Evaluating Audio Feature Extraction Methods for Identifying Bee Queen Presence

Huu-Tuong HO
FPT University
Da Nang, Vietnam
tuonghhde170471@fpt.edu.vn

Minh-Tien PHAM
FPT University
Da Nang, Vietnam
tienpmde170231@fpt.edu.vn

Quang-Duong TRAN
FPT University
Da Nang, Vietnam
duongtqde160638@fpt.edu.vn

Quang-Huy PHAM
FPT University
Da Nang, Vietnam
huypqde170011@fpt.edu.vn

Thi-Thu-Hong PHAN*
FPT University
Da Nang, Vietnam
HongPTT11@fe.edu.vn

ABSTRACT

Beehive monitoring is an essential task for beekeepers to keep under surveillance the health and productivity of their beehives. Traditional monitoring methods, such as visual inspection, are labor-intensive, time-consuming, and have negative effects on bee colonies. Recently, machine learning (ML) algorithms have emerged as a powerful tool for automated monitoring beehives using bee sounds. To apply the ML methods, the first main step is to extract important features from the original audio data. In this study, we examine the performance of various ML algorithms using six different audio feature extraction methods. Experiments have been conducted on an audio dataset collected in Vietnam when the queen bee is present or absent. The experiment results indicate that the audio-based approach can effectively monitor beehives and by choosing the suitable feature extraction technique the performance of the ML methods for detecting the absence of the bee queen can be improved significantly.

CCS CONCEPTS

• **Computing methodologies → Machine learning algorithms**; • **Hardware → Digital signal processing**.

KEYWORDS

Audio feature extraction, Queen bee detection, MFCCs, Machine learning, KNN, SVM, Ensemble learning

ACM Reference Format:

Huu-Tuong HO, Minh-Tien PHAM, Quang-Duong TRAN, Quang-Huy PHAM, and Thi-Thu-Hong PHAN*. 2023. Evaluating Audio Feature Extraction Methods for Identifying Bee Queen Presence . In *The 12th International Symposium on Information and Communication Technology (SOICT 2023), December 7–8, 2023, Ho Chi Minh, Vietnam.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3628797.3628852

SOICT 2023, December 7–8, 2023, Ho Chi Minh, Vietnam

ACM ISBN 979-8-4007-0891-6/23/12
https://doi.org/10.1145/3628797.3628852

1 INTRODUCTION

Honeybees play a crucial role in our lives thanks to pollinating many of the world's food crops, wild plants and contributing to the diversity of natural ecosystems. Breeze et al indicated that about 35% of global crop productions are pollinated by bees [3]. For economics, numerous products provided by honeybees like honey, royal jelly, and beeswax have been popularly used as food, supported human health, and brought high economic value [1]. As a result of the great advantages for our real life, keeping an eye on the health of honeybees is a fundamental task for beekeepers. It could be checking to see if the queen bee is in the hive or if bees can get sick or be infected with viruses, etc. Traditionally, these tasks have been done manually by opening and closing a hive regularly to inspect the inside of the hive. However, this can be time-consuming, and labor-intensive and also causes stress or even panic in the bees. This has prompted the need to develop automatic monitoring systems to reduce negative impacts on bee colonies. These systems use advanced technologies such as the Internet of Things (IoT) devices to continuously collect data from the beehive, then process the data and apply artificial intelligence algorithms for analyzing/detecting the status and behavior of the honeybees. In recent years, there has been a growing interest in using acoustic monitoring to track the health of beehives. This is based on the fact that honeybees produce various sounds that contain a lot of useful information and can be used to identify swarming, queenlessness, and other problems in the beehive [5, 6, 10].

In 2008, Ferrari et al. pioneered the creation of acoustic monitoring systems that leveraged bee sounds, temperature, and humidity data. These systems were designed to ascertain the state of bee colonies, identify swarming preparations, and utilize predictive capabilities to proactively avert swarming events [10]. Howard et al. investigated various frequency analysis methods including spectrograms, fast Fourier transform and Stockwell transform or S-transform using a type of artificial neural network known as Kohonen Self Organising Map [13] to detect the present queen bee. In the study of Nolasco et al. [17], the authors implemented SVM and CNN models to check the existence of the queen bee in a beehive. They applied MFCCs, Mel spectrograms, and the Hilbert Huang Transform (HHT) to extract meaningful features. In another study to detect the queen presence [6], Cejrowski et al. developed

[1] https://www.statista.com/statistics/933928/global-market-value-of-honey/

linear predictive coding to extract features and fed them to the SVM algorithm for identifying the acoustic patterns of healthy and the queenless state. Orloswska et al. investigated the transformation of spectrograms converting from audio data in order to improve the performance of determining queenless state [25]. The authors pointed out that this transformation could better represent a generalization of data and increase the ability of the CNN. In the work conducted by Ruvinga et al., the research centers around the utilization of LSTM, Multi-Layer Perceptron Neural Networks, and Logistic Regression techniques to identify the absence of a queen bee within a beehive. They used only MFCCs for extracting features and then provided them to ML models. Notably, preliminary findings, particularly those stemming from the LSTM approach, demonstrate a strong and promising outcome [26].

To apply machine learning methods, raw audio data cannot be used. It is necessary to employ techniques to extract key features from raw audio data previously providing them to ML methods. Nolasco [17] and Phan [19] pointed out that selecting the right approaches to extract key features can increase the performance of ML algorithms significantly. These extraction algorithms impressively decrease the dimensionality of the data while still maintaining important information representing the data.

