Myllena Caetano de Oliveira
Orientadora: Deborah Maria Vieira Magalhães

Extração Manual De Características Para Classificação De Sons Urbanos

Picos - PI
11 de janeiro de 2021

Myllena Caetano de Oliveira
Orientadora: Deborah Maria Vieira Magalhães

Extração Manual De Características Para Classificação De Sons Urbanos

Monografia submetida ao Curso de Bacharelado em Sistemas de Informação como requisito parcial para obtenção de grau de Bacharel em Sistemas de Informação. Orientadora: Profa. Dra. Deborah Maria Vieira Magalhães.

Universidade Federal do Piauí
Campus Senador Helvídio Nunes de Barros
Bacharelado em Sistemas de Informação

Picos - PI
11 de janeiro de 2021

Agradecimentos

Agradeço primeiramente a Deus, pois sei que tudo o que já tenho alcançado foi por sua graça e misericórdia.

Aos meus pais, Ailton e Rosinery, que sempre forneceram o que eu precisava materialmente, emocionalmente e espiritualmente. Aos meu irmãos Mayra, Matheus e Maryanne, que se esforçaram ao máximo para fazer silêncio enquanto eu estudava.

Aos meus amigos da "Panelinha", que me ajudaram a chegar onde cheguei sempre me dando forças para seguir em frente. Em especial aos meus amigos Patrick e Vitória, que mesmo quando eu já havia desistido eles não desistiram. E ao meu amigo Jederson, companheiro nessa jornada.

A minha orientadora e amiga Deborah, que sempre me ajudou a enxergar do que sou capaz.

E a todos que contribuíram direta ou indiretamente para a conclusão desse trabalho.

Deus os abençoe!

Porque o senhor é bom, e eterna, a sua misericórdia; e a sua verdade estende-se de geração a geração.

Salmos 100:5

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Campus Senador Helvídio Nunes de Barros
Biblioteca Setorial José Albano de Macêdo
Serviço de Processamento Técnico

O482e Oliveira, Myllena Caetano de

Extração manual de características para classificação de sons urbanos / Myllena Caetano de Oliveira – 2021.

37 f.; CD-ROM 4 ¾ pol.

Monografia (Graduação em Sistemas de Informação) – Universidade Federal do Piauí, Picos-PI, 2021.

"Orientadora: Profª. Deborah Maria Vieira Magalhães"

1. Processamento de áudio. 2. Som-Seleção de características. 3. Eventos sonoros. I. Título.

CDD 006

Elaborada por Maria José Rodrigues de Castro CRB 3: CE-001510/O

EXTRAÇÃO MANUAL DE CARACTERÍSTICAS PARA CLASSIFICAÇÃO DE SONS URBANOS

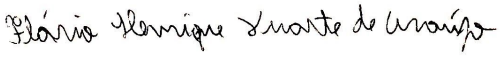
MYLLENA CAETANO DE OLIVEIRA

Monografia Aprovada como exigência parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Data de Aprovação

Picos – PI, 19 de janeiro de 2021

Prof(a). Deborah Maria Vieira Magalhães



Prof. Flávio Henrique Duarte de Araújo

Prof. Romuere Rodrigues Veloso e Silva

Resumo

O som é uma importante fonte de informações. Em relação aos sons urbanos, a perspectiva de conectividade, captura, análise e classificação automáticas dos sons podem auxiliar na redução da poluição sonora nas cidades, computação sensível ao contexto e vigilância urbana. Para que essas informações possam ser obtidas é necessário que o áudio passe por algumas etapas, dentre elas, a extração de características. Este trabalho propõe o desenvolvimento de um descritor, através da extração manual e seleção de características, capaz de realizar a classificação de sons urbanos. As características foram extraídas no domínio do tempo, frequência e cepstral, a seleção das mesmas foi feita com base na lista ranqueada gerada pelo *Information Gain*. O resultado obtido foi um descritor compacto e preciso capaz de caracterizar sons urbanos com 94,2% de acurácia para o conjunto UrbanSound8K e 78,7% para o conjunto ESC (urban noises). Em comparação aos estudos relacionados, o método proposto apresentou resultados promissores, apresentando um desempenho superior até mesmo que descritores baseados em abordagem de aprendizado profundo. Os resultados alcançados enfatizam o desempenho das características manuais para a representação dos sinais acústicos. Dessa forma, o método proposto mostra-se eficiente para integrar aplicações de classificação de eventos sonoros em espaços urbanos.

Palavras-chaves: Processamento de Áudio. Seleção de Características. Eventos Sonoros.

Abstract

The sound is a valuable source of information concerning connectivity, capture, analysis, and automatic classification of urban sounds. It can help reduce noise pollution in cities, context-aware computing, and urban surveillance. However, to obtain relevant information from audio, some steps are needed, including feature extraction. This work an urban sounds descriptor, through feature extraction and selection, capable of performing urban sounds classification. The features were extracted in the time, frequency, and cepstral domains, their selection was made based on the ranked list generated by Information Gain. The result obtained was a compact and efficient descriptor capable of characterizing urban sounds with 94.2% accuracy for the UrbanSound8K dataset and 78.7% for the ESC (urban noises) dataset. The proposed method presented promising results compared to related works, presenting a classification performance superior even to deep descriptors. The results achieved emphasize the performance of handcrafted features for the representation of acoustic signals. In this way, the proposed method proves to be efficient for integrating sound event classification applications in urban spaces that improve health and security, essential aspects of urban life.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

CENS	*Chroma Energy Normalized*
CQT	*Constant-Q Transform*
CNN	*Convolutional Neural Network*
IoT	Internet of Things
Mel	*Mel spectrogram*
MFCC	*Mel Frequency Cepstral Coefficient*
RF	*Random Forest*
RMS	*Root Mean Square*
STFT	*Short-Time Fourier Transform*
SVM	*Support Vector Machine*
ZCR	*Zero Crossing Rate*

Sumário

1 Introdução

Em 2018, cerca de 55,3% da população mundial vivia em espaços urbanos, esse número chegou a 80% quando tratamos da Europa e América do Norte. A China apresentou um crescimento de 40% em sua parcela de habitantes urbanos nos últimos 50 anos[1]. Até 2030, as áreas urbanas devem abrigar 60% das pessoas em todo o mundo e uma em cada três pessoas viverá em cidades com pelo menos meio milhão de habitantes. O aumento da densidade populacional estressa a infraestrutura dos centros urbanos e traz consigo diversos desafios relacionados à mobilidade urbana, segurança e saúde pública (SOUZA et al., 2018).

Neste âmbito, as cidades inteligentes despontam como uma oportunidade para guiar políticas públicas visando oferecer melhores serviços e infraestrutura aos cidadãos (RATHORE et al., 2016). As cidades inteligentes se apoiam na expansão da conectividade com a Internet das Coisas (*IoT*, do inglês *Internet of Things*) para o monitoramento ubíquo e inteligente através de redes de sensores e *smartphones*. Sendo o som uma importante fonte de informação a respeito da vida urbana (EDWARDS, 2018), a perspectiva de conectividade, captura, análise e classificação automáticas dos sons urbanos pode auxiliar na redução da poluição sonora nas cidades e na tarefa de vigilância urbana.

Alguns fatores motivadores do desenvolvimento de novas pesquisas focadas na identificação de eventos sonoros a partir do processamento de *streams* de áudio são:

- O aumento da densidade populacional em centros urbanos projeta, entre outros desafios, a dificuldade de realização de policiamento efetivo e proteção de espaços públicos. Nos últimos anos, enfrenta-se no mundo o crescimento da criminalidade e terrorismo (BELLO; MYDLARZ; SALAMON, 2018). No Brasil, em 2012, as grandes cidades apresentaram altos níveis de criminalidade nas categorias de roubo, estupro, fraude e roubos residenciais (BORGES et al., 2017);

- O crescimento das redes de sensores acústicos aliado à ampla adoção dos *smartphones* impulsionou a utilização do microfone como dispositivo vantajoso para monitorar espaços urbanos, pois este é menor e mais barato que a câmera. Além disso, o microfone é mais robusto às condições ambientes, como neblina, poluição, chuva e mudanças de luminosidade, e realiza o monitoramento omnidirecional, sendo menos susceptível à oclusão (BELLO; MYDLARZ; SALAMON, 2018). No geral, a captura de som exige menos bateria dos dispositivos envolvidos (VIRTANEN; PLUMBLEY; ELLIS, 2018) quando comparados a aplicações de monitoramento por vídeo.

[1] https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS

1.1 Problema

O processamento e análise de fontes acústicas podem contribuir para localização e rastreamento de atos criminosos, terroristas e tumultos e, a partir dessa identificação, definir características como nível, duração e frequência de tais eventos. Isso pode, por sua vez, oferecer maior compreensão nas áreas de ciências sociais e políticas públicas sobre a relação entre o som urbano e a criminalidade em uma determinada região, melhorando a efetividade das intervenções tanto da polícia como dos órgãos que prestam socorro às vítimas.

Para a maioria dos humanos e animais, a habilidade de escutar eventos sonoros é uma tarefa trivial, mas desenvolver algoritmos que sejam capazes de automaticamente reconhecer um som é desafiador (VIRTANEN; PLUMBLEY; ELLIS, 2018). Quando trata-se de ambientes urbanos, onde são encontrados sons naturais não biológicos (vento, chuva), sons naturais biológicos (animais em geral) e sons mecânicos (tráfego, construção, sinais, máquinas, instrumentos musicais), tem-se um ambiente altamente heterogêneo que pode possuir ilimitados sons.

Ainda, em ambientes realistas, o evento sonoro que se deseja classificar se sobrepõe aos demais presentes no ambiente. Portanto, o áudio capturado é uma superposição de todas as fontes presentes. Além disso, em várias aplicações de monitoramento de eventos sonoros, os microfones usados para capturar áudio costumam ficar significativamente mais distantes das fontes sonoras, o que aumenta a interferência no sinal capturado. Todos esses fatores previnem a correspondência do sinal com os modelos desenvolvidos para reconhecê-lo. Essa situação é bem diferente das aplicações de fala, entre elas, reconhecer a sequência de palavras na fala ou reconhecer a identidade da pessoa que está falando, onde comumente os microfones são utilizados em ambientes fechados e controlados. Portanto, a análise e compreensão de eventos acústicos em espaços urbanos é ainda mais desafiadora.

A realização de monitoramento sonoro exige custo financeiro e disponibilidade de tempo para instalação, configuração e monitoramento do som. Ao se adquirir uma base de dados pública, tais desafios são contornados, surgindo, no entanto, novos desafios. Sendo estes a impossibilidade de controlar a qualidade do sinal, número de amostras insuficientes para em algumas classes, dados provenientes de diferentes sensores e muito heterogêneos.

