

Bianca Sousa Soares
Orientadora: Deborah Maria Vieira Magalhães

Deteccção da presença da abelha rainha através da classificação e processamento de áudio

Picos - PI
19 de janeiro de 2024

Bianca Sousa Soares
Orientadora: Deborah Maria Vieira Magalhães

Detecção da presença da abelha rainha através da classificação e processamento de áudio

Monografia submetida ao Curso de Bacharelado em Sistemas de Informação como requisito parcial para a obtenção do grau de bacharel em Sistemas de Informação.

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

Picos - PI
19 de janeiro de 2024

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca José Albano de Macêdo

S676d Soares, Bianca Sousa

Detecção da presença da abelha rainha através da classificação e processamento de áudio./ Bianca Sousa Soares. – 2024.
42 f.

1 Arquivo em PDF

Indexado no catálogo *online* da biblioteca José Albano de Macêdo-CSHNB
Aberto a pesquisadores, com restrições da Biblioteca

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do Piauí, Curso de Bacharelado em Sistemas de Informação, Picos, 2024.
“Orientadora: Profa. Deborah Maria Vieira Magalhães”

1. Apicultura. 2. Colméia-monitoramento. 3. Abelha-estudo.
4. Bioacústica. I. Magalhães, Deborah Maria Vieira. II. Título.

CDD 638.1

Elaborado por Sérvulo Fernandes da Silva Neto CRB 15/603

DETECÇÃO DA PRESENÇA DA ABELHA RAINHA ATRAVÉS
DA CLASSIFICAÇÃO E PROCESSAMENTO DE ÁUDIO

BIANCA SOUSA SOARES

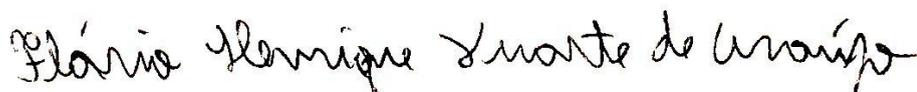
Monografia aprovada como exigência parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Data de Aprovação

Picos – PI, 26 de janeiro de 2024



Profª. Deborah Maria Vieira Magalhães



Prof. Flávio Henrique Duarte de Araújo



Prof. Romuere Rodrigues Veloso e Silva

Agradecimentos

Gostaria de expressar meus sinceros agradecimentos a todos que contribuíram de forma direta ou indireta para a conclusão desse trabalho.

Em primeiro lugar, agradeço a minha família, em especial, agradeço aos meus pais Luiza e Franciz, ao meu irmão Bruno, meus tios Maria, Edimilson, José e Agnaldo. O apoio e incentivo de vocês foi essencial para que eu pudesse superar os desafios e conquistar essa realização.

Aos meus amigos que conheci graças a esse curso: Matheus Victor, Milene, Millena, Vinicius e Eurico. Obrigada por cada momento compartilhado e pela amizade que se desenvolveu ao longo dessa jornada.

Por fim, quero expressar minha gratidão à minha orientadora e amiga Deborah, pelo seu apoio e orientação durante todo o processo de desenvolvimento desse trabalho, seus conselhos e críticas construtivas foram fundamentais para que eu pudesse atingir meus objetivos.

Em algum lugar, algo incrível está esperando para ser descoberto.

Carl Sagan

Resumo

Monitorar e detectar a saúde de uma colmeia são aspectos cruciais para o desenvolvimento de estratégias voltadas à conservação do ecossistema e à prática sustentável da apicultura. A ausência da abelha rainha representa um sinal de alerta, indicando uma condição anormal na colmeia. Na literatura, encontram-se abordagens promissoras de aprendizado de máquina e aprendizado profundo para detectar a presença da abelha rainha por meio da análise espectral temporal dos sons emitidos pela colmeia. Apesar de existirem estudos que consideram características manuais e profundas para abordar essa problemática, a falta de uma seleção específica de características dificulta a identificação dos descritores mais discriminativos para a presença/ausência da abelha rainha. Além disso, as técnicas de aprendizado profundo frequentemente resultam em descritores de alta dimensão e exigem considerável capacidade computacional. Nesse contexto, este trabalho realiza uma análise de técnicas de extração e seleção de características visando obter descritores eficientes e compactos capazes de realizar classificações em cenários de monitoramento em tempo real. Os resultados indicam que a combinação de características cepstrais, temporais e de frequência alcançaram 0,99 nas métricas acurácia, kappa e sensibilidade, superando diversos modelos de ponta para a classificação da presença da abelha rainha, inclusive modelos de redes neurais convolucionais. A etapa de seleção de características reduz significativamente o tamanho do descritor, preservando a eficácia da classificação. Os descritores de 15 e 31 MFCCs apresentam baixa dimensão e baseiam-se em apenas uma característica, reduzindo assim a carga computacional e dependências de bibliotecas. Com base nesses resultados, esses descritores podem ser implementados em dispositivos de processamento limitado para diagnósticos não intrusivos e em tempo real da saúde das colmeias.

Palavras-chaves: Monitoramento de colmeias, Processamento de áudio, Aprendizado de máquina, Seleção de características, Bioacústica.

Abstract

Monitoring and detecting the health of a beehive are crucial aspects for the development of strategies aimed at ecosystem conservation and sustainable beekeeping practices. The absence of the queen bee serves as a warning sign, indicating an abnormal condition within the hive. In the literature, promising approaches from machine learning and deep learning are found for detecting the presence of the queen bee through the temporal spectral analysis of sounds emitted by the hive. Despite studies considering manual and deep features to address this issue, the lack of specific feature selection hinders the identification of the most discriminative descriptors for the presence/absence of the queen bee. Additionally, deep learning techniques often result in high-dimensional descriptors and require considerable computational capacity. In this context, this work performs an analysis of feature extraction and selection techniques to obtain efficient and compact descriptors capable of real-time monitoring classifications. The results indicate that the combination of cepstral, temporal, and frequency features achieves an accuracy, kappa, and sensitivity of 0.99, surpassing various state-of-the-art models for classifying the presence of the queen bee, including convolutional neural network models. The feature selection stage significantly reduces the descriptor's size while preserving classification effectiveness. Descriptors with 15 and 31 Mel-Frequency Cepstral Coefficients (MFCCs) have low dimensions and are based on only one feature, thereby reducing computational load and library dependencies. Based on these results, these descriptors can be implemented on low-processing devices for non-intrusive, real-time diagnostics of beehive health.

Keywords: Beehive monitoring, Audio processing, Machine learning, Feature selection, Bioacoustics.

Lista de ilustrações

Figura 1 – Representação da onda sonora.	15
Figura 2 – Metodologia para detecção da presença da abelha rainha.	24
Figura 3 – Distribuição das 58 características extraídas das amostras de 2018. . .	29

Lista de tabelas

Tabela 1	–	Comparação dos trabalhos relacionados.	22
Tabela 2	–	Distribuição por ano das amostras de áudio.	24
Tabela 3	–	Parâmetros para a geração dos Mel-espectrogramas.	25
Tabela 4	–	Descrição das características extraídas nos domínios cepstral, frequência e tempo. O número entre parênteses na coluna <i>Característica</i> indica a dimensão correspondente da característica.	26
Tabela 5	–	Configuração dos hiperparâmetros dos classificadores otimizada por meio do algoritmo <i>Grid Search</i>	27
Tabela 6	–	Resultados de classificação dos descritores extraídos das amostras de 2018.	29
Tabela 7	–	Resultados de classificação dos descritores extraídos das amostras de 2017 com a estratégia de validação cruzada de 5 conjuntos. O número entre parênteses indica o desvio padrão correspondente.	30
Tabela 8	–	Resultados de classificação das características profundas extraídas pelos modelos CNN. Os melhores resultados são destacados em negrito, e os números entre parênteses correspondem ao tamanho do descritor.	30
Tabela 9	–	Comparação do método proposto com os trabalhos relacionados.	31

Lista de abreviaturas e siglas

CENS	Chroma Energy Normalized
CQT	Constant-Q Transform
CNN	Convolutional Neural Network
CV	Cross Validation
DL	Deep Learning
IoT	Internet of Things
HHT	Hilbert-Hang Transform
Mel	Mel spectrogram
MFCC	Mel Frequency Cepstral Coefficient
ML	Machine Learning
RF	Random Forest
RFE	Recursive Feature Elimination
RFECV	Recursive Feature Elimination with Cross Validation
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
ZCR	Zero Crossing Rate

Sumário

1	Introdução	11
1.1	Problema	12
1.2	Objetivos	12
1.3	Organização do trabalho	13
2	Referencial Teórico	14
2.1	Monitoramento de colmeias	14
2.2	Som	14
2.3	Pré-processamento	15
2.4	Extração de características	16
2.4.1	Características manuais	16
2.4.2	Características Profundas	17
2.5	Seleção de características	18
2.6	Classificação	18
2.7	Métricas de avaliação	19
3	Trabalhos Relacionados	21
4	Materiais e métodos	23
4.1	Aquisição dos áudios	23
4.2	Pré Processamento	23
4.3	Extração de características	25
4.3.1	Características manuais	25
4.3.2	Características profundas	25
4.4	Seleção de características	25
4.5	Classificação	26
4.6	Validação	27
5	Resultados	28
5.1	Seleção de características	28
5.2	Classificação	29
5.3	Discussão	30
5.4	Comparação com a literatura	31
6	Conclusão	33
7	Publicações	34

Referências **35**

1 Introdução

Além das abelhas produzirem mel, cera, geléia real e própolis, elas têm um papel essencial na polinização de florestas e plantações (NOLASCO et al., 2019). No mundo, a taxa de crescimento de culturas dependente de polinização cresceu 300%, enquanto a população de abelhas está em declínio, muitas vezes as abelhas adultas abandonam repentinamente suas colmeias em um fenômeno chamado de transtorno do colapso das colônias (FAROOQUI, 2013). Vários fatores podem estar associados a esse fenômeno, entre eles: doenças e o uso de pesticidas.