Therefore, in this study, we investigate the performance of various feature extraction techniques combined with diverse ML methods for the task of identifying the presence of the queen bee or not. Numerous feature extraction techniques have been conducted and evaluated in terms of classification performance, including the Mel frequency cepstral coefficients (MFCCs), the Short-time Fourier transform (STFT), the Fast Fourier Transform (FFT), the Constant-Q transform (CQT), the Spectral contrast (SC), and the Chroma. The paper is organized as follows. Section 2 describes the proposed approach, including the feature extraction and classification methodologies used. Section 3 explains the data obtained from beehives and discusses about results. Finally, section 4 shows conclusions and future work.

2 METHODOLOGY

Figure 1 depicts the schema of recognizing states of beehive based on bee sound using machine learning (ML) algorithms. This process includes 2 phases: i) model training and ii) model testing. In order to apply ML techniques, the first step is to extract features from the raw signals. Feature extraction is the process of transforming raw data into a set of features that are more relevant for machine learning algorithms. The goal of feature extraction is to reduce the dimensionality of the data while preserving the most important information. These features are then provided to machine learning algorithms as input to classify the states of the bee colony. In this study, we investigate different techniques for extracting features and then provide them to ML methods to find out which feature extraction algorithm is suitable for detecting the queen bee in a beehive. A summary of these methods is described below.

2.1 Feature extraction techniques

Different methods for extracting features namely, Mel-frequency cepstral coefficients (MFCCs), Short-time Fourier transform (STFT), Fast Fourier transform(FFT), Constant-Q Transform (CQT), Spectral Contrast (SC), and Chroma are studied in this paper.

2.1.1 Mel-frequency cepstral coefficients (MFCCs). The Mel - Frequency Cepstral Coefficients (MFCCs) is one of the most popular techniques deployed for extracting important features from sound signals [30]. Using this technique, the audio is divided into short frames, analyzed with a frequency using a Mel filterbank to capture relevant frequency information, the logarithm of the filterbank energies, and applied a Discrete Cosine Transform to obtain a set of MFCCs coefficients. The main steps are presented in figure 2 and summarized in the table 1.

Table 1: MFCCs Computation Process.

Step	Description
1	Pre-emphasis: Apply a high-pass filter to emphasize higher frequencies.
2	Framing: Split the pre-emphasis signal into small-equal frames duration.
3	Windowing: Multiply each frame by a windowing function (e.g., Hamming or Hanning)
4	Fast Fourier Transform (FFT): Compute the magnitude spectrum using FFT
5	Mel Filterbank: Apply a Mel filterbank to the magnitude spectrum
6	Logarithm: Transform the filterbank outputs using a logarithmic operation
7	Discrete Cosine Transform (DCT): Apply DCT to obtain final MFCCs

The formula for calculating Mel Frequency Cepstral Coefficients (MFCCs) is as follows:

$$C_n = \sum_{m=1}^{M} \log\left(\frac{1}{E}\sum_{k=1}^{N} X(k)\cdot\cos\left(\frac{\pi m}{N}\left(k-\frac{1}{2}\right)\right)\right) \tag{1}$$

Where:

- C_n represents the n-th MFCCs coefficient, which captures a specific aspect of the spectral content of the input signal.
- M represents the total number of Mel filters utilized in the calculation.
- $X(k)$ represents the magnitude spectrum or filterbank output at frequency bin k, indicating the spectral content of the input signal at that specific frequency bin.
- The term $\cos\left(\frac{\pi m}{N}\left(k-\frac{1}{2}\right)\right)$ represents the DCT basis function, which determines the weight or contribution of each frequency bin k to the m-th MFCCs coefficient.
- E is an optional energy normalization factor used to normalize the overall energy or power of the signal, ensuring that MFCCs are not affected by variations in signal intensity.

2.1.2 Short-time Fourier transform (STFT). STFT also called the time-frequency analysis, is one of the most fundamental and powerful tools in audio signal processing [35]. The STFT converts the filtered time-domain signals into the time-frequency domain during each short-time section. Because a filtered time-domain signal is

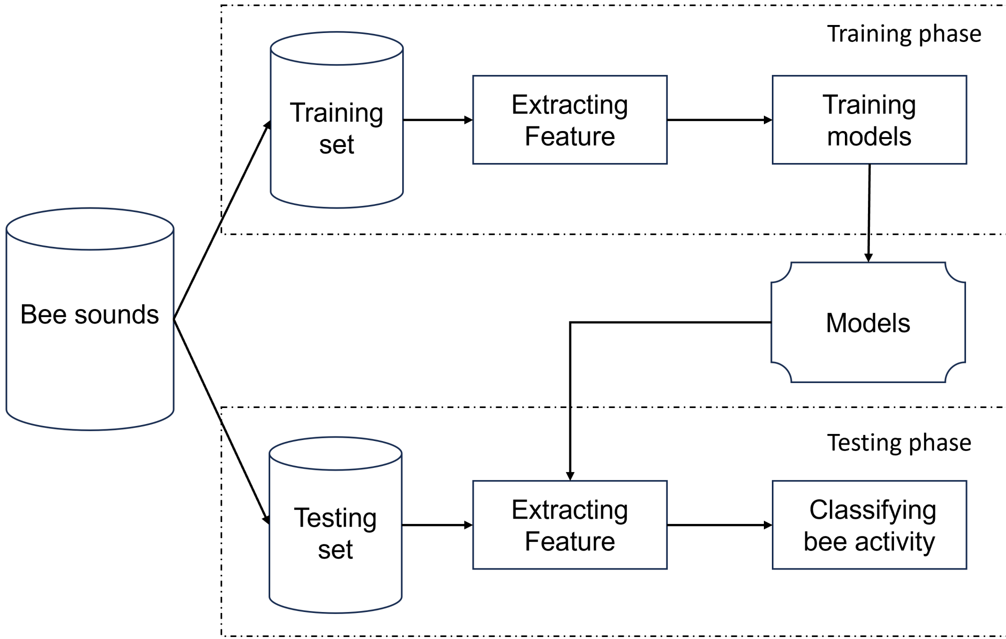


Figure 1: Bee audio classification pipeline.

