
Perceptron multicamadas (MLP) / MLP + outros	-	-	11	1	-
Fatoração de Matrizes Não-Negativas (NMF) /NMF + outros	-	1	2	-	-
Máquina de Vetores de Suporte	11	10	10	1	1

Nota-se, a partir da Tabela 1, que as pesquisas recentes e que constituem o estado-da-arte têm focado na utilização de atributos como mel-espectrograma e modelos de *machine learning* com complexidade elevada, tais como Redes Neurais Profundas e Redes Neurais Convolucionais. Em relação à Transformada Wavelet, apenas 6 dos 324 sistemas submetidos em todos os eventos utilizaram esse método alguma versão dessa ferramenta. Em nenhum destes sistemas foi utilizada a Transformada Wavelet Contínua (TWC) no processo de extração de atributos. Este fato reforça a necessidade de investigação a respeito de tais atributos.

Analisando a Tabela 1 de uma forma mais minuciosa, pode-se observar também dois momentos importantes da CCA: 1º) utilização predominante de algoritmos clássicos de aprendizagem de máquina, principalmente o classificador SVM juntamente com os atributos MFCCs, até a edição de 2016; 2º) popularização de abordagens de Aprendizagem Profunda, na língua inglesa *Deep Learning*, caracterizada na Tabela 1 pela utilização de Redes Neurais Convolucionais tendo como entrada o mel-espectrograma.

A fim de obter o mel-espectrograma de um arquivo de áudio, primeiro obtém-se o espectrograma do sinal, que consiste em uma matriz cujas colunas armazenam o espectro de amplitude ou de potência do conteúdo de cada janela, e, posteriormente, converte-se o eixo de frequência para a escala mel (Sakashita e Aono, 2018). É comum, também, converter o mel-espectrograma para uma escala logarítmica, denotado por log-mel-espectrograma. A escala Mel está relacionada à capacidade do ouvido humano em distinguir sons com frequências parecidas, uma vez que o ouvido humano não percebe as frequências em uma escala linear, como fornecida pelo espectrograma. Já os MFCCs são obtidos a partir do cálculo da Transformada de Cosseno Discreta sobre o mel-espectrograma (Barchiesi et al. 2015).

MFCCs constituem uma representação mais compressiva do sinal de áudio, quando comparado ao mel-espectrograma e são frequentemente utilizados com modelos lineares, como Modelos de Misturas Gaussianas e SVM. Para utilização com algoritmos de *Deep Learning*, que em geral utilizam modelos mais complexos e necessitam de uma quantidade maior de dados para treinamento, o mel-espectrograma fornece melhores resultados (Sakashita e Aono, 2018).

## Aplicações

Nesta seção, alguns exemplos de aplicações que a CCA pode possibilitar são apresentados. De forma geral, um sistema de CCA possibilita a aplicação de computação sensível ao contexto, dispositivos portáteis inteligentes, sistemas de navegação robótica, gerenciamento de arquivos de áudio, sistemas de segurança, entre outras (Barchiesi et al., 2015).

De maneira mais específica, pode-se destacar a possibilidade de desenvolvimento de aplicativos que permitam que *smartphones* mudem seu modo de operação (mudando para o modo silencioso, por exemplo) dependendo da localização do usuário; cadeiras de rodas inteligentes que ajustam seu funcionamento baseado no reconhecimento de ambientes abertos ou fechados; ajuste automático de implantes cocleares (IC) e aparelhos auditivos baseado nas características acústicas do ambiente em que o usuário se encontra (Hansen et al., 2019); em robótica, a integração da CCA com a visão computacional pode fornecer informações adicionais, principalmente onde informações visuais não são suficientes (Chu et al., 2006); em aplicações de segurança, a classificação de eventos sonoros suspeitos, como por exemplo, detecção de gritos e batidas pode integrar o sistema (Radhakrishnan et al., 2005).