1.2 Objetivos

Diante do exposto, o objetivo geral deste trabalho é a construção de uma metodologia para a extração de características que ofereça uma boa descrição de eventos sonoros, contribuindo para detecção de situações de risco à segurança dos habitantes, melhoria da saúde e gerenciamento do tráfego em espaços urbanos. Portanto, os objetivos específicos

são:

1. Investigar a viabilidade da utilização de características manuais para realizar a extração de atributos dos sinais acústicos nos domínios de tempo, frequência e cepstrais;
2. Aplicar uma etapa de seleção de características para utilizar na posterior etapa de classificação apenas as características mais representativas, reduzindo a dimensionalidade do descritor sem perder informações importantes;
3. Desenvolver um descritor preciso e compacto que possa ser empregado em soluções embarcadas, requisito relevante no contexto de cidades inteligentes; e
4. Aplicar algoritmos de classificação clássicos como, *Random Forest* e *Suport Vector Machine*, a fim de avaliar o descritor proposto.

1.3 Organização do Trabalho

Este trabalho está organizado da seguinte maneira: o Capítulo 2 fundamenta os conceitos envolvidos no desenvolvimento do trabalho; o Capítulo 3 detalha as etapas realizadas na execução do método proposto; o Capítulo 4 apresenta e discute os resultados alcançados em todos os testes realizados com a abordagem proposta; o Capítulo 5 apresenta a conclusão deste trabalho, bem como os trabalhos futuros; e, por fim, o Capítulo 6 lista as publicações alcançadas com o desenvolvimento deste trabalho.

2 Referencial Teórico

Com o intuito de fornecer um melhor entendimento sobre os temas abordados neste trabalho, este capítulo apresenta uma breve revisão literária. Os conceitos explanados referem-se ao som (Seção 2.1), às técnicas de pré-processamento (Seção 2.2), às etapas de extração (Seção 2.3) e seleção (Seção 2.4) de características sonoras, à classificação das características (Seção 2.5) e às métricas de avaliação dos modelos de classificação (Seção 2.6).

2.1 Som

O som é o resultado de vibrações no ar ou na água, que se propaga através ondas chamadas de ondas sonoras (VIRTANEN; PLUMBLEY; ELLIS, 2018). É possível concluir, então, que o som é uma onda do tipo mecânica, já que precisa de um meio para propagar-se, e que a mesma é tridimensional, já que pode ser percebida de qualquer direção. Existem 3 características principais que definem um som: intensidade, altura e timbre (OLIVEIRA, 2002). Na Figura 1, tem-se uma representação de onda sonora.

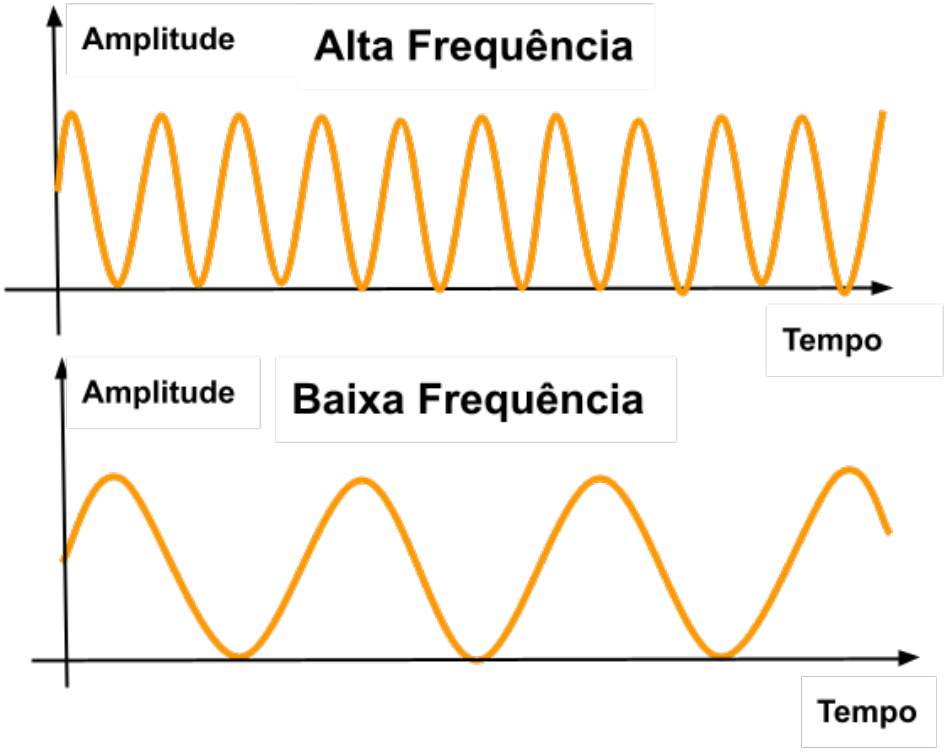


Figura 1 – Representação da onda sonora.

A intensidade do som está relacionada à energia de vibração da fonte que emite o som, ela é provocada pela pressão que a onda exerce sobre o ouvido, ou seja, quanto maior a

pressão maior será a intensidade (IAZZETTA, 2010). A unidade de medida utilizada para representar essa propriedade é o bel (decibel). A altura é o que diferencia um som entre grave e agudo, sendo que o som grave é o que identificamos como mais “grosso” enquanto o agudo é o mais “fino". Como demostrado na Figura 1, essa característica é definida pela frequência da onda sonora. Um som com baixa frequência é dito som grave e o som com altas frequências é dito som agudo (OLIVEIRA, 2002). Finalmente, o timbre é o que nos permite diferenciar os sons de mesma frequência e mesma intensidade. Isso ocorre porque o timbre faz com que as ondas sonoras produzidas por esses sons tenham um formato completamente distinto (OLIVEIRA, 2002). Por exemplo, se dois instrumentos diferentes, como um piano e um saxofone, estiverem tocando uma mesma nota musical, a distinção desses instrumentos será possível graças ao timbre.

Uma onda sonora pode ser representada em um gráfico bidimensional, onde o eixo horizontal representa a passagem do tempo e o vertical a variação de pressão. Esse tipo de gráfico pode fornecer várias informações sobre o som (IAZZETTA, 2010), e é baseado nisso que as características sonoras são extraídas.

2.2 Pré-processamento

Para que a extração de características possa ser realizada é necessário que um áudio passe por alguns processos. O principal papel da etapa de pré-processamento é aprimorar determinadas características do sinal recebido, a fim de maximizar o desempenho da análise de áudio nas fases posteriores do sistema de análise. O pré-processamento também busca garantir a uniformização dos dados adquiridos, de modo que todos possuam as mesmas configurações de taxa de amostragem, quantização e números de canais.

A taxa de amostragem consiste no número de amostras por unidade de tempo medido em hertz (Hz), sendo uma amostra uma medida de amplitude do sinal em um intervalo de tempo. A taxa de amostragem influencia diretamente na semelhança do sinal digital com o analógico (KOIZUMI et al., 2019), dessa maneira, quanto maior for a taxa de amostragem, mais precisa é a representação do sinal. A quantização trata de discretizar um áudio em valores inteiros dentro de um intervalo de representações, sendo os comprimentos mais usados de 16-bit, 24-bit e 32-bit. A quantização interfere na qualidade do som, quanto maior o comprimento, maior a qualidade, e por consequência, maior o tamanho do arquivo (PICHLMAIR; KAYALI, 2007). O número de canais de áudio se refere a quantidade de canais que são responsáveis pela captura do som. Em um sistema monofônico, por exemplo, todas as informações do áudio são registradas em um mesmo canal (EPPOLITO, 2008).

Os valores típicos para a qualidade de um CD de áudio são uma taxa de amostragem de $f/s = 44.1$ Hz e uma quantização em 16 bits por amostra levando a uma taxa de bits de 705.600 bit/s para um sinal de áudio de canal único. Padrões de qualidade mais altos

incluem taxas de amostragem de 48, 96 ou 192 Hz e quantização em 24 bits (SERIZEL et al., 2018).

2.3 Extração de Características

A fase de extração de características busca extrair de um som características que melhor o representam e que, através dessas, seja possível a um classificador distinguir esse som de acordo com a classe a qual ele pertence. Neste trabalho, foram extraídas características nos domínios de tempo, frequência e cepstral.

2.3.1 Características no domínio do tempo

As características de tempo estão diretamente ligadas com a onda temporal formada pelo áudio (SERIZEL et al., 2018). No domínio do tempo, foram extraídas as características: *Tempogram*, *Zero Crossing Rate* (*ZCR*) (SHARMA; UMAPATHY; KRISHNAN, 2020), *Spectral Rolloff* (XIE; ZHU, 2019), *Spectral Centroid* e *Root Mean Square* (*RMS*), sendo as características *Spectral Centroid* e *RMS* baseadas em energia (BARKER; VIRTANEN, 2016).

Entre elas, destaca-se o *Tempogram*, uma representação de tempo que está relacionada ao ritmo, matematicamente representado pela equação 2.1, onde t representa o tempo de duração em segundos e l o tempo de atraso (*lag*) (GROSCHE; MÜLLER; KURTH, 2010). A característica *ZCR* (Equação 2.2), assim como o nome sugere, irá contar a quantidade de vezes em que o sinal passa pelo ponto zero, ou seja, quando o valor do sinal muda de positivo para negativo ou de negativo para positivo. O *Spectral Rolloff* é definido como a frequência sob a qual uma porcentagem predefinida (normalmente entre 85% e 99%) da energia espectral total está presente.

$$A(t,l) = \frac{\sum_{n \exists \mathbb{Z}} \Delta(n)\Delta(n+l).W(n-t)}{2N+1-l} \tag{2.1}$$

$$ZCR = \frac{1}{2}\sum_{N=1}^{n} |sign(x[n]) - sign(x[n-1]) \tag{2.2}$$

Para se conseguir o *Spectral Centroid*, cada quadro de um espectrograma de magnitude é normalizado e tratado como uma distribuição sobre *bins* de frequência, da qual a média (centroid) é extraída por quadro (Equação 2.4) (GIANNAKOPOULOS; SPYROU; PERANTONIS, 2019). O *RMS* (*root-mean-square*) é calculado a partir do sinal de som bruto no domínio do tempo. Matematicamente, o *RMS* é representado pela equação 2.3 (LIU; TSENG; TRAN, 2019).

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2} \tag{2.3}$$

$$centroid(t) = \frac{\sum_k S(k,t).freq(k)}{\sum_j S(j,t)} \tag{2.4}$$

2.3.2 Características no domínio da frequência

No domínio da frequência, foram empregadas características relacionadas ao cromagrama, sendo elas: *Chroma STFT* (*Short-Time Fourier Transform* - transformada de Fourier de curto tempo) (ELLIS, 2007), *Chroma CQT* (*Constant-Q Transform* - transformada de constante-Q) (SCHÖRKHUBER; KLAPURI, 2010) e *Chroma CENS* (*Chroma Energy Normalized* - Energia de croma normalizada) (MÜLLER; EWERT, 2011). Também no domínio da frequência foram utilizadas as características espectrais: *Entropy* (AL-NASHERI et al., 2017), *Spectral bandwidth* (WU; VINTON, 2017), *Spectral Flatness* (DUBNOV, 2004) e *Spectral contrast* (JIANG et al., 2002). Por fim, foram extraídos os coeficientes polinomiais de enésima ordem (*Poly*) (YANG et al., 2019) e *Tonnez* (HARTE; SANDLER; GASSER, 2006), que está relacionado à tonalidade.