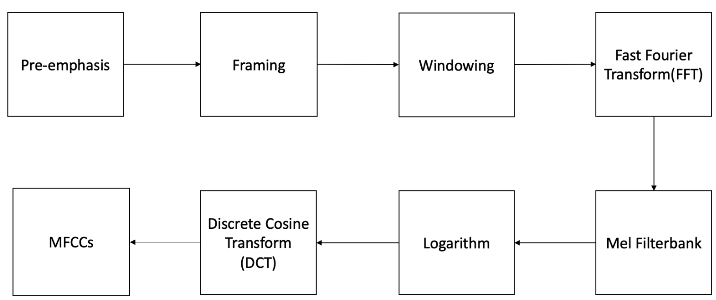


Figure 2: The process of computing MFCCs.

a time-varying signal, its amplitude changes over time. This can make it challenging to analyze, as the changing amplitude can obscure the features of the signal. Time-frequency (TF) captures the time-varying nature of signals, localizes frequency content, provides visualization of spectral energy, enables feature extraction, and supports a wide range of signal processing applications [12]. Assume the original sound signal $x(t)$ is short-term stationary, the window function is $h(t-\tau)$, and the short-time Fourier transform of signal $x(t)$ is computed as follows:

$$\mathbf{STFT}\{x[t]\}(\tau,\omega) = S(\tau,\omega) = \int_{-\infty}^{\infty} x(t)h(t-\tau)e^{-j\omega t}\,dt \quad (2)$$

With the output of equation 2, the original sound signal $x(t)$ will be transformed into a function in the time-frequency plane $h(t-\tau)$. As a result, $S(\tau,\omega)$ stores information from the time domain and of the frequency domain for the sound signal [21].

2.1.3 Fast Fourier transform(FFT). FFT is an efficient algorithm for computing the Discrete Fourier Transform (DFT) of a long signal in a short time [33] and its inverse. DFT is a math transform of a signal time cyclic to signal frequency cyclic [31]. DFT is applied in signal processing, image processing, and signal encoding [23]. It has many ways to calculate DFT but the FFT is the high performance [33].

FFT 1D is an algorithm that can calculate the DFT of a 1D signal. It is a recursion algorithm with a number of compute liners with the size of the signal. The FFT 1D works based on splitting the root signal into the small signal and then calculating it all. Finally, we coop all that result and have the final result of the root signal [32].

The formal application of FFT 1D in sound processing is to analyze the audio spectrum [22]. Audio spectrum is the chart that can present the intensity of audio. Audio spectrum can be used to distinguish different sounds, FFT 1D calculation formula is:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{nk}{N}} \quad (3)$$

where:

- $X[k]$ is the DFT of the signal $x[n]$ at frequency k.
- $x[n]$ is the input signal.
- N is the length of the signal.
- k is the frequency index.

2.1.4 Constant-Q Transform (CQT). The CQT is a famous signal processing technique often used in music analysis and audio processing. CQT is a type of Fourier transform that applies a logarithmic frequency scale to represent musical pitches and harmonics accurately. The CQT transform of a discrete time-domain signal x(n) is defined as follows [27]:

$$X^{C,Q}(k,n) = \sum_{j=n-\lfloor \frac{N_k}{2} \rfloor}^{n+\lfloor \frac{N_k}{2} \rfloor} x(j) a_k^*(j - n + \frac{N_k}{2}) \quad (4)$$

where:

- $k = 1, 2, 3, ...K$ represents the frequency bin index
- n represents the time frame index
- $a_k^*(n)$ denotes the complex conjugate of the CQT window function $a(n)$

2.1.5 Spectral Contrast (SC). SC is a feature extraction method used for audio signal processing and music analysis to obtain the perceptual contrast between different frequency bands in an audio signal. This method calculates the peak differences and valleys in the spectrum. The initial idea was proposed first time by Boers [2], where spectrum levels have been squared up, then normalizing the amplitude. To compute the SC features, firstly, the audio signal is divided into a series of frequency bands, typically using a Mel filter bank. For each frequency band, the energy/ magnitude of the signal within that band is computed. This energy is calculated by aggregating the squared values of the signal's Fourier coefficients within the band. The energy in each band is then compared to the energy in adjoining bands to calculate SC.

2.1.6 Chroma. The Chroma feature extraction method is a technique used in music and audio signal processing to represent the harmonic content and tonal characteristics of audio signals. Based on human perception of pitch, a pitch is divided into two components: tone height and chroma [16]. The tone height refers to the octave number and the chroma is a set of 12 pitches [29], each of which is one of the whole numbers of pitch classes, beginning with pitch class C and ending with pitch class B is C, C#, D, D#, E, F, F#, G, G#, A, A#, B [28]. The chroma features are derived from the spectrogram representation of the audio signal, and the chroma features extraction processing is depicted in figure 3.

