UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Visual *active learning* para rotulagem por características discriminantes em paisagens acústicas

Liz Maribel Huancapaza Hilasaca

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Liz Maribel Huancapaza Hilasaca

Visual *active learning* para rotulagem por características
discriminantes em paisagens acústicas

Dissertação apresentada ao Instituto de Ciências
Matemáticas e de Computação – ICMC-USP,
como parte dos requisitos para obtenção do título
de Mestra em Ciências – Ciências de Computação e
Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e
Matemática Computacional

Orientadora: Dra. Rosane Minghim

USP – São Carlos
Julho de 2020

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

H641v Hilasaca, Liz Maribel Huancapaza
Visual active learning para rotulagem por
características discriminantes em paisagens
acústicas / Liz Maribel Huancapaza Hilasaca;
orientador Rosane Minghim. -- São Carlos, 2020.
111 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2020.

1. Visualização. 2. Características
Discriminantes. 3. Paisagens Acústicas. 4.
Rotulagem de Dados. 5. Aprendizado Ativo. I.
Minghim, Rosane, orient. II. Título.

Liz Maribel Huancapaza Hilasaca

**Visual active learning for labeling in soundscape by
discriminant features**

Dissertation submitted to the Institute of Mathematics
and Computer Sciences – ICMC-USP – in
accordance with the requirements of the Computer
and Mathematical Sciences Graduate Program, for
the degree of Master in Science. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Dra. Rosane Minghim

USP – São Carlos
July 2020

A meus pais, Cirilo e Virginia.

A meu filho Santiago.

E ao meu amore Ivar.

AGRADECIMENTOS

a Deus pelas bênçãos concedidas sempre;

aos meus queridos pais, Cirilo e Virgínia, pelo amor e apoio incondicional, sabendo que todos os agradecimentos não seriam suficientes; e ao meu querido irmão “philosophus”, Joel;

ao meu esposo Ivar pela cooperação permanente nesta árdua tarefa, pelas conversas sempre interessantes da área, a compreensão e amor brindado durante o processo;

a Juan e a Trifona queines sempre serão considerados como pais no meu coração;

à minha orientadora Dra. Rosane Minghim, pela oportunidade de realizar meu mestrado no ICMC-USP, pelo incentivo, seus conselhos, sua amizade, e orientações de forma exigente, crítica e precisa;

ao Instituto De Ciências Matemáticas e de Computação (ICMC), aos seus professores, em especial a Dra. Maria Cristina Ferreira de Oliveira (ICMC-USP) e Dr. Moacir Antonelli Ponti (ICMC-USP). Também ao Dr. Vladimir Molchanov da Universidade de Münster (WWU Münster), pelos comentários e sugestões;

ao professor Dr. Milton Cezar Ribeiro da UNESP pela colaboração e todo o suporte dado durante a pesquisa;

aos colegas do grupo de Visualização, Imagens e Computação Gráfica (VICG - ICMC/USP) e do grupo Ecologia Espacial e Conservação (LEEC - UNESP) pelas experiências e companheirismo;

ao professor Dr. César Beltrán Castañón pelo incentivo à pesquisa durante a graduação;

aos amigos por todo o seu apoio e tempo compartilhados, em especial a minha amiga Carolina Cerqueira Barbosa (USP);

à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro concedido no início deste trabalho. Processo PROEX-10423932/M;

ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de mestrado no decorrer desta pesquisa. Processo 133718/2018-2.

RESUMO

HUANCAPAZA, H. L. **Visual active learning para rotulagem por características discriminantes em paisagens acústicas**. 2020. 111 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Para a análise de Paisagens Acústicas os especialistas precisam conhecer os eventos em um determinado ambiente. Uma das formas de se adquirir esse conhecimento é através do som, isto é, de paisagens acústicas. Atualmente a tecnologia facilita as gravações de ambientes distintos, produzindo centenas de milhares de gravações de áudio todos os anos. Mas, a capacidade de análise desses sinais coletados continuamente é reduzida, sugerindo a necessidade de se automatizar esse processo, sendo necessário utilizar técnicas computacionais que auxiliem a rotulação, a exploração, a análise e o entendimento desses dados. Neste trabalho, é apresentada uma metodologia baseada em conceitos de ciência de dados, aprendizado máquina e visualização de informação, para identificar e validar um conjunto de características mais discriminantes num ambiente de aprendizado ativo, de modo que se consiga segregar categorias de eventos de interesse em paisagens acústicas, isto é, grupos de eventos similares. Na primeira parte do trabalho foram identificadas e validadas características mais discriminantes, para isto, formulou-se uma estratégia de *ranking* de importância de características, no qual por meio da combinação das n primeiras características do ranking permitiu determinar as características mais discriminantes. Na segunda parte do trabalho foram validadas as características discriminantes em um ambiente de aprendizado visual ativo, que incorpora a utilização de técnicas de projeção como t-SNE e UMAP para visualizar e analisar a segregação entre categorias de eventos da paisagem analisada. Além disso, a visualização de espectrogramas denominada *Time-Line-Spectrogram* foi incorporada para auxiliar as tarefas de rotulação do usuário no aprendizado ativo. Os resultados confirmam a efetividade das características para segregar categorias de eventos acústicos, no conjunto de dados real denominado *Cantareira-Mantiqueira* que trata de 3 categorias de eventos: anuros, pássaros e insetos. Na identificação de características discriminantes os resultados de acurácia de classificação foram: 89.91% (com 30 características para anuros, pássaros e insetos), 82.64% (com 30 características para anuros e pássaros), 100.00% (com 46 características para anuros e insetos) e 99.40% (com 14 características para pássaros e insetos). Para a rotulagem de eventos sonoros com a metodologia visual de aprendizado ativo proposta, os resultados de acurácia atingidos utilizando as características discriminantes e uma determinada configuração de parâmetros foram: 75.53% (para anuros, pássaros e insetos), 71.85% (para anuros e pássaros), 90.57% (para anuros e insetos) e 93.99% (para pássaros e insetos).

Palavras-chave: Visualização, Características Discriminantes, Paisagem Acústica, Rotulagem de Dados, Aprendizado Ativo.

ABSTRACT

HUANCAPAZA, H. L. **Visual active learning for labeling in soundscape by discriminant features**. 2020. 111 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

For the analysis of soundscapes, experts need to know the events in a given environment. One way to acquire this knowledge is through sound, that is, acoustic landscapes. Today's technology makes it easy to record in different environments, producing hundreds of thousands of audio recordings every year. However, the ability to analyze continuously collected signals is reduced, suggesting the need to automate process, it is necessary to use computational techniques that aid in the labeling, exploration, analysis and understanding of these data. In this work, a methodology based on data science, machine learning and information visualization concepts is presented to identify and validate a set of most discriminant features in an active learning environment so that interest categories of events can be segregated in soundscapes, that is, segregated groups of similar events. In the first part of the work, the most discriminant features were identified and validated. For this, a ranking strategy of the importance of features was formulated, combining the first n ranking features it was possible to determine the most discriminant features. The second part of the work validated the discriminant features of an active visual learning environment that incorporates the use of projection techniques such as t-SNE and UMAP to visualize and analyze the segregation between event categories of the analyzed landscape. In addition, a spectrogram visualization strategy called Time-Line-Spectrogram has been incorporated to assist the user's labeling tasks in active learning. The results confirm the effectiveness of the features to segregate acoustic event categories in the real data set named *Cantareira-Mantiqueira*, which deals with 3 event categories: anurans, birds and insects. In identifying discriminant features the accuracy results (10% test) were: 89.91% (with 30 features for anurans, birds and insects), 89.91% (with 30 features for anurans and birds), 100.00% (with 46 features for anurans and insects) and 99.40% (with 14 features for birds and insects). For the labeling of sound events with proposed visual active learning methodology, the accuracy results achieved using the discriminant characteristics and a certain parameters setting were: 75.53% (for frogs, birds and insects), 71.85% (for anurans and birds), 90.57% (for anurans and insects) and 93.99% (for birds and insects).

Keywords: Visualization, Discriminant Features, Soundscape, Labeling Data, Active Learning.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de espectrograma normal, com vários eventos acontecendo ao mesmo tempo.	29
Figura 2 – Gravações de um minuto de duração de eventos como: (a) anuros, (b) pássaros e (c) insetos.	29
Figura 3 – Representação da onda sonora: as características entre os picos possuem o mesmo comprimento de onda e amplitude diferente. O comprimento entre os picos de uma onda (designado por λ), e o tamanho da onda é dado pela amplitude de pico. A frequência (f) é derivada do comprimento de onda e da velocidade do som (v), como $f = v/\lambda$	34
Figura 4 – Representação básica da propagação do som. a) Fonte (fontes de som), propagação e receptor. b) Fontes de som dentro de uma paisagem acústica (geofonia, biofonia e antropofonia).	34
Figura 5 – Exemplos de espectrograma gravado em uma área úmida florestada no Campus da Universidade de Purdue às 22:00.	36
Figura 6 – Divisões das caixas de tempo e frequência para obter o ACI.	39
Figura 7 – GLCM: A figura ilustra um exemplo de como inserir valores, normalizar e computar as características na matriz de co-ocorrência $P(i, j)$	43
Figura 8 – BIC: a figura mostra um exemplo para classificar pixels como bordas (B) ou como interiores (I).	43
Figura 9 – O paradigma de aprendizado ativo conhecido como piscina ou (pool)	44
Figura 10 – Exemplo de visualização com t-SNE: os pontos coloridos representam a 6.000 dígitos do conjunto de dados MNIST que contém no total 60.000 imagens em escala de cinza de dígitos em formato de manuscritos.	45
Figura 11 – Exemplo de visualização com UMAP do conjunto de dados de dígitos(0-9) MNIST, de 784 dimensões com 70000 amostras de dados.	46
Figura 12 – Na direita o espectrograma de falsa cor foi obtido pela combinação dos índices conhecidos como ACI, H(t) e CVR nas cores vermelho, verde e azul, respectivamente.	54

Figura 13 – Visualização de dados acústicos no espectrograma de falsa cor de 24 horas de gravação: a) espectrograma de 24h de gravação gerado pelo audacity. b) espectrograma de 24h com falsa cor para conjunto de dados coletados do SERF. c) espectrograma de 24h com falsa cor para conjuntos de dados coletado do Mt Byron. d) espectrograma com agrupamentos determinados pela técnica <i>Self-Organising Map</i> SOM e <i>Fuzzy c-Mean Clustering</i>	55
Figura 14 – Impressão digital do ambiente acústico no SPL-gram: linhas indicam a hora do dia, colunas representam os dias estudados e os valores de SPL são mapeados para barras de cores. As linhas ponteadas indicam o nascer e o pôr-do-sol e os pontos marcadores indicam as atividades do coro.	56
Figura 15 – A interface do sistema apresenta: uma matriz de correlação no lado esquerdo, também oferece uma visualização <i>Radviz</i> de amostras de áudio e ancoras com as quais o usuário pode interagir, também permite a seleção de características com seus respectivos controles localizados no centro da janela. A interface facilita a exploração de dados e análise de características	56
Figura 16 – Exemplo de projeção xHiPP de conjunto de dados acústico. Cada lugar de coleta é identificado por uma cor como: roxo (<i>CostaRica1</i>), cinza (<i>CostaRica2</i>), azul claro (<i>Ilheus</i>) e marrom claro (<i>Laje</i>). Também podem ser visualizados os espectrogramas onde são destacadas algumas áreas indicando padrões sonoros distintos.	57
Figura 17 – Visualização com <i>Diel Plot</i> : nesta visualização encontra-se 13 meses de gravação de áudio. Identificando-se com as linhas amarelas pontilhadas o amanhecer e o pôr do sol da esquerda para a direita. As cores identificam espécies: aves-verdes, ortóptero-amarelas, cigarra-laranjas, vento-azuis claros, e chuva-azuis escuras, silencio-cinza	58
Figura 18 – Visualização no Histograma Polar: corresponde a 13 meses de gravação. Na legenda da imagem são identificados eventos por alguma cor específica. . .	58
Figura 19 – Metodologia proposta com o esquema para identificar as características discriminantes	62
Figura 20 – Área de coleta de dados. Realizado no Continnum Ecológico Cantareira-Mantiqueira, localizado entre o sul de Minas Gerais e o nordeste de São Paulo. A área é composta por pântanos, água, estradas, aldeias e principalmente por florestas (BARROS <i>et al.</i> , 2019)	68

Figura 21 – Um exemplo de análise visual das características do conjunto de dados DS1 (tabela 2) : no lado esquerdo (a) <i>boxplots</i> permite visualizar a variação das características com relação a os eventos do conjunto de dados, nos eixos (x = características, y = valores de características), e no lado direito (b) histogramas permite visualizar a distribuição das características, no caso índices acústicos, nos eixos (x = valores de características, y = frequência ou repetição das características).	69
Figura 22 – Ranking de características: de cima para baixo são apresentados resultados de ranking de características para os conjuntos de dados DS1 (anuros, pássaros e insetos), DS2 (anuros e pássaros), DS3 (anuros e insetos) e DS4 (pássaros e insetos). As cores no <i>ranking</i> indicam o tipo de característica segundo a nossa categorização a cor: roxo (descritores de índices acústicos), verde (características cepstrais) e laranja (descritores de imagens).	71
Figura 23 – Melhores características: de cima para baixo são apresentados resultados de acurácia média para as n melhores características dos conjuntos de dados: a) DS1 (anuros, pássaros e insetos) com 30 características com 89,91% de acurácia, b) DS2 (anuros e pássaros) com 30 características com 82,64% de acurácia, c) DS3 (anuros e insetos) com 46 características com 100,0% de acurácia e d) DS4 (pássaros e insetos) com 31 características com 99,40% de acurácia. As linhas verticais indicam para cada modelo o máximo valor de acurácia média atingida. Na legenda de cada resultado estão os 4 modelos de aprendizagem utilizados	72
Figura 24 – Visualizações dos 4 conjuntos de dados: da esquerda para direita resultados para os conjuntos de dados DS1, DS2, DS3 e DS4. Para cada conjunto projeções t-SNE e UMAP com 30°, 30°, 46° e 41° primeiras melhores características (1° e 2° linha). Na 3° e 4° linha projeções t-SNE e UMAP com todas as 238 características originais.	76
Figura 25 – Metodologia proposta para rotular dados da paisagem acústica com base no aprendizado ativo.	78
Figura 26 – Exemplo visual dos tipos de amostras consideradas. As amostras do tipo aleatório foram extraídos de maneira arbitrária. As amostras do tipo medóide foram extraídas em relação às instâncias mais próximas do centroides dos agrupamentos. Finalmente as amostras do tipo contorno, foram extraídas em relação as instâncias mais afastadas dos centroides dos agrupamentos. C1 representa o grupo 1 e C2 representa o grupo 2.	79

Figura 27 – Comparação dos resultados de acurácia (em percentagem 100%) segundo os tipos de características e os tipos de amostras: nos gráficos são ilustrados resultados quando são utilizadas as n primeiras características (linhas contínuas) e quando são utilizadas todas as 102 características (linhas tracejadas). Esta comparação é feita considerando os tipos de amostras r , rc e c , com 10 amostras por agrupamento para os conjuntos de dados DS1(a), DS2(b), DS3(c) e DS4(d). Nos gráficos, o eixo X corresponde aos valores de agrupamentos e eixo Y corresponde aos valores de acurácia.	92
Figura 28 – Visualizações dos resultados da predição dos rótulos: as linhas de cima para baixo indicam os conjunto de dados DS1 (anuros, pássaros e insetos), DS2 (anuros e pássaros), DS3 (anuros e insetos) e DS4 (pássaros e insetos). Junto as colunas de esquerda para direita indicam as projeções para: <i>Clustering</i> ; <i>Sampling</i> e <i>Listening-annotation</i> ; <i>Learning-prediction</i> ; <i>Ground-truth</i> . Indicando o número de agrupamentos (k), o número de amostras no total (p), a acurácia alcançada na predição (AC) e finalmente o número de amostras definidas por grupo (pk) para os 4 conjuntos de dados é igual a 5.	94
Figura 29 – Capturas de tela de suas interfaces do <i>Framework Soundscape-X</i>	96
Figura 30 – Exemplo de interação visual: (a) sem interação, (b) com interação. A interação é a ação feita por o usuário em selecionar um conjunto de pontos da projeção para serem mapeados de forma sincronizada com a linha de tempo do espectrograma, quando a ação de iteração é executada, espaços da linha do tempo localizada na parte superior serão coloridos de cor verde.	97

LISTA DE ALGORITMOS

Algoritmo 1 – Algoritmo de análise visual características discriminantes	67
Algoritmo 2 – Algoritmo de aprendizado ativo visual ou <i>visual active learning</i>	81

LISTA DE TABELAS

Tabela 1 – Resumo dos trabalhos relacionados.	60
Tabela 2 – Definição e descrição dos conjuntos de dados	70
Tabela 3 – Resultados da classificação: são apresentados resultados de acurácia para os 4 conjunto de dados junto como os 4 modelos de aprendizagem e para variações de percentagens para o teste (validação). Também são apresentados o número de características (.) que foram precisas para obter a acurácia. . .	73
Tabela 4 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS1 com as n primeiras melhores características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{ \cdot }$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras <i>random</i> (r), <i>medoid</i> (m), <i>contour</i> (c) e suas combinações (rm , rc , mc e rmc). No <i>heatmap</i> , o cor amarelo indica aos máximos valores de acurácia alcançada (predição de rótulos), o cor azul indica aos mínimos valores, os valores intermediários de acurácia são visualizados com uma gradiente de cores entre as cores amarela, laranja e azul.	84
Tabela 5 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS2 com as n primeiras melhores características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{ \cdot }$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras <i>random</i> (r), <i>medoid</i> (m), <i>contour</i> (c) e suas combinações (rm , rc , mc e rmc).	85

Tabela 6 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS3 com as n primeiras melhores características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc). 86

Tabela 7 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS4 com as n primeiras melhores características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc). 87

Tabela 8 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS1 com 102 características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc). 88

Tabela 9 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS2 com 102 características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc). 89

Tabela 10 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS3 com 102 características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{ \cdot }$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras <i>random</i> (r), <i>medoid</i> (m), <i>contour</i> (c) e suas combinações (rm , rc , mc e rmc).	90
Tabela 11 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS4 com 102 características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{ \cdot }$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras <i>random</i> (r), <i>medoid</i> (m), <i>contour</i> (c) e suas combinações (rm , rc , mc e rmc).	91

LISTA DE ABREVIATURAS E SIGLAS

A	<i>Anthrophony</i>
AA	<i>Acoustic Activity</i>
ACI	<i>Acoustic Complexity Index</i>
AL	<i>Active Learning</i>
ASA	<i>Average Signal Amplitude</i>
B	<i>Biophony</i>
BGN	<i>Background Noise</i>
BIC	<i>BorderInterior Pixel Classification</i>
ETC	<i>Extra Trees Classifier</i>
FFT	<i>Fast Fourier Transform</i>
FT	<i>Fourier Transform</i>
GFCC	<i>Gammatone Frequency Cepstral Coefficients</i>
GLCM	<i>Grey Level Co-occurrence Matrix</i>
H	<i>Acoustic Entropy</i>
H'	<i>Shannon Index</i>
Hs	<i>Spectral Entropy</i>
Ht	<i>Temporal Entropy</i>
LPC	<i>Linear Predictive Coefficients</i>
M	<i>Median Of Amplitude Envelope</i>
MBA	<i>Mid Band Activity</i>
MDP	<i>Multidimensional Projection</i>
MEL	<i>Mel Scale</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
NAE	<i>Number of Acoustic Events</i>
NDSI	<i>Normalized Difference Soundscape Index</i>
PA	<i>Paisagem Acústica</i>
S	<i>Silhouette</i>
SNR	<i>Singal to Noise Ratio</i>
t-SNE	<i>t-Distributed Stochastic Neighbor Embedding</i>
UMAP	<i>Uniform Manifold Approximation and Projection</i>
WAV	<i>WAVEform audio format</i>

SUMÁRIO

1	INTRODUÇÃO	27
1.1	Contextualização e Motivação	27
1.2	Objetivos	31
1.3	Organização do Trabalho	31
2	CONCEITOS FUNDAMENTAIS	33
2.1	Considerações Iniciais	33
2.2	Processamento do Sinal de Áudio	33
2.2.1	<i>O som</i>	33
2.2.2	<i>Propagação do Som</i>	34
2.2.3	<i>Transformada de Fourier</i>	35
2.2.4	<i>Espectrograma</i>	35
2.3	Descritores de Características	35
2.3.1	<i>Índices Acústicos (Temporal - Frequência)</i>	37
2.3.2	<i>Características Cepstral</i>	40
2.3.3	<i>Características da Imagem</i>	41
2.4	Aprendizado Ativo	43
2.5	Projeções Multidimensionais	44
2.5.1	<i>Avaliação das Projeções</i>	45
2.6	Considerações Finais	47
3	TRABALHOS RELACIONADOS	49
3.1	Considerações Iniciais	49
3.2	Análise de Paisagem Acústica	49
3.3	Aprendizagem Ativa em Paisagem Acústica	53
3.4	Visualização em Paisagens Acústicas	54
3.5	Considerações Finais	57
4	ANALISE VISUAL DE CARACTERÍSTICAS DISCRIMINANTES	61
4.1	Considerações Iniciais	61
4.2	Metodologia	61
4.2.1	<i>Pre-processamento dos Dados</i>	62
4.2.2	<i>Descrição de Características</i>	62
4.2.3	<i>Análise de Características Discriminantes</i>	64

4.2.4	<i>Visualização</i>	66
4.3	Conjunto de Dados	68
4.4	Resultados e Avaliação	68
4.4.1	<i>Ranking de Características</i>	70
4.4.2	<i>Seleção das Características mais Discriminantes</i>	70
4.4.3	<i>Visualização Empregando Características mais Discriminantes</i>	74
4.5	Considerações Finais	74
5	VISUAL ACTIVE LEARNING PARA ROTULAÇÃO	77
5.1	Considerações Iniciais	77
5.2	Metodologia	77
5.2.1	<i>Clustering</i>	78
5.2.2	<i>Sampling</i>	78
5.2.3	<i>Listening-annotation</i>	79
5.2.4	<i>Learning-Prediction</i>	80
5.2.5	<i>Validation</i>	80
5.3	Resultados e Avaliação	80
5.3.1	<i>Análise de Clustering e Sampling</i>	81
5.3.2	<i>Análises das Visualizações</i>	92
5.3.3	<i>Framework - Rotulador de Paisagens Acústicas</i>	93
5.4	Considerações Finais	95
6	CONCLUSÕES	99
6.1	Contribuições	99
6.2	Limitações	101
6.3	Trabalhos Futuros	102
REFERÊNCIAS		103

INTRODUÇÃO

1.1 Contextualização e Motivação

Qual a situação do nosso meio ambiente? Um dos meios para responder essa pergunta é o som. “O som é uma propriedade perpétua e dinâmica de um determinado ambiente e está intimamente ligado à caracterização desse ambiente” (PIJANOWSKI *et al.*, 2011a). Podemos ouvir sons em qualquer área, espaço ou ambiente, mas os sons em determinados ambientes são geralmente complexos, já que múltiplas fontes localizadas em diferentes lugares do ambiente emitem sons em diferentes instantes de tempo com diferentes intensidades. Além disso, o comportamento acústico de um determinado ambiente pode refletir algum tipo de mudança no seu meio (VILLANUEVA-RIVERA *et al.*, 2011a). Por essa razão o som pode ser utilizado como um indicador para mensurar a saúde e biodiversidade de um ambiente (SERVICK, 2014).

Os dados de áudios gravados podem contribuir para uma série de pesquisas ecológicas, especificamente no monitoramento de atividade animal e de acontecimentos climáticos em ambientes terrestres e aquáticos. Os avanços rápidos em hardware e poder de computação, permitem instalar sensores acústicos autônomos em locais expostos por várias semanas ou meses de gravação contínua. Como consequência, existem grandes quantidades de dados de áudio ainda não exploradas e analisadas por falta de recursos que otimizem a análise e exploração.

Como exemplo, laboratórios como da Universidade *Purdue* em *West Lafayette Indiana*, da Universidade de Tecnologia de *Queensland* em *Brisbane Austrália*, coletaram durante anos milhares de terabyte de dados de Paisagem Acústica (PA) ao redor do mundo ((SERVICK, 2014), (TOWSEY *et al.*, 2014c), (SANKUPELLAY *et al.*, 2015a)). Para fazer análise de gravações de distintas paisagens precisa-se de observadores treinados para reconhecer que evento está acontecendo em um determinado tempo da gravação, além de longo período para adquirir conhecimento e fazer uma tomada de decisão oportuna ou adequada. Assim, os especialistas

necessitam de estratégias e técnicas para análises mais eficientes que possam agilizar a aquisição de conhecimento desses dados. A análise de gravações acústicas é uma técnica cada vez mais frequente e indispensável para entender ambientes naturais e avaliar o impacto da atividade humana e das mudanças climáticas sobre eles.

O trabalho de (SOUTHWORTH, 1969) apresenta um dos primeiros usos do termo *Soundscape* na literatura. Anos depois Pijanowski *et al.* (2011a), definem o *Soundscape*¹, como o termo utilizado para descrever a relação entre uma paisagem e seus sons. Assim também outras definições para paisagem acústica podem ser encontrados nos trabalhos de Krause (2002) e Farina (2006). Nesse contexto, a paisagem acústica ecológica (*Soundscape ecology*) é definida como a relação entre os sons, da biofonia², geofonia³ e antrofonia⁴, no espaço e no tempo (PIJANOWSKI *et al.*, 2011b).

A análise das gravações de paisagens acústicas desempenham um papel cada vez mais importante para entender o comportamento de determinados ambientes. Alguns exemplos são: o reconhecimento e a classificação de espécies de aves com base em sílabas de cantos (RAGHURAM *et al.*, 2016) (STOWELL; PLUMBLEY, 2014) (CHOU; LEE; NI, 2007), o estudo ecológico sobre espécies ameaçadas (sapos nativos) e espécies invasoras (HU *et al.*, 2009) (XIE *et al.*, 2016), a determinação da influência do desenvolvimento populacional urbano na vida e no comportamento dos animais (JOO; GAGE; KASTEN, 2011), a mensuração da biodiversidade especificamente no mar (PARKS; MIKSIS-OLDS; DENES, 2014) e a segregação entre paisagens acústicas diferentes (SANKUPELLAY *et al.*, 2015b).

Os áudios das paisagens podem ser tratados em domínios distintos (MITROVIC; ZEPPELZAUER; BREITENEDER, 2010), sendo os mais estudados: domínio de tempo e frequência. Cada domínio de tratamento do áudio pode ter vantagens e desvantagens segundo a complexidade do som. A complexidade do som nas gravações de paisagem acústica está nas diferentes amplitudes e frequências para um instante de tempo em que os eventos acústicos acontecem, considerando que podem existir eventos de interesse mascarados por ruídos de diferentes intensidades (exemplo: Figura 1). Além disso, as similaridades entre sons de diferentes contextos aumentam essa complexidade. Por outro lado a complexidade do som cresce quando dois eventos sonoros acontecem em paralelo num determinado tempo (manhã, tarde e noite) e habita animal, por exemplo, sapos e insetos geralmente ocorrem (acusticamente ativos) à noite e seus sons são simultâneos (PIJANOWSKI *et al.*, 2011a), aderindo assim, uma carga adicional à classificação das categorias de eventos como anuros, pássaros e insetos sendo seus padrões sonoros como segue nas Figuras 2a, 2b e 2c respectivamente.

¹ A tradução de *Soundscape* para português é paisagem sonora ou paisagem acústica

² A biofonia é todo som que é gerado por organismos vivos, não humanos, no habitat.

³ A geofonia se refere aos sons não biológicos, como o vento nas árvores, água em uma correnteza, ondas nas praias, movimentos da Terra.

⁴ A antrofonia se refere a todos os sons produzidos por humanos, sejam sons controlados, como a música ou teatro, ou caóticos e incoerentes, como a maioria de nossos barulhos.

Figura 1 – Exemplo de espectrograma normal, com vários eventos acontecendo ao mesmo tempo.

Fonte: <https://www.ecosounds.org/>

Figura 2 – Gravações de um minuto de duração de eventos como: (a) anuros, (b) pássaros e (c) insetos.

Fonte: Elaborada pelo autor.

Dentro da análise das gravações de paisagens acústicas, alguns trabalhos recentes chegaram a distinguir locais ou cenários das paisagens acústicas através de suas características (SANKUPELLAY *et al.*, 2015a), (DIAS, 2018), (PHILLIPS; TOWSEY; ROE, 2018). Mas, de fato, na análise ainda é importante para o usuário conhecer e saber, dentro desses locais, as situações ambientais diferentes num nível específico, por exemplo: distinguir a faixa temporal que pode ser manhã, tarde e noite, distinguir grupos de animais, identificar algum tipo de fenômeno natural, ou localizar qualquer categoria de eventos distintos que existem e estão presentes nesses ambientes.

O estudo em paisagens acústicas está focado principalmente na utilização de um conjunto de características extraídas do sinal do áudio ou da imagem do espectrograma para analisar ou quantificar ambientes. Para fazer uma análise da paisagem acústica por intermédio das suas características precisa-se de algum tipo de técnica que auxilie a exploração, análise e entendimento desses dados. Nesse contexto recorrer à visualização de dados multidimensionais⁵, uma área do conhecimento cujo propósito é facilitar a interpretação de dados por meio de representações gráficas interativas expõem propriedades relevantes dos dados (CARD; MACKINLAY; SHNEIDERMAN, 1999), técnicas geralmente usadas em qualquer problema que culmine na análise. Assim, as técnicas de visualização proporcionam maneiras visuais de identificar padrões e tendências num determinado conjunto de dados, pois acredita-se que a visualização desempenha

⁵ dados multidimensionais são aqueles instâncias que constam de múltiplos atributos ou características.

sempre um papel importante no desenho, no ajuste e na validação na implementação de soluções a problemas do mundo real.

Especificamente no artigo de [Liu et al. \(2017\)](#) descreve-se um resumo de técnicas de visualização multidimensional que podem ser utilizadas em cada um dos três estágios da visualização (transformação de dados, mapeamento visual e transformação da vista). Um exemplo dessas técnicas de visualização são as projeções multidimensionais, que tratam principalmente da redução de um espaço multidimensional de m dimensões para outro espaço reduzido com p dimensões ($X^m \Rightarrow X^p, p < m$), como por exemplo 2 ou 3 dimensões. Existem vários trabalhos com domínios diferentes, e as projeções multidimensionais são empregadas como um meio de visualização para realizar tarefas de análise e exploração de dados ([Nonato; Aupetit, 2019](#); [Wang et al., 2019](#); [ESPADOTO; HIRATA; TELEA, 2019](#)). Assim, fica claro que a abordagem de visualização de informação pode auxiliar na formulação de soluções para problemas de análise e exploração de dados multidimensionais provenientes de paisagens acústicas.

À análise das características é o estágio mais importante na análise de paisagem acústica, no Capítulo 3, são escassos os trabalhos que fazem um estudo da capacidade discriminativa das características considerando a importância ou da combinação de um certo conjunto de características para identificação de fenômenos ou espécies.

Além disso, é preciso reconhecer que um dos problemas mais relevantes da análise das paisagens acústicas é como distinguir categorias de eventos. Este problema é pela falta de conhecimento da capacidade discriminativa de um certo conjunto de características. Neste cenário, pode-se notar a necessidade de uma metodologia para identificar as características mais discriminantes para elementos dentro de uma determinada paisagem acústica.

Por outro lado, os usuários especialistas precisam conhecer os eventos que acontecem dentro de uma paisagem, para extrair conhecimentos e tomar decisões. Tempo e esforço do usuário são necessários para realização de tarefas como escutar e rotular os eventos de interesse. Nesse cenário, são necessárias novas metodologias para rotulação automática e para apoio na rotulação de eventos acústicos baseados em modelos computacionais de aprendizado.

Um dos métodos para rotular grandes volumes de dados provenientes de paisagens acústicas é o aprendizado ativo, conhecido em inglês como *Active Learning* ([SETTLES, 2009](#)), o diferencial importante deste tipo de aprendizagem é a participação do usuário de forma interativa para aprender com base em uma estratégia e conseguir resultados em novos conjuntos de dados. Alguns trabalhos com relação a paisagem acústica ecológica que empregam o paradigma de aprendizado ativo para enfrentar o problema da rotulação de grandes conjuntos de dados são apresentados por [Han et al. \(2016\)](#), [Shuyang, Heittola e Virtanen \(2017\)](#), [Qian et al. \(2017\)](#).

Em abordagens de aprendizado ativo a participação do usuário influencia na rotulação de um pequeno conjunto de amostras. Em seguida, um modelo de aprendizagem é treinado com a informação dos rótulos. Finalmente, esse modelo é utilizado para a predição de rótulos de uma

grande quantidade de instâncias do conjunto de dados. Nesse sentido, nos trabalhos que fazem rotulação de dados acústicos por meio do aprendizado ativo, pode-se notar a falta de estudos de como as técnicas de visualização podem assistir aos usuários nas tarefas de rotulação e auxiliar na segregação de eventos.

