

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS
GERAIS**

Engenharia de Controle e Automação

Bruno Santos Gollo do Amaral

**UTILIZAÇÃO DA ANÁLISE QUANTITATIVA DE RECORRÊNCIA E
COEFICIENTES CEPSTRAIS DE FREQUÊNCIA MEL PARA RECONHECIMENTO
DE INCÊNDIO EM VEGETAÇÃO USANDO SINAL DE ÁUDIO**

**Betim, MG
Dezembro, 2019**

Bruno Santos Gollo do Amaral

**UTILIZAÇÃO DA ANÁLISE QUANTITATIVA DE RECORRÊNCIA E
COEFICIENTES CEPSTRAIS DE FREQUÊNCIA MEL PARA RECONHECIMENTO
DE INCÊNDIO EM VEGETAÇÃO USANDO SINAL DE ÁUDIO**

Trabalho de conclusão de curso apresentado ao curso de Engenharia de Controle e Automação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais campus Betim como parte dos requisitos para obtenção do título de Bacharel em Engenharia de Controle e Automação.

Orientador: Prof. Dr. Leandro Freitas

**Betim, MG
Dezembro, 2019**

FICHA CATALOGRÁFICA

A485u Amaral, Bruno Santos Gollo do
Utilização da análise quantitativa de recorrência e
coeficientes cepstrais de frequência mel para reconhecimento
de incêndio em vegetação usando sinal de áudio / Bruno
Santos Gollo do Amaral. – 2019.

55 f. : il.

Trabalho de conclusão de curso (Bacharelado em
Engenharia de Controle e Automação) - Instituto Federal de
Educação, Ciência e Tecnologia de Minas Gerais, Câmpus
Betim, 2019.

Orientação: prof. Leandro Freitas de Abreu.

1. Incêndio em Vegetação. 2. Análise Quantitativa de
Recorrência. 3. Coeficientes Cepstrais de Frequência Mel. 4.
Máquina de Vetores de Suporte. I. Amaral, Bruno Santos Gollo
do. II. Título.

CDU: 614.84

Bruno Santos Gollo do Amaral

**UTILIZAÇÃO DA ANÁLISE QUANTITATIVA DE RECORRÊNCIA E
COEFICIENTES CEPSTRAIS DE FREQUÊNCIA MEL PARA
RECONHECIMENTO DE INCÊNDIO EM VEGETAÇÃO USANDO
SINAL DE ÁUDIO**

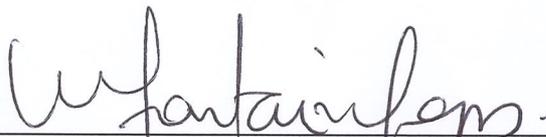
Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Controle e Automação do Instituto Federal de Minas Gerais Campus Betim como requisito parcial à obtenção do grau de Bacharel em Engenharia de Controle e Automação.

Betim, 12 de dezembro de 2019.

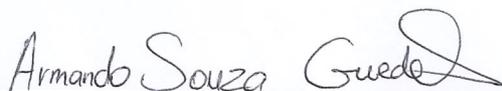
BANCA EXAMINADORA



Prof. Dr. Leandro Freitas
DAUTI – IFMG Campus Betim (orientador)



Prof. Me. Welinton La Fontaine Lopes
DAUTI – IFMG Campus Betim



Prof. Dr. Armando Souza Guedes
DAUTI – IFMG Campus Betim

*Este trabalho é dedicado à minha família e todos aqueles que,
de alguma forma, contribuíram para os meus estudos até agora.*

AGRADECIMENTOS

A todos que contribuíra para a realização deste trabalho, fica expressa aqui a minha gratidão, especialmente:

Ao Professor Leandro Freitas, pela orientação, pelo aprendizado e apoio nos momentos necessários.

Aos meus amigos e colegas de faculdade, pela rica troca de experiências, pelo apoio e paciência nos momentos necessários.

A todos que, de alguma forma, contribuíram para essa construção.

*“Em todas as coisas da natureza
existe algo de maravilhoso.”
(Aristóteles, 384 - 322 a.C.)*

RESUMO

Neste trabalho buscou-se avaliar a possibilidade de detectar incêndios em ambientes de vegetação por meio de sinais de áudio. Para isso, idealizou-se a utilização de um algoritmo de aprendizagem, mais especificamente uma Máquina de Vetores de Suporte, que realiza a tarefa de classificação binária entre áudios correspondentes ou não a fogo de acordo com as características que o alimentam. Para o treinamento dessa máquina foram obtidos na internet áudios de incêndio em vegetação e sons típicos dos ambientes de mata e campo como sons de animais, vento, água corrente e chuva por exemplo. A principal técnica proposta para a extração de características dos sinais é a Análise Quantitativa de Recorrência que é voltada para a análise de sistemas caóticos ou não lineares. Aliada a esta, empregaram-se os Coeficientes Cepstrais de Frequência Mel que facilitaram a aplicação daquela técnica, mas também contribuíram para o aumento da acurácia do algoritmo de classificação que, nos melhores modelos treinados, atingiram 95,1% de acurácia com uma taxa de falso negativo de 3%. Desta forma, concluiu-se que os métodos aqui utilizados são realmente capazes de realizar a detecção de incêndio em sinais de áudio de forma satisfatória quando comparados com resultados de outros trabalhos.

Palavras-chave: Incêndio em Vegetação. Análise Quantitativa de Recorrência. Coeficientes Cepstrais de Frequência Mel. Máquina de Vetores de Suporte.

ABSTRACT

In this work it was sought to analyze the possibility of detecting wildfire through audio signals. Therefore, it was idealized the usage of a learning algorithm, more specifically a Support Vector Machine, that performs the task of binary classification between audios corresponding or not to wildfire according to the features that feed him. To train this machine were obtained on the internet audios of wildfire and typical sounds of forest and field environments like sounds of animals, wind, running water and rain for exemple. The main technique proposed to extract the features from the signals is the Recurrence Quantification Analisis which is mainly designed for the analisis of chaotic or nonlinear systems. Allied to that, it was employed the Mel-Frequency Cepstral Coefficients which made the application of that technique easier and also contributed increasing the accuracy of the classification algorithm that, in the best trained models, achieved 95.1% of accuracy and a false negative rate of 3%. Thus, it was concluded that the utilized methods are satisfactorily able to detect wildfire in audio signals when compared to the results exposed in other papers.

Keywords: Wildfire. Recurrence Quantification Analisis. Mel-Frequency Cepstral Coefficients. Support Vector Machine.

LISTA DE ILUSTRAÇÕES

Figura 1 – Fogo Terrestre (CHOWDARY; GUPTA, 2018).	23
Figura 2 – Fogo de Superfície (CHOWDARY; GUPTA, 2018).	24
Figura 3 – Fogo de Copa (CHOWDARY; GUPTA, 2018).	24
Figura 4 – Rede de sensores sem fio. Adaptado (SINGH; SHARMA, 2017).	26
Figura 5 – Metodologia proposta para a detecção de anomalias acústicas. (ALMAA-DEED et al., 2018). Traduzido do inglês.	28
Figura 6 – Espectro de potência e linha de tendência para o ruído de fogo de copa. (KHAMUKHIN; BERTOLDO, 2016). Traduzido do inglês.	29
Figura 7 – Espectro de potência e linha de tendência para o ruído de fogo de superfície. (KHAMUKHIN; BERTOLDO, 2016). Traduzido do inglês.	29
Figura 8 – Valores médios e intervalos de confiança da acurácia dos métodos para classificação de cenas acústicas avaliados no conjunto de dados privados do DCASE usando validação cruzada de 5-fold. Os retângulos agrupam métodos que não se podem julgar de desempenhos diferentes com um intervalo de confiança de 95%. ‘H’ indica a acurácia média humana obtida por Krijnders e Holt, (2013). (BARCHIESI et al., 2015). Traduzido do inglês.	30
Figura 9 – Filtros utilizados na computação dos MFCCs (RABINER; SCHAFER, 2007).	32
Figura 10 – Representação dos filtros definido pela Equação 2 (HUANG; ACERO; HON, 2001).	33
Figura 11 – Gráficos de Recorrência de (A) um movimento periódico com uma frequência, (B) do sistema caótico de Rössler (com parâmetros $a=b=0.2$ e $c=5.7$) e (C) de um ruído uniformemente distribuído (MARWAN; CARMEN ROMANO et al., 2007). Traduzido do inglês.	35
Figura 12 – Representação da ideia de um hiperplano ótimo para padrões linearmente separáveis. Os pontos em azul são os vetores de suporte (HAYKIN, 2009). Traduzido do inglês.	39
Figura 13 – Gráficos de margens de separação suaves. À esquerda, o dado (quadrado) dentro da margem de separação está do lado correto. À direita, todavia, o dado (círculo) se encontra no lado incorreto (HAYKIN, 2009). Traduzido do inglês.	40
Figura 14 – Ilustração das duas operações de mapeamento numa SVM para a classificação de padrões (HAYKIN, 2009). Traduzido do inglês.	41
Figura 15 – Ilustrativo da metodologia utilizada.	45
Figura 16 – Matriz de confusão dos modelos 1.1, 3.1 e 3.3.	47
Figura 17 – Matriz de confusão dos modelos 1.2 e 3.2.	48
Figura 18 – Matriz de confusão do modelo 5.2.	48
Figura 19 – Matriz de confusão do modelo 4.1.	49

Figura 20 – Matriz de confusão dos modelos 1.1 e 3.1 para dos dados de teste. 50
Figura 21 – Matriz de confusão dos modelos 3.3 e 4.1 para dos dados de teste. 50

LISTA DE TABELAS

Tabela 1 – Resumo dos Kernels de Mercer (HAYKIN, 2009). Traduzido do inglês. . . .	38
Tabela 2 – Resultados dos modelos criados na etapa de treinamento e validação	46
Tabela 3 – Resultados dos melhores modelos para os dados de teste	49

LISTA DE ABREVIATURAS E SIGLAS

DCASE	Challenge on Detection and Classification of Acoustic Scenes and Events
DFT	Transformada Discreta de Fourier do inglês Discrete Fourier Transform
IEEE	Instituto de Engenheiros Elétricos e Eletrônicos
LIDAR	Light Detection And Range
MFCC	Mel-Frequency Cepstral Coefficients
RP	Recurrence Plot
RQA	Recurrence Quantification Analysis
SVM	Support Vector Machine
V-C	Vapnik-Chervonenkis
WSN	Wireless sensor network

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Os Tipos de Incêndios Florestais	23
1.2	Técnicas para a Detecção de Incêndios Florestais	24
1.3	Motivação	25
1.4	Objetivos	26
1.4.1	Objetivos Gerais	26
1.4.2	Objetivos Específicos	26
1.5	Estrutura do Trabalho	26
2	RECONHECIMENTO DE EVENTOS EM SINAIS DE ÁUDIO	28
3	EXTRAÇÃO DE CARACTERÍSTICAS E MEDIDAS	31
3.1	Coefficientes Cepstrais de Frequência Mel	31
3.2	Análise Quantitativa de Recorrência	33
3.2.1	Gráficos de Recorrência	33
3.2.2	Medidas de Complexidade	35
4	MÁQUINA DE VETORES DE SUPORTE	38
4.1	Hiperplano Ótimo para Padrões não Separáveis	39
4.2	A Filosofia por trás da Máquina de Vetores de Suporte para Reconhecimento de Padrões	41
5	METODOLOGIA	43
6	RESULTADOS E DISCUSSÃO	46
7	CONSIDERAÇÕES FINAIS	51
	Referências	52

1 INTRODUÇÃO

1.1 Os Tipos de Incêndios Florestais

Os incêndios florestais são um fenômeno que já há alguns anos despertam a preocupação da sociedade em função dos danos potenciais. De acordo com o relatório anual de incêndios florestais na Europa (SAN-MIGUEL-AYANZ et al., 2018), mais de 1,2 milhões de hectares de vegetação natural foram queimadas, 127 pessoas morreram e inúmeras riquezas em biodiversidade foram perdidas por causa desses incidentes no ano de 2017. Além disso, estima-se que foram gastos em torno de 10 bilhões de euros por causa desses incêndios.

As queimadas podem ser causadas por vários fatores como, por exemplo, atividade humana, condições climáticas, tradições culturais. Dentre eles, o mais comum é o causado pelo homem, mas os fatores que potencializam o alastramento do fogo geralmente estão relacionados às condições de vento e umidade. Segundo Chowdary e Gupta (2018), os incêndios florestais podem ser classificados em pelo menos três categorias:

1. Fogo terrestre: consome matéria enterrada (Figura 1);
2. Fogo de superfície: consomem folhas secas, galhos e vegetação rasteira (Figura 2); e
3. Fogo de copa: ocorre nas partes altas de árvores (Figura 3).

O fogo terrestre, ilustrado na Figura 1, costuma ser raro e é bastante complicado de se detectar e combater. Enquanto isso, os outros dois tipos são mais comuns. O de superfície, mostrado na Figura 2, deixa formatos bem definidos e seu alastramento é previsível. Por fim, o fogo de copa, exposto na Figura 3, apresenta as temperaturas mais altas e pode se espalhar muito rapidamente dependendo do vento.