Chroma STFT é um atributo poderoso de propósito geral para o processamento de sinal de áudio. Ele define uma classe particularmente útil de distribuições de tempo-frequência que especifica a amplitude complexa versus tempo e frequência para qualquer sinal. A *Fourier transform* é uma característica bem conhecida para analisar a distribuição de frequência de um sinal (ELLIS, 2007). Então o *STFT* discreto (D) sobre uma janela g com suporte compacto função pode ser escrita como:

$$STFT = F_g f\left[n,k\right] = \sum_{m=0}^{m-1} f\left[n-m\right] g\left[m\right] \epsilon_k\left[m\right], \tag{2.5}$$

onde,

$$\epsilon_k = e^{-2\pi m \frac{k}{N}}, \tag{2.6}$$

m é o comprimento da janela de g e n é o número de amostras em f. Este algoritmo pode ser interpretado como uma sucessiva avaliação das transformadas de *Fourier* sobre segmentos curtos de todo o sinal. Além disso, as frequências podem ser visualmente representadas exibindo a magnitude quadrada de *Fourier* coeficientes em cada seção. Este diagrama é chamado de espectrograma de f.

O *CQT*, segundo Brown (1991), é uma representação tempo-frequência que, diferente da *STFT*, apresenta um espectro de frequências com fator de seletividade (também conhecido como fator Q) constante. Devido a isto, as componentes de frequência estão espaçadas geometricamente, ou seja, uma componente f_k é dada pela Equação 2.7.

$$CQT = f_{k-1}\left(\frac{1}{Q}+1\right) = f_{min}\left(\frac{1}{Q}+1\right)^k \tag{2.7}$$

Adicionando um outro grau de abstração considerando o curto espaço de tempo estatísticas sobre distribuições de energia dentro as bandas de croma, obtém-se *CENS* (Es-

tatísticas normalizadas de energia cromada), que constitue uma família de características de áudio escaláveis e robustos. Estas características acabaram sendo muito úteis na correspondência e recuperação de áudio dos formulários. Na extração do *CENS*, uma quantização é aplicada com base em limites escolhidos logaritmicamente (MULLER; KURTH; CLAUSEN, 2005).

Inicialmente, uma quantidade que descreve o grau de desordem em um sistema termodinâmico, a entropia é posteriormente amplamente utilizada para avaliar a incerteza de um sistema. Da perspectiva da teoria da informação, entropia é a quantidade de informação contida em uma distribuição de probabilidade generalizada. Enquanto o paramétrico não linear que quantifica a complexidade de uma série temporal, pode ser usado para descrever sinais dinâmicos não lineares e instáveis (Equação 2.8) (AL-NASHERI et al., 2017).

$$Entropy = -\sum_{i=0}^{N-1} P_i * log_2\left(P_i\right) \tag{2.8}$$

Spectral bandwidth de banda e frequência dominante de uma potência de *Fourier* espectro para fornecer a base para a introdução de medidas espectrais instantâneas correspondentes. Essas quantidades são então reduzidas a atributos de rastreamento complexos facilmente computáveis. O objetivo de apresentar esses atributos nesse maneira é enfatizar seu papel como medidas variantes no tempo de propriedades espectrais médias. Isso os empresta intuitivamente significado atraente e sugere aplicações úteis (WU; VINTON, 2017).

Como conhecido da teoria de codificação, o ganho máximo que pode ser recuperado por redução de redundância usando métodos de codificação preditivos ou codificação de transformação é determinado pela desplanicidade da densidade espectral de potência do sinal e está relacionado com o chamado *Spectral Flatness* (DUBNOV, 2004).

A característica *Spectral contrast* pode refletir aproximadamente a distribuição relativa do componentes harmônicos e não harmônicos no espectro. Características anteriores, como *MFCC*, calculam a média do espectro distribuição em cada sub-banda, e assim perder o espectro espectral relativo em formação (JIANG et al., 2002).

O Tonnetz é uma representação plana dos sons através de um conjunto de linhas paralelas criadas a partir da circunferência do "círculo das quintas". Ele busca determinar a distância entre a tonalidade de cada som (Equação 2.9) (HARTE; SANDLER; GASSER, 2006).

$$t = \sqrt{\frac{1}{4} + d^2} > 1 \tag{2.9}$$

2.3.3 Características cepstrais

As características cepstrais tem relação com a forma com que a audição humana percebe os sons, principalmente a fala. As características cepstrais mais comuns são os coeficientes cepstrais de frequência mel (*MFCCs*, do inglês *Mel Frequency Cepstral Coefficients*), (VIRTANEN; PLUMBLEY; ELLIS, 2018). Os *MFCCs* na realidade são um conjunto de características, geralmente de 10 a 20, que descrevem de forma concisa a forma geral de um envelope espectral, ou seja, o limite em que o espectro do sinal está contido. Matematicamente, os *MFCCs* são resultado da transformação inversa da discreta do cosseno da energia do *log* em bandas de frequência de mel, representado pela Equação 2.10 (SERIZEL et al., 2018).

$$MFCC(t,c) = \sqrt{\frac{2}{M_{mfcc}}} \sum_{m=1}^{M_{mfcc}} \log(|X|_m(t)) \cos(\frac{c(m-\frac{1}{2})}{M_{mfcc}}), \tag{2.10}$$

onde M_{mfcc} é o número de bandas de frequência mel, m o índice de banda de frequência, $|X|_m(t)$ é a energia na *Mnésima* banda de frequência mel e c é o índice de do coeficiente cepstral.

Além disso, através do cálculo da primeira e segunda derivada dos *MFCCs*, foram obtidas as características também cepstrais, *delta1* (Equação 2.11) e *delta2* (Equação 2.12).

$$Delta1 = f'(MFCC) \tag{2.11}$$

$$Delta2 = f''(MFCC) \tag{2.12}$$

2.4 Seleção de Características

A fim de encontrar inconsistência ou redundância nas características obtidas na etapa de extração, foi aplicado uma seleção de características sobre o descritor gerado. Isso é feito pois a qualidade das características pode afetar no desempenho da classificação. Além disso, ao se identificar essas características, é possível reduzir a dimensionalidade do descritor, tornando o treinamento uma tarefa mais rápida e menos custosa computacionalmente.

O objetivo desta etapa é identificar quais características poderiam contribuir melhor para a discriminação dos sons urbanos. Para isso, foi utilizado o algoritmo de seleção de atributos *Information Gain* ou Ganho de Informação que funciona da seguinte forma: os valores de entrada variam de 0 (sem informação) a 1 (informação máxima), as características que contribuem com mais informações terão um maior valor de ganho, enquanto aquelas que não adicionam muita informação terão uma pontuação mais baixa (KAUR, 2019). Considerando um conjunto de dados S com s amostras pertencentes a m classes

distintas, a Equação 2.13 apresenta o *information* (VOGADO et al., 2018) utilizado para classificar uma dada amostra.

$$Info(S) = -\sum_{i=1}^{m} p_i \log_2(p_i), \tag{2.13}$$

onde p_i é a probabilidade de uma amostra pertencer a uma classe C_i. A entropia de um determinado atributo A que tem valores v é mostrada na Equação 2.14.

$$Entropy(A) = -\sum_{i=1}^{m} Info(S) \frac{s_{1i} + s_{2i} + \cdots + s_{mi}}{s}, \tag{2.14}$$

e a variável S_{ij} representa o número de amostras pertencentes à classe C_i do subconjunto S_i. O atributo gain A é representado pela Equação 2.15:

$$Gain(A) = Info(S) - Entropy(A). \tag{2.15}$$

2.5 Classificação

Para o reconhecimento dos padrões presentes nas características extraídas através dos métodos descritos anteriormente, foram utilizados dois classificadores amplamente aplicados pela literatura, o *Random Forest* (RF) (ESMAIL; AHMED; ELTAYEB, 2019) e o *Support Vector Machine (SVM)* (CORTES; VAPNIK, 1995). Um outro fator importante para a escolha dos mesmos, está na sua arquitetura estrutural, pois o *RF* resolve o problema com auxílio de árvores, e o *SVM* com auxílio de vetores se suporte, maximizando, assim, a fronteira de decisão. A seguir, está uma breve descrição dos classificadores utilizados pelo presente trabalho.

- O algoritmo de classificação *RF*, como o próprio nome sugere, gera um conjunto de árvores de decisão, independentes entre si, com predições aleatórias. Ou seja, uma floresta formada por árvores de decisão, contendo cada uma um sub-conjunto de indicadores escolhidos aleatoriamente. Aquela que obtiver o maior número de votos será escolhida como modelo de classificação (ESMAIL; AHMED; ELTAYEB, 2019).

- O *SVM*, proposto por Cortes e Vapnik (1995), é um algoritmo utilizado em diversos problemas, fazendo parte das classes de algoritmos de aprendizado supervisionados, que analisam os dados e reconhecem padrões existentes nos mesmos, sendo amplamente utilizado para classificação e regressão. O *SVM* toma como entrada um conjunto de dados e realiza uma predição para cada entrada dada, analisando qual das classes anteriormente definidas a nova entrada adequa-se melhor. Sendo assim, a proposta de Cortes e Vapnik (1995) torna-se um classificador linear binário não probabilístico. Um modelo *SVM* pode ser imaginado como composto por exemplos

de pontos espaciais, mapeados de maneira que os exemplos de cada categoria sejam divididos por um espaço claro que seja tão amplo quanto possível. Os novos exemplos são então mapeados no mesmo espaço e preditos como pertencentes a uma categoria baseados em qual o lado do espaço eles são colocados. De forma geral, o intuito do *SVM* é buscar uma linha de separação, mais comumente chamada de *hiperplano* entre os dados categoricamente separados anteriormente. Essa linha procura maximizar a distância entre os pontos mais próximos em relação a cada uma das classes (CORTES; VAPNIK, 1995).

2.6 Métricas de Validação

Para avaliar os resultados da classificação dos áudios, foram utilizadas medidas estatísticas baseadas na análise da matriz de confusão, que é calculada com base em quatro valores: verdadeiro positivo (TP), falso positivo (FP), falso negativo (FN) e verdadeiro negativo (TN). Esses valores servem para indicar o número de amostras classificadas correta e incorretamente. As medidas utilizadas foram: Acurácia (Acc), Índice *kappa* (κ) e a area sobre a curva ROC (AUC). A métrica de acurácia, representa a porcentagem de elementos que foram classificados corretamente (BARATLOO et al., 2015). Ela é descrita como:

$$Acc = \frac{TN + TP}{TN + TP + FN + FP}. \tag{2.16}$$

O índice *kappa* indica como os classificadores utilizados são capazes de superar o classificador que simplesmente adivinha aleatoriamente de acordo com a frequência de cada classe x. De acordo com Cohen (1960), há uma categorização sobre os níveis de desempenho da classificação segundo o índice *kappa*, apresentado na Tabela 1. A Equação 2.17 apresenta a fórmula para calcular essa métrica.