A pesquisa foi aplicada em dados cedidos pelo nosso parceiro o professor Milton Ribeiro do Laboratório de Ecologia Espacial e Conservação (LEEC) da Universidade Estadual Paulista "Júlio de Mesquita Filho"(UNESP - Rio Claro).

1.2 Objetivos

De modo a prover uma solução para os problemas citados anteriormente, considerando o cenário de análise de paisagens acústicas, o objetivo principal deste trabalho de mestrado é desenvolver metodologias para apoio à etiquetagem de categorias de eventos sonoros em paisagens acústicas ecológicas que envolva o computador e o usuário no processo, baseado na identificação de um conjunto de características mais discriminantes. De modo que por meio desse conjunto de características se consiga segregar categorias de eventos de interesse em paisagens acústicas. Para atingir este objetivo são definidos os seguintes objetivos específicos:

- (i) Desenvolver uma metodologia para análise e identificação das características mais discriminantes que permitam distinguir ou segregar categorias de eventos em paisagens acústicas.
- (ii) Desenvolver uma metodologia para rotular categorias de eventos acústicos com base no paradigma de aprendizado ativo e visualização com projeções multidimensionais.
- (iii) Avaliar a capacidade das características mais discriminantes em tarefas de classificação e rotulação de categoria de eventos acústicos.

1.3 Organização do Trabalho

Esta dissertação está organizada da seguinte maneira:

No **capítulo 2** são apresentados alguns conceitos fundamentais que fazem parte deste estudo para entendimento da pesquisa;

No **capítulo 3** é apresentada a revisão bibliográfica referente ao tema desta dissertação, com foco particular na análise da paisagem acústica.

No **capítulo 4** são apresentados a metodologia e os resultados obtidos na análise das características discriminantes.

No **capítulo 5** são apresentados a metodologia e os resultados obtidos na rotulação de dados acústicos baseada no aprendizado ativo.

No **capítulo 6** apresentam-se as conclusões, contribuições, limitações e sugestões de trabalhos futuros desta abordagem.

CONCEITOS FUNDAMENTAIS

2.1 Considerações Iniciais

Neste capítulo são apresentados conceitos das ferramentas e metodologias utilizadas no processamento de sinais de áudio (Seção 2.2), descritores de características (Seção 2.3), modelos de aprendizagem (Seção 2.4) e técnicas de visualização (Seção 2.5). As definições são dadas como referências, para serem consultadas nos capítulos posteriores.

2.2 Processamento do Sinal de Áudio

Nesta seção são apresentadas conceitos e ferramentas utilizadas na análise e tratamento de sinais e áudio de maneira sucinta junto às referências para seu aprofundamento.

2.2.1 O som

O som é uma característica dinâmica e permanente de um determinado ambiente e está intimamente ligado à descrição desse ambiente (PIJANOWSKI *et al.*, 2011a). O tom do som é ilustrado por meio de onda sinusoidal propagada ao longo do tempo, sendo a frequência (f) a onda sonora medida como o número de repetições de onda por período de tempo (ver Figura 3), geralmente quantificada por segundo e denominada por hertz (Hz). A intensidade é expressa em decibéis (ou dB) (VILLANUEVA-RIVERA *et al.*, 2011a). No trabalho de Everest (2001) encontram-se alguns exemplos de sons e os níveis de dB que produzem, como exemplo, o farfalhar das folhas (20 dB), a conversação de humanos (60 dB), tráfego intenso (80 dB), e foguete Saturno (190 dB). O som detectado por microfones pode ser medido pelas gravações acústicas digitais que armazenam informação do tempo e a intensidade em decibéis.

Figura 3 – Representação da onda sonora: as características entre os picos possuem o mesmo comprimento de onda e amplitude diferente. O comprimento entre os picos de uma onda (designado por λ), e o tamanho da onda é dado pela amplitude de pico. A frequência (f) é derivada do comprimento de onda e da velocidade do som (v), como $f = v/\lambda$.

Fonte: Villanueva-Rivera *et al.* (2011a).

2.2.2 Propagação do Som

Segundo Marten, Quine e Marler (1977) para a propagação de som (ver Figura 4) precisa-se de um trio definido como: Fonte, Propagação (ambiente físico na formação do sinal) e Receptor. No caso da fonte precisa-se das suas características biofísicas e sua intenção de mensagem. A propagação vai depender do meio no qual passa (ar, água, meio sólido, etc.) e de superfícies (vegetação, edifícios, água, etc.). Finalmente, do lado do receptor a percepção do sinal é influenciado pelo seu alcance auditivo.

Figura 4 – Representação básica da propagação do som. a) Fonte (fontes de som), propagação e receptor. b) Fontes de som dentro de uma paisagem acústica (geofonia, biofonia e antropofonia).

Fonte: Pijanowski *et al.* (2011b).

2.2.3 Transformada de Fourier

A transformada de *Fourier - Fourier Transform* (FT), foi denominada assim em referência ao matemático francês *Joseph Fourier* (1768-1830). A FT decompõe uma série temporal em uma soma de séries finitas de funções seno ou cosseno com diferentes frequências, cada uma multiplicada por um coeficiente diferente (Série de Fourier). As funções não periódicas podem ser descritas como uma integral de senos ou cossenos multiplicados por alguma função de ponderação (Transformada de Fourier) (Equação 2.1) (GONZALEZ; WOODS, 2010). Uma função expressada pela Transformada de *Fourier* pode ser reconstruída por meio do processo inverso (Equação 2.2).

$$f(u) = \int_{-\infty}^{\infty} f(x)e^{-j2\pi ux} dx \quad (2.1)$$

$$f(x) = \int_{-\infty}^{\infty} f(u)e^{j2\pi ux} du \quad (2.2)$$

onde, $f(x)$ é o sinal original e $f(u)$ é o resultado da FT (também conhecido como espectro de frequências) sobre um determinado sinal $f(x)$. Para computar a FT, existe um algoritmo que possui um baixo custo computacional, conhecido como *Fast Fourier Transform* (FFT), que calcula de forma rápida a transformada discreta de Fourier (*Discrete Fourier Transform - DFT*) (Wang et al., 2011).

2.2.4 Espectrograma

Um espectrograma é uma representação visual do espectro das frequências em uma amostra sonora. Assim, os espectrogramas são criados com base na transformada de *Fourier*. Na Figura 5, pode-se observar exemplos de espectrogramas, no qual os eixos horizontais representam o tempo, o eixos verticais representam a frequência e a cor representa a amplitude.

Para a criação do espectrograma, inicialmente os dados de som são armazenados como informações expressas como uma onda. Para converter os dados de som para um formato mais útil, a FT é aplicada à onda. Desta maneira o sinal de onda é convertido em níveis de amplitude por frequência. Esses dados, por sua vez, são usados para obter um gráfico de energia por frequência e por tempo chamado de espectrograma, (VILLANUEVA-RIVERA et al., 2011a)

2.3 Descritores de Características

Os descritores servem para extrair características de uma imagem ou dado em particular. Nesta seção são apresentados descritores de áudio e de imagem (descritores utilizados no processamento de imagens).

Figura 5 – Exemplos de espectrograma gravado em uma área úmida florestada no Campus da Universidade de Purdue às 22:00.

Fonte: Villanueva-Rivera *et al.* (2011a).

No âmbito das características do áudio, Mitrovic, Zeppelzauer e Breiteneder (2010) propõem uma maneira de classificar as características do áudio considerando 5 domínios: temporal, frequência, domínio cepstral, frequência de modulação, *eigendomain*, e espaço de fase. Durante o desenvolvimento deste trabalho as fontes de extração de características são representadas nos domínios: temporal, de frequência e cepstral. A seguir é dada uma breve descrição do que significa cada um destes domínios:

- (i) **Domínio Temporal:** São representadas as mudanças de sinal no período do tempo explicitamente em forma de onda. Na literatura, uma forma de algoritmo de extração de características neste domínio é denominada índices acústicos, extraídos diretamente desde o sinal do áudio. A Seção 2.3.1 são descritas alguns índices acústicos utilizados.
- (ii) **Domínio da Frequência:** Neste domínio é considerada a distribuição espectral de um sinal. Nesta forma diversos tipos de análise podem ser executados por faixas de frequência. Para passar do domínio do tempo ao domínio da frequência podem ser empregadas as

transformações mais conhecidas como a transformada de *Fourier* (DFT), transformada de cosseno (DCT) e *Wavelettransform* (MITROVIC; ZEPPELZAUER; BREITENEDER, 2010). Outros tipos de transformações podem ser feitas com banco de filtros passa-banda e filtros com escala Mel, dentre outros. Nesta categoria pode-se considerar alguns índices acústicos e outras características baseados no espectro.

- (iii) **Domínio Cepstral (Cepstrum):** É considerada a distribuição periódica em um espectro de frequências; cepstrum originalmente utilizado para a detecção de ecos das ondas tectônicas provenientes de terremotos (BOGERT, 1963). Uma representação no domínio *cepstral* é definido como a transformada discreta de Fourier (DFT) inversa da magnitude logarítmica da DFT de um sinal. Também é importante saber que uma representação *cepstral* é uma maneira de calcular a aproximação da forma (envelope) do espectro. Neste domínio pode-se considerar as características como *Mel Frequency Cepstral Coefficients* MFCC, *Gammatone Frequency Cepstral Coefficients* GFCC e outros.

2.3.1 Índices Acústicos (Temporal - Frequência)

Um índice acústico é uma estatística que resume alguns aspectos do som, como estrutura e distribuição de energia acústica (TOWSEY *et al.*, 2014a). Os índices descrevem as propriedades acústicas de gravações de áudio do ambiente natural. A maioria dos índices foi desenvolvido com a finalidade de estimar e quantificar riqueza, uniformidade, regularidade, abundância de espécies, etc, para uma paisagem sonora (SUEUR *et al.*, 2014). Além disso, os índices acústicos podem ajudar aos ecologistas em uma série de problemas ecológicos. Alguns índices são descritos a seguir:

- (i) **Acoustic Activity (AA):** considera-se como uma atividade acústica a uma fração de *frames* dentro de um segmento de um minuto, em que o envelope de sinal está a mais de 3dB acima do nível de ruído de fundo (TOWSEY *et al.*, 2014b).
- (ii) **Number of Acoustic Events (NAE):** é determinado pelo número de vezes que o envelope do sinal ultrapassa o limite de 3 dB (TOWSEY *et al.*, 2014b).
- (iii) **Shannon Index (H’):** segundo Villanueva-Rivera *et al.* (2011b) o índice de uniformidade de Shannon (H’) é um dos mais utilizados na avaliação da biodiversidade. É preciso indicar que este índice possui fundamentação na Teoria da Informação para a quantificação da entropia de *string* em textos (Shannon, 1948). Este índice geralmente apresenta bons resultados quando atividade animal é alta e o ruído (tudo aquilo que não é atividade animal) baixo. Este índice é calculado por meio da Equação 2.3:

$$H' = \sum_{i=1}^S p_i \ln p_i \quad (2.3)$$

onde, S representa a um conjunto de espécies, p_i é a proporção de indivíduos encontrados na i th especie.

- (iv) **Average Signal Amplitude (ASA)**: conhecido como amplitude média do sinal, é calculado como a média da amplitude do envelope da onda, depois este valor é convertido em decibéis (TOWSEY *et al.*, 2014b).
- (v) **Temporal Entropy (Ht)**: segundo Sueur *et al.* (2008) é calculado pela aplicação da uniformidade de Shannon, e baseado no envelope de amplitude que obtido com a transformada de Hilbert. Desta maneira buscar visualizar a energia acústica sobre as amplitudes.
- (vi) **Background Noise (BGN)**: o ruído do fundo é calculado a partir do envelope de onda (um minuto de gravação acústica) utilizando o método de Lamel (Lamel *et al.*, 1981), finalmente o valor é dado em decibéis (TOWSEY *et al.*, 2014b).
- (vii) **Singal to Noise Ratio (SNR)**: este índice é calculado dentro de um minuto de áudio pela diferença de decibéis entre a amplitude máxima do envelope e o ruído de fundo, para qualquer segmento de minuto (TOWSEY *et al.*, 2014b).
- (viii) **Spectral Entropy (Hs)**: para calcular a entropia espectral (Hs) precisa-se identificar uma faixa de frequência determinada desde os 482 Hz a 8820 Hz do espectrograma. A Hs é calculado pela aplicação da uniformidade de Shannon (Shannon, 1948) sobre o espectrograma por meio da transformada de Fourier, procurando a concentração de energia sobre as frequências (SUEUR *et al.*, 2009).
- (ix) **Acoustic Entropy (H)**: o índice da entropia acústica resulta da multiplicação da entropia espectral (Hs) pela entropia temporal (Ht). Os valores da entropia acústica podem oscilar entre zero e um ($0 \leq H \leq 1$) (SUEUR *et al.*, 2009).
- (x) **Anthrophony (A)**: mede os níveis de antropofonia na paisagem acústica e é calculado em função da porção de componentes acústicos gerados pelos humanos (KASTEN *et al.*, 2012).
- (xi) **Biophony (B)**: mede os níveis de biofonia na paisagem acústica, calculando a proção de componentes acústicos gerados pelos entes biológicos (KASTEN *et al.*, 2012).
- (xii) **Median Of Amplitude Envelope (M)**: este índice é calculado por meio do cálculo da mediana nos envelopes de amplitude. Assim, os seus valores oscilam entre zero e um ($0 \leq M \leq 1$) (DEPRAETERE *et al.*, 2012). O índice é calculado por meio da Equação 2.4:

$$M = \text{median}(A(t)) \times 2^{(1-\text{depth})} \quad (2.4)$$

onde $A(t)$ é amplitude do envelope, depth é a profundidade da digitalização do sinal (por exemplo 16 bits).

- (xiii) **Mid Band Activity (MBA)**: esta configuração procura a fracção de fragmentos do espectrograma entre os valores 482 Hz e 3500 Hz, onde a amplitude espectral excede 0.015, isto para um minuto de áudio (TOWSEY *et al.*, 2014a).
- (xiv) **Normalized Difference Soundscape Index (NDSI)**: o objetivo deste índice é estimar o nível de distúrbio causado pelos sons do tipo antropofônicos (KASTEN *et al.*, 2012). O NDSI é calculado a partir da proporção entre sons do tipo antropológicos (humanos) e biológicos (especies animais). Os valores do NDSI variam entre $(0 \leq MDSI \leq 1)$, indicando presença de sons antropológicos quando o valor é perto de -1 , presença de sons biológicos para valores perto de $+1$, e presença de sons antropológicos e biológicos ao mesmo tempo, quando o valor esta bem perto de 0. Este índice é calculado por meio da Equação 2.5:

$$MSDI = \frac{(\beta - \alpha)}{(\beta + \alpha)} \quad (2.5)$$

- (xv) **Acoustic Complexity Index (ACI)**: o índice de complexidade acústica (ACI) serve para quantificar direta e rapidamente as atividades acústicas da paisagem acústica (PIERETTI; FARINA; MORRI, 2011). O cálculo de ACI está baseado nas diferenças na variabilidade das intensidades produzida por sons bióticos e sons produzidos pelo homem, indicando-se que os sons bióticos apresentam uma variabilidade intrínseca de intensidades, por outro lado, os sons produzidos pelo homem apresentam valores de intensidade constantes. Então o cálculo de ACI começa por encontrar a diferença absoluta entre os valores adjacentes de intensidades (caixas de frequências) pertencentes a um conjunto de tempo como é ilustrado na Figura 6. No final, o índice ACI é determinado pela média de todas as caixas das frequências (PIERETTI; FARINA; MORRI, 2011). Na figura 6, no eixo do tempo (t), observa-se um conjunto de tempo (j) que esta dividido em n unidades de δt_k , banda de frequência l , e d_k que é a diferença de intensidades entre as frequências adjacentes.

Figura 6 – Divisões das caixas de tempo e frequência para obter o ACI.

Fonte: Pieretti, Farina e Morri (2011).

2.3.2 Características Cepstral

O conceito de Cepstrum foi introduzido originalmente por [Bogert \(1963\)](#) para a detecção de e análises do sinal sísmico. Estas características estão sendo aplicadas no contexto de recuperação e análise de áudio de fala, música e meio ambiente. As características do domínio *cepstral* segundo [Mitrovic, Zeppelzauer e Breiteneder \(2010\)](#) são classificadas em 3 categorias: filtro de banco perceptivo (*Perceptual Filter Bank*) (MFCCs, GFCCs), modelos auditórios avançados, auto-regressão (*Linear Predictive Coefficients* PLP, *Linear Predictive Cepstral Coefficients* LPCCs). Algumas das características *cepstral* são detalhadas a seguir:

- (i) **Mel Scale (MEL)**: a escala de *Mel* foi nomeada por [Stevens, Volkman e Newman \(1937\)](#).

O nome *mel* vem da palavra melodia para indicar que a escala é baseada em comparações de afinação. Uma formulação comumente utilizada para converter de Hertz para Mel é dada por meio da Equação 2.6:

$$F_{mel} = \frac{1000}{\log_{10}} \times \log_{10} \left(1 + \frac{F_{hertz}}{1000} \right) \quad (2.6)$$

onde F_{mel} é a frequência resultante na escala de mel medida em mels e F_{Hertz} é a frequência normal medida em Hz.

- (ii) **Mel Frequency Cepstral Coefficients (MFCC)** os coeficientes MFCC são conhecidos por sua robustez em diferentes tarefas de reconhecimento de padrões referentes à fala ([ON et al., 2006](#)). Estes coeficientes apresentam vantagens como: precisão em tarefas de reconhecimento, taxa de desempenho alta e baixa complexidade. Assim também apresenta desvantagens quando o áudio contém muito ruído de fundo, fornecendo resultados imprecisos. Os coeficientes MFCC geralmente são utilizados para classificação de áudio e recuperação baseada em conteúdo. No entanto verificar a robustez do MFCC em dados de paisagem acústica ecológica ainda é necessário, especificamente na hora de categorizar eventos sonoros existentes dentro da paisagem. Para mais aprofundamento do MFCCs pode-se consultar os trabalhos de [Jaafar, Ramli e Shahrudin \(2013\)](#) e [Jhawar, Nagraj e Mahalakshmi \(2016\)](#).

- (iii) **Gammatone Frequency Cepstral Coefficients (GFCC)** : os filtros Gammatone foram introduzidos por [Aertsen, Johannesma e Hermes \(1980\)](#). Estudos recentes mostraram que os coeficientes GFCC podem fornecer desempenho igual ou melhor que o MFCC devido às características aprimoradas das respostas do filtro, isto usada principalmente na análise de fala para tom de voz. Por outro lado é escasso as evidências dessa eficiência em dados de paisagem acústica ecológica. Os coeficientes GFCC podem ser aprofundados nos trabalhos de [Valero e Alias \(2012\)](#) e [Schluter et al. \(2007\)](#). Também um exemplo da aplicação dos coeficientes GFCC em paisagem acústica pode ser revisado no trabalho de [Agrawal et al. \(2017\)](#)

- (iv) **Linear Predictive Coefficients (LPC)**: os coeficientes LPC são amplamente utilizados em muitos campos do processamento digital de sinais de fala e música por serem considerados robustos para este tipo de tarefa. No caso do LPC, se uma amostra de sinal sonoro for fornecida no tempo n , $([n])$, essa amostra pode ser expressa como uma combinação linear das p amostras passadas dos sinais sonoros (Göğüş; TEZEL, 2018), tal como pode ser observado na Equação 2.7:

$$s[n] = a_1s[n-1] + a_2s[n-2] + \dots + a_ps[n-p] \quad (2.7)$$

onde, $(a_1, a_2, \dots, a_p)$ são os coeficientes de previsão determinados por p , que define o grau LPC.

Uma grande parte dos índices acústicos descritos podem ser computados utilizando as ferramentas como *essentia*¹ que está implementado na linguagem *Python* e o pacote *soundecology*² implementado na linguagem *R*. Estas ferramentas foram incluídas como parte das implementações dos descritores durante o desenvolvimento da pesquisa.

2.3.3 Características da Imagem

As características da imagem com base na sua informação visual, são extraídas por meio de descritores de características. Estes descritores podem ser conceitualizados como algoritmos que extraem vetores de características desde a imagem e podem ser classificados em função ao tipo de informação que extraem, como: cor, textura, forma, etc. (PENATTI; VALLE; TORRES, 2012).

A identificação de um objeto por meio de textura até pode ser trivial quando é realizada por humanos, mas, deixa de ser trivial, quando este trabalho é realizado por algoritmos no computador. A importância de descrever a textura é dada porque permite diferenciar regiões na imagem. Intuitivamente segundo Gonzalez e Woods (2010) esse tipo de descritor fornece medidas de propriedades como suavidade, rugosidade e regularidade. Também a cor é uma das características extraídas que constantemente são utilizadas na literatura de processamento de imagens. As características que descrevem a cor da imagem são muito importantes, pelo fato que ao ser comparadas com a visão humana, pousem um alto poder discriminatório, isto permite diferenciar e identificar as imagens (PENATTI; VALLE; TORRES, 2012). A seguir são apresentados dois descritores de características, um para textura e outro para a cor:

- (i) **Grey Level Co-occurrence Matrix (GLCM)**: a Matriz de Co-ocorrência é uma das técnicas mais conhecidas para analisar a textura numa imagem. Este descritor foi proposto por Haralick (1979), e considera a relação de ocorrência entre as intensidades de pixels

¹ Disponível em <https://essentia.upf.edu>

² Disponível em <https://cran.r-project.org/web/packages/soundecology/>

vizinhos. Este descritor estima características de textura da imagem por meio de coeficientes de *Haralick*, as características resultantes podem ser usadas para classificação e segmentação de imagens.

Uma matriz de co-ocorrência de níveis de cinza $P_{(dx,dy)}(i, j)$ é uma matriz quadrada, que descreve as frequências relativas entre os tons de cinza para cada par de pixels i e j separados por uma distância $d = (dx, dy)$ (SALHI; JAARA; ALAOU, 2016). Matematicamente, a GLCM pode ser definida como:

$$P_{(dx,dy)}(i, j) = \sum_{x=1}^n \sum_{y=1}^m \begin{cases} 1, & \text{se } I(x,y) = i \text{ e } I(x+dx, y+dy) = j \\ 0, & \text{caso contrario} \end{cases} \quad (2.8)$$

onde, $I(x,y)$ representa a intensidade de um determinado pixel da posição x,y . Após, as características de textura são computadas geralmente por 6 medidas ou coeficientes de *Haralick* (HARALICK, 1979; SALHI; JAARA; ALAOU, 2016), estas medidas são: *Homogeneity* (Eq.2.9), *Energy* (Eq.2.10), *Entropy* (Eq.2.11), *Contrast* (Eq.2.12), *Correlation* (Eq.2.13) e *Maximum Probability* (Eq.2.14):

$$f_1 = \sum_{i,j} \frac{P(i, j)}{1 + |i - j|} \quad (2.9)$$

$$f_2 = \sum_{i,j} P(i, j)^2 \quad (2.10)$$

$$f_3 = \sum_{i,j} P(i, j) \cdot \ln[P(i, j)] \quad (2.11)$$

$$f_4 = \sum_{i,j} (i - j)^2 \cdot P(i, j) \quad (2.12)$$

$$f_5 = \sum_{i,j} \frac{(i - \mu_x)(j - \mu_y) \cdot P(i, j)}{\sigma_x \cdot \sigma_y} \quad (2.13)$$

$$f_6 = \max_{i,j} (P(i, j)) \quad (2.14)$$

onde i e j representam a linha e a coluna da matriz, μ_x e μ_y são os valores de médias para x e y , σ_x e σ_y são os valores de desvio padrão para x e y . Complementarmente a Figura 7, ilustra um exemplo de como inserir valores na matriz $P(i, j)$.

- (ii) **BorderInterior Pixel Classification (BIC)**: o descritor BIC (STEHLING; NASCIMENTO; aO, 2002) gera uma representação da distribuição da cor da imagem computando 2 histogramas: um para os pixels classificados como bordas e outro para pixels calculados como interiores. Um pixel é classificado como borda se um dos seus vizinhos tem uma quantização diferente de cor. É classificado como interior no caso contrário. O vetor é computado concatenando-se os 2 histogramas, totalizando $2C$ dimensões. Para uma quantização de cores em 64 cores o descritor final tem uma dimensão de 128 características. Na Figura 8 é possível observar um exemplo para identificar pixels localizados na borda e no interior.

Figura 7 – GLCM: A figura ilustra um exemplo de como inserir valores, normalizar e computar as características na matriz de co-ocorrência $P(i, j)$

Fonte: Löfstedt *et al.* (2019).

Figura 8 – BIC: a figura mostra um exemplo para classificar pixels como bordas (B) ou como interiores (I).

Fonte: Elaborada pelo autor.

2.4 Aprendizado Ativo

O aprendizado ativo - *Active Learning* (AL) (SETTLES, 2009) também é chamado de aprendizagem de consulta. O AL é um tipo especial de aprendizagem semi-supervisionada desenvolvido nas áreas de aprendizado máquina e inteligência artificial, para grandes conjuntos de dados não rotulados. Nos últimos anos este paradigma de aprendizado está sendo aplicado para reduzir o esforço da rotulação em dados acústicos (Shuyang; Heittola; Virtanen, 2017; KHOLGHI *et al.*, 2018; Shuyang; Heittola; Virtanen, 2018) e em outros domínios que possam também ter uma grande quantidade de dados a rotular (TORRES; CATANIA; VEAS, 2019; HALDER; KUMAR, 2019; DEKA; BHATTACHARYYA; KALITA, 2019). Dentro do processo de aprendizado ativo uma das etapas mais importantes é o fornecimento de rótulos a partir de uma pequena amostra para gerar a aprendizagem. Muitas vezes rotular essas pequenas instâncias pode-se tornar muito difícil, demorada ou até computacionalmente cara. A rotulação desse conjunto de amostras é feito por humanos nomeados como usuários especialistas. Na Figura 9 mostra-se o modelo de aprendizado ativo. E para aprofundar a teoria de AL pode-se consultar o trabalho de Settles (2009).

Figura 9 – O paradigma de aprendizado ativo conhecido como piscina ou (pool)

2.5 Projeções Multidimensionais

As técnicas de projeção multidimensional - *Multidimensional Projection* (MDP) são conhecidas também como técnicas de posicionamento de pontos no plano, e são uma solução diferente para a visualização de dados multidimensionais. O objetivo da projeção é representar os itens de dados que podem ter dezenas ou até milhares de dimensões m de entrada em um espaço de menor dimensão p , ($X^m \Rightarrow X^p, p < m$), de modo que, certas propriedades da estrutura do conjunto de dados sejam preservadas o mais fielmente possível. As técnicas de MDP continuam sendo uma boa alternativa para exploração de dados de grande dimensionalidade como foi descrito por [Tejada, Minghim e Nonato \(2003\)](#). Com base na qualidade dos resultados, duas técnicas de projeção que foram empregadas no desenvolvimento deste projeto, e são descritos a seguir:

- (i) ***t-Distributed Stochastic Neighbor Embedding (t-SNE)***: foi proposto por [Maaten e Hinton \(2008\)](#) para conjuntos de dados de alta dimensão. O t-SNE é uma técnica para redução de dimensionalidade que é particularmente adequada para a visualização de conjuntos de dados multidimensionais. A perda de informação é menor do que acontece com outras técnicas como *Principal Component Analysis* (PCA), além disso, o t-SNE não é totalmente determinista, isto porque pode-se estar projetando o mesmo conjunto de dados, mas os resultados podem variar em cada execução do algoritmo. O t-SNE junto outras técnicas como *Least Square Projection* (LSP) ([Paulovich et al., 2008](#)) e *Stochastic Neighbor Embedding* (SNE) ([HINTON; ROWEIS, 2003](#)) estão caracterizados por uma boa preservação de relações de vizinhança refletidas nas projeções. Os autores [Maaten e Hinton \(2008\)](#) testaram o t-SNE com um conjunto de dados com até 30 milhões de exemplos. O t-SNE pode apresentar variantes e podem ser aprofundado nos trabalhos de [Maaten \(2009\)](#), [Maaten](#)

e Hinton (2012), Maaten (2014). A figura 10 ilustra um exemplo de visualização com a técnica t-SNE.

Figura 10 – Exemplo de visualização com t-SNE: os pontos coloridos representam a 6.000 dígitos do conjunto de dados MNIST que contém no total 60.000 imagens em escala de cinza de dígitos em formato de manuscritos.

Fonte: Maaten e Hinton (2008).

- (ii) **Uniform Manifold Approximation and Projection (UMAP)**: UMAP também é uma técnica de redução da dimensionalidade de recente publicação (MCINNES; HEALY, 2018). Esta técnica também pode ser utilizada para visualizar informação multidimensional. O algoritmo do UMAP é baseado na teoria da geometria de *Riemann* e na topologia algébrica. Esta combinação faz que o algoritmo seja escalável e prático para visualizar grandes conjuntos de dados. Segundo o autor McInnes e Healy (2018) por meio do UMAP podem-se aplicar para dados do mundo real, sendo competitivo com o t-SNE, considerando a qualidade da visualização, a preservação da informação, e no tempo de execução. A figura 11 ilustra um exemplo de visualização com o UMAP.

2.5.1 Avaliação das Projeções

Uma das formas para avaliar as projeções é por meio da métrica de *Stress* (KRUSKAL, 1964) e o coeficiente *Silhouette* (S) (ROUSSEEUW, 1987).

O *Stress* que é definida na Equação 2.15 permite avaliar numericamente a qualidade visual das projeções. Os valores do *Stress* indicam, que quanto mais próximo de zero, maior é a

Figura 11 – Exemplo de visualização com UMAP do conjunto de dados de dígitos(0-9) MNIST, de 784 dimensões com 70000 amostras de dados.

Fonte: [McInnes e Healy \(2018\)](#).

preservação de distância entre o espaço original e o espaço projetado, consequentemente isto se traduz como uma melhor visualização da projeção.

$$stress = \frac{\sum_{ij} (d_{ij} - d_{ij}^*)^2}{\sum_{ij} d_{ij}^2} \quad (2.15)$$

onde, d_{ij} e d_{ij}^* são respectivamente as distâncias no espaço original e projetado das instâncias i e j . É desta forma que o *Stress* avalia a preservação de distâncias entre as características do espaço original e no espaço projeto.

O *Silhouette* é definida na Equação 2.16 para avaliar a coesão e a separação entre instâncias agrupadas no espaço visual.

$$S = \frac{1}{p} \sum_i \frac{(b_i - a_i)}{\max\{a_i, b_i\}} \quad (2.16)$$

onde, a_i é a distância média entre y_i (no p espaço visual projetado) e todas as outras instâncias da mesma classe de y_i e b_i é a distância mínima entre y_i e todas as outras instâncias nos outros grupos. S possui valores no intervalo $[-1, 1]$ e, para valores próximos a 1 o valor de S é melhor em termos de coesão e separabilidade.

2.6 Considerações Finais

Neste capítulo foram descritos conceitos e definições úteis para o desenvolvimento do presente trabalho de mestrado. Entre os conceitos fundamentais destaca-se a descrição de 3 categorias de descritores de características que podem ser empregados para extrair características do som. Estas categorias de descritores identificados são: os baseados nos índices acústicos, em características ceptrais e em características de textura e cor extraídas das imagens; neste caso extraídas da imagem dos espectrogramas. Também foi descrito o paradigma de aprendizado ativo, que atualmente é aplicado para tarefas de rotulação de grandes conjuntos de dados. Finalmente, foram descritas as técnicas de projeções multidimensionais (MDP), que na realidade são técnicas de redução da dimensionalidade.

TRABALHOS RELACIONADOS

3.1 Considerações Iniciais

Neste capítulo são descritos os trabalhos que têm relação com a pesquisa. Assim, os trabalhos que fizeram análise da paisagem acústica com base no estudo de diversos descritores são descritos na Seção 3.2. Também um estudo de trabalhos que abordam a metodologia de aprendizado ativo para rotulação de dados eco-acústicos, são descritos na Seção 3.3. Por fim, uma revisão de trabalhos que incluem a visualização em dados acústicos ambientais é apresentada na Seção 3.4. Para concluir é apresentada uma breve discussão como considerações finais sobre os artigos citados no capítulo, na Seção 3.5.

3.2 Análise de Paisagem Acústica

Os trabalhos descritos nesta seção apresentam resultados sobre a utilização dos índices acústicos, métricas acústicas espectrais e imagens do espectrograma como meios para descrever ou sumarizar informações acústicas das gravações. Estas informações permitem a análise das paisagens acústicas, descrevem diferenças entre áreas ambientais acústicas distintas e descrevem alguns eventos acústicos específicos.

Muitos trabalhos relacionados à análise das paisagens acústicas em ecologia estão baseados nos índices acústicos. Dentre de este tipo de trabalhos, pode-se citar o trabalho de [Pieretti, Farina e Morri \(2011\)](#), cujo objetivo é avaliar a capacidade do índice de complexidade acústica (ACI) na quantificação de som biótico (aves) no que se refere à dinâmica diária com relação ao tempo e intervalos de frequência. Além disso, destacaram que o ACI serve para avaliar a filtragem de outros sons não bióticos dentro de uma paisagem acústica.

