Journal of Internet Services and Applications, 2024, 15:1, doi: 10.5753/jisa.2024.4015

Cepstral and Deep Features for Apis mellifera Hive Strength Classification

Jederson S. Luz [**Federal University of Piauí** | ***jedersonluz@ufpi.edu.br***]
Myllena C. de Oliveira [**Federal University of Piauí** | ***myllenaoliveira@ufpi.edu.br***]
Fábia de M. Pereira [**Embrapa Mid-North** | ***fabia.pereira@embrapa.br***]
Flávio H. D. de Araújo [**Federal University of Piauí** | ***flavio86@ufpi.edu.br***]
Deborah M. V. Magalhães [**Federal University of Piauí** | ***deborah.vm@ufpi.edu.br***]

Postgraduate Program in Electrical Engineering, Federal University of Piauí, Teresina, PI, Brazil.

Received: 12 January 2024 • **Accepted:** 17 September 2024 • **Published:** 06 December 2024

Abstract Regular management practices are crucial to assessing colonies' conditions and implementing measures to improve their strength. However, constant revisions can induce stress and even contribute to swarm loss. Therefore, effective management that considers the well-being of the bees is necessary. In order to assist the beekeeper in managing the hives, this study proposes a noninvasive approach integrating *Apis mellifera* L., 1758 (Hymenoptera: Apidae) colony sound processing with machine learning and deep learning techniques to identify colony strength, essential for the productivity of apiculture. We developed an audio acquisition process focused on colony strength, resulting in a dataset with 3702 samples. We explored features extracted by CNNs, including VGG16, ResNet50, MobileNet, and YOLO, comparing them with cepstral features such as Mel-Frequency cepstral coefficients (MFCCs). Cepstral features significantly outperformed those extracted by CNN, with MFCCs achieving an accuracy of 95.53%, compared to the 78.99% achieved by the best-performing CNN. These results highlight the effectiveness of MFCCs in accurately identifying hive strength. This work differs from literature because it presents a protocol for categorizing beehives as either weak or strong, with a focus on reducing intervention time. It also includes a public dataset containing MFCCs and Deep Features extracted from audio recorded at different apiaries. Additionally, it offers a method for automatically classifying hives based on their strength. These contributions aim to serve as a knowledge base for the scientific community and to support beekeepers in non-invasive and cost-effective apiary management.

Keywords: Precision beekeeping, Machine learning, Feature extraction, Audio processing.

1 Introduction

Bees play an indispensable role in agriculture by providing essential products such as honey, wax, pollen, and propolis, contributing approximately U$12 billion to the Brazilian economy through pollination services [Vieira *et al.*, 2021]. Beekeeping productivity depends on the strength of hives due to their high correlation with honey production. Colonies with larger populations generally perform better than those with smaller populations [Kumar and Mall, 2018]. Hives with decreased honey bee populations are more susceptible to attacks from natural enemies [Mazepa and Laurenti, 2022].

Regular management practices are crucial to assess colony conditions and implement necessary measures to improve their strength [Gorroi *et al.*, 2020]. However, constant revisions can induce stress, and during periods of food scarcity, they may contribute to swarm loss. On the other hand, the absence of management and weak colonies' existence in apiaries negatively impact activity [Oliveira Costa *et al.*, 2016]. Therefore, beekeepers consistently have to decide between carrying out management practices that stress the colonies and needing to assess hive conditions to implement necessary measures.

Management involves opening each beehive and observing the condition of the combs, presence of the queen, laying pattern, presence of predators, and the amount of stored food [Gorroi *et al.*, 2020]. For large-scale beekeepers, periodic inspections may be impractical due to time constraints. Performing these reviews safely and without causing stress to the colonies is crucial for developing apiculture activities.

Bees use sound to communicate within the hives. Experienced beekeepers can perceive characteristic sounds in queenless hives or colonies preparing to swarm [Phan *et al.*, 2023]. Research has shown that the sound patterns produced by bees are efficient indicators for monitoring the conditions and needs of hives [Rustam *et al.*, 2024]. As a result, studies have investigated noninvasive methods for apiary monitoring based on hive sounds for various applications such as counting the bee's entrance in the hive [Heise *et al.*, 2020], detection of queen bees [Ruvinga *et al.*, 2021; Barbisan *et al.*, 2024], swarming [Zgank, 2021], circadian rhythm [Kim *et al.*, 2021], presence of air pollutants [Sharif *et al.*, 2020], foraging period [Shostak and Prodeus, 2019], and strength of colony [Zhang *et al.*, 2021].

These studies have shown promising results by demonstrating that the use of audio processing techniques combined with artificial intelligence can assist in identifying colony demands. This is achieved through the development of models that can automatically recognize acoustic patterns. This approach has the potential to reduce the necessity for daily physical inspections, lower costs, and improve the overall efficiency of beekeeping management.

Based on this, the present research aims to automatically detect the strength of hives based on acoustic patterns. The main hypothesis is that cepstral and deep features extracted from sounds produced by colonies can discriminate hives based on their strength. This hypothesis is grounded in the literature, suggesting a relationship between bee sounds and colony activity level, health status, and behavior [Rustam *et al.*, 2024].

This work provides to the academic and beekeeper community: (1) a protocol to inspect the beehives and categorize them as either weak or strong focus on reducing the intervention time and, consequently, the colony stress;(2) the availability of a public dataset with MFCCs and Deep Features extracted from audio collected from different classes of hives, resulting in an extension of de Oliveira *et al.* [2023] with 18 new audio samples with approximately 30 minutes each; and (3) a methodology to automatically classify hives according their strength that adopts cepstral and deep features extracted from sounds produced by colonies to characterize strength patterns. This methodology can be directly extrapolated for different scenarios such as queen absence detection, *Varroa* destructor detection, and detection of pollutants, among others. These contributions serve as a knowledge base for the scientific community and support beekeepers in non-invasive and cost-effective apiary management.

This work provides to the academic and beekeeper community the following: (1) A protocol for inspecting beehives and classifying them as weak or strong, with a focus on reducing intervention time and colony stress; (2) A public dataset containing MFCCs and Deep Features extracted from audio collected from different classes of hives, including 18 new audio samples, each approximately 30 minutes long, extending the work of de Oliveira *et al.* [2023]; and (3) A methodology for automatically classifying hives based on their strength, which utilizes cepstral and deep features extracted from sounds produced by colonies to characterize strength patterns. This methodology can be extrapolated to various scenarios, including queen absence detection, *Varroa* destructor detection, and identification of pollutants, among others. These contributions serve as a knowledge base for the scientific community and support beekeepers in non-invasive and cost-effective apiary management.

This work is organized as follows. Section 2 presents a state-of-the-art survey of works related to the proposed theme. Section 3 details the methodology proposed in this work, outlining all process stages. Section 4 presents the results obtained with the methodology of this work and a comparison with state-of-the-art results, followed by a brief discussion. Finally, Section 5 provides the final considerations of the work, addressing the main topics.

2 Related Work

Noninvasive methods for monitoring bees have been a subject of investigation in various studies, with audio analysis and processing applied in various applications: counting the entry and exit of bees in the hive [Heise *et al.*, 2020], detection of queen bees [Ruvinga *et al.*, 2021; Barbisan *et al.*, 2024], swarming [Zgank, 2021], circadian rhythm [Kim *et al.*, 2021], presence of air pollutants Sharif *et al.* [2020], estimation of the peak foraging activity period of bees [Shostak and Prodeus, 2019], and strength of colony [Zhang *et al.*, 2021]. Additionally, Abdollahi *et al.* [2022] identified approximately 60 studies investigating audio processing for hive monitoring.

The Mel Frequency Cepstral Coefficients (MFCCs) are highly relevant for sound classification in apiary monitoring. For instance, Ruvinga *et al.* [2021] achieved a 92% accuracy for bee queen detection using 13 MFCCs and log energy. The authors used an LSTM and tested the MLP model for classification, which reached 90% accuracy. The work of Soares *et al.* [2022] used the combination of cepstral, time, and frequency characteristics to classify the presence or absence of the queen in the hive, achieving their best result with a feature vector of 58 dimensions. The author used the SVM classifier, obtaining an accuracy of 99%. Similarly, Kulyukin [2021] utilized MFCCs as features to classify between bees, crickets, or noise. The author employed 13 MFCCs and achieved an accuracy of 98.43% with Random Forest (RF) and 98% with Support Vector Machine (SVM) classifiers. Although both studies show promising results, there is still room to improve the generalization capacity of the models, using audio captured from different hives and at different periods of the day, bringing it closer to the real apiary scenario. In addition, it would be beneficial to explore other types of features in the extraction of audio, complementing the use of MFCCs, to ensure greater reliability in the results obtained.