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{2.17}$$

onde po é o resultado alcançado e pe é o resultado esperado.

Tabela 1 – Categorização do desempenho da classificação, de acordo com o índice *kappa*.

Kappa	**Qualidade**
$\kappa < 0{,}2$	Ruim
$0{,}2 \leq \kappa < 0{,}4$	Razoável
$0{,}4 \leq \kappa < 0{,}6$	Bom
$0{,}6 \leq \kappa < 0{,}8$	Muito bom
$\kappa \geq 0{,}8$	Excelente

A métrica *AUC* fornece informações sobre a capacidade da robustez do modelo em separar as classes usando o gráfico de desempenho fornecido pela curva *ROC* (*Receiver*

Operating Characteristic). A curva *ROC* tem-se seu cálculo através de diferentes limiares, analisando duas métricas calculadas através da matriz de confusão: a Taxa de Verdadeiro Positivo (TVP) e Taxa de Falso Positivo (TFP), sendo assim uma métrica sensível a proporção de verdadeiros negativos corretamente classificados. Quanto mais próximo de 1 o valor da *AUC* for, melhor é o desempenho do método na distinção das classes (HANLEY; MCNEIL, 1982).

2.7 Trabalhos Relacionados

A literatura dispõe de estudos recentes sobre extração de características sonoras que possam ser utilizadas em algoritmos de classificação dos sinais de áudio. Neste Capítulo são apresentados trabalhos encontrados na literatura que, assim como este trabalho, também utilizam a técnica de extração manual de características para descrever os atributos relevantes dos áudios.

Silva et al. (2019) avaliou técnicas de aprendizagem de máquina para a classificação de sons urbanos em dispositivos embarcados. Eles aplicaram uma abordagem baseada na extração manual de características (*Mel Frequency Cepstral Coefficients* (MFCC), informação espectral, *root mean square* (RMS), e *zero crossing rate*) e classificação. Os autores adotaram o algoritmo *Information Gain* na fase de seleção de características, resultando em um vetor de características com dimensão total de 90. A proposta alcançou uma precisão de classificação de 46,90% com k-NN para o conjunto de dados ESC-50 e 53,50% com Naive Bayes para o conjunto de dados UrbanSound8K.

Giannakopoulos, Spyrou e Perantonis (2019) utilizou *Convolutional Neural Networks* (*CNNs*) para a extração de características e mel-espectrogramas como entrada para as *CNNs* na tarefa de reconhecimento de eventos sonoro em ambientes urbanos. O autor extraiu características manualmente diretamente da onda sonora e as combinou com as características extraídas pela *CNN*. A combinação das duas metodologias resultou em um almento de 11% no desempenho de classificação com SVM. Eles alcançaram uma acurácia máxima de 79,3 e 54,2% para a UrbanSound8K e a ESC-50, respectivamente. O autor não implementou a fase de seleção de características, resultando em um vetor com uma dimensão total de 580, para cada áudio.

Su et al. (2019) propôs um sistema composto por três componentes: extração e combinação de características, treinamento de uma *CNN* e fusão em nível de decisão baseada na teoria de DS (Dempster—Shafer), para o reconhecimento inteligente de eventos sonoros. O autor combinou as características extraídas em dois conjuntos destintos, LMC, composto por: log-mel spectogram, chroma, spectral contrast e tonnetz; e MC, composto por: MFCC, chroma, spectral contrast e tonnetz. Em seguida, ambos os conjuntos são combinados e um espectrograma é criado, sendo utilizado para o treinamento da *CNN*. O modelo TSCNN-DS, proposto pelo autor, atinge uma acurácia de classificação de 97,2%

para o conjuntos de dados UrbanSound8K.

Mushtaq e Su (2020) utilizou três diferentes características sonoras: *Mel spectrogram* (*Mel*), *Mel Frequency Cepstral Coefficient* (*MFCC*) e *Log-Mel*. Tais características foram consideradas, pelo autor, como entrada para o modelo de *CNN* proposto. Para a ESC-10, a ESC-50 e a UrbanSound8K, a acurácia alcançada foi de 94,94%, 89,28%, e 95,37% respectivamente. O autor também adotou técnicas de deformação para o aumento dos dados.

Alves et al. (2020) utilizou a extração de características manuais para a tarefa de classificação de sons urbanos em sistemas embarcados. O autor extraiu as características M*el Frequency Cepstral Coefficients* (*MFCC*), Spectral rolloff, Spectral centrid e *zero crossing rate* (*ZCR*). Gerando um vetor com dimensão total de 46 por áudio. As caracteristicas foram classificadas utiizando os algoritmos de classificação SVM e K-NN alcançando uma acurácia de 73% para o conjunto UrbanSound8K.

Tabela 2 – Comparação dos trabalhos relacionados elencados.

Trabalho	Objetivo	Características	Seleção de características	Bases	Representação do som	Acurácia (Esc-10/Esc-50/UrbanSound8K)
Silva et al. (2019)	Utilizar características extraídas manualmente para classificação com aprendizagem de máquina.	MFCCs Informação cepstral RMS ZCR	Sim	ESC-50/UrbanSound8K	Onda sonora	-/46,9%/53,5%
Giannakopoulos, Spyrou e Perantonis (2019)	Concatenar características manuais e extraídas pela CNN para maximizar os resultados.	ZCR Energia Entropia da energia Spectral centroid Spectral spread Spectral entropy Spectral flux Spectral rollof MFCCs Chroma vector Chroma deviation	Não	ESC-50/UrbanSound8K	Melspectrogramas	-/54,2%/79,3%
Su et al. (2019)	Testar diferentes conjuntos de características manuais no treinamento de CNNs.	MFCCs Chroma Spectral Contrast Log-mel spectrogram Tonnetz	Não	UrbanSound8K	Espectrogramas	-/-/97,2%
Mushtaq e Su (2020)	Usar espectogramas extraídos manualmente como entrada de CNNs.	MFCCs Mel spectrogram Log-Mel	Não	*Esc-10*/ESC-50/UrbanSound8K	Espectrogramas	94,94%/89,28%/95,37%
Alves et al. (2020)	Concatenar características manuais e extraídas pela CNN para maximizar os resultados.	MFCCs Spectral rolloff spectral centroid ZCR	Não	UrbanSound8K	Espectrogramas	73%

A tabela 2 apresenta um breve resumo dos trabalhos relacionados. Dentre os trabalhos apresentados apenas o Silva et al. (2019) realizou a etapa de seleção de característica, o que mostra que a maioria dos trabalhos não buscou reduzir a dimensionalidade de seus descritores. Também, nenhum dos trabalhos apresentados realizou uma analise para definir quais características melhor contribuem na classificação de sons urbanos.

3 Proposta

As cidades inteligentes se apoiam na expansão da conectividade com a *IoT* para o monitoramento ubíquo e inteligente através de redes de sensores e *smartphones*. Sendo o som uma importante fonte de informação a respeito da vida urbana (EDWARDS, 2018), a classificação automática dos sons urbanos pode auxiliar na detecção de situações de risco à segurança dos habitantes, melhoria da saúde e gerenciamento do tráfego em espaços urbanos. Diante do exposto, a proposta deste trabalho é a construção de um descritor, através da extração manual de características, capaz de fazer a distinção entre diferentes classes de sons urbanos. E aplicando uma seleção de características, transformar esse descritor em um descritor compacto e preciso com um menor custo computacional, tornando-o adequado para a utilização em aplicativos de reconhecimento de som móvel ou sistemas embarcados. Para isso, foi seguido o passo à passo definido na Figura 2.

3.1 Materiais e Métodos

Para o desenvolvimento do descritor proposto neste trabalho, foi utilizada uma metodologia composta por 6 etapas sequenciais, sendo assim, cada etapa utiliza o resultado da etapa anterior. O fluxograma exibido na Figura 2 detalha a metodologia utilizada. A primeira etapa consiste na aquisição dos áudios, adquiridos através de duas bases de dados públicas: UrbanSound8K (SALAMON; JACOBY; BELLO, 2014), e ESC-50 (PICZAK, 2015). Em seguida, a etapa de pré-processamento busca garantir a uniformização de todos os áudios, o que é de grande importância para etapa de extração de características. Na quarta etapa ocorre a seleção das características extraídas, onde foram identificadas aquelas que não agregam informações relevantes na discriminação dos áudios. A etapa 5 consiste no treinamento dos classificadores para que possam distinguir as diferentes classes existentes. Finalmente, na etapa 6 os resultados obtidos pelos classificadores são avaliados com diferentes métricas de avaliação. As seções a seguir discutem cada uma dessas etapas.

3.1.1 Aquisição dos áudios

Dois dos mais populares conjuntos de dados públicos para reconhecimento de sons urbanos foram utilizadas para a aquisição dos dados: UrbanSound8K (SALAMON; JACOBY; BELLO, 2014) e ESC-50 (PICZAK, 2015). O conjunto ESC-50 é composto por 50 classes de sons diferentes, no entanto, apenas 10 classes são diretamente relacionadas a ambientes urbanos. Assim, neste trabalho foi adotado o subconjunto ESC (urban noises), formado pelas 10 classes de sons urbanos do ESC-50.

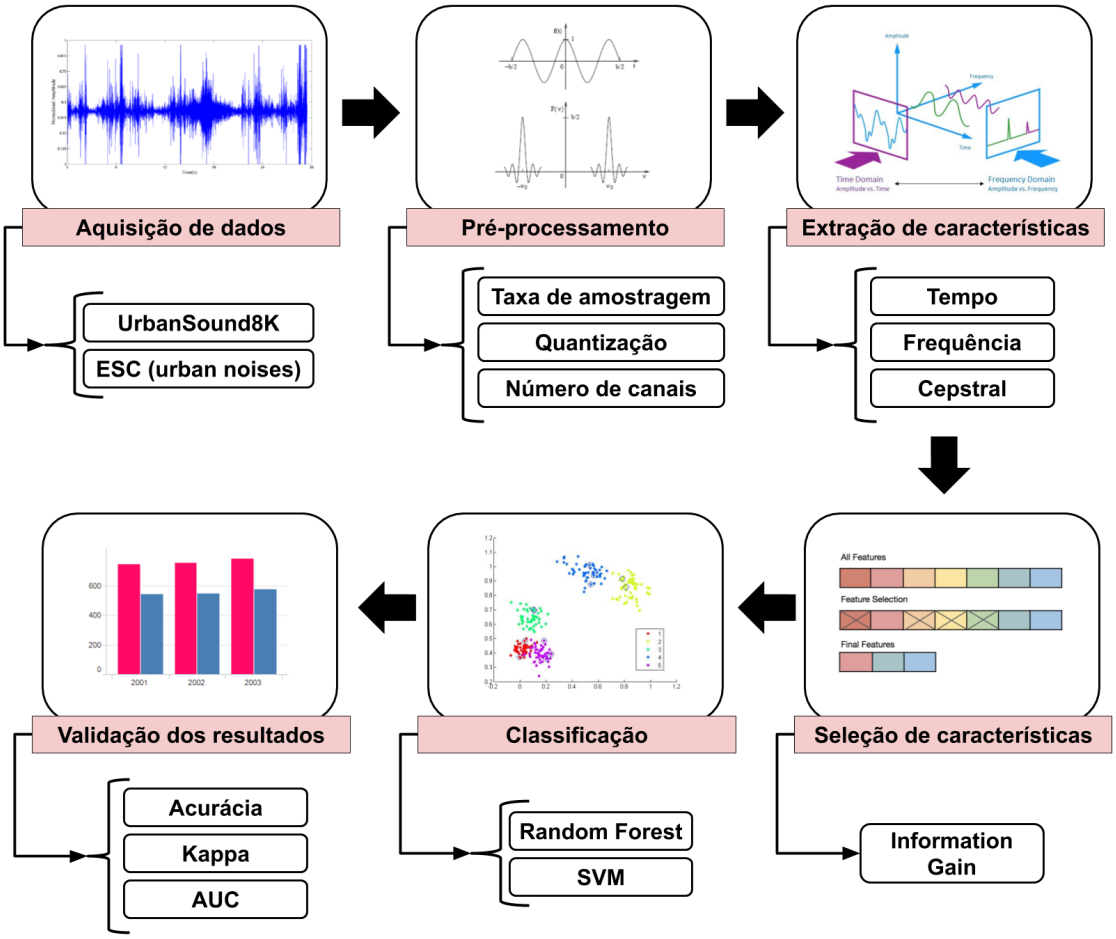


Figura 2 – Metodologia para classificação de sons urbanos.