Figura 1 – Fogo Terrestre (CHOWDARY; GUPTA, 2018).



Figura 2 – Fogo de Superfície (CHOWDARY; GUPTA, 2018).



Figura 3 – Fogo de Copa (CHOWDARY; GUPTA, 2018).

Os tipos de fogo de interesse neste trabalho serão os de superfície e de copa, mas sem realizar diferenciação entre eles. Serão exploradas na próxima seção algumas das técnicas empregadas para a detecção de incêndios em vegetação.

1.2 Técnicas para a Detecção de Incêndios Florestais

As técnicas para a detecção de incêndios florestais são as mais diversas. Cada uma delas com seus prós e contras. No entanto, a detecção precoce do foco do incêndio é um objetivo em comum, pois contribui para uma resposta mais rápida de equipes de combate a incêndio antes que o fogo se alastre por áreas muito extensas.

Uma das formas mais simples de identificar fogo em mata é feita pela observação do homem posicionado em torres. Entretanto, essa é uma forma muito limitada e desgastante do ponto de vista físico do observador, pois depende da sua atenção constante de grandes áreas e

que podem estar em locais inabitados. Assim, os meios automatizados são a chave mitigar esse problema.

Lemos (2015) apresenta o método LIDAR (do inglês *Light Detection And Range*) que se baseia na retrodifusão de um laser pulsante. Este faz uma varredura dos arredores e os dados recebidos são processados através de um sistema de reconhecimento de padrões. Essa técnica tem como atrativo o baixo índice de alarmes falsos. No entanto, possui limitações de alcance e potência do laser, bem como o risco de contato direto aos olhos.

A detecção através de satélite também é uma tecnologia bastante utilizada e é uma das mais comuns. Guang e Xu (2017) mostram um estudo de caso de uma queimada em Esperance, Oeste da Austrália que ocorreu em 2015 e demonstra a robustez do satélite geoestacionário Himawari-8 na detecção e no monitoramento com fumaça e poucas nuvens de foco de incêndio e com sensibilidade para a detecção precoce. No entanto, uma camada de nuvens pode impedir a identificação prematura.

Grande parte das técnicas utilizadas se baseia em processamento de vídeo e imagem. Poobalan e Liew (2015) propuseram um algoritmo de detecção baseado em técnicas de processamento de imagem que identifica a coloração e as bordas do fogo. Esse método apresentou acurácia de mais de 93,61%. Outro exemplo é dado por Frizzi et al. (2016) que implementa uma rede neural convolucional para identificar fogo e fumaça em vídeo com acurácia de 97,9%.

Outro método interessante para a detecção de incêndios é a utilização de redes de sensores sem fio (WSNs). Singh e Sharma (2017) defendem que as técnicas mais tradicionais como a utilização de satélites não são muito eficientes devido a fatores como a longa duração de escaneamento e a sua baixa resolução. Nesse sentido, eles afirmam que uma WSN pode detectar e monitorar em tempo real sendo mais eficiente na detecção precoce de incêndio. No entanto, é ressaltado que existem desafios com relação ao protocolo de comunicação, *software* de análise de dados e *hardware*. Esses sensores trabalham detectando fumaça, medindo temperatura, umidade e outros dados e os enviam para os outros sensores até o gateway como está ilustrado na [Figura 4](#).

1.3 Motivação

Os fatores que justificam a escolha do tema estão relacionados tanto com a área ambiental quanto com a tecnológica. As áreas de matas e florestas oferecem serviços de extrema relevância como, por exemplo, a limpeza do ar, regulação do ciclo da água e promoção de habitat para a biodiversidade. Assim, o desenvolvimento de formas proteção dessas é de grande relevância. Nesse sentido, empregam-se muitas técnicas para a detecção do incêndio sendo a maioria baseada em vídeo e imagem, mas cada uma delas possui vantagens e desvantagens dependendo do cenário. Assim, a oportunidade visualizada é a de avaliar a capacidade do método baseado em sinais de áudio para a detecção e monitoramento de queimadas em vegetação, pois pode ser uma solução mais barata, com menores exigências de processamento computacional que o vídeo por exemplo, não está sujeito a interferências de barreiras visuais, mas que também poderia trabalhar em conjunto com outras técnicas.

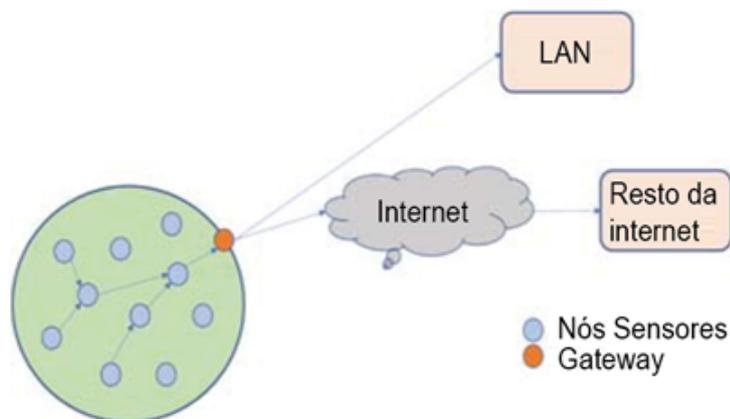


Figura 4 – Rede de sensores sem fio. Adaptado (SINGH; SHARMA, 2017).

Traduzido do inglês.

1.4 Objetivos

1.4.1 Objetivos Gerais

Este trabalho tem como objetivo geral realizar o reconhecimento de focos de incêndios em vegetação a partir de sinais de áudio utilizando a técnica da Análise Quantitativa de Recorrência (*Recurrence Quantification Analysis - RQA*) e os Coeficientes Cepstrais de Frequência Mel (*Mel-Frequency Cepstral Coefficients - MFCC*) para a extração de características. Além disso, a decisão classificatória sobre o tipo de evento presente no sinal será feita com o algoritmo de aprendizagem chamado Máquina de Vetores de Suporte (*Support Vector Machine - SVM*).

1.4.2 Objetivos Específicos

Como objetivos específicos destacam-se:

- Estudar, compreender e aplicar as técnicas RQA e MFCC;
- Estudar, compreender e implementar o sistema de classificação SVM;
- Treinar diferentes modelos de SVM para comparar seus desempenhos;
- Analisar o desempenho da metodologia utilizada comparando com o resultado de outros autores e técnicas.

1.5 Estrutura do Trabalho

Nos próximos três capítulos deste trabalho é apresentado o referencial teórico. No [Capítulo 2](#) é feita uma breve revisão bibliográfica sobre reconhecimento de eventos em sinais de áudio. Em seguida, são abordadas as teorias por trás dos Coeficientes Cepstrais de Frequência

Mel e da Análise Quantitativa de Recorrência no [Capítulo 3](#). Continuando, os conceitos sobre as Máquinas de Vetores de Suporte são descritos no [Capítulo 4](#). A metodologia para resolver o problema proposto é exposta no [Capítulo 5](#). Logo em seguida, no [Capítulo 6](#), são apresentados os resultados obtidos e as ponderações sobre eles. Por fim, no [Capítulo 7](#), são feitas as considerações finais sobre as metas propostas no trabalho e sugestões para trabalhos futuros.

2 RECONHECIMENTO DE EVENTOS EM SINAIS DE ÁUDIO

As técnicas para a análise e o processamento de áudio para o reconhecimento de padrões são diversas. Almaadeed et al. (2018) propôs a adição de microfones ao sistema de monitoramento de estradas para a detecção e a classificação de eventos. Esta técnica apresentou bons resultados, melhorando em 7% a exatidão do conjunto. Os microfones detectam sons de colisões e derrapadas em meio ao barulho característico desse ambiente. Para realizar a classificação, são extraídas dos áudios propriedades do domínio do tempo e da frequência como média, variância, fluxo espectral, curtose, assimetria, entropia, frequência média, taxa de cruzamento por zero, centroide espectral, entre outros. Assim, uma Máquina de Vetores de Suporte é alimentada com as propriedades julgadas mais interessantes para realizar a decisão classificatória. Os autores destacam ainda que o áudio deve ser dividido em pequenos quadros que precisam ser dimensionados levando em consideração o compromisso entre as componentes de alta e baixa frequência do sinal. O processo é ilustrado na Figura 5.

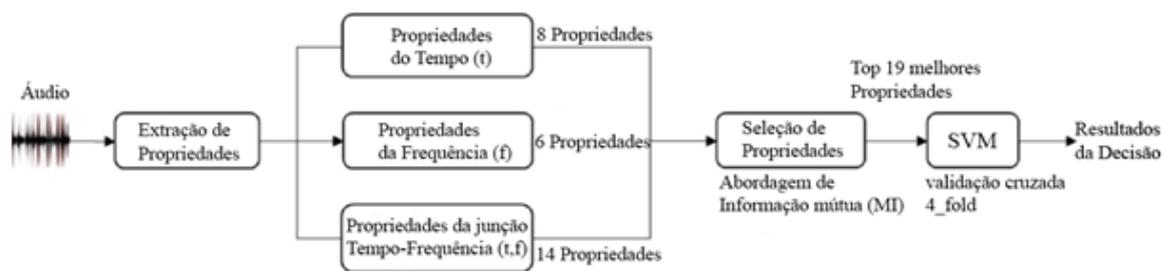


Figura 5 – Metodologia proposta para a detecção de anomalias acústicas. (ALMAADEED et al., 2018). Traduzido do inglês.

Para a diferenciação de vários tipos de sons como, por exemplo, de pássaros, aplausos, tempestades, entre outros, Rong (2017) também sugere a utilização de uma SVM com um núcleo Gaussiano para classificação de forma eficiente. Para alimentar essa, ele extrai do arquivo de áudio propriedades como a energia de curto prazo, a taxa de cruzamento de zero e os coeficientes cepstrais de frequência mel. Como resultado, obteve-se uma acurácia de mais de 80% em média.

Khamukhin e Bertoldo (2016) realizaram a análise do espectro de potência do ruído de incêndios florestais para determinar se o incêndio é de copa ou de superfície. Nesse estudo, eles observaram que a linha de tendência do espectro do fogo de copa se assemelha a uma Gaussiana localizada entre as frequências de 250 à 450Hz. Enquanto isso, para casos entre forte incêndio de superfície e copa, ou seja, intermediários, a linha de tendência apresenta características entre o ruído vermelho e o gaussiano. Exemplos desses dois casos podem ser visualizados nas Figuras 6 e 7.

Uma técnica relativamente nova é utilizada por Roma, Nogueira e Herrera (2013) para o reconhecimento sonoro de ambientes. Eles utilizam uma técnica chamada Análise Quantitativa de Recorrência que obtém um conjunto de medidas de complexidade, retiradas de Gráficos

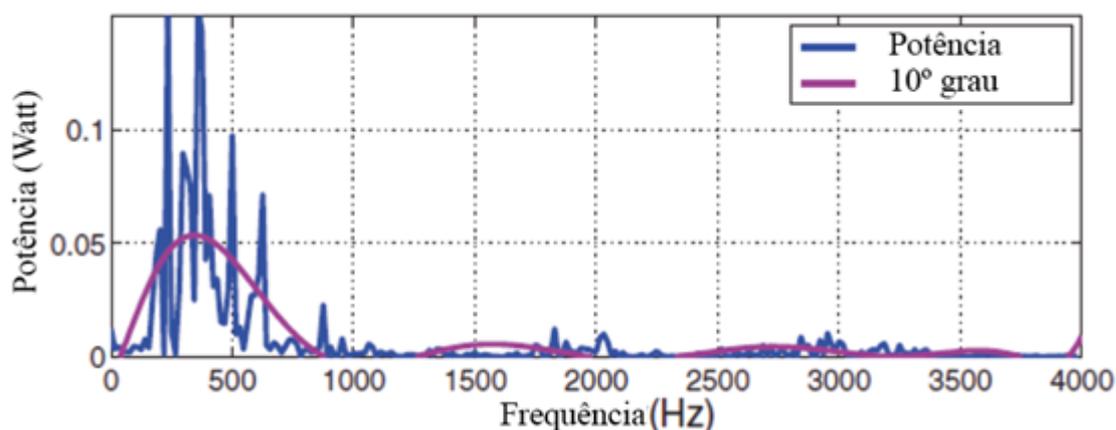


Figura 6 – Espectro de potência e linha de tendência para o ruído de fogo de copa. (KHAMUKHIN; BERTOLDO, 2016). Traduzido do inglês.

