# Short-Time Fourier Transform for detecting the queen bee state

Thi-Thu-Hong Phan[1]⋆, Duc Minh Nguyen[1]⋆⋆, Ho Nhat Huy Nguyen[1], Trung Quan Hoang[1], Tri Nhan Nguyen[1], Hong Thai Pham[2], and Mai The Vu[3]

[1] Artificial Intelligence Department, FPT University, Danang, Vietnam
[2] Research Center for Tropical Bees and Beekeeping, Vietnam National University of Agriculture, Hanoi, Vietnam
[3] Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, South Korea

Email: hongptt11@fe.edu.vn, ducminhnhue@gmail.com, huyhods@gmail.com, quanht782619@gmail.com, nhannguyentri11052003@gmail.com, phthai@vnua.edu.vn, maithevu90@sejong.ac.kr

**Abstract.** This study explores the potential of sound analysis to detect the queenless in beehives, a significant challenge for beekeepers. Comparing Mel-Frequency Cepstral Coefficients (MFCCs) and Short-Time Fourier Transform (STFT) as feature extraction methods, our proposal shows that machine learning models utilizing low-frequency (below 1200 Hz) features extracted by STFT outperform these models using MFCCs derived from the full spectrum for the queen bee presence detection in beehives. The emphasis on low frequencies provides several benefits such as improved classification accuracy and enhanced computational efficiency by focusing on informative features. This paper contributes valuable insights for developing effective bee colony health monitoring systems.

**Keywords:** Queenless · STFT · MFCCs · Machine learning methods· low-frequency.

## 1 Introduction

Honey bees are vital pollinators, crucial for ecosystem stability and agricultural productivity. Their efficient foraging aids in plant reproduction, including essential crops, while their hive products like honey and beeswax serve diverse human needs. Additionally, honey bee populations reflect broader environmental conditions, underscoring their importance as ecosystem indicators, warranting conservation efforts.

Monitoring bee colony health is paramount for beekeepers. They need to be vigilant against issues like swarming, Varroa mites, missing queens, etc. The queen bee is essential for colony survival and emits pheromones that unify the

⋆ Corresponding author: hongptt11@fe.edu.vn
⋆⋆ Corresponding author: ducminhnhue@gmail.com

colony and maintain cohesion. Her absence can lead to decreased productivity and even cause the collapse of the bee colony. Therefore, regularly checking the presence and assessment of queen health is necessary for beekeepers to uphold colony health and productivity. However, traditional methods often involve daily hive inspections which can be time-consuming and disrupt the bees' behaviors and potentially harm the queen. This highlights the need for non-invasive approaches to detect queen absence and facilitate interventions. With strong developments, new technology provides new bee-friendly alternatives: sensors are mounted inside the beehive to collect information such as temperature, humidity, sound, etc. These data are then analyzed to predict behaviors of the bees. In particular, analyzing bee sounds has shown promise in distinguishing queenless from queenright state. In the literature, many studies have applied machine learning methods using sound as input to detect the queenless in the beehive. Ruvinga et al. (2021) [15] used LSTM networks with 13 Mel-frequency cepstral coefficients (MFCCs) and one log energy feature to detect "queenlessness" in honeybee hives, achieving 0.92 accuracy. Nolasco et al. (2019) [11] employed Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) for queen identification, achieving 0.91 accuracy with MFCCs and improving to 0.94 with MFCCs combined with Hilbert Huang Transform (HHT). Barbisan and Riente (2023) developed a machine learning framework using 20 MFCC features, achieving high accuracy with Neural Networks (NN) at 0.9885 and SVM at 0.9886 [1]. Rustam et al. (2023) [14] analyzed audio data using selective features and achieved 0.82 accuracy with Random Forest (RF) and 0.83 with K-Nearest Neighbors (KNN). Fourer et al. [**?**] investigated queen presence detection using mean Short-Time Fourier Transform (STFT) representations corresponding to distinct time bins with CNNs, achieving 96% accuracy. In a related study by Ho et al. (2023) [8], queen bee presence was successfully identified using a real dataset with a peak accuracy of 91.75% achieved by employing MFCCs for feature extraction.