## MATERIAL E MÉTODOS

### Transformada Wavelet Contínua

A Transformada Wavelet Contínua (TWC) é uma transformação de um espaço real unidimensional para um espaço real bidimensional. Dado um sinal  $f(t) \in L^2(\mathbb{R})$ , onde  $L^2(\mathbb{R})$  é o espaço real das funções de quadrado integrável, a TWC é o mapeamento  $Wf(s, u): L^2(\mathbb{R}) \mapsto L^2(\mathbb{R}^+ \times \mathbb{R})$  (Daubechies, 1992; Mallat, 2009).

$$Wf(s, u) = \int_{-\infty}^{\infty} f(t) \psi_{s,u}^*(t) dt, \quad (1)$$

onde  $s \in \mathbb{R}^+$  e  $u \in \mathbb{R}$  são os parâmetros de escala e translação, respectivamente, e  $*$  denota o conjugado complexo da família de funções *wavelets*  $\psi_{s,u}$ , dada por

$$\psi_{s,u}(t) = s^{-0.5} \psi\left(\frac{t-u}{s}\right). \quad (2)$$

A norma Euclidiana de tais funções devem ser unitária, ou seja,  $\|\psi_{s,u}\| = 1$ . Além disso, uma função *wavelet* deve satisfazer a condição de admissibilidade  $C_\psi = 2\pi \int_{-\infty}^{\infty} |\hat{\psi}(\omega)|^2 / \omega d\omega < \infty$ , onde  $\hat{\psi}(\omega)$  é a Transformada de Fourier de  $\psi(t)$ .

Algumas funções *wavelets* têm suporte compacto, como a *wavelet* de Haar e outras. Apesar de algumas nunca convergirem para zero, têm um suporte que pode ser considerado compacto, pois o decaimento é rápido, por exemplo, a *wavelet* Ricker. A condição de admissibilidade acima mencionada, implica que as funções *wavelets* têm média nula, logo, as áreas abaixo e acima do eixo temporal se anulam,

$\int_{-\infty}^{\infty} \psi(t)dt = 0$ , e isto caracteriza sua forma em “ondinhas”, que não oscila indefinidamente como uma senóide, mas apenas no seu intervalo suporte.

Dado seu caráter ondulatório e seu suporte compacto, as funções *wavelets* podem ser vistas como janelas de análise. Neste sentido, a TWC consiste na obtenção da similaridade entre o sinal  $f(t)$  e a função analisadora *wavelet*  $\psi(t)$ , que é transladada sobre o sinal. Considerando uma *wavelet* com suporte igual a  $M$ , ao transladar essa função pelo sinal, alterando o parâmetro  $u$ , e realizar a multiplicação da função transladada por segmento de comprimento  $M$  do sinal, aqueles segmentos que possuem maior similaridade com a forma de onda da *wavelet* resultarão em coeficientes *wavelets* com maior amplitude. Ao alterar a escala  $s$ , a função wavelet é comprimida quando  $0 < s < 1$  e dilatada quando  $s > 1$ , implicando que seu intervalo suporte também é modificado e, conseqüentemente, a sua forma de onda, acarretando na detecção de outras similaridades entre o sinal  $f(t)$  e a função *wavelet* escalonada.

A Transformada de Fourier das funções *wavelets* mostra que as *wavelets* têm características de filtros passa-banda, cuja banda de passagem depende da escala escolhida. Assim, se no domínio de Fourier uma função *wavelet* é nula fora intervalo  $]\alpha, \beta[$ , ou seja, somente as frequências dentro deste intervalo passam pelo filtro *wavelet*, então, ao alterar o parâmetro de escala  $s$ , este intervalo é modificado para  $]\frac{\alpha}{s}, \frac{\beta}{s}[$ . Isso implica que, quanto maior a escala, menor será a resolução em frequência.