2.2 Machine learning methods

2.2.1 K-Nearest Neighbors (KNN). KNN is a supervised learning algorithm that does not require training data during the training phase and can be applied for classification and regression problems [9]. For the classification complication, the prediction for a new data point's label (or the result of an exam question) is determined by directly analyzing the K closest data points in the training set. The label of test data can be ascertained through either majority voting, where each nearest point contributes one vote, or by making inferences based on different weightings assigned to each of the nearest points [7].

2.2.2 Support Vector Machine (SVM). SVM is a type of supervised machine learning algorithm used for classification and regression problems. SVM was introduced by Vladimir Vapnik and Alexey Chervonenkis in 1996 [34]. Figure 5 demonstrates the idea of this method. SVM works by looking for an ideal hyperplane that can split data into two classes while maximizing the margin - which is the distance of the hyperplane and the nearest data points from each class - between them. SVM is initially devoted to binary classification problems. Still, they can be extended to handle effectively in high-dimensional spaces for both linear and non-linear classification by using different kernel functions. Some popular kernel functions are presented as follows:

A polynomial kernel is commonly described as:

$$k(x,z) = (r + \gamma x^T z)^d \quad (5)$$

dd is a degree of kernel function.

Radial Basic Function (RBF) kernel or Gaussian kernel is the most used in practice, it is defined by:

$$k(x,z) = \exp\left(-\gamma \|x - z\|_2^2\right), \gamma > 0 \quad (6)$$

$\gamma, rrandddarekernelparameters$[18].

In this article, we only use RBF kernel, because its performance has shown good in some recent studies and papers [24] [14].

2.2.3 Random Forest (RF). RF, a famous ensemble learning, was proposed by Leo Breiman [4] in 2001. RF is a machine learning algorithm that depends on combining the results of many simple decision trees and can handle problems of classification and regression. The RF algorithm builds a set of decision trees, each tree being built on a randomly sampled subset of the original dataset.

The prediction of the random forest is the result of the consensus of the decision trees. RF uses a random selection of features to split each node in the trees and is more robust with respect to noise. This indicates that just a randomly chosen subset of the features is taken into consideration for splitting at each node in the tree. This helps to reduce the variance of the individual trees, which in turn leads to a more accurate and robust [4].

2.2.4 Extra trees (ET) . The ET model is an ensemble machine learning algorithm that uses decision trees and unpruned decisions [11]. It differs by three points from other tree-based ensemble models. First, the sampling of the training data is done completely randomly. Additionally, the splitting criteria for each node in the tree by choosing cut points are also entirely random. Finally, It uses the total learning sample instead of a bootstrap replica to grow the trees. ET creates an ensemble of extremely randomized trees. A random subset of training data is used to train each tree, and at each node, the following process is carried out: from the whole feature vector, first, choose a subset of KK features at random. Second, employ the KK characteristics to determine the node's ideal split. The final choice should be the one that maximizes the entropy Shannon information

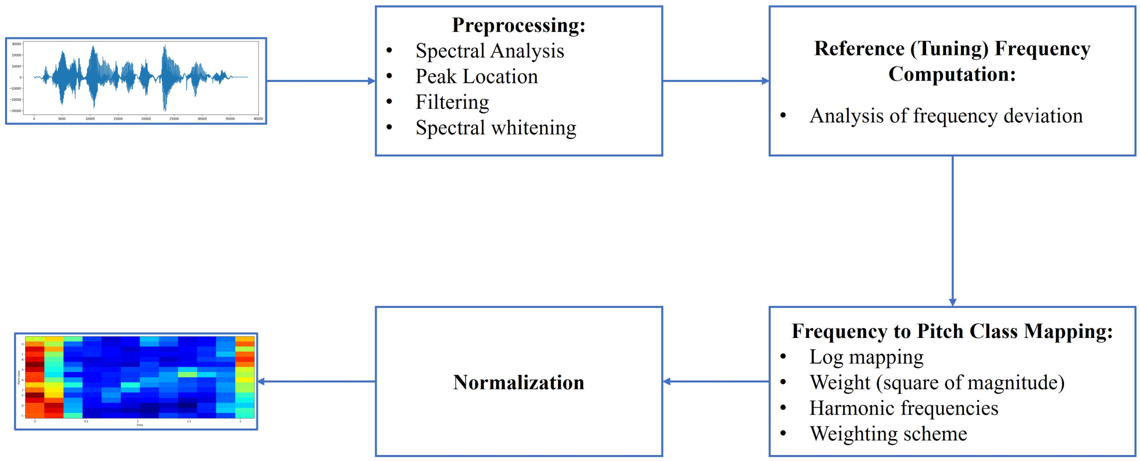


Figure 3: The Chroma features extraction pipeline [20].

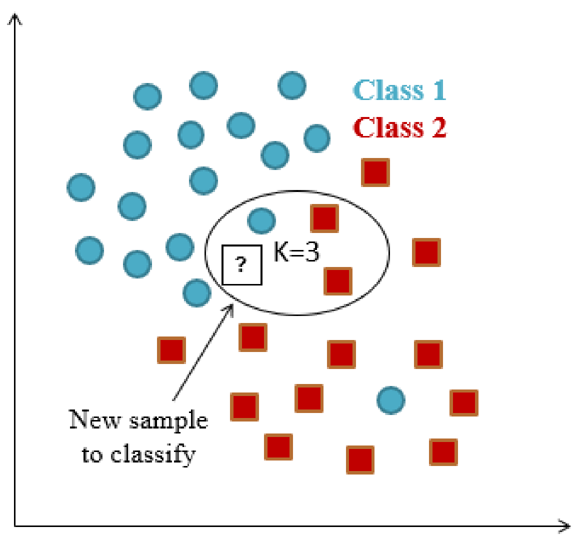


Figure 4: Example of two-class classification using KNN.