In parallel with these advancements in machine learning techniques for monitoring bee colonies, researchers have also delved into the acoustic characteristics of honeybee behaviors. This aims to leverage sound-based monitoring as an additional tool for detecting and understanding various aspects of bee colony health and behavior. Seeley and Tautz (2001) [16] distinguished two worker piping forms in honeybees: "wings-together piping" and "wings-apart piping." The former, prevalent in swarms, features a frequency-modulated sound transitioning from low (100-200 Hz) to mixed higher frequencies (200-2000 Hz) as workers pull their wings together. Conversely, "wings-apart piping," found in hives, emits primarily low frequencies (300-400 Hz) as workers slightly spread their wings. Duran et al. (2013) [9] identified queen-less bee colonies through "warble" (225-285Hz) and "moaning" (165-285Hz) in beehive audio, indicative of queen inactivity and absence, respectively, using Fast Fourier Transform (FFT) and S-transform. Robles-Guerrero et al. (2017) [13] differentiated queenless colonies by observing unique FFT spectrum patterns centered around 400 Hz, validated through Logistic Regression with two crucial features. Cejrowski et al. (2018) [3]

focused on hardware, monitoring systems, and data analysis algorithms. Their equipment included bandpass filters to render the microphone sensitive to bee sounds within the frequency range of 20 to 2000 Hz. Utilizing the t-SNE algorithm, they observed distinct behavioral patterns between bees with and without the queen. Shostak and Prodeus (2019) [17] classified bee colony conditions based on power spectrum density, highlighting frequencies of 200 Hz and 250 Hz as significant features.

This study investigates a novel approach that focuses on low-frequency components extracted from the STFT method instead of analyzing every detail of an audio sample. The objective is to condense the data into a lower-dimensional representation that retains essential information, thereby improving computational efficiency and making it easier to understand how the sound characteristics influence the ML models.

The rest of the paper is organized as follows. Section 2 is to describe shortly the methods applied in the study. The experiments and obtained results have been presented in section 3. Some concluding remarks are given in section 4.

## 2 Methodology

Figure 1 illustrates the process for detecting the absence of the queen bee within a colony. Initially, bee sound data are collected from various beehives. Subsequently, important features are extracted from these data using well-established methods including MFCC and STFT. Notably, this study particularly emphasizes the low-frequency range of bee sound data. Therefore, only audio features within the 0-1200Hz range are retained to feed as input for the machine learning methods. Finally, these trained models are then utilized to predict new sound patterns, indicating the presence or absence of the queen bee within the colony.

In this analysis, we primarily examine features at lower frequencies. In figure 2, we observe a notable similarity in amplitude among frequency components approximately above 1200 Hz. This consistency across various features suggests potential redundancy within the dataset. Redundant features offer limited additional information to the model and may increase complexity without improving performance. Consequently, this could extend training time and heighten the likelihood of overfitting. Conversely, focusing on a narrower frequency spectrum simplifies the feature set and reduces model complexity, thereby potentially preventing overfitting and improving generalization to new data.

The following section provides a brief overview of the STFT algorithm and how to select low-frequency components. Additionally, we discuss different ML algorithms utilized to identify the queenless state.

### 2.1 Feature extraction algorithm: Short-Time Fourier Transform

STFT is a cornerstone technique in signal processing, widely applied in various domains such as audio processing, speech recognition, and biomedical engineering [5]. It provides a time-frequency representation of signals, allowing