Barbisan *et al.* [2024] detected the presence of the queen bee as well. The study utilized neural networks and support vector machine (SVM) models to classify audio signals captured inside the hive. Features were extracted using Mel-Frequency Cepstral Coefficients (MFCC) and Short-Time Fourier Transform (STFT). The models achieved F1-score exceeding 98%. However, convolutional neural networks (CNNs) were not tested, and the approach relies on sensor deployment, which may pose implementation and maintenance costs for small-scale beekeepers.

Cejrowski *et al.* [2020] proposed a monitoring approach to detect the circadian rhythm of bees, determining their nighttime interval through the energy level of the signal emitted by the hive. They concluded that the nighttime period for *Buckfast Apis mellifera* is between 23:00 and 04:00. To achieve this, the authors extracted the Mel Frequency Cepstral Coefficients (MFCCs) and used them as input for the Support Vector Machine (SVM) classifier, ultimately obtaining an accuracy of 81.14%. Despite the promising results, the data was collected in just two hives and only at one time of the year. Evaluating the proposed approach in productive apiaries and with a greater number of hives and different weather conditions is valuable.

In the study by Sharif *et al.* [2020], three different sets of characteristics were analyzed for detecting pollutants in beehives, specifically focusing on the organic compound Trichloromethane (CHCl3). The authors used Mel Frequency Cepstral Coefficients (MFCCs) and sound landscape indices as the feature sets. The features extracted were input into the Random Forest classifier, resulting in an accuracy of 91.66% with sound landscape indices and 80% with MFCC. It's worth noting that this evaluation is limited to binary classification: distinguishing between blank air and Trichloromethane. Additionally, the study proposes a differential method for determining the optimal sample size for classification based on seasonal characteristics.

Bromenshenk *et al*. [2009] developed acoustic recording equipment to observe that the sound emitted by bees can not only detect pollutants but also identify the type of pollutant present. However, the equipment comes at a higher acquisition cost compared to other, more affordable alternatives, which may increase the beekeeper's production costs.

Shostak and Prodeus [2019] conducted a study to determine the honey harvesting period. The research involved analyzing spectral density from audio recordings to assess whether the beehive was prepared for honey collection. Classification was performed using a dividing curve equation to separate the different categories in spatial space. Model validation was carried out based on the probability of correct classification, resulting in a 96% accuracy rate. Additionally, the authors suggested combining the power spectrum density estimate at 200 Hz or 250 Hz with a Bayesian decision rule. However, this method produced lower classification accuracy compared to the initial approach, and the authors did not evaluate its ability to differentiate between bee colonies.

Zhang *et al*. [2021] modeled beehives strength using a semi-supervised deep learning approach. They proposed a hardware system to collect audio and environmental data from beehives and used a hierarchical generative-prediction network to model hive strength based on audio. The model achieved 78.1% accuracy, improving performance when trained on labeled and unlabeled data. Despite these advances, the complexity of the sensors and the need for continuous human annotations further restrict the model's applicability on a larger scale or in smaller beekeeping operations.

Among the works discussed, MFCC extraction combined with machine learning techniques like RF or SVM has shown promising results for various apiary monitoring purposes. Only Kulyukin [2021] used a publicly available database, the lack of publicly accessible acoustic databases presents a challenge as it hampers the advancement of research in this area. The works that utilize private data are limited to collecting audio from a few hives in restricted periods and testing the models in limited climatic variation conditions.

Our study discusses automatic audio classification in a scenario that few have explored in the literature, only Zhang *et al*. [2021] approached the strength of the hive detection. Additionally, we aim to address one of the gaps in the current state of the art by making the database of extracted MFCCs publicly available, thereby contributing to the accessibility and replicability of the research. This study used a smartphone microphone to record audio, which reduced costs compared to studies using external sensors. Furthermore, we adopted a real acquisition scenario with production hives rather than manipulating the condition of the hive, considering different apiaries under varying climatic conditions.

3 Materials and Methods

We aim to develop a solution for the processing and classification of bioacoustic signals that characterize the hive's state as strong or weak, assisting beekeepers in understanding the condition of their hives. Depending on environmental conditions, strong hives may be ready for honey collection, while weak hives may require management to strengthen. Figure 1 illustrates the steps to identify the hive.

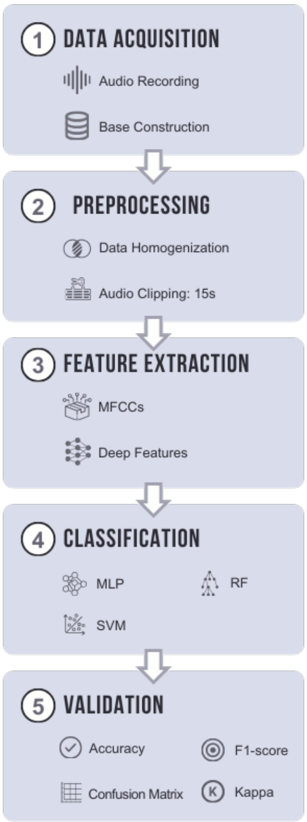

Figure 1. Methodology steps: (1) Data acquisition: Collection of raw data from apiaries; (2) Preprocessing: standardization and normalization for consistent and comparable features; (3) Feature extraction: identification and extraction of pertinent features from preprocessed data; (4) Classification: models evaluation to identify the most suitable for the dataset, and finally, (5) Validation: different metrics to assess the model's generalizability.

3.1 Data Acquisition

Motivated by the lack of publicly available databases related to hive sounds of *Apis mellifera* L., 1758 (Hymenoptera: Apidae), we constructed our database focusing on hive strength. Additionally, inspired by Al-Tikrity *et al*. [1971], DeGrandi-Hoffman *et al*. [2008] and Vallenas-Sánchez *et al*. [2023], we developed a new methodology to characterize the colonies as strong and weak, focusing on reducing the intervention time and, consequently, the stress caused in the classification colony. We consider the following aspects to classify colonies: population density, availability of honey and pollen, broods quantity, queen posture quality, and presence of natural enemies.

It is important to emphasize that the available studies on the analysis of the sound produced by bees were carried out by manipulating the hives for the characteristics studied, such as the presence and absence of queens [Rustam *et al.*, 2024], colony density [Di *et al.*, 2023] and information on air pollutants [Yu *et al.*, 2023]. In this study, the sound was collected in beekeepers' hives used for honey production. There was no manipulation for the desired characteristics, but rather monitoring of the natural development of the beehives used for production.

Initially, we managed the hives, installed in two apiaries in the rural area of Teresina, Piauí, Brazil (5°05′21″ S; 42°48′06″ W), observing and noting the number of frames not used by bees and the number of frames occupied by adult bees, worker bee brood, honey and pollen. We also observed the queen's posture and the appearance of the offspring, looking for symptoms of diseases and infestations by natural enemies. These characteristics affect the hive population, the number of offspring, and the availability of food and, together, more efficiently reflect the classification of the hive into strong and weak.

Subsequently, the data was analyzed to calculate the occupancy rate of the hives based on the available space (Occupancy rate = (number of occupied frames/number of unoccupied frames) x 100). The hives were classified as strong and weak according to this rate. Hives with an occupancy between 75% and 100% were classified as strong. Hives with occupancy below 74% were classified as weak. Symptoms of natural enemy attacks and queen posture (regular or irregular) contributed to the decision-making process in cases where hive occupancy was above the established rate limit. All hives classified as strong had queens with assessed egg-laying capacities rated as good or very good. Both strong and weak hives were situated in the same apiaries and subjected to identical environmental conditions and availability of bee plants. All hives adopted in the experiments had queens and no presence of natural enemies.

The beekeeper used a smartphone to collect audio recordings via the Voice Recorder app[1] with the following configuration: 11,000 Hz sampling rate, single channel (mono), and the recording format set to Windows Wave (wav).

The development of honey bee colonies presents distinct seasonal states related to the availability of food collected in flowers. The recordings were conducted on different dates, considering the seasonality of the hives: (i) period of food shortage (November and December); (ii) beginning of flowering (February); and (iii) period of high food availability (April and July). In November 2022, three collections were made on the 22nd and 28th. Between December 1 and 6 of the same year, three additional collections were performed. Finally, four new collections were conducted on February 1, 2023.