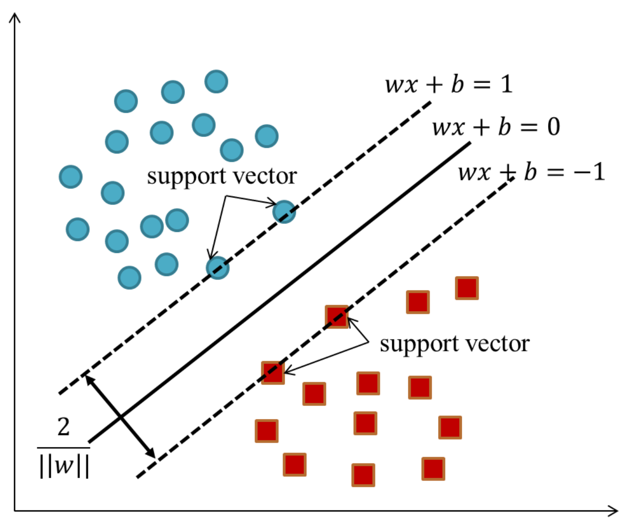


Figure 5: Diagram of SVM algorithm for linear classification.

gain score [15]. KK is a parameter that allows you to adjust the degree of randomness in tree generation: a high value for KK allows you to filter out unimportant factors, making the tree classification process more challenging. A high KK number will generate more dependable trees and reveal more correlations between trees in the forest; a low KK value will generate more random trees and hence lessen the correlation in the forest. However, the set of trees will help the classification process.

The ET has some advantages, such as a robust algorithm that is less prone to overfitting than other ensemble methods; it is fast to train, even on large datasets, and highly generalizable. Therefore, the Extra tree is a good choice for this problem.

2.2.5 Logistic regression (LR). LR is a widely used and interpretable classification algorithm that predicts the probability of a binary outcome (e.g., yes/no, 1/0, true/false). Although it is a form of regression analysis, LR predicts categorical variables such as binomial, ordinal, or multinomial variables instead of continuous variables like linear regression with some modifications. The LR model is fit by determining the coefficient values that maximize the likelihood of the observed data [1]. The coefficients represent each independent variable's effect on the log odds of the dependent variable. Logistic regression is a statistical model that uses the logistic function, which maps the dependent variables by the sigmoid function of independent variables and is represented by equation 7:

$$y = p(x) = \frac{1}{1 + e^{(\alpha + \Sigma B_k X_k)}} \tag{7}$$

The value of the dependent variable y is always bounded by the values 0 and 1 when using the sigmoid function [8].

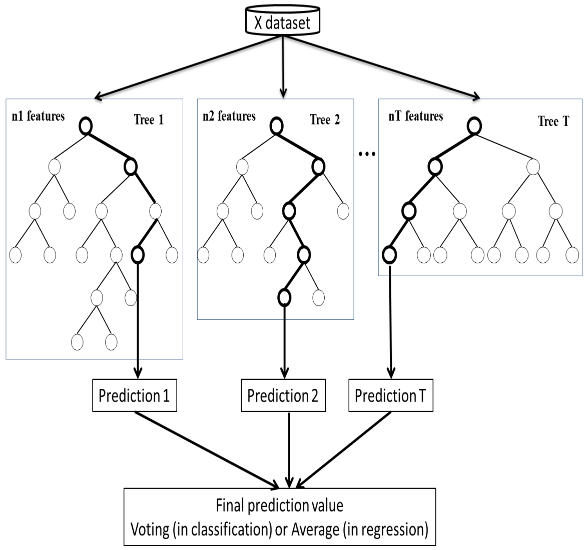


Figure 6: Diagram of random forest algorithm.

3 EXPERIMENTS AND RESULTS

3.1 Data description

For evaluating feature extraction algorithms, we use an audio dataset collected at the Research Center for Tropical Bees and Beekeeping, Vietnam National University of Agriculture in 2022. The data were recorded from various beehives under different environments and noise conditions such as construction, wind, traffic, and other ambient noises. Audio data were captured continuously at each beehive in two states: 1 - normal state (i.e. queen bee is present in the hive) and 2 - queenless state (i.e. no queen bee in the hive). The original data were then clipped into 20,000 2-second audio clips with 1-second overlaps. Finally, this dataset was divided into train, and test sets, with a ratio of 7:3 (e.g. 14.000 samples for training and 6.000 samples for testing). Examples of spectrogram visualization for audio clips from each state are presented in figure 7.

3.2 Experimental results

In this study, six different algorithms have been investigated for extracting important features from the original dataset, including the FFT features, the STFT features, the MFCCs features, the CQT features, the Chroma features, and the Spectral contrast features. Then these important features have been provided to various machine learning methods to identify the sounds when the queen bee is present or absent in the beehive.