O trabalho apresentado por [Gasc *et al.* \(2013\)](#) tem como objetivo provar teoricamente

se os índices de diversidade¹ podem ser aplicados para refletir a diversidade filogenética² e funcional³ em comunidades de aves. Os resultados exibiram que os índices de diversidade podem refletir a diversidade filogenética e funcional, os índices aplicados foram: índice de *Shannon* (H'), *Entropia* (H) e *Gini-simpson* (G). Os autores concluíram que o componente espectral da diversidade acústica parece mais apropriado para revelar a diversidade filogenética das aves e o componente temporal parece mais adequado para revelar a diversidade funcional de uma comunidade de aves.

No artigo de [Towsey et al. \(2014b\)](#) o objetivo foi estimar a riqueza de espécies aviárias com mais eficiência (número de espécies identificadas) em comparação de contagens tradicionais. Para tal propósito os autores prepuseram a combinação de diversos índices acústicos que segundo os autores podem produzir mais informações ecologicamente úteis do que índices únicos. A metodologia foi aplicada em dois conjuntos de dados coletados em PAs diferentes. Os conjuntos de dados tinham segmentos de gravação de 1 minuto, contendo todo tipo de eventos que mascaravam o evento de interesse no caso das aves, demonstrando assim que a abordagem pode ser aplicada em dados do contexto real. Também pode-se observar nesse trabalho uma avaliação da robustez de combinação de um conjunto de características na hora de identificar espécies nas gravações, além disso, eles alcançaram vários resultados entre eles a obtenção de um ganho de 87% a mais na identificação de aves do que o jeito tradicional no campo, isto pela combinação de 5 tipos de índices acústicos (*Spectral diversity* ($Sp.div$), ACI , *Entropy of the spectral variance* $H(v)$, *Entropy of spectral maxima* $H(m)$, H).

Outro trabalho baseado na utilização de índices acústicos é apresentado por [Sankupellay et al. \(2015b\)](#). Os objetivos desse trabalho foram conseguir uma representação visual acústica que caracterizasse um ciclo de 24 horas de informação acústica significativa, o qual foi alcançado pela determinação e utilização de 3 índices espectrais para gerar um espectrograma de falsa cor. Outro objetivo foi realizar comparações quantitativas de diferentes ambientes ou áreas de onde foram coletados os dados. Para isso, foram computados 14 índices acústicos dos quais somente 9 índices foram escolhidos por serem mais apropriados para identificar diferenças entre dois ambientes distintos. O trabalho de [Sueur et al. \(2014\)](#) faz uma análise de 21 índices acústicos, para depois fornecer um conjunto de recomendações de quais índices podem ser utilizados na identificação de espécies. Os autores classificaram os índices em duas grandes classes (α e β), que são utilizadas para análise de eventos únicos e para análise entre grupos de eventos, respectivamente. Os índices classificados na classe α foram: *Relative avian abundance*, *Temporal Entropy* (H_t), *Spectral Entropy* (H_f), *Acoustic Entropy Index* (H), *Ratio of biophony to anthrophony* (p), *Acoustic Complexity Index* (ACI), *Biophony* (B), *Biophony peak* ($bioPeak$), *Acoustic Entropy Index* (AEI), *Shannon's Index* (H'), *Acoustic Richness* (AR), *Median of amplitude envelope* (M), *Normalised Difference Soundscape Index* ($NDSI$), *Acoustic Diversity Index* (ADI), *Sound*

¹ Os índices de diversidade denotam riqueza acústica

² Que trata da dispersão das espécies em uma árvore de filogenia

³ É medida por características fisiológicas, morfológicas, ecológicas, comportamentais, etc.

pressure level parameters (L), *Number of peaks (NP)*, *Mid-band activity*, *Entropy of spectral maxima (Hm)*, *Entropy of spectral variance (Hv)*, *Spectral diversity* e *Spectral persistence*. Os índices classificados na classe β formam: *Spectral Dissimilarity (Df)*, *Temporal Dissimilarity (Dt)*, *Acoustic Dissimilarity Index (D)*, *Kolmogorov-Smirnov distance (KS)*, *Kullback-Leibler distance (KL)*, *Vectorial correlation coefficient (RV)* e *Cumulative Dissimilarity (Dcf)*.

No trabalho de [Phillips, Towsey e Roe \(2018\)](#) áudios de um minuto de duração são descritos com base em 12 índices acústicos, os quais são: *Background Noise (BGN)*, *Signal to Noise (SNR)*, *Events per Second (EVN)*, *Activity (ACT)*, *Low-frequency Cover (LFC)*, *Mid-frequency Cover (MFC)*, *High-frequency Cover (HFC)*, *Entropy of the Peaks Spectrum (EPS)*, *Entropy of the Average Spectrum (EAS)*, *Entropy of the Spectrum of Coefficients of Variation (ECV)*, *Acoustic Complexity Index (ACI)* e *Cluster Count*. Os autores tinham como objetivo utilizar os índices para encontrar informações acústicas relevantes dentro de gravações de longo tempo. Para isto os autores empregaram métodos para gerar agrupamentos, e desta forma reduzir a análise de grandes quantidades de dados, e em consequência facilitar as investigações ecológicas. Além disso, 60 agrupamentos acústicos gerados foram codificados por meio de cores para representar dados como meses e anos, desta forma os áudios são facilmente monitorados e visualizados.

Outro meio importante para descrever características de uma paisagem acústica é utilizando técnicas e abordagens da área de processamento de imagens. Neste tipo de abordagens a análise é realizada sobre a imagem do espectrograma. [Xie et al. \(2015\)](#), motivados pela revisão visual de espectrograma, utilizaram também técnicas de processamento de imagem para classificação de eventos de canto de 15 espécies diferentes de sapos. Os autores avaliaram e categorizaram 3 conjuntos de características, e concluíram que as características extraídas da imagem baseadas na detecção de cume seriam as mais adequadas para este tipo de análise. Outro trabalho, que foi apresentado por [Thwe e War \(2017\)](#), visa usar o espectrograma para extrair características de textura por meio do descritor *Local Binary Patter (LBP)*. Com as características os autores realizam a classificação de múltiplas espécies do conjunto de dados conhecido como ESC-10, que é uma variante do ESC-50 ([PICZAK, 2015](#)).

Também foi identificado que o trabalho de [Dias \(2018\)](#) faz a análise de características de paisagem acústica, extraíndo características de áudio utilizando os índices acústicos (ACI, ADI, AEI, AR, BIO, H, Hf, Ht, M, NDSI), junto com características extraídas desde a imagem do espectrograma (*Gray-Level Co-Occurrence Matrix (GLCM)*, *Local Binary Patterns (LBP)*) e cor (*Global Color Histogram (GCH)*, *Color Correlogram (ACC)*, *Color Coherence Vector (CCV)* e *Border/Interior Pixel Classification (BIC)*). Também foi considerada a extração de características empregando um descritor *autoencoder*, mas este não apresentou bons resultados. O objetivo do trabalho foi identificar características capazes de segregar ambientes diferentes de onde foram coletados os dados, no caso foram coletados de 4 áreas distintas, o autor concluiu no caso das características extraídas da imagem, que descritores de cor conseguem distinguir

as áreas denominadas CostaRica1 e CostaRica2, melhor do que as de textura. De outra forma, para o conjunto de dados das áreas denominadas Ilhéus1 e Ilhéus2 não foram separados por nenhum grupo de características extraídas da imagem do espectrograma. Não era esperado que separassem, uma vez que são arquivos descrevendo o mesmo ambiente.

Também foram identificados trabalhos que classificam o som, para subespécies específicas e fazem indiretamente análise de certo conjunto de características. Por exemplo, se podem referenciar alguns trabalhos que classificam subespécies de sapos (XIE *et al.*, 2018) (HAN; MUNIANDY; DAYOU, 2011), (XIE *et al.*, 2016), anuros (NODA; TRAVIESO; SÁNCHEZ-RODRÍGUEZ, 2016) e pássaros (RAGHURAM *et al.*, 2016) (STOWELL; PLUMBLEY, 2014) (Qian *et al.*, 2015).

Existem também as características cepstrais ou *cepstrum*, que são utilizados para descrever os áudios. Dentre os tipos algumas muito conhecidas que são: MFCC, *Linear Predictive Cepstral Coefficients* (LPCC) e GFCC que em sua grande maioria foram aplicados para análise e reconhecimento de fala. Por exemplo, no trabalho de Dias (2018) foi concluído que as características MFCC são muito eficientes na hora de realizar tarefas de segregação de paisagens acústicas. No trabalho de Reis, Santos e Oliveira (2018) é apresentado um *Framework* visual aplicado ao estudo da paisagem acústica para apoiar em tarefas de análise de características. Nos estudos de caso apresentados foram considerados os índices acústicos e as características espectrais MFCC para serem analisados no *Framework*. No final, os autores conseguiram informações e conhecimentos específicos do conjunto de dados acústico explorados.

Outro trabalho de análise de características cepstrais é apresentado por Agrawal *et al.* (2017), onde se faz a análise de características cepstrais como GTCC e MFCC em tarefas de classificação de som ambiental. Para a análise, foram utilizados os conjuntos de dados *ESC-50* (PICZAK, 2015) e *UrbanSond8K* (SALAMON; JACOBY; BELLO, 2014), estes conjuntos possuem instâncias de áudios de 4 e 5 segundos de duração respectivamente. O objetivo principal desse trabalho foi avaliar e comparar a capacidade das características GTCC e MFCC com uma proposta que combina o banco de filtros *Gammatone* modificado com o *Teager Energy Operator* (TEO) nomeado como TEO-GTCC. Nesse trabalho verifico-se a superioridade do TEO-GTCC sobre o MFCC em tarefas de classificação de qualquer dos dois conjuntos de dados já mencionados. Por outro lado pode-se verificar nos resultados a capacidade da combinação de características como TEO-GTCC e MFCC ao classificar o conjunto de dados *ESC-50* com maior acurácia. Os autores concluíram que as características *Gammatone* baseados em TEO contêm informações complementares que são úteis em tarefas de classificação de som ambiental.

Na revisão de trabalhos que fazem o análise de paisagens acústicas, as características são estudadas de forma independente, sendo limitados somente a identificar o tipo de característica na análise. Os trabalhos não fazem um estudo que permita identificar explicitamente que características têm mais relevância nas tarefas de análise de PA e explicitamente na segregação de eventos acústicos de interesse. Também na revisão os trabalhos não incluem estudos que avaliem

as combinações de características (extraídas a partir de sinal e imagem do áudio) e avaliação do desempenho nas tarefas de análise. Uma proposta para identificar as características mais discriminantes no contexto de PA, que incluía a combinações de um conjunto de características para segregar os eventos acústicos é abordado na presente pesquisa.

3.3 Aprendizagem Ativa em Paisagem Acústica

Nesta seção são apresentados alguns trabalhos que fizeram estudos da capacidade do aprendizado ativo (AA) na redução do esforço em tarefas de anotação de dados acústicos, e na verificação dos resultados obtidos em tarefas de predição de etiquetas.

O trabalho de [Han et al. \(2016\)](#) tem como objetivo minimizar o esforço de anotação manual dos usuários para o treinamento de modelo de classificação de som. Nesse sentido, aborda-se uma metodologia baseada na combinação de aprendizado ativo e auto-treinamento com base no nível de confiança das instâncias. Assim, as instâncias com valores baixos de confiança são entregados aos usuários especialistas e as instâncias com valores altos de confiança utilizados na predição automaticamente. Finalmente, os autores afirmam que conseguiram o mesmo desempenho que outras abordagens só que exigindo menor número de instâncias rotuladas.

[Shuyang, Heittola e Virtanen \(2017\)](#) propõem um novo método de aprendizado ativo tendo como objetivo também minimizar os esforços de anotação de instâncias para treinar classificadores sonoros. Utilizando o conjunto de dados *UrbanSound8k*, os autores propuseram gerar agrupamentos por meio de *K-Medoids Clustering*, depois, os *medoids* dos agrupamentos são apresentados aos usuários especialistas para serem rotulados. Após de obter os rótulos, as instâncias são derivadas a um classificador como treinamento supervisionado.

Por outro lado, [Kholghi et al. \(2018\)](#) propõem uma abordagem baseada no aprendizado ativo, que trata sobre a classificação de gravações ecoacústicas de longa duração divididas em áudios de 1 minuto. O objetivo desse trabalho foi investigar a eficácia do AA na redução do esforço de anotação manual necessário para rotular um grande volume de dados acústicos. Segundo a sua metodologia, primeiro foram gerados 60 agrupamentos de todo o conjunto de dados. Depois selecionaram aleatoriamente 10 instâncias de cada grupo para serem rotuladas por usuário especialista, isto permitiu conhecer o nome da classe para cada agrupamento. Nos experimentos foram empregados 12 índices para extrair características do áudio. Os autores concluíram que a abordagem de AA, precisa entre 5 e 16 horas de esforço de anotação manual para criar um classificador que anote automaticamente áudios não identificados.

Ainda, o trabalho de ([Shuyang; Heittola; Virtanen, 2018](#)) propõe um método baseado em AA para controlar um processo de rotulagem de anotação eficiente. Os autores propuseram um AA com 2 estágios. (i) Primeiro geraram agrupamentos por meio de *K-medoids clustering* sobre um conjunto de dados não rotulado, depois os medóides são apresentados a um anotador

para anotação manual do rótulo. A partir dos medóides são propagados os rótulos para outros vizinhos dos agrupamentos. (ii) No segundo estágio propõem uma seleção de amostras iterativas procurando segmentos com rótulos incorretos, assim rótulos corrigidos são atualizadas nos agrupamentos.

A revisão de trabalhos de aprendizagem ativa para análise e rotulação de eventos de paisagem acústica começa na extração de amostras para serem rotuladas de forma manual. Nos trabalhos revisados a seleção de amostras torna-se crucial e geralmente os mais utilizados são os tipos de amostragem: aleatório e por medóide. Neste contexto, a presente pesquisa procura aprimorar o estudo da exploração de novos tipos de amostragens para melhorar a tarefa de rotular eventos da paisagem acústica. Por outro lado a participação do usuário especialista é fundamental para rotular manualmente um conjunto de amostras mais informativas do conjunto total do dados, para isso é necessário incluir estratégias de visualização. Neste contexto, a presente pesquisa busca enriquecer a metodologia proposta com a inclusão de estratégias visuais no modelo de aprendizado ativo.

3.4 Visualização em Paisagens Acústicas

Nesta seção são considerados trabalhos que fazem visualização e exploração de gravações de paisagem acústica.

O trabalho apresentado por (TOWSEY *et al.*, 2014c), tem como objetivo visualizar dados acústicos de longa duração por meio de espectrogramas. Para isso, os autores propuseram empregar a combinação de 3 tipos de índices para gerar um espectrograma de falsa cor. Os índices utilizados foram o ACI, H(t) e *Acoustic Cover* (CVR), que foram mapeados para o 3 canais de cor RGB respectivamente. Um exemplo dos resultados alcançados é apresentado na Figura 12, onde podemos verificar uma maior facilidade em detectar a atividade acústica na imagem de falsa cor, em comparação à imagem original do espectrograma.

Figura 12 – Na direita o espectrograma de falsa cor foi obtido pela combinação dos índices conhecidos como ACI, H(t) e CVR nas cores vermelho, verde e azul, respectivamente.

Fonte: Towsey *et al.* (2014c).

Continuando com a mesma ideia anterior, no trabalho de Sankupellay *et al.* (2015a) combinaram outros índices espectrais (ACI, Ht, EVN) para a geração de espectrograma de falsa cor, conseguindo desta forma visualizar gravações de 24 horas (ver figura 13). Os autores

também conseguiram visualizar as diferenças entre os locais (SERF e Mt Byron) e entre os dias da coleta dos dados, por meio de uma visualização hierárquica. O objetivo das visualizações é comparar ambientes acústicos distintos e facilitar a exploração.

Figura 13 – Visualização de dados acústicos no espectrograma de falsa cor de 24 horas de gravação: a) espectrograma de 24h de gravação gerado pelo audacity. b) espectrograma de 24h com falsa cor para conjunto de dados coletados do SERF. c) espectrograma de 24h com falsa cor para conjuntos de dados coletado do Mt Byron. d) espectrograma com agrupamentos determinados pela técnica *Self-Organising Map SOM* e *Fuzzy c-Mean Clustering*.

Fonte: Sankupellay *et al.* (2015a).

O trabalho proposto por Sánchez-Gendriz e Padovese (2017) busca explorar visualmente grandes quantidades de áudios. Os autores utilizaram técnicas tradicionais como o espectrograma meio e o *Sound Pressure Levels* SPL para explorar visualmente as gravações acústicas de longo tempo. Assim, os autores apresentaram uma nova visualização nomeada SPL-gram, que é ilustrada na figura 14. E por meio das visualizações foram detectados coros de animais ao amanhecer e no anoitecer.

No trabalho de Reis, Santos e Oliveira (2018) é apresentado uma abordagem visual de análise de características para tarefas de exploração e análise de dados de gravações ecológicas. Os autores empregaram as técnicas de visualização como: dendrogramas que é utilizado para representar um modelo de agrupamento hierárquico, também é utilizado o *Heatmap* e o *Radvis* (RV) (Hoffman *et al.*, 1997) para visualizar características. No caso do RV, permite uma interação para explorar a alternância ou combinação das caraterísticas que, ao mesmo tempo permitem

Figura 14 – Impressão digital do ambiente acústico no SPL-gram: linhas indicam a hora do dia, colunas representam os dias estudados e os valores de SPL são mapeados para barras de cores. As linhas ponteadas indicam o nascer e o pôr-do-sol e os pontos marcadores indicam as atividades do coro.

Fonte: Sánchez-Gendriz e Padovese (2017).

visualizar o grau visual da separabilidade ou agrupamento entre as instâncias das classes. Uma das limitações ocorre quando são analisadas uma grande quantidade de características (≥ 200), produzindo sobreposição nas âncoras. A figura 15 ilustra um exemplo de interface do *Framework*.

Figura 15 – A interface do sistema apresenta: uma matriz de correlação no lado esquerdo, também oferece uma visualização *Radviz* de amostras de áudio e âncoras com as quais o usuário pode interagir, também permite a seleção de características com seus respectivos controles localizados no centro da janela. A interface facilita a exploração de dados e análise de características

Fonte: Reis, Santos e Oliveira (2018).

No trabalho de Dias e Minghim (2018) é apresentado o *eXtended Hierarchical Point Placement Strategy* (xHiPP), que é uma extensão da primeira versão do HiPP (Paulovich; Minghim, 2008). O objetivo de xHiPP é melhorar a capacidade de exploração do HiPP e aumentar a sua flexibilidade na hora de analisar dados que não se limitam só para dados acústicos. No trabalho foram apresentados casos de visualização com dados acústicos de 4 lugares distintos, que são nomeados como: *CostaRica1*, *CostaRica2*, *Ilheus* e *Laje*; as duas primeiras foram coletadas de áreas de Costa Rica e o resto foram coletados no Brasil. Os autores conseguiram

demonstrar que o xHiPP consegue visualizar o agrupamento hierárquico exibindo a separabilidade entre grupos, identificando assim a segregação de ambientes acústicos distintos pelo som. A figura 16 ilustra um exemplo do xHiPP.

Figura 16 – Exemplo de projeção xHiPP de conjunto de dados acústico. Cada lugar de coleta é identificado por uma cor como: roxo (*CostaRica1*), cinza (*CostaRica2*), azul claro (*Ilheus*) e marrom claro (*Laje*). Também podem ser visualizados os espectrogramas onde são destacadas algumas áreas indicando padrões sonoros distintos.

Fonte: Dias e Minghim (2018).

O trabalho de Phillips, Towsey e Roe (2018) tem como objetivo revelar o conteúdo acústico de longa duração por meio de técnicas de visualização. Algumas das técnicas utilizadas para visualizar *clusters* nos dados foram o *Radar Plot* e *Rose Plot* que auxiliam na interpretação do áudio ambiental por meio de *clusters* codificados por cores. Os resultados da visualização de paisagem acústica foram apresentados por meio do *Diel Plot* (ver Figura 17) e do Histograma Polar (ver Figura 18) onde conseguiram sumarizar 13 meses de gravação. Este tipo de abordagem pode desempenhar um papel importante no monitoramento de mudanças de longo prazo dentro da paisagem.

3.5 Considerações Finais

Na seção 3.2 pode-se perceber a utilização de índices acústicos como uma das principais fontes na hora de caracterizar uma paisagem acústica (PA), algumas destacando-se mais do que outras para certas tarefas. Outras formas de análise de PA também são por meio de características ceprais, e por características extraídas da imagem do espectrograma. Foi identificado somente um trabalho que faz a análise da PA para segregar cenários e que inclui 3 categorias de características:

Figura 17 – Visualização com *Diel Plot*: nesta visualização encontra-se 13 meses de gravação de áudio. Identificando-se com as linhas amarelas pontilhadas o amanhecer e o pôr do sol da esquerda para a direita. As cores identificam espécies: aves-verdes, ortóptero-amarelas, cigarra-laranjas, vento-azuis claros, e chuva-azuis escuras, silêncio-cinza

Fonte: Phillips, Towsey e Roe (2018).

Figura 18 – Visualização no Histograma Polar: corresponde a 13 meses de gravação. Na legenda da imagem são identificados eventos por alguma cor específica.

Fonte: Phillips, Towsey e Roe (2018).

de imagem do espectrograma, ceprais e índices acústicos (DIAS, 2018). Algumas das abordagens dos trabalhos apresentados somente foram aplicados a conjunto de dados com instâncias de áudio de curta duração, então pode-se inferir que eles não teriam o mesmo resultado com

dados mais reais, como por exemplo instâncias de um minuto de áudio que podem conter ruído mascarando ao evento de interesse. Também foram identificados em alguns trabalhos a avaliação das características segundo sua robustez (capacidade) em distintas tarefas relacionadas a análise da PA, dentre esses trabalhos destaca-se aquele que faz uma análise das combinações de características para reconhecer coros de sapos e outros fenômenos naturais (TOWSEY *et al.*, 2014b). Considera-se pelo revisado até o momento, importante realizar além de uma avaliação da capacidade das características, uma análise e utilização da combinação de um certo conjunto de características para encontrar segregações de categorias de eventos dentro da paisagem acústica ambiental.

Por outro lado, os trabalhos que incluem uma metodologia de aprendizado ativo (AA) para rotulação (seção 3.3), indicam que a AA pode ser um dos métodos mais capazes na redução do esforço pelo usuário na anotação de dados acústicos. Pela revisão foi conferido que os trabalhos não apresentam uma análise nem recomendações em alguns estágios do AA, como por exemplo: o tipo de amostras que o usuário tem que rotular, a análise de *cluster* adequado, modelos de aprendizado, iteratividade e visualização.

Finalmente na seção 3.4 são descritos trabalhos que fazem visualizações em PA. Pode-se perceber que em alguns trabalhos (TOWSEY *et al.*, 2014c; SANKUPELLAY *et al.*, 2015a) que geram um espectrograma de falsa cor baseado numa tripla de índices acústicos espectrais, não fizeram uma análise nem recomendações do porque empregar um determinado conjunto de tripla. O método do xHIPP (DIAS; MINGHIM, 2018) é uma alternativa para fazer análise e exploração, para gerar uma primeira perspectiva dos dados. No caso do *Framework* apresentado por Reis, Santos e Oliveira (2018) é considerado como uma alternativa para análise visual de características, mas para uma quantidade limitada de características. Até o encerramento da revisão bibliográfica, não foram encontrados abordagens de visualização com aprendizado ativo aplicado em paisagens acústicas.

Na Tabela 1 é apresentado um resumo dos trabalhos relacionados. Na tabela podem-se visualizar para cada trabalho informações como: os tipos de características extraídas desde os áudios, o trabalho apresenta um ranking de características, realizam combinação das características, se utilizam técnicas de visualização dos dados como apoio nos seus processos e se incluem o aprendizado ativo para rotulação de dados de paisagens acústicas ecológicas.

Em função à revisão de trabalhos que fazem análise da paisagem acústica, a presente pesquisa visa estudar a formulação de metodologias para segregar eventos acústicos por meio do análise e identificação de características mais discriminantes baseado nas estratégias do aprendizado ativo e as técnicas de visualização como suporte na rotulação de dados de PA.

Tabela 1 – Resumo dos trabalhos relacionados.

Trabalho	Tipo de característica			Ranking de características	Combinação de características	Visualização	Aprendizado ativo
	Índices acústicos	Imagem	Ceptrais				
Pieretti, Farina e Morri (2011)	✓	×	×	×	×	×	×
Gasc <i>et al.</i> (2013)	✓	×	×	×	×	×	×
Towsey <i>et al.</i> (2014b)	✓	×	×	✓	✓	×	×
Sankupellay <i>et al.</i> (2015b)	✓	×	×	×	×	✓	×
Sueur <i>et al.</i> (2014)	✓	×	×	×	×	×	×
Phillips, Towsey e Roe (2018)	✓	×	×	×	×	✓	×
Xie <i>et al.</i> (2015)	×	✓	✓	✓	×	×	×
Dias (2018)	✓	✓	✓	✓	×	✓	×
Agrawal <i>et al.</i> (2017)	×	×	✓	✓	✓	✓	×
Reis, Santos e Oliveira (2018)	✓	×	✓	✓	×	✓	×
Towsey <i>et al.</i> (2014c)	✓	×	×	×	×	✓	×
Sankupellay <i>et al.</i> (2015a)	✓	×	×	×	×	✓	×
Dias e Minghim (2018)	✓	×	✓	×	×	✓	×
Shuyang, Heittola e Virtanen (2017)	×	×	✓	×	×	×	✓
Kholghi <i>et al.</i> (2018)	✓	×	×	×	×	×	✓
Shuyang, Heittola e Virtanen (2018)	×	×	✓	×	×	×	✓

ANALISE VISUAL DE CARACTERÍSTICAS DISCRIMINANTES

4.1 Considerações Iniciais

Como parte do cumprimento do primeiro objetivo ¹ deste trabalho de mestrado, neste capítulo são apresentados todos processos e resultados da análise das características mais discriminantes para uma determinada paisagem acústica. Na seção 4.2 descreve-se a metodologia de análise e determinação das características mais discriminantes. Na seção 4.3 descreve-se o conjunto de dado empregado nos testes da metodologia. Na seção 4.4 são apresentados os resultados e avaliação. Finalmente, na seção 4.5 são apresentadas as considerações finais.

4.2 Metodologia

A figura 19 ilustra as etapas do processo para determinar as características mais discriminantes. No início é definida a etapa de pré-processamento de dados; depois são aplicados descritores de características baseados em imagens, no espectro e no sinal do áudio para extrair características apropriadamente. Em seguida, é definida a etapa de análise das características para identificar a capacidade de discriminação entre categorias de eventos na paisagem acústica. Para atingir este objetivo são empregadas diversas abordagens visuais e numéricas que assistiram na avaliação das características mais discriminantes. Os passos da metodologia são dados no Algoritmo (1). Cada etapa é descrita nas seções seguintes.

¹ Desenvolver uma metodologia para análise e identificação das características mais discriminantes que permita distinguir ou segregar categorias de eventos em paisagens acústicas.

Figura 19 – Metodologia proposta com o esquema para identificar as características discriminantes

Fonte: Elaborada pelo autor.

4.2.1 Pre-processamento dos Dados

Nesta etapa são realizadas 3 tarefas:

- (i) **Leitura de Arquivos:** Cada um dos 2277 arquivos foram convertidos em formato mono canal empregados 40124 (*samples/second*): As principais *libraries* empregadas são *soundfile*² versão 0.10.3, *librosa*³ versão 0.8.0 e *essentia*⁴ versão 2.1, todas em Python.
- (ii) **Criação do Espectro e do Espectrograma:** O espectro foi computado por FFT com a seguinte configuração: ($n_fft = 512$, $hop_length = 2048$, $win_length = 512$). Finalmente o espectrograma é obtido somente com os valores reais do espectro.
- (iii) **Criação da Imagem do Espectrograma:** A partir dos espectrogramas são criadas imagens coloridas de dimensões 100×100 , nas quais as intensidades do espectrograma são normalizadas na faixa de valores reais de $(0 - 1)$, e as cores estabelecidas para os *pixels* são dadas em função do mapeamento de um determinado mapa de cor. Neste trabalho emprega-se os mapas de cores *nipy spectral* e *inferno*, os quais são obtidos pela *library matplotlib*⁵ do Python.

4.2.2 Descrição de Características

Nesta etapa são utilizados diversos descritores de características tendo como fonte diretamente o sinal de áudio, o espectro e a imagem do espectrograma. Para cada instância de áudio, são extraídas as características diretamente da sinal do áudio com descritores que sumarizam a paisagem acústica e que são denominados índices acústicos (ver capítulo 2). As características a partir do espectro são extraídas por meio dos coeficientes cepstrais. Das imagens

² Disponível em: <https://pypi.org/project/SoundFile>

³ Disponível em: <https://librosa.github.io/librosa>

⁴ Disponível em: <https://essentia.upf.edu>

⁵ Disponível em: <https://matplotlib.org/>

dos espectrogramas foram extraídas as características de textura e cor por meio de descritores utilizados em processamento de imagens. A seguir são descritos os 3 tipos de descritores utilizados para a extração de características.

(i) **Descritores baseados nos índices acústicos:** As características são extraídas do sinal do áudio por meio dos sumarizadores matemáticos conhecidos como índices acústicos. Para descrever a paisagem acústica neste trabalho são utilizados 15 tipos índices acústicos: *Average Signal Amplitude (ASA)*, *Background Noise (BGN)*, *Singal to Noise Ratio (SNR)*, *Acoustic Activity (AA)*, *Number of Acoustic Events (NAE)*, *Temporal Entropy (Ht)*, *Spectral Entropy (Hs)*, *Acoustic Entropy (H)*, *Antrhophony (A)*, *Biophony (B)*, *Normalized Difference Soundscape Index (NDSI)*, *Acoustic Complexity Index (ACI)*, *Shannon Index (H')*, *Median Of Amplitude Envelope (M)* e *Mid Band Activity (MBA)*.

(ii) **Descritores baseados no cepstrum:**

São extraídas 59 características no total, as quais correspondem aos seguintes descritores: *Frequency Cepstral Coefficients (MFCC)*, *Gammatone Frequency Cepstral Coefficients (GFCC)*, *Mels-Scaled Spectrogram (MEL)*, *Linear Predictive Coefficients (LPC)*. As configurações e procedimentos para extrair este conjunto de características são as seguintes:

- **características MFCC:** Empregando o espectro são computados os primeiros 16 *Frequency Cepstral Coefficients*.
- **características GFCC:** De forma similar às características MFCC, a partir do espectro são computados os primeiros 16 *Gammatone Frequency Cepstral Coefficients*.
- **características MEL:** Diretamente do espectro são extraídas os primeiros 16 *Mels-Scaled Spectrogram (MEL)*.
- **características LPC:** Diretamente dos *frames* são computadas 11 características do tipo *Linear Predictive Coefficients*.

(iii) **Descritores Baseados na Imagem:** No total 164 características são extraídas da imagem do espectrograma. Os descritores utilizados são *Gray Level Co-ocurrence Matrix (GLCM)*, para a descrição baseada na textura, e *Border Interior Pixel Classification (BIC)* para a descrição baseada na cor. As configurações para extrair este conjunto de características da imagem são como seguem:

- **Características de Textura GLCM** Para descrever a textura nas imagens dos espectrogramas são computadas características do tipo *Gray Level Co-ocurrence Matrix*. Primeiro a imagem colorida é convertida para escalas de cinza. Depois são computadas 6 matrizes de co-ocorrência do descritor com os seguintes deslocamentos $d_{(xy)} = \{(0, 1), (0, 3), (0, 5), (1, 0), (3, 0), (5, 0)\}$. As características de textura são obtidas computando para cada matriz de co-ocorrência as 6 medidas de *Haralik: Energy*,

Entropy, *Contrast*, *Correlation*, *Homogeneity* e *Maximun Probability*. No final, o total de características extraídas são 36.

- **Características de Cor com BIC** Empregando as imagens coloridas do espectrograma são extraídas as características do tipo *Border Interior Pixel Classification*. Nesse caso as cores das imagens foram quantizadas em 64 cores. Assim foram computados dois vetores de características para classificar as cores dos pixels, pixels localizados nas bordas e pixels localizados no interior das regiões. O total de características extraídas são 128.

Na extração de características são processadas no total 238 características para 2277 instâncias de áudio, o conjunto de características é representado como X^m , onde X denota o número de instâncias e m o número de características por instância)

4.2.3 Análise de Características Discriminantes

Para esta etapa são executadas as seguintes tarefas.

- (i) **Análise visual das características:** Para analisar as características existem diversas abordagens. Uma análise visual combinada com numérica utilizando: histogramas com 10 *bins*, *boxplots* e *scatterplots*, para uma análise das características com relação a sua distribuição, variabilidade, e correlação, respectivamente. Essas abordagens são detalhadas a seguir:

- **Histogramas:** Colaboram na identificação visual da distribuição e comportamento de cada categoria de características. Segundo [Faceli et al. \(2011\)](#), para conhecer a distribuição dos valores dos dados num histograma é importante associar ao valor da obliquidade. De acordo com essa informação, em histogramas com obliquidade igual a zero a distribuição é denotada como normal ou simétrica, e se a obliquidade é maior ou menor que zero a distribuição seria denotada como positiva ou negativa respectivamente.
- **Boxplot:** Segundo [Haemer \(1948\)](#) os *Boxplots* são úteis para visualizar a variação das características por meio de quartis, sendo também úteis para expressar medidas centrais como mediana, média, modo e desvio padrão.
- **Scatterplots:** É basicamente um meio visual para observar a associação entre pares de variáveis ou atributos, permitindo observar se existe uma correlação ou dispersão entre as duas variáveis ([Bachthaler; Weiskopf, 2008](#)). Alguns resultados que pode-se identificar neste tipo de gráfico é correlação positiva ou correlação negativa.