O ESC (urban noises) é o subconjunto formado a partir do conjunto de dados ESC-50, ele possui 400 amostras de áudio distribuídas em 10 classes, que, diferentemente do UrbanSound8K, estão balanceadas, logo, tem o mesmo número de amostras por classe. No entanto, o pequeno número de amostras por classe, cuja divisão pode ser observada na Tabela 3, pode impactar negativamente no desempenho dos classificadores. O ESC (urban noises) é um conjunto de dados homogêneo, onde todas as amostras têm uma duração fixa de 5 segundos e um único canal.

O UrbanSound8K é um conjunto de dados formado por 8732 áudios, exclusivamente urbanos, com duração igual ou inferior a 4 segundos agrupados em 10 classes diferentes. No entanto, suas classes são desbalanceadas, logo, o número de amostras de uma classe pode chegar a ser três vezes maior que o de outra, assim como demonstrado na Tabela 3. O UrbanSound8K é um conjunto de dados heterogêneo, visto que a taxa de amostragem, a codificação e o número de canais podem variar entre as amostras. Na Tabela 3, as classes comuns às duas bases estão em destaque.

Tabela 3 – Divisão de classes por conjunto de dados.

UrbanSound8K			**ESC (urban noises)**		
Classes	**Número de amostras**	**Duração por classe (s)**	**Classes**	**Número de amostras**	**Duração por classe (s)**
Ar condicionado	1000	3994.9287	Helicóptero	40	200
Buzina de carro	**429**	**1053.9533**	**Buzina de carro**	**40**	**200**
Crianças brincando	1000	3961.8745	Motoserra	40	200
Latido	1000	3148.7496	Fogos de artifício	40	200
Perfurador	1000	3548.244	Motor	40	200
Motor ocioso	1000	3935.9925	Trem	40	200
Tiro	374	616.7965	Sinos de igreja	40	200
Britadeira	1000	3610.9747	Avião	40	200
Sirene	**929**	**3632.7016**	**Sirene**	**40**	**200**
Música de rua	1000	4000.0	Serrote	40	200

3.1.2 Pré-processamento

Para realizar a manipulação dos áudios nas bases, foi utilizada a biblioteca *LibROSA* (MCFEE et al., 2015). Esta biblioteca normaliza por padrão os dados no intervalo [-1,1], e transforma os sinais em mono ao fazer a média das amostras entre os canais. Além disso, ela garante que todas as amostras possuam uma mesma taxa de amostragem, de 22050 Hz e uma quantização de 16 bits (MCFEE et al., 2015) (Tabela 4).

Ambos os conjuntos de dados passaram pelo pré-processamento mencionado, dessa forma, nenhuma das amostras utilizadas na etapa de extração de características, tampouco as etapas seguintes, possuem uma configuração diferente. A Tabela 4 mostra a configuração utilizada.

Apesar de no conjunto de dados UrbanSound8K as amostras possuírem diferentes tamanhos de duração, isso não impacta na extração de características, pois a dimensão de cada característica se mantém independente da duração do áudio. As características com variações em sua dimensão, foram transformadas em um único valor calculando sua média. Por esse motivo a duração das amostras não foi uma das configurações modificadas.

Tabela 4 – Configuração das amostras.

Medida	**Valor**
Taxa de amostragem	22050 Hz
Quantização	16 bits
Canais	1

3.1.3 Extração de características

Na fase de extração de características, também foi aplicada a biblioteca *LibROSA*, que possui uma extensa documentação com inúmeras funções para viabilizar o processamento de áudio. Ela fornece os componentes necessários para criar sistemas de recuperação de informações sonoras. Através das funções da *LibROSA*, foram extraídas características

nos diferentes domínios: tempo, frequência e cepstral. Todas as características extraídas foram descritas na Seção 2.3.

Ao final da etapa de extração, obteve-se um descritor com uma dimensão total de 82 posições, ou seja, um conjunto de diferentes características capaz de descrever um som, de modo que seja possível à um classificador agrupar em classes, que, no caso deste trabalho, são diferentes classes de sons urbanos. O resumo das características extraídas neste trabalho pode ser visto na Tabela 5.

Tabela 5 – Características em seus diferentes domínios e dimensões.

Característica	**Domínio**	**Dimensão**
MFCC	Cepstral	60
Delta1	Cepstral	1
Delta2	Cepstral	1
ZCR	Tempo	1
Spectral rolloff	Tempo	1
Spectral centroid	Tempo	1
RMS	Tempo	1
Tempogram	Tempo	1
Chroma STFT	Frequência	1
Choma CQT	Frequência	1
Chroma CENS	Frequência	1
Entropy	Frequência	1
Spectral Flatness	Frequência	1
Spectral bandwidth	Frequência	1
Spectral contrast	Frequência	7
Poly	Frequência	1
Tonnez	Frequência	1

3.1.4 Seleção de características

A seleção das características que deveriam ser mantidas foi feita baseada na lista ranqueada gerada pelo algoritmo de seleção *Information Gain*. Com base em sua posição no ranque, as características foram consideradas como possuindo maior ou menor relevância para distinção dos sons, sendo consideradas com maior relevância as que possuíam posições mais altas no ranque e menor relevância as que possuíam posições mais baixas.

Para verificar a influência das características e dos subconjuntos de características, os testes foram realizados agrupando as características em grupos de 5, variando em um intervalo de 5 a 82, com acréscimos de 5. Após ser definido dentro desse intervalo qual o menor número de características que oferecia uma acurácia maior ou igual a acurácia alcançada com todas elas, um novo intervalo é testado, a partir do valor encontrado no intervalo anterior, testando agora em incrementos de 1, o desempenho com as 5 características anteriores e as posteriores a este valor.

Todos os teste realizados nesta etapa foram executados usando o software *WEKA*, uma suíte de algoritmos de mineração de dados e *Machine Learning*, desenvolvida pela Univer-

sity of Waikato, Nova Zelândia. Essa suíte contém ferramentas para pré-processamento de dados, seleção, classificação, regressão, agrupamento, regras de associação e visualização (HALL et al., 2009).

3.1.5 Classificação

Para verificar se as características resultantes da etapa de seleção possuem de fato eficiência para distinção das classes de áudio, foram utilizados os seguintes classificadores: *RF* e *SVM*. Esses algoritmos foram escolhidos levando em considerando sua ampla adoção na literatura relacionada, como mencionado no Capítulo 2. Os resultados foram avaliados utilizando 20% do conjunto de dados total para testes e 80% para treinamento, divididos aleatoriamente. Uma vez divididos, os mesmos conjuntos foram utilizados em todos os testes.

Assim como na etapa anterior, também foi utilizada a ferramenta WEKA nesta etapa, já que ela disponibiliza os algoritmos de classificação utilizados nesta metodologia. Em relação aos parâmetros de cada classificador, foram utilizados os valores padrão da ferramenta WEKA. Contudo, alguns parâmetros foram variados durante a execução do método com o intuito de encontrar a melhor configuração para os modelos. Esses parâmetros foram o número de árvores, no *RF*, e o *kernel*, no *SVM*. As variações desses parâmetros estão detalhadas no Capítulo 4.2.

3.1.6 Validação dos resultados

Como mencionado no Capítulo 2, para avaliar os modelos de classificação, foram utilizadas três métricas de validação comumente aplicadas na literatura: acurácia, índice *kappa* e *AUC*. As métricas escolhidas para realizar a validação dos resultados obtidos pelo presente trabalho são amplamente utilizadas por conseguirem demostrar de forma clara a proporção de erros e acertos obtidos pelos classificadores, por esse motivo foram utilizadas neste trabalho. As três métricas estão disponíveis na avaliação dos classificadores pela ferramenta WEKA, sendo assim, os valores foram obtidos através dela.

4 Resultados

Levando em consideração a metodologia descrita no Capítulo 3.1, este capítulo apresenta os resultados obtidos pela extração, seleção e classificação das características presentes no descritor proposto por esse trabalho. O descritor proposto resultou em um vetor com dimensão total de 82 posições, composto pelas características apresentadas na Tabela 5.

A discussão dos resultados é dividida em três seções: na Seção 4.1 são apresentados os resultados obtidos com a seleção de características do descritor, enfatizando as características mais relevantes para a classificação; na Seção 4.2 são apresentados os resultados alcançados por cada classificador; por fim, na Seção 4.3, os resultados da proposta são comparados com os trabalhos correlatos.

4.1 Seleção de Características

A etapa de seleção permitiu a redução da dimensionalidade do descritor através da seleção de características importantes para a caracterização dos sons urbano. A Figura 3 apresenta a acurácia alcançada pelo algoritmo de classificação *Randon Forest* (*RF*) em função do número de características usadas nos testes, bem como os dois conjuntos de dados se comportaram durante os testes de seleção, sendo os algoritmos de classificação treinados com o conjunto de características selecionadas pelo algoritmo de seleção *Information Gain*.

Observa-se, através da Figura 3, que os conjuntos de dados apresentam padrões diferentes. Para o conjunto de dados UrbanSound8K, a acurácia melhora à medida que novas características são adicionadas, consequentemente, a maior acurácia é obtida com todas as 82 características, o que corresponde a 94,2153%. No entanto, com um subconjunto de 35 características, o *RF* atinge 93,2417% de acurácia. O conjunto de dados ESC (urban noises) mostra um padrão irregular, o que pode sugerir que certos subconjuntos de características podem conter informações mais inconsistentes com impacto no desempenho da classificação. A melhor acurácia foi de de 72,5% com todas as características, por outro lado, com 20 características, a acurácia é de 71,25%. De acordo com a Figura 3, com um grupo de pelo menos 35 características foi possível alcançar um desempenho significativo para ambos os conjuntos de dados.