Table 2 presents valuable insights into the performance of various feature extraction methods and machine learning algorithms on the test set. Among these methods, it is evident that MFCCs feature stands out as the top performer, consistently delivering remarkable accuracy rates across different experiments. Specifically, using MFCCs feature as input, all the ML algorithms achieve the highest accuracy among all the types of feature, with an impressive accuracy score when combining MFCCs features with the KNN classified algorithm (91.75%), demonstrating notable proficiency in classifying the queen bee audio data. With this feature type, we can see that most ML methods achieve accuracy higher than 90%, specifically 91.75%, 91.25%, 90.62%, and 91.7% respectively for KNN, SVM, RF, and ET methods. However, it's noteworthy that there is one exception to this trend, where LR only achieves 78.32% of accuracy using MFCCs features. Although LR has a lower performance than other ML methods based on MFCCs but when comparing the performance of this method combined with the other feature types, LR remains the top-performing algorithm as paired with MFCCs. This shows that the MFCCs feature is a very effective feature extraction technique and also emphasizes the significant impact of feature extraction type on the performance of classified models.

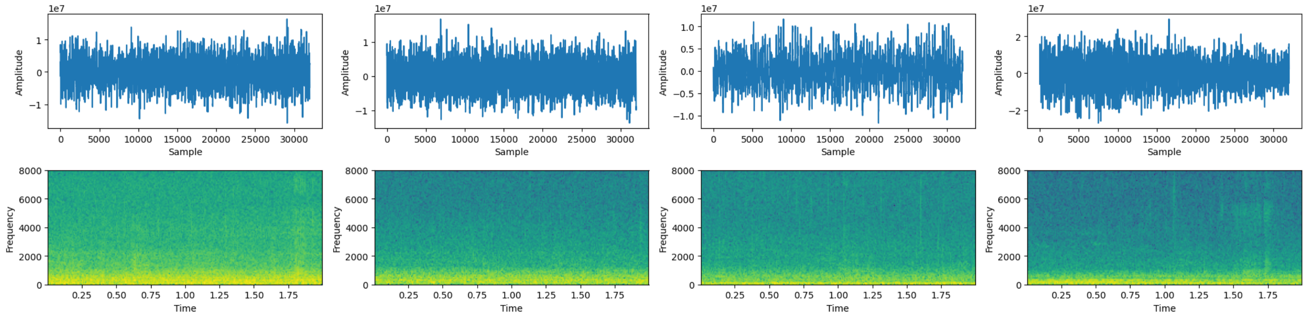
When looking at the FFT features type, it clearly shows that combining FFT with ET yields the highest accuracy result, reaching an impressive 91.33%. This achievement, although slightly below the best result of KNN and MFCCs (91.75%), showcases the efficacy of FFT when integrated with ET for the task of identifying the bee queen's presence. However, it is worth noting that SVM does not perform optimally with this particular feature type, as it only achieves an accuracy of 74.05%. On the other hand, the KNN and RF algorithms demonstrate strong performance, with both approaching an accuracy of approximately 89%.

When considering the STFT feature type, overall, the experiments show that the performance of this feature type is comparable to that of FFT, albeit with a slight reduction in accuracy. Across various machine learning algorithms, KNN consistently exhibited a commendable accuracy rate of 89.13%, this accuracy outperforms FFT combined with KNN by approximately 1.5%. However, the real strength of STFT becomes evident when combined with the ET classifier, where it gets its highest accuracy of 90.55%.

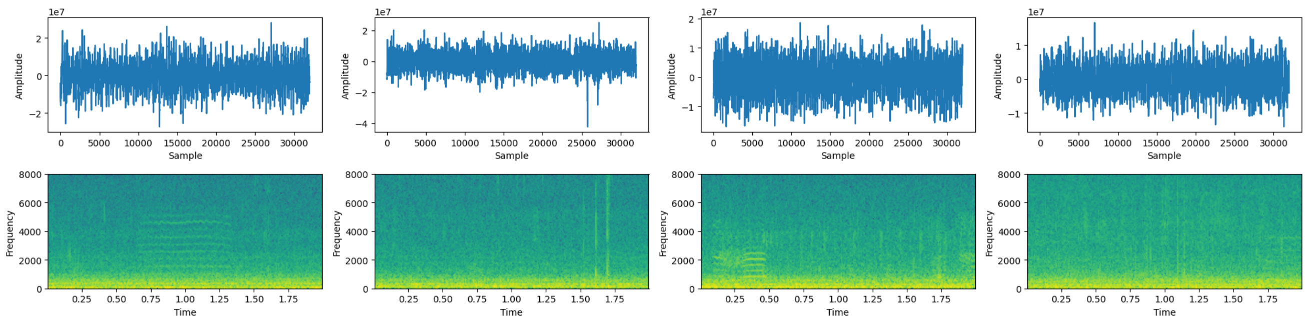
For the spectrum contrast feature type, it is noteworthy that across various machine learning algorithms, the accuracy yields using SC hovers around of 80%. However, it is important to highlight an exception, as the Logistic Regression classifier exhibits a significantly lower accuracy of just 69.02% when utilizing Spectral Contrast features.

For Chroma-based feature extraction, the findings indicate a consistent but relatively lower level of accuracy compared to some other methods. For almost all machine learning algorithms, the accuracy scores remain consistently below 80%. This means that these results are uniform, demonstrating a stable performance for Chroma-based features. The highest accuracy achieved using Chroma features was 79.52% when employing the ET classifier.

The final feature extraction method in this study is CQT, which unfortunately reaches the poorest performance among the techniques considered. For three classifiers, namely KNN, SVM, and LR, the accuracy achieved using CQT features is particularly low, approximately 50%, 53.32%, 58.07%, and 51.23%, respectively. In contrast, RF and ET obtain comparatively better accuracy of 70.15% and 70.25%, respectively, when utilizing CQT features. One potential explanation for the suboptimal performance of CQT features is that they primarily represent the tonal content of a musical audio signal in a condensed form. Given the unique characteristics and complexity of bee audio data, the tonal features captured by CQT may not be as informative or discriminative for distinguishing between different bee audio patterns.