- (ii) **Limpeza de características:** De forma similar como é feito no tratamento dos dados em ciência de dados, nesta tarefa são eliminadas as características que têm baixa variabilidade ou valores constantes, como aquelas características que têm todos os valores fixados em

zero. Depois da tarefa de limpeza, espera-se uma considerável redução no número de características.

- (iii) **Normalização de características:** Nesta tarefa os dados são normalizados utilizando os métodos conhecidos como *Min-Max* (Equação 4.1) e *Z-Score* (Equação 4.2). O objetivo de normalizar é transformar todas as variáveis no mesmo intervalo de variação.

$$X_{changed} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.1)$$

onde, X é o valor para normalizar, X_{min} e X_{max} , são o valor mínimo e máximo de um conjunto de valores da característica.

$$Z = \frac{x - \mu}{\sigma} \quad (4.2)$$

onde, x é um valor para normalizar, μ e σ são a média e o desvio padrão de um conjunto de valores de uma característica.

- (iv) **Seleção das características:** Esta tarefa tem como objetivo empregar métodos de seleção de características para abordar o problema da grande dimensionalidade das características e ao mesmo tempo para influenciar positivamente no desempenho de procedimentos dedicados a identificar segregação de eventos.

O objetivo da seleção de características é estabelecer um subconjunto distinto de características relevantes ou discriminantes que influencie no alcance de desempenho do modelo de aprendizado máquina. A seleção de características pode ser denotada como $X^m \Rightarrow X^p$, onde m denota o conjunto original de características e p denota do subconjunto ótimo de características, tal que $p \leq m$.

Segundo Miao e Niu (2016) os métodos de seleção de características podem ser categorizados em *filter*, *wrapper* e *embedded*. Nos métodos da categoria *filter* as características são selecionadas com base em suas pontuações em vários testes estatísticos para serem ordenadas de acordo a sua importância do conjunto de características. Para métodos de seleção de características da categoria *wrappers*, um subconjunto de características são selecionados para depois serem utilizadas para treinar num modelo. Com base na inferência do modelo, podemos tomar a decisão de adicionar ou remover características do subconjunto. Não é recomendado que esse método seja usado para um alto número de características, já que tem um custo computacional caro. E finalmente o método *embutido* nomeado também incorporado por ser um método de seleção variável. Este método resulta da combinação dos métodos *filter* e *wrapper*, onde a seleção das características é realizada internamente pelo próprio algoritmo de aprendizado, que dizer, que a tarefa de seleção está embutida no algoritmo de aprendizado máquina.

Na presente pesquisa o método de seleção de características empregado é o *embedded Extra Trees Classifier* (ETC) por ser uma solução intermediária. Este método apresenta risco

menor de sobre-ajuste em comparação com os demais métodos. Ao empregar o método espera-se como saída um ranking de importância das características.

Identificação das características mais discriminantes

Nesta tarefa é empregado o ranking da importância das características processadas na tarefa anterior de seleção de características. Para identificar as características mais discriminantes são computadas para as n primeiras combinações de características tarefas de classificação (aprendizagem supervisionada). As instâncias são divididas em treino e teste, sendo empregada a validação cruzada $5-k-fold$ para Treinamento e sendo a percentagem de instâncias de Teste empregadas para a validação final do aprendizado. Os modelos de classificação utilizados foram: *Random Forest (RF)* (BREIMAN, 2001), *K-Nearest Neighbor Classifier (KNNC)* (Fukunaga; Narendra, 1975), *Support Vector Classifier (SVC)* (PLATT, 1999) e *Extreme Gradient Boosting Classifier (XGBC)* (CHEN; GUESTRIN, 2016). As características mais discriminantes são determinadas pelos resultados de maior acurácia da classificação das n combinações de recursos.

- (v) **Avaliação:** Conhecendo que as características serão utilizadas em tarefas classificação, a avaliação da capacidade discriminatória das características será realizada empregando a métrica *Accuracy (AC)*. Esta métrica computa a relação do número de instâncias corretamente classificadas (ou preditas) sobre o número total de instâncias do conjunto de dados. Esta métrica de avaliação está formalmente definida na Equação 4.3.

$$AC = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.3)$$

onde, TP são os verdadeiros positivos, TN são os verdadeiros negativos, FP são os falsos positivos e FN são os falsos negativos.

4.2.4 Visualização

Nesta etapa são empregadas projeções multidimensionais para evidenciar visualmente a qualidade de discriminação das características discriminantes identificadas na coesão ou separabilidade de classes. Desta forma, o conjunto de instâncias e características X^p são transformadas para 2 dimensões $p = 2$. As técnicas de projeção empregadas foram: *t-SNE* (MAATEN; HINTON, 2008) e *UMAP* (MCINNES; HEALY, 2018). Espera-se que ao empregar as características mais discriminantes para gerar as projeções multidimensionais poderá ao mesmo tempo evidenciar a qualidade das visualizações, em relação à aproximação de objetos similares no layout. Para avaliar a qualidade da visualização das projeções é empregada a métrica de *Stress* que foi definida na equação 2.15 e o *Silhouette* que foi definida na Equação 2.15

—formula silhouette—

Algoritmo 1 – Algoritmo de análise visual características discriminantes

Input:

data: conjunto rotulado de arquivos de som de uma paisagem acústica.
model_embedded: modelo ETC para determinar a importância das características.
model_learn: modelo de aprendizagem ex. *RFC*, *SVC*, *KNNC* ou *XBGC*.

Output:

$X_{discriminat}$: características mais discriminante para uma determinada paisagem acústica.

```

1:  $Y \leftarrow \text{Labels}(data)$  //rótulos dos arquivos de som
2:  $repeat = True$ 
3: enquanto  $repeat == True$  faça
    // Stage (1): pre-processing data
4:  $\hat{X} \leftarrow \text{PreprocessinData}(data)$  //pre-processamento dos dados
    // Stage (2): feature description
5:  $X_{index} \leftarrow \text{IndexFetureDescription}(\hat{X})$  //características de índices acústicos
6:  $X_{spectrum} \leftarrow \text{SpectrumFetureDescription}(\hat{X})$  //características do espectro
7:  $X_{image} \leftarrow \text{ImageFetureDescription}(\hat{X})$  //características da imagem do espectrograma
8:  $X_{raw} \leftarrow (X_{index} \cup X_{spectrum} \cup X_{image})$  //total de características
    // Stage (3): discriminant feature analysis
9:  $X_{cleaned} \leftarrow \text{FeatureCleaning}(X_{raw})$  //limpeza das características
10:  $X_{normalized} \leftarrow \text{FeatureNormalization}(X_{cleaned})$  //normalização das características
11:  $X_{analyzed} \leftarrow \text{FeatureAnalysis}(X_{normalized})$  //análise das características
12:  $X_{important} \leftarrow \text{FeatureSelection}(model\_embedded, X_{analyzed})$  //importância das
    características
13:  $X_{ranking} \leftarrow \text{FeatureRanking}(X_{important})$  //ranking das características
14:  $X_{discriminant} \leftarrow [\emptyset]$  //para armazenar as características mais discriminantes
15:  $max\_acc \leftarrow 0$  //armazena a acurácia máxima
16: for each  $i \in \{0, \dots, |X_{ranking}|\}$  faça //para cada características do ranking
17:    $X \leftarrow (X_{ranking}^0, \dots, X_{ranking}^i)$  //obter as n primeiras características do ranking
18:    $acc \leftarrow \text{Learning}(model\_learn, X, Y)$  //treinar um modelo e computar acurácia do teste (acc)
19:   se  $acc > max\_acc$  então //computar o máximo valor de acurácia
20:      $X_{discriminant} \leftarrow X$ 
21:      $max\_acc \leftarrow acc$ 
22:   fim se
23: fim para
    // Stage (4): visualization
24:  $view \leftarrow \text{Projection}(X_{discriminant})$  //visualizar os dados
25:  $repeat \leftarrow \text{UserVisualEvaluation}(view)$  //o usuário avalia a repetição
26: fim enquanto

```

4.3 Conjunto de Dados

Para os experimentos deste trabalho foi considerado o conjunto de dados fornecido por nosso parceiro de pesquisa, denominado Cantareira-Mantiqueira (CM3).

O estudo foi realizado no Continuum Ecológico Cantareira-Mantiqueira, localizado entre o sul de Minas Gerais e o nordeste de São Paulo (Figura 20). A área de coleta se caracteriza por estar entre duas montanhas (Cantareira-Mantiqueira) altamente arborizadas, em consequência uma região com uma alta prioridade de conservação da Mata Atlântica (BOSCOLO *et al.*, 2017). Este conjunto de dados foi cedido pelo professor Milton Ribeiro do Laboratório de Ecologia Espacial e Conservação (LEEC) da Universidade Estadual Paulista (UNESP). Este conjunto consta com mais de 40000 arquivos de áudio de um minuto cada. Todos os arquivos possuem o formato WAVEform *audio format* (WAV). Um subconjunto é selecionado pelos parceiros para serem rotulados manualmente por especialistas. Segundo a rotulação, cada minuto de áudio foi etiquetado em três grandes categorias de eventos como: Anuros, pássaros e Insetos. Para os testes foram utilizados 615 instâncias para a classe anuros, 822 instâncias para a classe pássaros e 840 instâncias para a classe insetos, completando um total de 2277 instâncias de áudio.

Figura 20 – Área de coleta de dados. Realizado no Continuum Ecológico Cantareira-Mantiqueira, localizado entre o sul de Minas Gerais e o nordeste de São Paulo. A área é composta por pântanos, água, estradas, aldeias e principalmente por florestas (BARROS *et al.*, 2019)

4.4 Resultados e Avaliação

Dado o conjunto de dados CM3 descrito na Seção 4.3, inicialmente as características são obtidas seguindo as duas primeiras etapas da metodologia, pré-processamento de dados e descrição de características. Assim foram obtidas 238 características e depois foram processadas as tarefas de análise visual, limpeza e normalização de dados da etapa de análise discriminante de características. Na tarefa de análise visual das características foram examinadas a variação e

distribuição das características como pode-se observar na Figura 21 que ilustra alguns exemplos empregados para esta tarefa de análise visual. Depois da análise visual das características a decisão foi a realizar tarefas limpeza, para isso foram eliminadas as características constantes e com variabilidade baixa, empregando um desvio padrão (std) inferior ou igual a 0.015. Assim as características foram reduzidas de 238 para 102 características. Consecutivamente as características foram normalizadas para cada conjunto de dado empregando os métodos de Mix-Max e Z-Score separadamente.

Figura 21 – Um exemplo de análise visual das características do conjunto de dados DS1 (tabela 2) : no lado esquerdo (a) *boxplots* permite visualizar a variação das características com relação a os eventos do conjunto de dados, nos eixos (x = características, y = valores de características), e no lado direito (b) histogramas permite visualizar a distribuição das características, no caso índices acústicos, nos eixos (x = valores de características, y = frequência ou repetição das características).

Fonte: Elaborada pelo autor.

Nesta seção a avaliação das características são realizadas por meio de tarefas de classificação (aprendizagem supervisionado). Para ampliar a análise de características também é incluso a classificação binária. Para isto, é definido 3 subconjuntos que resultam da combinação dupla entre as instâncias das classes: anuros, pássaros e insetos. Desta maneira, para os próximos experimentos são empregados no total 4 conjuntos de dados, são eles: (DS1) anuros, pássaros e insetos, com 2277 instâncias; (DS2) anuros e pássaros, com 1437 instâncias, (DS3) anuros e insetos, com 1455 instâncias; e (DS4) pássaros e insetos, com 1662 instâncias. Esta nova configuração de conjunto de dados é resumida na Tabela 2. Nas seguintes seções são descritos os experimentos para completar o análises da identificação das características mais discriminantes, os experimentos são parte da etapa de análise de características discriminantes, e também da etapa de visualização.

Tabela 2 – Definição e descrição dos conjuntos de dados

Conjunto de dados	Classes	Arquivos por classe	Total de arquivos
DS1	Anuros	615	2277
	Pássaros	822	
	Insetos	840	
DS2	Anuros	615	1437
	Pássaros	822	
DS3	Anuros	615	1455
	Insetos	840	
DS4	Pássaros	822	1662
	Insetos	840	

4.4.1 Ranking de Características

Para cada conjunto de dados e seguindo a metodologia da etapa de análises de características discriminantes, a técnica de seleção de características (*embedded*), *Extra Trees Classifier* (ETC), foi empregada para obter a importância (valores) das características. Em seguida, para obter o *ranking*, os valores de importância das características foram ordenados de forma decrescente (de maior para menor). Os resultados do ranking de características são apresentados na Figura 22. Nos rankings podemos observar, além da ordem de importância das características, a sua categorização, sendo a cor laranja para características baseadas em imagens, roxo para características baseadas no espectro e verde para características baseadas nos índices acústicos.

Neste experimento pode ser notado que as características do descritor de cor BIC não apresentam maior importância em comparação às características do tipo textura GLCM. Isso pode acontecer porque o mapa de cor não providenciaria informação relevante para descrever aos espectrogramas no caso tratado, ou pela quantização de cores. Por outro lado os descritores de textura são mais importantes e podem descrever melhor a informação dos espectrogramas. Ao mesmo tempo as características de textura destacam-se como as mais importantes do que a maioria das 128 características de cor do tipo BIC.

Dentro das 10 melhores características pode-se notar que as características do tipo espectrais como MFCC e MEL junto com algumas de textura do tipo GLCM são as primeiras no *ranking*. Considerando o conjunto de características de índices acústicos, as mais importantes características são: *Temporal Entropy* (Ht), *Spectral Entropy* (Hs), *Acoustic Entropy* (H), *Acoustic Complexity Index* (ACI) e *Shannon Index* (H'). Esses índices (características) geralmente estão localizados nas 5 primeiras posições no *ranking*.

4.4.2 Seleção das Características mais Discriminantes

Considerando o *ranking* da importância de características em cada conjunto de dados, foi avaliado o poder discriminatório para as n primeiras características. Neste experimento, em cada

Figura 22 – Ranking de características: de cima para baixo são apresentados resultados de ranking de características para os conjuntos de dados DS1 (anuros, pássaros e insetos), DS2 (anuros e pássaros), DS3 (anuros e insetos) e DS4 (pássaros e insetos). As cores no *ranking* indicam o tipo de característica segundo a nossa categorização a cor: roxo (descritores de índices acústicos), verde (características cepstrais) e laranja (descritores de imagens).

Fonte: Elaborada pelo autor.

combinação de características, a acurácia nos subconjuntos de treino e teste foram computadas usando respectivamente a estratégia *5-k-fold cross-validation* e o 10% de instâncias por classe. Na Figura 23 são apresentados os resultados de acurácia média para os subconjunto de teste de 10% com características normalizadas usando Z-Score. Também foram computados resultados com as características originais e normalizadas com o método *Min-Max*, no entanto, com a normalização Z-Score foram obtidos os melhores resultados. Os resultados foram obtidos para 4 modelos de aprendizagem: *Random Forest Classifier* (RF), *K-Nearest Neighbor classifier* (KNNC), *Support Vector Classifier* (SVC) e *Extreme-Gradient Boosting Classifier* (XGBC).

Neste experimento, para o conjunto de dados DS1, os melhores resultados foram obtidos empregando o modelo de aprendizagem *Random Forest Classifier* (RFC) com 30 características. Nos resultados pode-se evidenciar geralmente que o modelo RF é o mais destacado e constante com seus resultados nos conjuntos de dados tratados por esse motivo, podemos inferir que é o modelo mais adequado de aprendizagem de dados acústicos. Por outro lado, os demais modelos

Figura 23 – Melhores características: de cima para baixo são apresentados resultados de acurácia média para as n melhores características dos conjuntos de dados: a) DS1 (anuros, pássaros e insetos) com 30 características com 89,91% de acurácia, b) DS2 (anuros e pássaros) com 30 características com 82,64% de acurácia, c) DS3 (anuros e insetos) com 46 características com 100,0% de acurácia e d) DS4 (pássaros e insetos) com 31 características com 99,40% de acurácia. As linhas verticais indicam para cada modelo o máximo valor de acurácia média atingida. Na legenda de cada resultado estão os 4 modelos de aprendizagem utilizados

Fonte: Elaborada pelo autor.

apresentam resultados similares ao do RFC como é no caso do DS3 e DS4. Também podem-se observar que resultados com valores mais alto de acurácia apresentados no DS3 pelo modelo SVC é igual ao obtido por RFC com a diferença no numero de características. Assim também podem-se observar que resultados com valor mais alto de acurácia apresentados no DS4 pelo modelo KNNC é igual ao obtido por RFC com a diferença no numero de características. Por meio desses resultados identifica-se o conjunto de características mais discriminantes para segregação de categorias de eventos sonoros tratados em cada um dos conjuntos de dados.

Para uma análise mais detalhada da qualidade das características, os resultados de uma extensão do experimento anterior são apresentados na Tabela 3. Os resultados correspondem ao mesmo procedimento para os 4 conjuntos de dados com acurácia de classificação no Treino e no Teste com suas respectivas melhores n primeiras características. Neste experimento é considerada a variação da percentagem (10%, 20%, 30%, 40% e 50%) na conformação das instâncias como Teste (validação). O restante das instâncias empregadas para a tarefa de Treino, onde também é utilizado *5-k-fold cross-validation* como estratégia para validar a aprendizagem, os resultados do *test* dos *5-k-fold* encontram-se na coluna do *Train* na tabela 3. Na maioria dos resultados de acurácia no Teste podemos comprovar ainda a superioridade do modelo de classificação RFC. Nas variações de percentagens de Teste, podemos observar também a preservação dos resultados de acurácia, isto valida a qualidade das n primeiras características selecionadas do ranking.

Tabela 3 – Resultados da classificação: são apresentados resultados de acurácia para os 4 conjunto de dados junto como os 4 modelos de aprendizagem e para variações de percentagens para o teste (validação). Também são apresentados o número de características (.) que foram precisas para obter a acurácia.

Dataset	Model	Test 10%		Test 20%		Test 30%		Test 40%		Test 50%	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
DS1 Frogs Birds Insects (2277)	RFC	86.43 (38) ±0.017	89.91 (30)	85.94 (48) ±0.008	87.28 (32)	85.12 (57) ±0.012	85.09(59)	84.48 (36) ±0.017	84.74(51)	82.60(45) ±0.017	83.41 (59)
	KNNC	81.89(26) ±0.018	84.65(42)	83.14(23) ±0.011	83.11(19)	81.48(29) ±0.016	82.16(21)	81.99(26) ±0.019	80.90(15)	81.02(35) ±0.019	81.30(32)
	SVC	85.55(54) ±0.011	87.28(39)	85.12(27) ±0.015	85.53(22)	84.62(52) ±0.016	85.53 (44)	83.75(59) ±0.031	84.52(57)	84.19 (52) ±0.022	83.14(49)
	XGBC	84.77(76) ±0.007	86.84(52)	84.51(54) ±0.003	84.65(89)	83.24(66) ±0.009	85.38(76)	83.02(60) ±0.021	84.85 (67)	82.78(52) ±0.018	82.62(34)
DS2 Frogs Birds (1437)	RFC	81.59 (46) ±0.022	82.64 (30)	81.03 (55) ±0.012	82.64 (37)	81.00 (47) ±0.020	82.41 (74)	78.76(42) ±0.025	81.91 (83)	78.68 (31) ±0.029	81.36 (48)
	KNNC	76.64(28) ±0.015	79.86(45)	75.80(26) ±0.030	77.43(33)	76.72(26) ±0.048	75.69(32)	74.59(39) ±0.013	77.91(21)	74.51(43) ±0.022	78.03(48)
	SVC	80.82(47) ±0.017	80.56(13)	79.20(22) ±0.019	80.56(46)	79.90(57) ±0.029	80.09(37)	78.77 (45) ±0.017	80.70(30)	78.13(56) ±0.015	79.69(25)
	XGBC	81.05(46) ±0.020	81.25(38)	79.72(41) ±0.017	78.47(49)	79.40(47) ±0.012	78.94(49)	78.30(47) ±0.020	79.65(59)	77.58(42) ±0.020	79.00(77)
DS3 Frogs Insects (1455)	RFC	98.32(41) ±0.005	100.00 (46)	98.28 (55) ±0.013	99.31 (32)	97.54(41) ±0.018	99.31 (67)	97.25(51) ±0.017	99.14 (53)	96.83(42) ±0.014	96.70(27)
	KNNC	96.79(37) ±0.011	97.26(11)	97.34(12) ±0.008	98.28(13)	96.17(31) ±0.013	95.65(39)	96.11(26) ±0.017	96.74(13)	96.29(26) ±0.011	95.05(17)
	SVC	98.40 (38) ±0.009	100.00 (23)	98.11(27) ±0.006	98.63(28)	98.13 (44) ±0.015	97.25(39)	98.17 (48) ±0.004	97.59(28)	97.11 (53) ±0.019	97.39 (44)
	XGBC	97.33(44) ±0.008	98.63(18)	97.59(60) ±0.008	98.28(61)	96.56(64) ±0.021	97.03(22)	96.68(51) ±0.007	97.25(26)	96.84(53) ±0.003	96.15(54)
DS4 Birds Insects (1662)	RFC	98.60(52) ±0.004	98.80(31)	98.42(62) ±0.008	98.80(34)	98.19 (40) ±0.005	99.00(77)	98.19(24) ±0.004	98.50(33)	98.07(45) ±0.013	98.44(78)
	KNNC	98.60(16) ±0.007	99.40 (14)	98.19(18) ±0.006	99.10 (18)	98.02(22) ±0.011	99.20 (17)	98.30 (23) ±0.007	98.95 (20)	97.71(24) ±0.005	98.07(16)
	SVC	98.66 (27) ±0.008	98.80(28)	98.50 (31) ±0.004	99.10 (25)	98.19 (58) ±0.010	98.20(41)	98.29(38) ±0.007	98.65(45)	98.43 (45) ±0.012	98.56 (28)
	XGBC	97.93(56) ±0.007	98.20(41)	98.19(73) ±0.007	98.80(51)	98.19 (41) ±0.006	98.40(48)	98.30 (46) ±0.007	98.50(51)	98.08(27) ±0.006	97.95(61)

4.4.3 Visualização Empregando Características mais Discriminantes

Depois de identificar as melhores características foram empregadas técnicas de projeção multidimensional para visualizar o conjunto de dados. Dessa forma também pretende-se avaliar visualmente a qualidade das características. Quando a visualização apresenta maior separabilidade pode-se inferir que as características utilizadas nessa visualização são as mais discriminantes. A Figura 24 mostra os resultados das visualizações empregando as técnicas de projeções multidimensionais t-SNE e UMAP. Os resultados são apresentados comparando as visualizações quando são empregadas as características mais discriminantes (identificadas no estágio anterior) e as 238 características originais. Nos gráficos os pontos representam instâncias do conjunto de arquivos de som, sendo a cor azul para instâncias da classe insetos, rosa para anuros e verde para instâncias da classe pássaros.

Visualmente, os melhores resultados foram obtidos com o t-SNE, onde pode ser observada uma maior separabilidade das classes. Por outro lado, quando é empregado UMAP os resultados visuais não são favoráveis; isto pode ser em razão de empregar somente 2277 instâncias, sendo que segundo a literatura afirma que melhores resultados são obtidos ao empregar uma maior quantidade de instâncias por classe (MCINNES; HEALY, 2018). Os resultados de *Stress* e *Silhouette* das projeções também validam a qualidade das visualizações. Para o *Stress* valores próximos a zero mostram que as visualizações apresentam maior preservação de distância entre o espaço original e o espaço projetado, isto pode evidenciar a qualidade da visualização utilizando as características mais discriminantes. Para o *Silhouette* (S) valores próximos a um (1) mostram que as visualizações têm boa qualidade em termos de coesão e separabilidade das classes, os resultados evidenciam a capacidades das características mais discriminantes na qualidade da visualização. A partir dos resultados, ao mesmo tempo, podemos afirmar que uma boa visualização é produzida a partir das melhores características extraídas e selecionadas.

4.5 Considerações Finais

Neste capítulo foi apresentada uma metodologia para identificar as características mais discriminantes para o conjunto de CM3. Os resultados atingidos com as características identificadas foram de 89.91% para a classificação de 3 classes. Para ampliar a análise nos experimentos, nós incluímos a classificação binária entre as instâncias de duas classes, além de variações de sub-conjuntos de treino e teste. No caso da classificação binária os resultados de acurácia atingiram 82.64%, 100.0% e 99.40% para os pares anuros-pássaros, anuros-insetos e pássaros-insetos, respectivamente.

A análise nos permitiu identificar como conjunto de características mais discriminantes: Temporal Entropy (Ht), Spectral Entropy (Hs), Acoustic Entropy (H), Acoustic Complexity Index (ACI) and the Shannon Index (H'), além, das características MFCC, MEL e GLCM, que também denotaram importância quando foram empregadas em tarefas de classificação supervisionada. As

características identificadas por meio da metodologia foram empregadas no próximo estágio da pesquisa para fins de apoio na rotulação de eventos acústicos, que é apresentado no Capítulo 5.

Figura 24 – Visualizações dos 4 conjuntos de dados: da esquerda para direita resultados para os conjuntos de dados DS1, DS2, DS3 e DS4. Para cada conjunto projeções t-SNE e UMAP com 30°, 30°, 46° e 41° primeiras melhores características (1° e 2° linha). Na 3° e 4° linha projeções t-SNE e UMAP com todas as 238 características originais.

Fonte: Elaborada pelo autor.

VISUAL ACTIVE LEARNING PARA ROTULAÇÃO

5.1 Considerações Iniciais

Pela grande quantidade de dados de paisagens acústica, a necessidade de rotulá-los é cada vez mais notória e relevante na hora de realizar tarefas de análise. Uma dessas tarefas é a segregação de eventos. Idealmente, nesse cenário, as tarefas de rotulação destas grandes quantidades de dados poderiam ser realizadas a partir de um número reduzido de instâncias. Como parte do cumprimento do segundo objetivo ¹ do presente trabalho de mestrado, neste capítulo é apresentada uma metodologia para rotular dados de paisagens acústica, incluindo o paradigma de aprendizado ativo com base na visualização. Assim, na seção 5.2 são descritas as etapas da metodologia para rotular dados de paisagens acústicas. Na seção 5.3 são apresentados os resultados e avaliação. Finalmente, na seção 5.4 são apresentadas as considerações finais.

5.2 Metodologia

A Figura 25 ilustra as etapas da metodologia estabelecida para a rotulação de dados de paisagens acústica. Dado um conjunto de dados que não apresenta rótulos, inicialmente é executada a etapa *Clustering*. Em seguida foram extraídas amostras dos agrupamentos na etapa nomeada como *Sampling*. Os usuários especialistas interagem com as amostras escutando e rotulando os áudios na etapa *Listening-annotation*. Considerando as amostras rotuladas pelos especialistas, em seguida foi treinado um modelo de aprendizagem (classificador), o qual, na sequência, foi utilizado para prever o restante dos rótulos das instâncias do conjunto de dados, excluindo as amostras. Essas tarefas foram executadas na etapa *Learning-prediction*. Finalmente

¹ Definir e criar uma metodologia para rotular categorias de eventos acústicos baseado no paradigma de aprendizado ativo e visualização com projeções multidimensionais

os resultados obtidos com a metodologia foram avaliados. Os passos das etapas da metodologia são apresentados no Algoritmo (2) e são descritos a seguir.

Figura 25 – Metodologia proposta para rotular dados da paisagem acústica com base no aprendizado ativo.

Fonte: Elaborada pelo autor.

5.2.1 Clustering

Para inicializar esta etapa foi necessário um conjunto de dados de paisagens acústicas com instâncias não rotuladas e que possuam características previamente extraídas. Assim, nesta etapa foram formados k grupos de instâncias do conjunto de dados. Para a executar esta etapa foram empregados os seguintes algoritmos de agrupamento: *Hierarchical Agglomerative Clustering* e *K-Means*, ambos implementados na *library* Scikit-learn². Espera-se que por meio dos agrupamentos se permita identificar padrões entre as instâncias de áudios, padrões que permitam a segregação das categorias de eventos em paisagens acústicas.

5.2.2 Sampling

O principal objetivo desta etapa foi extrair amostras mais representativas dos grupos formados, para depois serem empregadas em tarefa futura de aprendizagem. Assim, nesta etapa foram extraídas p amostras de cada um dos agrupamentos. Para isto foram propostos como métodos de extração de amostras: aleatório ou *random* (r), medóide ou *medoid* (m) e contorno ou *contour* (c); além das suas combinações (rm, rc, mc e rmc). Para extrair amostras do tipo *random* foram extraídas amostras aleatoriamente. No caso das amostras do tipo medóide, as amostras foram extraídas em relação às instâncias mais próximas dos centroides dos agrupamentos. No caso de amostras do tipo *contour*, as amostras foram extraídas entre as instâncias mais afastadas dos centroides dos agrupamentos. A Figura 26 mostra os 3 tipos de amostras a serem empregadas.

² Disponível em scikit-learn: <https://scikit-learn.org/>

Figura 26 – Exemplo visual dos tipos de amostras consideradas. As amostras do tipo aleatório foram extraídas de maneira arbitrária. As amostras do tipo medóide foram extraídas em relação às instâncias mais próximas do centroides dos agrupamentos. Finalmente as amostras do tipo contorno, foram extraídas em relação as instâncias mais afastadas dos centroides dos agrupamentos. C1 representa o grupo 1 e C2 representa o grupo 2.

Fonte: Elaborada pelo autor.

5.2.3 Listening-annotation

Nesta etapa foi iniciada a indução da aprendizagem do paradigma de aprendizado ativo por meio da interação de usuários especialistas. O objetivo desta etapa foi abstrair a informação dos usuários especialistas em reconhecer e diferenciar as categorias dos som. Assim, esta etapa trata das tarefas de escutar e rotular arquivos de áudios correspondentes a amostra mais representativa. Para executar estas tarefas foram empregadas projeções multidimensionais, as quais possibilitariam ao usuário a interação e visualização entre as amostras como pode-se observar no item a da Figura 29. Ao mesmo tempo, as projeções podem facilitar a visualização dos agrupamentos previamente gerados entre as instâncias.

De modo a auxiliar a interação do usuário com as projeções, foram empregadas também visualizações de espectrogramas na forma de linha de tempo, denominada como *Time-Line-Spectrogram*. O objetivo deste tipo de visualização foi prover maior informação visual no momento que o usuário interage com as projeções, nas tarefas de escutar e rotular instâncias de áudio.

A interação do usuário com as projeções em coordenação com as visualização dos espectrogramas na forma de linha de tempo, durante as tarefas de escutar e rotular instâncias, foi uma nova proposta de aprendizado ativo visual, sendo o objetivo incrementar a aprendizagem. É desta maneira os usuários especialistas podem rotular as instâncias de amostras mais representativas do conjunto de dados.

5.2.4 Learning-Prediction

Esta etapa tem como objetivo treinar um modelo de aprendizado a partir das características e dos rótulos das amostras. Essa aprendizagem foi utilizada para a predição de rótulos das demais instâncias do conjunto de dados. Assim, para dar continuidade ao processo de aprendizagem ativa, as instâncias de amostras previamente rotuladas foram utilizadas para treinar um modelo, depois por meio desta aprendizagem foram rotulados a maior parte das instâncias do conjunto de dados. Para o cumprimento desta etapa foi preciso executar as tarefas a seguir:

- (i) *Learning*: Um modelo de aprendizado foi treinado utilizando as características e os rótulos das instâncias consideradas como amostras. Neste caso, o modelo empregado foi o *Random Forest Classifier* (RFC).
- (ii) *Prediction*: Após gerada a aprendizagem, o modelo já treinado foi utilizado para prever os rótulos das demais instâncias do conjunto de dados (isto é, desconsiderando as amostras). Em função aos resultados das visualizações e por critério do especialista as etapas da metodologia podem ser repetidas, desde a etapa de *Clustering* como podemos observar no item a da Figura 29.

Como pode ser observado na Figura 25 existe uma zeta que indica iteração desde o estágio *Learning-Prediction* a *Clustering*, nesse cenário a iteração na metodologia serve para analisar e determinar os parâmetros mais adequados segundo os valores de acurácia alcançada ao realizar a rotulação automática .

5.2.5 Validation

Para validar a metodologia foi preciso empregar um conjunto de dados que possua rótulos verdadeiros, foi verificar quantas as instâncias dos áudios foram rotuladas como seriam por usuários especialistas. Dessa forma a validação nesta etapa da metodologia consiste em comparar os rótulos verdadeiros com os rótulos que foram determinados na etapa *Learning-prediction*. Para validar o grau de predição foi utilizada a métrica que avalia a acurácia de classificação *Accuracy* (AC) definida na Equação 4.3 no Capítulo 4.