Audio recordings were made between 6:30 AM and 8:30 AM on sunny, windless days. This standardization was performed because the number of bees inside the hive depends on the environmental conditions and the time of day. In the early hours of sunny days, there are more bees collecting food outside the hive. To prevent the analysis of the collected audios from being influenced by the number of bees inside the hives due to their food-gathering behavior, we collected the sound at the same times and under the same environmental conditions. The audios were collected in two different apiaries, 11 recordings were made in Apiary A and 2 in Apiary B. In total, 10 audios were recorded with an average duration of 30 minutes and a standard deviation of 11.46. The audios from the first collection have been made available and published in a previous study [de Oliveira *et al.*, 2023]. In the second stage of audio capturing, the audio recordings were captured between 1 April and 7, 2023, totaling 8, and between 5 July and 21, 2023, totaling 10. All samples from the second collection belong only to Apiary A, located in Teresina. The subsequent collection followed the same pattern as the first, including 18 new recordings, each approximately 30 minutes in duration, with a standard deviation of 4.8. Consequently, the final dataset was composed of 28 audio recordings. The distribution of the number of samples per class after the collection is detailed in Table 1.

Class	Samples (before cut)	Samples (after cut)	Total time (s)
Weak	11	1880	30255
Strong	17	1822	27840

Table 1. Distribution of dataset samples based on colony strength, considering the number of recordings before and after the original audio cutting in 15-second slices and the total recording time in seconds. This results in an unbalanced dataset, with the longest duration of samples coming from weaker colonies.

Before recording the audio, the hives received puffs of smoke at the entrance of the hive to prevent highly defensive behavior and, with the cover slightly open, over the frames. After the smoke had taken effect, the smartphone was placed between the cover and the frames, with half of the device containing the microphone inside the hive and the other half outside, allowing activation of the recording button, as illustrated in Figure 2. Once recording began, the smartphone remained in the hive for approximately 30 minutes. The beekeepers distanced themselves from the collection site to avoid interference with captured sound.

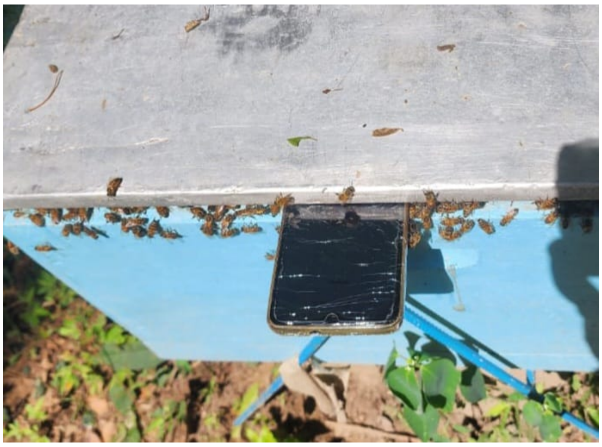

Figure 2. Hive audio acquisition in a real-world apiary setting. Audio is recorded by placing a smartphone with its microphone facing inside the hive, positioned between the lid and the box. The smartphone is wrapped in plastic film to prevent bees from covering the device with propolis.

3.2 Pre-processing

The preprocessing step was performed to achieve uniformity in the acquired data, ensuring that all recordings had the same

[1] https://bit.ly/voice-recorder-app

settings in terms of duration, sampling rate, quantization, and number of channels. Additionally, as the captured audio recordings had an average duration of 30 minutes, they were cut to generate new audio clips with a duration of 15 seconds, testing both with and without overlap. The decision to use 15-second clips was based on the goal of finding the shortest audio segment that would still provide satisfactory classification performance. Considering a scenario with a large number of hives, it is relevant to minimize the audio recording to reduce the management time. We experimented several clip lengths, including up to 1 minute, but found that 15 seconds was efficient, since longer clips did not improve the results. These shortened clips were used as individual samples in the subsequent steps. The number of samples after the cut can be observed in Table 1.

The manipulation of the audio recordings captured in the previous stage was performed using the *LibROSA* library [McFee *et al.*, 2015]. By default, this library normalizes the data to the range [-1,1] and converts the signals to mono. Furthermore, it ensures that all samples have the same sampling rate of 11,000 Hz, a parameterized value, and a quantization of 16 bits [McFee *et al.*, 2015].

3.3 Feature Extraction

In this phase, the goal was to extract features that allow the classifier to better distinguish between the classes to which the sound belongs. We evaluated the performance of cepstral features and features extracted through Convolutional Neural Networks (CNNs) to differentiate the sound classes associated with colony strength.

3.3.1 Mel Frequency Cepstral Coefficients (MFCCs)

Cepstral features are related to how the human auditory system perceives sounds, especially speech. The most common are the Mel frequency spectral coefficients (MFCCs) [Virtanen *et al.*, 2018]. MFCCs concisely describe the overall shape of a spectral envelope, representing the boundaries within which the signal's spectrum is contained.

The *LibROSA* library was employed to extract 40 MFCCs. The work of Soares *et al.* [2022] inspired this choice, which utilized MFCCs, among other features, to classify sound scenarios involving bees, specifically the absence or presence of the queen in the hive. In that study, the MFCCs were ranked among the top 40 most relevant features. The extracted MFCCs are publicly available on [2].

3.3.2 Deep features

The extraction of features through Convolutional Neural Networks (CNNs) involves a detailed process to translate audio into visual representations. Initially, we used the *LibROSA* library to generate Mel spectrograms from the audio, producing images scaled to 224x224 pixels, as demonstrated in Figures 4 and 5. This resolution was selected to balance the trade-off between computational efficiency and the preservation of relevant spectral details necessary for accurate pattern recognition. To this end, it is necessary to follow the steps as presented in Figure 3, where the audio samples were split into smaller windows (2048) to ensure a fine-grained analysis of the time-frequency components. The window size of 2048 was chosen as it is a standard in audio processing, providing a good resolution of frequency components while maintaining manageable data sizes [Oppenheim and Schafer, 2009]. After signal partitioning, the discrete Fourier transforms (DFT) over the overlapped windows were computed using the Blackman-Harris windowing function. This specific window function was selected due to its superior ability to minimize spectral leakage, which is crucial for obtaining accurate frequency representations [Blackman and Harris, 1967]. Besides, we adopted an overlap of 50% of the windows (hop length) to guarantee statistical dependence between the windows, enhancing the model's capacity to capture temporal correlations in the audio signal.

These spectrograms served as input for different CNN architectures, each specialized in extracting specific patterns. We utilized models such as VGG16 [Simonyan and Zisserman, 2014], ResNet50 [He *et al.*, 2016], MobileNet [Howard *et al.*, 2017], and YOLO [Zhang *et al.*, 2022] to process the spectrograms, resulting in feature vectors of distinct sizes: 512 for VGG16, 2048 for ResNet50, 1024 for MobileNet, and 1728 for YOLO. These architectures were selected based on their established performance in visual pattern recognition tasks. Each model presents a distinct balance between model complexity, parameter count, and the capacity to generalize across diverse datasets. All models were utilized in both pretrained and fine-tuned versions.

We have chosen VGG16 and ResNet50 based on their established effectiveness in feature extraction, as demonstrated in the original papers by Simonyan and Zisserman [2014] and He *et al.* [2016], respectively. VGG16 is renowned for its deep architecture and outstanding performance in image classification, while ResNet50's residual connections effectively tackle the degradation problem in deeper networks. As for MobileNet and YOLO, we have opted for them due to their impressive computational efficiency and real-time performance, making them particularly well-suited for mobile and embedded applications. MobileNet, as outlined by Howard *et al.* [2017], is purposefully crafted to be lightweight and efficient, while YOLO, particularly the improved YOLOv5 version discussed by Zhang *et al.* [2022], excels in swift object detection.

We used transfer learning principles to improve audio feature extraction. Specifically, we fine-tuned preexisting Convolutional Neural Networks (CNNs) such as VGG16, ResNet50, MobileNet, and YOLO. We used the Mel spectrograms mentioned in the previous paragraph as input to the CNNs. We utilized pre-trained weights from ImageNet to VGG16, ResNet50 and MobilNet, and ultralytics to YOLO to serve as the foundation for our fine-tuning methodology. The work of Yosinski *et al.* [2014] inspired our method.