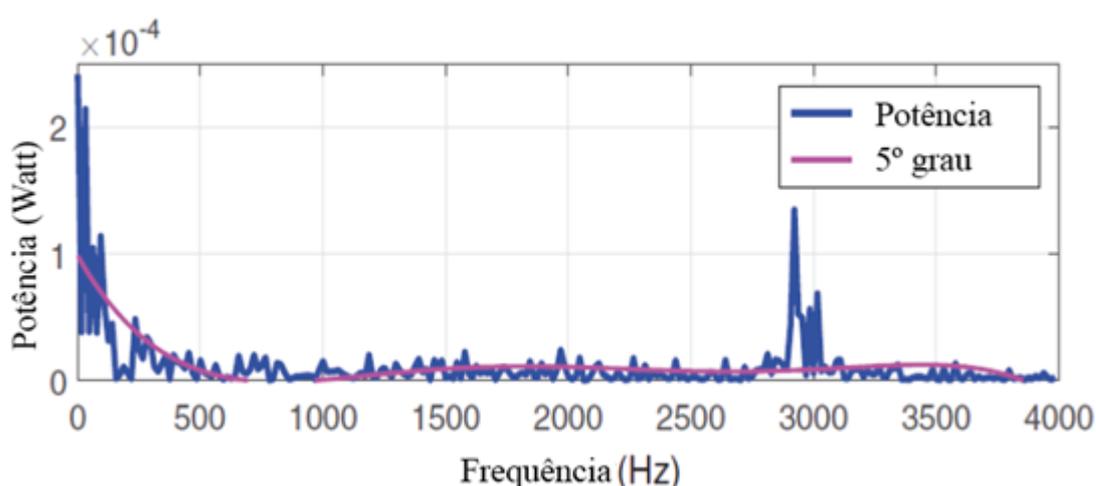


Figura 7 – Espectro de potência e linha de tendência para o ruído de fogo de superfície. (KHAMUKHIN; BERTOLDO, 2016). Traduzido do inglês.

de Recorrência (*Recurrence Plots* - RPs) (MARWAN; CARMEN ROMANO et al., 2007), para extrair características dos sinais de áudio que serão utilizadas na classificação do evento. Aliadas a isso, também são obtidas as médias e variâncias de vários coeficientes MFCCs provenientes do áudio. Esse conjunto de características é usado como entrada de um sistema de classificação que utiliza uma SVM. Esse trabalho destacou-se dentre os melhores algoritmos de classificação de cenários no desafio organizado sobre a guarda do Comitê Técnico de Processamento de Sinais de Áudio e Acústicos do IEEE, de Detecção e Classificação de Cenas e Eventos Acústicos (DCASE) (BARCHIESI et al., 2015). O comparativo do desempenho desse algoritmo, que é representado pelo acrônimo RNH, com vários outros presentes nesse desafio é apresentado na Figura 8 em que se pode observar que ele atingiu uma acurácia de aproximadamente 80%.

Nos dois próximos capítulos serão abordados de forma mais detalhada os conceitos

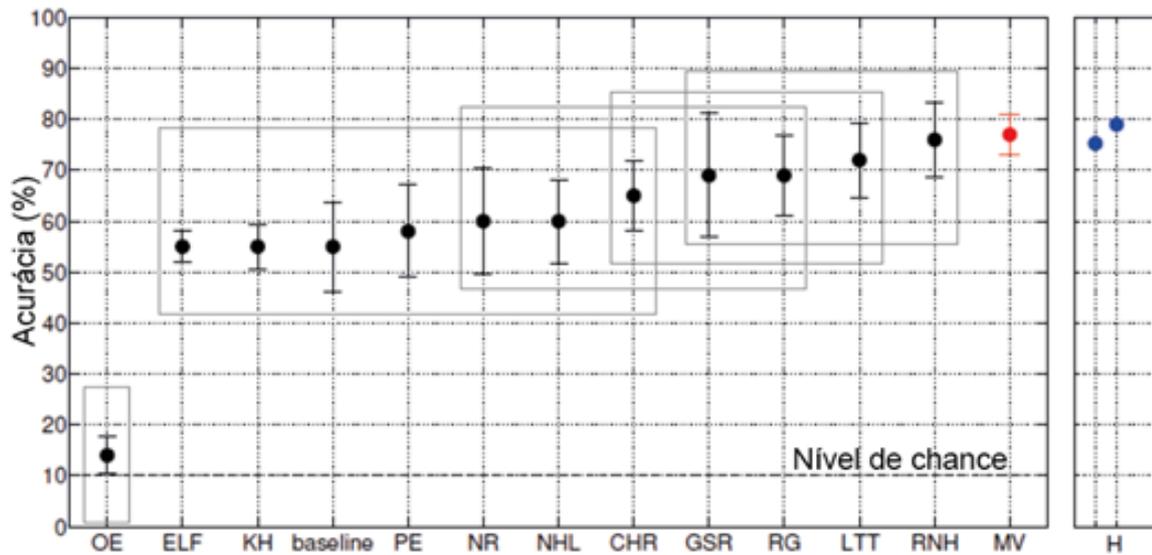


Figura 8 – Valores médios e intervalos de confiança da acurácia dos métodos para classificação de cenas acústicas avaliados no conjunto de dados privados do DCASE usando validação cruzada de 5-fold. Os retângulos agrupam métodos que não se podem julgar de desempenhos diferentes com um intervalo de confiança de 95%. ‘H’ indica a acurácia média humana obtida por Krijnders e Holt, (2013). (BARCHIESI et al., 2015). Traduzido do inglês.

dos Coeficientes Cepstrais de Frequência Mel, da Análise Quantitativa de Recorrência e da Máquina de Vetores de Suporte.

3 EXTRAÇÃO DE CARACTERÍSTICAS E MEDIDAS

O processamento de áudio é um tema extremamente trabalhado já há alguns anos com aplicações nas mais diversas áreas como, por exemplo, no reconhecimento de fala, de cenário, de música, dentre outras. O fato é que para que um sinal de áudio possa ser analisado em um ambiente computadorizado é importante empregar ferramentas e estratégias que permitam extrair informações, parâmetros e medidas relevantes sobre o sinal, seja no domínio da frequência e ou do tempo. No entanto, características do domínio do tempo costumam ser muito menos precisas do que aquelas do domínio da frequência como os Coeficientes Cepstrais de Frequência Mel (HUANG; ACERO; HON, 2001).

3.1 Coeficientes Cepstrais de Frequência Mel

Os Coeficientes Cepstrais de Frequência Mel são largamente utilizados na área de reconhecimento de fala e foram desenvolvidos em função de observações nessa área de pesquisa (MOHAMED, 2014). Estas mostraram que as larguras de banda, ao longo do eixo da frequência, para o ouvido humano não são espaçadas de forma uniforme. Nesse sentido, Davis e Mermelstein (1980) desenvolveram uma nova forma de representação Cepstral utilizando a escala Mel que se apresentou como uma boa alternativa para representar o comportamento do sistema auditivo, pois ela é linear em baixas frequências e logarítmica em altas frequências.

De forma prática, o que foi proposto por Davis e Mermelstein (1980) é realizar a análise de frequência baseada em uma série de filtros passa-faixa espaçados ao longo do eixo da frequência como pode ser visualizado na Figura 9. Esses filtros computam a média espectral ou a energia total em torno do seu centro com larguras de banda crescentes. O processo de extração dos coeficientes envolve a segmentação do sinal em pequenos quadros e a utilização da Transformada de Fourier de Tempo Discreto (*Discrete Fourier Transform* - DFT) (RABINER; SCHAFER, 2007). Nesse sentido, uma janela como, por exemplo, a de Hamming ou Hanning é aplicada a cada quadro do sinal de áudio para que se minimize descontinuidades nas bordas dos segmentos e para que se reduza o efeito de vazamento espectral (BÄCKSTRÖM, 2019; SAHIDULLAH; SAHA, 2013).

Matematicamente, o processo para obtenção dos MFCCs ocorre conforme a seguir. Inicialmente, é obtida a DFT do sinal de entrada janelado com

$$X_a[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi nk/N}, \quad 0 \leq k \leq N. \quad (1)$$

Onde N é o tamanho da DFT e $x[n]$ o sinal de entrada janelado. Em seguida são definidos M filtros ($m = 1, 2, \dots, M$), em que m é um filtro triangular como mostrado na

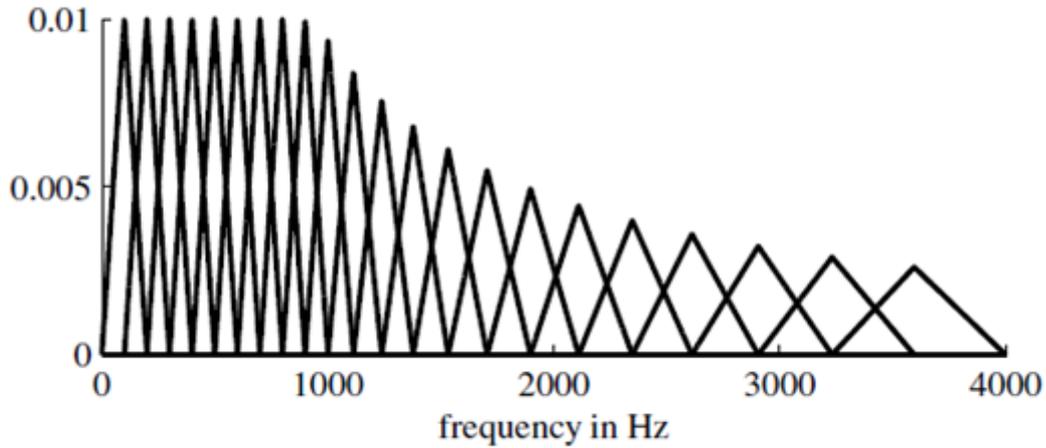


Figura 9 – Filtros utilizados na computação dos MFCCs (RABINER; SCHAFER, 2007).

Figura 10 que é definido como:

$$H_m[k] = \begin{cases} 0 & k < f[m-1], \\ \frac{2(k-f[m-1])}{(f[m+1]-f[m-1])(f[m]-f[m-1])} & f[m-1] \leq k \leq f[m], \\ \frac{2(f[m+1]-k)}{(f[m+1]-f[m-1])(f[m+1]-f[m])} & f[m] \leq k \leq f[m+1], \\ 0 & k > f[m+1], \end{cases} \quad (2)$$

em que $f[m]$ representa as frequências em Hz das bordas dos filtros uniformemente espaçados na escala Mel conforme:

$$f[m] = \left(\frac{N}{F_s}\right) B^{-1} \left(B(f_l) + m \frac{B(f_h) - B(f_l)}{M+1} \right). \quad (3)$$

Onde F_s é a frequência de amostragem em Hz , f_l e f_h são a menor e a maior frequência respectivamente do banco de filtros e B é a escala Mel que pode ser aproximada em função da frequência f por:

$$B(f) = 1125 \ln(1 + f/700). \quad (4)$$

Outra etapa do processo de obtenção dos coeficientes envolve tirar o logaritmo dos valores de saída dos filtros. Isso se deve ao fato de que o sistema auditivo humano não responde linearmente à amplitude do sinal de entrada como apontado por pesquisas na área de reconhecimento de fala (MOHAMED, 2014) e é computado como:

$$S[m] = \ln \left[\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k] \right], \quad 0 \leq m < M. \quad (5)$$

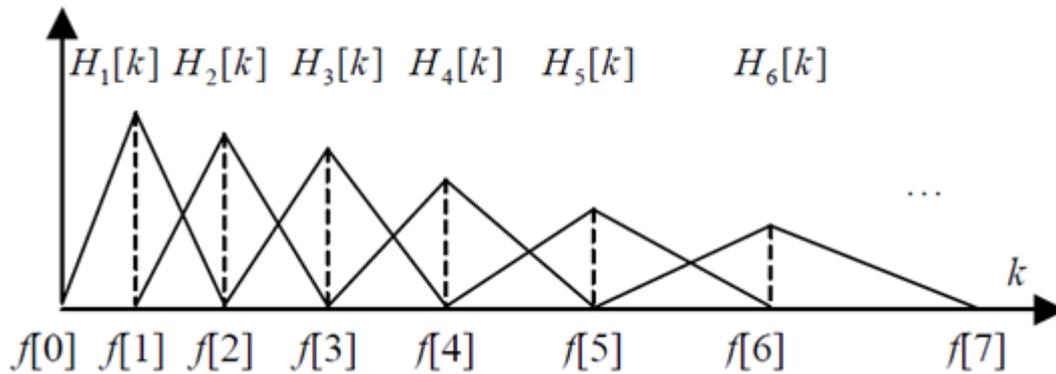


Figura 10 – Representação dos filtros definido pela Equação 2 (HUANG; ACERO; HON, 2001).

Finalmente, os MFCC são obtidos pela transformada discreta do cosseno das saídas dos M filtros:

$$c[n] = \sum_{m=0}^{M-1} S[m] \cos\left(\frac{\pi n(m + \frac{1}{2})}{M}\right) \quad 0 \leq m < M. \quad (6)$$

A quantidade de filtros M costuma variar entre 20 e 40 dentre diferentes implementações (HUANG; ACERO; HON, 2001). Esse número também leva em consideração a extensão do intervalo de frequência abrangido. Apesar de tantos filtros, tipicamente são usados apenas os 13 primeiros coeficientes cepstrais por indicações de que esses já contêm as informações mais importantes sobre o sinal (HUANG; ACERO; HON, 2001). Vale mencionar também que é comum aliar a estes os coeficientes de regressão de primeira e segunda ordem (delta e delta-delta respectivamente) que medem as mudanças dos coeficientes cepstrais ao longo do tempo (HUANG; ACERO; HON, 2001).

