Archives of Acoustics Online First May 28, 2024 doi: 10.24425/aoa.2024.148796

Research Paper

Drone Flight Detection at an Entrance to a Beehive Based on Audio Signals

Urszula LIBAL*, Pawel BIERNACKI

Faculty of Electronics, Photonics and Microsystems
Department of Acoustics, Multimedia and Signal Processing
Wroclaw University of Science and Technology
Wrocław, Poland; e-mail: pawel.biernacki@pwr.edu.pl

*Corresponding Author e-mail: urszula.libal@pwr.edu.pl

(*received June 28, 2023; accepted March 5, 2024; published online May 28, 2024*)

Spotting a significant number of drones flying near the entrance of a beehive during late Spring could indicate the occurrence of swarming mood, as the the surge in drone presence is related to an overcrowded hive. Swarming refers to a natural reproductive process witnessed in honey bees, wherein half of the bee colony departs from their hive alongside the aging queen. In the paper, we propose an early swarming detection mechanism that relies on the behavior of the drones. The proposed method is based on audio signals registered in a close proximity to the beehive entrance. A comparative study was performed to find the most effective preprocessing method for the audio signals for the detection problem. We have compared the results for three different power spectrum density coefficients estimation methods, which are used as an input of an autoencoder neural network to discriminate drones from worker bees. Through simulations employing real-life signals, it has been demonstrated that drone detection based solely on audio signals is indeed feasible. The attained level of detection accuracy enables the creation of an efficient alarm system for beekeepers.

Keywords: signal processing; machine learning; neural networks; anomaly detection; autoencoder; honey bee swarming; drone detection.

1. Introduction

Swarming is a natural phenomenon that occurs when a honey bee colony reproduces and divides into multiple colonies. Swarming typically occurs during the late spring and early summer months (Wright, 1913; Ostrowska, 1980). Early detection of swarming (Zgank, 2011; Hadjur *et al.*, 2022) in honey bees is essential for swarming prevention, colony health monitoring, queen management, swarm capture, and effective population management. It allows beekeepers to take timely actions to maintain healthy colonies, prevent the loss of bees, and optimize honey production.

There are several ways to predict when honey bees will swarm. Most of them require interference in the hive. Here are a few:
– queen cells present,
– the old age of the queen,
– a hive becomes too crowded,
– increased foraging activity, more drones coming and going.

Thus, the detection of swarming without interfering in the hive can be based on the detection of drones' activity around the hive. This can be accomplished by analyzing the sounds around the hive and identifying drones.

Honey bees use sound as means of communication, both within the hive and with other bees outside the hive. The sound produced by bees is a form of vibration created by the rapid beating of their wings and is used to convey information about the location of food, the presence of danger, and other important information. One of the most well-known sounds produced by bees is the buzzing sound that is heard when they are in flight. This sound is created by the rapid beating of their wings, which can occur at a rate of up to 200 beats per second. The frequency of the buzzing sound can vary depending on the size and species of the bee (Kawakita, Ichikawa, 2019). The typical range of frequencies generated by different bees is piping signal in the range: 100 Hz–500 Hz (Seeley, Tautz, 2001), with a fundamental frequency of 384 Hz (Sarma *et al.*, 2002), tooting: 400 Hz–500 Hz, and quacking: around 350 Hz.

Distinguishing between bees and drones, based on the sounds they make, can be done using the fact that the drones are bigger, have bigger wings. The results showed that body shape or wing size can be correlated with fundamental frequency (Gradišek *et al.*, 2017), and the duration of the buzzes has also been shown to differ with body size (larger bees producing shorter buzzes). Moreover, using the amplitude, frequency, and duration of flight, one can distinguish between bees and drones by analyzing the frequency spectrum of their sounds (Ribeiro *et al.*, 2021).

This paper is divided into the following sections. Section 1 briefly provides information why distinguishing between worker bees and drones is important to control beehive environment and its health. Section 2 presents related works on the bee sound detection and classification. In Sec. 3 the methodology of the proposed solution is discussed. We focus on feature extraction and the implementation of machine learning techniques. Section 4 discusses the identification results based on the collected datasets. The final conclusions are in Sec. 5.