5.3 Resultados e Avaliação

Para a execução dos experimentos: foram empregados os conjuntos de dados DS1, DS2, DS3 e DS4 já definidos na Seção 4.3 do Capítulo 4. As características dos conjuntos de dados foram computadas por meio da metodologia descrita no Capítulo 4. A predição dos rótulos dos conjuntos de dados foi computada executando todas as etapas da metodologia descrita na Seção 5.2, as quais são: *Clustering*, *Sampling*, *Listening-annotation* e *Learning-prediction*. Nesse cenário é importante destacar que a interação do usuário na etapa *Listening-annotation* é simulada

Algoritmo 2 – Algoritmo de aprendizado ativo visual ou *visual active learning*

Input:

data: conjunto não rotulado de arquivos de som de uma paisagem acústica.
model_clust: método de agrupamento ex: *KM* ou *HAC*.
k: número de agrupamentos.
model_smp: tipo de amostragem ex. *r*, *m*, *c*, *rm*, *rc*, *mc* ou *rmc*.
size_smp: número de amostras por agrupamento.
model_learn: modelo de aprendizagem ex. *RFC*, *SVC*, *KNNC* ou *XBGC*.

Output:

$\widehat{data}$: dados rotulados, onde os rótulos representam as categorias de eventos segregados.

```

1: repeat == True
2: enquanto user_decision = True faça
3:   X ← Fetures(data)                                //descrição das características
4:   Y ← Labels(data)                                  //para referenciar aos rótulos
   // Etapa (1)
5:   clusters ← Clustering(model_clust, X, k)            //fazer o agrupamento
   // Etapa (2)
6:   samples ← Sampling(clusters, model_smp, size_smp)    //fazer a amostragem
7:   unsamples ← (X – samples)                          //obter as instâncias que não são amostras
   // Etapa (3)
8:   view_p ← Projection(X, samples)                    //visualizar os conjunto de dados
9:   view_s ← TimeLineSpectrogram(X, samples)           //visualizar os conjunto de dados
10:  Y_samples ← ListenedAnotatted(view_p, view_s,)      //escutar e rotular as amostrar
   // Etapa (4)
11:  Learning(model_learn, X_samples, Y_samples)          //treinar o modelo
12:  y_unsamples ← Prediction(model_learn, X_unsamples)   //estimar os rótulos
13:  UpdadeLabels(view, Y_unsamples)                     //atualizar os rótulos na projeção
14:  repeat ← UserVisualEvaluation(view)                 //o usuário avalia a repetição
15: fim enquanto

```

utilizando os rótulos verdadeiros unicamente para as amostras, garantindo aprendizagem no contexto dos experimentos.

Na etapa de *Learning-prediction* o classificador utilizado foi o *Random Forest* (RFC) por mostrar-se robusto em tarefas de classificação em dados de paisagens acústica, como verificado nos experimentos relatados no Capítulo 4. Os experimentos realizados para analisar as etapas da metodologia são descritos nas seções a seguir.

5.3.1 Análise de Clustering e Sampling

Com o intuito de explorar e conhecer as melhores configurações dos parâmetros da metodologia, o objetivo deste experimento foi avaliar as etapas de *Clustering* e *Sampling*, gerando subsídios para avaliar as subsequentes etapas da metodologia. Assim, as análises das

melhores configurações de parâmetros providenciaram informações relevantes que contribuem ao incremento da acurácia na predição dos rótulos.

Neste experimento foram avaliados: o número de agrupamentos (k), o número total de amostras (p), o número de amostras por agrupamento (pk) e os métodos para extrair amostras (r , m , c , rm , rc , mc e rmc). A configuração do experimento foi definida como segue: foram utilizados os 4 conjunto de dados DS1, DS2, DS3 e DS4. Com características extraídas utilizando a metodologia para extrair as n primeiras melhores características do Capítulo 4. O número de agrupamentos foi analisado na faixa de $k = \{5, 6, \dots, 49, 50\}$. O número de amostras por agrupamento foi estabelecido considerando o menor número possível de instâncias por agrupamento, desta forma o número de amostras por agrupamento foi fixado em $pk = \{5, 10, \sqrt{|\cdot|}\}$, onde $|\cdot|$ representa ao total de instâncias por agrupamento.

Os resultados para os 4 conjuntos de dados são apresentados, respectivamente, nas tabelas *heatmap* 4, 5, 6 e 7, cor amarelo indica aos máximos valores de acurácia alcançada (predição de rótulos), cor azul indica os mínimos valores, os valores intermediários de acurácia são visualizados com uma gradiente de cores entre as cores amarela, laranja e azul. Nos experimentos, a metodologia de *visual active learning*, foi executada no total 3864 vezes. Nas tabelas, cada linha de resultados com os 7 tipos de amostras (r , m , c , rm , rc , mc , rmc) são apresentadas na forma visual de *heatmaps*, onde os valores máximos de acurácia são visualizados com a cor amarela, os valores mínimos de acurácia são visualizados com a cor azul claro e os valores intermediários de acurácia são visualizados com uma gradiente de cores entre as cores amarela, laranja e azul claro.

Inicialmente os agrupamentos foram computados empregando os algoritmos *K-Means* (KM) e *Hierarchical Agglomerative Clustering* (HAC), mas, os melhores resultados foram obtidos com HAC para agrupamentos aproximadamente maiores a 20 (grupos). Em relação às amostras, pode-se observar que em quanto maior seja o número de amostras a acurácia tende a aumentar, porém, idealmente espera-se utilizar poucas amostras para depois predizer a maior parte das instâncias. No entanto, empregar poucas amostras não permite obter altos resultados de acurácia, por outro lado, o desafio é utilizar o menor número de amostra possível e obter resultados acurados. Neste cenário a partir dos resultados observados nas tabelas *heatmap* podemos inferir que a proposta de fixar o número amostras por agrupamento em $\sqrt{|\cdot|}$ é a melhor opção dentre as testadas porque seleciona uma porcentagem ou quantidade de amostras adequada ao numero de elementos por cada agrupamento. Em relação aos métodos para extrair amostras, a análise procura descobrir e definir quais dessas são as mais informativas e geram uma acurácia maior na hora de predizer rótulos. Depois de fazer uma análise visual da informação das tabelas por meio dos *heatmaps*, pode-se notar que os métodos mais adequados para extrair amostras em ordem (de melhor para pior) foram: r , rc , rm , mc , rmc , m , e c .

Para validar a qualidade das n primeiras melhores características e também as melhores configurações de parâmetros, apresentadas nas tabelas 4, 5, 6 e 7, foi executado de novo o

experimento anterior, mas, empregando para este caso, todas as 102 características. Assim, nas tabelas 8, 9, 10 e 11 são apresentados os resultados de acurácia obtidos, os quais são muito parecidos aos resultados obtidos quando foram empregadas as n primeiras melhores características. Porém, de modo a fazer uma comparação mais clara e específica, a Figura 27 ilustra a comparação dos resultados entre as n primeiras características e as 102 características, considerando especificamente: os 2 melhores métodos de amostras (r e rc) e o pior método de amostra (c) com 10 amostras por agrupamento. Nos resultados apresentados na Figura 27 pode-se verificar a superioridade das amostras r e rc para os 4 conjuntos de dados. Desta forma, por meio dos experimentos no contexto de aprendizado ativo, verificou-se a capacidade discriminatória das características em função de uma configuração desejável de parâmetros.

Tabela 4 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS1 com as n primeiras melhores características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc). No *heatmap*, o cor amarelo indica aos máximos valores de acurácia alcançada (predição de rótulos), o cor azul indica aos mínimos valores, os valores intermediários de acurácia são visualizados com uma gradiente de cores entre as cores amarela, laranja e azul.

k	pk = 5									pk = 10									pk = $\sqrt{ \cdot }$								
	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)
5	25	1.10	62.79	58.30	32.50	59.64	47.38	58.53	62.79	50	2.20	64.66	64.12	31.88	60.66	58.69	53.84	58.15	103	4.52	67.25	62.79	34.64	66.65	62.93	55.24	57.41
6	30	1.32	65.24	62.44	31.51	60.21	43.21	59.28	60.17	60	2.64	64.10	63.78	33.60	62.07	58.05	52.64	58.37	112	4.92	64.43	63.74	40.05	67.48	69.24	55.66	57.14
7	35	1.54	71.19	61.46	38.85	67.53	59.28	62.36	62.62	70	3.07	71.95	60.67	37.20	66.38	65.43	57.59	63.25	119	5.23	72.85	63.30	37.26	68.40	67.98	55.28	61.03
8	40	1.76	65.62	61.42	37.82	65.53	59.10	64.68	64.33	80	3.51	67.96	62.90	32.13	64.54	67.05	60.31	61.72	126	5.53	67.55	63.69	39.01	68.62	68.06	61.13	64.53
9	45	1.98	60.26	64.61	40.14	62.46	64.25	66.71	66.58	90	3.95	67.08	64.61	33.74	68.54	68.22	65.11	65.57	136	5.97	69.87	64.27	41.71	71.18	63.80	64.83	66.14
10	50	2.20	63.36	64.89	43.65	62.73	66.91	67.85	67.49	100	4.39	70.56	64.17	40.19	66.65	68.08	67.98	65.55	142	6.24	71.52	65.43	48.01	73.40	71.05	68.99	68.29
11	55	2.42	64.13	62.15	46.98	65.84	67.96	66.61	65.98	110	4.83	68.76	63.54	43.79	68.94	70.14	67.74	65.39	149	6.54	73.50	64.29	43.70	71.95	69.41	69.08	66.17
12	60	2.64	70.55	60.62	45.20	67.34	68.20	63.78	65.31	120	5.27	75.15	62.87	46.13	64.72	71.12	67.45	64.86	155	6.81	76.48	64.89	49.62	69.32	72.20	67.86	66.21
13	65	2.85	67.18	61.35	44.71	66.59	60.67	64.87	67.09	130	5.71	68.70	62.55	50.30	67.54	68.37	68.70	65.16	160	7.03	70.00	64.38	60.23	68.02	70.15	69.25	66.98
14	70	3.07	69.87	60.17	58.13	67.20	66.38	65.84	65.29	140	6.15	69.91	62.56	60.60	68.79	68.69	68.65	64.25	166	7.29	73.14	62.72	59.02	66.51	76.69	70.82	67.69
15	75	3.29	68.94	61.22	54.13	66.17	65.67	63.40	63.90	150	6.59	71.98	59.71	58.44	68.97	71.37	69.44	65.26	171	7.51	72.98	62.68	58.97	70.85	73.12	69.94	67.57
16	80	3.51	66.64	61.36	56.94	66.18	67.59	64.54	63.31	160	7.03	70.81	62.21	57.96	67.45	73.50	69.49	67.60	177	7.77	70.48	64.38	60.86	71.10	75.24	69.95	69.14
17	85	3.73	67.38	62.55	61.82	69.21	68.48	67.47	65.88	170	7.47	71.95	62.70	59.80	68.15	71.86	71.29	67.68	180	7.91	73.72	63.61	57.51	71.01	73.06	71.44	69.19
18	90	3.95	66.94	62.87	59.35	68.27	67.86	68.77	64.29	180	7.91	72.96	63.09	60.75	71.63	72.10	71.39	68.43	187	8.21	73.25	64.31	62.54	67.13	72.87	70.29	68.04
19	95	4.17	69.43	62.74	61.18	67.51	68.65	68.70	66.68	190	8.34	73.84	63.44	61.04	70.96	73.74	71.01	68.28	190	8.34	74.89	65.26	62.48	71.35	72.40	70.29	69.38
20	100	4.39	70.56	63.30	57.83	67.98	63.99	69.13	68.26	200	8.78	72.85	64.52	63.31	72.22	73.57	73.18	72.32	196	8.61	72.80	64.44	62.09	70.98	74.20	72.03	70.40
21	105	4.61	67.50	63.49	61.60	66.44	69.11	69.20	66.85	210	9.22	72.57	64.25	61.44	71.02	71.46	73.39	72.18	202	8.87	70.80	64.00	58.31	69.54	73.78	72.96	71.28
22	110	4.83	69.59	62.67	64.28	67.10	68.90	68.44	67.01	220	9.66	76.52	63.93	62.08	71.66	75.79	71.85	69.42	205	9.00	73.94	64.43	57.77	72.39	75.87	71.43	69.88
23	115	5.05	69.29	61.89	62.86	65.54	70.40	68.13	65.82	230	10.10	72.74	62.77	61.65	71.18	76.80	72.79	71.52	208	9.13	72.21	64.52	58.05	71.53	73.66	71.58	71.00
24	120	5.27	70.47	62.31	63.65	67.22	72.60	68.47	66.34	240	10.54	72.46	63.48	65.44	71.18	74.37	73.10	71.28	213	9.35	74.13	64.58	63.81	71.32	75.53	72.29	70.69
25	125	5.49	71.28	63.48	64.92	64.78	69.56	68.77	67.80	250	10.98	74.49	65.02	66.01	70.35	76.76	72.72	71.88	216	9.49	73.70	64.68	60.89	69.00	76.71	72.68	72.97
26	130	5.71	71.50	64.28	62.23	69.21	71.63	70.42	67.26	260	11.42	73.13	66.83	64.95	70.70	75.31	74.62	73.33	220	9.66	73.94	66.80	61.84	69.52	75.55	73.65	71.46
27	135	5.93	71.99	64.43	61.06	67.04	69.33	69.09	66.29	270	11.86	73.39	66.72	65.62	70.30	76.23	73.59	71.10	227	9.97	72.59	65.66	61.41	72.73	74.54	73.07	71.41
28	140	6.15	71.31	64.16	60.93	69.35	69.07	70.47	67.01	280	12.30	73.46	66.80	64.15	73.56	74.71	73.86	71.71	230	10.10	72.50	66.05	62.29	72.69	74.50	73.82	70.35
29	145	6.37	73.26	64.59	58.72	65.95	71.62	70.73	67.68	290	12.74	75.74	66.53	64.47	73.28	76.60	73.73	71.97	235	10.32	75.37	64.84	62.10	69.59	75.81	72.97	70.23
30	150	6.59	73.25	64.27	64.17	67.47	72.21	69.25	68.36	300	13.18	75.37	66.26	66.87	70.97	77.14	74.00	73.88	237	10.41	73.48	66.96	63.43	71.27	74.85	71.76	69.56
31	155	6.81	72.01	63.81	64.14	69.93	74.46	70.45	68.94	310	13.61	72.70	67.16	66.50	71.73	75.19	73.97	73.84	241	10.58	73.04	66.70	65.28	72.59	73.04	73.43	71.41
32	160	7.03	73.36	64.67	66.23	69.44	71.99	70.29	68.63	320	14.05	77.11	67.25	68.73	72.15	77.72	73.89	72.84	245	10.76	75.84	67.47	64.22	73.13	76.28	73.38	72.59
33	165	7.25	72.02	65.34	65.53	66.76	74.76	70.74	69.74	330	14.49	74.47	67.28	67.95	72.37	77.40	75.30	74.65	249	10.94	74.51	66.96	67.50	71.65	72.68	73.82	73.67
34	170	7.47	72.95	65.54	67.44	70.72	76.70	70.00	70.53	340	14.93	77.59	66.91	68.97	71.66	75.32	75.12	74.57	254	11.16	75.43	66.58	69.25	72.96	78.25	72.66	72.86
35	175	7.69	74.41	65.08	67.98	72.36	72.84	69.36	69.89	350	15.37	76.49	68.19	71.56	74.26	78.31	75.30	74.13	257	11.29	74.60	67.67	70.25	74.01	75.54	74.50	71.63
36	180	7.91	72.53	64.43	68.14	72.25	71.20	70.15	69.19	360	15.81	76.21	66.82	71.36	72.98	77.78	75.59	75.14	260	11.42	73.67	67.48	71.00	70.90	77.94	73.28	70.80
37	185	8.12	75.05	66.68	66.30	72.28	75.00	71.94	70.98	370	16.25	77.24	69.01	70.79	74.04	78.66	75.62	74.99	262	11.51	75.88	68.44	73.15	73.10	75.93	73.40	72.16
38	190	8.34	75.13	64.64	66.94	71.63	73.26	71.73	70.92	380	16.69	76.70	68.90	72.32	75.28	76.07	75.12	75.18	264	11.59	74.42	68.21	71.98	75.56	74.96	72.78	72.48
39	195	8.56	73.15	64.94	66.71	74.02	75.22	71.85	71.66	390	17.13	77.16	68.73	70.48	74.03	76.31	74.46	74.31	268	11.77	75.96	68.19	70.68	72.62	77.20	72.32	71.38
40	200	8.78	73.95	64.28	67.60	75.16	75.16	71.98	71.27	399	17.52	79.07	68.00	70.02	73.48	76.89	74.81	73.54	271	11.90	77.52	66.15	71.44	75.62	74.63	72.78	72.38
41	205	9.00	71.72	64.29	67.37	71.09	76.11	72.10	70.24	409	17.96	77.46	67.93	71.15	75.54	76.28	74.20	74.57	274	12.03	75.09	67.30	70.00	72.74	74.84	73.34	71.24
42	210	9.22	71.60	64.97	65.99	70.39	76.15	72.09	70.16	419	18.40	76.64	69.05	69.54	71.91	78.53	74.87	74.49	277	12.17	74.35	67.95	68.75	72.60	77.60	75.30	71.85
43	215	9.44	75.46	66.25	66.88	72.41	75.56	71.92	69.95	429	18.84	78.35	68.72	69.05	72.73	78.90	74.57	73.76	280	12.30	74.81	67.80	70.36	74.01	77.52	73.86	70.56
44	220	9.66	75.21	66.65	67.87	72.78	73.36	71.91																			

Tabela 5 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS2 com as n primeiras melhores características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc).

k	pk = 5									pk = 10									pk = √ ·								
	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)
5	25	1.74	56.66	65.44	49.01	63.53	55.10	59.49	61.40	50	3.48	67.12	66.62	57.03	65.90	63.88	63.73	64.46	79	5.50	64.58	68.19	63.40	69.37	68.48	65.61	65.17
6	30	2.09	69.72	66.67	53.66	63.26	67.31	64.32	65.67	60	4.18	70.52	70.44	57.88	69.06	64.71	63.40	66.74	88	6.12	70.72	68.20	57.15	70.42	67.98	67.68	68.72
7	35	2.44	62.62	68.97	63.77	68.83	65.62	62.41	69.33	70	4.87	67.23	70.23	63.86	68.32	67.08	69.20	68.47	94	6.54	70.96	71.03	65.97	72.08	64.48	70.66	70.44
8	40	2.78	69.15	67.50	61.42	66.93	63.64	65.93	70.65	80	5.57	69.64	69.79	66.32	70.82	71.78	69.64	68.75	99	6.89	71.60	71.38	65.40	70.25	71.45	69.51	70.70
9	45	3.13	66.02	66.74	48.78	71.70	63.94	63.00	66.74	90	6.26	72.83	71.34	65.40	71.64	66.22	68.30	68.15	104	7.24	70.82	70.67	67.07	69.84	65.64	69.92	72.24
10	50	3.48	68.85	69.36	48.38	66.62	70.01	63.30	68.28	100	6.96	65.74	70.01	64.47	71.58	65.00	67.69	66.34	109	7.59	70.48	69.20	66.94	71.91	64.23	68.45	68.75
11	55	3.83	63.89	67.08	59.12	66.43	65.70	63.17	67.29	110	7.65	68.27	70.84	64.13	69.56	62.17	66.62	67.45	115	8.00	66.41	69.97	64.07	67.55	68.76	68.76	69.21
12	60	4.18	67.39	67.90	67.68	68.63	68.41	66.52	66.96	120	8.35	67.96	70.46	65.83	72.44	70.84	67.20	67.27	120	8.35	67.58	70.62	69.63	68.41	71.15	69.02	68.64
13	65	4.52	70.04	67.27	64.72	66.76	69.17	69.31	67.78	130	9.05	71.61	68.55	66.18	68.48	71.84	69.09	68.71	125	8.70	72.48	69.74	67.23	69.36	72.18	69.89	70.50
14	70	4.87	69.50	66.72	65.03	68.40	66.50	68.69	65.54	140	9.74	71.40	67.62	65.07	70.24	70.24	69.08	69.08	128	8.91	71.20	68.37	68.30	68.37	69.29	69.21	69.06
15	75	5.22	69.38	67.91	66.74	69.46	66.67	67.55	66.23	150	10.44	69.77	67.13	68.76	71.33	70.86	68.07	68.69	131	9.12	69.30	68.91	68.53	70.14	69.91	68.91	69.68
16	80	5.57	70.67	65.95	68.02	69.64	71.41	65.66	63.96	160	11.13	72.20	66.41	67.97	70.71	70.01	68.75	67.74	138	9.60	72.44	67.51	69.75	69.59	70.90	69.36	71.75
17	85	5.92	72.63	67.01	66.72	67.23	70.71	66.86	67.68	170	11.83	71.90	69.06	68.11	69.53	70.17	69.22	70.56	143	9.95	72.64	67.08	69.17	69.63	68.32	69.55	71.41
18	90	6.26	70.53	67.56	67.41	70.82	70.53	69.49	66.30	180	12.53	71.12	66.75	70.01	71.84	72.16	68.58	70.96	146	10.16	71.96	67.39	69.95	71.96	70.72	70.26	72.35
19	95	6.61	68.26	68.11	67.59	67.66	68.26	71.39	67.36	190	13.22	69.61	66.96	69.45	72.33	71.37	69.21	70.25	151	10.51	68.58	68.74	71.85	68.04	74.49	70.76	70.76
20	100	6.96	70.98	69.56	64.77	69.56	67.09	70.68	68.66	200	13.92	72.84	66.53	69.36	70.01	70.49	71.38	71.54	155	10.79	73.87	70.83	69.50	70.90	68.49	72.78	72.62
21	105	7.31	67.94	70.80	66.29	71.25	70.05	71.92	70.12	210	14.61	70.99	71.39	70.09	73.19	70.82	71.23	72.62	158	11.00	72.09	72.24	69.19	73.18	72.63	69.82	72.09
22	110	7.65	73.02	71.51	66.54	69.86	69.63	70.61	70.54	220	15.31	76.01	70.09	70.50	74.45	73.21	70.75	72.14	162	11.27	73.33	72.16	70.75	72.63	73.10	70.43	72.08
23	115	8.00	71.10	71.10	65.43	69.67	71.10	71.41	69.21	230	16.01	74.07	72.58	69.43	74.90	72.08	73.82	71.67	166	11.55	72.07	71.60	71.05	72.78	70.73	71.13	73.41
24	120	8.35	71.07	71.98	66.59	72.59	72.36	72.06	69.78	240	16.70	73.60	72.93	70.59	74.44	72.51	72.68	72.10	169	11.76	72.40	71.21	72.24	70.98	71.85	70.43	73.03
25	125	8.70	72.79	70.20	67.99	70.88	70.88	71.95	70.12	250	17.40	73.13	72.96	70.43	71.44	72.37	70.94	71.26	171	11.90	73.62	73.46	69.67	74.33	72.35	70.14	72.83
26	130	9.05	72.46	70.16	65.65	71.23	71.46	70.54	70.70	260	18.09	72.90	72.56	70.26	74.34	75.62	70.86	72.87	173	12.04	73.26	72.07	72.07	68.91	71.28	70.49	71.76
27	135	9.39	68.97	69.43	66.67	72.04	68.97	70.74	70.51	270	18.79	75.24	74.29	70.09	74.38	73.78	71.98	71.87	175	12.18	71.16	72.82	71.79	71.24	74.17	70.76	72.98
28	140	9.74	70.70	69.31	64.69	71.32	72.01	70.93	69.16	280	19.49	73.90	73.21	70.01	72.60	73.81	71.56	72.08	178	12.39	72.99	72.12	72.36	71.25	72.92	71.01	72.12
29	145	10.09	71.75	70.51	67.49	72.76	70.90	70.59	70.82	290	20.18	74.28	73.06	70.36	74.02	71.23	72.62	71.58	181	12.60	74.20	71.66	71.50	71.18	71.18	71.34	71.42
30	150	10.44	69.54	69.46	67.60	71.48	68.69	69.77	70.40	300	20.88	73.00	70.45	71.68	75.11	71.94	73.09	72.45	184	12.80	70.95	71.51	72.47	71.27	71.99	70.47	72.87
31	155	10.79	73.95	70.05	65.68	71.14	71.22	71.53	69.27	310	21.57	75.07	71.52	70.90	74.27	72.40	72.76	71.68	187	13.01	73.12	71.36	72.48	73.28	73.60	72.00	73.68
32	160	11.13	71.34	70.01	67.19	71.18	72.75	70.95	70.16	320	22.27	72.69	70.81	70.28	71.62	73.23	71.53	70.33	189	13.15	72.44	71.23	71.55	76.44	71.31	71.63	71.47
33	165	11.48	70.52	69.65	69.89	71.86	72.41	71.62	70.36	330	22.96	73.53	74.07	70.91	75.52	73.98	73.08	71.49	191	13.29	73.43	72.07	71.43	73.27	72.55	69.90	72.63
34	170	11.83	69.77	69.22	69.77	72.14	72.30	72.38	71.11	340	23.66	74.38	74.29	69.55	74.75	73.84	74.48	73.13	194	13.50	73.05	71.84	70.72	73.45	68.70	73.29	74.50
35	175	12.18	72.82	69.65	71.55	74.41	73.22	72.90	72.66	350	24.36	73.97	75.07	70.56	75.16	73.97	74.98	73.61	197	13.71	72.82	72.66	71.85	72.74	72.90	72.90	74.19
36	180	12.53	72.79	71.60	69.45	72.63	72.39	73.43	73.51	360	25.05	74.56	74.28	70.10	74.00	74.84	73.72	74.75	201	13.99	73.71	72.17	72.09	71.36	68.77	73.71	73.38
37	185	12.87	72.44	69.65	71.41	73.16	71.88	72.60	72.52	370	25.75	73.38	72.82	69.73	77.13	73.01	73.76	73.71	202	14.06	71.90	72.39	71.26	72.79	71.90	72.39	74.82
38	190	13.22	74.82	71.69	71.21	71.69	70.09	73.62	72.01	380	26.44	73.51	72.28	70.01	73.13	74.08	73.89	73.79	204	14.20	74.86	72.99	69.83	73.97	71.61	73.80	74.53
39	195	13.57	71.66	70.45	72.54	72.22	72.62	72.46	73.27	390	27.14	72.68	74.12	71.54	77.17	77.36	74.21	73.03	205	14.27	72.73	73.21	70.94	73.70	72.81	73.70	75.16
40	200	13.92	72.92	69.60	72.03	74.21	73.89	73.16	72.03	400	27.84	73.48	75.89	71.75	74.45	74.93	74.54	74.86	207	14.41	74.88	72.03	71.71	73.25	71.87	72.44	75.20
41	205	14.27	72.89	70.21	72.89	73.13	72.08	73.78	71.29	407	28.32	73.59	74.76	72.91	73.69	74.85	74.95	74.62	208	14.47	74.94	71.28	72.34	73.72	71.85	72.42	73.15
42	210	14.61	72.13	70.58	70.74	70.90	72.45	73.84	72.15	417	29.02	72.84	73.53	71.76	75.69	73.92	74.71	73.35	211	14.68	73.08	70.72	72.19	74.80	70.55	73.74	75.29
43	215	14.96	72.18	71.60	71.52	75.20	72.09	75.04	74.24	427	29.71	76.04	72.87	73.76	75.74	74.16	75.15	74									

Tabela 6 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS3 com as n primeiras melhores características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc).

k	pk = 5									pk = 10								pk = $\sqrt{ \cdot }$									
	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)
5	25	1.72	72.31	74.27	41.82	73.43	76.43	73.64	75.10	50	3.44	81.85	73.67	46.83	76.23	57.58	65.05	72.81	76	5.22	82.02	78.03	42.78	84.12	82.60	75.71	74.69
6	30	2.06	78.18	72.98	56.49	71.16	77.61	73.19	74.53	60	4.12	75.20	70.75	51.54	76.06	69.32	64.09	71.83	83	5.70	83.67	77.84	47.96	75.95	80.32	74.49	74.34
7	35	2.41	75.42	69.51	54.44	69.44	73.17	72.54	74.51	70	4.81	79.42	65.27	55.38	75.67	84.69	72.27	71.91	92	6.32	81.44	80.70	49.82	84.30	82.17	71.68	76.52
8	40	2.75	81.91	74.70	52.79	75.27	76.40	77.03	78.52	80	5.50	87.78	73.75	57.45	81.89	81.45	74.18	76.65	100	6.87	88.63	81.18	49.15	85.24	86.79	75.42	78.67
9	45	3.09	85.25	71.42	61.77	83.33	80.14	76.95	76.17	90	6.19	86.74	78.17	55.53	74.07	81.10	72.97	77.51	106	7.29	85.84	80.13	47.66	83.17	81.10	74.43	79.84
10	50	3.44	84.48	78.29	47.12	83.63	81.21	75.44	77.08	100	6.87	86.64	75.79	48.78	83.62	84.28	73.87	78.45	111	7.63	84.97	80.65	49.40	83.85	91.15	75.00	78.87
11	55	3.78	82.14	79.07	51.29	82.00	81.36	78.43	78.71	110	7.56	87.66	78.51	53.83	84.16	88.85	79.78	81.26	113	7.77	86.07	79.14	50.37	85.39	86.36	77.50	82.41
12	60	4.12	83.37	74.48	49.18	75.84	83.08	78.92	79.78	120	8.25	86.14	76.55	51.61	83.90	88.09	80.15	81.27	119	8.18	88.10	79.42	49.63	83.83	84.51	77.92	80.46
13	65	4.47	82.59	77.27	48.56	81.58	81.29	75.32	78.85	130	8.93	87.77	76.98	52.15	84.30	91.02	83.85	81.96	127	8.73	87.42	77.71	51.96	82.83	88.93	81.17	79.22
14	70	4.81	83.75	77.04	48.59	83.90	78.41	78.27	80.72	140	9.62	87.15	76.27	50.87	86.77	94.45	87.07	84.33	131	9.00	87.54	77.49	50.68	84.74	91.09	85.35	85.88
15	75	5.15	87.54	75.87	48.04	87.68	77.39	80.58	80.22	150	10.31	90.11	76.32	53.33	83.07	90.11	86.28	87.13	135	9.28	90.08	77.50	50.23	85.38	88.94	85.00	87.20
16	80	5.50	83.42	74.47	48.95	82.33	86.11	82.18	81.02	160	11.00	90.04	77.07	52.59	87.49	93.20	85.10	86.72	140	9.62	87.98	76.88	49.96	85.78	88.59	86.24	86.54
17	85	5.84	80.58	77.59	53.21	79.12	87.15	82.12	77.96	170	11.68	86.69	79.46	61.56	80.47	93.62	85.21	84.36	143	9.83	88.26	79.88	53.12	82.32	91.84	85.14	85.44
18	90	6.19	86.37	77.22	60.15	82.64	89.38	84.18	82.86	180	12.37	90.82	77.80	71.37	84.63	92.39	86.75	85.57	148	10.17	91.20	77.58	63.81	88.14	90.05	84.54	84.09
19	95	6.53	84.12	76.03	57.79	85.22	87.28	84.93	81.62	190	13.06	90.28	77.45	63.79	86.01	91.86	88.14	88.22	153	10.52	88.56	78.03	62.06	85.79	91.94	87.33	86.33
20	100	6.87	85.17	75.65	57.56	87.01	89.00	84.50	81.11	200	13.75	87.25	76.81	63.43	86.45	92.03	88.05	85.42	156	10.72	87.99	78.68	58.58	86.91	90.61	88.30	87.45
21	105	7.22	81.56	78.81	62.07	80.67	87.04	84.59	80.00	210	14.43	87.15	75.42	69.08	88.67	90.52	87.63	87.79	159	10.93	84.57	77.47	63.89	86.27	90.35	86.96	85.49
22	110	7.56	85.72	80.59	63.27	82.53	88.18	83.72	82.68	220	15.12	92.23	77.25	68.10	88.50	90.28	87.85	87.13	163	11.20	90.48	79.64	67.26	87.54	89.55	87.00	85.84
23	115	7.90	90.52	78.96	68.21	85.75	86.49	84.70	82.16	230	15.81	92.16	77.31	70.37	89.71	91.10	88.24	86.94	167	11.48	90.92	80.12	66.85	85.79	93.25	87.66	86.65
24	120	8.25	89.29	78.43	68.46	84.79	91.09	84.12	82.70	240	16.49	93.17	78.11	71.36	89.38	91.85	88.89	87.49	172	11.82	90.96	78.72	71.08	88.46	90.57	87.92	86.52
25	125	8.59	88.87	77.52	68.20	88.35	91.58	86.54	83.98	250	17.18	93.11	76.76	71.87	88.63	95.27	92.28	88.96	176	12.10	92.34	77.48	69.35	85.07	89.52	89.29	87.57
26	130	8.93	90.57	78.79	62.11	83.85	90.57	87.40	85.36	260	17.87	93.97	76.99	73.47	88.95	92.64	91.72	89.04	179	12.30	92.55	77.19	69.28	88.79	89.18	90.36	88.79
27	135	9.28	90.45	78.64	61.89	80.23	89.62	87.05	85.15	270	18.56	92.57	75.95	76.37	89.70	93.92	92.91	88.69	182	12.51	92.07	76.51	68.26	87.75	92.07	90.81	88.22
28	140	9.62	89.05	78.48	66.84	81.90	90.19	87.07	84.41	280	19.24	93.36	75.23	79.15	89.53	92.60	92.94	88.87	184	12.65	91.74	76.32	70.42	87.18	92.13	90.32	87.57
29	145	9.97	90.69	79.69	68.32	82.29	90.76	86.41	85.57	290	19.93	93.65	78.71	78.11	91.59	92.70	93.13	88.97	186	12.78	92.12	77.23	70.53	90.54	91.96	91.57	87.16
30	150	10.31	91.72	79.62	67.43	83.60	88.43	86.36	84.29	300	20.62	93.07	77.84	78.53	89.61	93.85	92.38	88.88	189	12.99	92.18	77.41	70.46	87.60	89.02	90.60	87.60
31	155	10.65	91.54	80.85	71.00	88.62	89.46	87.38	86.62	310	21.31	94.06	80.44	77.12	91.62	93.45	92.05	89.14	192	13.20	92.48	78.46	69.68	89.23	91.05	91.05	87.89
32	160	11.00	92.28	81.47	69.27	88.26	88.65	87.72	86.65	317	21.79	94.02	80.67	78.21	89.46	94.02	92.27	89.34	194	13.33	91.67	77.72	71.85	86.04	87.87	91.99	88.26
33	165	11.34	91.32	84.03	77.05	88.84	89.07	88.14	86.99	327	22.47	94.24	85.11	82.98	89.54	94.59	92.29	89.43	197	13.54	92.29	83.31	77.66	91.26	88.79	90.62	88.31
34	170	11.68	91.98	84.28	75.80	86.54	92.92	88.02	87.33	337	23.16	94.01	83.01	82.38	89.62	94.54	92.31	88.98	201	13.81	91.55	83.33	74.56	88.44	92.34	90.35	89.23
35	175	12.03	88.67	85.39	75.78	86.88	91.41	89.06	87.05	346	23.78	93.51	82.78	85.03	90.53	93.69	92.06	89.64	203	13.95	91.37	83.31	79.71	88.10	91.77	90.81	89.46
36	180	12.37	87.92	83.22	76.00	89.10	90.90	88.16	86.45	356	24.47	94.18	83.17	84.90	89.35	91.99	92.36	90.28	205	14.09	90.88	83.20	75.44	89.36	90.88	91.04	88.08
37	185	12.71	90.47	83.62	76.14	90.87	91.81	88.58	87.50	366	25.15	92.93	82.37	84.21	93.20	92.93	92.65	90.04	207	14.23	91.03	83.49	79.89	88.94	94.47	91.03	88.70
38	190	13.06	91.86	84.03	74.86	90.83	90.75	89.96	88.09	374	25.70	93.62	82.61	82.70	92.41	93.99	93.15	90.53	209	14.36	92.05	83.15	77.37	88.84	91.09	91.89	89.57
39	195	13.40	90.79	82.14	80.16	87.86	92.54	89.21	87.17	384	26.39	93.65	81.79	85.43	89.54	94.68	93.74	90.54	211	14.50	91.56	80.95	81.11	89.71	93.09	92.04	89.39
40	200	13.75	92.59	82.07	81.04	87.89	90.68	88.69	87.84	394	27.08	94.63	81.06	87.37	90.57	93.97	93.50	90.29	213	14.64	92.35	81.96	81.24	88.89	93.16	92.27	90.10
41	205	14.09	90.96	82.72	81.68	86.64	91.60	88.08	87.39	404	27.77	95.05	82.49	87.73	91.25	95.62	94.39	90.96	215	14.78	90.73	82.34	80.32	92.74	93.71	91.45	89.52
42	210	14.43	91.41	82.65	83.21	88.27	91.89	87.55	87.82	414	28.45	94.14	82.80	87.32	88.76	96.73	94.04	91.17	216	14.85	91.61	82.32	81.84	86.92	90.64	91.85	89.10
43	215	14.78	92.74	82.10	80.00	87.42	91.61	87.74	86.00	424	29.14	95.83	82.54	87.68	9												