Tendo C_k como o vetor de coeficientes, delta (ΔC_k) e delta-delta ($\Delta\Delta C_k$) são computados como:

$$\Delta C_k = C_{k+2} - C_{k-2}, \quad (7)$$

$$\Delta\Delta C_k = \Delta C_{k+1} - \Delta C_{k-1}. \quad (8)$$

3.2 Análise Quantitativa de Recorrência

3.2.1 Gráficos de Recorrência

Antes de explicar mais sobre a Análise Quantitativa de Recorrência é preciso fazer primeiro uma breve abordagem sobre o conceito de recorrência e a ferramenta conhecida como Gráfico de Recorrência (*Recurrence Plot* - RP).

Talvez um dos exemplos mais simples e de fácil compreensão para falar sobre recorrência seja o da previsão de um evento climático como uma tempestade por exemplo. Essa

pode ser "pressentida" por uma pessoa em um dia úmido e quente no verão sem precisar de instrumentos complexos de análise. Isso é devido ao fato de que situações similares a essa ocorreram múltiplas vezes e geralmente evoluem de maneira similar. É claro que para sistemas determinísticos que podem ser modelados por equações diferenciais, é possível prever o estado de um sistema desde que se conheçam as condições iniciais e os valores dos parâmetros. No entanto, de acordo a teoria do Caos, mesmo sistemas determinísticos podem ser bastante sensíveis a perturbações das condições iniciais e tornar impossíveis previsões em longa escala de tempo (MARWAN; CARMEN ROMANO et al., 2007). Apesar disso, previsões de pequena escala ainda são possíveis mesmo em sistemas caóticos.

Previsões de uma tempestade baseada em temperatura e umidade só são possíveis em sistemas que recorrem para estados anteriores como o atmosférico. Essa forma de recorrência é uma característica intrínseca de diversos sistemas dinâmicos e pode ser usada para estudo nas mais diversas áreas como astrofísica, engenharia, mercado financeiro, biologia (MARWAN; CARMEN ROMANO et al., 2007).

Visando explorar essa característica de recorrência nos sistemas dinâmicos, Eckmann, Kamphorst e Ruelle (1987) propuseram o método denominado: Gráficos de Recorrência que traduz esses eventos de forma que possam ser visualizados através de uma representação bidimensional mesmo que o espaço de fase possua múltiplas dimensões (WEBBER; MARWAN, 2015). Eles medem como um vetor (\vec{x}_i), com sua determinada trajetória num espaço abstrato matemático chamado espaço de fase, recorre a estados visitados anteriormente dentro de uma margem de erro ε . Assim, os RPs podem ser expressos matematicamente pela matriz

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|\vec{x}_i - \vec{x}_j\|), \quad i, j = 1, \dots, N, \quad (9)$$

onde N representa o número de pontos \vec{x}_i medidos, $\Theta(\cdot)$ é a função degrau (se $w < 0$, então $\Theta(w) = 0$, caso contrário $\Theta(w) = 1$) e $\|\cdot\|$ é uma norma que define a geometria da vizinhança que contém os pontos de recorrência (MARWAN; CARMEN ROMANO et al., 2007). Nesse sentido, o RP é obtido plotando-se essa matriz com pontos nas coordenadas (i, j) utilizando-se cores diferentes para os casos em que $R_{i,j} = 0$ ou que $R_{i,j} = 1$.

Um exemplo de como os RPs são capazes de ilustrar os comportamentos de diferentes sistemas pode ser visto na Figura 11. Nesta, são mostrados RPs para um sistema de movimento circular periódico (A), um sistema caótico de Rössler (B) e para um ruído uniformemente distribuído (C). Como pode ser observado, cada sistema apresenta um padrão diferente que reflete suas propriedades como é o caso das extensas linhas diagonais para o movimento periódico, por exemplo, ou a falta de diagonais, à exceção da diagonal principal que é uma linha de simetria presente em todos RPs, no caso do ruído que reflete sua imprevisibilidade (MARWAN; CARMEN ROMANO et al., 2007).

Várias formas de explorar e aplicar o que os RPs tem a oferecer já foram desenvolvidas para conhecer mais sobre o comportamento de sistemas dinâmicos e dados medidos. Dentre essas, destaca-se a Análise Quantitativa de Recorrência (RQA) que será utilizada nesse projeto.

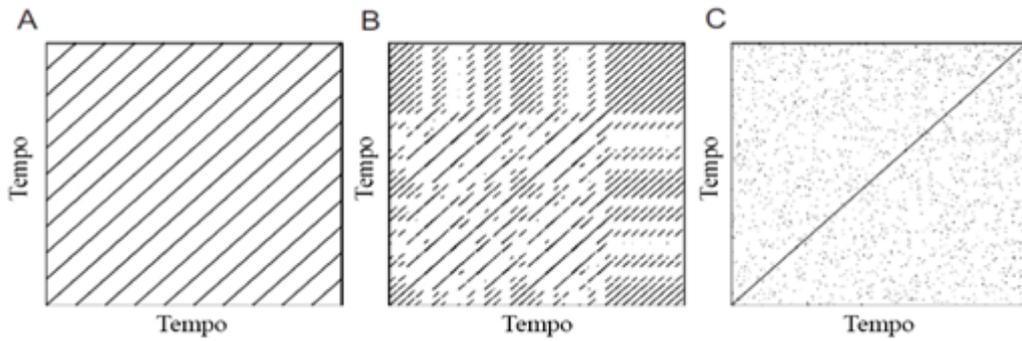


Figura 11 – Gráficos de Recorrência de (A) um movimento periódico com uma frequência, (B) do sistema caótico de Rössler (com parâmetros $a=b=0.2$ e $c=5.7$) e (C) de um ruído uniformemente distribuído (MARWAN; CARMEN ROMANO et al., 2007). Traduzido do inglês.

Na seguinte seção, esta técnica será explicada em maiores detalhes.

3.2.2 Medidas de Complexidade

Apesar dos RPs permitirem que sejam inferidas algumas características sobre a dinâmica de um sistema como, por exemplo, a periodicidade, a interpretação dos padrões neles pode estar sujeito a certa subjetividade. Nesse sentido, para ir além de apenas impressões visuais, no início dos anos 90, Zbilut e Webber propuseram um conjunto de cinco medidas de complexidade baseadas nas linhas diagonais que quantificam as estruturas nos RPs (WEBBER; MARWAN, 2015). Além disso, Marwan et al. (2002) complementou esse conjunto introduzindo mais outras três medidas que tratam das linhas verticais.

A primeira das variáveis que compõem a RQA a se destacar é a taxa de recorrência (*Recurrence Rate – RR*)

$$RR(\varepsilon, N) = \frac{1}{N^2 - N} \sum_{i \neq j=1}^N R_{i,j}^{m,\varepsilon}. \quad (10)$$

A segunda variável em RQA é a porcentagem determinística (*DET*) e representa a porcentagem de pontos que formam as linhas diagonais. Para isso, é interessante levar em consideração primeiro o histograma dos comprimentos (l) das diagonais

$$H_D(l) = \sum_{i,j=1}^N (1 - R_{i-j,j-1})(1 - R_{i+j,j+1}) \prod_{k=0}^{l-1} R_{i+k,j+k}, \quad (11)$$

assim, a variável *DET* pode ser determinada como:

$$DET = \frac{\sum_{l=dmin}^N l H_D(l)}{\sum_{i,j=1}^N R_{i,j}}. \quad (12)$$

Outra variável que deriva dessas duas primeiras é a razão entre DET e RR , isto é, o $RATIO$ que tem utilidade na detecção de transições dinâmicas (WEBBER; MARWAN, 2015)

$$RATIO = N^2 \frac{\sum_{l=l_{min}}^N l H_D(l)}{(\sum_{l=1}^N l H_D(l))^2}. \quad (13)$$

A terceira medida é basicamente o comprimento D_{max} da maior linha diagonal encontrada no RP

$$D_{max} = \arg \max_l H_D(l). \quad (14)$$

A quarta variável é referente à entropia de Shannon da frequência de distribuição dos comprimentos das linhas diagonais (ENT)

$$ENT = - \sum_{l=d_{min}}^N p(l) \ln p(l) \quad \text{com} \quad p(l) = \frac{H_D(l)}{\sum_{l=d_{min}}^N H_D(l)}. \quad (15)$$

A quinta medida RQA é a tendência (TND), que é um coeficiente de regressão linear sobre a densidade de pontos de recorrência RR_* das diagonais paralelas à principal

$$TND = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)(RR_i - \langle RR_i \rangle)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2}. \quad (16)$$

Como pode ser observado, até agora somente foram elencadas medidas relacionadas às linhas diagonais. No entanto, os RPs não são compostos apenas por linhas diagonais e informações adicionais podem ser obtidas das linhas verticais como proposto por Marwan et al. (2002). Assim, a sexta medida da RQA é a laminaridade (LAM)

$$LAM = \frac{\sum_{l=V_{min}}^N l H_V(l)}{\sum_{i,j=1}^N R_{i,j}}, \quad (17)$$

com

$$H_V(l) = \sum_{i,j=1}^N (1 - R_{i,j-1})(1 - R_{i,j+1}) \prod_{k=0}^{l-1} R_{i,j+k}, \quad (18)$$

que, de forma análoga à DET , relata a porcentagem de pontos de recorrência nas linhas verticais.

A sétima medida definida como o comprimento médio das estruturas verticais que é chamada de tempo de aprisionamento TT

$$TT = \frac{\sum_{l=V_{min}}^N v H_V(l)}{\sum_{l=V_{min}}^N H_V^e(l)}. \quad (19)$$

Por fim, a oitava medida RQA, de forma semelhante à medida diagonal D_{max} , é o comprimento máximo das estruturas verticais V_{max}

$$V_{max} = \arg \max_l H_V(l). \quad (20)$$

que, apesar de não ter uma interpretação dinâmica clara (WEBBER; MARWAN, 2015), pode ser relacionado a estados singulares em que o sistema fica preso de forma que sejam impressos

retângulos no RP. O conjunto de medidas citadas constitui a RQA, mas também existem outras medidas relacionadas àquelas. Além disso, é importante ressaltar que existem ainda outros métodos de medidas de complexidade dos RPs. Neste trabalho serão utilizadas as medidas retornadas pela função "crqa()" da ferramenta CRP toolbox ([MARWAN, 2019](#)) que são 13 no total incluindo as oito descritas nesta seção.

4 MÁQUINA DE VETORES DE SUPORTE

As Máquinas de Vetores de Suporte fazem parte da categoria dos algoritmos de aprendizagem alimentados adiante, ou seja, sem a retropropagação do erro (HAYKIN, 2009). Segundo Vapnik, (2000), a SVM é baseada no princípio da minimização de risco estrutural. Diferentemente de outras máquinas de aprendizado, como redes neurais, que assumem uma complexidade e buscam minimizar o risco, ela define o nível baixo de risco para então escolher a estrutura com a menor complexidade para atingir esse risco (VAPNIK, 2000). Esse princípio de funcionamento baseia-se no fato de que, em uma máquina de aprendizagem, a taxa de erro sobre os dados de teste fica restringida pela soma da taxa de erro de treinamento e por um termo que depende da chamada dimensão de Vapnik-Chervonenkis. Elas podem ser utilizadas em problemas de regressão linear/não-linear e classificação de padrões linearmente separáveis ou não.

A ideia base de uma SVM é a criação de um hiperplano de separação ótima, isto é, tomando como exemplo um problema de padrões separáveis, uma superfície em que a margem de separação entre os exemplos positivos e negativos seja máxima. Esse princípio é de certa forma estendido para lidar com problemas mais complexos. Outro ponto importante no algoritmo dessa máquina é o chamado *kernel* ou o núcleo do produto interno entre um “vetor de suporte” x_i e o vetor x proveniente do espaço de entrada. De acordo com o tipo de núcleo pode-se construir diferentes tipos de máquinas. Dentre os *kernels* mais comuns estão os polinomiais, gaussianos e os sigmoidais que são apresentados na Tabela 1.

Tabela 1 – Resumo dos Kernels de Mercer (HAYKIN, 2009). Traduzido do inglês.

Tipo da máquina de vetor de suporte	Kernel de Mercer $k(\mathbf{x}, \mathbf{x}_i), i = 1, 2, \dots, N$	Comentários
Máquina Polinomial	$(\mathbf{x}^T \mathbf{x}_i + 1)^p$	O expoente p é especificado a priori pelo usuário
Rede de função de base radial	$\exp\left(-\frac{1}{2\sigma^2} \ \mathbf{x} - \mathbf{x}_i\ ^2\right)$	A largura σ^2 , comum a todos kernels, é especificada a priori pelo usuário
Perceptron de duas camadas	$\tanh(\beta_0 \mathbf{x}^t \mathbf{x}_i + \beta_1)$	O teorema de Mercer é satisfeito apenas para alguns valores de β_0 e β_1

É preciso esclarecer ainda o que são os vetores de suporte. Eles são um pequeno conjunto de pontos extraídos pelo algoritmo de aprendizagem do próprio conjunto de amostras de treinamento. De forma mais específica, eles são os pontos que se encontram mais próximos da superfície de decisão (Figura 12) e, conseqüentemente, os mais difíceis de classificar (HAYKIN, 2009).