2. Related work

2.1. Audio analysis methods

Mel-frequency cepstral coefficients (MFCCs) (Davis, Mermelstein, 1980; Soares *et al.*, 2022; Libal, Biernacki, 2024) are the most common set of features used in studies that exploited a machine learning framework. Many studies analyzed MFCCs to extract information for bee detection, queen absence and swarming detection, and bee species identification, as well as environmental effects, with the three first coefficients showing the greatest discrimination. Peng *et al.* (2020) used the so-called improved MFCC (IMFCC) proposed in (Yegnanarayana *et al.*, 2005) to capture additional information from the higher-frequency part of the spectrum that is typically ignored by traditional MFCC. This has been shown to improve the classification accuracy for queenless hive detection tasks. In (Zlatkova *et al.*, 2020), the short-time Fourier transform (STFT) calculated with filter banks and the overlapping method was used to detect swarming events. The STFT has been calculated using 128, 256, 512, and 1024 bins to investigate the impact of window width. In the study (Gourisaria *et al.*, 2024), the MFCC approach was compared with the STFT.

2.2. Machine learning algorithms

A typical machine learning framework encompasses signal measurement, preprocessing, feature extraction, and finally classification. In the area of acoustic analysis of bee sounds, many different classifiers have been explored. The most common classifier is a support vector machine (SVM) (Cejrowski *et al.*, 2020; Nolasco *et al.*, 2018), a kernel-based method that projects data into higher dimensions in which a hyperplane can separate the classes.

More recent deep learning neural network-based methods are being introduced. In (Ruvinga *et al.*, 2021; Kulyukin *et al.*, 2018) the use of so-called long short-term memory (LSTM) recurrent neural networks (RNN) for the queen bee presence detection is proposed. A comparison between LSTM, a multi-layer perceptron (MLP) neural network, and logistic regression was made, and it showed the power of the LSTM for the task at hand. Recently, convolutional neural networks (CNN) (Nolasco *et al.*, 2018; Kim *et al.*, 2021), have gained popularity, especially within computer vision tasks. To make them directly applicable to the bee acoustics analysis, researchers have relied on image-like inputs, such as spectrograms, mel-scaled spectrograms, or other two-dimensional time-frequency representations of the audio signals.

3. Methods

Presented in the article results based on a selected set of audio recordings acquired in the context of the beehive monitoring system capable of identifying and predicting certain events and states of the hive that are of interest to the beekeeper. All recordings are sampled with frequency of 44 100 Hz and saved in WAV format without any compression. For signal processing the recordings were divided into 1 second long samples. The data set used for simulations consists of 10 000 samples of bee flight sound and around 1700 samples of a flying drone sound. To record the audio samples, we used a directional microphone mounted on top of the hive and aimed at its entrance.

The whole detection process is divided into two parts: feature extraction and classification by autoencoder neural network. The signal processing flow chart is shown in Fig. 1.

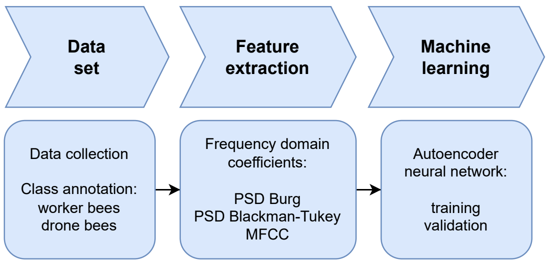

Fig. 1. Signal processing flow chart.

3.1. Power spectrum density coefficients

Power spectrum coefficients are the type of features that can be extracted from an audio signal to characterize its spectral content. They are calculated by taking the squared magnitude of the Fourier transform of the signal, which represents the power or energy content of the signal at each frequency component. The power spectrum coefficients can then be used as a feature vector to identify or classify different types of audio signals. We decided to employ power spectrum density (PSD) coefficients as features used in the detection phase of the identification.

Power spectral density estimation techniques can be divided into parametric and nonparametric methods. The non-parametric methods estimate PSD explicitly from signal samples, without any assumptions about a particular process structure. Parametric approaches assume that the signal can be described as the stationary process (MA – moving average, AR – autoregressive, or ARMA – autoregressive moving average) of the order m. The power spectral density is then calculated using estimated model parameters. This paper presents PSD estimated with the parametric approaches (the Burg method) and nonparametric methods (the Blackman-Tukey method).