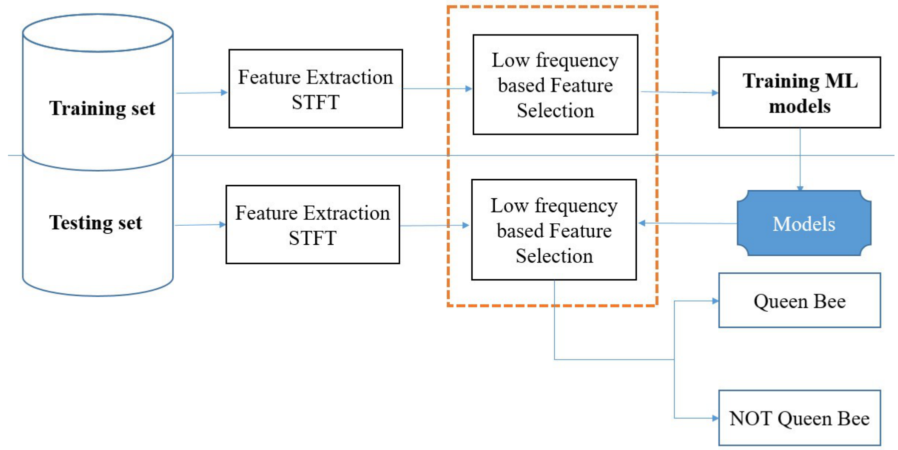


**Fig. 1.** Bee audio classification pipeline using STFT feature

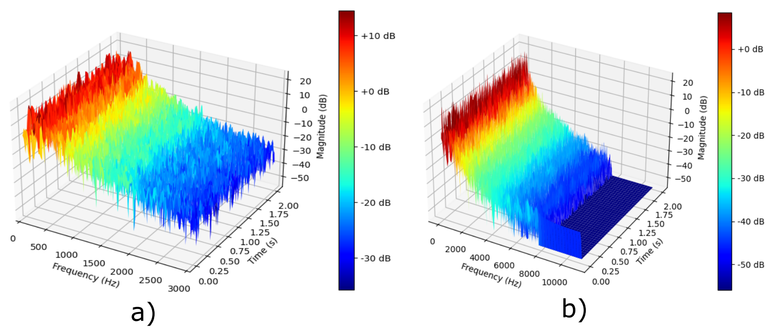


**Fig. 2.** Comparison of Frequency Distributions of a queen-less beehive: a) Full-Frequency Spectrum versus b) Low-Frequency Spectrum

for the analysis of evolving spectral content. Unlike traditional Fourier analysis for stationary signals, STFT computes localized Fourier transforms over small, overlapping windows, extending its capability to dynamic signals.

Mathematically, the STFT of a discrete signal $x[n]$ is defined as:

$$X[m, \omega] = \sum_{n=0}^{N-1} x[n] w[n-m] e^{-j\omega n}$$

Where $x[n]$ is the discrete input signal, $w[n]$ is the window function, $m$ represents the time shift or window position, $\omega$ denotes the frequency, $N$ is the length of the signal.

The magnitude or amplitude spectrum can be calculated to extract meaningful information from the complex-valued STFT output, followed by conversion to the decibel (dB) scale, which offers a logarithmic representation of signal

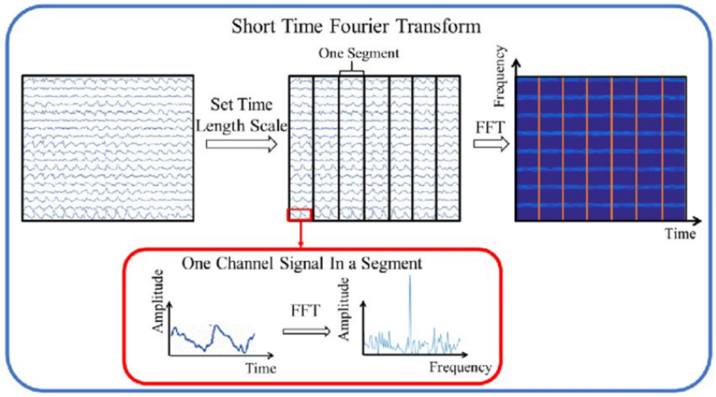


**Fig. 3.** Short-Time Fourier Transform(STFT)

power relative to a reference level. This conversion enhances the interpretability of the frequency content, particularly in applications such as audio processing and spectrum analysis.

The STFT produces a 2D matrix, with rows representing different time instances (called frames) and columns denoting various frequency bins. The number of frequency bins is determined by the formula:

$$\text{Number of frequency bins} = \frac{\mathrm{F}}{2} + 1$$

Meanwhile, the quantity of frames can be computed using the formula:

$$\text{Number of frames} = \left(\frac{\mathrm{S} - \mathrm{F}}{\mathrm{H}}\right) + 1$$

where $S$ denotes the total number of samples in the input signal, $F$ represents the number of samples in each windowed segment, and $H$ indicates the number of samples by which consecutive windows are shifted.

### 2.2 Machine learning models

**$K$-nearest Neighbors** is a versatile supervised learning algorithm used for both classification and regression tasks [4]. It operates by assigning a label to a new sample based on the labels of its $K$ closest neighbors in the training set. Typically, the class label is determined through a majority vote of these neighbors, with closer neighbors potentially exerting a higher influence (weighted vote). $K$NN relies on a distance metric, commonly Euclidean distance, to quantify the proximity between data points. This method offers the advantage of efficiently adapting to new data without requiring model retraining. However, it exhibits poor performance in high-dimensional spaces and is sensitive to noise and missing data within the training set.