In fine-tuning, we freeze all base model layers and introduce a fully connected layer. This hybrid architecture underwent an initial ten epochs training phase with a conservative learning rate of 0.0001. The choice of ten epochs for initial training is consistent with common practices in transfer learning, where a limited number of epochs is sufficient to adjust the new layers without overfitting, especially when the

[2] https://bit.ly/mfcc-and-deep-features-dataset

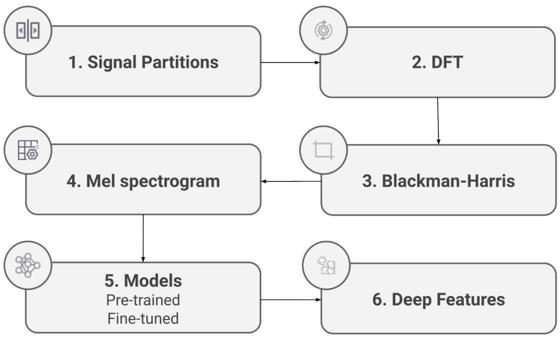

Figure 3. Steps to generate Deep Features: (1) Signal Partition: Dividing the input signal into smaller windows. (2) DFT: Transforming the signal from the time domain to the frequency domain. (3) Mel Spectrogram: Converting DFT into Mel-frequency representations for human-like perception. (4) Blackman-Harris: Applying window function for enhanced resolution and noise reduction. (5) Models Pre-training and Fine-tuning: Using pre-existing models, then adjusting them to specific tasks. (6) Deep Features: High-level abstract representations extracted from pre-trained and fine-tuned models.

dataset is relatively small [Yosinski *et al.*, 2014]. The learning rate of 0.0001 was selected to ensure stable convergence during this phase, as it is generally considered a safe starting point that prevents drastic updates to the weights, which could destabilize the training [Smith, 2017].

Building on this foundation, we strategically unfroze the final layers of the base model and continued training for an additional ten epochs, employing a reduced learning rate of 0.00001. This second phase of fine-tuning, where specific layers are unfrozen, typically requires a smaller learning rate to allow for fine-grained adjustments without overshooting the optimal weights, especially when the model is close to convergence [Yosinski *et al.*, 2014]. The reduction in the learning rate to 0.00001 aligns with best practices in fine-tuning, which advocate for progressively smaller learning rates as training progresses to refine the model's performance [He *et al.*, 2016].

The resulting fine-tuned CNN models were used to extract features from Mel spectrograms, generating feature vectors of varying sizes (512 for VGG16, 2048 for ResNet50, 1024 for MobileNet, and 768 for YOLO). Their derived feature vectors serve as inputs for machine learning models.

3.4 Classification

Three distinct classifiers were adopted: Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF). These classifiers were selected to provide insights into the performance of different types of models. MLP is a neural network capable of learning complex and non-linear relationships, making it suitable for capturing patterns in audio features [Haykin, 2001]. We conducted manual parameter tuning on the Multilayer Perceptron (MLP), focusing on the number of epochs, batch size, and the number of hidden layers. SVMs are particularly effective for handling non-linear data through kernel functions, which map the data into higher-dimensional spaces, enabling the construction of non-linear decision boundaries [Cortes and Vapnik, 1995]. In the context of SVM, we conducted tuning on the kernel type and the regularization parameter C to optimize the performance of the decision boundary. Based on decision trees, RF was chosen for its robustness and reduced risk of overfitting, especially in cases with noisy or imbalanced data [Breiman, 2001]. In RF, we optimized the number of trees (n_estimators) and the splitting criterion (e.g., Gini impurity, entropy) to enhance the model's accuracy and stability. Table 2 shows the parameters range of classifiers, the best values of parameters are different for each experimental scenario. The selected values are presented in section 4. Results and Discussion.

To define the training and test sets and avoid data leakage, we first split the 28 audio recordings before any preprocessing steps were performed, that is, before segmenting the audios into smaller 15-second samples. We aimed for a split close to 80/20 proportion, taking into account the duration

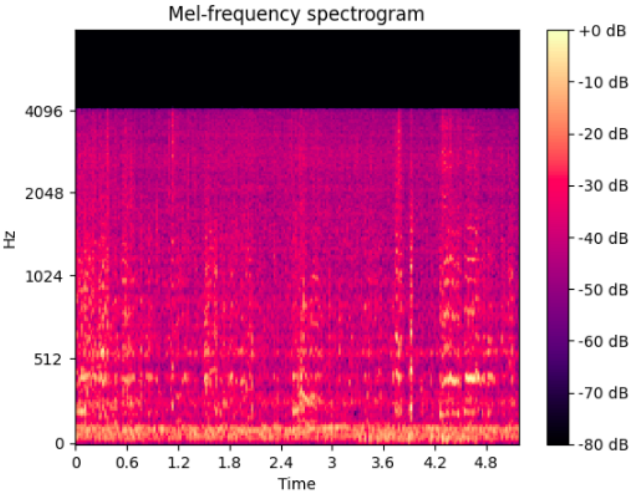

Figure 4. The strong hive produces a distinct sound pattern with a prominent peak occurring at around 0.6 seconds. This pattern repeats consistently over the entire duration, with segments characterized by 0 dB power. Over a duration of 5.4 seconds, the sound decreases in the higher frequency range (above 1024 Hz), while its main power lies within the lower frequency range (0 to around 512 Hz).

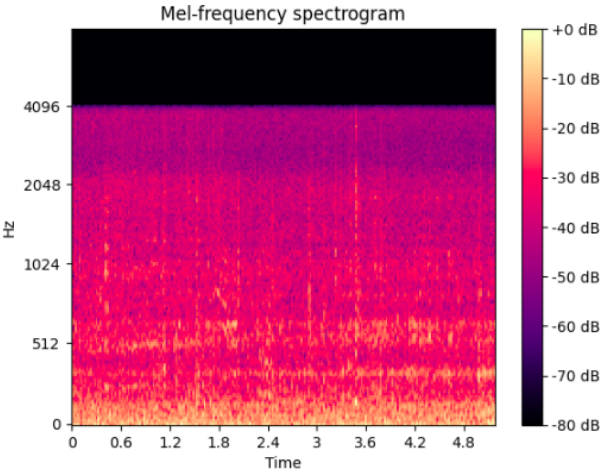

Figure 5. Melspectogram of a weaker hive, that exhibits a consistent sound pattern over time. Within a 5.4-second time window, the sound is significantly attenuated in the higher frequency range (above 2048 Hz), while most of its power is concentrated in the lower frequency range (0 to approximately 512 Hz).

Classifier	Parameters	Values
MLP	No. of epochs	20, 25, 30, 35
	Batch-size	1, 4, 8
SVM	Kernel	poly, rbf, sigmoid, linear
	C	1 - 11
RF	Criterion	gini, entropy
	N-estimators	10,20,30,40,50,60

Table 2. Range of values adopted for classifier parametrization.

of each recording to maintain balance between the sets

For the test set, we selected 2 recordings from weak hives and 2 from strong hives, ensuring that the duration of these recordings was similar. The training set included the remaining 24 recordings, with 15 from weak hives and 9 from strong hives. Although there was a difference in the number of recordings per class, the total duration of audio for each class in the training set was kept as close as possible.

This approach ensured that no part of the training set was present in the test set, and no part of the test set was present in the training set. The preprocessing step, including the segmentation into 15-second clips, was conducted separately for each set, further preventing any potential data leakage. As a result, the test set consisted of 303 samples from strong hives and 235 samples from weak hives, while the training set contained 1525 samples from strong hives and 1662 samples from weak hives.

3.5 Validation

To evaluate the results of audio classification, statistical measures widely used in the literature were adopted: Accuracy (Acc), Kappa index (κ), F1 score, and confusion matrix, which presents the proportion of errors and correct predictions obtained by the classifiers [Wardhani *et al.*, 2019]. The F1 score balances precision and recall metrics, where the ideal value is 1. The Kappa index was selected due to the class imbalance, as shown in Table 1.

4 Results and Discussion

We conducted a comparative analysis of the performance of Mel Frequency Cepstral Coefficients (MFCCs) and Convolutional Neural Networks (CNNs) in feature extraction. We evaluated both extraction methods using samples with and without overlap. Moreover, we evaluated the performance of pre-trained networks and those that underwent fine-tuning for CNN-based feature extraction.

Tables 3 and 4 summarize the results of experiments conducted with MFCCs, using non-overlapping and overlapping data, respectively. Upon analyzing the results, we observed that overlapping data did not enhance the classifiers' performance for MFCCs. The best accuracy was achieved with non-overlapping data, with 1.65% accuracy higher than the best result obtained with overlapping data. Notably, the kappa and F1-score metrics showed significant improvement without overlapping data. Considering the MFCCs, the overlapping could raise noisy information, making audio events classification a challenge [Leng *et al.*, 2015].

Based on the outcomes presented in Table 3, the MLP exhibited the highest accuracy among other classifiers. This was expected as the MLP can create complex models and adjust to various data types. Additionally, Table 4 indicates that Random Forest (RF) was the most effective classifier when dealing with overlapping. The tree voting system of RF can operate more reliably even in the presence of noise. [Breiman, 2001].

Model	Parameters	Accuracy	f1-score	kappa
MLP (1 layer: 6)	**epc:30, bs:4**	**0.9553**	**0.9550**	**0.9082**
SVM	k:rbf, C:10	0.9182	0.9184	0.8348
RF	est:20, crit:entropy	0.8996	0.9000	0.799

Table 3. Colony strength classification results are presented in terms of accuracy, f-score, and kappa. MFCCs are used as input features without data overlapping, and different classifiers with manual parameter optimization are employed. The best results are highlighted in bold.