3.1.1. Burg algorithm

The Burg algorithm (Kay, 1988; Orfanidis, 1995) assumes that a signal can be described as an autoregressive (AR) process of the order m:

$$\widehat{x} = -\sum_{k=1}^{m} a_m(k)x(n-k). \tag{1}$$

The Burg algorithm solves the ordinary least squares problem. AR parameters a_m are estimated by minimizing the prediction forward and backward errors which are referred to as the error between the actual value signal and the corresponding estimators in forward and backward:

$$\mathrm{PSD}_x^{\mathrm{Burg}}(f) = \frac{E_m}{\left|1+\sum_{k=1}^{m} a_m(k)e^{-j2\pi fk}\right|^2}. \tag{2}$$

3.1.2. Blackman-Tukey method

The Blackman-Tukey (Blackman, Tukey, 1958; Cooley, 1997) power spectrum estimate is calculated with the use of the fast Fourier transform (FFT) in the following way:

$$\mathrm{PSD}_x^{B-T}(f) = |\mathrm{FFT}\{w(n) * R(n)\}|, \tag{3}$$

where $w(n)$ is a window, $R(n)$ is an autocorrelation of the input signal $x(n)$.

The signal processing scheme for the Blackman-Tukey estimation method of PSD is shown in Fig. 2.

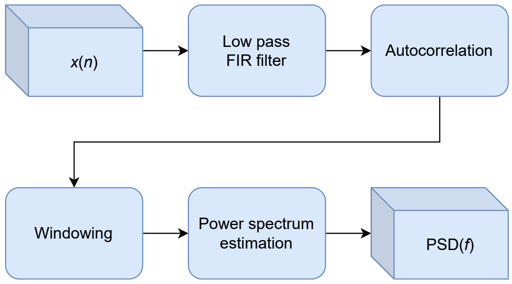

Fig. 2. Schema of Blackman-Tukey method of power spectrum estimation.

The lowpass FIR filter is used to adjust the bandwidth of the signal to investigate its influence on identification. Filter coefficients were changed to obtain the desired filter characteristic. Power spectrum estimation requires the Fourier transform calculation. To minimize leakage of spectrum a windowing procedure was implemented. Different windows were investigated (Hanning, Hamming, Kaiser).

During trait extraction, we noticed some differences in PSD coefficients to distinguish bees from drones. It can be observed in Fig. 3.

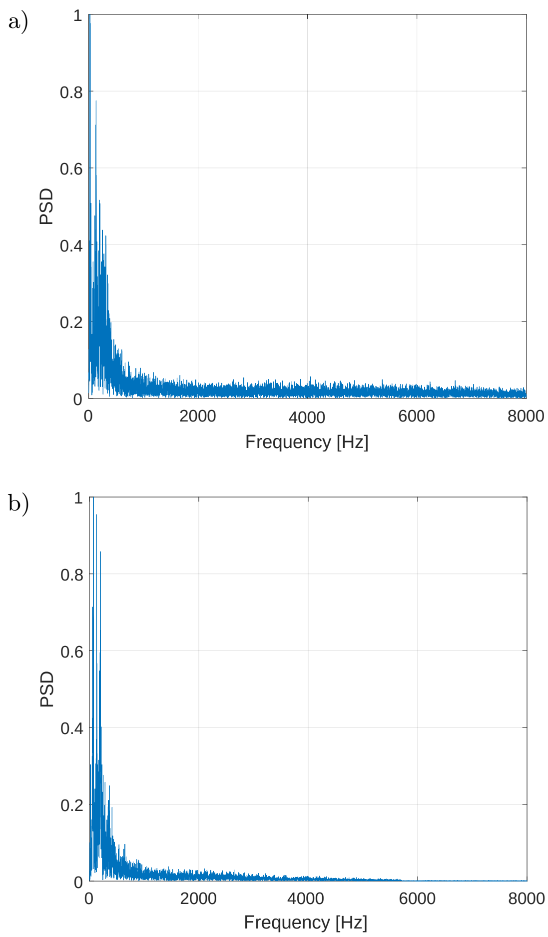

Fig. 3. PSD: a) worker bees; b) drones.

3.2. Mel-frequency cepstral coefficient

The motivating idea of MFCCs is to compress information about the vocal tract (smoothed spectrum) into a small number of coefficients based on an understanding of the cochlea in the ear. The basic steps to calculate MFCC are shown in Fig. 4.