Uma das ferramentas mais empregadas da TWC é o Escalograma, que é obtido tomando o quadrado do valor absoluto dos coeficientes *wavelets*, isto é  $|Wf(s, u)|^2$ . Ele representa o espectro de energia dos coeficientes *wavelets* em uma certa escala  $s$  em função do tempo  $u$ , utilizando intensidade de cores para designar a energia de cada coeficiente. A fim de entender a distribuição de energia por escala, recorre-se ao Espectro Wavelet Global (*Global Wavelet Spectrum - GWS*) que é calculado como a média temporal dos coeficientes para uma escala selecionada, de acordo com a equação (4), onde consideramos uma versão discreta da TWC, dada pela equação (3). Para um sinal  $g[n]$  de comprimento  $N$ , obtido a partir da amostragem de  $f(t)$ , sendo  $a$  a escala e  $b$  o parâmetro de translação discretizados, temos (Alessio, 2016):

$$Wg(a, b) \approx \sum_{n=0}^{N-1} g[n] \psi^* \left[ \frac{n-b}{a} \right], \quad (3)$$

$$GWS(a) = \frac{1}{N} \sum_{n=0}^{N-1} |Wg(a, b)|^2. \quad (4)$$

Na equação (4), o GWS pode ser calculado utilizando a potência, da forma como está definido, ou utilizando a amplitude. Em outras palavras, pode-se utilizar apenas o valor absoluto dos coeficientes *wavelets*. Neste trabalho, após a realização de algumas simulações, optou-se por utilizar apenas a magnitude.

## Implementação do sistema

O sistema principal implementado para avaliar os atributos é baseado no voto majoritário, que pode ser descrito da seguinte forma: para cada sinal de áudio a ser classificado, realiza-se um janelamento de Hanning com 1024 pontos; para cada janela, os descritores são extraídos, normalizados pela média e variância, e então utilizados para alimentar um SVM com *kernel* linear; para cada janela, uma classificação é emitida. A classificação final do arquivo é emitida observando a classe que recebeu a maior quantidade de votos pelo classificador. Vale destacar que o sistema implementado não foi desenvolvido para competir com os sistemas vencedores do evento, pelo contrário, o intuito é utilizar uma abordagem padrão de aprendizagem de máquina, que favoreça e simplifique a comparação dos atributos avaliados.

A base de dados escolhida é a fornecida para o DCASE-2013. Esta base de dados contém 100 arquivos de áudio divididos em dez classes, no formato .wav, com taxa de amostragem de 44.100 Hz, 16 bits, estéreo. Bases de dados fornecidas para edições mais recentes do DCASE contém milhares de arquivos de áudios, sendo desenvolvidas principalmente para avaliação de algoritmos de *Deep Learning*. Utilizar tais bases de dados fogem ao escopo do presente trabalho.

Com relação aos protocolos da simulação, a validação cruzada com 5 pastas é utilizada a fim de que todos os sinais da base de dados sejam classificados. Para cada pasta, oito arquivos de áudio por classe são utilizados para treino e dois arquivos para teste. Para implementação da Transformada Wavelet Contínua, a biblioteca para Python denominada PyWavelets (Gregory et al., 2019) foi utilizada. A biblioteca librosa (Mcfee et al., 2015) foi utilizada para extração dos atributos MFCCs e do mel-espectrograma.

A Tabela 2 apresenta algumas variações do sistema principal, destacando suas principais configurações, escolhidas para fornecer as análises pretendidas. Resultados das simulações com diferentes parâmetros dos algoritmos são apresentados na próxima seção.

**Tabela 2.** Características dos sistemas implementados. Fonte: os autores.

Sistema	Atributos extraídos	Família Wavelet	Pós-processamento dos atributos
1	MFCCs	-	Filtragem cepstral
2	GWS	Gauss de ordem 8	Filtragem recursiva
3	Escalograma + Hog (9 orientações)	Gauss de ordem 8	-
4	Mel-espectrograma	-	-
5	Log-mel-espectrograma	-	-