Model	Parameters	Accuracy	f1-score	kappa
MLP (1 layer: 6)	epc:30, bs:4	0.7573	0.6045	0.2106
SVM	k:rbf, C:10	0.6652	0.6620	0.3354
RF	**est:20, crit:entropy**	**0.7574**	**0.7566**	**0.5173**

Table 4. Colony strength classification results are presented in terms of accuracy, f-score, and kappa. MFCCs are used as input features with data overlapping, and different classifiers with manual parameter optimization are employed. The best results are highlighted in bold.

In our study, we assess the impact of fine-tuning pre-trained CNNs for colony strength classification. The results showed similar performance, suggesting that pre-trained models without fine-tuning may be relevant for this specific scenario. The fine-tuning offered fast overfitting, even with an extremely low learning rate. Fine-tuning can lead to instability and overfitting, particularly for small datasets [Dong *et al.*, 2021].

Extraction model	Classification model	Parameters	Accuracy	F1-score	kappa
VGG16	MLP (1 layer: 6)	epc:30, bs:4	0.7147	0.7098	0.4348
VGG16	SVM	k:linear, d:3, C:1	0.7707	0.7621	0.5476
VGG16	RF	est:50, crit:entropy	0.6349	0.6349	0.2703
ResNet50	**MLP (1 layer: 6)**	**epc:30, bs:4**	**0.7860**	**0.7798**	**0.5771**
ResNet50	SVM	k:linear, d:3, C:1	0.7809	0.7747	0.5670
ResNet50	RF	est:50, crit:entropy	0.7436	0.7434	0.4870
MobileNet	MLP (1 layer: 6)	epc:30, bs:4	0.6621	0.6614	0.3268
MobileNet	SVM	k:linear, d:3, C:1	0.6587	0.6587	0.3186
MobileNet	RF	est:50, crit:entropy	0.7215	0.7214	0.4431
YOLO	MLP (1 layer: 6)	epc:30, bs:4	0.6994	0.6994	0.3998
YOLO	SVM	k:linear, d:3, C:1	0.6332	0.6327	0.2691
YOLO	RF	est:50, crit:entropy	0.6842	0.6841	0.3683

Table 5. Colony strength classification results are presented in terms of accuracy, f-score, and kappa. Features extracted from no overlapping data through pre-trained deep networks are used as input of different classifiers with manual parameter optimization. The best results are highlighted in bold.

In the context of this study, MFCCs demonstrated superior performance over the deep features extracted from CNNs due to several key factors. Pre-trained CNN models are inherently designed with generic feature representations, optimized for a wide spectrum of image classification tasks. However, the original training datasets of these models do not align closely with the unique characteristics of our domain, which involves analyzing audio signals to assess colony strength. This significant domain shift implies that the deep features extracted by these CNNs may lack the

Extraction model	Classification model	Parameters	Accuracy	F1-score	kappa
VGG16	MLP (1 layer: 6)	epc:30, bs:4	0.7053	0.7049	0.4128
VGG16	SVM	k:linear, d:3, C:1	0.7677	0.7591	0.5417
VGG16	RF	est:50, crit:entropy	0.6037	0.6034	0.2099
ResNet50	**MLP (1 layer: 6)**	**epc:30, bs:4**	**0.7967**	**0.7917**	**0.5980**
ResNet50	SVM	k:linear, d:3, C:1	0.7890	0.7839	0.5828
ResNet50	RF	est:50, crit:entropy	0,7284	0,7280	0.4561
MobileNet	MLP (1 layer: 6)	epc:30, bs:4	0.6515	0.6514	0.3031
MobileNet	SVM	k:linear, d:3, C:1	0.6823	0.6822	0.3647
MobileNet	RF	est:50, crit:entropy	0.7062	0.7061	0.4123
YOLO	MLP (1 layer: 6)	epc:30, bs:4	0.6020	0.5971	0.2107
YOLO	SVM	k:linear, d:3, C:1	0.6285	0.6282	0.2591
YOLO	RF	est:50, crit:entropy	0.6464	0.6464	0.2931

Table 6. Colony strength classification results are presented in terms of accuracy, f-score, and kappa. Features extracted from overlapping data through pre-trained deep networks are used as input of different classifiers with manual parameter optimization. The best results are highlighted in bold.

Extraction model	Classification model	Parameters	Accuracy	F1-score	kappa
VGG16	MLP (1 layer: 6)	epc:30, bs:4	0.7181	0.7124	0.449
VGG16	**SVM**	**k:linear, d:3, C:1**	**0.7843**	**0.7787**	**0.5735**
VGG16	RF	est:50, crit:entropy	0.6519	0.6511	0.307
ResNet50	MLP (1 layer: 6)	epc:30, bs:4	0.7403	0.7301	0.4874
ResNet50	SVM	k:linear, d:3, C:1	0.7060	0.7516	0.5275
ResNet50	RF	est:50, crit:entropy	0.7623	0.7622	0.5247
MobileNet	MLP (1 layer: 6)	epc:30, bs:4	0.7164	0.7164	0.4335
MobileNet	SVM	k:linear, d:3, C:1	0.672	0.6670	0.3362
MobileNet	RF	est:50, crit:entropy	0.7249	0.7248	0.4512
YOLO	MLP (1 layer: 6)	epc:30, bs:4	0.6757	0.6756	0.3528
YOLO	SVM	k:linear, d:3, C:1	0.6366	0.6356	0.2765
YOLO	RF	est:50, crit:entropy	0.6570	0.6567	0.3160

Table 7. Colony strength classification results are presented in terms of accuracy, f-score, and kappa. Features extracted from no overlapping data through fine-tuning deep networks are used as input of different classifiers with manual parameter optimization. The best results are highlighted in bold.

Extraction model	Classification model	Parameters	Accuracy	F1-score	kappa
VGG16	MLP (1 layer: 6)	epc:30, bs:4	0.5978	0.5830	0.2063
VGG16	SVM	k:linear, d:3, C:1	0.7412	0.7297	0.4900
VGG16	RF	est:50, crit:entropy	0.6310	0.6291	0.2665
ResNet50	MLP (1 layer: 6)	epc:30, bs:4	0.7011	0.6896	0.4104
ResNet50	**SVM**	**k:linear, d:3, C:1**	**0.7899**	**0.7847**	**0.5845**
ResNet50	RF	est:50, crit:entropy	0.7591	0.7589	0.5179
MobileNet	MLP (1 layer: 6)	epc:30, bs:4	0.7344	0.7319	0.4660
MobileNet	SVM	k:linear, d:3, C:1	0.6797	0.6797	0.3603
MobileNet	RF	est:50, crit:entropy	0.7002	0.6997	0.4028
YOLO	MLP (1 layer: 6)	epc:30, bs:4	0.6746	0.6745	0.3507
YOLO	SVM	k:linear, d:3, C:1	0.6703	0.6702	0.3422
YOLO	RF	est:50, crit:entropy	0.6831	0.6831	0.3668

Table 8. Colony strength classification results are presented in terms of accuracy, f-score, and kappa. Features extracted from overlapping data through fine-tuning deep networks are used as input of different classifiers with manual parameter optimization. The best results are highlighted in bold.

specificity required to capture the nuances essential for this classification task.

Additionally, the fine-tuning process, typically employed to adapt pre-trained models to new tasks, presented challenges due to the limited size of our dataset. Fine-tuning under such conditions requires a delicate balance to avoid overfitting, where the model learns the noise of the training data rather than generalizable patterns. Despite attempts at fine-tuning, the CNN models tended to overfit rapidly, contributing to their inferior performance compared to the more stable and robust MFCCs.

Even lighter CNN architectures, such as MobileNet and YOLO, which are less prone to overfitting due to their smaller number of parameters, showed inferior performance compared to MFCCs. The t-SNE plots in the Figures 7, 8, 9, and 10 reveal that CNNs, even with mel-spectrograms as inputs, exhibit significant class overlap, indicating less distinct feature representations. In contrast, MFCCs, as illustrated in Figure 6, provide greater class separability, which is crucial for distinguishing varying levels of colony strength in our context.

The performance gap underscores the limitations of using pre-trained CNNs and highlights the significance of traditional acoustic features like MFCCs, which effectively capture the tonal characteristics of audio signals. The challenges encountered with CNNs underscore the necessity for deep models specifically trained with domain-specific data, such as beehive sounds, to achieve optimal performance. Future research should consider expanding the dataset to explore how it might influence the comparative effectiveness of CNNs versus traditional features in audio classification tasks.