Tabela 7 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS4 com as n primeiras melhores características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc).

k	pk = 5									pk = 10									pk = $\sqrt{ \cdot }$								
	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)
5	25	1.50	78.44	84.91	74.47	87.35	89.55	88.21	86.44	50	3.01	85.30	86.97	76.30	86.48	87.78	87.97	86.23	85	5.11	92.77	88.02	71.78	91.69	87.89	86.30	88.52
6	30	1.81	85.05	86.27	74.26	88.42	89.22	86.83	86.03	60	3.61	92.51	86.95	73.10	87.70	91.01	87.95	89.08	93	5.60	92.67	88.40	69.15	88.78	87.00	88.21	89.74
7	35	2.11	90.53	86.17	72.34	90.78	86.29	86.72	87.89	70	4.21	92.71	87.06	67.90	88.44	86.62	88.07	88.88	101	6.08	93.21	88.02	65.98	92.18	89.81	88.98	90.01
8	40	2.41	82.06	87.11	74.29	86.81	87.18	87.36	87.36	80	4.81	87.29	87.48	66.81	90.64	89.51	87.29	88.62	107	6.44	91.38	88.62	65.72	90.68	90.42	87.27	88.55
9	45	2.71	90.11	87.20	73.16	89.36	88.68	88.25	88.87	90	5.42	91.86	88.23	67.24	90.52	90.08	86.70	89.25	114	6.86	92.96	88.24	72.35	90.76	91.09	87.02	89.02
10	50	3.01	88.77	87.72	72.33	90.26	83.56	88.34	90.63	100	6.02	88.80	88.60	72.22	87.77	88.67	87.13	88.48	119	7.16	90.67	87.95	73.95	92.03	89.76	86.91	89.05
11	55	3.31	92.47	87.37	72.62	90.23	89.61	87.24	90.85	110	6.62	92.33	89.18	70.75	90.21	91.56	87.18	88.27	127	7.64	94.53	91.14	74.20	91.60	91.47	88.01	88.60
12	60	3.61	91.20	88.89	69.79	91.64	85.96	84.33	88.89	120	7.22	92.87	87.55	68.22	90.99	91.96	83.01	87.16	133	8.00	91.96	90.26	71.55	91.96	92.28	84.04	87.51
13	65	3.91	90.98	87.91	76.71	90.17	88.29	90.54	92.11	130	7.82	94.26	89.30	76.96	93.73	92.75	91.38	90.60	138	8.30	95.01	87.93	77.69	92.32	94.16	90.55	90.42
14	70	4.21	88.32	90.89	78.08	89.82	88.51	91.83	92.46	140	8.42	92.71	90.74	79.30	93.56	92.31	92.38	92.05	142	8.54	93.03	90.66	79.61	93.22	90.39	92.17	91.32
15	75	4.51	92.19	91.62	84.81	90.99	91.37	92.38	91.93	150	9.03	93.12	90.41	86.90	93.19	92.66	92.53	91.07	146	8.78	93.01	89.97	85.49	91.49	94.39	91.36	91.62
16	80	4.81	91.85	92.23	82.74	90.20	92.04	92.23	92.29	160	9.63	92.14	90.88	87.22	95.21	92.94	92.41	91.08	150	9.03	92.79	90.67	86.51	94.05	92.86	91.93	92.00
17	85	5.11	91.69	90.93	82.94	92.14	89.60	93.09	91.06	170	10.23	93.90	90.88	84.85	93.57	92.90	93.16	91.29	156	9.39	94.22	90.70	87.98	93.69	92.10	91.50	92.03
18	90	5.42	89.31	91.54	81.55	90.59	89.06	91.48	91.35	180	10.83	93.86	91.77	86.30	94.94	94.67	93.12	92.24	160	9.63	94.41	88.75	87.35	94.14	94.34	93.14	91.15
19	95	5.72	92.60	90.62	83.85	92.21	92.09	92.21	91.00	190	11.43	93.48	90.76	86.89	92.60	94.02	91.85	91.92	163	9.81	92.73	91.13	88.86	93.53	92.66	94.06	92.86
20	100	6.02	92.00	90.01	84.76	91.61	90.20	92.77	91.29	200	12.03	93.71	90.49	86.46	93.37	92.07	91.66	92.20	166	9.99	94.45	90.91	90.91	90.84	94.25	93.72	92.25
21	105	6.32	90.04	90.11	84.78	91.07	92.36	91.84	89.79	210	12.64	92.98	89.88	89.67	92.98	92.08	91.87	90.84	171	10.29	92.56	90.07	91.01	92.22	91.82	93.49	92.35
22	110	6.62	91.37	89.63	84.92	93.23	93.04	91.49	92.27	220	13.24	92.23	90.85	90.29	93.27	92.72	91.68	91.05	173	10.41	92.01	90.26	89.19	92.48	95.10	94.02	93.49
23	115	6.92	92.63	90.37	86.23	93.08	94.25	93.21	92.37	230	13.84	92.39	90.57	90.50	92.74	93.72	93.58	93.99	179	10.77	92.18	90.76	91.77	93.32	92.31	94.88	93.80
24	120	7.22	92.15	91.12	86.38	94.29	93.58	92.09	91.37	240	14.44	93.67	91.56	89.24	93.53	94.80	93.11	93.25	181	10.89	93.79	92.84	90.14	93.11	93.99	93.18	93.18
25	125	7.52	92.13	91.80	88.09	93.04	92.39	92.91	92.26	250	15.04	94.48	91.71	90.01	92.42	93.27	94.05	93.98	186	11.19	95.05	90.99	89.63	94.17	95.26	94.99	93.36
26	130	7.82	92.43	91.91	87.60	92.30	93.99	92.56	92.36	260	15.64	92.94	92.65	90.37	92.08	95.29	94.15	93.72	188	11.31	93.49	92.61	90.57	94.10	96.68	94.91	94.10
27	135	8.12	92.86	90.83	88.15	92.53	94.37	92.01	91.36	270	16.25	95.69	93.18	91.45	92.96	94.25	95.26	94.54	191	11.49	94.97	92.05	90.55	93.47	95.99	95.65	94.97
28	140	8.42	90.08	91.98	85.87	90.54	93.63	92.38	91.52	280	16.85	92.11	92.19	91.17	92.19	94.14	94.65	94.29	194	11.67	91.21	92.78	91.89	93.87	94.41	95.37	94.82
29	145	8.72	93.87	91.10	91.43	90.90	94.26	91.69	92.22	290	17.45	92.06	92.71	92.64	92.27	94.75	95.12	93.74	199	11.97	93.10	93.16	93.10	93.23	94.81	94.33	93.92
30	150	9.03	93.98	91.34	90.21	91.47	94.25	92.46	91.80	300	18.05	94.20	92.44	93.39	95.52	95.52	95.30	94.87	202	12.15	95.55	91.71	92.74	95.75	95.75	95.34	93.97
31	155	9.33	93.56	91.90	91.24	91.37	91.84	91.97	91.97	310	18.65	96.01	93.57	93.42	94.08	95.64	94.82	93.80	205	12.33	96.57	93.07	93.27	93.89	93.75	94.85	94.10
32	160	9.63	93.08	91.88	90.95	94.34	94.81	93.61	93.08	320	19.25	94.71	92.70	92.40	94.41	95.23	96.65	94.64	208	12.52	94.22	92.98	91.54	93.60	96.84	95.39	94.91
33	165	9.93	93.85	91.72	90.65	92.59	91.72	93.25	92.25	330	19.86	94.07	92.49	91.89	93.99	94.89	95.72	94.84	210	12.64	95.11	93.25	92.49	94.08	95.45	95.45	94.08
34	170	10.23	93.10	91.29	88.87	93.50	94.44	93.90	93.03	340	20.46	94.93	93.57	90.70	94.78	96.75	95.84	95.70	214	12.88	94.89	93.37	91.57	94.13	95.23	95.17	95.65
35	175	10.53	94.49	92.40	88.37	93.34	93.88	94.28	93.54	350	21.06	96.34	95.35	90.93	95.81	97.03	95.66	96.50	218	13.12	95.71	93.91	91.76	93.28	95.78	95.08	94.53
36	180	10.83	91.43	92.65	86.98	93.52	94.87	93.93	93.05	360	21.66	96.08	94.16	90.40	95.31	96.77	95.70	96.09	221	13.30	94.38	93.96	90.91	95.28	95.42	95.21	94.24
37	185	11.13	94.18	94.18	87.47	94.31	93.43	94.58	93.64	370	22.26	97.06	95.59	90.63	95.20	96.75	95.98	95.83	224	13.48	93.88	93.95	90.26	96.24	95.13	94.78	95.41
38	190	11.43	95.18	93.14	89.33	93.75	93.82	94.50	92.93	380	22.86	97.11	94.23	91.73	95.40	96.72	96.41	96.11	227	13.66	96.03	93.03	91.50	93.17	95.89	95.75	94.29
39	195	11.73	95.57	93.66	93.32	94.00	94.75	95.09	94.07	390	23.47	97.25	94.65	94.03	96.15	97.17	96.54	96.25	230	13.84	96.16	93.78	92.81	95.53	95.60	95.32	94.27
40	200	12.03	94.94	93.71	93.64	94.73	95.96	94.39	94.05	400	24.07	96.43	94.14	94.14	96.04	97.07	96.43	96.14	231	13.90	94.90	94.06	93.43	95.46	95.04	95.32	95.04
41	205	12.33	94.99	93.14	93.34	94.65	95.61	94.78	95.06	410	24.67	95.61	94.73	94.09	96.41	97.12	96.09	96.27	235	14.14	95.23	94.74	92.78	94.88	95.52	94.74	94.53
42	210	12.64	96.07	93.32	94.49	94.49	96.35	95.45	94.63	420	25.27	96.54	95.25	96.86	95.57	98.31	97.58	96.88	237	14.26	96.77	93.82	94.18	95.30	95.65	95.93	94.18
43	215	12.94	94.19	93.30	94.19	94.47	96.34	95.16	94.34	428	25.75	95.79	95.62	96.03	96.43	97.41	97.41	96.									

Tabela 8 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS1 com 102 características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc).

k	pk = 5								pk = 10								pk = $\sqrt{ \cdot }$										
	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)
5	21	0.92	59.22	59.44	40.43	56.34	51.29	62.99	58.78	41	1.80	65.52	62.43	28.00	59.75	60.15	59.70	59.57	93	4.08	69.05	60.76	42.22	65.29	65.71	62.23	63.46
6	22	0.97	55.03	60.80	36.14	58.63	40.62	61.42	60.27	42	1.84	64.70	63.31	27.16	62.64	61.39	60.85	61.34	94	4.13	68.53	59.60	43.43	66.56	68.94	61.98	62.02
7	27	1.19	57.42	62.84	32.58	63.51	48.09	63.51	64.36	52	2.28	55.28	63.28	38.61	62.79	58.29	64.18	62.79	100	4.39	68.21	58.25	43.09	64.63	65.41	62.20	63.62
8	32	1.41	64.63	64.05	38.93	66.06	60.58	64.37	65.26	62	2.72	65.55	61.90	47.49	64.33	65.10	66.23	66.09	108	4.74	68.33	57.81	45.74	65.38	69.34	64.41	63.95
9	37	1.62	63.30	60.45	52.95	60.36	65.04	61.16	62.50	72	3.16	70.25	63.27	45.94	65.58	65.76	66.67	65.49	119	5.23	71.46	63.44	49.81	67.38	69.69	67.01	66.73
10	38	1.67	59.36	63.33	47.74	64.49	59.85	63.38	61.50	73	3.21	66.92	64.07	48.19	65.83	59.17	66.70	67.11	120	5.27	63.61	63.47	48.49	68.43	66.57	65.09	66.25
11	43	1.89	60.56	64.64	49.06	63.29	59.40	66.97	64.23	83	3.65	63.58	64.63	48.13	64.40	59.75	65.82	66.23	127	5.58	67.81	64.47	50.51	68.70	68.88	66.88	68.56
12	48	2.11	66.49	65.63	44.32	63.97	58.82	65.10	65.14	93	4.08	68.09	66.67	44.92	67.17	68.22	66.35	66.70	130	5.71	70.98	65.02	51.75	68.14	66.14	67.58	67.02
13	53	2.33	61.06	66.01	43.88	65.33	58.41	64.52	66.10	103	4.52	67.34	65.46	47.01	66.24	66.84	67.53	64.61	135	5.93	69.65	64.89	51.63	68.25	66.15	69.14	67.97
14	54	2.37	69.41	66.44	46.42	66.04	64.28	64.51	65.77	104	4.57	71.01	66.59	46.80	68.02	61.90	66.59	65.75	136	5.97	69.08	64.69	52.59	67.49	65.30	69.27	67.49
15	59	2.59	64.16	65.87	46.44	65.06	61.41	65.06	65.01	114	5.01	67.82	65.51	43.69	64.31	66.57	65.42	64.90	145	6.37	70.17	64.12	48.22	69.14	71.20	66.28	66.98
16	64	2.81	67.10	65.16	46.63	64.93	62.72	63.58	62.99	124	5.45	69.25	65.12	50.95	65.07	67.44	64.75	64.97	151	6.63	68.06	62.94	49.53	68.63	68.02	65.38	65.66
17	69	3.03	64.09	65.04	47.01	68.07	64.67	63.72	62.86	134	5.88	69.34	65.61	54.27	67.94	70.28	65.84	64.62	153	6.72	70.57	63.70	52.64	67.37	69.16	66.24	65.58
18	74	3.25	68.86	64.87	44.62	64.55	62.05	63.00	62.82	144	6.32	69.20	65.35	52.79	67.74	68.40	67.60	64.73	158	6.94	70.22	64.61	54.22	68.00	70.36	66.07	66.30
19	75	3.29	68.85	66.12	45.10	65.53	67.94	62.22	65.71	145	6.37	71.86	66.04	53.56	71.11	68.06	66.46	65.09	159	6.98	71.58	65.53	52.74	67.04	66.62	67.66	66.15
20	80	3.51	67.73	64.54	47.02	66.73	61.08	62.36	62.18	155	6.81	72.24	64.33	53.11	68.43	67.39	67.15	65.63	167	7.33	72.37	65.78	53.22	68.48	71.28	66.68	66.54
21	84	3.69	69.27	62.56	44.69	64.20	66.71	65.16	63.81	159	6.98	70.59	63.74	54.77	69.45	69.22	66.19	66.85	169	7.42	69.69	66.22	49.43	70.26	68.98	66.60	66.13
22	89	3.91	69.84	65.36	46.44	66.32	60.51	65.13	63.79	167	7.33	71.52	66.35	52.46	69.38	70.38	66.78	66.67	171	7.51	69.09	65.67	51.85	69.80	66.33	65.24	66.29
23	94	4.13	69.58	65.69	44.53	64.91	66.10	64.59	64.99	177	7.77	71.52	65.81	55.14	67.62	70.71	66.00	67.17	177	7.77	71.33	65.71	51.95	69.14	72.52	66.86	65.62
24	99	4.35	66.85	66.35	43.71	65.89	66.67	64.05	65.64	187	8.21	69.62	66.89	55.50	69.00	70.81	67.13	67.21	183	8.04	71.25	67.67	51.53	69.05	69.91	67.38	66.91
25	104	4.57	67.74	66.50	42.94	65.07	62.31	66.73	67.03	197	8.65	70.38	69.23	55.19	69.52	72.50	66.30	68.24	186	8.17	69.73	67.91	51.22	67.81	68.77	66.76	67.86
26	107	4.70	68.48	65.90	43.92	67.33	66.91	65.07	66.59	200	8.78	70.29	68.08	55.61	69.04	71.40	67.40	68.63	187	8.21	72.06	65.89	50.05	69.38	70.14	67.13	68.09
27	110	4.83	64.70	65.44	43.56	68.44	67.88	65.39	67.02	200	8.78	69.81	66.20	58.06	70.10	70.73	67.40	68.12	189	8.30	70.64	66.86	49.28	70.11	70.45	67.48	67.15
28	115	5.05	69.89	67.07	43.34	67.44	68.36	64.85	66.85	210	9.22	73.34	67.59	54.91	69.42	70.88	67.83	69.27	195	8.56	75.60	68.01	50.48	71.47	72.72	67.87	67.77
29	120	5.27	71.26	65.74	44.09	69.31	65.88	66.76	67.24	220	9.66	74.53	69.47	56.30	70.54	71.22	67.96	70.86	202	8.87	74.17	67.33	51.42	69.20	73.45	67.76	69.06
30	125	5.49	71.65	66.26	45.96	68.73	67.89	69.10	66.60	230	10.10	72.74	69.37	59.40	68.88	73.82	69.71	70.96	206	9.05	72.57	67.94	52.68	72.09	72.72	69.48	69.53
31	130	5.71	69.73	67.54	44.53	69.40	65.95	68.84	66.91	237	10.41	74.26	69.75	62.50	72.25	71.42	69.51	69.61	207	9.09	73.67	67.83	55.27	71.93	71.26	69.61	70.68
32	135	5.93	73.44	68.53	45.00	68.81	61.81	68.53	66.84	247	10.85	71.38	69.26	59.66	69.95	75.12	68.92	71.53	210	9.22	73.10	68.60	53.60	71.75	72.91	70.63	68.94
33	140	6.15	70.38	66.35	48.01	69.82	65.65	68.69	66.29	257	11.29	74.26	66.83	60.40	70.54	72.92	69.21	70.01	215	9.44	72.65	69.30	54.75	69.93	72.94	70.76	68.67
34	145	6.37	70.64	66.93	48.17	67.73	71.01	71.53	67.24	267	11.73	74.98	67.21	62.94	71.69	73.88	69.40	70.56	220	9.66	73.55	69.86	55.03	70.39	70.93	69.71	69.62
35	150	6.59	68.88	68.12	48.47	69.58	72.31	70.76	67.81	273	11.99	75.25	69.41	62.13	70.76	73.45	71.46	69.74	221	9.71	75.05	70.38	57.59	71.30	74.56	69.41	70.43
36	152	6.68	72.19	67.15	43.91	68.42	68.56	69.79	68.48	275	12.08	74.38	69.73	62.74	71.88	74.88	70.33	71.00	222	9.75	71.53	68.86	57.27	70.41	74.45	69.15	69.44
37	157	6.90	71.32	67.22	50.90	67.31	73.63	70.66	66.90	285	12.52	73.80	69.38	61.45	72.99	73.80	70.68	71.04	225	9.88	72.08	69.93	56.09	71.20	74.03	71.00	70.47
38	162	7.11	72.39	69.27	49.27	70.21	72.25	70.17	69.18	295	12.96	76.59	68.82	61.91	73.36	71.59	72.45	71.39	232	10.19	74.03	69.10	56.77	71.34	73.59	71.49	69.49
39	167	7.33	70.57	67.20	47.68	71.85	69.62	70.24	69.44	305	13.39	74.39	68.15	64.30	71.45	74.29	72.16	71.25	235	10.32	72.92	69.34	56.86	71.94	71.45	71.60	70.18
40	172	7.55	70.12	68.41	53.35	73.16	70.50	70.40	68.04	315	13.83	71.36	68.60	68.45	70.64	76.35	72.73	72.01	238	10.45	71.02	68.51	62.92	73.96	75.38	73.17	71.90
41	174	7.64	70.23	68.66	56.06	69.66	71.37	71.04	69.19	317	13.92	71.73	68.37	69.80	72.55	76.31	73.21	71.68	239	10.50	70.41	69.04	62.90	74.04	75.27	72.62	71.59
42	179	7.86	72.07	66.35	57.67	70.35	69.40	72.07	69.55	327	14.36	73.28	68.26	69.85	73.33	74.56	73.18	72.45	243	10.67	74.53	69.47	64.95	74.24	76.50	71.39	71.29
43	184	8.08	72.72	68.04	55.71	69.04	71.48	72.58	69.14	337	14.80	76.44	68.87	70.62	71.19	74.18	73.71	72.36	245	10.76</							

Tabela 9 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS2 com 102 características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc).

k	pk = 5									pk = 10									pk = $\sqrt{ \cdot }$								
	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)
5	17	1.18	52.18	60.70	62.18	56.97	54.37	58.52	60.70	32	2.23	61.07	60.57	58.22	59.50	59.43	61.99	59.93	66	4.59	60.54	64.62	62.14	65.28	67.76	59.30	59.23
6	22	1.53	57.67	62.05	62.40	56.61	58.45	57.88	62.19	42	2.92	62.51	62.58	64.16	65.59	62.94	65.30	62.87	74	5.15	68.31	62.14	61.48	71.31	63.76	63.39	61.92
7	27	1.88	58.44	58.87	62.84	61.99	62.41	61.56	58.16	52	3.62	65.20	60.51	62.96	65.05	65.49	63.25	62.38	82	5.71	67.68	68.27	63.32	64.35	68.34	66.94	63.84
8	28	1.95	60.75	61.67	62.24	62.38	66.50	56.85	59.97	53	3.69	62.72	63.01	62.14	64.67	56.65	63.73	63.66	83	5.78	67.13	67.95	63.37	67.87	62.33	64.33	64.92
9	33	2.30	63.03	59.83	53.92	61.54	49.72	61.18	54.84	63	4.38	64.48	67.90	62.15	66.38	62.30	62.37	64.63	91	6.33	66.94	66.57	63.67	66.64	60.85	65.16	65.23
10	34	2.37	65.22	61.58	58.09	63.36	49.47	63.86	57.02	64	4.45	67.73	68.61	63.73	65.77	62.49	62.78	65.55	92	6.40	67.88	66.17	63.57	65.43	64.24	64.68	64.24
11	39	2.71	63.09	62.95	60.09	61.09	59.30	62.30	61.23	74	5.15	65.37	67.50	60.67	66.18	68.82	64.86	64.57	95	6.61	64.01	67.51	64.38	65.80	64.01	65.35	65.80
12	44	3.06	59.94	62.38	58.36	66.62	64.97	62.74	61.38	84	5.85	64.97	67.78	64.23	66.00	64.52	66.08	69.15	98	6.82	68.04	65.80	64.75	65.94	65.12	66.17	66.24
13	45	3.13	47.92	59.77	60.70	65.01	58.84	62.64	58.98	85	5.92	60.87	67.83	63.24	67.23	66.35	66.05	66.03	99	6.89	67.64	66.29	63.68	68.68	67.79	65.02	67.34
14	50	3.48	58.83	63.09	62.22	59.41	68.06	64.38	61.93	95	6.61	65.72	65.20	64.16	67.21	65.20	64.98	64.74	103	7.17	68.52	67.47	64.32	68.74	67.54	63.04	65.37
15	55	3.83	69.10	65.05	60.71	63.97	65.99	61.79	63.75	105	7.31	65.77	67.34	64.41	69.07	66.82	67.57	64.17	107	7.45	68.05	67.37	63.16	67.07	65.11	66.47	66.62
16	59	4.11	65.38	66.69	60.30	68.58	64.66	61.90	62.15	109	7.59	65.89	67.47	64.68	65.89	63.86	65.81	66.79	109	7.59	66.79	68.60	63.03	68.52	68.07	67.77	65.44
17	64	4.45	65.11	67.66	59.07	66.79	66.86	63.07	65.38	117	8.14	67.12	67.95	64.47	67.42	62.88	66.06	67.24	110	7.65	69.33	67.97	63.15	67.90	65.64	67.14	66.69
18	68	4.73	57.49	68.15	60.63	65.96	67.79	64.79	66.62	121	8.42	65.65	68.24	65.27	70.82	63.30	66.95	67.30	111	7.72	62.82	68.70	62.14	68.93	67.12	66.67	66.97
19	73	5.08	66.50	67.52	59.46	65.91	67.01	64.44	68.32	131	9.12	69.07	68.38	64.55	69.14	66.23	69.83	66.59	119	8.28	70.71	67.53	63.58	68.21	63.73	68.06	67.60
20	77	5.36	68.46	68.24	61.03	67.94	67.43	65.59	68.33	135	9.39	69.89	67.82	64.82	72.58	69.59	68.51	67.35	121	8.42	69.45	66.95	64.36	70.67	65.88	68.69	68.54
21	82	5.71	66.20	66.72	62.21	64.35	61.18	65.83	67.13	142	9.88	67.64	67.95	65.33	67.88	67.18	68.19	63.52	123	8.56	65.60	67.66	63.70	68.57	69.41	66.36	68.04
22	84	5.85	64.52	64.01	63.12	62.68	66.44	63.49	64.36	144	10.02	68.52	66.36	65.51	68.83	69.68	68.45	65.69	124	8.63	69.15	68.39	65.12	69.23	69.15	65.88	68.01
23	89	6.19	64.39	57.64	64.09	62.83	58.16	61.72	63.71	154	10.72	67.19	66.02	65.32	71.63	66.48	68.90	63.96	130	9.05	66.72	66.11	64.27	68.40	67.33	67.02	67.64
24	94	6.54	68.80	63.44	63.14	67.01	64.85	65.97	67.36	164	11.41	72.66	67.48	66.30	71.09	66.69	69.52	67.34	135	9.39	72.96	63.75	62.52	69.05	64.98	67.51	68.89
25	99	6.89	66.59	60.76	66.89	69.06	69.21	63.60	65.97	174	12.11	66.35	66.83	71.73	71.10	68.80	67.06	66.61	139	9.67	68.41	64.79	67.03	72.57	69.72	67.41	68.10
26	104	7.24	67.29	59.71	67.22	65.57	69.39	66.54	60.15	180	12.53	70.25	67.14	72.24	68.18	70.49	68.34	66.72	140	9.74	68.00	66.38	67.23	71.32	68.93	67.46	67.15
27	109	7.59	63.63	61.22	66.49	63.63	67.70	65.96	64.49	190	13.22	69.77	68.89	71.13	72.57	67.44	69.93	69.48	143	9.95	72.18	63.76	67.54	71.17	67.54	68.01	67.85
28	114	7.93	69.84	64.25	67.04	67.20	66.36	67.50	65.86	200	13.92	71.06	69.60	71.46	72.59	69.68	69.20	69.31	149	10.37	73.21	67.24	67.55	68.17	66.85	67.86	67.39
29	119	8.28	69.42	66.24	67.37	65.86	67.15	65.17	67.25	210	14.61	71.15	70.25	71.72	71.80	69.85	70.42	68.26	152	10.58	69.34	68.09	68.09	69.88	68.95	64.82	67.16
30	124	8.63	70.75	66.87	68.32	68.47	63.98	66.26	67.05	220	15.31	74.36	71.82	70.91	69.60	70.91	71.32	68.97	158	11.00	72.24	67.08	69.51	70.21	72.79	67.01	67.94
31	129	8.98	69.19	65.60	67.05	69.72	67.51	66.28	64.79	230	16.01	72.66	71.67	73.07	69.01	70.26	71.00	69.70	162	11.27	69.25	69.25	70.59	70.04	68.71	68.55	67.45
32	130	9.05	71.08	66.34	67.02	66.72	65.80	66.64	68.57	231	16.08	73.63	69.98	72.31	71.72	73.05	67.91	69.60	163	11.34	71.66	67.50	67.97	72.61	69.70	67.50	69.00
33	135	9.39	71.81	66.97	68.28	70.89	67.82	67.59	68.75	241	16.77	71.15	71.74	73.24	73.08	73.24	71.15	71.33	167	11.62	70.24	68.43	70.16	72.13	72.28	70.79	70.63
34	140	9.74	72.55	69.31	69.01	68.85	70.32	68.31	69.10	251	17.47	71.59	70.66	72.93	69.98	71.33	70.24	70.43	171	11.90	72.04	68.33	69.83	70.54	67.54	67.69	68.56
35	145	10.09	68.81	67.18	69.66	66.72	70.12	69.12	68.51	261	18.16	72.36	72.19	72.36	71.17	72.36	69.13	71.44	173	12.04	74.21	67.25	69.46	68.99	69.94	68.35	70.73
36	146	10.16	69.48	66.69	69.48	67.00	70.57	70.33	69.41	262	18.23	75.74	71.91	73.11	74.04	69.96	69.79	72.18	174	12.11	73.95	67.54	71.10	70.31	69.20	68.09	69.04
37	151	10.51	71.46	69.36	70.14	69.75	66.80	69.36	68.62	269	18.72	73.54	72.95	72.43	69.61	71.58	72.35	70.17	175	12.18	73.30	67.43	70.13	74.17	69.65	69.73	69.73
38	156	10.86	72.52	69.71	70.65	67.45	69.40	70.10	70.44	279	19.42	73.92	73.06	72.45	72.11	73.14	72.63	71.90	178	12.39	72.99	68.63	70.21	72.44	71.17	69.82	70.77
39	157	10.93	70.86	69.30	70.86	68.52	69.61	70.00	70.19	280	19.49	71.56	72.77	73.47	74.50	71.74	72.60	71.54	179	12.46	70.03	68.36	69.16	71.14	71.14	70.27	70.59
40	158	11.00	69.19	71.31	71.54	71.07	70.84	70.05	69.70	280	19.49	73.29	73.98	72.60	75.28	72.26	73.64	72.11	180	12.53	71.84	68.50	70.01	72.08	70.17	69.13	69.93
41	163	11.34	71.43	69.47	70.57	70.49	70.64	66.09	71.06	290	20.18	73.84	72.80	72.54	70.71	71.23	71.58	72.04	183	12.73	70.18	67.38	70.89	71.29	70.57	70.81	71.13
42	167	11.62	71.97	70.47	70.63	72.28	71.42	68.74	70.13	294	20.46	72.62	74.10	72.70	74.02	71.65	72.35	70.91	185	12.87	71.01	69.81	69.01	69.89	71.49	71.41	70.53
43	168	11.69	72.03	69.90	69.42	72.18	69.66	68.79	71.36	295	20.53	73.47	73.82	71.89	74.61	71.19	72.77	71.14									