A Filtragem cepstral, mencionada na Tabela 2, é realizada da seguinte forma:

$$S[n, m] = S[n, m] \left( 1 + \text{sen} \left( \frac{\pi(n+1)}{\beta} \right) \right) \frac{\beta}{2}, \tag{4}$$

onde  $S$  é a matriz que armazena a sequência de MFCCs em suas colunas,  $n = 1, 2, \dots, N_{MFCCs}$  e  $m = 1, 2, \dots, M$ , com  $N_{MFCCs}$  e  $M$  sendo, respectivamente, o número de MFCCs extraídos e a quantidade de janelas analisadas. Utilizou-se  $\beta = 26$ . Já a filtragem recursiva utilizada para suavizar a sequência de GWS, extraída de cada janela, é realizada como segue:

$$GWS_m = \alpha GWS_{m-1} + (1 - \alpha) GWS_m, \tag{5}$$

$\alpha$  é o parâmetro que controla o quão suave é a transição do vetor de atributos entre janelas consecutivas. Definiu-se empiricamente  $\alpha = 0,6$ .

Os Sistemas 3 e 4 utilizam o descritor HOG, sigla padrão para denotar Histograma de Gradientes Orientados, calculado sobre o escalograma gerado pela TWC. O descritor HOG é muito utilizado para localização de objetos em uma imagem digital, pela sua capacidade em detectar bordas e contornos. Quando aplicado sobre o escalograma, extrai informação relacionada a distribuição da energia do sinal pelas escalas (Mattesco e Abreu, 2019).

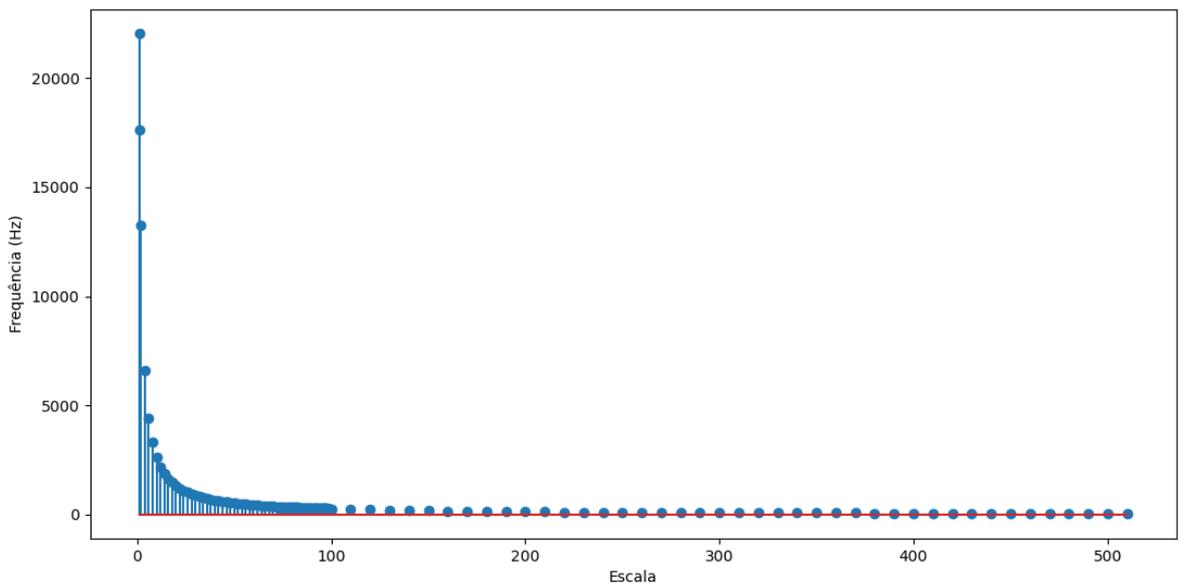
Na Tabela 3 constam mais detalhes sobre a análise *wavelet* realizada pela CWT sobre a base de dados DCASE-2013, sendo apresentadas a sequência de escalas utilizadas para a decomposição e as respectivas frequências alcançadas. As frequências correspondentes para cada escala são estimadas levando em consideração a função *wavelet* escolhida e a frequência de amostragem do sinal, de forma teórica, podendo ser utilizada para orientar a escolha das escalas (Oliveira et al., 2018).