Desse modo, o som pode atuar como uma importante fonte de informação a respeito da saúde da colmeia, pois, as abelhas produzem assinaturas de sons únicas, em resposta a temperaturas internas elevadas da colmeia, ausência da rainha, presença de ameaças e exposição a produtos nocivos. A ausência da rainha em uma colmeia é um forte indicador da necessidade de intervenção do apicultor. Porém, procurar manualmente a rainha é uma tarefa que demanda tempo e interrompe o ciclo de vida da colmeia, além de ser uma fonte de estresse para as abelhas. (NOLASCO et al., 2019).

A agricultura digital é uma das principais tecnologias que vem ganhando apoio através de diferentes iniciativas de pesquisa na última década (ZGANK, 2020), o monitoramento do áudio em tempo real é uma alternativa não-intrusiva para a análise do comportamento das colmeias, fazendo uso do aprendizado de máquina e redes de sensores. Ele pode ser usado em diferentes aplicações, entre elas, previsão da enxameação, detecção de parasitas como o ácaro varroa (*Varroa destructor*), e na detecção da presença da rainha.

O monitoramento acústico em uma colmeia é desafiador, pois existem diversos fatores que podem interferir na coleta dos dados, por exemplo, posicionamento do microfone, o tipo de dispositivo utilizado na captura, o espaço de propagação acústica, entre outros. Ainda, os fenômenos na natureza são dinâmicos e intermitentes. Além disso, a qualidade e quantidade dos dados coletados impactam diretamente no desempenho do algoritmo de classificação. Esses fatores desafiam pesquisas na área, impactando na disponibilidade de bases públicas.

Técnicas de aprendizado profundo foram aplicadas com sucesso para detectar a saúde da colmeia ((KIM; OH; HEO, 2021); (NOLASCO et al., 2019)), porém resultaram em descritores com grandes dimensões e alto custo computacional. Quando esses descritores são implantados em dispositivos com restrição de recursos computacionais, seu tempo de treinamento e classificação será maior em comparação com algoritmos tradicionais de aprendizado de máquina, o que dificulta a aplicação para o monitoramento em tempo real.

1.1 Problema

O papel das abelhas como agentes polinizadores é essencial para a manutenção de sistemas naturais e agrícolas (TERENZI et al., 2019). Da perspectiva agrícola, no Brasil 68% das 53 principais culturas comerciais dependem em algum grau da polinização por animais. Sua perda representaria um prejuízo na ordem de bilhões e tal impacto seria absorvido pelos agricultores familiares, que representam 74,4% da força de trabalho agrícola no país (NOVAIS et al., 2016). Dessa forma, é fundamental monitorar a saúde da colmeia de modo que o apicultor possa manejar as abelhas de maneira eficaz.

O uso de IoT (*Internet of Things*) tem sido explorado na apicultura a fim de reduzir perdas na produção e evitar redução da população das colmeias (BRAGA et al., 2020). Nesse contexto, o microfone desponta com um dispositivo de baixo custo, robusto e que consome menos bateria. O monitoramento acústico vem sendo amplamente explorado na detecção do estado de saúde de colônias de abelhas por ser um processo menos invasivo comparado a inspeção manual e, assim, reduz o estresse ao qual as abelhas são submetidas.

Embora as técnicas de aprendizado profundo tenham alcançado um desempenho significativo para extração e classificação de características na bioacústica, elas resultam em descritores de alta dimensão. As CNNs apresentam restrições de tempo que comprometem sua viabilidade para monitoramento em tempo real com dispositivos com poder de processamento limitado. Kulyukin, Mukherjee e Amlathe (2018) compararam o desempenho das CNNs com algoritmos clássicos de aprendizado de máquina (ML - *machine learning*) para diferenciar sons de abelhas, grilos e ruídos. Os métodos de ML obtiveram um descritor com tamanho 193, enquanto o vetor de entrada para a CNN tinha um tamanho de 20.000, correspondendo ao sinal de áudio. Os autores concluíram que uma CNN poderia operar com sucesso em um computador Raspberry Pi, porém, levou dezesseis vezes mais tempo para o treinamento do que os métodos de aprendizado de máquina. Logo, não seria uma alternativa viável para o monitoramento em tempo real, devido ao alto custo de processamento.

Diante disso, é importante investigar técnicas de extração e seleção de características para produzir descritores eficientes e compactos, que possam realizar a classificação em cenários de tempo real. Um descritor compacto pode ser embarcado em vários tipos de dispositivos, permitindo que um aplicativo para smartphones seja desenvolvido, facilitando o uso do descritor de forma acessível.

1.2 Objetivos

Este trabalho tem como objetivo geral o desenvolvimento de uma metodologia para extração de características e classificação de dados acústicos que possam contribuir na detecção da presença da abelha rainha em uma colmeia.

Os objetivos específicos são:

1. Aplicar técnicas de pré-processamento nos áudios, com a finalidade de melhorar sua representação;
2. Avaliar o impacto da seleção de características, de modo a definir um grupo de características que seja relevante para discriminação entre as classes de sons e, ainda sim, reduza o custo computacional;
3. Avaliar o desempenho das características extraídas com algoritmos de classificação clássicos.

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 apresenta os trabalhos relacionados, o Capítulo 4 detalha as etapas realizadas na execução do método proposto; o Capítulo 5 apresenta e discute os resultados alcançados em todos os testes realizados com a abordagem proposta; o Capítulo 6 apresenta a conclusão deste trabalho, bem como os trabalhos futuros; e, por fim, o Capítulo 7 lista as publicações alcançadas com o desenvolvimento deste trabalho.

2 Referencial Teórico

Este capítulo busca fornecer uma breve revisão dos tópicos que são relevantes para este trabalho. Estando sub-dividida em abordagens sobre o monitoramento de colmeias (Seção 2.1), o som (Seção 2.2), o pré-processamento dos áudios (Seção 2.3), a extração (Seção 2.4) e seleção (Seção 2.5) de características, à classificação das características extraídas (Seção 2.6) e as métricas de avaliação dos algoritmos de classificação (Seção 2.7).

2.1 Monitoramento de colmeias

O monitoramento de colmeias por meio de tecnologias avançadas tem se mostrado uma ferramenta fundamental para a apicultura, oferecendo uma série de vantagens em comparação com a inspeção manual tradicional (BILIK et al., 2024). Ao recorrer a sensores e dispositivos de monitoramento remoto, torna-se possível realizar um acompanhamento em tempo real das condições no interior da colmeia, incluindo a temperatura, umidade e a atividade das abelhas. Esses dados fornecem informações valiosas sobre a saúde das abelhas, permitindo que os apicultores identifiquem rapidamente qualquer problema e intervenham antes que evoluam para estágios mais críticos.

Além disso, o monitoramento automatizado proporciona uma eficiência significativa em termos de tempo e recursos. Enquanto a inspeção manual demanda que o apicultor abra a colmeia, expondo as abelhas ao estresse e criando riscos de perturbação do ambiente interno, as tecnologias de monitoramento permitem a coleta de dados sem a necessidade de intervenção direta (MEIKLE; HOLST, 2015). Isso reduz o impacto sobre as abelhas, minimiza o tempo gasto pelos apicultores e possibilita uma gestão mais precisa e proativa das colmeias.

Nesse contexto, a análise do áudio das colmeias permite detectar padrões sonoros que indicam mudanças na comunicação das abelhas, sugerindo a ausência da rainha, a ocorrência de enxames ou até mesmo sinais de estresse nas abelhas (ABDOLLAHI; GIOVENAZZO; FALK, 2022). Assim, a capacidade de monitorar o áudio das colmeias oferece aos apicultores uma ferramenta adicional para compreender o estado das colônias, complementando outras formas de monitoramento.

2.2 Som

O som é uma forma de energia que se propaga através de meios materiais, como o ar, por meio de ondas de compressão e rarefação (VIRTANEN; PLUMBLEY; ELLIS, 2018).