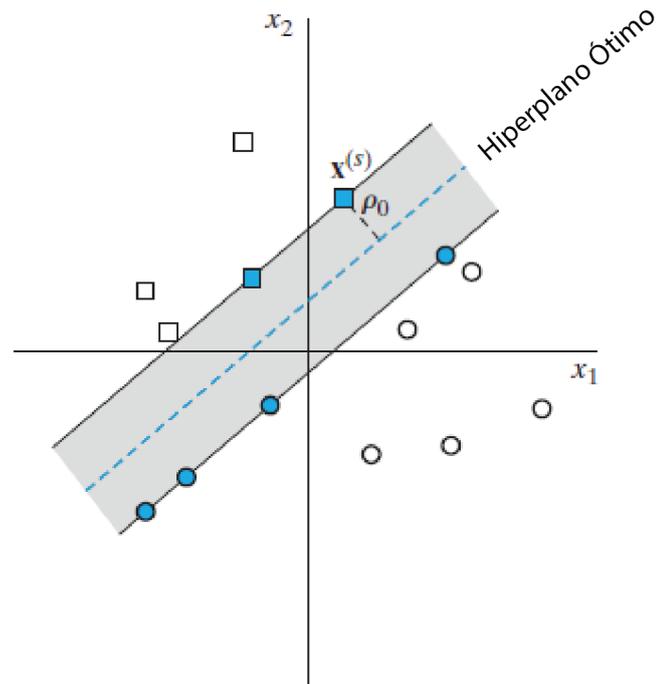


Figura 12 – Representação da ideia de um hiperplano ótimo para padrões linearmente separáveis. Os pontos em azul são os vetores de suporte (HAYKIN, 2009). Traduzido do inglês.

4.1 Hiperplano Ótimo para Padrões não Separáveis

Infelizmente nem todos os problemas se resumem a padrões linearmente separáveis como demonstrado na Figura 12. Assim, não se podem criar hiperplanos capazes de separar esses tipos de dados sem cometer erros de classificação. Nesse sentido, é necessário encontrar um hiperplano que minimiza a probabilidade desses erros.

A região em cinza vista na Figura 12 é a chamada região de separação mencionada anteriormente que se busca ser máxima. No entanto, pontos podem aparecer nessa região para os casos não separáveis. Matematicamente, é dito que a margem de separação é suave quando o ponto (\mathbf{x}_i, d_i) , onde \mathbf{x}_i é o padrão de entrada para o i -ésimo exemplo e d_i é a resposta desejada, viola a condição:

$$d_i = (\mathbf{w}^t \mathbf{x}_i + b) \geq +1, \quad i = 1, 2, \dots, N \quad (1)$$

Onde \mathbf{w} é um vetor peso ajustável e b é um bias. Essa violação pode ocorrer quando o ponto de dado se encontra dentro da margem e no lado correto da superfície de decisão, como demonstrado na Figura 13a, ou quando ele se encontra do lado errado da superfície de decisão como na Figura 13b.

A estratégia adotada para esses casos é a introdução de um conjunto de variáveis não negativas $\{\xi_i\}_{i=1}^N$ também conhecidas como variáveis soltas (HAYKIN, 2009) que servem para medir o desvio de um ponto das condições ideais de separabilidade de padrões. Assim, a Equação

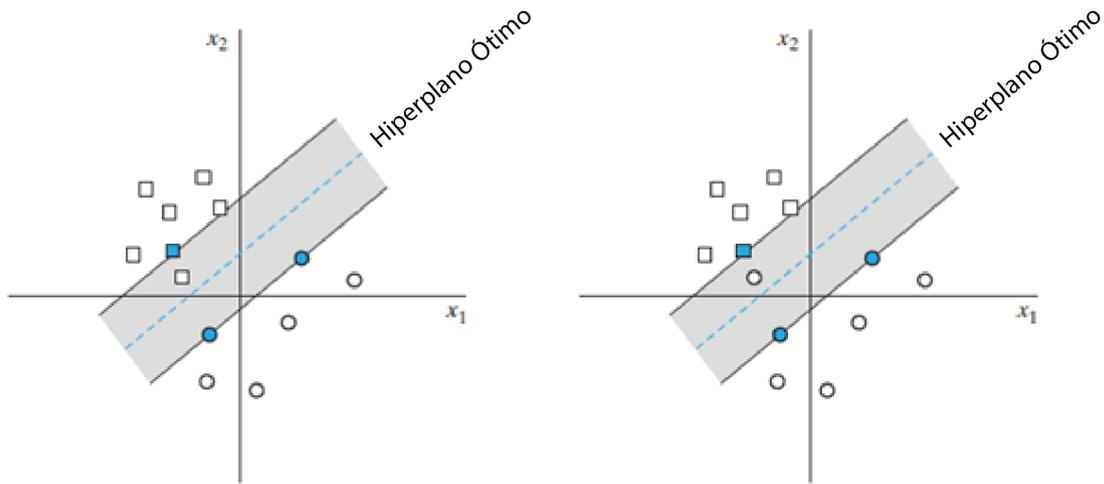


Figura 13 – Gráficos de margens de separação suaves. À esquerda, o dado (quadrado) dentro da margem de separação está do lado correto. À direita, todavia, o dado (círculo) se encontra no lado incorreto (HAYKIN, 2009). Traduzido do inglês.

(1) se torna:

$$d_i = (\mathbf{w}^t \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N. \quad (2)$$

Novamente, pretende-se encontrar um hiperplano para o qual o erro de classificação é minimizado. Para isso, minimiza-se o funcional

$$\Phi(\xi) = \sum_{i=1}^N I(\xi_i - 1). \quad (3)$$

Em relação ao vetor \mathbf{w} , tendo a Equação (2) e $\|\mathbf{w}\|^2$ como restrições. A função $I(\xi)$ é chamada função indicadora (HAYKIN, 2009), definida por

$$I(\xi) = \begin{cases} 0, & \text{se } \xi \leq 0 \\ 1, & \text{se } \xi > 0 \end{cases}. \quad (4)$$

Por se tratar de um problema de otimização não-convexo é preciso realizar uma aproximação do funcional para torná-lo matematicamente tratável. Assim, escreve-se $\Phi(\xi)$ como

$$\Phi(\xi) = \sum_{i=1}^N \xi_i. \quad (5)$$

Além disso, se o funcional for expresso em relação ao vetor peso \mathbf{w} , tem-se:

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i, \quad (6)$$

Onde C é um parâmetro positivo indicado pelo usuário que regula o compromisso entre a complexidade da máquina e o número de pontos não separáveis (HAYKIN, 2009).

Um ponto interessante de se observar é que fazer $\xi_i = 0$ para todo i nas Equações (2) e (6) resulta no caso de um problema linearmente separável.

Por fim, para resolver o problema de otimização pode-se proceder com o método dos multiplicadores de Lagrange como foi feito em Haykin (2009, p.279).

4.2 A Filosofia por trás da Máquina de Vetores de Suporte para Reconhecimento de Padrões

De forma resumida, o funcionamento de uma SVM recai em um processo que envolve duas etapas conforme demonstrado na Figura 14

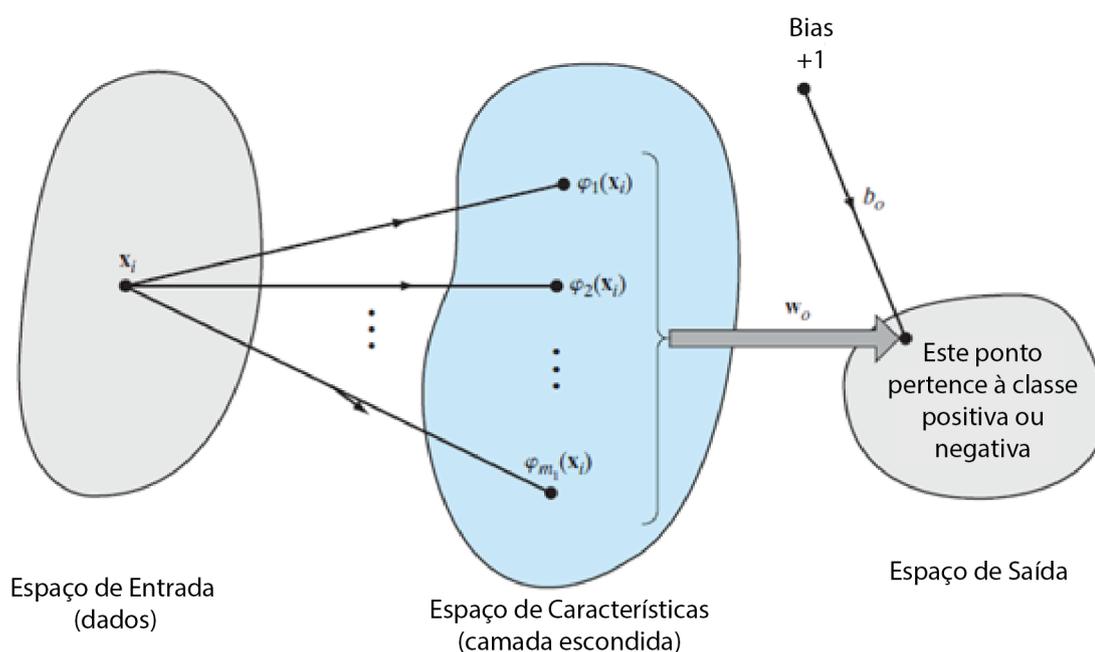


Figura 14 – Ilustração das duas operações de mapeamento numa SVM para a classificação de padrões (HAYKIN, 2009). Traduzido do inglês.

A primeira etapa diz respeito ao mapeamento não linear de um vetor de entrada em um espaço de alta dimensionalidade, isto é, os padrões do espaço de entrada são transformados em um novo espaço de características de maior dimensionalidade em que esses agora são linearmente separáveis e com alta probabilidade de acordo com o teorema de Cover (HAYKIN, 2009). Nesse sentido, o vetor $\varphi(x)$ seria de certa forma a “imagem” do vetor de entrada no espaço de características.

A segunda e última etapa trata da criação do hiperplano de separação ótimo conforme descrito na seção anterior. No entanto, ao invés do vetor x_i , agora é utilizado o vetor $\varphi(x_i)$ para a construção do hiperplano.

Devido a sua ótima capacidade de generalização, entregar desempenho robusto em problemas de classificação e regressão, facilidade de uso e uma base teórica matemática rigorosa, a máquina de vetor de suporte se estabeleceu como o algoritmo de aprendizagem

mais amplamente utilizado e é considerado o estado da arte ([HAYKIN, 2009](#)). No entanto, deve-se atentar para a principal limitação desse tipo de máquina que é o rápido aumento do consumo computacional e armazenamento de acordo com o aumento do número de amostras de treinamento.

No próximo capítulo será apresentada a metodologia utilizada para a extração de características de arquivos de áudio que deverão alimentar uma Máquina de Vetores de Suporte que realizará a tarefa de classificação.

5 METODOLOGIA

O primeiro passo que deveria ser dado para solucionar o problema proposto era a obtenção de amostras de sinais de áudio que representassem o cenário de vegetação em combustão. Além disso, também eram necessárias amostras falsas, isto é, sinais de áudio que não configurem o cenário mencionado, mas que sejam comuns ao ambiente de vegetação e mata. Assim, foram coletadas diferentes amostras de incêndios em vegetação, de ambiente de mata com sons de animais, inclusive insetos, amostras de sons de chuva, tempestade, água corrente, rodovias, trovões e vento em sítios da internet ((FONT; ROMA; SERRA, c2013);(PARTNERS IN RHYME INC, c2019); (MCKINNEY, c2018); (GOOGLE LCC, c2006)). Vale mencionar que não se fez diferenciação entre incêndio de superfície e de copa.

Uma parcela dos arquivos encontravam-se em formatos variados. Assim, optou-se por padronizar os dados no formato “.wav” que era original da maioria das amostras. Além disso, os arquivos mais extensos (com duração acima de 10 segundos) foram divididos em partes com duração em torno de seis a oito segundos, pois isso aumenta o número de amostras e evita problemas com alta demanda de memória. Por fim, garantiu-se que os arquivos estivessem em mono, isto é, sem repartição espacial como acontece no estéreo, para que as informações de um sinal estivessem armazenadas em apenas um vetor. Aqueles sinais que estavam em estéreo foram convertidos para mono. Para realizar todas essas modificações nos arquivos de áudio foi utilizado o software livre Audacity.