A Figura 4 apresenta a distribuição das primeiras 35 características classificadas pelo *Information Gain*. As características são apresentadas em grupos de 5, variando de 5 a 35. O *Information Gain* selecionou 14 de 20 e 27 de 35 das características em comum entre os conjuntos de dados. A Figura 4 revela a predominância de características de forma espectral entre as primeiras 15 posições de características para ambos os conjuntos de

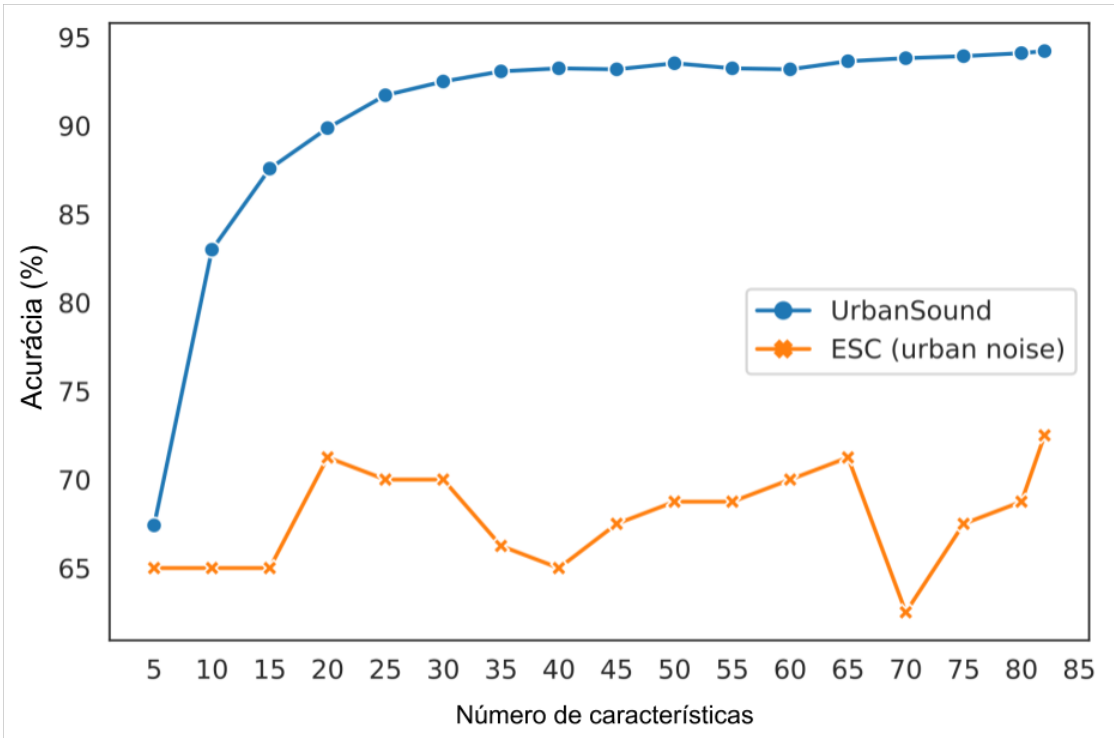


Figura 3 – Acurácia da seleção de características com o classificador RF.

dados. Além disso, o *Tempogram* e as características relacionadas ao *Chroma* ficaram nas primeiras 5 posições selecionadas.

Os *MFCCs* foram divididos em 3 grupos, cada um contendo 20 coeficientes. O primeiro, segundo e terceiro grupo de *MFCCs* foram nomeados como MFCC (0-20), MFCC (20-40) e MFCC (40-60), respectivamente. O MFCC (0-20) passa a ser selecionado entre a sexta e a décima posição e atinge uma frequência alta entre 20ª e 25ª posição. O MFCC (20-40) só aparece após a 15ª posição para ESC (urban noises) e a 20ª posição para UrbanSound8K. Os últimos 20 coeficientes *MFCC*, o conjunto MFCC (40-60), aparecem na 20ª característica. Ambos MFCC (20-40) e MFCC (40-60) são menos frequentes do que MFCC (0-20). Além disso, a característica *Spectral roll-off*, de domínio do tempo, aparece nos subconjuntos que oferecem uma acurácia significativa para ambos os conjuntos de dados.

4.2 Classificação

A Tabela 6 apresenta os resultados alcançados pelo descritor resultante das etapas de extração e seleção de características, utilizando os classificadores *Random Forest* (RF) e SVM, descritos anteriormente na Seção 3.1.5.

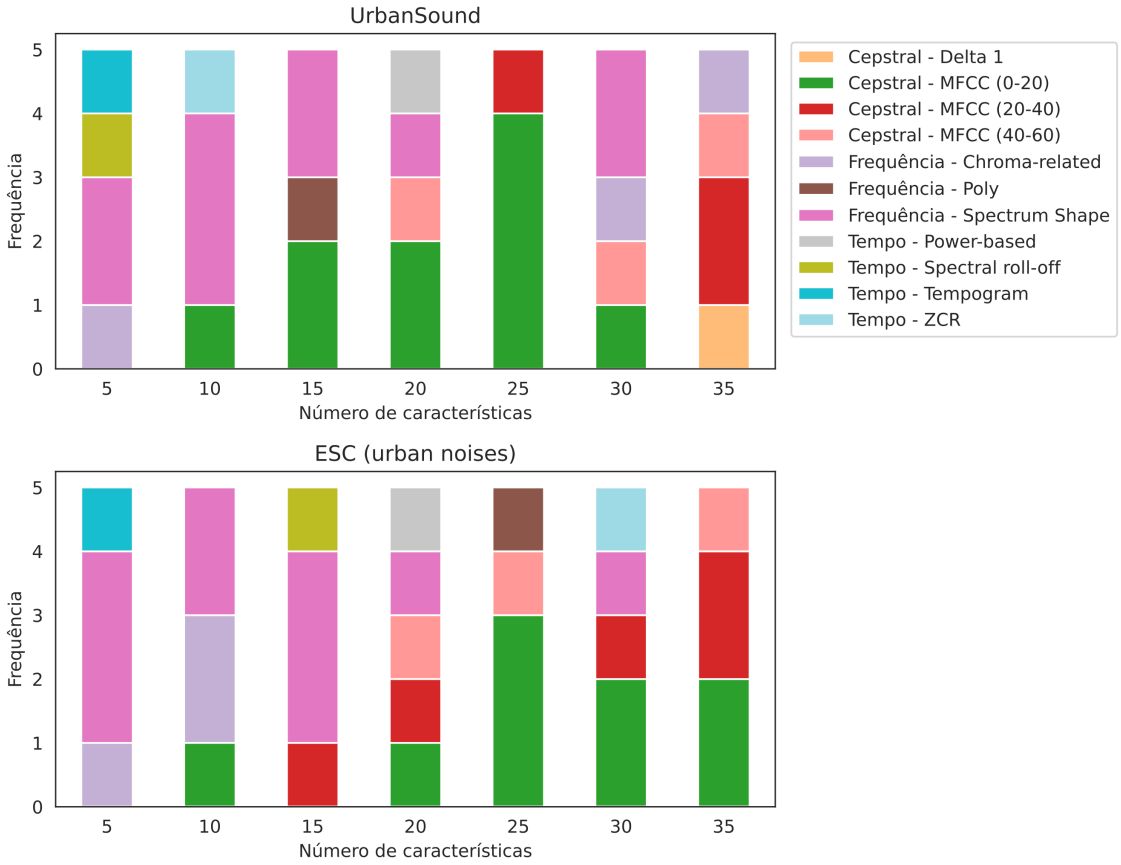


Figura 4 – Distribuição das 35 melhores características selecionadas pelo *Information Gain*.

Para definir o número de árvores a ser utilizado, foram realizados testes variando na faixa de 50 a 450, incrementado em 50. Depois de definir o número mais adequado nesta escala, foi avaliado o desempenho do classificador em incrementos de 10. O algoritmo de classificação *SVM*, usado para classificação e análise de regressão, teve seu desempenho avaliado usando quatro funções de *kernel*: *radial, linear, polynomial e sigmoid*. O *kernel sigmoid* obteve os melhores resultados para os dois conjuntos de dados.

A Figura 5 apresenta o desempenho do classificador *Random Forest* de acordo com a variação do número de árvores selecionadas. Com 350 árvores, o RF obteve o melhor desempenho, de 94,21%, para a UrbanSound. Considerando o conjunto de dados ESC (urban noises), o *Random Forest* obteve a maior acurácia em 50 árvores, com 71,25%. No entanto, os incrementos de 10 mostraram que 70 árvores apresentaram melhor desempenho, com 72,5% de acerto. O número inferior de amostras na base ESC (urban noises) faz com que seja necessário um numero menor de árvores, isso porque, aumentando o número de arvores, aumenta também as chances de overfitting para o conjunto.

A configuração usada para todos os experimentos do conjunto de treinamento UrbanSound foi, 350 árvores e *seed* (semente) igual a 1, para o classificador *Random Forest*, e *kernel sigmoid* para o classificador *SVM*. Já para o conjunto ESC (urban noises), o *Random Forest* com 70 árvores e *seed* igual a 1, e o *SVM* também com o *kernel sigmoid*.

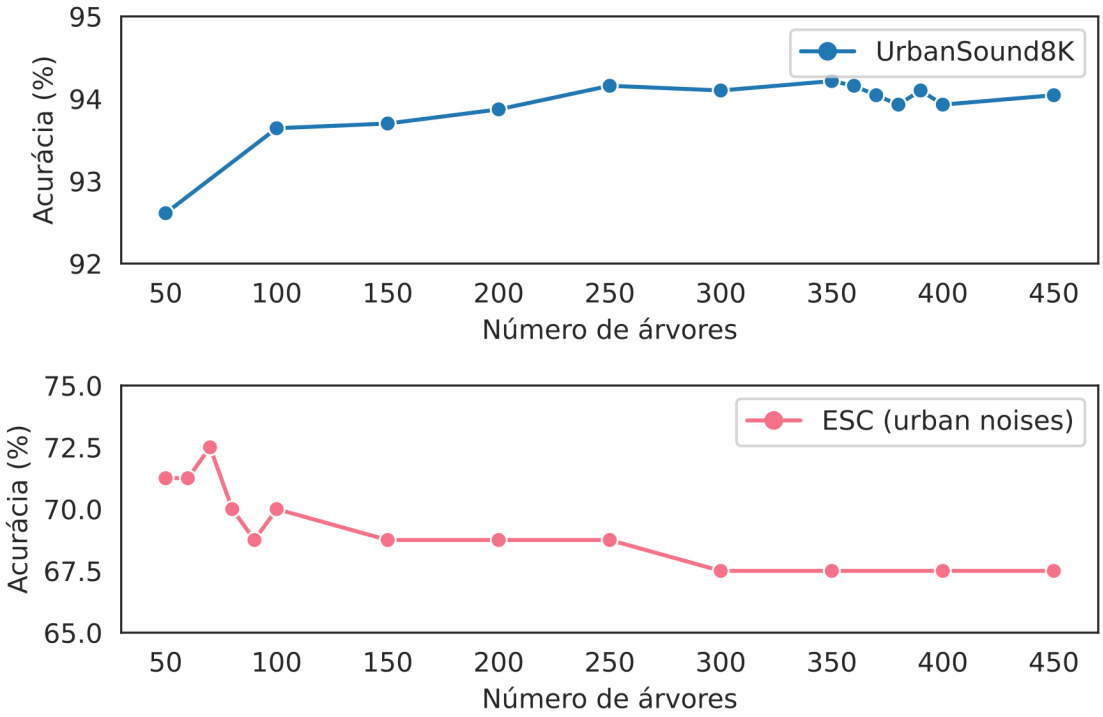


Figura 5 – Acurácia de acordo com o número de árvores do classificador *RF*.