Características, como o timbre, a intensidade e a altura, são cruciais para sua identificação e classificação (MESAROS et al., 2021).

O timbre permite diferenciar sons provenientes de fontes distintas, mesmo quando apresentam a mesma altura e intensidade. A intensidade, medida em decibéis, representa a amplitude das ondas sonoras e está diretamente relacionada à percepção de volume. Já a altura, expressa em hertz (Hz), está associada à frequência das ondas e determina se um som é agudo ou grave.

A representação gráfica de uma onda sonora em um gráfico bidimensional é essencial para a análise e compreensão de suas características (ZÖLZER, 2022). Através dessa representação, é possível extrair informações cruciais para o processamento e classificação de áudio. Na Figura 1 tem-se a representação de uma onda sonora, onde o eixo horizontal representa o tempo, enquanto o vertical representa a amplitude da onda.

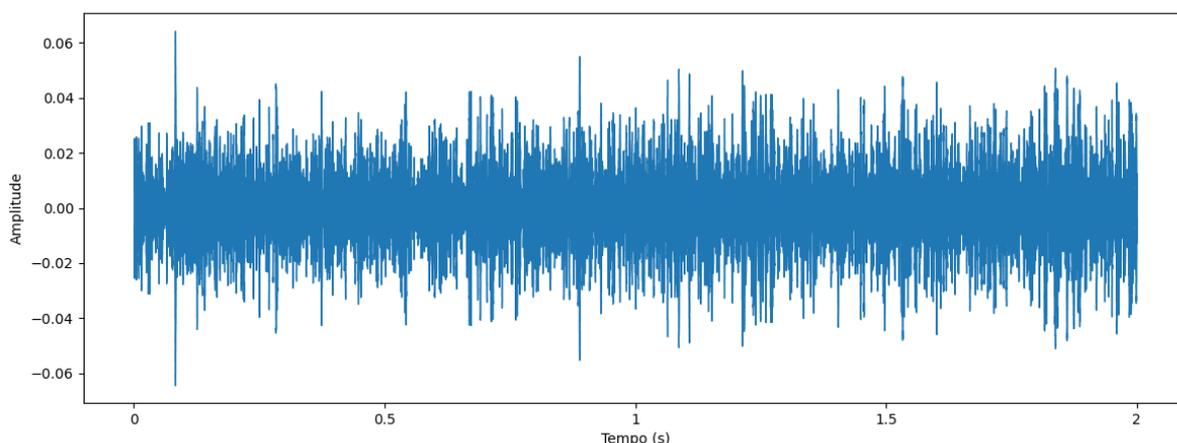


Figura 1 – Representação da onda sonora.

2.3 Pré-processamento

Para que a extração de características possa ser realizada é necessário que o áudio passe por alguns processos. O principal papel dessa etapa é aprimorar determinadas características do sinal recebido, a fim de maximizar o desempenho da análise de áudio nas fases posteriores do sistema de análise. Isso é conseguido reduzindo os efeitos do ruído ou enfatizando os sons alvo no sinal.

Dentre os processos que ocorrem no pré-processamento está o recorte de frame, onde o sinal de áudio é recortado em pequenos quadros de mesmo tamanho ao longo do sinal, onde tipicamente são quantificados entre 20 e 60 ms, tendo um deslocamento, também de tamanho fixo, entre um quadro e outro, de modo que o final de um frame sempre se cruze com o início de outro, com esse cruzamento dos frames sendo geralmente de 50% (VIRTANEN; PLUMBLEY; ELLIS, 2018), este processo se denomina *overlapping*. Os frames resultantes do processo anterior são suavizados com uma função de janelamento

com o intuito de evitar mudanças abruptas nos limites dos frames, pois podem gerar distorções no espectro.

Outra função da fase de pré-processamento é a uniformização dos sinais de áudio, para que se tornem homogêneos. Como os áudios são capturados de diferentes fontes eles podem ter diferentes configurações como, taxa de amostragem, número de canais e quantização, por isso é necessário realizar a uniformização dos dados, para garantir que todos os áudios tenham a mesma configuração e que possam passar pelo treinamento (SERIZEL et al., 2018).

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. 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.

2.4 Extração de características

A análise de áudio geralmente é baseada em características extraídas do sinal de áudio para representar o áudio de uma maneira compacta e não redundante. A extração de características pode ser realizada manualmente, envolvendo a seleção de atributos específicos do sinal, ou automaticamente, através de métodos computacionais .

A extração manual pode incluir características baseadas em aspectos com a intensidade e o timbre, enquanto a extração automática, frequentemente envolve a utilização de características profundas extraídas por redes neurais. As características profundas são representações aprendidas automaticamente a partir dos dados durante o treinamento do modelo, capturando padrões complexos e hierárquicos no sinal de áudio (SHARMA; UMAPATHY; KRISHNAN, 2020).

Para algoritmos de reconhecimento, a propriedade necessária das características sonoras é baixa variabilidade entre as características extraídas de exemplos pertencentes à mesma classe e, ao mesmo tempo, alta variabilidade, permitindo a distinção entre as características extraídas de exemplos pertencentes à diferentes classes (VIRTANEN; PLUMBLEY; ELLIS, 2018).

2.4.1 Características manuais

As características sonoras fornecem uma representação numérica do conteúdo do sinal relevante para o aprendizado de máquina, caracterizando o sinal com valores que têm conexão com suas propriedades físicas, por exemplo, energia do sinal, sua distribuição em

frequência e alterações ao longo do tempo. Existem diversos tipos de características que podem ser extraídas de um áudio, dentre elas estão características dos domínios cepstral, frequência e tempo.

As características cepstrais tem relação com a forma com que a audição humana percebe os sons, principalmente a fala. A característica cepstral mais comum são os Coeficientes cepstrais de frequência mel (MFCCs) (TIWARI, 2010). Os MFCCs são um conjunto de características, geralmente de 10 à 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 discreta do cosseno da energia do log em bandas de frequência de mel, representado pela Equação 2.1.

$$MFCC(t, c) = \sqrt{\frac{2}{M_{mfcc}}} \sum_{m=1}^{M_{mfcc}} \log(\tilde{X}_m(t)) \cos\left(\frac{c(m - \frac{1}{2})\pi}{M_{mfcc}}\right) \quad (2.1)$$

onde M_{mfcc} é o número total de coeficientes calculados e $X_m(t)$ é a magnitude do componente de frequência para um quadro de tempo.

As características do domínio da frequência estão relacionadas com o conteúdo da frequência sonora, um exemplo é o Chroma STFT (Short-Time Fourier Transform - Transformada de Fourier de Curto-Tempo) (LEISTIKOW et al., 2004) que representa informações sobre a classificação do tom e da estrutura do sinal, sendo calculado pela Equação 2.2, onde m é o comprimento da janela de g e n é o número de amostras em f .

$$STFT = F_g f[n, k] = \sum_{m=0}^{m-1} f[n - m]g[m]\epsilon_k[m] \quad (2.2)$$

As características de tempo estão diretamente ligadas com a onda temporal formada pelo áudio, a mesma é computada e as características são extraídas através disso. Uma das características mais utilizadas é o Zero Crossing Rate (ZCR - Taxa de Cruzamento por Zero), que, 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 (GOUYON et al., 2000). A fórmula do ZCR é representada na Equação 2.3, onde N é o número total de amostras no sinal de áudio e $X[n]$ representa a amostra atual do sinal.

$$ZCR = \frac{1}{2} \sum_{n=1}^N |sgn(X[n]) - sgn(X[n - 1])| \quad (2.3)$$

2.4.2 Características Profundas

A extração de características profundas de áudio utilizando redes neurais convolucionais tem se destacado como uma abordagem eficaz para a análise e representação de sinais sonoros (SHARMA; UMAPATHY; KRISHNAN, 2020). Ao utilizar técnicas como o mel-espectrograma como entrada para as CNNs, torna-se possível codificar informações fundamentais relacionadas à distribuição espectral e modulações temporais do áudio.

As CNNs, conhecidas por sua eficiência na análise de padrões espaciais, têm mostrado uma notável capacidade em identificar características complexas em sinais de áudio, como frequências específicas e mudanças temporais cruciais para a compreensão de eventos sonoros (PURWINS et al., 2019). Isso possibilita uma representação mais rica e discriminativa do áudio, superando as limitações de métodos tradicionais de extração de características.

À medida que as camadas da CNN processam o sinal, características mais abstratas e complexas são derivadas, permitindo uma compreensão mais profunda das nuances sonoras presentes no áudio. Essa hierarquia de representações é crucial para a captura de padrões complexos e a modelagem eficiente de relações entre diferentes elementos acústicos (PEETERS; RICHARD, 2021). Como resultado, as CNNs se destacam na extração de características que não apenas identificam a presença de determinados elementos sonoros, mas também entendem a interação e a textura entre eles (WANG et al., 2022).