**Tabela 3.** Escalas utilizadas para o cálculo da CWT e respectivas frequências correspondentes para a função *wavelet* Gaussiana de ordem 8 e frequência de amostragem do sinal 44.100 Hz. Fonte: os autores.

Escalas	1,2 1,5 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 82,5 83, 83,5 84, 84,5 85, 85,5 86, 86,5 87, 87,5 88, 88,5 89, 89,5 90, 90,5 91, 91,5 92, 92,5 93, 93,5 94, 94,5 95, 95,5 96, 96,5 97, 97,5 98, 98,5 99, 99,5 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510,
Frequências (Hz)	22050, 17640, 13230, 6615, 4410, 3307,5 2646, 2205, 1890, 1653,75 1470, 1323, 1202,73 1102,5 1017,69 945, 882, 826,88 778,24 735, 696,32 661,5 630, 601,36 575,22 551,25 529,2 508,85 490, 472,5 456,21 441, 426,77 413,44 400,91 389,12 378, 367,5 362,47 357,57 352,8 348,16 343,64 339,23 334,94 330,75 326,67 322,68 320,73 318,8

	316,89	315,	313,14	311,29	309,47	307,67	305,9	304,14	302,4	300,68
	298,98	297,3	295,64	294,	292,38	290,77	289,18	287,61	286,05	284,52
	282,99	281,49	280,	278,53	277,07	275,62	274,2	272,78	271,38	270,
	268,63	267,27	265,93	264,6	240,55	220,5	203,54	189,	176,4	165,38
	155,65	147,	139,26	132,3	126,	120,27	115,04	110,25	105,84	101,77
	98,	94,5	91,24	88,2	85,35	82,69	80,18	77,82	75,6	73,5
	69,63	67,85	66,15	64,54	63,	61,53	60,14	58,8	57,52	56,3
	54,	52,92	51,88							

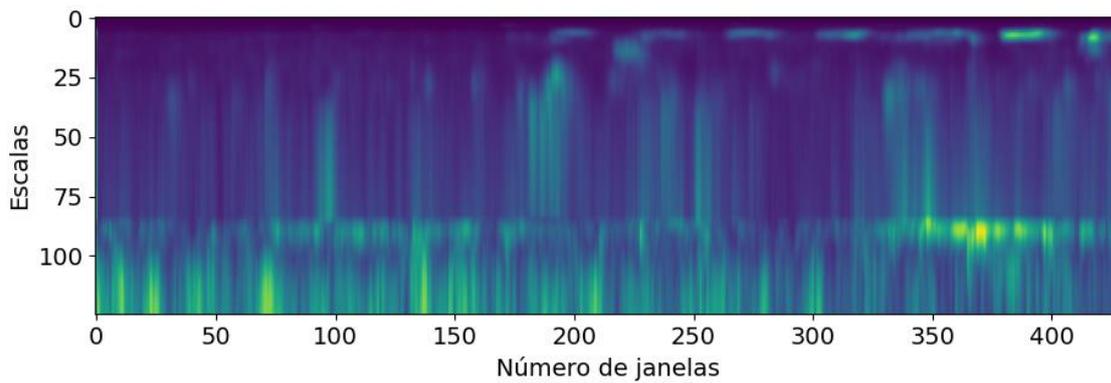
Complementarmente à Tabela 3, a Figura 2 apresenta um gráfico com a sequência de escalas escolhidas no eixo das abscissas e as respectivas frequências no eixo das ordenadas.



**Figura 2.** Escalas utilizadas para o cálculo da CWT e respectivas frequências alcançadas. Fonte: os autores.

Note na Tabela 2 que a escala 1,2 foi escolhida para que se contemple a maior frequência do sinal, que neste caso é de 22050 Hz, conforme o Teorema de Nyquist. As demais escalas foram escolhidas de forma que se tenha uma distribuição suave das frequências analisadas e uma maior resolução em baixas frequências, conforme pode ser visto na Figura 2.