**Support Vector Machines (SVM)** is a potent supervised learning algorithm employed for both classification and regression tasks. Inspired by statistical learning theory, SVM aims to identify the optimal hyperplane in a high-dimensional space that best separates different categories of data (classes) with the maximum possible margin [6]. This hyperplane is determined by a subset of data points known as support vectors, crucial for maximizing class separation. The effectiveness of SVM lies in its ability to handle linearly inseparable data using the kernel trick, which implicitly maps input features into higher-dimensional spaces, facilitating non-linear classification. In this study, we exclusively utilize the radial basis function (RBF) kernel due to its proven efficacy, supported by scholarly investigations and publications.

**Random Forest (RF)** is a robust machine learning algorithm that leverages the combined knowledge of the collective wisdom of multiple decision trees. Each tree in the forest is constructed using a random subset of the training data (a process known as bootstrapping), which promotes diversity among the trees and helps prevent overfitting [2]. The trees independently learn to classify or predict outcomes based on optimal splitting points (using metrics like Gini impurity). Each tree's predictions are then combined through aggregation: for regression tasks, the results are averaged, while for classification tasks, the majority vote determines the final prediction.

**Extra Trees (ET)** or Extremely Randomized Trees, similar to the RF algorithm, exploit an ensemble of decision trees for prediction [7]. However, ET differs by utilizing the whole original sample, which helps in reducing bias. In addition to using the full sample, ET introduces randomness by selecting split points randomly instead of computing the local optimum using metrics like Gini impurity or entropy. This random selection of split points adds diversity and reduces correlation among the trees, contributing to the effectiveness of the algorithm. The ET al algorithm is known for its speed and ability to mitigate overfitting, making it particularly well-suited for large datasets with a high number of features.

**Logistic Regression (LR)** is a powerful method primarily used for classification tasks rather than continuous value prediction. This technique performs best when the relationship between the independent variables and the dependent variable (often binary) can be modeled linearly [12]. LR is particularly useful for estimating the probability of an event occurring. The coefficients produced by LR indicate the extent to which each independent variable influences the likelihood of a specific outcome. Its simplicity and effectiveness make it a popular choice for binary classification problems where understanding variable influences is essential.

## 3 Experiments and Results

### 3.1 Data description

This assessment utilizes a real dataset of audio recordings captured in 2022 from various beehives at the Research Center for Tropical Bees and Beekeeping-Vietnam National University of Agriculture, introduced in the previous study [8].

The recordings, taken under varying ambient noise levels, are categorized into two groups: those with the presence of the queen bee and those without. The audio data are then segmented into 20,000 2-second clips with 1-second overlaps for consistent analysis. The method was akin to that used in [10], where each 30-second audio clip was divided into 2-second segments with a 1-second overlap, yielding 28 2-second segments per 30-second audio file. To evaluate machine learning methods, we employ the hold-out strategy, where the dataset is split into a training set (67%) and a testing set (33%) for model training and evaluation, respectively. An example of Spectrographic visualizations of audio clips for each category (presence/absence of queen bee) is shown in figure 2.

### 3.2 The performance of ML methods for recognizing bee sound state

Table 1 shows the performance of various machine learning algorithms (RF, ET, KNN, SVM, and LR) for detecting the queen bee within beehives. The table evaluates their accuracy across seven frequency ranges, ranging from 0-100 Hz to 0-1200 Hz. Additionally, the table shows the number of features extracted for each frequency range. This indicates the number of characteristics derived from the beehive audio data used by the algorithms. KNN achieves the highest accuracy in most frequency ranges, reaching up to 96.7% when using STFT features extracted from audio samples captured between 0-1200 Hz. Following KNN are RF and ET with consistently high accuracy, exceeding 90% in most ranges. SVM and LR have the lowest accuracy among the algorithms, generally below 92.5%.