2.5 Seleção de características

A fim de encontrar inconsistência ou redundância nas características obtidas na etapa de extração, é necessário aplicar uma seleção de características sobre o descritor que foi gerado, 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. Dois algoritmos de seleção de características conhecidos são o RFE (Recursive Feature Elimination - Eliminação Recursiva de Características) e o RFECV (Recursive Feature Elimination with Cross Validation - Eliminação Recursiva de Características com Validação Cruzada)

O RFE (CHEN et al., 2018) é um algoritmo de seleção que classifica as características por importância, descartando as menos importantes e reajustando o modelo. Um estimador (algoritmo de classificação) define a importância das características; o RFE repete o processo recursivamente até que metade das características permaneçam. O RFECV (KOUL; MANVI, 2019), por outro lado, usa o desempenho das características selecionadas no conjunto de dados de validação para definir sua importância com base em uma pontuação.

2.6 Classificação

Para verificar se as características extraídas possuem de fato eficiência para distinção das classes de áudio, são utilizados algoritmos de classificação, que são modelos matemáticos que utilizam uma abordagem supervisionada. O Random Forest (RF) e o SVM (Support Vector Machine) são exemplos de algoritmos que possuem uma ampla adoção na literatura.

O Random Forest (ESMAIL; AHMED; ELTAYEB, 2019), como o próprio nome sugere, gera um conjunto de árvores de decisão, independentes entre si, com predições aleatórias. Aquela que obtiver o maior número de votos será escolhida como modelo de classificação. O SVM (PATHIVADA; VEDAGIRI, 2022), usado para classificação e análise de regressão, nesse algoritmo, cada item de dados é representado como um ponto no espaço n -dimensional (onde n é o número de características), com o valor de cada característica sendo o valor de uma determinada coordenada. Então, é executada a classificação encontrando o hiperplano que melhor diferencia as duas classes.

2.7 Métricas de avaliação

As métricas de avaliação desempenham um papel fundamental na análise e aprimoramento de modelos de classificação, sendo essenciais para a compreensão de seu desempenho e eficácia. Elas proporcionam uma avaliação quantitativa e objetiva do quão eficaz um modelo é na tarefa de classificar dados em diferentes categorias.

A matriz de confusão é uma abordagem fundamental para avaliar o desempenho de modelos de classificação, pois permite a compreensão das relações entre os resultados previstos e os valores reais (LUQUE et al., 2019). Esta matriz organiza as previsões de um modelo em quatro categorias distintas: verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN). A partir das análises desses elementos, foram utilizadas as seguintes medidas estatísticas: Acurácia, Índice kappa e a Sensitividade.

A acurácia (BARATLOO et al., 2015) fornece uma medida geral de precisão do modelo. Calculada como a proporção de previsões corretas em relação ao total de predições (Equação 2.4) e quanto mais o valor se aproxima a 1, melhor é o seu o resultado. Essa métrica é amplamente utilizada para avaliar o desempenho global de um modelo de classificação.

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.4)$$

O índice Kappa (COHEN, 1960) avalia a concordância entre as previsões do modelo e as classes reais, ajustando para a concordância esperada ao acaso. Os valores do índice são classificados como: ruim ($K \leq 0,2$), razoável ($0,21 \leq K \leq 0,4$), bom ($0,41 \leq K \leq 0,6$), muito bom ($0,61 \leq K \leq 0,8$) e excelente ($K \geq 0,81$). Ao considerar a probabilidade de acerto, o kappa oferece uma medida mais robusta da concordância, sendo particularmente útil em situações de desequilíbrio nas classes. A Equação 2.5 apresenta a formula para calcular essa métrica, onde p_o é o resultado alcançado e p_e é o resultado esperado.

$$k = \frac{p_o - p_e}{1 - p_e} \quad (2.5)$$

A sensibilidade (OLSON; DELEN, 2008) avalia a capacidade do modelo identificar corretamente todas as instâncias positivas (VP) em relação ao total real de positivos, sendo representado pela Equação 2.6.

$$Sen = \frac{VP}{VP + FN} \quad (2.6)$$

3 Trabalhos Relacionados

Neste capítulo são apresentados os trabalhos que têm alguma relação com este. A Tabela 1 resume um comparativo dos trabalhos relacionados sobre quatro aspectos: modelo usado, tamanho do descritor, a utilização de base pública e métricas.

No trabalho de [Cejrowski et al. \(2018\)](#) foi coletado o som produzido por uma colônia de abelhas durante o período de fevereiro a agosto de 2017. Como não houve evidência de enxame ou ataque, os autores retiraram a rainha da colmeia e coletaram novas amostras, resultando em duas classes: com rainha e sem rainha. Os dados foram adquiridos em intervalos de 15 minutos, e um vetor de características foi extraído de cada amostra usando a técnica de codificação preditiva linear (LPC) ([O'SHAUGHNESSY, 1988](#)). Tais vetores foram submetidos ao algoritmo de incorporação de vizinho estocástico distribuído em t (t-SNE) ([MAATEN; HINTON, 2008](#)) para fins de redução de dimensionalidade. Em seguida, os vetores de características serviram como entrada para o classificador SVM. Os resultados alcançados apontam para uma taxa de erro igual a 2,17%. Apesar da redução da dimensionalidade, sua análise não identificou características irrelevantes ou qual grupo de características diferencia a presença/ausência da rainha.

No trabalho de [Nolasco et al. \(2019\)](#) os autores propuseram uma solução para detectar abelhas rainhas por meio de processamento de áudio com auxílio das redes neurais convolucionais (CNN) e técnicas clássicas de aprendizagem de máquina. Os áudios utilizados foram extraídos de base pública do projeto Nu-Hive ([CECCHI et al., 2018](#)). As amostras foram normalizadas através do Z-score e quebradas em intervalos de 60 segundos. A partir de cada amostra original, foram geradas 3 sintéticas através da variação aleatória do semitom dentro do intervalo -1 e 1. Os MFCCs, Mel-Espectrogramas e a transformada de Hilbert-Huang (HHT) foram extraídos das amostras normalizadas. A CNN foi adotada na etapa de extração de características e, posteriormente os descritores foram submetidos ao classificador SVM. O melhor resultado foi alcançado por uma combinação de MFCCs e HHT, atingindo uma AUC de 94%. Apesar de considerar características diferentes, os autores não aplicaram uma etapa de seleção de características para avaliar se a combinação de componentes de diferentes grupos poderia resultar em uma melhor representação.

No trabalho de [Robles-Guerrero et al. \(2019\)](#) foram monitoradas cinco colônias de abelhas (*Apis mellifera carnica*) e realizaram dois experimentos: (i) monitoramento de uma colônia com uma população reduzida e sem a rainha e quatro colmeias saudáveis, nesse primeiro caso, a rainha morreu naturalmente; (ii) monitoramento de três colônias saudáveis e duas foram selecionadas para remoção manual da rainha com o objetivo de simular a perda da rainha. Foram coletadas amostras com duração de 30 segundos em intervalos de 10 minutos durante 24 horas. A partir dos dados coletados foram extraídos os MFCCs e em seguida, calcularam a média, desvio padrão, variância, assimetria, mediana

e curtose dos coeficientes para cada amostra. A regressão logística Lasso (YU et al., 2015) foi utilizada na seleção de características, e concluíram que as características mais relevantes são a média e a mediana dos MFCCs. O algoritmo de decomposição de valor singular (SVD) (WALL; RECHTSTEINER; ROCHA, 2003) foi adotado para identificar a ausência de rainha, obtendo um AUC de 99%.

No trabalho de Kim, Oh e Heo (2021) os autores investigaram o desempenho de algoritmos de aprendizado de máquina (SVM, Random Forest e extreme gradient boosting (XGBoost)) e dois modelos CNN (*shallow* CNN e VGG-13) para distinguir sons como abelhas e não abelhas, eles usaram dados de som de colmeias do projeto Open Source Bee-hives (OSBH)¹. Os autores obtiveram uma acurácia de 91,93% usando a arquitetura VGG-13 combinada com MFCCs como dados de entrada, sendo a mais adequada para detectar sons não abelhas. No entanto, gerou um descritor de alta dimensão com tamanho de 32.768, e não foi aplicada nenhuma técnica de seleção de características ou redução de dimensionalidade.

Os trabalhos discutidos neste capítulo apresentam resultados promissores de classificação das técnicas de aprendizado de máquina e redes neurais para detecção da presença da abelha rainha. Além disso, apontam o MFCC como uma característica relevante para a diferenciação entre a presença e ausência da abelha rainha. Nenhum dos trabalhos apresentados realizou uma etapa de pré-processamento nos áudios, que fornece filtragem de ruído e detecção de intervalo irrelevante, o que impacta no desempenho e tempo de classificação. Embora a seleção de características permita a identificação de características mais coerentes e alcance descritores mais compactos, apenas Robles-Guerrero et al. (2019) selecionou características entre os trabalhos relacionados.