A Figura 3 apresenta a sequência de GWS obtida durante a análise de 5 segundos de um arquivo de áudio da classe ônibus. O GWS obtido em cada janela foi armazenado em uma coluna da matriz de saída. Para fins de visualização, a matriz obtida foi plotada como imagem.



**Figura 3.** Sequência de GWS obtidos após análise de uma porção de cinco segundos de um arquivo de áudio da classe ônibus. Fonte: os autores.

## RESULTADOS E DISCUSSÃO

Como mencionado anteriormente, o principal objetivo deste trabalho é apresentar uma comparação justa entre os descritores fornecidos pela TWC, os MFCCs e o mel-espectrograma. Estes dois últimos são considerados descritores padrão para análise de áudio. Dessa forma, a análise apresentada se limita a dimensão dos descritores e não da abordagem de aprendizagem em si. Por este motivo, uma abordagem tradicional que utiliza um classificador padrão foi implementada.

É importante destacar também que os resultados apresentados nesta seção são naturalmente inferiores aos resultados encontrados pelos competidores dos eventos DCASE, visto que, de forma geral, os competidores buscam a construção de uma abordagem de aprendizagem complexa e inovadora, normalmente utilizando arquiteturas de redes neurais avançadas em conjunto com os descritores padrão apresentados anteriormente.

**Tabela 3.** Resultados das simulações sobre a base de dados DCASE-2013.

Sistema	Dimensão do vetor de atributos	Porção do sinal utilizado (s)	Acurácia (%)
1	19	0,5	37
1	19	5	49
2	125	0,5	45
2	125	5	47
3	4032	0,5	41
3	4032	5	43
4	32	0,5	39
4	32	5	41
4	128	0,5	37
4	128	5	37
5	32	0,5	36
5	32	5	47

Na Tabela 3, os resultados apresentados na coluna Acurácia representam a média das acurácias obtidas em cada uma das cinco pastas da validação cruzada. Os três melhores resultados estão destacados em cinza escuro e cinza claro, respectivamente. De maneira geral, nota-se que o Sistema 1, que utiliza os MFCCs como atributos, alcançou o melhor resultado, sendo ligeiramente melhor que o Sistema 2 e 5, que são baseados no atributo *wavelet* GWS e log-mel-espectrograma, respectivamente. Em seguida, aparece o Sistema 3, que utiliza o atributo *wavelet* baseado no descritor HOG. O sistema que apresentou o pior resultado foi o Sistema 4, que é baseado no mel-escalograma.

Com relação ao Sistema 1, utilizou-se 19 MFCCs, já considerando que o coeficiente de ordem zero foi descartado. Este número foi escolhido empiricamente e forneceu os melhores resultados. Nota-se que os resultados são melhores quando se utiliza uma porção maior do sinal, alcançando uma acurácia média de 49%.

De forma similar, o Sistema 2 alcançou seu melhor resultado quando utilizado 5 segundos do sinal, alcançando 47% de acurácia. Este sistema utiliza o GWS calculado utilizando a função *wavelet* de Gauss de ordem 8. Várias simulações foram realizadas alternando a função *wavelet*, porém a Gaussiana de ordem 8 se mostrou mais adequada para analisar os sinais de áudio da base DCASE-2013. Uma explicação sugerida é o fato de sua forma de onda apresentar mais similaridades com o sinal analisado do que as demais funções testadas.

O descritor HOG utilizado no Sistema 3 fornece um vetor de atributos com alta dimensionalidade, quando comparado aos demais descritores. Apesar de ser superado pelos MFCCs, log-mel-espectrograma e pelo GWS, gerou resultados ligeiramente superiores ao mel-espectrograma, alcançando 43% de acurácia.