Tabela 10 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS3 com 102 características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc).

k	pk = 5									pk = 10									pk = $\sqrt{ \cdot }$								
	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)
5	21	1.44	77.20	61.16	41.63	73.08	75.31	78.31	73.64	41	2.82	76.10	69.17	40.95	70.30	76.59	75.74	73.97	72	4.95	81.34	73.75	49.67	77.22	78.38	73.61	77.37
6	26	1.79	75.65	69.49	41.85	69.00	68.02	75.72	79.36	51	3.51	80.91	74.72	44.37	75.00	81.98	80.98	80.91	80	5.50	85.09	75.42	53.31	82.98	78.11	77.89	79.05
7	31	2.13	74.02	68.75	47.61	74.93	79.49	72.47	76.12	61	4.19	79.20	72.09	65.42	75.47	79.63	79.12	78.26	85	5.84	83.43	72.34	67.74	81.97	81.46	76.79	79.64
8	36	2.47	77.52	67.51	42.21	68.50	70.68	75.69	75.83	71	4.88	78.97	72.69	57.08	76.37	86.20	77.67	76.88	91	6.25	81.67	72.36	60.78	81.16	82.18	77.49	75.51
9	37	2.54	76.73	66.08	43.23	79.34	79.48	72.21	74.96	72	4.95	79.97	74.04	57.63	78.24	78.16	77.37	76.50	92	6.32	86.65	73.07	62.36	79.16	79.97	75.86	76.30
10	42	2.89	77.92	69.64	42.60	75.80	80.89	73.74	72.19	82	5.64	76.91	77.93	60.89	79.97	80.92	76.04	79.83	101	6.94	84.79	74.82	60.27	85.01	82.57	79.47	78.73
11	47	3.23	81.75	71.66	42.76	76.85	72.66	74.36	74.79	92	6.32	87.01	81.07	63.83	78.87	77.55	77.55	82.10	106	7.29	88.66	76.87	67.68	79.91	87.77	80.06	77.91
12	52	3.57	83.89	74.06	44.05	78.19	78.33	79.19	77.33	102	7.01	85.96	78.20	71.54	81.89	85.81	79.97	80.86	115	7.90	85.60	77.39	67.01	86.04	85.52	81.42	80.60
13	57	3.92	86.12	75.25	43.85	77.54	73.03	77.40	74.27	111	7.63	86.68	78.65	71.50	78.72	85.12	79.61	79.21	118	8.11	85.79	77.26	69.41	81.30	89.23	85.04	79.66
14	62	4.26	81.41	79.11	42.14	79.40	71.72	78.61	77.76	121	8.32	85.53	79.99	72.56	83.28	84.71	80.66	80.93	120	8.25	88.31	80.97	64.57	85.09	88.46	84.49	82.02
15	64	4.40	85.12	79.37	45.00	80.95	82.31	77.79	77.66	123	8.45	83.48	79.88	73.35	81.61	80.93	83.03	82.17	121	8.32	86.81	80.88	63.49	85.91	88.31	83.13	82.98
16	69	4.74	85.57	78.93	47.91	78.72	82.40	77.42	78.80	133	9.14	88.43	78.82	76.55	84.42	87.07	83.66	83.40	126	8.66	86.91	77.20	67.57	81.94	90.59	84.57	82.47
17	74	5.09	83.35	81.90	45.18	78.49	85.08	84.65	83.50	143	9.83	85.52	82.39	76.22	87.42	84.53	85.75	88.44	131	9.00	86.56	82.40	62.76	87.46	90.03	85.88	87.08
18	77	5.29	82.66	82.08	47.02	84.54	81.57	83.38	83.06	143	9.83	86.59	83.00	75.53	86.20	90.55	86.13	88.12	131	9.00	86.86	83.23	62.92	87.01	87.24	85.57	87.31
19	82	5.64	83.25	82.88	47.78	84.92	85.43	82.30	80.68	150	10.31	85.06	83.52	78.70	85.59	88.28	86.05	86.76	132	9.07	85.41	82.54	56.76	86.17	92.37	85.79	85.94
20	87	5.98	85.45	83.85	44.66	79.82	79.75	81.07	82.00	160	11.00	88.03	85.33	72.20	87.10	83.40	84.25	87.05	137	9.42	90.21	83.08	60.55	88.54	88.24	85.96	86.57
21	92	6.32	86.57	83.71	47.91	83.35	86.72	81.73	82.15	170	11.68	87.47	84.12	72.53	85.84	89.42	84.98	87.57	139	9.55	89.44	84.65	66.41	85.26	91.26	87.16	85.79
22	97	6.67	84.76	82.25	63.99	86.38	85.35	83.65	84.21	180	12.37	88.55	84.63	87.06	85.57	88.94	86.90	89.66	143	9.83	87.88	83.69	69.66	81.55	89.02	88.57	86.81
23	102	7.01	88.62	83.22	64.97	85.29	84.55	83.67	81.60	185	12.71	87.56	84.25	87.64	84.96	89.76	87.17	89.64	144	9.90	89.32	84.06	71.17	86.19	87.19	87.95	88.10
24	107	7.35	87.91	83.83	73.37	83.98	92.06	86.28	83.46	195	13.40	88.73	84.44	85.71	86.11	90.24	87.54	92.25	148	10.17	89.90	83.70	72.38	89.29	89.90	89.90	87.76
25	112	7.70	85.85	84.14	76.10	84.96	87.71	86.67	83.77	205	14.09	90.24	84.40	84.00	87.52	86.96	89.04	89.95	153	10.52	87.86	84.79	71.04	87.94	87.86	88.48	88.79
26	114	7.84	86.65	84.49	72.04	84.94	89.11	88.44	84.71	207	14.23	91.75	84.62	86.38	85.42	88.78	88.70	90.42	154	10.58	90.08	84.40	72.18	85.70	91.70	88.62	89.55
27	119	8.18	82.78	84.28	74.48	86.15	90.04	86.68	84.28	217	14.91	90.55	82.63	85.46	85.22	91.92	88.45	89.69	159	10.93	83.87	83.33	71.60	89.27	89.97	89.74	87.81
28	124	8.52	88.88	84.45	71.07	84.67	86.93	86.70	84.97	227	15.60	91.37	84.28	86.73	85.59	93.08	88.44	90.34	164	11.27	90.78	84.35	77.30	86.44	90.70	89.78	87.99
29	129	8.87	87.33	84.62	62.29	85.67	90.12	88.31	87.24	237	16.29	90.80	83.99	84.56	87.19	92.12	89.08	91.00	169	11.62	88.18	82.97	76.21	88.34	90.36	87.64	88.41
30	130	8.93	82.94	85.06	66.79	85.21	91.09	90.26	86.48	238	16.36	89.81	84.55	83.98	87.02	92.11	88.66	91.73	170	11.68	84.98	83.42	78.83	87.86	89.96	88.79	87.70
31	135	9.28	84.55	85.38	75.23	87.20	89.39	87.73	86.89	245	16.84	90.66	84.96	85.54	88.60	90.08	89.50	92.92	172	11.82	87.69	83.24	73.89	89.40	89.24	89.01	89.01
32	140	9.62	87.68	85.32	73.69	87.38	91.63	88.90	87.14	255	17.53	89.83	86.17	89.00	87.58	90.67	90.00	91.56	174	11.96	86.73	84.47	72.99	87.35	89.54	88.91	89.46
33	143	9.83	90.09	85.67	77.36	89.48	89.02	89.94	87.80	258	17.73	90.48	86.13	87.22	91.48	91.56	89.64	91.87	175	12.03	89.06	85.62	71.09	86.48	90.47	88.05	88.44
34	148	10.17	87.53	86.00	66.49	87.76	89.36	89.06	87.00	268	18.42	89.39	86.77	86.02	89.39	90.14	90.23	90.63	178	12.23	88.80	84.34	71.10	86.06	88.72	88.57	90.13
35	153	10.52	87.17	84.33	79.49	86.64	91.47	88.33	87.02	278	19.11	91.67	86.15	85.98	86.83	91.50	89.72	90.97	181	12.44	91.29	86.81	77.71	88.38	92.23	88.54	88.15
36	153	10.52	87.40	85.87	78.03	83.72	90.48	88.63	87.47	278	19.11	89.72	86.92	87.94	88.62	92.27	88.95	91.05	182	12.51	89.24	85.47	75.65	86.80	89.47	88.61	88.22
37	158	10.86	88.51	84.97	76.64	87.66	90.98	87.82	87.19	288	19.79	89.72	86.89	87.40	88.35	90.23	89.97	90.64	186	12.78	89.99	85.42	77.15	89.52	89.05	88.57	88.65
38	162	11.13	90.49	84.22	78.11	87.86	88.94	88.86	87.10	289	19.86	90.99	85.51	86.62	89.19	92.80	89.11	91.13	187	12.85	92.51	86.20	75.32	87.62	89.67	88.96	88.80
39	167	11.48	85.25	84.24	73.60	87.73	93.17	89.98	87.20	299	20.55	90.22	86.33	89.45	89.27	94.81	89.79	91.91	192	13.20	85.43	84.96	78.54	86.06	89.23	89.94	88.20
40	172	11.82	89.71	84.02	71.94	88.23	87.22	89.87	86.76	309	21.24	91.80	85.25	88.13	89.44	92.50	89.44	90.28	194	13.33	89.69	85.09	76.21	87.63	91.51	89.14	87.23
41	177	12.16	91.55	83.33	77.07	86.15	90.92	89.98	87.41	319	21.92	94.28	83.27	89.88	89.00	94.81	92.61	91.43	195	13.40	91.19	83.65	78.73	87.46	92.46	90.16	88.73
42	181	12.44	89.40	83.20	71.90	86.89	90.89	91.21	87.46	323	22.20	90.99	83.30	89.93	88.34	96.02	91.78	91.23	197	13.54	91.18	84.50	75.04	87.36	91.65	90.54	87.84
43	186	12.78	90.07	82.27	72.50	88.26	90.94	91.73	86.63	333	22.89	93.05	83.33	89.75	89.13	92.78											

Tabela 11 – Resultados de acurácia (em percentagem 100%) da rotulação para o conjunto DS4 com 102 características: nas linhas de cima para baixo são indicados o número de agrupamentos (k). Nas colunas de esquerda para direita são indicados 3 grupos de colunas para indicar os números de amostras por agrupamento (5, 10 e $\sqrt{|\cdot|}$). Em cada grupo de colunas de esquerda para direita são indicados: o número de amostras no total (p), a percentagem (%) em relação ao total de instâncias do conjunto de dados, e finalmente a acurácia para os métodos de extração de amostras *random* (r), *medoid* (m), *contour* (c) e suas combinações (rm , rc , mc e rmc).

k	pk = 5									pk = 10									pk = $\sqrt{ \cdot }$								
	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)	(p)	(%)	(r)	(m)	(c)	(rm)	(rc)	(mc)	(rmc)
5	21	1.26	75.32	71.30	59.35	87.93	81.96	83.67	76.05	41	2.47	88.83	80.32	63.42	86.37	87.85	88.28	87.54	79	4.75	93.11	89.45	67.53	90.40	83.83	89.77	89.07
6	26	1.56	84.96	87.65	74.21	87.35	86.25	88.88	86.19	51	3.07	89.01	89.26	67.91	89.20	85.04	88.45	90.13	88	5.29	87.80	89.45	70.90	91.17	89.45	89.58	88.88
7	31	1.87	89.39	86.82	74.19	81.73	86.45	87.06	88.11	61	3.67	88.82	87.38	71.08	88.94	88.13	87.26	89.19	96	5.78	92.53	89.72	74.97	90.10	88.51	89.91	88.19
8	32	1.93	88.22	86.75	75.58	86.44	87.91	88.10	88.83	62	3.73	88.00	88.31	71.19	89.75	89.56	87.19	87.75	97	5.84	91.95	89.90	74.89	91.57	87.73	89.58	88.37
9	33	1.99	88.95	87.42	72.07	87.91	84.04	87.29	86.62	63	3.79	89.56	88.87	72.80	91.43	92.18	87.93	88.56	98	5.90	91.69	89.90	78.01	89.07	88.68	89.19	88.43
10	38	2.29	86.70	86.70	74.01	85.71	87.81	86.76	87.13	73	4.39	92.26	89.49	72.69	88.92	89.80	86.53	88.29	103	6.20	91.66	89.80	75.88	90.19	84.93	89.80	88.77
11	43	2.59	87.40	88.57	70.23	88.20	91.85	86.97	88.88	83	4.99	90.50	90.18	79.35	86.32	87.21	88.66	89.11	110	6.62	91.56	89.88	80.93	92.85	88.92	89.95	88.72
12	48	2.89	87.61	89.10	69.08	89.53	89.28	87.92	87.48	93	5.60	91.14	91.27	77.69	88.72	90.57	87.38	88.53	118	7.10	93.39	89.51	80.05	89.96	92.03	88.99	87.44
13	53	3.19	86.89	87.76	70.91	87.82	89.75	87.94	88.88	103	6.20	93.07	88.33	75.82	90.25	91.85	88.58	90.25	123	7.40	91.10	90.06	78.69	91.62	89.60	89.93	88.24
14	56	3.37	85.55	88.73	70.55	88.04	89.73	88.61	89.17	106	6.38	91.52	89.27	75.90	90.36	91.26	88.75	90.23	124	7.46	90.31	89.73	78.41	90.70	89.40	90.77	89.86
15	61	3.67	91.01	88.07	68.21	88.57	90.38	87.76	88.09	112	6.74	91.94	89.03	76.39	92.26	89.42	89.16	90.46	125	7.52	94.08	89.92	80.81	90.63	92.52	89.07	90.24
16	66	3.97	90.04	88.10	67.11	89.91	88.10	87.53	88.81	122	7.34	92.86	89.29	75.06	90.84	86.95	87.40	88.97	129	7.76	94.13	89.69	78.41	91.52	88.00	89.24	89.30
17	71	4.27	90.76	88.31	69.96	87.99	90.57	88.18	88.60	128	7.70	92.44	90.22	83.31	90.35	89.90	87.22	89.26	130	7.82	92.69	89.03	78.98	91.19	89.43	89.75	89.36
18	76	4.57	89.91	88.02	71.25	89.85	89.66	88.40	89.57	138	8.30	92.26	89.63	84.19	90.03	89.90	88.91	89.65	136	8.18	92.60	89.58	80.14	92.92	93.58	89.84	89.45
19	77	4.63	90.35	89.09	72.56	90.09	86.75	89.09	89.43	138	8.30	90.29	89.83	84.91	89.11	91.67	89.17	88.79	136	8.18	93.25	89.25	79.55	91.22	88.86	89.78	89.45
20	82	4.93	89.62	89.43	72.91	88.80	89.05	88.86	89.21	148	8.90	91.61	90.36	85.80	90.55	88.84	88.97	88.98	139	8.36	91.99	89.76	80.83	92.06	89.76	90.54	89.43
21	87	5.23	91.75	89.33	72.13	88.83	86.48	88.32	89.11	158	9.51	91.76	89.69	83.78	91.22	87.97	88.56	89.50	143	8.60	91.77	89.47	78.87	90.78	91.31	89.86	89.20
22	90	5.42	89.76	88.74	72.26	90.20	88.93	88.30	88.09	161	9.69	90.54	89.61	82.41	90.01	90.54	88.21	88.68	144	8.66	91.57	89.79	78.19	91.50	91.11	89.72	89.72
23	95	5.72	89.79	89.47	69.94	90.17	90.04	88.77	88.43	171	10.29	90.01	89.87	79.75	91.82	91.35	87.86	88.34	149	8.97	92.33	89.69	79.64	90.42	89.56	90.22	89.23
24	100	6.02	92.13	89.31	71.19	89.69	86.68	88.86	88.27	181	10.89	93.38	89.20	82.11	90.14	91.15	87.71	88.66	153	9.21	92.45	89.66	80.05	92.98	91.72	90.19	89.33
25	105	6.32	91.84	89.92	74.63	92.81	88.82	88.89	89.76	191	11.49	92.86	90.28	85.93	91.30	90.75	88.31	88.86	158	9.51	92.95	89.83	82.58	91.82	91.16	89.76	89.63
26	110	6.62	91.37	89.37	73.45	88.92	89.95	88.72	88.96	201	12.09	93.16	89.05	84.26	89.94	90.83	89.05	88.58	164	9.87	94.19	89.65	79.17	91.72	92.92	89.25	88.58
27	115	6.92	92.50	91.14	76.28	91.47	91.21	89.59	91.11	211	12.70	93.38	91.32	84.36	93.31	91.59	90.01	90.08	167	10.05	93.78	91.10	81.07	91.44	91.84	90.70	90.70
28	120	7.22	91.50	90.73	80.87	92.48	91.63	89.56	90.89	221	13.30	92.44	91.46	89.31	92.44	92.09	90.56	91.05	170	10.23	94.64	90.75	86.33	92.36	91.09	91.15	90.75
29	125	7.52	92.52	91.87	82.69	92.13	91.67	89.98	91.90	231	13.90	93.36	91.96	90.22	94.27	92.52	90.92	91.62	174	10.47	93.21	90.59	85.82	91.80	91.80	91.60	91.80
30	130	7.82	93.28	92.04	84.40	93.21	91.64	90.14	91.61	241	14.50	94.72	92.61	91.13	92.89	92.33	91.48	92.69	178	10.71	93.73	91.71	85.92	91.98	94.14	92.65	92.39
31	135	8.12	91.49	91.49	83.76	93.32	91.55	90.57	91.85	251	15.10	93.69	92.13	90.86	93.27	93.55	92.56	92.49	182	10.95	95.00	91.55	86.01	93.45	93.24	92.43	92.70
32	140	8.42	91.52	91.79	81.34	92.64	91.13	90.93	91.82	261	15.70	93.36	92.01	90.44	92.79	93.08	92.01	92.02	184	11.07	93.23	90.93	85.86	92.96	91.47	92.49	92.22
33	145	8.72	92.16	92.16	82.53	92.68	92.09	91.56	92.45	271	16.31	93.75	92.67	91.59	91.73	94.61	91.73	92.32	188	11.31	94.84	91.72	86.84	93.62	94.71	92.13	93.15
34	150	9.03	94.51	93.78	84.99	93.92	91.34	91.53	92.82	281	16.91	94.35	93.70	91.46	93.77	94.28	93.77	93.35	193	11.61	95.17	93.33	87.81	91.97	94.55	93.06	93.46
35	155	9.33	92.37	93.30	86.60	92.63	91.24	91.37	93.39	291	17.51	94.09	93.73	92.34	93.65	93.22	93.44	93.52	198	11.91	91.94	92.96	86.89	93.92	93.44	92.55	93.72
36	158	9.51	93.75	93.22	86.24	94.22	91.95	91.89	93.38	294	17.69	95.39	92.98	92.32	94.81	94.30	93.20	93.36	199	11.97	94.12	92.62	87.35	93.64	92.96	92.34	92.75
37	163	9.81	94.73	92.86	86.06	93.06	92.93	92.39	93.69	304	18.29	95.73	93.00	92.12	95.66	95.73	93.30	93.98	202	12.15	94.52	93.08	87.40	94.52	91.99	92.60	93.29
38	164	9.87	94.13	92.92	86.38	92.06	94.19	92.06	93.82	304	18.29	95.73	93.23	93.37	95.66	94.99	93.96	94.19	203	12.21	95.27	92.73	86.63	93.35	92.19	92.53	94.11
39	169	10.17	93.90	93.50	86.20	93.17	93.03	91.43	93.87	314	18.89	95.47	93.25	92.51	94.29	95.55	93.77	94.15	207	12.45	96.08	93.13	85.77	93.54	94.78	93.06	93.54
40	174	10.47	94.62	93.21	89.11	94.29	92.54	91.60	93.98	324	19.49	95.59	93.80	92.90	95.07	94.99	94.39	95.00	210	12.64	94.83	93.25	88.09	94.90	94.56	92.49	94.08
41	176	10.59	93.74	93.54	87.89	92.93	94.48	91.52	93.77	326	19.61	95.28	92.96	93.34	95.36	93.64	94.31	94.69	211	12.70	94.21	93.04	88.42	94.42	93.38	93.11	93.94
42	177	10.65	94.01	93.40	89.43	93.94	92.93	91.92	93.83	327	19.68	94.91	93.26	93.56	94.16	94.08	93.86	94.54	212	12.76	94.07	93.52	87.31	93.59	93.45	93.31	93.93
43	182	10.95	93.65	92.91	87.97	93.85	91.08	92.30	93.68	337	20.28	93.81	93.66	93.51	95.09	94.87	94.04										

Figura 27 – Comparação dos resultados de acurácia (em percentagem 100%) segundo os tipos de características e os tipos de amostras: nos gráficos são ilustrados resultados quando são utilizadas as n primeiras características (linhas contínuas) e quando são utilizadas todas as 102 características (linhas tracejadas). Esta comparação é feita considerando os tipos de amostras r , rc e c , com 10 amostras por agrupamento para os conjuntos de dados DS1(a), DS2(b), DS3(c) e DS4(d). Nos gráficos, o eixo X corresponde aos valores de agrupamentos e eixo Y corresponde aos valores de acurácia.

Fonte: Elaborada pelo autor.

5.3.2 Análises das Visualizações

Nesta seção são analisados visualmente os resultados obtidos nas tabelas 4, 5, 6 e 7. Para estas análises visuais desde os resultados quantitativos das tabelas, foi selecionado arbitrariamente uma configuração de resultados, que corresponde a: $k = 24$ como número de agrupamentos, $pk = 5$ como número de amostras por agrupamento e rc foi selecionado como o método para extração de amostras. Os resultados são apresentados por meio de projeções multidimensionais t-SNE. Dessa forma, na Figura 28 são apresentados os resultados visuais da configuração selecionada, onde podemos verificar visualmente as etapas da metodologia: *Clustering*, *Sampling*, *Listening-annotation* e *Learning-prediction*. Para cada conjunto de dados, os pontos de cores variados nas Figuras 28a, 28e, 28i e, 28m representam os 24 agrupamentos gerados na etapa de *Clustering*. Os pontos coloridos com até 3 cores das Figuras 28b 28f 28j e 28n representam as amostras das etapas de *Sampling*, especificamente as cores representam a interação do usuário da etapa *Listening-annotation*. Os pontos coloridos das Figuras 28c, 28g, 28k e, 28o representam as instâncias com rótulos que foram determinados pela predição na etapa

Learning-prediction. É preciso indicar que para a etapa de aprendizagem foram considerados os rótulos das amostras, e na etapa de predição de rótulos foi utilizada o restante das instâncias (numero total de instâncias menos as instâncias amostra).

Finalmente as cores dos pontos nas Figuras 28d, 28h, 28l e, 28p representam os verdadeiros rótulos das instâncias dos conjuntos de dados. Visualmente podemos apreciar a similitude entre os pares de instâncias rotuladas nas Figuras 28c-28d, 28g-28h, 28k-28l e 28o-28p, que na realidade se traduz como o grau visual dos rótulos na predição em relação aos rótulos verdadeiros respectivamente para os 4 conjuntos de dados.

5.3.3 Framework - Rotulador de Paisagens Acústicas

Um *Framework* foi desenvolvido pela necessidade que surgiu em testar as tarefas de todas etapas da metodologia descrita neste capítulo, também, as etapas da metodologia para a descrição das características do Capítulo 4. O *Framework* esta baseado num conceito de análise visual de dados, para isto, são empregadas projeções multidimensionais como t-SNE e UMAP. Dentro das principais tarefas que se podem desenvolver no *Framework* estão: a descrição das características, onde as características são extraídas por meio de descritores previamente estabelecidos, como já foi descrito no Capítulo 4.

Para determinar a predição dos rótulos a partir do *Framework*, podem-se realizar as tarefas como: gerar agrupamentos, visualizar amostras, garantir a interação do usuário na aprendizagem ativa, por meio de ações da etapa definida como *Listening-annotation* e assim também no *Framework* pode-se visualizar os resultados da predição dos rótulos com relação aos verdadeiros rótulos (*ground truth*) dos conjuntos de dados, isto permitiu fazer uma análise visual na comparação dos resultados. Assim na Figura 29 são apresentadas algumas capturas de tela das interfaces do *Framework*. No primeiro caso (a) para mostrar visualmente os estágios da metodologia *Listening-annotation* e *Learning-prediction*. No segundo caso (b) a interface principal do *Framework Soundscape-X*.

Por outro lado no *Framework* são desenvolvidas estratégias incluindo as projeções multidimensionais como um meio visual para assistir a rotulação de categorias de eventos em um contexto de aprendizado ativo. Isso permite ao usuário especialista interagir com as projeções, os espectrogramas e o som. Desta forma, o método visual tem como objetivo apoiar a tarefas de rotulação, mas também permite conhecer se um conjunto de eventos sonoros (pontos da projeção) ocorrem sequencialmente ou em horários diferentes. Um exemplo da interação visual foi descrito na Figura 30.

O *Framework* foi implementado sob uma arquitetura cliente/servidor, para isto, foi utilizado a *library Tornado*³ na linguagem de programação *python*. Para as tarefas de predição foram implementados módulos que utilizam algoritmos de aprendizado máquina da *library*

³ Disponível em: <https://www.tornadoweb.org/>

Figura 28 – Visualizações dos resultados da predição dos rótulos: as linhas de cima para baixo indicam os conjunto de dados DS1 (anuros, pássaros e insetos), DS2 (anuros e pássaros), DS3 (anuros e insetos) e DS4 (pássaros e insetos). Junto as colunas de esquerda para direita indicam as projeções para: *Clustering*; *Sampling* e *Listening-annotation*; *Learning-prediction*; *Ground-truth*. Indicando o número de agrupamentos (k), o número de amostras no total (p), a acurácia alcançada na predição (AC) e finalmente o número de amostras definidas por grupo (pk) para os 4 conjuntos de dados é igual a 5.

Fonte: Elaborada pelo autor.

*scikit-learn*⁴. A descrição das características foram computadas utilizando as *libraries Librosa*⁵, *Essentia*⁶ e também *Pandas*⁷ para algumas tarefas de limpeza de dados nas características. Do lado do cliente, as visualizações foram implementadas utilizando a biblioteca *D3.js* na linguagem *JavaScript*.

5.4 Considerações Finais

Neste capítulo foi descrita uma abordagem para rotular dados de paisagem acústica, que responde ao cumprimento do segundo objetivo principal da pesquisa. A metodologia proposta envolveu a definição das etapas: *Clustering*, *Sampling*, *Listening-annotation* e *Learning-prediction*. Nos experimentos a metodologia foi executada 3864 vezes, nesses experimentos foram avaliados os seguintes parâmetros: o número de agrupamentos, o número de amostras por agrupamento e o tipo método para extrair amostras.

Na etapa de *Clustering* foram testados os métodos *Hierarchical Agglomerative Clustering* e *K-Means*, destaca-se os resultados alcançados por *Hierarchical Agglomerative Clustering*. Na seleção de amostras foram testados sete métodos (r, m, c, rm, rc, mc e rmc) sendo evidenciado como estratégia adequada para determinar amostras mais representativas os métodos r e rc. A técnica de aprendizado que mostrou maior eficácia no tratamento de dados de paisagens acústicas foi a RFC, sendo que foi testada também a SVC, KNNC ou XBGC. Os testes foram executados com a seleção de características descrita no capítulo 4, e não foi apresentada melhora significativa no uso de todas as características em comparação às características selecionadas como as mais discriminantes.

⁴ Disponível em: <https://scikit-learn.org/>

⁵ Disponível em: <https://librosa.github.io/>

⁶ Disponível em: <https://essentia.upf.edu/>

⁷ Disponível em: <https://pandas.pydata.org/>

Figura 29 – Capturas de tela de suas interfaces do *Framework Soundscape-X*.

(a) Com fines ilustrativos a partir do conjunto de dados DS1 (anuros, pássaros e insetos) é apresentado no lado esquerdo os estágios de *Sampling* e *Listening-annotation* evidenciando que as amostras foram identificadas e depois rotuladas pelo usuário experto, na direita o estágio *Learning-prediction* que apresenta a visualização de resto dos áudios como resultado da predição de rótulos.

(b) A interface principal está composta por: A) Inicia o processo no painel de configuração com o conjunto de dados a trabalhar, conjunto de características, tipo de normalização, técnica de agrupamento, finalmente a técnica de visualização; B) Botões que permite interagir com a projeção; C) A região apresenta o espectrograma do total dos áudios, na forma de linha de tempo, que foi denominada como *Time-Line-Spectrogram*. (oferece uma visão geral); D) Várias visualizações coordenadas podem ser exploradas como a projeção do agrupamento, depois as amostras selecionadas para serem rotuladas pelo usuário especialista, finalmente se visualiza a predição dos rótulos; E) Painel de configuração de tipo de amostras, quantidade de amostras e atualização de rótulos pelo usuário; F) Histograma polar das características de acordo a sua capacidade discriminante; G) Mapeamento por cor das características de acordo o tipo de categoria; H) Tipos de eventos sonoros nos áudios; I) O *ground truth* aonde idealmente espera-se alcançar, assim é apresentado a comparação entre os rótulos da predição (D) e os rótulos verdadeiros (I).

Fonte: Elaborada pelo autor.

Figura 30 – Exemplo de interação visual: (a) sem interação, (b) com interação. A interação é a ação feita por o usuário em selecionar um conjunto de pontos da projeção para serem mapeados de forma sincronizada com a linha de tempo do espectrograma, quando a ação de interação é executada, espaços da linha do tempo localizada na parte superior serão coloridos de cor verde.

Fonte: Elaborada pelo autor.

CONCLUSÕES

O objetivo principal deste trabalho de mestrado foi: desenvolver metodologias para apoio à etiquetagem de categorias de eventos sonoros em paisagens acústicas ecológicas que envolva o computador e o usuário no processo, baseado na identificação de um conjunto de características mais discriminantes.

Esse objetivo central motivou estudos e contribuições nas áreas de análise de características, visualização e rotulação de categoria de eventos em paisagens acústicas. As principais contribuições deste trabalho são apresentadas na Seção 6.1. As limitações dos resultados desta dissertação são apresentadas na Seção 6.2. Os desenvolvimentos futuros que podem dar continuidade à abordagem proposta são apresentados na Seção 6.3.

6.1 Contribuições

A metodologia desenvolvida para análise visual de características, nos permitiram identificar e conhecer um conjunto de características com maior capacidade discriminatória. As estratégias visuais (histogramas, *boxplots* e *scatterplots*) utilizadas nos primeiros estágios permitiu tomar a decisão de reduzir ou eliminar um numero grande de características (136) as quais eram de valores contantes ou igual a zero, isso reduziu a carga computacional no processamento dos conjuntos de dados. Os modelos de classificação utilizados além de permitir conhecer as características mais discriminantes pelos valores de acurácia alcançado também deram a conhecer certos aspectos e utilidades das características encontradas.

O *Support Vector Classifier* (SVC) sempre corta o espaço por hiperplanos, isso indica que, se o espaço dos características estiver muito bem formado e não apresentar sobreposição, ele gera resultados satisfatórios. Portanto, como foi observado no conjunto de dados DS3 as características mais discriminantes (primeiros 23) identificados com SVC fornecem informações relevantes ao classificar anuros e insetos, sem recorrer a métodos mais avançados, como RF. Por

outro lado, pela maneira como o modelo *K-Nearest Neighbor classifier* (KNNC) executa a tarefa de classificação, podemos inferir que, as características identificadas como discriminantes pelo modelo, podem ser usados em tarefas de recuperação baseado em conteúdo. No conjunto de dados DS4 o KNNC alcançou acurácia maior junto com o RFC, em diante essas características identificadas poderiam ser utilizadas em tarefas de recuperação para esse tipo de eventos sonoros. Neste estudo evidenciou-se que o *Extreme-Gradient Boosting Classifier* (XGBC) não é um modelo eficiente com dados que contem ruído. Por fim, no caso de *Random Forest Classifier* (RFC) conhece-se que particiona o espaço de forma recursiva, fazendo cortes ortogonais, isso nos leva a inferir que o RF captura melhor as informações quando as classes de eventos sonoros são mais difundidas, justificando os valores de acurácia alcançados neste estudo.

Por meio da metodologia desenvolvida de aprendizado visual ativo, permitiu-se também avaliar o poder discriminatório das características, como apoio às tarefas de rotulação por segregação de categoria de evento acústico. Durante a avaliação, foi identificada uma configuração de parâmetros ótima para a metodologia que indicam melhores resultados nos dados tratados em tarefas de rotulação. A estratégia apresentada nomeada como *contours* e especificamente a combinação de *random* e *contours* como seleção de amostra evidenciou-se como estratégias de seleção de amostras mais informativas no estudo, isto devido a que as amostras aleatórias y as adquiridas do contorno dos *clusters* são as mais apropriadas.