Tabela 1 – Comparação dos trabalhos relacionados.

Trabalho	Modelo	Tamanho do Descritor	Base Pública	Métricas
(CEJROWSKI et al., 2018)	LPC + t-SNE + SVM	16	Não	Error (2,17%)
(NOLASCO et al., 2019)	(MFCC+HHT) + SVM	40	Sim	AUC (94%)
(ROBLES-GUERRERO et al., 2019)	Lasso logistic regression + SVD	12	Não	AUC (99%)
(KIM; OH; HEO, 2021)	MFCC + VGG-13	32768	Sim	Acc (91%)

¹ <https://www.osbeehives.com/>

4 Materiais e métodos

Para o desenvolvimento do descritor proposto neste trabalho, foi utilizada uma metodologia composta por seis 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 uma base de dados pública. A segunda etapa envolve o processamento de amostras de áudio para garantir uniformidade em suas propriedades. As etapas três e quatro correspondem à extração e seleção de características, identificando aquelas que não contribuem de maneira relevante para a discriminação dos áudios. Por fim, nas etapas cinco e seis, as características selecionadas são submetidas à classificação, e diversas métricas são empregadas para avaliar o desempenho do modelo de classificação, respectivamente.

4.1 Aquisição dos áudios

Para a aquisição dos áudios foi adotado um conjunto de dados públicos sobre a presença da abelha rainha na colmeia. As gravações de áudio são do projeto Open Source Beehive ¹ e do conjunto de dados do projeto NU-Hive (CECCHI et al., 2018) e foram rotuladas por (NOLASCO; BENETOS, 2018). Duas colmeias foram monitoradas com e sem a rainha em um dia após a remoção da abelha rainha. Neste trabalho foi utilizada 508 amostras com tempo médio de 10 minutos².

4.2 Pré Processamento

Na etapa de pré-processamento as 508 amostras de áudio foram divididas em intervalos de 15 segundos, totalizando 19.597 amostras, os áudios insuficientes para compor novas amostras foram descartados. A Tabela 2 mostra a distribuição das amostras por ano e classe e a duração total das gravações de áudio. A biblioteca LibROSA (MCFEE et al., 2015) foi utilizada para realizar a manipulação dos áudios, esta biblioteca normaliza por padrão os dados no intervalo $[-1, 1]$, faz a transformação dos sinais em mono ao fazer a média das amostras entre os canais e garante que todas as amostras possuam uma mesma taxa de amostragem, de 22050Hz e uma quantização de 16 bits. Após a padronização dos dados, foram gerados Mel-espectrogramas em escala logarítmica para serem usados como entrada para as CNNs, a Tabela 3 apresenta os parâmetros dos Mel-espectrogramas com dimensão de 163×163 .

¹ <https://www.osbeehives.com/>

² <https://zenodo.org/record/1321278.YC6ueXVKhH4>



Figura 2 – Metodologia para detecção da presença da abelha rainha.

Tabela 2 – Distribuição por ano das amostras de áudio.

Classe	2017		2018	
	Número de Amostras	Tempo Total (s)	Número de Amostras	Tempo Total (s)
Rainha	5501	82515	5613	84195
Sem rainha	2867	43005	5616	84240
Total	8368	125520	11229	168435

Tabela 3 – Parâmetros para a geração dos Mel-espectrogramas.

Parâmetros	Valores
Taxa de amostragem	22050
Tamanho da janela	4096
Função de janela	Hann

4.3 Extração de características

4.3.1 Características manuais

Com base no estudo realizado por Luz et al. (2021), foram extraídas 13 características dos domínios de tempo, frequência e cepstral através das funções da LibROSA. As características selecionadas, conforme apresentado na Tabela 4, foram escolhidas por sua capacidade de representar informações relevantes sobre a qualidade do som. Ao final da extração, obteve-se um descritor com uma dimensão total de 58 posições, ou seja, um conjunto de diferentes características capaz de descrever um som. Essa abordagem possibilita que um classificador agrupe os sons em classes, sendo que, neste trabalho, as classes correspondem a "Rainha" e "Sem Rainha".

4.3.2 Características profundas

A extração de características profundas foi realizada utilizando a biblioteca Keras, uma interface de alto nível para redes neurais, que se beneficia do backend do TensorFlow (KETKAR; KETKAR, 2017). Foram adotados três modelos CNN pré-treinados com a base de dados ImageNet (RUSSAKOVSKY et al., 2015): ResNet50 (HE et al., 2016), VGG16 (ZHANG et al., 2015) e VGG19 (SIMONYAN; ZISSERMAN, 2014). Em todos os modelos, as características foram obtidas da última camada de pooling, enquanto as camadas totalmente conectadas foram desconsideradas. Posteriormente, aplicou-se o método de média global de pooling ao mapa de características resultante, reduzindo o descritor para a profundidade da última camada de pooling. A entrada foi padronizada para todas as arquiteturas, mantendo dimensões iguais de $163 \times 163 \times 3$. Contudo, os descritores resultantes possuem dimensões distintas de acordo com o modelo: ResNet50 (2048), VGG16 (512) e VGG19 (512).

4.4 Seleção de características

A seleção das características manuais que deveriam ser mantidas baseou-se no ranqueamento proporcionado pelo algoritmo RFECV, implementado por meio da biblioteca *scikit-learn* (PEDREGOSA et al., 2011). Essa etapa foi direcionada exclusivamente às características manuais, não abrangendo as características profundas extraídas anterior-

Tabela 4 – Descrição das características extraídas nos domínios cepstral, frequência e tempo. O número entre parênteses na coluna *Característica* indica a dimensão correspondente da característica.

Domínio	Característica	Descrição
Cepstral	MFCC (40)	Representação cepstral onde as bandas de frequência não são lineares, mas distribuídas de acordo com a escala mel. (TIWARI, 2010)
	Chroma CENS (1)	Calcula a variante cromática de energia normalizada. (MÜLLER, 2007)
	Chroma CQT (1)	Calcula a variante cromática da transformada Q constante. (DAS et al., 2020)
	Chroma STFT (1)	Calcula um cromagrama a partir de uma forma de onda ou um espectrograma de potência. (LEISTIKOW et al., 2004)
Frequência	Spectral Bandwidth (1)	Calcula a largura de banda espectral. (TZANETAKIS; COOK, 2002)
	Spectral Contrast (7)	Calcula o contraste espectral. (JIANG et al., 2002)
	Spectral Flatness (1)	Calcula o coeficiente de tonalidade. (DUBNOV, 2004)
	Poly (1)	Obtém coeficientes para ajustar um polinômio de ordem n às colunas de um espectrograma. (AGCAOGLU; SANTHANAM; HAYAT, 2013)
	Tonnetz (1)	Calcula as características do centróide tonal. (HARTE; SANDLER; GASSER, 2006)
	Spectral Centroid (1)	Representa o centro de gravidade do espectro. (LE et al., 2011)
Tempo	Spectral Rolloff (1)	Calcula a frequência <i>roll-off</i> . (STOLAR et al., 2018)
	Tempogram (1)	Calcula uma representação para um sinal dependente do tempo dado. (GROSCHKE; MÜLLER; KURTH, 2010)
	ZCR (1)	Representa a taxa de mudanças de sinal durante o tempo de duração de um determinado quadro (GOUYON et al., 2000)

mente, assegurando uma análise direcionada e otimizada das características manuais, sem comprometer a integridade ou complexidade das características profundas, que podem ser mais sensíveis à redução de dimensionalidade.

4.5 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, implementados com a biblioteca *scikit-learn*: RF e SVM. O RF foi aplicado devido aos seus resultados robustos em muitos problemas de reconhecimento de padrões, e o SVM foi selecionado, já que é o classificador mais utilizado em outros trabalhos relacionados à presença da abelha rainha.

Os parâmetros dos classificadores foram definidos por meio do otimizador de busca em grade (*Grid Search*) (BERGSTRA; BENGIO, 2012). A Tabela 5 detalha o intervalo

estabelecido para cada parâmetro e o melhor conjunto encontrado pelo otimizador. Dessa maneira, foram selecionados o kernel linear e o parâmetro de regularização (C) igual a 0,1 para o SVM e 90 árvores para o Random Forest para as características manuais. Já para os descritores profundos, foram selecionados 100 árvores para o RF, e o kernel linear e o parâmetro de regularização foram iguais a 0,1 para o SVM.

Tabela 5 – Configuração dos hiperparâmetros dos classificadores otimizada por meio do algoritmo *Grid Search*.

Classificadores	Intervalo dos parâmetros	Características Manuais	Características Profundas
RF	*estimators: [70, 80, 90, 100, 110]	90	100
SVM	C: [0.1, 0.5, 0.7, 1]	0.1	0.1
	kernel: [poly, rbf, sigmoid, linear]	linear	linear

* Número de estimators

Nos experimentos envolvendo as características manuais, inicialmente foi aplicada uma divisão de treino/teste na proporção de 80/20 para o subconjunto de 2018, resultando em 8983 amostras destinadas ao treinamento e 2246 amostras destinadas ao teste. Posteriormente, avaliou-se a generalização dos descritores para o subconjunto de 2017 através de uma validação cruzada, com $k = 5$.