Por fim, analisando os resultados dos Sistemas 4 e 5, observa-se que utilizar um número elevado de bandas mel não é adequado. A melhor configuração é encontrada utilizando apenas 32 bandas, alcançando 41% e 47% de acurácia, respectivamente. Observando os resultados utilizando o mel-espectrograma e o log-mel-espectrograma, nota-se que para 0,5 segundos de áudio analisado, o primeiro forneceu um melhor desempenho. No entanto, utilizando um período maior do sinal, a versão em escala logarítmica apresentou um desempenho superior. Comparando ambas as versões com os MFCCs, nota-se um melhor desempenho para o último. Este resultado vem de encontro com o discutido na seção anterior: tanto o mel-espectrograma, quanto o log-mel-espectrograma, geram melhores resultados quando utilizado em metodologias de *Deep Learning*.

Do ponto de vista de aplicações de tempo real, é desejável que se utilize um período curto para emitir uma classificação do ambiente acústico. Nesse sentido, o atributo que gerou o melhor desempenho quando utilizado uma porção de 0,5 segundos do arquivo de áudio foi o GWS.

Em termos gerais, pode-se concluir que a TWC é uma ferramenta que pode fornecer resultados semelhantes, ou até mesmo superiores aos descritores popularmente utilizados para a CCA. Este fato indica que trabalhos futuros devem se concentrar na proposição de uma abordagem de aprendizagem mais complexa e profunda, que possa explorar de forma mais adequada e significativa o escalograma.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Alessio SM (2016). *Digital Signal Processing and Spectral Analysis for Scientists: Concepts and Applications*. New York: Springer, 909p.
- Chachada S, Kuo J (2014). Environmental Sound Recognition: A Survey. *Apsipa Transactions On Signal And Information Processing*, 3. DOI: 10.1017/ATSIP.2014.12.
- Chu S et al. (2006). Where Am I? Scene Recognition For Mobile Robots Using Audio Features. Toronto, Canada: 885–888. DOI: 10.1109/Icme.2006.262661.
- Daubechies I (1992). *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics, 61, 343p.
- Domingues MO et al (2016). Explorando A Transformada Wavelet Contínua. *Rev. Bras. Ensino Fís.*, São Paulo, 38(3). DOI: 10.1590/1806-9126-RBEF-2016-0019.
- Gregory RL et al. (2019). PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software*, 4(36). DOI: 10.21105/joss.01237.
- Jhl Hansen et al. (2019). Cci-Mobile: Design And Evaluation Of A Cochlear Implant And Hearing Aid Research Platform For Speech Scientists And Engineers. In: *IEEE Embs International Conference On Biomedical & Health Informatics (BHI)*, Chicago, Il, EUA: 1-4, DOI: 10.1109/BHI.2019.8834652.
- Mattesco L, Abreu CCE (2019). O Uso do Descritor HOG no Problema da Classificação de Cenas Acústicas sobre o Domínio Wavelet. *Anais da X Escola Regional de Informática de Mato Grosso*. SBC. DOI: 10.5753/eri-mt.2019.8618.
- Mallat S (1989). A Theory For Multiresolution Signal Decomposition - The Wavelet Representation. *Ieee On Pattern Analysis And Machine Intelligence*, 11(7): 674-693. DOI: 10.1109/34.192463.
- Mallat S (2009). *A Wavelet Tour Of Signal Processing: The Sparse Way*. 3 ed. New York: Elsevier. DOI: 10.1016/B978-0-12-374370-1.X0001-8.
- Mesaros A, Heittola T, Virtanen T (2018). Acoustic Scene Classification: An Overview Of Dcase 2017 Challenge Entries. 2018 16th International Workshop On Acoustic Signal Enhancement (Iwaenc), Tokyo: 411-415. DOI: 10.1109/Iwaenc.2018.8521242.
- McFee B et al (2015). *Librosa: Audio and music signal analysis in python* (2015). In: *Proceedings of the 14th python in science conference*: 18-25. DOI: 10.25080/Majora-7b98e3ed-003.