**Table 1.** Accuracy in case feature extraction using low-frequency STFT

| **No.Features** | **10** | **20** | **29** | **38** | **57** | **80** | **112** |
|---|---|---|---|---|---|---|---|
| **Frequency (Hz)** | **0-100** | **0-200** | **0-300** | **0-400** | **0-600** | **0-850** | **0-1200** |
| **RF** | 0.810 | 0.905 | 0.931 | 0.937 | 0.945 | 0.943 | 0.947 |
| **ET** | 0.816 | 0.908 | 0.939 | 0.944 | 0.946 | 0.949 | 0.9545 |
| **KNN** | 0.781 | 0.900 | 0.944 | 0.953 | 0.961 | 0.965 | **0.967** |
| **SVM** | 0.769 | 0.857 | 0.896 | 0.903 | 0.917 | 0.921 | 0.9245 |
| **LR** | 0.622 | 0.721 | 0.768 | 0.794 | 0.818 | 0.823 | 0.835 |

Table 2 presents a comparative analysis of different machine learning algorithms for the task of queenless detection, emphasizing the use of MFCCs for feature extraction. MFCCs are a method to transform raw audio data into a format that highlights frequencies most important to human hearing. The performance of the algorithms seems to be somewhat dependent on the MFCC variations used. For example, RF performs best with 13 MFCCs, while ET performs best with both 13 MFCCs and 26 MFCCs. Overall, ET appears to be the top-performing algorithm across all MFCC variations, with accuracy values ranging from 93.2% to 91%.

This table also indicates that using fewer features (13 MFCCs) performs better than using more features (26 or 39 MFCCs). This might suggest that higher-order MFCC features (beyond the 1st and 2nd derivatives ) capture less significant information for the specific task. In essence, MFCC features are supposed to capture the spectral envelope of a sound, which is the distribution of energy across different frequencies. The first few MFCC coefficients capture the most important information in the spectrum, while higher-order coefficients capture finer details. However, these finer details may not be relevant for the task of queenless detection, and may even introduce noise or irrelevant information.

**Table 2.** Accuracy in case feature extraction using MFCCs

| Method | 13 MFCCs | 26 MFCCs | 39 MFCCs |
|---|---|---|---|
| **RF** | 0.921 | 0.910 | 0.894 |
| **ET** | 0.932 | 0.91 | 0.887 |
| **SVM** | 0.908 | 0.869 | 0.851 |
| **KNN** | 0.926 | 0.818 | 0.774 |
| **LR** | 0.723 | 0.720 | 0.720 |

Results from tables 1 and 2 show that the utilization of low-frequency STFT demonstrates enhancements in classification accuracy when using MFCCs derived from the full spectrum. Optimal classification accuracy with STFT is achieved by focusing on frequencies below 1200 Hz in conjunction with the KNN algorithm. On the other hand, MFCCs exhibit their best performance when employing the first 13 coefficients in combination with the ET method. The accuracy achieved using the low-frequency STFT method reaches 96.7%, showcasing a notable improvement of 3.475% over the MFCC approach. This can be explained upon examining the 3D spectrogram 2. We have observed heightened amplitudes in the lower frequency spectrum (below approximately 1200 Hz). This suggests that these frequencies might hold greater significance in distinguishing the presence or absence of queen bees. Amplified signals within this range could offer clearer indications of queen bee activity.

Moreover, the queen bee and its activities may predominantly emit sounds in the lower frequency spectrum. The characteristic sounds associated with queen bee presence are more pronounced in lower frequencies, prioritizing these bins can effectively capture the relevant information for classification. Besides, higher frequency bins may contain more noise or irrelevant information that could potentially confuse the machine learning models. Filtering out frequency bins exceeding 1200 Hz can diminish noise within our dataset, enabling the model to focus on the most relevant features for classification.

In addition to comparing the performance of two feature extraction algorithms to detect the queenless state, we also compare the training time of ML algorithms when using these two types of features. Figure 4 presents the training time of various ML methods when using MFCCs (39 features) and STFT

(112 features). Although SFTF has more features, the computational time is not excessively complex and remains manageable.