Com as amostras nos padrões descritos anteriormente é possível iniciar a etapa de extração de características. Nesse sentido, foi escrito um código na plataforma Matlab que carrega um arquivo de áudio para passar pelo processo de extração de características. As técnicas utilizadas nesse processo visam obter os Coeficientes Cepstrais de Frequência Mel e as medidas de complexidade da Análise Quantitativa de Recorrência. Para obter os MFCCs empregou-se a função "mfcc()" que está disponível no software utilizado voltada para esse objetivo que retorna, dentre outras medidas, 13 coeficientes cepstrais por padrão. Para as medidas RQA, foi necessário instalar um pacote de licença livre no Matlab chamado CRP toolbox, desenvolvido e disponibilizado por Marwan (2019).

O vetor que armazena o sinal de áudio pode ser muito extenso mesmo para sinais de poucos segundos de duração. Isso dificulta a aplicação da técnica de análise quantitativa de recorrência diretamente sobre o vetor do sinal, pois exige altíssima quantidade de memória para o processamento, fato constatado durante o desenvolvimento deste trabalho. Por outro lado, na extração dos MFCCs o sinal é dividido em quadros e analisado quadro a quadro. Assim, é retornada uma matriz com os 14 coeficientes para cada quadro. O tamanho de cada quadro e a sobreposição entre quadros subsequentes utilizados foram os valores padrões da própria função para todos os áudios.

Os MFCCs obtidos são o resultado de um processo de filtragem do sinal quadro a quadro sendo um extrato de informações bastante relevantes de acordo com a revisão bibliográfica

feita. Eles são armazenados numa matriz com um número de linhas, que corresponde ao número de quadros, muitas vezes menor do que o comprimento do vetor do sinal. Assim, decidiu-se aplicar a RQA sobre essa matriz uma vez que se reduziu consideravelmente os requisitos de processamento.

Como mencionado em seção anterior que abordava os MFCCs, costuma-se não utilizar o primeiro coeficiente cepstral por conter maior parte dos ruídos. Desta forma, foram empregados os últimos 13. Esse dado foi passado para a função de cálculo da RQA "crqa()" como o número de dimensões do sistema, ou seja, 13 dimensões. Outro dado passado é o tempo de atraso que, aliado ao número de dimensões, seria utilizado para a reconstrução do sinal no espaço de fase de acordo com o teorema de Takens (1981) que é necessário para criar o gráfico de recorrência e consequentemente obter as medidas RQA, mas nesse caso o número de dimensões passado é o mesmo do número de coeficientes passados, logo não ocorre reconstrução e esse valor perde relevância. O valor informado para esse parâmetro foi o atraso de uma amostra. Além disso, uma importante escolha feita foi a utilização de um raio adaptativo para manter uma taxa de recorrência de 10% uma vez que observaram-se muitos gráficos de recorrência das amostras quase que completamente preenchidos com pontos de recorrência o que poderia dificultar sua análise, mas também por ser uma taxa recomendada em trabalhos de especialistas (MARWAN, 2019).

Da função "crqa()" são retornadas 13 medidas de complexidade, mas uma delas é a taxa de recorrência que foi mantida constante para todas as amostras o que a torna irrelevante para o treinamento. Desta forma, restaram 12 medidas que deverão alimentar um algoritmo de classificação. Além disso, também são fornecidas medidas estatísticas dos MFCCs encontrados que são o valor médio e o desvio padrão de cada coeficiente ao longo de vários quadros. Assim, totalizando 38 parâmetros de entrada para esse algoritmo. Para a realização do treinamento supervisionado, os parâmetros de todas as amostras foram armazenados em uma planilha com suas respectivas classificações, isto é, *fire* para amostras de incêndio e *nonfire* para os demais casos. É muito importante mencionar também que variou-se a quantidade de parâmetros MFCC de entrada visando encontrar o melhor desempenho utilizando o mínimo de coeficientes. O critério para selecionar quais coeficientes dispensar foi baseado no exposto por Khamukhin e Bertoldo (2016) que, como apresentado nas Figuras 6 e 7, mostram que as informações mais importantes estão nas frequências mais baixas. Assim, deu-se preferência para as medidas estatísticas dos MFCCs provenientes dos primeiros filtros. Em suma, foram treinados modelos utilizando:

- Tipo 1. Todas as 38 características;
- Tipo 2. Apenas as 12 medidas RQA;
- Tipo 3. As 12 RQA mais 12 medidas estatísticas MFCCs provenientes dos primeiros 6 filtros; e

- Tipo 4. As 12 RQA mais 6 medidas estatísticas MFCCs provenientes dos primeiros 3 filtros.
- Tipo 5. Apenas as 26 medidas estatísticas MFCC.

Uma vez que as características das amostras se encontram organizadas em uma tabela, o algoritmo de aprendizagem pode ser treinado. O tipo de algoritmo utilizado foi a conhecida Máquina de Vetores de Suporte (SVM). Para criá-la e treiná-la empregou-se um aplicativo disponível no ambiente do Matlab voltado para a área de aprendizagem de máquina. Nesse sentido, foram treinados seis modelos de SVMs com os diferentes *kernels* suportados pelo aplicativo. Além disso, a validação dos treinamentos foi feita utilizando o método de validação cruzada com 5 subconjuntos (*5-fold cross-validation*) que visa evitar o problema de “*overfitting*”, ou seja, busca garantir a capacidade de generalização da máquina, mas também por ser um bom método para casos em que se tem um pequeno conjunto de amostras, pois todas as repartições serão usadas tanto para o treinamento quanto para a validação de alguns dos modelos criados no decorrer do processo. O resumo da metodologia utilizada pode ser conferido na [Figura 15](#).

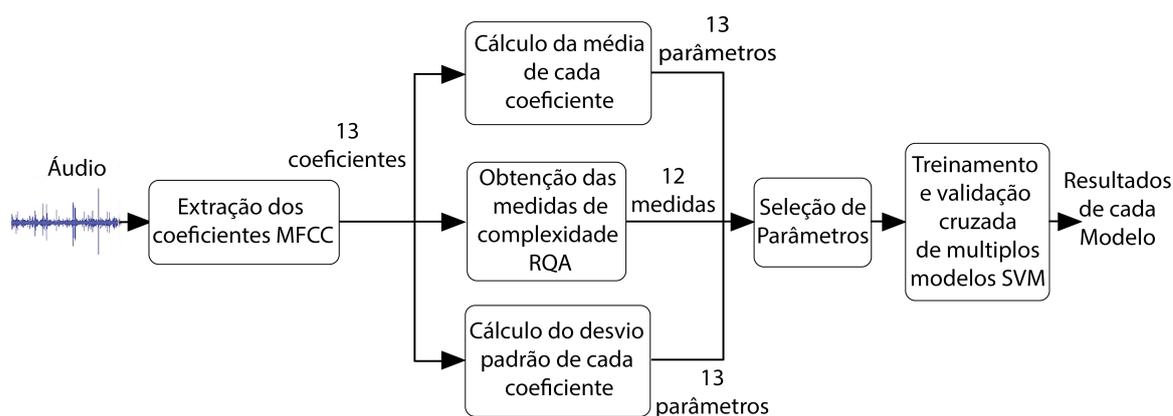


Figura 15 – Ilustrativo da metodologia utilizada.

O computador utilizado para execução de todo esse processo é um *desktop* equipado com CPU AMD Ryzen 1600, 16GB de memória RAM, placa de vídeo Radeon RX570 4GB. No próximo capítulo são expostos e comentados os resultados obtidos com a metodologia adotada.

6 RESULTADOS E DISCUSSÃO

Os resultados obtidos para cada um dos modelos de SVM treinados com um total de 81 amostras são expostos na [Tabela 2](#).

Tabela 2 – Resultados dos modelos criados na etapa de treinamento e validação

Quantidade de Características Utilizadas	Modelo	Tipo de Classificador	Acurácia (%)	Tempo de Treinamento (s)	Velocidade de Predição Aproximada (Observações/s)
Todas as 38	1.1	Linear SVM	95,1	9,2344	770
	1.2	Quadratic SVM	93,8	7,7086	730
	1.3	Cubic SVM	92,6	7,5845	640
	1.4	Fine Gaussian SVM	76,5	7,5063	660
	1.5	Medium SVM	92,6	7,3975	750
	1.6	Coarse Gaussian SVM	82,7	7,2011	730
Apenas as 12 RQA	2.1	Linear SVM	88,9	1,0528	3400
	2.2	Quadratic SVM	88,9	0,7701	2900
	2.3	Cubic SVM	88,9	0,6905	2700
	2.4	Fine Gaussian SVM	76,5	0,5991	2600
	2.5	Medium SVM	88,9	0,5154	3000
	2.6	Coarse Gaussian SVM	87,7	0,4494	3000
12 RQA + 12 MFCC	3.1	Linear SVM	95,1	0,9552	3200
	3.2	Quadratic SVM	93,8	0,7208	2400
	3.3	Cubic SVM	95,1	0,6429	2500
	3.4	Fine Gaussian SVM	76,5	0,5624	3100
	3.5	Medium SVM	92,6	0,4730	3100
	3.6	Coarse Gaussian SVM	86,4	0,3698	4100
12 RQA + 6 MFCC	4.1	Linear SVM	95,1	1,0014	3700
	4.2	Quadratic SVM	92,6	0,7513	2800
	4.3	Cubic SVM	92,6	0,6680	3100
	4.4	Fine Gaussian SVM	76,5	0,5858	2800
	4.5	Medium SVM	92,6	0,5000	3100
	4.6	Coarse Gaussian SVM	86,4	0,4108	3600
Apenas as 26 MFCC	5.1	Linear SVM	87,7	1,0195	3500
	5.2	Quadratic SVM	93,8	0,7499	2500
	5.3	Cubic SVM	91,4	0,6590	2700
	5.4	Fine Gaussian SVM	79,0	0,5583	2900
	5.5	Medium SVM	90,1	0,4597	3200
	5.6	Coarse Gaussian SVM	72,8	0,9576	3700

Como pode ser observado na [Tabela 2](#), os modelos que apresentaram melhor acurácia foram aqueles com os *kernels* linear, quadrático e cúbico que atingiram acurácias entre 93,8% e 95,1%. Isso sugere que esses apresentaram maior margem de separabilidade entre as classes analisadas conforme a teoria da SVM tratada no [Capítulo 4](#), mas também aponta a forma como os dados devem estar distribuídos no espaço multidimensional. Esses níveis de acurácia condizem com o esperado, pois na tarefa de classificação de múltiplos cenários acústicos, demonstrados na [Figura 8](#), os melhores algoritmos chegaram próximos à acurácia de 80% e o desenvolvido neste trabalho que é uma tarefa mais simples de classificação binária atingiu, naturalmente, um nível

maior. Outro ponto importante foi a diferença entre os resultados variando-se as características utilizadas para alimentar o classificador. Nota-se que apenas com as medidas RQA ou MFCC não foi possível atingir os níveis mais altos de acurácia, mas também não é necessário utilizar todas as 38 características extraídas. Nesse sentido, os modelos de destaque foram as SVMs treinadas com 6 ou 12 medidas estatísticas dos MFCCs além das 12 RQAs. Essa redução de entradas é positiva no sentido de que vai contra o sobre-ajuste e impacta em melhor desempenho computacional como pode ser visto nas últimas duas colunas da [Tabela 2](#) em que se nota uma redução expressiva no tempo de treinamento e grande aumento da velocidade de predição em relação ao cenário que se utilizou todas as 38 características.

Existem casos em que também é importante avaliar os tipos de erros de cada modelo, isto é, falsos positivos e falsos negativos. Na área da medicina, por exemplo, o resultado de um exame que seja um falso negativo pode evitar que o devido tratamento seja realizado. Para o caso de um incêndio, um falso positivo mobilizaria uma equipe de combate que viria a constatar que se tratava apenas de um alarme falso, porém, um falso negativo permitiria que o incêndio evoluísse causando maiores danos. Assim, é importante que a taxa de erro de falsos negativos nesses casos seja mínima. Tendo isso em mente, foram criadas as matrizes de confusão dos modelos com as maiores acurácias para comparação. Essas podem ser conferidas nas [Figuras 16, 17, 18 e 19](#).

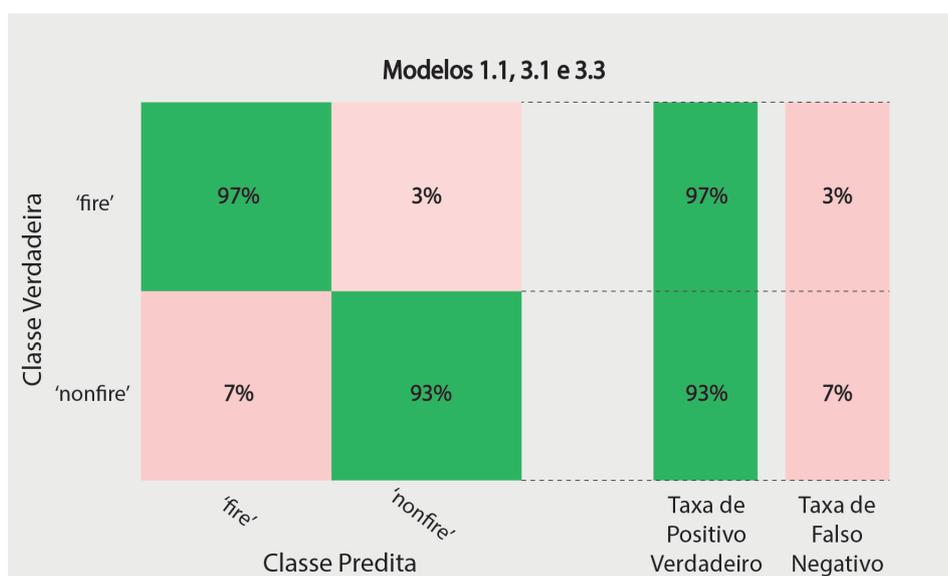


Figura 16 – Matriz de confusão dos modelos 1.1, 3.1 e 3.3.