In addition, based on our experiment presented in Tables 3, 4, 5, 6, 7, and 8, it was found that introducing a 50% overlap between audio samples did not have a significant impact on the classification outcomes. Thus, it can be inferred that the temporal overlap of samples does not substantially affect the performance of the classification models used in our study.

It is important to note that all results obtained with features extracted through CNNs were inferior to those obtained with MFCCs. Additionally, the descriptors produced by the evaluated CNNs have higher dimensionality. This comparison suggests that, for this specific context, CNN-derived features may not be as effective as MFCCs. This may be justified by Figures 7, 8, 9, and 10, which demonstrate the complexity of data dispersion when using these features as descriptors.

Our analysis provides insights into the applicability of different feature extraction methods in the classification of colony strength and the effectiveness of overlapping data. These findings not only contribute to the specific understanding of colony strength classification but also provide broader insights into the applicability of deep learning techniques in complex acoustic contexts. They enrich the ongoing discourse and lay a solid foundation for the development of future research and advancements in the field.

4.1 Comparison with the State of the Art

Table 9 presents the different applications for acoustic hive monitoring such as queen bee presence, bee detection, bees' circadian rhythm, hive pollutants, honey harvest period, and colony strength. The latter is addressed in this work. Table 9 summarizes the related work regarding the features extracted, the size of the descriptor, the classifiers, and the respective results in terms of accuracy, a metric common to all works.

Table 9 shows the relevance of MCCFs, which are the most commonly used feature and contribute to high classification accuracies in various applications. Each study also has a variety of descriptor sizes, ranging from 2 to 193 features. Although there is no strict correlation between the descriptor size and accuracy, we emphasize their importance because they could impact the performance of the embedded application on devices with limited computational capabilities. SVM was the most commonly employed classifier, followed by neural networks (NN), which yielded promising results.

Our proposal achieved an accuracy of 95.53% in detecting the strength of the hive using a descriptor of 40 MFCCs and

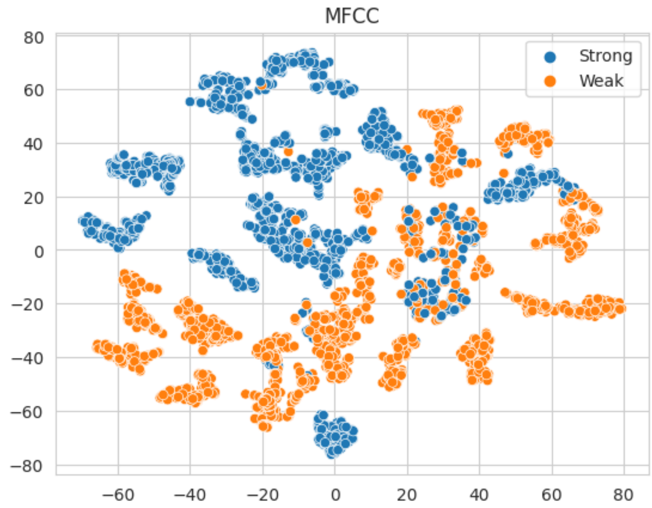

Figure 6. TSNE scatterplot illustrates the separation between the weak and strong classes based on the MFCC features. It is important to note the limited areas of overlap between classes, which further validates the promising classification results.

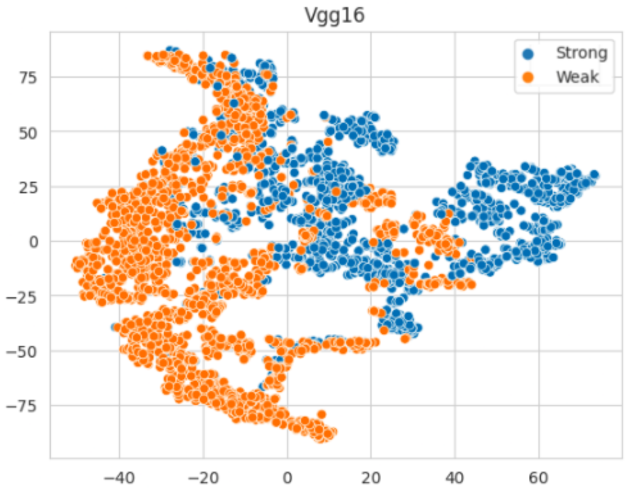

Figure 7. TSNE scatterplot demonstrates the separation between the weak and strong classes based on the features extracted by VGG16. Although there are some overlapping areas, there is a considerable distinction between the classes, which reinforces the high classification performance of this descriptor compared to descriptors from lighter architectures such as YOLO and MobileNet.

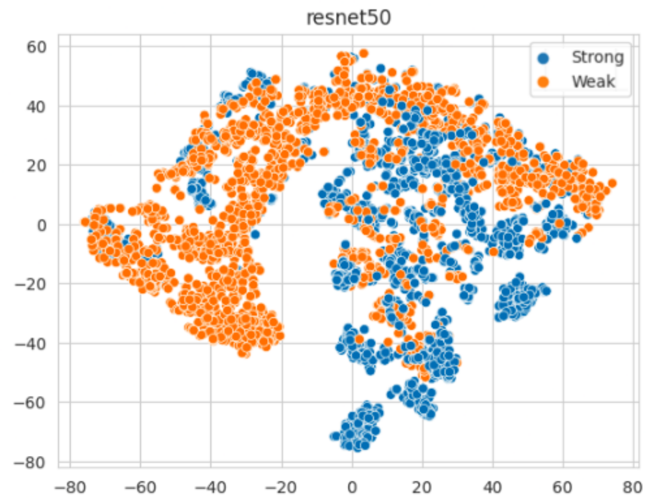

Figure 8. TSNE scatterplot demonstrates the separation between the weak and strong classes based on the features extracted by Resnet50. Although there are overlapping areas, there is a considerable distinction between the classes, which reinforces the high classification performance of this descriptor compared to descriptors from lighter architectures such as YOLO and MobileNet.

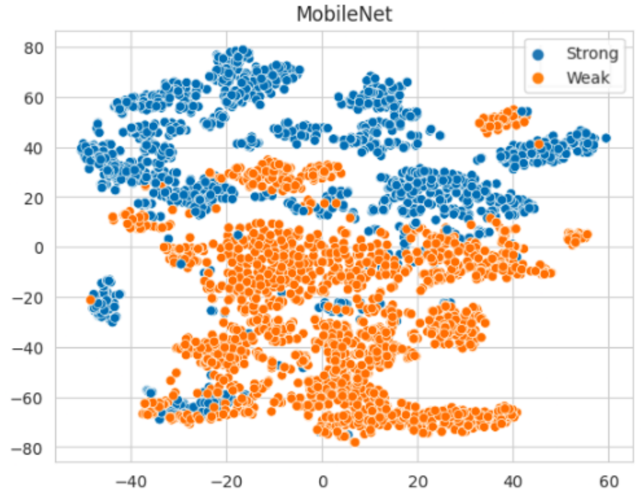

Figure 9. TSNE scatterplot demonstrates the separation between the weak and strong classes based on the features extracted by MobileNet. There is a large region of overlap between classes, reinforcing the poor classification performance presented by this descriptor compared to the descriptors from dense architectures such as Resnet50 and VGG16.

an MLP with only 1 layer and 6 neurons. The reduced size of the descriptor, based on a single feature and combined with a model of little depth, suggests a viable alternative for real-time use in an apiary on a device with limited computational power.

The variety of objectives, methodologies, and datasets makes direct comparisons challenging. It's worth noting that Zhang et al. (2021) conducted closely related work, focusing on assessing colony strength. Their classification accuracy was the lowest, but it's important to note that no assumptions can be made about the audio complexity, as the authors used a private dataset. These diversity highlights the complexity of using sound for hive state classification, suggesting a promising area for future investigations.

5 Conclusion

This research involved the collection of audio from beehives, exploring the potential of MFCCs and Mel spectrograms to describe colony strength. The main objective was to provide relevant information to beekeepers, assisting them, for example, in selecting hives for honey extraction. During development, we identified that to provide more practical guidance to beekeepers, classify the colony strength as strong or weak, simplify the decision-making process, and intervene when necessary. Additionally, audio capture in the apiaries allowed the construction of a labeled and public database, filling a gap in the literature.

The method proposed in this research achieved high accuracy using a descriptor with 40 MFCCs, overperforming CNN-based descriptors. Deeper models (VGG and ResNet) captured more useful features than the light models (MobileNet and YOLO), achieving better classification accuracy. These findings contribute to the specific understanding of colony strength classification and provide broader insights into the applicability of deep learning techniques in complex acoustic contexts. They enrich the ongoing discourse for future research and advancements in the field. The result suggests that a compact descriptor effectively identifies colony strength, offering a practical advantage: descriptors based on a single feature reduce the number of necessary calculations and extraction time. Moreover, they are more suitable for implementation on devices with limited computational power, often found in beehive monitoring environments in the apiaries. This consideration suggests that the proposed method is feasible for practical use in real-world conditions.