No contexto das características profundas, realizou-se a divisão treino/teste utilizando com todas as amostras disponíveis, alocando 80% para treinamento e reservando 20% para teste, com a finalidade de explorar a totalidade do conjunto de dados disponível e proporcionar uma avaliação abrangente do desempenho dos modelos em relação às características profundas.

4.6 Validação

Conforme abordado no Capítulo 2, a avaliação dos modelos de classificação no presente estudo empregou três métricas de validação amplamente utilizadas na literatura: acurácia, índice kappa e sensibilidade, que foram calculadas com a biblioteca *scikit-learn*. A seleção dessas métricas se deve à capacidade demonstrada de apresentar de maneira clara a proporção de erros e acertos alcançados pelos classificadores.

5 Resultados

Considerando a metodologia detalhada no Capítulo 4, 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 58 posições, composto pelas características apresentadas na Tabela 4. Esse capítulo é dividido em quatro seções: na Seção 5.1 são apresentados os resultados obtidos com a seleção de características do descritor, enfatizando as características mais relevantes para a classificação; na Seção 5.2 é detalhado os resultados alcançados por cada classificador; na Seção 5.3 ocorre a discussão dos resultados, por fim, na Seção 5.4, os resultados da proposta são comparados com os trabalhos relacionados.

5.1 Seleção de características

A seleção de características foi realizada com base nos dados de 2018, uma vez que essa porção do conjunto de dados apresentou quase o dobro de amostras para a classificação "Sem Rainha" em comparação com o ano de 2017 (Tabela 2). A distribuição das 58 características, exibida na Figura 3, está ordenada por nível de relevância, abrangendo uma escala de 6 a 60, onde as mais relevantes ocupam as primeiras posições. Os 40 MFCCs, foram divididos em dois grupos de 20 coeficientes cada.

Das primeiras 36 características selecionadas, 31 pertenciam ao domínio cepstral (15 MFCCs (0–20) e 16 MFCCs (21–40)), quatro ao domínio de frequência (Poly, Spectral Flatness e 2 Spectral Contrast), e apenas uma pertencia ao domínio de tempo (ZCR). As quatro últimas características incluíram Tonnetz, Chroma CENS, Chroma CQT e um MFCC (0–20). Assim, concluí-se que os MFCCs representaram a característica mais relevante para detectar a presença/ausência da abelha rainha por meio do algoritmo de seleção RFECV.

Ao considerar o ranking gerado pelo RFECV, conforme ilustrado na Figura 3, foram criados quatro descritores para a etapa de classificação. O primeiro inclui as 58 características, abrangendo o conjunto completo de atributos extraídos dos áudios. O segundo descritor consiste nas 36 características mais relevantes, enquanto no terceiro foram escolhidos os 31 MFCCs, considerados os mais relevantes pelo algoritmo de seleção. Por fim, o quarto descritor é composto pelos 15 MFCCs que ocupam as primeiras posições no ranking. Ao empregar esses diferentes descritores, busca-se não apenas avaliar a influência das características individuais na precisão da classificação, mas também compreender como diferentes conjuntos podem impactar a performance geral do modelo.

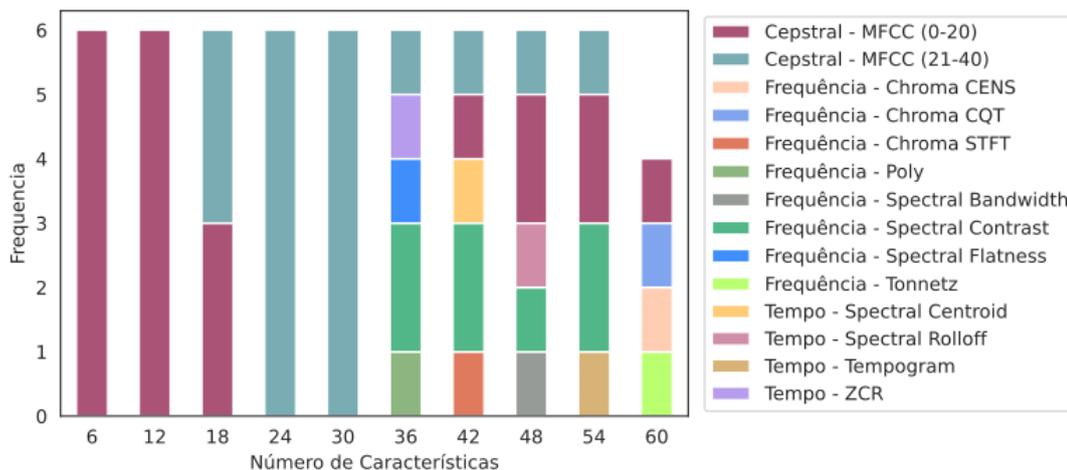


Figura 3 – Distribuição das 58 características extraídas das amostras de 2018.

5.2 Classificação

A Tabela 6 apresenta os resultados de classificação obtidos com os dados de 2018 por meio dos algoritmos Random Forest (RF) e Support Vector Machine (SVM) para os quatro descritores, os quais foram extraídos conforme detalhado na Seção 5.2. Todas as métricas alcançaram uma taxa de acerto de 0,99 para os descritores com 58 e 36 características em ambos os classificadores, evidenciando a robustez desses conjuntos de características. O RF manteve um desempenho de classificação de 0,99 para o descritor de 31 MFCCs. Já o descritor com 15 MFCCs apresentou uma redução no desempenho para o mesmo classificador, embora os resultados ainda sejam promissores. Em contraste, o desempenho do SVM foi inferior ao do RF para os descritores de 31 e 15 MFCCs, sendo a redução mais acentuada no conjunto com 15 MFCCs. Assim, ao utilizar exclusivamente os 31 MFCCs com o RF, observou-se uma precisão notável, ressaltando a importância das características do domínio cepstral. Mesmo diante de uma redução de 38% nas características, os resultados da classificação demonstram a relevância da seleção de características para este problema, especialmente em contextos embarcados.

Tabela 6 – Resultados de classificação dos descritores extraídos das amostras de 2018.

Classificadores	Métricas	Descritor (58)	Descritor (36)	Descritor (31)	Descritor (15)
RF	Acurácia	0,9973	0,9959	0,9951	0,9826
	Kappa	0,9946	0,9919	0,9902	0,9652
	Sensitividade	0,9964	0,9982	0,9955	0,9964
SVM	Acurácia	0,9959	0,9951	0,9835	0,9318
	Kappa	0,9919	0,9902	0,9670	0,8637
	Sensitividade	0,9973	0,9955	0,9991	0,9795

As características selecionadas a partir dos dados de 2018 foram extrapoladas para o subconjunto de 2017. Os resultados da classificação, utilizando a estratégia de validação cruzada com $k = 5$, são apresentados na Tabela 7. As métricas de desempenho revelam um valor médio de 0,99 e o maior desvio padrão de 0,006, considerando todas as caracte-

rísticas e o descritor com dimensão 36, para ambos os classificadores. Adicionalmente, o descritor de 31 MFCCs demonstrou um desempenho de 0,99 com o Random Forest (RF), corroborando os resultados da Tabela 6. Por outro lado, o SVM apresentou o desempenho menos satisfatório para o descritor de 15 MFCCs.

Tabela 7 – Resultados de classificação dos descritores extraídos das amostras de 2017 com a estratégia de validação cruzada de 5 conjuntos. O número entre parênteses indica o desvio padrão correspondente.

Classificadores	Métricas	Descritor (58)	Descritor (36)	Descritor (31)	Descritor (15)
RF	Acurácia	0,9984 (0,001)	0,9974 (0,002)	0,9972 (0,000)	0,9847 (0,002)
	Kappa	0,9973 (0,003)	0,9938 (0,006)	0,9925 (0,002)	0,9673 (0,012)
	Sensitividade	0,9998 (0,000)	0,9990 (0,001)	0,9985 (0,000)	0,9903 (0,001)
SVM	Acurácia	0,9991 (0,000)	0,9982 (0,002)	0,9674 (0,011)	0,9133 (0,007)
	Kappa	0,9981 (0,002)	0,9960 (0,005)	0,9279 (0,025)	0,8057 (0,014)
	Sensitividade	0,9987 (0,001)	0,9989 (0,000)	0,9996 (0,000)	0,9732 (0,003)

Os resultados de classificação para as características profundas são apresentados na Tabela 8. As arquiteturas VGG16 e VGG19 exibiram desempenhos semelhantes para ambos os classificadores. Já, a ResNet50 superou as demais arquiteturas em ambos os classificadores, destacando-se como a mais eficaz na detecção da presença de abelhas rainhas. Ao comparar esses resultados com os obtidos a partir das características manuais, destaca-se que o descritor de 31 MFCCs com o RF obteve desempenho superior aos modelos CNNs, incluindo a ResNet50 com um descritor de tamanho 2048. Conclui-se, portanto, que o descritor de 31 MFCCs proporciona uma caracterização precisa e de baixa dimensão da presença da abelha rainha, destacando-se como uma opção eficiente baseada em apenas uma característica.