Na [Figura 16](#) é apresentada a matriz de confusão dos modelos 1.1, 3.1 e 3.3 que apresentaram desempenhos idênticos nos quesitos acurácia e taxa de falso negativo mesmo com esses dois últimos utilizando 14 características a menos que o primeiro. Enquanto isso, o 2.1 e o 3.2, mostrados na [Figura 17](#), que apresentaram acurácia abaixo daqueles, mantiveram o bom desempenho no que diz respeito aos falsos negativos, mas apresentou mais alarmes falsos que os demais. Além disso, o modelo 5.2 treinado apenas com as medidas MFCC e que possui a

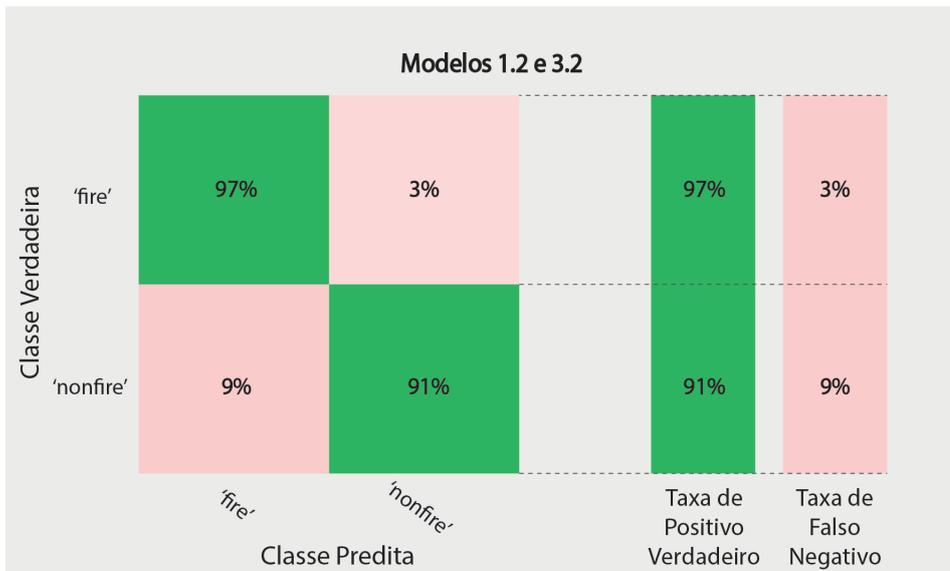


Figura 17 – Matriz de confusão dos modelos 1.2 e 3.2.

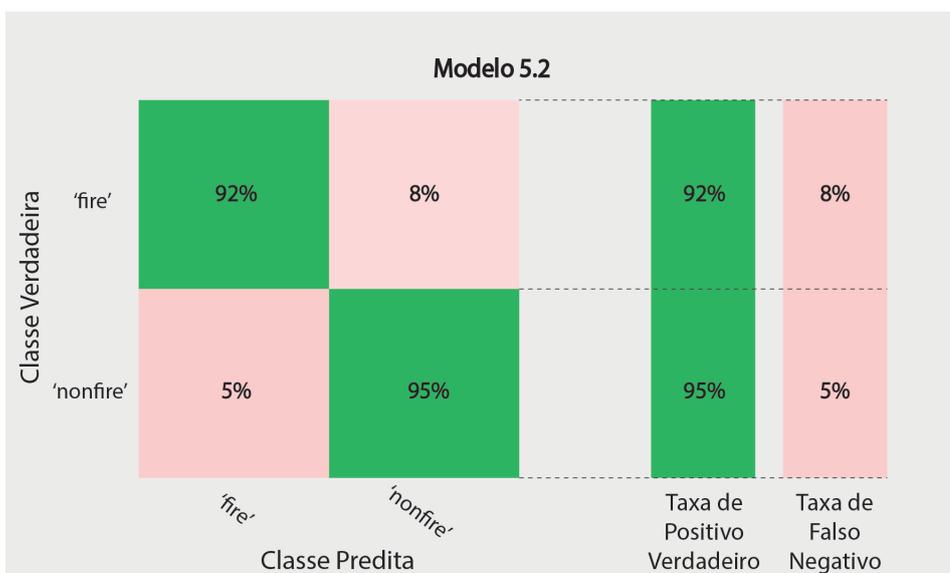


Figura 18 – Matriz de confusão do modelo 5.2.

mesma acurácia que o 2.1 e o 3.2 apresentou uma piora em relação a esses dois no quesito taxa de falso negativo para os casos em que há incêndio ('fire') como pode ser visto na [Figura 18](#). Algo similar acontece quando é feita a comparação entre o modelo 4.1, mostrado na [Figura 19](#), e aqueles da [Figura 16](#) que, apesar de possuírem a mesma acurácia, o 4.1 perde no quesito taxa de falsos negativos para os casos em que há incêndio.

Para reforçar os resultados apresentados até agora que foram obtidos com os dados de treinamento e de validação, foi realizado um último teste com os modelos 1.1, 3.1, 3.3 e 4.1, que foram os que apresentaram os melhores resultados, com relação à acurácia e à taxa de falso negativo, com um pequeno conjunto dados não visto anteriormente por nenhum dos modelos.

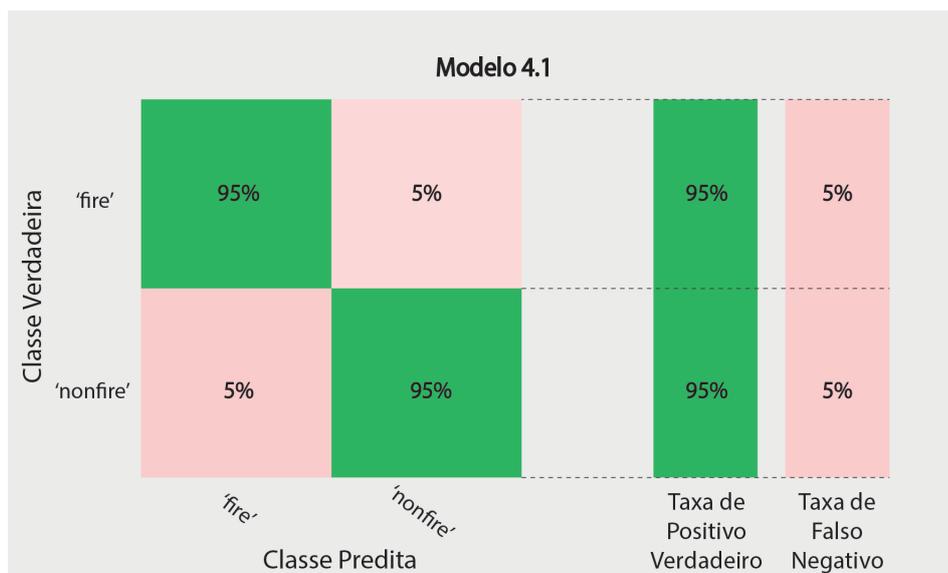


Figura 19 – Matriz de confusão do modelo 4.1.

Esse conjunto é constituído de 8 áudios de incêndio de vegetação e 12 áudios de outros eventos seguindo o mesmo método adotado anteriormente e as mesmas fontes. Os resultados desses quatro modelos para os dados de teste são apresentados na [Tabela 3](#) e nas Figuras [20](#) e [21](#).

Tabela 3 – Resultados dos melhores modelos para os dados de teste

Modelo	Quantidade de Parâmetros de Entrada	Velocidade de Predição Aproximada (Observações/s)	Acurácia (%)
1.1	38	770	90,0
3.1	24	3200	80,0
3.3	24	2500	90,0
4.1	18	3700	85,0

Observa-se nos resultados com os dados de teste expostos na [Tabela 3](#) que alguns dos modelos foram capazes de manter a acurácia próxima da atingida na etapa de validação mostrada na [Tabela 2](#). Isso demonstra então a boa capacidade de generalização desses modelos com exceção do 3.1 que apresentou grande queda de acurácia. Além disso, nas Figuras [20](#) e [21](#) é possível visualizar a forma como os modelos erraram por meio das suas matrizes de confusão. Nesse sentido, os modelos que apresentaram as menores taxas de falsos negativos foram o 3.3 e o 4.1. Diante disso, conclui-se que os melhores modelos para identificação de incêndio, levando em consideração a acurácia, quantidade de parâmetros de entrada, o tempo de treinamento, velocidade de predição e taxa de falso negativo para o caso em que há incêndio, são o modelo 3.3 e o 4.1.

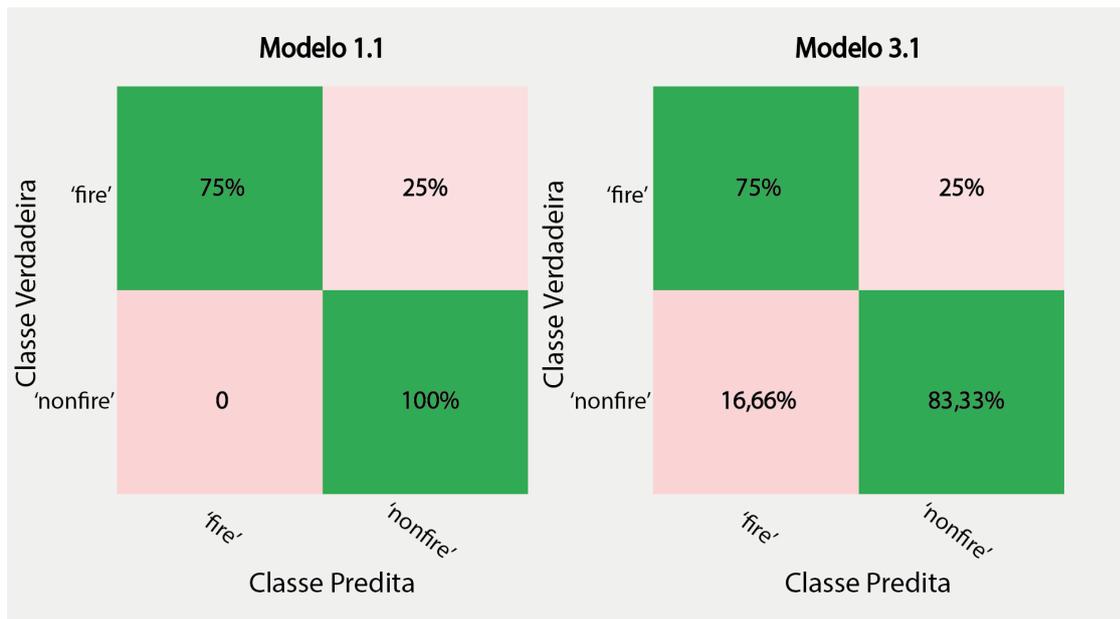


Figura 20 – Matriz de confusão dos modelos 1.1 e 3.1 para dos dados de teste.

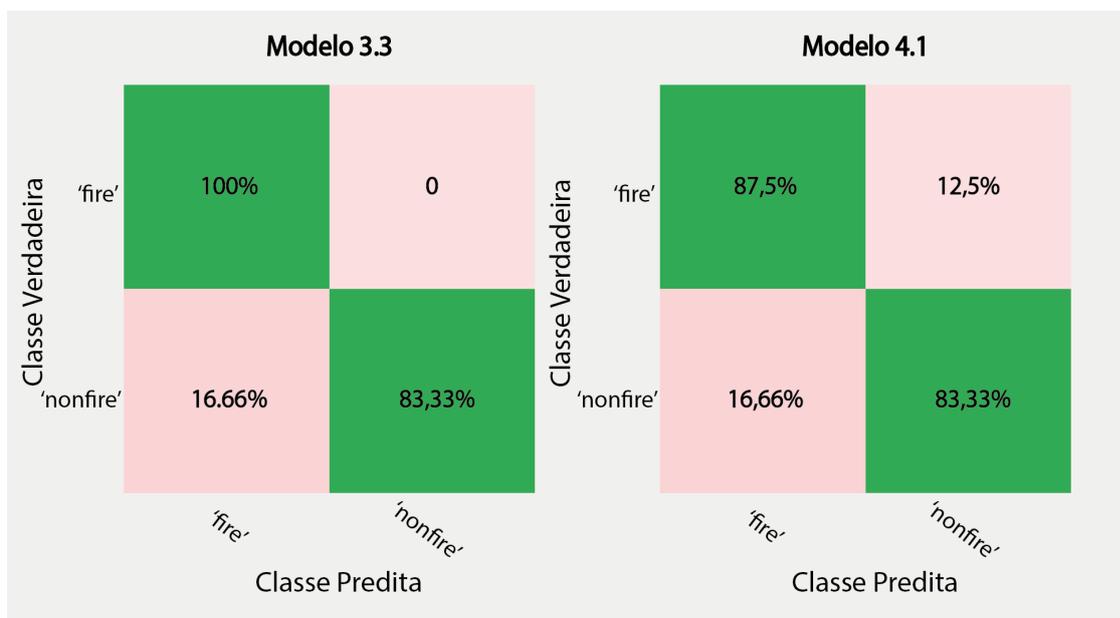


Figura 21 – Matriz de confusão dos modelos 3.3 e 4.1 para dos dados de teste.