The nature of the field of beekeeping is complex. Environmental conditions and various factors can lead to overlaps in hive characteristics. Addressing this challenge is valuable, as it mirrors the real inherent complexity of hive monitoring. Additionally, it is crucial to emphasize the difficulty of collecting new samples, as it requires appropriate conditions in the hives and underscores the commitment to obtaining high-quality data.

For future research, we will investigate how noise filtering impacts preprocessing to improve classification performance and extrapolate the descriptor to other scenarios of interest in the beekeeping chain, such as identifying the presence or absence of the queen in the hive, detecting invaders, monitoring hive temperature, among other applications. Additionally, we evaluate the computational cost of different classifiers, including Markov chains, to embed the classification model and seek a representation of the colony strength with intuitive numerical values for beekeepers.

Another future research direction could be to investigate how MFCCs can be integrated as input to CNNs and evaluate the performance of these models in comparison with currently employed methods, bearing in mind that in the current work, CNNs were used only for feature extraction and not for classification. This approach would not only broaden the scope of input feature analysis but also offer valuable insights into the effectiveness of MFCCs in deep learning contexts for the task at hand.

We also aim to expand the applicability of the methodology developed in this study by increasing the number of samples, covering a variety of scenarios in beekeeping. A natural extension would be to explore the system's ability to identify the presence or absence of the queen in hives. Additionally, we consider integrating information on hive temperatures, a critical variable for bee health and productivity, investigating how the methodology adapts to apiaries in different regions and whether it will provide valuable insights for beekeepers and researchers. We also intend to explore the optimization of neural network architecture and model parameters to enhance the system's accuracy and efficiency further. Finally, we will develop robust models capable of extrapolating to various scenarios within hives, providing a versatile and valuable tool for monitoring and effectively managing bee colonies.

This study has consolidated the application of machine learning predictive models as a valuable tool for improving observability in complex IT systems. The microservices-based architecture proved to be the right selection, with significant benefits in terms of scalability and maintenance. The GradientBoostingRegressor and RandomForestRegressor models proved to be particularly efficient, with the former achieving an R^2 Score of 0.86 when predicting HTTP request rates and the latter reducing the Mean Squared Error (MSE) by 2.06% for memory usage predictions when compared to traditional monitoring methods.

These advances highlight the models' ability to identify crucial patterns and anticipate anomalies with considerable accuracy, enabling more agile and informed interventions. However, challenges such as the need for fine-tuning models and improving training performance still persist. The complexity and computational cost of machine learning models demand special attention, indicating the need for ongoing research into optimization and efficiency.

Future work will explore strategies that can speed up the training process without compromising the accuracy of the models. This could include the application of more efficient algorithms, the use of specialized hardware, and data dimensionality reduction techniques. In addition, emphasis will be placed on implementing auto-tuning mechanisms that can simplify the selection of hyperparameters, making predictive models not only more agile but also accessible for wider adoption in IT production environments. Furthermore, modifying the application to be able to run more than one application on different servers is also mapped out future work.

Reference	Objective	Features	Descriptor Size	Classifiers	Results
Ruvinga *et al*. [2023]	Queen presence	MFCCs	14	LSTM	91.81% acc
Kulyukin [2021]	Bee or no bee	MFCCs	193	RF	98.43%
Cejrowski *et al*. [2020]	Circadian rhythm	MFCCs	13	SVM	81.14% acc
Sharif *et al*. [2020]	Pollutants detection	Landscape indices	4	RF	91.66% acc
Zhao *et al*. [2021]	Pollutants detection	MFCCs	39	SVM	93.7% acc
Shostak and Prodeus [2019]	Harvest period	Spectral density	2	Divisive curve equation	96% acc
Zhang *et al*. [2021]	Colony strength	Mel spectrograms, temperature, humidity, and pressure	96	GPN	78.1% acc
Barbisan *et al*. [2024]	Queen presence	MFCCs, STFT	1-50	NN, SVM	99.21% acc (STFT, NN)
Proposed Method	**Colony strength**	**MFCCs**	**40**	**MLP**	**95.53% acc**

Table 9. In comparing the use of audio processing in beehives across different literature, we emphasize the importance of the size of the descriptor and classifier. This is crucial as it affects the performance of the embedded application on devices with limited computational capabilities. Our proposal, highlighted in bold, achieves significant accuracy (acc) with a relatively small descriptor. However, it is important to note that the classification complexity can vary greatly depending on the specific application.

These future guidelines aim to strengthen the proposition that integrating machine learning into observability is a technical enhancement that can take IT systems management to a new level of proactivity and resilience.

Declarations

Funding

This research was partially funded by FAPEPI via Centelha 2 Piauí Program.

Authors' Contributions

Jederson Sousa Luz was the main contributor and writer of this manuscript. Fábia de Mello Pereira, Deborah Maria Vieira Magalhães, and Myllena Caetano de Oliveira contributed to the conception of the study and performed the review. Jederson Sousa Luz and Myllena Caetano de Oliveira conducted the experiments. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets generated and analysed during the current study are available in: `https://bit.ly/mfcc-and-deep-features-dataset`.

References

Abdollahi, M., Giovenazzo, P., and Falk, T. H. (2022). Automated beehive acoustics monitoring: a comprehensive review of the literature and recommendations for future work. *Applied Sciences*, 12(8):3920. DOI: 10.3390/app12083920.

Al-Tikrity, W., Hillmann, R., Benton, A., and Clarke, WW, J. (1971). A new instrument for brood measurement in a honeybee colony. Available at:`https://www.cabidigitallibrary.org/doi/full/10.5555/19740200801`.

Barbisan, L., Turvani, G., and Riente, F. (2024). A machine learning approach for queen bee detection through remote audio sensing to safeguard honeybee colonies. *IEEE Transactions on AgriFood Electronics*. DOI: 10.1109/TAFE.2024.3406648.

Blackman, S. and Harris, J. W. (1967). On the use of windows for harmonic analysis with the discrete fourier transform. *IEEE Transactions on Audio and Electroacoustics*, 15(2):236–241. DOI: 10.1109/PROC.1978.10837.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32. DOI: 10.1023/A:1010933404324.

Bromenshenk, J. J., Henderson, C. B., Seccomb, R. A., Rice, S. D., and Etter, R. T. (2009). Honey bee acoustic recording and analysis system for monitoring hive health. Available at:`https://patents.google.com/patent/US7549907B2/en`.

Cejrowski, T., Szymański, J., and Logofătu, D. (2020). Buzz-based recognition of the honeybee colony circadian rhythm. *Computers and Electronics in Agriculture*, 175:105586. DOI: 10.1016/j.compag.2020.105586.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297. Available at:`https://ise.ncsu.edu/wp-content/uploads/sites/9/2022/08/Cortes-Vapnik1995_Article_Support-vectorNetworks.pdf`.

de Oliveira, M. C., Pereira, F. d. M., de Moura, V. G., Brito, M. A., dos Santos, B. R., de Oliveira, M. C., and Magalhaes, D. M. (2023). Aquisição e classificação da intensidade da colmeia usando características cepstrais. In *Anais do XV Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 31–40. SBC. DOI: 10.5753/sbcup.2023.230536.

DeGrandi-Hoffman, G., Wardell, G., Ahumada-Segura, F., Rinderer, T., Danka, R., and Pettis, J. (2008). Comparisons of pollen substitute diets for honey bees: consumption rates by colonies and effects on brood and adult populations. *Journal of apicultural research*, 47(4):265–270. DOI: 10.1080/00218839.2008.11101473.

Di, N., Sharif, M. Z., Hu, Z., Xue, R., and Yu, B. (2023). Applicability of vggish embedding in bee colony monitoring: comparison with mfcc in colony sound classification. *PeerJ*, 11:e14696. DOI: 10.7717/peerj.14696.

Dong, X., Luu, A. T., Lin, M., Yan, S., and Zhang, H. (2021). How should pre-trained language models be fine-tuned towards adversarial robustness? *Advances in Neural Information Processing Systems*, 34:4356–4369. Available at:`https://proceedings.neurips.cc/paper/2021/hash/22b1f2e0983160db6f7bb9f62f4dbb39-Abstract.html`.

Gorroi, G., Freitas, L. P. V. d., and Assis, D. C. S. d. (2020). Apicultura: o manejo das abelhas do gênero apis. *Cad. técn. Vet. Zoot.*, pages 9–36.