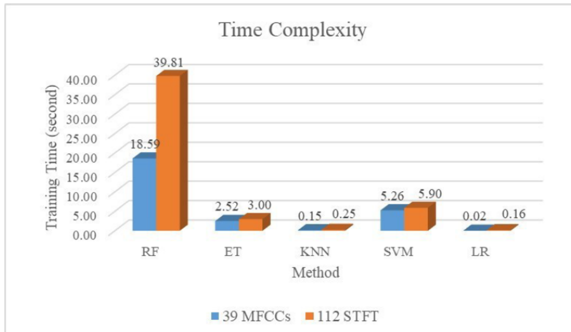


**Fig. 4.** Training time compare between 39 MFCCs and 112 STFT

## 4 Conclusion

This paper investigates the efficacy of sound analysis in detecting the queenless state in beehives. The experiment results demonstrate that utilizing low-frequency components extracted from STFT in beehives is a powerful tool for beehive monitoring. In comparison to a conventional method based on MFCCs, our approach achieves superior accuracy, particularly associating features with frequencies below 1200 Hz with the KNN algorithm. Examination of the 3D spectrogram highlights the importance of lower frequency amplitudes in identifying the queen bee, while higher frequencies exhibited uniformity, suggesting redundancy. This study also reveals that the information captured by the first few MFCC coefficients is more relevant for queen bee detection than using a wider range. Deep learning models have the potential to significantly improve the performance of this task in the future.

### Acknowledgment

This research was funded by the Vietnam national research project titled "Study and application on industry 4.0 technologies in the management of honey bee production for export and national consumption," Grant Number: KC4.0-20/19-25. The funders had no role in designing experiments, collecting and processing data, deciding to publish, or preparing the manuscript.

## References

1. Barbisan, L., Riente, F.: Machine learning framework for the acoustic detection of the queen bee presence. Acta Acustica (2023)

2. Breiman, L.: Random forests. Machine Learning **45**, 5–32 (2001)
3. Cejrowski, T., Szymanski, J., Mora, H.M., Gil, D.: Detection of the bee queen presence using sound analysis. Intelligent Information and Database Systems pp. 297–306 (2018)
4. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory (1967)
5. Durak Ata, L., Arikan, O.: Short-time fourier transform: Two fundamental properties and an optimal implementation. Signal Processing, IEEE Transactions on **51**, 1231 – 1242 (06 2003). https://doi.org/10.1109/TSP.2003.810293
6. Evgeniou, T., Pontil, M.: Support vector machines: Theory and applications. Machine Learning and Its Applications, Advanced Lectures (2001)
7. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach Learn **63**, 3–42 (2006)
8. Ho, H.T., Pham, M.T., Tran, Q.D., Pham, Q.H., Phan, T.T.H.: Evaluating audio feature extraction methods for identifying bee queen presence. In: SOICT '23: Proceedings of the 12th International Symposium on Information and Communication Technology. pp. 93–100 (12 2023). https://doi.org/10.1145/3628797.3628852
9. Howard, D., Duran, O., Hunter, G., Stebel, K.: Signal processing the acoustics of honeybees (APIS MELLIFERA) to identify the "queenless" state in hives. Proceedings of the Institute of Acoustics (2013)
10. Kulyukin, V.A., Mukherjee, S., Amlathe, P.: Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples. Applied Sciences (2018), https://api.semanticscholar.org/CorpusID:55421005
11. Nolasco, I., Terenzi, A., Cecchi, S., Orcioni, S., Bear, H.L., Benetos, E.: Audio-based Identification of Beehive States. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 8256–8260 (May 2019). https://doi.org/10.1109/ICASSP.2019.8682981, https://ieeexplore.ieee.org/document/8682981, iSSN: 2379-190X
12. Peng, J., Lee, K.L., Ingersoll, G.M.: An introduction to logistic regression analysis and reporting. The Journal of Educational Research **96**(1), 3–14 (2002)
13. Robles-Guerrero, A., Saucedo-Anaya, T., González-Ramírez, E., Galván-Tejada, C.E.: Frequency analysis of honey bee buzz for automatic recognition of health status: A preliminary study. Research in Computing Science **142** (2017)
14. Rustam, F., Sharif, M.Z., Aljedaani, W., Lee, E., Ashraf, I.: Bee detection in bee hives using selective features from acoustic data. Multimedia Tools and Applications **83**, 23269–23296 (2023)
15. Ruvinga, S., Hunter, G.J., Duran, O., Nebel, J.C.: Use of LSTM Networks to Identify "Queenlessness" in Honeybee Hives from Audio Signals. In: 2021 17th International Conference on Intelligent Environments (IE). pp. 1–4. IEEE (Jun 2021). https://doi.org/10.1109/IE51775.2021.9486575, https://ieeexplore.ieee.org/document/9486575/
16. Seeley, T., Tautz, J.: Worker piping in honey bee swarms and its role in preparing for liftoff. Journal of Comparative Physiology A (2001)
17. Shostak, S., Prodeus, A.: Classification of the bee colony condition using spectral features. 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (2019)