7 CONSIDERAÇÕES FINAIS

Apesar de a quantidade de amostras utilizadas neste trabalho ser muito menor do que se gostaria, os resultados apontam para a factibilidade da detecção de incêndio em vegetação por meio de sinais de áudio utilizando as técnicas de extração de características exploradas aqui. Vale mencionar que o algoritmo de aprendizagem empregado tem reconhecida capacidade de generalização o que em conjunto com a técnica de validação cruzada K-fold contribui para a redução do impacto do número limitado de amostras sobre os resultados.

Os melhores modelos criados apresentaram acurácias de 95,1% e 90,0% nas etapas de validação e teste respectivamente. Isso os coloca próximo dos resultados apresentados por Poobalan e Liew (2015) e por Frizzi et al. (2016) que realizam detecção de fogo e fumaça em vídeo e imagem e obtiveram acurácias de 93,61% e 97,9% respectivamente. Quando comparados com os resultados obtidos por Rong (2017) e aqueles do desafio DCASE (BARCHIESI et al., 2015) em que o objetivo era a detecção de vários cenários acústicos, os melhores modelos criados neste trabalho, com destaque para o modelo 3.3, apresentaram até certa vantagem. No entanto, acredita-se que isso ocorra justamente porque a tarefa proposta neste trabalho é apenas uma classificação binária e não múltipla.

Propostas para trabalhos futuros seriam a obtenção de amostras em campo variando inclusive o tipo de vegetação como madeira e mato por exemplo para criar conjuntos mais robustos de dados uma vez que encontrou-se muita dificuldade em encontrar conjuntos de amostras relacionados com o tema explorado. Um conjunto de dados mais robusto poderia inclusive melhorar os desempenhos dos modelos apresentados. Além disso, outras propostas seriam a criação de modelos capazes de identificar o tipo e a intensidade do incêndio, a aplicação de um método de seleção de parâmetros como, por exemplo, a Análise de Componentes Principais e o desenvolvimento de dispositivos para testar funcionamento do método proposto em campo. É importante mencionar ainda que há uma quantidade extensa de parâmetros e configurações relacionadas às técnicas utilizadas neste trabalho que possuem margem para ajustes e testes.

REFERÊNCIAS

- ALMAADEED, Noor et al. Automatic Detection and Classification of Audio Events for Road Surveillance Applications. **Sensors (Basel, Switzerland)**, v. 18, n. 6, jun. 2018. ISSN 1424-8220. DOI: <10.3390/s18061858>. Disponível em: <<<http://www.ncbi.nlm.nih.gov/pubmed/29882825>%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6022152>>.
- BÄCKSTRÖM, Tom. **Windowing - Introduction to Speech Processing - Aalto University Wiki**. 2019. Disponível em: <<<https://wiki.aalto.fi/display/ITSP/Windowing>>>. Acesso em: 22 dez. 2019.
- BARCHIESI, Daniele et al. Acoustic Scene Classification: Classifying environments from the sounds they produce. **IEEE Signal Processing Magazine**, Institute of Electrical e Electronics Engineers Inc., v. 32, n. 3, p. 16–34, 2015. ISSN 10535888. DOI: <10.1109/MSP.2014.2326181>.
- CHOWDARY, Vinay; GUPTA, Mukul Kumar. Automatic forest fire detection and monitoring techniques: A survey. In: **ADVANCES in Intelligent Systems and Computing**. Springer Verlag, 2018. v. 624. p. 1111–1117. ISBN 9789811059025. DOI: <10.1007/978-981-10-5903-2_116>.
- DAVIS, Steven B.; MERMELSTEIN, Paul. **Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences**. v. 28. IEEE, 1980. p. 357–366. DOI: <10.1109/TASSP.1980.1163420>.
- ECKMANN, J. P.; OLIFFSON KAMPHORST, O.; RUELLE, D. Recurrence plots of dynamical systems. **Europhysics Letters**, v. 4, n. 9, p. 973–977, nov. 1987. ISSN 12864854. DOI: <10.1209/0295-5075/4/9/004>.
- FONT, Frederic; ROMA, Gerard; SERRA, Xavier. Freesound technical demo. In: **MM 2013 - Proceedings of the 2013 ACM Multimedia Conference**. c2013. p. 411–412. ISBN 9781450324045. DOI: <10.1145/2502081.2502245>. Disponível em: <<<https://freesound.org/>>>. Acesso em: 19 dez. 2019.
- FRIZZI, Sebastien et al. Convolutional neural network for video fire and smoke detection. In: **IECON Proceedings (Industrial Electronics Conference)**. IEEE Computer Society, dez. 2016. p. 877–882. ISBN 9781509034741. DOI: <10.1109/IECON.2016.7793196>.
- GOOGLE LCC. **YouTube**. c2006. Disponível em: <<<https://www.youtube.com/>>>. Acesso em: 19 dez. 2019.
- HAYKIN, Simon. **Neural networks and learning**. 3rd: Person Prentice Hall, 2009. v. 127, p. 7–12. ISBN 0750302259.

HUANG, Xuedong; ACERO, Alex; HON, Hsiao-Wuen. **Spoken Language Processing: A Guide to Theory, Algorithm and System Development**. Redmond: Prentice Hall, 2001. p. 1008. ISBN 9780130226167. Disponível em:

<<<https://www.pearson.com/us/higher-education/program/Huang-Spoken-Language-Processing-A-Guide-to-Theory-Algorithm-and-System-Development/PGM232550.html>>>.

KHAMUKHIN, Alexander A.; BERTOLDO, Silvano. Spectral analysis of forest fire noise for early detection using wireless sensor networks. In: 2016 International Siberian Conference on Control and Communications, SIBCON 2016 - Proceedings. Institute of Electrical e Electronics Engineers Inc., jun. 2016. ISBN 9781467383837. DOI: <10.1109/SIBCON.2016.7491654>.

KRIJNDERS, Johannes D; TEN HOLT, Gineke A. **Tone-fit and MFCC Scene Classification compared to Human Recognition**. New Paltz, 2013. Disponível em:

<<<https://www.researchgate.net/publication/255823915>>>.

LEMONS, Luiz Cesar Maia. **ESTUDO COMPARATIVO DE TECNOLOGIAS PARA DETECÇÃO PRECOCE DE INCÊNDIOS FLORESTAIS EM ÁREAS DE REFLORESTAMENTO NO BRASIL | Maia Lemos | Projetos e Dissertações em Sistemas de Informação e Gestão do Conhecimento**. 2015. Tese (Doutorado) – FUMEC. Disponível em: <<<http://www.fumec.br/revistas/sigc/article/view/2921>>>.

MARWAN, Norbert. **Cross Recurrence Plot Toolbox for MATLAB®, Ver. 5.22 (R32.3)**. 2019. Disponível em: <<<http://tocsy.pik-potsdam.de/CRPtoolbox/>>>. Acesso em: 23 nov. 2019.

MARWAN, Norbert; CARMEN ROMANO, M. et al. **Recurrence plots for the analysis of complex systems**. v. 438. Jan. 2007. p. 237–329. DOI: <10.1016/j.physrep.2006.11.001>.

MARWAN, Norbert; WESSEL, Niels et al. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. **Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics**, v. 66, n. 2, ago. 2002. ISSN 1063651X. DOI: <10.1103/PhysRevE.66.026702>. arXiv: <0201064> [physics].

MCKINNEY, Alan. **Zapsplat: free sound effects and music library**. c2018. Disponível em: <<<https://www.zapsplat.com/>>>. Acesso em: 19 dez. 2019.

MOHAMED, Abdel-Rahman. **Deep Neural Network acoustic models for ASR**. 2014. Doctoral Thesis – University of Toronto. Disponível em: <<<http://hdl.handle.net/1807/44123>>>.

PARTNERS IN RHYME INC. **Free Sound Effects.com**. c2019. Disponível em: <<<https://www.freesoundeffects.com/>>>. Acesso em: 19 dez. 2019.

POOBALAN, Kumarguru; LIEW, Siau-Chuin. FIRE DETECTION ALGORITHM USING IMAGE PROCESSING TECHNIQUES. In: INTERNATIONAL Conference on Artificial Intelligence and Computer Science, At Penang, Malaysia. Penang, 2015. Disponível em: <<https://www.researchgate.net/publication/285580944%7B%5C_%7DFIRE%7B%5C_%7DDETECTION%7B%5C_%7DALGORITHM%7B%5C_%7DUSING%7B%5C_%7DIMAGE%7B%5C_%7DPROCESSING%7B%5C_%7DTECHNIQUES>>.

RABINER, Lawrence R.; SCHAFER, Ronald W. Introduction to Digital Speech Processing. **Foundations and Trends in Signal Processing**, v. 1, n. 1, p. 1–194, 2007. ISSN 19328346. DOI: <[10.1561/20000000001](https://doi.org/10.1561/20000000001)>.

ROMA, Gerard; NOGUEIRA, Waldo; HERRERA, Perfecto. Recurrence quantification analysis features for environmental sound recognition. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. 2013. ISBN 9781479909728. DOI: <[10.1109/WASPAA.2013.6701890](https://doi.org/10.1109/WASPAA.2013.6701890)>.

RONG, Feng. Audio classification method based on machine learning. In: PROCEEDINGS - 2016 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2016. Institute of Electrical e Electronics Engineers Inc., set. 2017. p. 81–84. ISBN 9781509060610. DOI: <[10.1109/ICITBS.2016.98](https://doi.org/10.1109/ICITBS.2016.98)>.

SAHIDULLAH, Md; SAHA, Goutam. A novel windowing technique for efficient computation of MFCC for speaker recognition. **IEEE Signal Processing Letters**, v. 20, n. 2, p. 149–152, 2013. ISSN 10709908. DOI: <[10.1109/LSP.2012.2235067](https://doi.org/10.1109/LSP.2012.2235067)>. arXiv: <[1206.2437](https://arxiv.org/abs/1206.2437)>.

SAN-MIGUEL-AYANZ, Jesus et al. **Forest Fires in Europe, Middle East and North Africa 2017**. 2018. ISBN 978-92-79-92831-4. DOI: <[10.2760/663443](https://doi.org/10.2760/663443)>. Disponível em: <<<https://ec.europa.eu/jrc/en/publication/forest-fires-europe-middle-east-and-north-africa-2017>>>.

SINGH, Pradeep Kumar; SHARMA, Amit. An insight to forest fire detection techniques using wireless sensor networks. In: 4TH IEEE International Conference on Signal Processing, Computing and Control, ISPCC 2017. Institute of Electrical e Electronics Engineers Inc., set. 2017. 2017-January, p. 647–653. ISBN 9781509058389. DOI: <[10.1109/ISPCC.2017.8269757](https://doi.org/10.1109/ISPCC.2017.8269757)>.

TAKENS, Floris. Detecting strange attractors in turbulence. In: RAND D., Young LS. (eds) Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics, vol 898. Springer, Berlin, Heidelberg, 1981. p. 366–381. ISBN 978-3-540-38945-3. DOI: <[10.1007/bfb0091924](https://doi.org/10.1007/bfb0091924)>.

VAPNIK, Vladimir N. **The Nature of Statistical Learning Theory**. 2. ed. New York, NY: Springer-Verlag New York, 2000. p. 314. ISBN 978-1-4419-3160-3. DOI: <[10.1007/978-1-4757-3264-1](https://doi.org/10.1007/978-1-4757-3264-1)>. Disponível em: <<<http://link.springer.com/10.1007/978-1-4757-3264-1>>>.

WEBBER, Charles L. Jr.; MARWAN, Norbert. **Recurrence Quantification Analysis - Theory and Best Practices**. Edição: Charles L. Jr. Webber e Norbert Marwan. 1st: Springer International Publishing, 2015. p. 421. ISBN 978-3-319-07155-8. DOI: <[10.1007/978-3-319-07155-8](https://doi.org/10.1007/978-3-319-07155-8)>. Disponível em: <<<https://www.springer.com/gp/book/9783319071541>>>.

XU, Guang; ZHONG, Xu. Real-time wildfire detection and tracking in Australia using geostationary satellite: Himawari-8. **Remote Sensing Letters**, Taylor e Francis Ltd., v. 8, n. 11, p. 1052–1061, nov. 2017. ISSN 21507058. DOI: <[10.1080/2150704X.2017.1350303](https://doi.org/10.1080/2150704X.2017.1350303)>.