Tabela 6 – Resultados com e sem seleção de características. Os números entre parênteses representam a dimensão do vetor.

		Sem seleção de características			**Com seleção de características**		
Conjunto de dados	**Classificador**	**Acurácia**	**Kappa**	**AUC**	**Acurácia**	**Kappa**	**AUC**
UrbanSound8K	**RF**	**94,215% (82)**	0,9351	0,996	**94,100% (80)**	0,9338	0,996
UrbanSound8K	**SVM**	84,822% (82)	0,8296	0,914	84,937% (80)	0,8309	0,915
ESC (urban noises)	**RF**	**72,500% (82)**	0,6927	0,952	72,500% (18)	0,6917	0,961
ESC (urban noises)	**SVM**	71,250% (82)	0,6768	0,839	**78,750% (70)**	0,7606	0,881

Observando a Tabela 6, é possível notar que, aplicando a seleção de características sobre o descritor, pode-se chegar a uma redução de dimensionalidade. Isso pode ser melhor observado no caso do conjunto de dados ESC (urban noises), na classificação através do RF, que, apesar de não ter alcançado melhores resultados de acurácia, teve sua dimensionalidade reduzida de 82 posições para apenas 18. Também no conjunto de dados ESC (urban noises), porém com classificador *SVM*, a seleção de características não só garantiu a redução de dimensionalidade, passando de 82 para 70 características no descritor, como também melhorou o desempenho de classificação com um aumento de 7,5% na acurácia, indo de 71,250% para 78,750%.

Já o conjunto de dados UrbanSound8K não mostrou nenhuma mudança significativa após a seleção de características, o que já era esperado baseado no padrão observado na Figura 3, onde, de acordo com que novas características iam sendo acrescentadas, sua acurácia aumentava. Também na Figura 3, é possível observar que, diferente do comportamento do conjunto ESC (urban noises), que se mostrou irregular tendo grandes mudan-

ças de acurácia dependendo do subconjunto de características, o conjunto UrbanSound8K se manteve sempre constante. Isso pode ser justificado pela diferença na quantidade de amostras por base.

4.3 Comparação com a literatura

A Tabela 7 sumariza uma comparação entre os resultados da abordagem proposta neste trabalho e os resultados dos estudos descritos no Capítulo 2.7. Os critérios de comparação, além dos resultados, foram a dimensão do descritor e o fato de utilizar ou não a seleção de características. Vale ressaltar que, alguns dos trabalhos relacionados possuem valores de acurácia para base *ESC*, relacionados a todo o conjuntos de dados ou um subconjunto de dez classes (não relacionado ao som urbano). Já este trabalho apresenta valores de acurácia da *ESC (urban noise)*, subconjunto formado pelas 10 classes de sons urbanos da *Esc-50*.

Tabela 7 – Comparação do método proposto com os trabalhos relacionados.

Trabalho	**Seleção de características**	**Extração com CNN**	**Dimenção do vetor**	**UrbanSound8K (Acurácia)**	**Esc-10/Esc-50 (Acurácia)**
Silva et al. (2019)	Sim	Não	90	53.5%	-/46.9%
Giannakopoulos, Spyrou e Perantonis (2019)	Não	Sim	580	79.3%	-/54.2%
Su et al. (2019)	Não	Sim	LMC(41 x 85) MC(41 × 85)	97.2%	-
Mushtaq e Su (2020)	Não	Sim	Melspectrogram-(28 x 128) Logmel(128x128) MFCC(20x128)	95.37%	94.94%/89.28%
Alves et al. (2020)	Não	Não	46	73%	-
Este trabalho	Sim	Não	82	94.2%	72,500%/-

A Tabela 7 mostra que o descritor proposto neste trabalho consegue representar, de maneira satisfatória, os dados para a classificação entre os diferentes tipos de sons urbanos, tendo em vista que a acurácia alcançada foi mais alta até mesmo que os modelos que utilizaram *CNNs* para a extração de características. Além disso, o descritor proposto apresentou uma dimensionalidade menor que os estudos relacionados, o que contribui para uma otimização na integração dessas características em ambientes de produção.

O trabalho de Mushtaq e Su (2020) apresentou uma acurácia superior ao método proposto neste trabalho. Contudo, os autores aplicaram representações diferentes para os sinais acústicos, com exceção do *MFCC*, que também foi utilizado neste trabalho. Ademais, por utilizar uma *CNN*, as representações utilizadas pelos autores podem ter impactado positivamente os resultados, proporcionando um melhor desempenho. Todavia, esta abordagem traz consigo um aumento no custo computacional em virtude da utilização da *CNN*, em contraste ao nosso método, que utiliza um descritor com uma dimensionalidade menor, mas que apresentou resultados promissores.

4.4 Discussão

Com base nos resultados apresentados na Seção 4, é possível perceber que diferentes conjuntos de características podem proporcionar diferentes desempenhos ao classificador, não sendo o ranqueamento de características fornecido pelo *Information Gain* o único determinante do grau de importância das característica para distinção da classe sonora. Isso pode ser justificado através da observação do comportamento do conjunto de dados ESC (urban noises), uma vez que, mesmo adicionando características ranqueadas em boas posições pelo seletor, a acurácia sofre uma queda. A presença de informação com ruído nessas características pode explicar porque isso acontece.

O conjunto de dados *UrbanSound8K*, apesar de não ter obtido um melhor resultado com a redução de características, para um subconjunto de 35 características obteve uma acurácia muito próxima ao melhor resultado, com uma diferença um pouco maior que 1% . A redução da dimensionalidade produz um descritor mais compacto, isso pode ser favorável para adoção do descritor proposto em aplicações embarcadas, onde um menor custo computacional é desejado.

Percebe-se também que as características espectrais possuíram maior influência para a caracterização dos sons urbanos, tendo em vista a distribuição demostrada na Figura 4, onde é possível perceber que as características espectrais se encontram, em ambos os conjuntos de dados, entre as primeiras 15 características ranqueadas. Dentre as características *MFCCs*, que foram divididas em 3 grupos com 20 coeficientes cada, o primeiro conjunto também se mostrou presente, em ambos os conjuntos, entre as características ranqueadas como de maior importância.

Durante os testes de parâmetro para o algoritmo de classificação *RF*, nos conjuntos UrbanSound8K e ESC (urban noises), o modelo alcançou melhor desempenho com diferentes números de árvores, sendo 350 para o UrbanSound8K e 70 para o ESC (urban noises). É possível que o baixo número de amostras disponível no conjunto ESC (urban noises) tenha influenciado nas diferenças existentes entre os resultados alcançados pelas bases, já que, por possuir um menor número de amostras.

Detre os modelos de classificação aplicados o algoritmo *RF* se destacou obtendo os melhores resultados. Nos testes realizados com o descritor em sua dimensão total, o *RF* chegou a melhores valores de acurácia em ambos os conjuntos de dados, tendo uma diferença de 9,39% em comparação ao *SVM* no conjunto UrbanSound8K. Nos testes após ser aplicada a seleção de características o classificador *RF* também alcançou um melhor resultado que o *SVM* no conjunto UrbanSoun8K. No entanto, no conjunto de dados Esc (urban noises) o algoritmo de classificação *SVM* obteve uma melhor acurácia, com 6,25% a mais em relação ao *RF*.

5 Conclusão

Este trabalho apresentou uma metodologia para a extração de características manuais para classificação de som ambiental. Além disso, a etapa de seleção de características resultou em descritores compactos com dimensão 80 e 70 que alcançaram 94,1 e 78.75% de acurácia, respectivamente, para os conjuntos de dados UrbanSound8K e ESC (urban noises). Levando em consideração o resultado inicial alcançado pelo descritor para o conjunto Esc(urban noise) (71,25%) e o resultado após a seleção de características(78.75%) , percebe-se um aumento de 7,5% na acurácia. Pode-se concluir então que, a aplicação do método de seleção de características adequado pode melhorar o desempenho da classificação. Ao mesmo tempo, reduz notavelmente o tamanho do descritor, podendo diminuir em alguns casos até 78% do tamanho original, o que torna esses descritores adequados para serem usados em aplicativos de reconhecimento de som móvel ou sistemas embarcados.

Com base nos resultados alcançados, é possível concluir que a extração manual de características em conjunto com a etapa de seleção é, de fato, uma forma eficiente de se obter um descritor capaz de realizar a discriminação de diferentes classes de sons encontradas no ambiente urbano. Na base UrbanSound8K, o método proposto alcançou acurácia de 94,1%, *AUC* de 0,99, *Kappa* de 0,93 e um descritor com dimensão 80, sendo o índice *Kappa* considerado excelente de acordo com a categorização de Cohen (1960). Já na base ESC (urban noises), apesar das poucas amostras, o método obteve acurácia de 78,75%, *AUC* de 0,88, *Kappa* de 0,76 e um descritor com dimensão 70, com um *Kappa* considerado muito bom, de acordo com a categorização de Cohen (1960).

Apesar do melhor resultado na UrbanSound8K ter sido sem a seleção de características, a diferença entre os resultados com e sem seleção foi quase imperceptível, com apenas 0.11% de diferença. Outro ponto importante é que, no teste executado com 35 características, o método obteve uma acurácia de 93,24% na UrbanSound8K, bem próximo ao resultado máximo alcançado pelo método. Sendo assim, nota-se que a utilização da seleção torna-se mais relevante para o método, levando em consideração que a redução da dimensionalidade do descritor pode reduzir o custo computacional da classificação.

Em comparação aos estudos relacionados, a abordagem proposta neste trabalho mostrou-se promissora diante dos dois conjuntos de dados utilizados na avaliação. Além disso, o descritor proposto resultou em uma dimensionalidade menor de características, o que pode reduzir o custo computacional em comparação a outros métodos comumente utilizados, como no caso de algumas redes neurais. Em suma, este trabalho propôs um descritor com baixa dimensionalidade e com um significante desempenho de classificação , concentrando informações mais relevantes do problema em decorrência da seleção de características.