A inclusão de técnicas de visualização nos processos de aprendizado ativo em dados de paisagens acústicas permitiram extrair melhor o conhecimento em cada um dos estágios para a rotulação dos dados, além disso, a interação incluindo as projeções multidimensionais para assistir a rotulação de amostras representativas pelo usuário especialista. Uma das contribuições do trabalho foi o desenvolvimento de estratégias visuais interativas incluindo as projeções multidimensionais para assistir a rotulação de categorias de eventos em um contexto de aprendizado ativo. O desenvolvimento destas estratégias visuais constaram do estudo, projeto e implementação de um contexto visual para que o usuário especialista possa interagir com as projeções, os espectrogramas e o som.

As metodologias de análise de características e aprendizado visual ativo, não somente permitiram gerar informação adequada para obter melhores resultados. Os experimentos das metodologias desenvolvidas permitiram também gerar informações relevantes para futuras pesquisas teriam em consideração evitar certos procedimentos. Especificamente no contexto dos dados tratados neste trabalho, características identificadas como as mais discriminantes com KNNC podem ser utilizadas em trabalhos de recuperação por conteúdo sonoro, também as características identificadas com os diversos modelos de classificação podem ser utilizadas para treinar uma rede neural porque o trabalho fornece uma análise ampla e variada de características (*features*) das paisagens acústicas.

No desenvolvimento das metodologias para análise visual de características e aprendizado visual ativo para a rotulação de categoria de eventos, geraram informação para o desenvolvimento

de um *Framework* nomeado como *Soundscape-X*. Dentre as principais tarefas que podem ser executadas no *Framework* são: extração de características com descritores estabelecidos do tipos espectral, imagem, e índices acústicos, que podem gerar um conjunto de 238 características; tarefas de seleção de amostras com 7 tipos de métodos; visualização, interatividade na rotulação das amostras de som pelos usuários, acessíveis por meio de projeções multidimensionais; treino de um modelo de aprendizado baseado nas amostras; predição de rótulos de novas instâncias com o intuito de segregar categorias de eventos acústicos; e visualização do ranking de características.

Finalmente, são apresentados os resultados desta pesquisa na elaboração dos seguintes artigos:

- Huancapaza, L. H. ; Gaspar, L. P. ; Ribeiro, M. C. ; Minghim, R. Visualization and categorization of ecological acoustic events based on discriminant features. Journal **Ecological Indicators** (*submetido*).
- Huancapaza, L. H. ; Minghim, R. Visual active learning for labeling soundscape data. (*sendo finalizado*).

6.2 Limitações

A principal limitação da abordagem está baseada na escolha de categorias de eventos, onde não há garantias que as mesmas diretrizes que funcionam para um tipo (categoria de animais como aves, anuros ou insetos) irá funcionar para outras, por exemplo categoria de condições climatológicas como chuva ou vento; categoria de faixa de tempo como dia ou noite. Futuras pesquisas são necessárias para testar a diferença de discriminação entre eventos mais gerais e mais específicos. Um exemplo é reconhecer as espécies presentes dentro da categoria de aves, ou de insetos. Esse tipo de segregação de sub-categorias de eventos não foi abordada na pesquisa, por não ser parte dos objetivos.

Para alcançar melhores resultados na segregação de categorias de eventos acústicos, a metodologia de aprendizado visual ativo está preparada para que o usuário especialista possa repetir as etapas de *Clustering*, *Listening-Annotation* e *Learning-Prediction* de modo que o usuário possa avaliar variações em cada etapa e em cada repetição da metodologia. No entanto, no contexto de repetições da metodologia, não está contemplada a análise de informações tais como: quais amostras foram as mais representativas, quais amostras apresentam uma maior probabilidade de serem rotuladas incorretamente, que relação tem essas informações com a importância das características e a escalabilidade da metodologia na previsão de rótulos. Fazer uma análise dessas informações de forma iterativa e incremental, poderia ser determinante para melhorar os resultados de acurácia. Também é importante notar que, enquanto o foco deste trabalho esteve no conjunto de dados aqui analisado (de anuros, insetos e aves e suas combinações), não há garantias de que as mesmas características sejam tão discriminantes em

outros contextos. A metodologia, entretanto, é um bom ponto de partida para outros conjuntos de dados onde a tarefa é discriminar e identificar eventos na mesma escala dos eventos aqui tratados.

6.3 Trabalhos Futuros

Futuras pesquisas poderiam incluir o desenvolvimento de novas metodologias ou estratégias para segregar sub-categorias de eventos acústicos. Para a realização desses trabalhos, pode-se experimentar a execução recursiva das metodologias já desenvolvidas. Ou seja, a segregação de sub-categorias de eventos poderia ser obtida pela execução dos mesmos métodos de análise visual de características e aprendizado visual ativo, porém sobre as instâncias identificadas para cada categoria de evento acústico. Questões sobre como avaliar essa proposta de trabalho e como realizar as interações dos usuários poderiam ser os principais temas de pesquisa.

Com o intuito de melhorar o processo de aprendizagem desde o início da metodologia, futuras pesquisas deveriam incluir outras estratégias para a seleção de amostras. Estas estratégias teriam com objetivo antecipar e avaliar a probabilidade das instâncias serem rotuladas incorretamente. Ao identificar as instâncias com maior probabilidade em serem rotuladas incorretamente, poderia-se garantir uma melhoria no processo de treinamento do modelo de aprendizagem, e pelo tanto, incrementar a acurácia na segregação de categorias de eventos acústicos.

Futuros trabalhos onde se incorporem iterações entre as etapas de *Clustering*, *Listening-Annotation* e *Learning-Prediction* para melhorar o processo de aprendizagem e corrigir gradualmente os rótulos de instâncias provavelmente rotuladas de forma incorreta, poderia melhorar a tarefa de segregar categoria de eventos acústicos. Poderiam-se incluir estratégias baseadas em conceitos de incertezas e estratégias visuais para visualizar as instâncias rotuladas incorretamente nos processos iterativos.

Futuros trabalhos também poderiam explorar a possibilidade de incluir outros descritores de características, como por exemplo características extraídas por meio de metodologias baseadas em aprendizado profundo - *Deep Learning* ou em redes complexas - *Complex Networks*. Por outro lado, para a tarefa de treinamento da aprendizagem, além dos modelos de aprendizagem utilizados como *Random Forest Classifier* e *Support Vector Classifier*, poderiam-se incluir outros modelos de aprendizagem, como por exemplo modelos baseados no aprendizado profundo. Para considerar a inclusão de qualquer outro modelo no treinamento da aprendizagem, teria que ser considerada também a compatibilidade de interação com o aprendizado ativo.

REFERÊNCIAS

- AERTSEN, A. M. H. J.; JOHANNESMA, P. I. M.; HERMES, D. J. Spectro-temporal receptive fields of auditory neurons in the grassfrog. **Biological Cybernetics**, v. 38, n. 4, p. 235–248, Nov 1980. ISSN 1432-0770. Disponível em: <https://doi.org/10.1007/BF00337016>. Citado na página 40.
- Agrawal, D. M.; Sailor, H. B.; Soni, M. H.; Patil, H. A. Novel teo-based gammatone features for environmental sound classification. In: **2017 25th European Signal Processing Conference (EUSIPCO)**. [S.l.: s.n.], 2017. p. 1809–1813. ISSN 2076-1465. Citado nas páginas 40, 52 e 60.
- Bachthaler, S.; Weiskopf, D. Continuous scatterplots. **IEEE Transactions on Visualization and Computer Graphics**, v. 14, n. 6, p. 1428–1435, Nov 2008. Citado na página 64.
- BARROS, F.; MARTELLO, F.; PERES, C.; PIZO, M.; RIBEIRO, M. Matrix type and landscape attributes modulate avian taxonomic and functional spillover across habitat boundaries in the brazilian atlantic forest. **Oikos**, v. 128, 06 2019. Citado nas páginas 14 e 68.
- BOGERT, B. P. The quefrency analysis of time series for echoes : cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In: . [S.l.: s.n.], 1963. Citado nas páginas 37 e 40.
- BOSCOLO, D.; TOKUMOTO, P. M.; FERREIRA, P. A.; RIBEIRO, J. W.; SANTOS, J. S. dos. Positive responses of flower visiting bees to landscape heterogeneity depend on functional connectivity levels. **Perspectives in Ecology and Conservation**, v. 15, n. 1, p. 18 – 24, 2017. ISSN 2530-0644. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2530064417300329>. Citado na página 68.
- BREIMAN, L. Random forests. **Mach. Learn.**, Kluwer Academic Publishers, USA, v. 45, n. 1, p. 5–32, out. 2001. ISSN 0885-6125. Disponível em: <https://doi.org/10.1023/A:1010933404324>. Citado na página 66.
- CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. (Ed.). **Readings in Information Visualization: Using Vision to Think**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999. ISBN 1-55860-533-9. Citado na página 29.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining**. [S.l.: s.n.], 2016. p. 785–794. Citado na página 66.
- CHOU, C.-H.; LEE, C.-H.; NI, H.-W. Bird species recognition by comparing the hmms of the syllables. In: . [S.l.: s.n.], 2007. p. 143–143. ISBN 0-7695-2882-1. Citado na página 28.
- DEKA, R. K.; BHATTACHARYYA, D. K.; KALITA, J. K. Active learning to detect ddos attack using ranked features. **Computer Communications**, v. 145, p. 203 – 222, 2019. ISSN 0140-3664. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0140366419303858>. Citado na página 43.

DEPRAETERE, M.; PAVOINE, S.; JIGUET, F.; GASC, A.; DUVAIL, S.; SUEUR, J. Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland. **Ecological Indicators**, v. 13, n. 1, p. 46 – 54, 2012. ISSN 1470-160X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1470160X11001361>>. Citado na página 38.

DIAS, F. F. **Uma estratégia para análise visual de Paisagens Acústicas com base em seleção de características discriminantes**. Dissertação (Mestrado) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos - SP, 2018. Citado nas páginas 29, 51, 52, 58 e 60.

DIAS, F. F.; MINGHIM, R. xhipp: extended hierarchical point placement strategy. In: ROSS, A.; GASTAL, E. S. L.; JORGE, J. A.; QUEIROZ, R. L. d.; MINETTO, R.; SARKAR, S.; PAPA, J. P.; OLIVEIRA, M. M.; ARBELÁEZ, P.; MERY, D.; OLIVEIRA, M. C. F. d.; SPINA, T. V.; MENDES, C. M.; COSTA, H. S. G.; MEJAIL, M. E.; GEUS, K. d.; SCHEER, S. (Ed.). **Proceedings...** 2018. Disponível em: <<http://urlib.net/rep/8JMKD3MGPAW/3RPBD6H>>. Acesso em: 2018, Nov. 07. Citado nas páginas 56, 57, 59 e 60.

ESPADOTO, M.; HIRATA, N. S. T.; TELEA, A. C. Deep learning multidimensional projections. **CoRR**, abs/1902.07958, 2019. Disponível em: <<http://arxiv.org/abs/1902.07958>>. Citado na página 30.

EVEREST, F. A. Master handbook of acoustics, fourth edition. **The Journal of the Acoustical Society of America**, v. 110, n. 4, p. 1714–1715, 2001. Disponível em: <<https://doi.org/10.1121/1.1398048>>. Citado na página 33.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. F. d. **Inteligência artificial: uma abordagem de aprendizado de máquina**. [S.l.]: LTC, 2011. Citado na página 64.

FARINA, A. **Principles and Methods in Landscape Ecology**: Towards a science of the landscape. [S.l.]: Springer Netherlands, 2006. Citado na página 28.

Fukunaga, K.; Narendra, P. M. A branch and bound algorithm for computing k-nearest neighbors. **IEEE Transactions on Computers**, C-24, n. 7, p. 750–753, 1975. Citado na página 66.

GASC, A.; SUEUR, J.; JIGUET, F.; DEVICTOR, V.; GRANDCOLAS, P.; BURROW, C.; DEPRAETERE, M.; PAVOINE, S. Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? **Ecological Indicators**, v. 25, p. 279 – 287, 2013. ISSN 1470-160X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1470160X12003603>>. Citado nas páginas 49 e 60.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2010. ISBN 013168728X. Citado nas páginas 35 e 41.

Göğüş, F. Z.; TEZEL, G. Comparison of lpc based parametric techniques for respiratory sounds recognition. **International Journal of Signal Processing Systems**, v. 6, n. 1, p. 6–11, 2018. Citado na página 41.

HAEMER, K. W. Range-bar charts. **The American Statistician**, v. 2, n. 2, p. 23–23, 1948. Citado na página 64.

- HALDER, A.; KUMAR, A. Active learning using rough fuzzy classifier for cancer prediction from microarray gene expression data. **Journal of Biomedical Informatics**, v. 92, p. 103136, 2019. ISSN 1532-0464. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046419300541>>. Citado na página 43.
- HAN, N. C.; MUNIANDY, S. V.; DAYOU, J. Acoustic classification of australian anurans based on hybrid spectral-entropy approach. **Applied Acoustics**, v. 72, n. 9, p. 639 – 645, 2011. ISSN 0003-682X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0003682X11000314>>. Citado na página 52.
- HAN, W.; COUTINHO, E.; RUAN, H.; LI, H.; SCHULLER, B.; YU, X.; ZHU, X. Semi-supervised active learning for sound classification in hybrid learning environments. **PLOS ONE**, Public Library of Science, v. 11, n. 9, p. 1–23, 09 2016. Disponível em: <<https://doi.org/10.1371/journal.pone.0162075>>. Citado nas páginas 30 e 53.
- HARALICK, R. M. Statistical and structural approaches to texture. **Proceedings of the IEEE**, v. 67, n. 5, p. 786–804, May 1979. ISSN 0018-9219. Citado nas páginas 41 e 42.
- HINTON, G. E.; ROWEIS, S. T. Stochastic neighbor embedding. In: BECKER, S.; THRUN, S.; OBERMAYER, K. (Ed.). **Advances in Neural Information Processing Systems 15**. MIT Press, 2003. p. 857–864. Disponível em: <<http://papers.nips.cc/paper/2276-stochastic-neighbor-embedding.pdf>>. Citado na página 44.
- Hoffman, P.; Grinstein, G.; Marx, K.; Grosse, I.; Stanley, E. Dna visual and analytic data mining. In: **Proceedings. Visualization '97 (Cat. No. 97CB36155)**. [S.l.: s.n.], 1997. p. 437–441. Citado na página 55.
- HU, W.; BULUSU, N.; CHOU, C. T.; JHA, S.; TAYLOR, A.; TRAN, V. N. Design and evaluation of a hybrid sensor network for cane toad monitoring. **ACM Trans. Sen. Netw.**, ACM, New York, NY, USA, v. 5, n. 1, p. 4:1–4:28, fev. 2009. ISSN 1550-4859. Disponível em: <<http://doi.acm.org/10.1145/1464420.1464424>>. Citado na página 28.
- Jaafar, H.; Ramli, D. A.; Shahrudin, S. Mfcc based frog identification system in noisy environment. In: **2013 IEEE International Conference on Signal and Image Processing Applications**. [S.l.: s.n.], 2013. p. 123–127. Citado na página 40.
- Jhawar, G.; Nagraj, P.; Mahalakshmi, P. Speech disorder recognition using mfcc. In: **2016 International Conference on Communication and Signal Processing (ICCSP)**. [S.l.: s.n.], 2016. p. 0246–0250. Citado na página 40.
- JOO, W.; GAGE, S. H.; KASTEN, E. P. Analysis and interpretation of variability in soundscapes along an urban–rural gradient. **Landscape and Urban Planning**, v. 103, n. 3, p. 259 – 276, 2011. ISSN 0169-2046. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169204611002465>>. Citado na página 28.
- KASTEN, E. P.; GAGE, S. H.; FOX, J.; JOO, W. The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. **Ecological Informatics**, v. 12, p. 50 – 67, 2012. ISSN 1574-9541. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S157495411200088X>>. Citado nas páginas 38 e 39.
- KHOLGHI, M.; PHILLIPS, Y.; TOWSEY, M.; SITBON, L.; ROE, P. Active learning for classifying long-duration audio recordings of the environment. **Methods in Ecology and Evolution**, Wiley Online Library, v. 9, n. 9, p. 1948–1958, 2018. Citado nas páginas 43, 53 e 60.

KRAUSE, B. **Wild Soundscapes: Discovering the Voice of the Natural World : a Book and CD Recording**. Wilderness Press, 2002. (Wilderness Press Series). ISBN 9780899972961. Disponível em: <<https://books.google.com.br/books?id=kxW5wAEACAAJ>>. Citado na página 28.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. **Psychometrika**, v. 29, n. 1, p. 1–27, Mar 1964. ISSN 1860-0980. Disponível em: <<https://doi.org/10.1007/BF02289565>>. Citado na página 45.

Lamel, L.; Rabiner, L.; Rosenberg, A.; Wilpon, J. An improved endpoint detector for isolated word recognition. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 29, n. 4, p. 777–785, August 1981. Citado na página 38.

Liu, S.; Maljovec, D.; Wang, B.; Bremer, P.; Pascucci, V. Visualizing high-dimensional data: Advances in the past decade. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 3, p. 1249–1268, March 2017. ISSN 1077-2626. Citado na página 30.

LÖFSTEDT, T.; BRYNOLFSSON, P.; ASKLUND, T.; NYHOLM, T.; GARPEBRING, A. Gray-level invariant haralick texture features. **PLOS ONE**, Public Library of Science, v. 14, n. 2, p. 1–18, 02 2019. Disponível em: <<https://doi.org/10.1371/journal.pone.0212110>>. Citado na página 43.

MAATEN, L.; HINTON, G. Visualizing non-metric similarities in multiple maps. **Mach. Learn.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 87, n. 1, p. 33–55, abr. 2012. ISSN 0885-6125. Disponível em: <<https://doi.org/10.1007/s10994-011-5273-4>>. Citado nas páginas 44 e 45.

MAATEN, L. V. D. Accelerating t-sne using tree-based algorithms. **J. Mach. Learn. Res.**, JMLR.org, v. 15, n. 1, p. 3221–3245, jan. 2014. ISSN 1532-4435. Disponível em: <<http://dl.acm.org/citation.cfm?id=2627435.2697068>>. Citado nas páginas 44 e 45.

MAATEN, L. van der. Learning a parametric embedding by preserving local structure. In: DYK, D. van; WELLING, M. (Ed.). **Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics**. Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA: PMLR, 2009. (Proceedings of Machine Learning Research, v. 5), p. 384–391. Disponível em: <<http://proceedings.mlr.press/v5/maaten09a.html>>. Citado nas páginas 44 e 45.

MAATEN, L. van der; HINTON, G. Visualizing data using t-SNE. **Journal of Machine Learning Research**, v. 9, p. 2579–2605, 2008. Disponível em: <<http://www.jmlr.org/papers/v9/vandermaaten08a.html>>. Citado nas páginas 44, 45 e 66.

MARTEN, K.; QUINE, D.; MARLER, P. Sound transmission and its significance for animal vocalization. **Behavioral Ecology and Sociobiology**, v. 2, n. 3, p. 291–302, Sep 1977. ISSN 1432-0762. Disponível em: <<https://doi.org/10.1007/BF00299741>>. Citado na página 34.

MCINNES, L.; HEALY, J. Umap: Uniform manifold approximation and projection for dimension reduction. **CoRR**, abs/1802.03426, 2018. Citado nas páginas 45, 46, 66 e 74.

MIAO, J.; NIU, L. A survey on feature selection. In: . [S.l.: s.n.], 2016. Citado na página 65.

MITROVIC, D.; ZEPPELZAUER, M.; BREITENEDER, C. Features for content-based audio retrieval. **Advances in Computers**, v. 78, p. 71–150, 2010. Citado nas páginas 28, 36, 37 e 40.

NODA, J. J.; TRAVIESO, C. M.; SÁNCHEZ-RODRÍGUEZ, D. Methodology for automatic bio-acoustic classification of anurans based on feature fusion. **Expert Systems with Applications**, v. 50, p. 100 – 106, 2016. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417415008222>>. Citado na página 52.

Nonato, L. G.; Aupetit, M. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. **IEEE Transactions on Visualization and Computer Graphics**, v. 25, n. 8, p. 2650–2673, Aug 2019. ISSN 2160-9306. Citado na página 30.

ON, C. K.; PANDIYAN, P. M.; YAACOB, S.; SAUDI, A. Mel-frequency cepstral coefficient analysis in speech recognition. In: **2006 International Conference on Computing Informatics**. [S.l.: s.n.], 2006. p. 1–5. ISSN 2166-5710. Citado na página 40.

PARKS, S. E.; MIKSIS-OLDS, J. L.; DENES, S. L. Assessing marine ecosystem acoustic diversity across ocean basins. **Ecological Informatics**, v. 21, p. 81 – 88, 2014. ISSN 1574-9541. Ecological Acoustics. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1574954113001167>>. Citado na página 28.

Paulovich, F. V.; Minghim, R. Hipp: A novel hierarchical point placement strategy and its application to the exploration of document collections. **IEEE Transactions on Visualization and Computer Graphics**, v. 14, n. 6, p. 1229–1236, Nov 2008. ISSN 1077-2626. Citado na página 56.

Paulovich, F. V.; Nonato, L. G.; Minghim, R.; Levkowitz, H. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. **IEEE Transactions on Visualization and Computer Graphics**, v. 14, n. 3, p. 564–575, May 2008. ISSN 1077-2626. Citado na página 44.

PENATTI, O. A.; VALLE, E.; TORRES, R. da S. Comparative study of global color and texture descriptors for web image retrieval. **Journal of Visual Communication and Image Representation**, v. 23, n. 2, p. 359 – 380, 2012. ISSN 1047-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1047320311001465>>. Citado na página 41.

PHILLIPS, Y. F.; TOWSEY, M.; ROE, P. Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation. **PLOS ONE**, Public Library of Science, v. 13, n. 3, p. 1–27, 03 2018. Disponível em: <<https://doi.org/10.1371/journal.pone.0193345>>. Citado nas páginas 29, 51, 57, 58 e 60.

PICZAK, K. J. Esc: Dataset for environmental sound classification. In: **ACM Multimedia**. [S.l.: s.n.], 2015. Citado nas páginas 51 e 52.

PIERETTI, N.; FARINA, A.; MORRI, D. A new methodology to infer the singing activity of an avian community: The acoustic complexity index (aci). **Ecological Indicators**, v. 11, n. 3, p. 868 – 873, 2011. ISSN 1470-160X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1470160X10002037>>. Citado nas páginas 39, 49 e 60.

PIJANOWSKI, B. C.; FARINA, A.; GAGE, S. H.; DUMYAHN, S. L.; KRAUSE, B. L. What is soundscape ecology? an introduction and overview of an emerging new science. **Landscape Ecology**, v. 26, n. 9, p. 1213–1232, Nov 2011. ISSN 1572-9761. Disponível em: <<https://doi.org/10.1007/s10980-011-9600-8>>. Citado nas páginas 27, 28 e 33.

- PIJANOWSKI, B. C.; VILLANUEVA-RIVERA, L. J.; DUMYAHN, S. L.; FARINA, A.; KRAUSE, B. L.; NAPOLETANO, B. M.; GAGE, S. H.; PIERETTI, N. Soundscape ecology: The science of sound in the landscape. **BioScience**, Oxford University Press (OUP), v. 61, n. 3, p. 203–216, mar 2011. Citado nas páginas 28 e 34.
- PLATT, J. C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: **ADVANCES IN LARGE MARGIN CLASSIFIERS**. [S.l.]: MIT Press, 1999. p. 61–74. Citado na página 66.
- QIAN, K.; ZHANG, Z.; BAIRD, A.; SCHULLER, B. W. Active learning for bird sound classification via a kernel-based extreme learning machine. **The Journal of the Acoustical Society of America**, v. 142 4, p. 1796, 2017. Citado na página 30.
- Qian, K.; Zhang, Z.; Ringeval, F.; Schuller, B. Bird sounds classification by large scale acoustic features and extreme learning machine. In: **2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)**. [S.l.: s.n.], 2015. p. 1317–1321. Citado na página 52.
- RAGHURAM, M. A.; CHAVAN, N. R.; BELUR, R.; KOOLAGUDI, S. G. Bird classification based on their sound patterns. **International Journal of Speech Technology**, v. 19, n. 4, p. 791–804, Dec 2016. ISSN 1572-8110. Disponível em: <<https://doi.org/10.1007/s10772-016-9372-2>>. Citado nas páginas 28 e 52.
- REIS, C. D. G.; SANTOS, T. N.; OLIVEIRA, M. C. F. de. A visualization framework for feature investigation in soundscape recordings. **2018 22nd International Conference Information Visualisation (IV)**, p. 490–497, 2018. Citado nas páginas 52, 55, 56, 59 e 60.
- ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53 – 65, 1987. ISSN 0377-0427. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0377042787901257>>. Citado na página 45.
- SALAMON, J.; JACOBY, C.; BELLO, J. P. A dataset and taxonomy for urban sound research. In: **Proceedings of the 22Nd ACM International Conference on Multimedia**. New York, NY, USA: ACM, 2014. (MM '14), p. 1041–1044. ISBN 978-1-4503-3063-3. Disponível em: <<http://doi.acm.org/10.1145/2647868.2655045>>. Citado na página 52.
- SALHI, K.; JAARA, E. M.; ALAOU, M. T. Pretreatment approaches for texture image segmentation. In: **2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)**. [S.l.: s.n.], 2016. p. 221–225. Citado na página 42.
- SANKUPELLAY, M.; TOWSEY, M.; TRUSKINGER, A.; ROE, P. Visual fingerprints of the acoustic environment: The use of acoustic indices to characterise natural habitats. In: **2015 Big Data Visual Analytics (BDVA)**. [S.l.: s.n.], 2015. p. 1–8. Citado nas páginas 27, 29, 54, 55, 59 e 60.
- SANKUPELLAY, M.; TOWSEY, M. W.; TRUSKINGER, A.; ROE, P. Visual fingerprints of the acoustic environment: The use of acoustic indices to characterise natural habitats. **2015 Big Data Visual Analytics (BDVA)**, p. 1–8, 2015. Citado nas páginas 28, 50 e 60.
- Schluter, R.; Bezrukov, I.; Wagner, H.; Ney, H. Gammatone features and feature combination for large vocabulary speech recognition. In: **2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07**. [S.l.: s.n.], 2007. v. 4, p. IV–649–IV–652. Citado na página 40.

SERVICK, K. Eavesdropping on ecosystems. **Science**, American Association for the Advancement of Science, v. 343, n. 6173, p. 834–837, 2014. ISSN 0036-8075. Disponível em: <<http://science.sciencemag.org/content/343/6173/834>>. Citado na página 27.

SETTLES, B. **Active Learning Literature Survey**. [S.l.], 2009. Disponível em: <<http://axon.cs.byu.edu/~martinez/classes/778/Papers/settles.activelearning.pdf>>. Citado nas páginas 30, 43 e 44.

Shannon, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, n. 3, p. 379–423, July 1948. Citado nas páginas 37 e 38.

Shuyang, Z.; Heittola, T.; Virtanen, T. Active learning for sound event classification by clustering unlabeled data. In: **2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2017. p. 751–755. Citado nas páginas 30, 43, 53 e 60.

_____. An active learning method using clustering and committee-based sample selection for sound event classification. In: **2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)**. [S.l.: s.n.], 2018. p. 116–120. Citado nas páginas 43, 53 e 60.

SOUTHWORTH, M. The sonic environment of cities. **Environment and Behavior**, v. 1, n. 1, p. 49–70, 1969. Disponível em: <<https://doi.org/10.1177/001391656900100104>>. Citado na página 28.

STEHLLING, R. O.; NASCIMENTO, M. A.; aO, A. X. F. A compact and efficient image retrieval approach based on border/interior pixel classification. In: **Proceedings of the Eleventh International Conference on Information and Knowledge Management**. New York, NY, USA: ACM, 2002. (CIKM '02), p. 102–109. ISBN 1-58113-492-4. Disponível em: <<http://doi.acm.org/10.1145/584792.584812>>. Citado na página 42.

STEVENS, S. S.; VOLKMANN, J.; NEWMAN, E. B. A scale for the measurement of the psychological magnitude pitch. **The Journal of the Acoustical Society of America**, v. 8, n. 3, p. 185–190, 1937. Disponível em: <<https://doi.org/10.1121/1.1915893>>. Citado na página 40.

STOWELL, D.; PLUMBLEY, M. D. Audio-only bird classification using unsupervised feature learning. In: **CLEF**. [S.l.: s.n.], 2014. Citado nas páginas 28 e 52.

SUEUR, J.; FARINA, A.; GASC, A.; PIERETTI, N.; PAVOINE, S. Acoustic indices for biodiversity assessment and landscape investigation. In: . [S.l.: s.n.], 2014. Citado nas páginas 37, 50 e 60.

SUEUR, J.; PAVOINE, S.; HAMERLYNCK, O.; DUVAIL, S. Rapid acoustic survey for biodiversity appraisal. **PLoS ONE**, v. 3, 2008. Citado na página 38.

_____. Rapid acoustic survey for biodiversity appraisal. **PLOS ONE**, Public Library of Science, v. 3, n. 12, p. 1–9, 12 2009. Disponível em: <<https://doi.org/10.1371/journal.pone.0004065>>. Citado na página 38.

SÁNCHEZ-GENDRIZ, I.; PADOVESE, L. A methodology for analyzing biological choruses from long-term passive acoustic monitoring in natural areas. **Ecological Informatics**, v. 41, p. 1 – 10, 2017. ISSN 1574-9541. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1574954117300201>>. Citado nas páginas 55 e 56.

TEJADA, E.; MINGHIM, R.; NONATO, L. G. On improved projection techniques to support visual exploration of multidimensional data sets. **Information Visualization**, Palgrave Macmillan, v. 2, n. 4, p. 218–231, dez. 2003. ISSN 1473-8716. Disponível em: <<http://dx.doi.org/10.1057/palgrave.ivs.9500054>>. Citado na página 44.

Thwe, K. Z.; War, N. Environmental sound classification based on time-frequency representation. In: **2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)**. [S.l.: s.n.], 2017. p. 251–255. Citado na página 51.

TORRES, J. L. G.; CATANIA, C. A.; VEAS, E. Active learning approach to label network traffic datasets. **Journal of Information Security and Applications**, v. 49, p. 102388, 2019. ISSN 2214-2126. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2214212618304344>>. Citado na página 43.

TOWSEY, M.; WIMMER, J.; WILLIAMSON, I.; ROE, P. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. **Ecological Informatics**, v. 21, p. 110 – 119, 2014. ISSN 1574-9541. Ecological Acoustics. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1574954113001209>>. Citado nas páginas 37 e 39.

_____. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. **Ecological Informatics**, Elsevier, v. 21, p. 110–119, May 2014. Disponível em: <<https://eprints.qut.edu.au/65741/>>. Citado nas páginas 37, 38, 50, 59 e 60.

TOWSEY, M.; ZHANG, L.; COTTMAN-FIELDS, M.; WIMMER, J.; ZHANG, J.; ROE, P. Visualization of long-duration acoustic recordings of the environment. **Procedia Computer Science**, v. 29, p. 703 – 712, 2014. ISSN 1877-0509. 2014 International Conference on Computational Science. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877050914002403>>. Citado nas páginas 27, 54, 59 e 60.

Valero, X.; Alias, F. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. **IEEE Transactions on Multimedia**, v. 14, n. 6, p. 1684–1689, Dec 2012. Citado na página 40.

VILLANUEVA-RIVERA, L. J.; PIJANOWSKI, B. C.; DOUCETTE, J.; PEKIN, B. A primer of acoustic analysis for landscape ecologists. **Landscape Ecology**, v. 26, n. 9, p. 1233, Jul 2011. ISSN 1572-9761. Disponível em: <<https://doi.org/10.1007/s10980-011-9636-9>>. Citado nas páginas 27, 33, 34, 35 e 36.

VILLANUEVA-RIVERA, L. J.; PIJANOWSKI, B. C.; DOUCETTE, J.; PEKIN, B. K. A primer of acoustic analysis for landscape ecologists. **Landscape Ecology**, v. 26, p. 1233–1246, 2011. Citado na página 37.

Wang, F.; Wen, S.; Wu, C.; Zhang, Y.; Wang, H. Gait recognition based on the fast fourier transform and svm. In: **2011 Chinese Control and Decision Conference (CCDC)**. [S.l.: s.n.], 2011. p. 1091–1094. ISSN 1948-9439. Citado na página 35.

Wang, Z.; Haihong, E.; Song, M.; Ren, Z. Time-varying data visual analysis method based on parallel coordinate system. In: **2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)**. [S.l.: s.n.], 2019. p. 1256–1260. ISSN null. Citado na página 30.

XIE, J.; INDRASWARI, K.; SCHWARZKOPF, L.; TOWSEY, M.; ZHANG, J.; ROE, P. Acoustic classification of frog within-species and species-specific calls. **Applied Acoustics**, v. 131, p. 79 – 86, 2018. ISSN 0003-682X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0003682X17304024>>. Citado na página 52.

Xie, J.; Towsey, M.; Zhang, J.; Dong, X.; Roe, P. Application of image processing techniques for frog call classification. In: **2015 IEEE International Conference on Image Processing (ICIP)**. [S.l.: s.n.], 2015. p. 4190–4194. Citado nas páginas 51 e 60.

XIE, J.; TOWSEY, M.; ZHANG, J.; ROE, P. Acoustic classification of australian frogs based on enhanced features and machine learning algorithms. **Applied Acoustics**, v. 113, p. 193 – 201, 2016. ISSN 0003-682X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0003682X16301864>>. Citado nas páginas 28 e 52.