Tabela 8 – Resultados de classificação das características profundas extraídas pelos modelos CNN. Os melhores resultados são destacados em negrito, e os números entre parênteses correspondem ao tamanho do descritor.

Classificadores	Métricas	ResNet50 (2048)	VGG16 (512)	VGG19 (512)
RF	Acurácia	0,9666	0,9594	0,9570
	Kappa	0,9568	0,9294	0,9091
	Sensitividade	0,9610	0,9277	0,9096
SVM	Acurácia	0,9877	0,9666	0,9702
	Kappa	0,9746	0,9297	0,9372
	Sensitividade	0,9760	0,9397	0,9427

5.3 Discussão

Os resultados de 2018 e 2017 mostram que a combinação de características cepstrais, temporais e de frequência ofereceu um descritor eficiente para identificar a presença da abelha rainha. Os valores da sensitividade foram próximos aos valores da acurácia, indicando que o modelo proposto obteve um bom desempenho de classificação para ambas as classes ("Rainha" e "Sem Rainha").

A etapa de seleção de características notavelmente reduziu a dimensão do descritor, preservando, ao mesmo tempo, o desempenho da classificação. Além disso, evidenciou que o MFCC foi a característica mais relevante para distinguir entre a presença e ausência de abelhas rainhas. O descritor contendo 31 MFCCs mostrou-se significativo para diminuir o processamento de dados e o consumo de energia em um sistema de monitoramento da saúde da colmeia. Na próxima seção, serão comparados os resultados obtidos com os encontrados na literatura.

5.4 Comparação com a literatura

A Tabela 9 apresenta os modelos propostos na literatura para a detecção da presença/ausência da abelha rainha, incluindo informações sobre a dimensionalidade do descritor e as métricas de desempenho associadas. O descritor desenvolvido neste trabalho demonstrou um notável desempenho, atingindo uma acurácia, kappa e sensibilidade de 99%. Em comparação com [Nolasco et al. \(2019\)](#), que utilizou o mesmo conjunto de dados, o descritor proposto superou substancialmente os resultados, uma vez que eles alcançaram uma AUC de aproximadamente 94% com o classificador SVM. Destaca-se que, mesmo ao considerar os resultados específicos do descritor composto por 15 MFCCs apresentados na Tabela 7, todos os descritores propostos neste estudo superaram as métricas de [Nolasco et al. \(2019\)](#).

Tabela 9 – Comparação do método proposto com os trabalhos relacionados.

Modelo	Tamanho do descritor	Métricas
LPC + t-SNE + SVM (CEJROWSKI et al., 2018)	16	Error (2,17%)
(MFCC+HHT) + SVM (NOLASCO et al., 2019)	40	AUC (94%)
Lasso logistic regression + SVD (ROBLES-GUERRERO et al., 2019)	12	AUC (99%)
MFCC + VGG-13 (KIM; OH; HEO, 2021)	32768	Acc (91%)
MFCC + RF (proposed)	31	Acc (99%) Kappa (99%) Sensitividade (99%)

Os resultados obtidos por [Robles-Guerrero et al. \(2019\)](#) destacaram a eficiência de um descritor mais compacto, com tamanho 12, e alcançaram uma pontuação AUC de 99%. Essa representação foi derivada de cálculos estatísticos aplicados aos MFCCs, sendo a média e a mediana consideradas as características mais relevantes. A escolha de um descritor de baixa dimensionalidade revela-se crucial para otimizar o processamento de dados e reduzir o consumo energético em um sistema de monitoramento. Em contrapartida, [Kim, Oh e Heo \(2021\)](#) atingiram o maior descritor ao utilizar uma abordagem baseada em CNN, alcançando um AUC de 91%. Notavelmente, constataram que os MFCCs apresentam uma distinção mais eficaz entre sons de abelhas e ruídos do que o Mel-espectrograma e a Transformada Q-constante (CQT). Essa descoberta é de significativa relevância para este estudo, pois os MFCCs também podem discernir a presença de invasores, fator que

compromete a saúde das colmeias. Dessa forma, o descritor proposto pode ser extrapolado para outros contextos de monitoramento acústico.

Em resumo, este trabalho se destaca das demais contribuições na literatura pelos seguintes aspectos:

1. A combinação de características cepstrais, temporais e de frequência, proporcionando uma representação mais abrangente dos sinais de áudio;
2. A análise das características mais eficazes para a detecção da abelha rainha, destacando aquelas que desempenham um papel crucial nesse processo e aprofundando-se em sua relevância para a classificação;
3. A implementação de uma avaliação de desempenho robusta, fundamentada em três métricas distintas, onde os resultados dos descritores foram confirmados com dados de diferentes anos de aquisição.

6 Conclusão

Nesse trabalho foi investigado o uso de características manuais para detectar a presença da abelha rainha em uma colmeia. O principal objetivo foi alcançado, que era a proposta de um descritor compacto, eficiente e de baixo custo computacional que supera a maioria das abordagens de ponta para classificação da presença da abelha rainha, incluindo modelos baseados em aprendizado profundo.

As principais contribuições deste trabalho são a avaliação e combinação de características cepstrais, temporais e de frequência para identificar o grupo de características que podem diferenciar entre a presença e ausência da abelha rainha. A etapa de seleção de características resultou em um descritor compacto com 31 MFCCs, superando outras abordagens, como as arquiteturas CNN. Assim, concluí-se que a aplicação de um método adequado de seleção de características reduziu consideravelmente o tamanho do descritor e manteve o desempenho de classificação. Os descritores de 15 e 31 MFCCs têm dimensões reduzidas e são baseadas em apenas uma característica, o que reduz o número de cálculos e dependências de bibliotecas, tornando-os adequados para aplicações de reconhecimento de som em dispositivos móveis ou sistemas embarcados.

Devido à falta de base públicas de sons de abelhas, esse trabalho foi limitado por não conseguir extrapolar o modelo para outros dados. A ausência desses conjuntos de dados abrangentes comprometeu a capacidade de generalização do descritor desenvolvido, uma vez que a diversidade e representatividade dos sons de abelhas são cruciais para garantir a eficácia do modelo em diferentes cenários. Essa limitação comprometeu a capacidade de assegurar a robustez e a aplicabilidade mais ampla do modelo, dificultando a adaptação do descritor para diferentes ambientes ou espécies de abelhas.

Com o intuito de amenizar esse problema e aprimorar os resultados de forma abrangente, pretende-se realizar os seguintes trabalhos futuros:

- Investigar conjuntos de dados adicionais de insetos para enriquecer a diversidade e representatividade dos dados.
- Explorar técnicas de aumento de dados, como aplicação de máscaras temporais e de frequência, inserção de ruído de fundo e mudança de tom, visando aprimorar a capacidade do modelo em lidar com variações nos sons das abelhas.
- Implementar os descritores propostos em dispositivos de baixo custo, com capacidade computacional limitada, e avaliá-los considerando o tempo de processamento, precisão e tamanho dos dados, visando aplicabilidades práticas em ambientes com recursos limitados.

7 Publicações

Os resultados apresentados neste trabalho foram compilados em formato de artigo, nomeado *MFCC-based descriptor for bee queen presence detection*, e publicado na revista [Expert Systems with Applications](#), Qualis(A1), fator de impacto(8,5).

Referências

- ABDOLLAHI, M.; GIOVENAZZO, P.; FALK, T. H. Automated beehive acoustics monitoring: A comprehensive review of the literature and recommendations for future work. *Applied Sciences*, MDPI, v. 12, n. 8, p. 3920, 2022. Citado na página 14.
- AGCAOGLU, O.; SANTHANAM, B.; HAYAT, M. Improved spectrograms using the discrete fractional fourier transform. In: IEEE. *2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*. [S.l.], 2013. p. 80–85. Citado na página 26.
- 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 19.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of machine learning research*, v. 13, n. 2, 2012. Citado na página 26.
- BILIK, S. et al. Machine learning and computer vision techniques in continuous beehive monitoring applications: A survey. *Computers and Electronics in Agriculture*, Elsevier, v. 217, p. 108560, 2024. Citado na página 14.
- BRAGA, A. R. et al. A method for mining combined data from in-hive sensors, weather and apiary inspections to forecast the health status of honey bee colonies. *Computers and Electronics in Agriculture*, Elsevier, v. 169, p. 105161, 2020. Citado na página 12.
- CECCHI, S. et al. A preliminary study of sounds emitted by honey bees in a beehive. In: AUDIO ENGINEERING SOCIETY. *Audio Engineering Society Convention 144*. [S.l.], 2018. Citado 2 vezes nas páginas 21 e 23.
- CEJROWSKI, T. et al. Detection of the bee queen presence using sound analysis. In: SPRINGER. *Asian Conference on Intelligent Information and Database Systems*. [S.l.], 2018. p. 297–306. Citado 3 vezes nas páginas 21, 22 e 31.
- CHEN, Q. et al. Decision variants for the automatic determination of optimal feature subset in rf-rfe. *Genes*, MDPI, v. 9, n. 6, p. 301, 2018. 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 na página 19.
- DAS, J. K. et al. Urban sound classification using convolutional neural network and long short term memory based on multiple features. In: IEEE. *2020 Fourth International Conference On Intelligent Computing in Data Sciences*. [S.l.], 2020. p. 1–9. Citado na página 26.
- 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 na página 26.