- Nguyen TKT, Pernkopf F (2018). Acoustic Scene Classification Using A Convolutional Neural Network Ensemble And Nearest Neighbor Filters. In: Workshop On Detection And Classification Of Acoustic Scenes And Events.
- Oliveira BR et al. (2018). A wavelet-based method for power-line interference removal in ECG signals. *Res. Biomed. Eng*, 34(1): 73-86. DOI: 10.1590/2446-4740.01817.
- Radhakrishnan R et al. (2005). Audio Analysis For Surveillance Applications. *IEEE Workshop On Applications Of Signal Processing To Audio And Acoustics*: 158-161. DOI: 10.1109/ASPAA.2005.1540194.
- Sakashita Y, Aono M (2018). Acoustic scene classification by ensemble of spectrograms based on adaptive temporal divisions. *Detection and Classification of Acoustic Scenes and Events (DCASE) Challenge*.

## ÍNDICE REMISSIVO

### A

agrupamento, 49, 53, 54  
aprendizado de máquina, 8, 25, 28, 29, 59, 60, 61, 80  
Aprendizagem de Máquina, 50, 57, 83

### C

caso de uso, 41, 42, 44, 46  
CCA, 84, 85, 86, 87, 88, 95  
cenário, 39, 43, 44, 45, 85  
Cenas Acústicas, 83, 84, 95

### Ch

*checklist*, 34, 39, 41

### C

Ciência de Dados, 33  
classificação, 31, 39, 50, 56, 59, 60, 61, 62, 63, 68, 69, 72, 73, 74, 84, 85, 88, 90, 94  
clusterização, 50, 51, 54  
coeficiente, 10, 11, 89, 94  
comportamento, 8, 9, 22, 26, 30, 32, 37, 56, 70

### D

*Deep Learning*, 31, 32, 59, 81, 87, 90, 94  
desenvolvimento de SI, 35, 36, 39  
domótica, 25, 26

### E

Engenharia de Requisitos, 33, 47, 48  
escalograma, 89, 90  
espaço de atributos, 59, 60, 70, 73, 74, 76, 80, 81  
espectro, 86, 89  
espectrograma, 85, 86

### F

*Farthest First*, 49, 50, 53, 54, 56

filtro, 8, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21, 22, 89

### G

GWS, 89, 90, 91, 92, 93, 94

### I

Inteligência Artificial, 25

### M

*machine learning*, 30, 31, 84, 87  
mapeamento, 48, 60, 70, 73, 74, 77, 78, 80, 81, 88  
metodologia, 33, 34, 35, 36, 37, 39, 41, 42, 47  
MFCCs, 84, 86, 87, 90, 91, 93, 94

### O

obesidade, 49, 50, 55, 56, 57, 58

### P

polinômio, 10, 11, 16, 20, 21  
prototipação, 33, 35, 39, 41, 47

### Q

qualidade, 33, 34, 35, 39, 41, 47, 57

### R

Redes Neurais  
  Artificiais, 25, 26  
  LSTM, 26  
regressão logística, 59, 62, 63, 70  
requisitos, 28, 33, 34, 35, 39, 41, 48

### S

Savitzky-Golay, 8, 9, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23  
sensores, 26, 27, 29  
Série Temporal, 8, 9  
Sistema de Informação, 47, 48

*Software*, 47, 48, 95

superfície de decisão, 62, 77

## **T**

tendência, 8, 9, 10, 14, 15, 16, 17, 18, 19, 21,  
22, 57, 58

Transformada Wavelet, 8, 9, 10, 13, 70, 83, 84,  
87, 88, 90, 95



**Pantanal Editora**

Rua Abaete, 83, Sala B, Centro. CEP: 78690-000

Nova Xavantina – Mato Grosso – Brasil

Telefone (66) 99682-4165 (Whatsapp)

<https://www.editorapantanal.com.br>

[contato@editorapantanal.com.br](mailto:contato@editorapantanal.com.br)

ISBN 978-658831926-0



9 786588 319260