Em relação às limitações deste trabalho, vale destacar o não tratamento do desbalanceamento da base UrbanSound8K, que possui um número limitado de amostras nas

classes Buzina de carro e Tiro, o que pode impactar na classificação. Além disso, um aumento de dados poderia ser útil para melhorar os resultados na base ESC (urban noises), já que ela possui poucas amostras por classe, sendo esta uma dificuldade para obter um bom desempenho nos modelos de classificação. Com o intuito de mitigar esses problemas e melhorar os resultados no geral, pretende-se realizar os seguintes trabalhos futuros:

- Realizar o balanceamento da base UrbanSound8K através de técnicas de *undersampling* ou *oversampling*;
- Aplicar um aumento de dados na base ESC (urban noises) para ampliar o conjunto de dados em cada classe, podendo ser feito através da concatenação de classe ou aplicando técnicas de transformações nos áudios;
- Fundir as características manuais com características de *CNNs* para verificar se a união destes descritores proporcionam melhores resultados;
- Avaliar outros classificadores presentes na literatura, como o *XGBoost*, o *MLP* e o *KNN*;
- Avaliar outras bases de sons urbanos disponíveis publicamente;
- Avaliar outros algoritmos de seleção de características ou de redução da dimensionalidade, como o *PCA*.

6 Publicações

Os resultados apresentados neste trabalho foram compilados em formato de artigo, nomeado *Ensemble of Handcrafted and Deep Features for Urban Sound Classification*, e publicado na revista Applied Acoustics, Qualis(A2), fator de impacto(2,44).

A autora também obteve uma publicação no capítulo de livro intitulado "Processamento e Análise de Sinais Acústicos em Cenários Urbanos com ConvNets: Teoria e Prática", apresentado no evento ENUCOMPI - Encontro Unificado de Computação do Piauí, 2019.

Referências

AL-NASHERI, A. et al. Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions. *IEEE Access*, IEEE, v. 6, p. 6961–6974, 2017. Citado 2 vezes nas páginas 18 e 19.

ALVES, J. et al. Urban sound event detection and classification. *i-ETC: ISEL Academic Journal of Electronics Telecommunications and Computers*, v. 6, n. 1, p. 2, 2020. Citado 2 vezes nas páginas 24 e 34.

BARATLOO, A. et al. Part 1: simple definition and calculation of accuracy, sensitivity and specificity. ARCHIVES OF ACADEMIC EMERGENCY MEDICINE (EMERGENCY), 2015. Citado na página 22.

BARKER, T.; VIRTANEN, T. Blind separation of audio mixtures through nonnegative tensor factorization of modulation spectrograms. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, IEEE, v. 24, n. 12, p. 2377–2389, 2016. Citado na página 17.

BELLO, J. P.; MYDLARZ, C.; SALAMON, J. Sound analysis in smart cities. In: *Computational Analysis of Sound Scenes and Events*. [S.l.]: Springer, 2018. p. 373–397. Citado na página 12.

BORGES, J. et al. Feature engineering for crime hotspot detection. In: IEEE. *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*. [S.l.], 2017. p. 1–8. Citado na página 12.

BROWN, J. C. Calculation of a constant q spectral transform. *The Journal of the Acoustical Society of America*, Acoustical Society of America, v. 89, n. 1, p. 425–434, 1991. Citado na página 18.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 37–46, 1960. Citado 2 vezes nas páginas 22 e 36.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado 2 vezes nas páginas 21 e 22.

DUBNOV, S. Generalization of spectral flatness measure for non-gaussian linear processes. *IEEE Signal Processing Letters*, IEEE, v. 11, n. 8, p. 698–701, 2004. Citado 2 vezes nas páginas 18 e 19.

EDWARDS, J. Signal processing opens the internet of things to a new world of possibilities: Research leads to new internet of things technologies and applications [special reports]. *IEEE Signal Processing Magazine*, IEEE, v. 35, n. 5, p. 9–12, 2018. Citado 2 vezes nas páginas 12 e 25.

ELLIS, D. Chroma feature analysis and synthesis. *Resources of Laboratory for the Recognition and Organization of Speech and Audio-LabROSA*, 2007. Citado na página 18.

EPPOLITO, A. *Multi-channel sound panner*. [S.l.]: Google Patents, 2008. US Patent App. 11/786,863. Citado na página 16.

ESMAIL, M. Y.; AHMED, D. H.; ELTAYEB, M. Classification system for heart sounds based on random forests. *Journal of Clinical Engineering*, LWW, v. 44, n. 2, p. 76–80, 2019. Citado na página 21.

GIANNAKOPOULOS, T.; SPYROU, E.; PERANTONIS, S. J. Recognition of urban sound events using deep context-aware feature extractors and handcrafted features. In: SPRINGER. *IFIP International Conference on Artificial Intelligence Applications and Innovations*. [S.l.], 2019. p. 184–195. Citado 4 vezes nas páginas 17, 23, 24 e 34.

GROSCHE, P.; MÜLLER, M.; KURTH, F. Cyclic tempogram—a mid-level tempo representation for musicsignals. In: IEEE. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.], 2010. p. 5522–5525. Citado na página 17.

HALL, M. et al. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, ACM New York, NY, USA, v. 11, n. 1, p. 10–18, 2009. Citado na página 29.

HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, v. 143, n. 1, p. 29–36, 1982. Citado na página 23.

HARTE, C.; SANDLER, M.; GASSER, M. Detecting harmonic change in musical audio. In: *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*. [S.l.: s.n.], 2006. p. 21–26. Citado 2 vezes nas páginas 18 e 19.

IAZZETTA, F. *Tutoriais de áudio e acústica*. [S.l.]: São Paulo: Departamento de Música da ECA-USP. Disponível em http://www. eca . . . , 2010. Citado na página 16.

JIANG, D.-N. et al. Music type classification by spectral contrast feature. In: IEEE. *Proceedings. IEEE International Conference on Multimedia and Expo*. [S.l.], 2002. v. 1, p. 113–116. Citado 2 vezes nas páginas 18 e 19.

KAUR, K. Modified info gain attribute eval feature selection algorithm to increase efficiency of classification algorithms in data mining. *Journal of the Gujarat Research Society*, v. 21, n. 10s, p. 199–209, 2019. Citado na página 20.

KOIZUMI, Y. et al. Toyadmos: A dataset of miniature-machine operating sounds for anomalous sound detection. In: IEEE. *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. [S.l.], 2019. p. 313–317. Citado na página 16.

LIU, M.-K.; TSENG, Y.-H.; TRAN, M.-Q. Tool wear monitoring and prediction based on sound signal. *The International Journal of Advanced Manufacturing Technology*, Springer, v. 103, n. 9-12, p. 3361–3373, 2019. Citado na página 17.

MCFEE, B. et al. librosa: Audio and music signal analysis in python. In: *Proceedings of the 14th python in science conference*. [S.l.: s.n.], 2015. v. 8, p. 18–25. Citado na página 27.

MÜLLER, M.; EWERT, S. Chroma toolbox: Matlab implementations for extracting variants of chroma-based audio features. In: CITESEER. *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), 2011. hal-00727791, version 2-22 Oct 2012*. [S.l.], 2011. Citado na página 18.

MULLER, M.; KURTH, F.; CLAUSEN, M. Chroma-based statistical audio features for audio matching. In: IEEE. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.* [S.l.], 2005. p. 275–278. Citado na página 19.

MUSHTAQ, Z.; SU, S.-F. Environmental sound classification using a regularized deep convolutional neural network with data augmentation. *Applied Acoustics*, Elsevier, v. 167, p. 107389, 2020. Citado 2 vezes nas páginas 24 e 34.

OLIVEIRA, J. Z. d. Assimetria funcional dos hemisférios cerebrais na percepção de timbre, intensidade ou altura, em contexto musical. 2002. Citado 2 vezes nas páginas 15 e 16.

PICHLMAIR, M.; KAYALI, F. Levels of sound: On the principles of interactivity in music video games. In: CITESEER. *DiGRA Conference*. [S.l.], 2007. Citado na página 16.

PICZAK, K. J. Esc: Dataset for environmental sound classification. In: *Proceedings of the 23rd ACM international conference on Multimedia*. [S.l.: s.n.], 2015. p. 1015–1018. Citado na página 25.

RATHORE, M. M. et al. Urban planning and building smart cities based on the internet of things using big data analytics. *Computer Networks*, Elsevier, v. 101, p. 63–80, 2016. Citado na página 12.

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*. [S.l.: s.n.], 2014. p. 1041–1044. Citado na página 25.

SCHÖRKHUBER, C.; KLAPURI, A. Constant-q transform toolbox for music processing. In: *7th Sound and Music Computing Conference, Barcelona, Spain*. [S.l.: s.n.], 2010. p. 3–64. Citado na página 18.

SERIZEL, R. et al. Acoustic features for environmental sound analysis. In: *Computational analysis of sound scenes and events*. [S.l.]: Springer, 2018. p. 71–101. Citado 2 vezes nas páginas 17 e 20.

SHARMA, G.; UMAPATHY, K.; KRISHNAN, S. Trends in audio signal feature extraction methods. *Applied Acoustics*, Elsevier, v. 158, p. 107020, 2020. Citado na página 17.

SILVA, B. da et al. Evaluation of classical machine learning techniques towards urban sound recognition on embedded systems. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 9, n. 18, p. 3885, 2019. Citado 3 vezes nas páginas 23, 24 e 34.

SOUZA, T. I. et al. Um método para detecção e diagn óstico de outliers em dados urbanos via análise multidimensional. In: SBC. *Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. [S.l.], 2018. Citado na página 12.

SU, Y. et al. Environment sound classification using a two-stream cnn based on decision-level fusion. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 19, n. 7, p. 1733, 2019. Citado 3 vezes nas páginas 23, 24 e 34.

VIRTANEN, T.; PLUMBLEY, M. D.; ELLIS, D. Introduction to sound scene and event analysis. In: *Computational analysis of sound scenes and events*. [S.l.]: Springer, 2018. p. 3–12. Citado 4 vezes nas páginas 12, 13, 15 e 20.

VOGADO, L. H. et al. Leukemia diagnosis in blood slides using transfer learning in cnns and svm for classification. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 72, p. 415–422, 2018. Citado na página 21.

WU, C.-W.; VINTON, M. Blind bandwidth extension using k-means and support vector regression. In: IEEE. *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. [S.l.], 2017. p. 721–725. Citado 2 vezes nas páginas 18 e 19.

XIE, J.; ZHU, M. Investigation of acoustic and visual features for acoustic scene classification. *Expert Systems with Applications*, Elsevier, v. 126, p. 20–29, 2019. Citado na página 17.

YANG, X. et al. On the design of solfeggio audio machine assessment system. In: IEEE. *2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN)*. [S.l.], 2019. p. 234–238. Citado na página 18.

TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA "JOSÉ ALBANO DE MACEDO"

Identificação do Tipo de Documento

( ) Tese

( ) Dissertação

(X) Monografia

( ) Artigo

Eu, **Myllena Caetano de Oliveira**, autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de 02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar, gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação **Extração Manual De Características Para Classificação De Sons Urbanos** de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título de divulgação da produção científica gerada pela Universidade.

Picos-PI 26 de Julho de 2021.

Myllena Caetano de Oliveira

Assinatura