- 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 19.
- FAROOQUI, T. A potential link among biogenic amines-based pesticides, learning and memory, and colony collapse disorder: A unique hypothesis. *Neurochemistry international*, Elsevier, v. 62, n. 1, p. 122–136, 2013. Citado na página 11.
- GOUYON, F. et al. On the use of zero-crossing rate for an application of classification of percussive sounds. In: CITESEER. *Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00)*, Verona, Italy. [S.l.], 2000. v. 5, p. 16. Citado 2 vezes nas páginas 17 e 26.
- GROSCHÉ, P.; MÜLLER, M.; KURTH, F. Cyclic tempogram—a mid-level tempo representation for music signals. In: IEEE. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.], 2010. p. 5522–5525. Citado na página 26.
- 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 na página 26.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 25.
- JIANG, D. et al. Music type classification by spectral contrast feature. In: IEEE. *Proceedings. IEEE International Conference on Multimedia and Expo*. [S.l.], 2002. p. 113–116. Citado na página 26.
- KETKAR, N.; KETKAR, N. Introduction to keras. *Deep learning with python: a hands-on introduction*, Springer, p. 97–111, 2017. Citado na página 25.
- KIM, J.; OH, J.; HEO, T.-Y. Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and grad-cam. *Mathematical Problems in Engineering*, Hindawi, v. 2021, 2021. Citado 3 vezes nas páginas 11, 22 e 31.
- KOUL, N.; MANVI, S. S. A scheme for feature selection from gene expression data using recursive feature elimination with cross validation and unsupervised deep belief network classifier. In: IEEE. *2019 3rd International Conference on Computing and Communications Technologies (ICCCCT)*. [S.l.], 2019. p. 31–36. Citado na página 18.
- KULYUKIN, V.; MUKHERJEE, S.; AMLATHE, P. Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples. *Applied Sciences*, MDPI, v. 8, n. 9, p. 1573, 2018. Citado na página 12.
- LE, P. N. et al. Investigation of spectral centroid features for cognitive load classification. *Speech Communication*, Elsevier, v. 53, n. 4, p. 540–551, 2011. Citado na página 26.
- LEISTIKOW, R. J. et al. Bayesian identification of closely-spaced chords from single-frame stft peaks. In: *Proc. 7th Int. Conf. Digital Audio Effects (DAFx-04)*. [S.l.: s.n.], 2004. p. 228–233. Citado 2 vezes nas páginas 17 e 26.

- LUQUE, A. et al. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, Elsevier, v. 91, p. 216–231, 2019. Citado na página 19.
- LUZ, J. S. et al. Ensemble of handcrafted and deep features for urban sound classification. *Applied Acoustics*, v. 175, p. 107819, 2021. Citado na página 25.
- MAATEN, L. Van der; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. 11, 2008. Citado na página 21.
- MCFFEE, 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 23.
- MEIKLE, W.; HOLST, N. Application of continuous monitoring of honeybee colonies. *Apidologie*, Springer, v. 46, p. 10–22, 2015. Citado na página 14.
- MESAROS, A. et al. Sound event detection: A tutorial. *IEEE Signal Processing Magazine*, IEEE, v. 38, n. 5, p. 67–83, 2021. Citado na página 15.
- MÜLLER, M. *Information Retrieval for Music and Motion*. 1. ed. [S.l.]: Springer Berlin Heidelberg, 2007. 318 p. ISBN 9783540740483. Citado na página 26.
- NOLASCO, I.; BENETOS, E. To bee or not to bee: Investigating machine learning approaches for beehive sound recognition. *arXiv preprint arXiv:1811.06016*, 2018. Citado na página 23.
- NOLASCO, I. et al. Audio-based identification of beehive states. In: IEEE. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.], 2019. p. 8256–8260. Citado 4 vezes nas páginas 11, 21, 22 e 31.
- NOVAIS, S. M. et al. Effects of a possible pollinator crisis on food crop production in brazil. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 11, n. 11, p. e0167292, 2016. Citado na página 12.
- OLSON, D. L.; DELEN, D. *Advanced data mining techniques*. [S.l.]: Springer Science & Business Media, 2008. Citado na página 20.
- O'SHAUGHNESSY, D. Linear predictive coding, one popular technique of analyzing certain physical signals. *IEEE Potentials*, 1988. Citado na página 21.
- PATHIVADA, B. K.; VEDAGIRI, P. Investigating dilemma zone boundaries for mixed traffic conditions using support vector machines. *Transportation letters*, Taylor & Francis, v. 14, n. 4, p. 378–384, 2022. Citado na página 19.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, JMLR. org, v. 12, p. 2825–2830, 2011. Citado na página 25.
- PEETERS, G.; RICHARD, G. Deep learning for audio and music. *Multi-faceted Deep Learning: Models and Data*, Springer, p. 231–266, 2021. Citado na página 18.
- PURWINS, H. et al. Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, IEEE, v. 13, n. 2, p. 206–219, 2019. Citado na página 18.

- ROBLES-GUERRERO, A. et al. Analysis of a multiclass classification problem by lasso logistic regression and singular value decomposition to identify sound patterns in queenless bee colonies. *Computers and Electronics in Agriculture*, Elsevier, v. 159, p. 69–74, 2019. Citado 3 vezes nas páginas [21](#), [22](#) e [31](#).
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, Springer, v. 115, p. 211–252, 2015. Citado na página [25](#).
- 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 na página [16](#).
- SHARMA, G.; UMAPATHY, K.; KRISHNAN, S. Trends in audio signal feature extraction methods. *Applied Acoustics*, Elsevier, v. 158, p. 107020, 2020. Citado 2 vezes nas páginas [16](#) e [17](#).
- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014. Citado na página [25](#).
- STOLAR, M. N. et al. Detection of adolescent depression from speech using optimised spectral roll-off parameters. *Biomedical Journal*, v. 2, p. 10, 2018. Citado na página [26](#).
- TERENZI, A. et al. Features extraction applied to the analysis of the sounds emitted by honey bees in a beehive. In: *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*. [S.l.: s.n.], 2019. Citado na página [12](#).
- TIWARI, V. Mfcc and its applications in speaker recognition. *International journal on emerging technologies*, Citeseer, v. 1, n. 1, p. 19–22, 2010. Citado 2 vezes nas páginas [17](#) e [26](#).
- TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, IEEE, v. 10, n. 5, p. 293–302, 2002. Citado na página [26](#).
- VIRTANEN, T.; PLUMBLEY, M. D.; ELLIS, D. *Computational analysis of sound scenes and events*. [S.l.]: Springer, 2018. Citado 3 vezes nas páginas [14](#), [15](#) e [16](#).
- WALL, M. E.; RECHTSTEINER, A.; ROCHA, L. M. Singular value decomposition and principal component analysis. In: *A practical approach to microarray data analysis*. [S.l.]: Springer, 2003. p. 91–109. Citado na página [22](#).
- WANG, Y. et al. Deep learning in audio classification. In: SPRINGER. *International Conference on Information and Software Technologies*. [S.l.], 2022. p. 64–77. Citado na página [18](#).
- YU, D. et al. Classification of spectral data using fused lasso logistic regression. *Chemometrics and Intelligent Laboratory Systems*, Elsevier, v. 142, p. 70–77, 2015. Citado na página [22](#).
- ZGANK, A. Bee swarm activity acoustic classification for an iot-based farm service. *Sensors*, MDPI AG, v. 20, n. 1, p. 21, 2020. Citado na página [11](#).

ZHANG, X. et al. Accelerating very deep convolutional networks for classification and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 38, n. 10, p. 1943–1955, 2015. Citado na página 25.

ZÖLZER, U. *Digital audio signal processing*. [S.l.]: John Wiley & Sons, 2022. Citado na página 15.



**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA
“JOSÉ ALBANO DE MACEDO”**

Identificação do Tipo de Documento

- () Tese
() Dissertação
(X) Monografia
() Artigo

Eu, **Bianca Sousa Soares**, 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 **Deteção da presença da abelha rainha através da classificação e processamento de áudio** 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 02 de fevereiro de 2024.

Bianca Sousa Soares

Assinatura