(a)



(b)

Figure 7: Original audio samples and spectrogram of audio samples for (a) Queen and (b) Queenless.

Table 2: Performance of ML algorithms using various methods of feature extraction on testing set

Method	KNN	SVM	LR	RF	ET
FFT	87.58%	76.71%	68.86%	90.2%	91.33%
STFT	89.13%	74.05%	69.96%	89.88%	90.55%
MFCCs	91.75%	91.25%	78.32%	**92.62%**	90.7%
CQT	53.32%	58.07%	51.23%	70.15%	70.25%
Chroma	76.65%	76.88%	65.37%	78.68%	79.52%
Spectral Contrast	80.6%	81.1%	69.02%	82.3%	83.02%

In general, when considering the performance of ML methods, for most feature types, ensemble learning techniques have been shown to be more effective than other methods, except in the case of KNN with MFCCs.

In addition to comparing the classification accuracy of the ML algorithms, we also investigate the time computing of these methods using different types of features for the task of classifying the presence of the queen bee. Figure 8 provides the training time of RF using different kinds of features as input. Notably, a clear and strong correlation is evident between training time and the number of extracted features. The CQT displays the longest training time of 35.86 seconds associated with its 84 features. FFT requires the second longest training time with 24.81 seconds. Following are MFCCs, STFT, and Chroma methods respectively, with a training time of 12, 5.5, and 4.5 seconds. A notable exception is seen with spectral contrast, which completed training in a mere 2.39 seconds, attributable to its streamlined feature set of only 7.

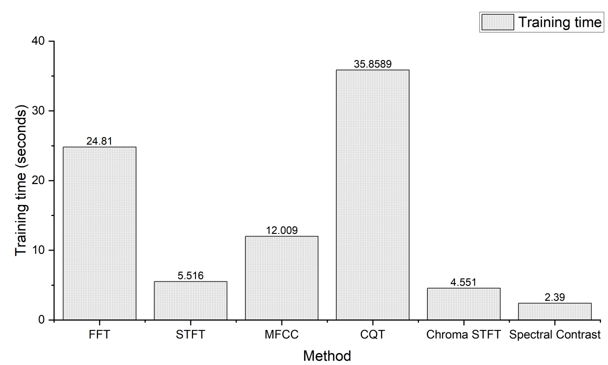


Figure 8: Training time comparison of RF using different types of features.

4 CONCLUSION

In this study, we have delved into the intricate realm of audio feature extraction, exploring six distinct methods, including FFT, STFT, MFCCs, CQT, Chroma, and SC, in conjunction with five diverse ML algorithms such as KNN, SVM, LR, RF, and ET. The aim is to reveal the optimal pairing that would yield the highest performance for the task of identifying whether the bee queen is present or missing in a beehive. Experiment results indicate superior classification results of machine learning algorithms when using features extracted by the MFCCs method. The most compelling combination distinguished by its outstanding performance, is the MFCC feature in tandem with the KNN classifier algorithm. In addition, the findings also show that with the appropriate selection of feature extraction methods combined with suitable machine learning algorithms, classification results are significantly improved. In the future, we plan to apply deep learning methods such as 1DCNN or LSTM to enhance the performance of identifying the presence of queen bees in the beehive.

ACKNOWLEDGMENTS

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] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

[2] PM Boers. 1980. Formant enhancement of speech for listeners with sensorineural hearing loss. *IPO annual progress report* 15 (1980), 21–28.

[3] T. D. Breeze, A. P. Bailey, K. G. Balcombe, and S. G. Potts. 2011. Pollination services in the UK: How important are honeybees? *Agriculture, Ecosystems & Environment* 142, 3 (Aug. 2011), 137–143.

[4] Leo Breiman. 2001. Random forests. *Machine learning* 45 (2001), 5–32.

[5] Jerry J Bromenshenk, Colin B Henderson, Robert A Seccomb, Steven D Rice, and Robert T Etter. 2009. Honey bee acoustic recording and analysis system for monitoring hive health. US Patent 7,549,907.

[6] Tymoteusz Cejrowski, Julian Szymański, Higinio Mora, and David Gil. 2018. Detection of the Bee Queen Presence Using Sound Analysis. In *Intelligent Information and Database Systems (Lecture Notes in Computer Science)*, Ngoc Thanh Nguyen, Duong Hung Hoang, Tzung-Pei Hong, Hoang Pham, and Bogdan Trawiński (Eds.). Springer International Publishing, Cham, 297–306. https://doi.org/10.1007/978-3-319-75420-8_28

[7] T. Cover and P. Hart. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13, 1 (1967), 21–27. https://doi.org/10.1109/TIT.1967.1053964

[8] Alfred DeMaris. 1995. A tutorial in logistic regression. *Journal of Marriage and the Family* (1995), 956–968.

[9] Shi Dong and Mudar Sarem. 2020. DDoS Attack Detection Method Based on Improved KNN With the Degree of DDoS Attack in Software-Defined Networks. *IEEE Access* 8 (2020), 5039–5048. https://doi.org/10.1109/ACCESS.2019.2963077

[10] Sara Ferrari, Mitchell Silva, Marcella Guarino, and Daniel Berckmans. 2008. Monitoring of swarming sounds in bee hives for early detection of the swarming period. *Computers and electronics in agriculture* 64, 1 (2008), 72–77.