Available at:https://vet.ufmg.br/ARQUIVOS/FCK/Cadernos%20T%C3%A9cnicos%20-%2096%20-%20para%20internet%20(1).pdf.

Haykin, S. (2001). *Redes neurais: princípios e prática*. Bookman Editora. Book.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778. Available at:https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

Heise, D., Miller, Z., Wallace, M., and Galen, C. (2020). Bumble bee traffic monitoring using acoustics. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE. DOI: 10.1109/I2MTC43012.2020.9129582.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. DOI: 10.48550/arXiv.1704.04861.

Kim, J., Oh, J., and Heo, T.-Y. (2021). Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and grad-cam. *Mathematical Problems in Engineering*, 2021:1–13. DOI: 10.1155/2021/5594498.

Kulyukin, V. (2021). Audio, image, video, and weather datasets for continuous electronic beehive monitoring. *Applied Sciences*, 11(10):4632. DOI: 10.3390/app11104632.

Kumar, R. and Mall, P. (2018). Important traits for the selection of honey bee (apis mellifera l.) colonies. *J. Entomol*, 6:906–909. Available at:https://www.entomoljournal.com/archives/2018/vol6issue3/PartM/6-2-267-642.pdf.

Leng, Y., Sun, C., Cheng, C., Xu, X., Li, S., Wan, H., Fang, J., and Li, D. (2015). Classification of overlapped audio events based on at, plsa, and the combination of them. *Radioengineering*, 24(2):593–603. Available at:https://www.radioeng.cz/fulltexts/2015/15_02_0593_0603.pdf.

Mazepa, C. I. and Laurenti, C. R. S. (2022). Evolución del estado sanitario en colmenas de apis mellifera l. bajo distinas condiciones de manejo y su relación con el aporte nutricional del polen. *Agrotecnia*, (32):34–56. DOI: 10.30972/agr.0326339.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, pages 18–25. Available at:https://www.researchgate.net/publication/328777063_librosa_Audio_and_Music_Signal_Analysis_in_Python.

Oliveira Costa, R., Bezerra, A. H. A., Ferreira, A. C., Pereira, B. B. M., Pimenta, T. A., and de Andrade, A. B. A. (2016). Análise hierárquica dos problemas existentes na produção de mel do estado da paraíba. *Revista Verde de Agroecologia e Desenvolvimento Sustentável*, 11(2):24–28. Available at:https://dialnet.unirioja.es/servlet/articulo?codigo=7264999.

Oppenheim, A. V. and Schafer, R. W. (2009). *Discrete-time signal processing*. Pearson Education. Book.

Phan, T.-T.-H., Nguyen-Doan, D., Nguyen-Huu, D., Nguyen-Van, H., and Pham-Hong, T. (2023). Investigation on new mel frequency cepstral coefficients features and hyper-parameters tuning technique for bee sound recognition. *Soft Computing*, 27(9):5873–5892. DOI: 10.1007/s00500-022-07596-6.

Rustam, F., Sharif, M. Z., Aljedaani, W., Lee, E., and Ashraf, I. (2024). Bee detection in bee hives using selective features from acoustic data. *Multimedia Tools and Applications*, 83(8):23269–23296. DOI: 10.1007/s11042-023-15192-5.

Ruvinga, S., Hunter, G., Duran, O., and Nebel, J.-C. (2023). Identifying queenlessness in honeybee hives from audio signals using machine learning. *Electronics*, 12(7):1627. DOI: 10.3390/electronics12071627.

Ruvinga, S., Hunter, G. J., Duran, O., and Nebel, J.-C. (2021). Use of lstm networks to identify "queenlessness" in honeybee hives from audio signals. In *2021 17th International Conference on Intelligent Environments (IE)*, pages 1–4. IEEE. DOI: 10.1109/IE51775.2021.9486575.

Sharif, M. Z., Wario, F., Di, N., Xue, R., and Liu, F. (2020). Soundscape indices: new features for classifying beehive audio samples. *Sociobiology*, 67(4):566–571. DOI: 10.13102/sociobiology.v67i4.5860.

Shostak, S. and Prodeus, A. (2019). Classification of the bee colony condition using spectral features. In *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, pages 737–740. IEEE. DOI: 10.1109/PICST47496.2019.9061441.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. DOI: 10.48550/arXiv.1409.1556.

Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (WACV)*, pages 464–472. IEEE. DOI: 10.1109/WACV.2017.58.

Soares, B. S., Luz, J. S., de Macêdo, V. F., e Silva, R. R. V., de Araújo, F. H. D., and Magalhães, D. M. V. (2022). Mfcc-based descriptor for bee queen presence detection. *Expert Systems with Applications*, 201:117104. DOI: 10.1016/j.eswa.2022.117104.

Vallenas-Sánchez, Y., Honorio-Javes, C. E., Valdivia-Camargo, V., and Rodríguez-Soto, J. C. (2023). Efecto de suplemento proteico sobre la postura y la población de colonias de abejas (apis mellifera l.) comerciales ubicadas en paisaje polifloral. *Ciencia y Tecnología Agropecuaria*, 24(2). DOI: 10.21930/rcta.vol24$_num2_art:3058$.

Vieira, F. R., Andrade, D. C., and Ribeiro, F. L. (2021). A polinização por abelhas sob a perspectiva da abordagem de serviços ecossistêmicos (ase). *Revista Ibero-Americana de Ciências Ambientais*, 12(4):544–560. DOI: 10.6008/CBPC2179-6858.2021.004.0042.

Virtanen, T., Plumbley, M. D., and Ellis, D. (2018). *Computational analysis of sound scenes and events*. Springer.

DOI: 10.1007/978-3-319-63450-0.

Wardhani, N. W. S., Rochayani, M. Y., Iriany, A., Sulistyono, A. D., and Lestantyo, P. (2019). Cross-validation metrics for evaluating classification performance on imbalanced data. In *International conference on computer, control, informatics and its applications*, pages 14–18. IEEE. DOI: 10.1109/IC3INA48034.2019.8949568.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27. Available at:`https://proceedings.neurips.cc/paper_files/paper/2014/hash/375c71349b295fbe2dcdca9206f20a06-Abstract.html`.

Yu, B., Huang, X., Sharif, M. Z., Jiang, X., Di, N., and Liu, F. (2023). A matter of the beehive sound: Can honey bees alert the pollution out of their hives? *Environmental Science and Pollution Research*, 30(6):16266–16276. DOI: 10.1007/s11356-022-23322-z.

Zgank, A. (2021). Iot-based bee swarm activity acoustic classification using deep neural networks. *Sensors*, 21(3):676. DOI: 10.3390/s21030676.

Zhang, T., Zmyslony, S., Nozdrenkov, S., Smith, M., and Hopkins, B. (2021). Semi-supervised audio representation learning for modeling beehive strengths. *arXiv preprint arXiv:2105.10536*. DOI: 10.48550/arXiv.2105.10536.

Zhang, Y., Guo, Z., Wu, J., Tian, Y., Tang, H., and Guo, X. (2022). Real-time vehicle detection based on improved yolo v5. *Sustainability*, 14(19):12274. DOI: 10.3390/su141912274.

Zhao, Y., Deng, G., Zhang, L., Di, N., Jiang, X., and Li, Z. (2021). Based investigate of beehive sound to detect air pollutants by machine learning. *Ecological Informatics*, 61:101246. DOI: 10.1016/j.ecoinf.2021.101246.

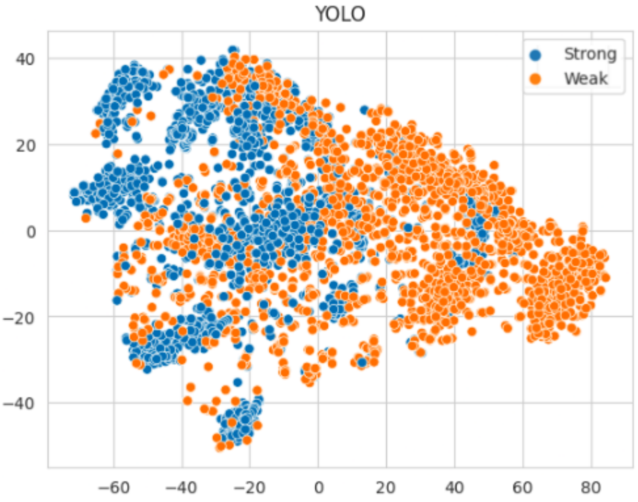

Figure 10. TSNE scatterplot illustrates the separation between the weak and strong classes based on the features extracted by YOLO. It is important to note the several regions of overlap between classes, reinforcing the poor classification performance presented by this descriptor compared to the descriptors from dense architectures such as Resnet50 and VGG16.