[11] Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine learning* 63 (2006), 3–42.

[12] Anupama Gupta and A. A. Bazil Rai. 2019. Feature Extraction of Intra-Pulse Modulated LPI Waveforms Using STFT. In *2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*. 742–746. https://doi.org/10.1109/RTEICT46194.2019.9016799

[13] D Howard O Duran G Hunter and K Stebel. 2013. Signal processing the acoustics of honeybees (Apis Mellifera) to identify the 'Queenless' state in hives. *Proceedings of the Institute of Acoustics* 35, 1 (2013), 290.

[14] Bor-Chen Kuo, Hsin-Hua Ho, Cheng-Hsuan Li, Chih-Cheng Hung, and Jin-Shiuh Taur. 2014. A Kernel-Based Feature Selection Method for SVM With RBF Kernel for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 1 (2014), 317–326. https://doi.org/10.1109/JSTARS.2013.2262926

[15] Eirini A. Leonidaki and Nikos D. Hatziargyriou. 2006. Investigation of Decision Trees (DTs) Parameters for Power System Voltage Stability Enhancement. In *Advances in Artificial Intelligence*, Grigoris Antoniou, George Potamias, Costas Spyropoulos, and Dimitris Plexousakis (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 181–191.

[16] Meinard Müller. 2015. *Fundamentals of music processing: Audio, analysis, algorithms, applications*. Vol. 5. Springer.

[17] Inês Nolasco and Emmanouil Benetos. 2018. To bee or not to bee: Investigating machine learning approaches for beehive sound recognition. *arXiv preprint arXiv:1811.06016*.

[18] Arti Patle and Deepak Singh Chouhan. 2013. SVM kernel functions for classification. In *2013 International Conference on Advances in Technology and Engineering (ICATE)*. 1–9. https://doi.org/10.1109/ICAdTE.2013.6524743

[19] Hong Phan, Huu-Du Nguyen, and Dong Nguyen Doan. 2022. Evaluation of Feature Extraction Methods for Bee Audio Classification. In *Intelligence of Things: Technologies and Applications*. 194–203. https://doi.org/10.1007/978-3-031-15063-0_18

[20] Thi-Thu-Hong Phan, Huu-Du Nguyen, and Dong Nguyen Doan. 2022. Evaluation of feature extraction methods for bee audio classification.

[21] M. Portnoff. 1980. Time-frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 1 (1980), 55–69. https://doi.org/10.1109/TASSP.1980.1163359

[22] Ville Pulkki. 2007. Spatial sound reproduction with directional audio coding. *Journal of the Audio Engineering Society* 55, 6 (2007), 503–516.

[23] Gagan Rath and Christine Guillemot. 2003. Performance analysis and recursive syndrome decoding of DFT codes for bursty erasure recovery. *IEEE Transactions on Signal Processing* 51, 5 (2003), 1335–1350.

[24] Matthias Ring and Bjoern M. Eskofier. 2016. An approximation of the Gaussian RBF kernel for efficient classification with SVMs. *Pattern Recognition Letters* 84 (2016), 107–113. https://doi.org/10.1016/j.patrec.2016.08.013

[25] Antonio Robles-Guerrero, Tonatiuh Saucedo-Anaya, Efren gonzalez ramirez, and Jose De la Rosa. 2019. Analysis of a multiclass classification problem by Lasso Logistic Regression and Singular Value Decomposition to identify sound patterns in queenless bee colonies. *Computers and Electronics in Agriculture* 159 (2019), 69–74.

[26] Stenford Ruvinga, Gordon JA Hunter, Olga Duran, and Jean-Christophe Nebel. 2021. Use of LSTM neural networks to identify'queenlessness' in honeybee hives from audio signals. (2021).

[27] Christian Schörkhuber and Anssi Klapuri. 2010. Constant-Q transform toolbox for music processing. In *7th sound and music computing conference, Barcelona, Spain*. 3–64.

[28] Ayush Shah, Manasi Kattel, Araju Nepal, and Dichha Shrestha. 2019. Chroma Feature Extraction.

[29] Leisi Shi, Chen Li, and Lihua Tian. 2019. Music Genre Classification Based on Chroma Features and Deep Learning. In *2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP)*. 81–86. https://doi.org/10.1109/ICICIP47338.2019.9012215

[30] Nilu Singh, RA Khan, and Raj Shree. 2012. MFCC and prosodic feature extraction techniques: a comparative study. *International Journal of Computer Applications* 54, 1 (2012).

[31] Duraisamy Sundararajan. 2001. *The discrete Fourier transform: theory, algorithms and applications*. World Scientific.

[32] KS Thyagarajan and KS Thyagarajan. 2019. Fast Fourier Transform. *Introduction to Digital Signal Processing Using MATLAB with Application to Digital Communications* (2019), 385–426.

[33] Isa Servan Uzun, Abbes Amira, and Ahmed Bouridane. 2005. FPGA implementations of fast Fourier transforms for real-time signal and image processing. *IEE Proceedings-Vision, Image and Signal Processing* 152, 3 (2005), 283–296.

[34] Vladimir Vovk, Harris Papadopoulos, and Alexander Gammerman. 2015. *Measures of Complexity*. Springer.

[35] Jingyu Wang, Ke Zhang, Kurosh Madani, and Christophe Sabourin. 2013. A visualized acoustic saliency feature extraction method for environment sound signal processing. In *2013 IEEE International Conference of IEEE Region 10 (TENCON 2013)*. 1–4. https://doi.org/10.1109/TENCON.2013.6718918