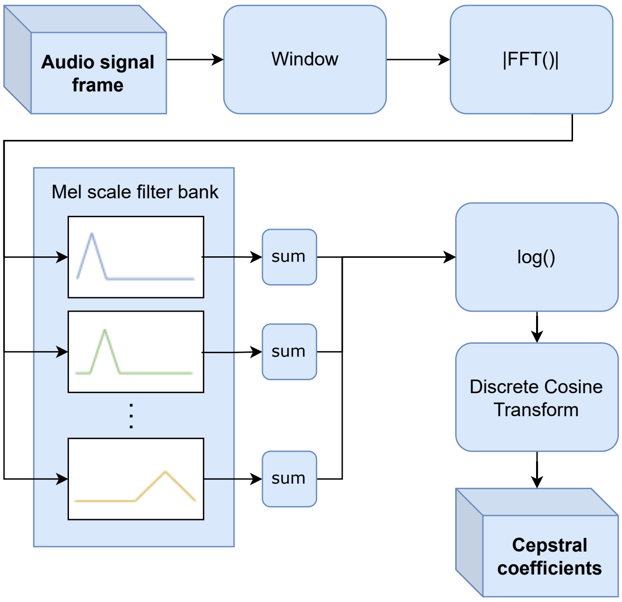

Fig. 4. MFCC calculation diagram.

The extraction of cepstral coefficients allowed for noticing some differences in MFCCs distinguishing worker bees from drones. It can be observed in Fig. 5 and in the calculated difference in Fig. 6.

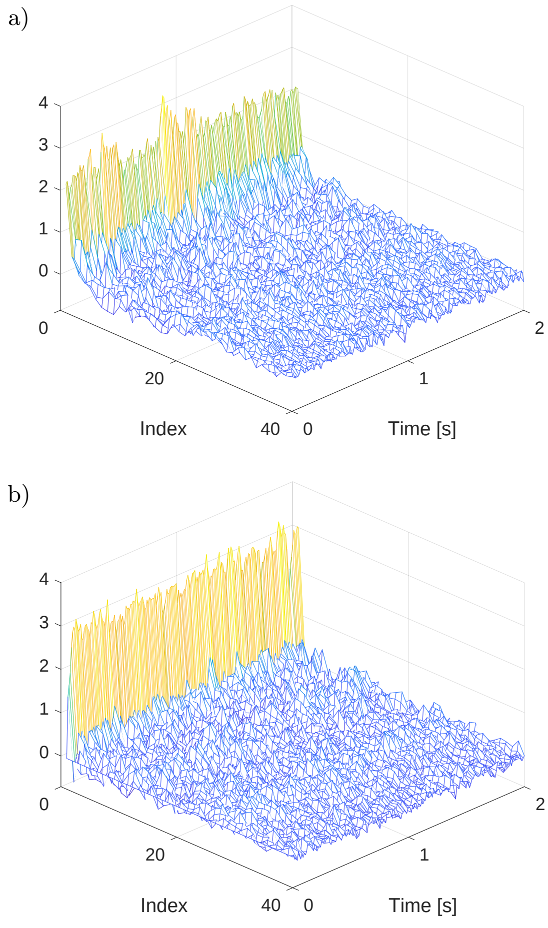

Fig. 5. MFCCs: a) worker bees; b) drones.

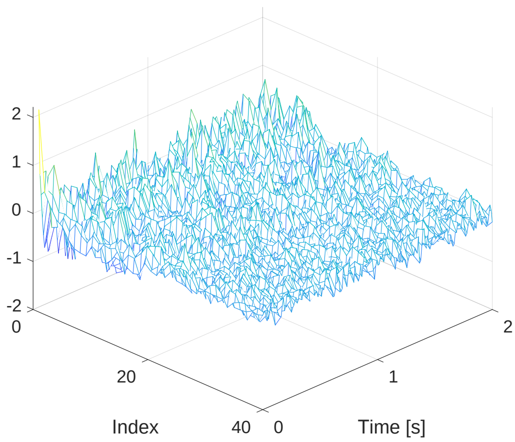

Fig. 6. Difference between MFCCs for worker bees and drones.

3.3. Autoencoder

An autoencoder neural network is a type of artificial neural network that is used for unsupervised learning of efficient data representations. The network consists of an encoder and a decoder, where the encoder maps the input data to a compressed representation (Hinton, Salakhutdinov, 2006), and the decoder maps the compressed representation back to the original data. The objective of the autoencoder is to minimize the difference between the input and output data, while also enforcing a constraint on the dimensionality of the compressed representation.

In audio signal identification, autoencoder neural networks can be used to learn compact representations of audio signals that capture their essential features. This can be useful for tasks such as pattern recognition, classification, identification, anomaly detection or noise reduction. Autoencoder neural networks have been used in a variety of applications in speech recognition, speaker identification or music genre classification. By training an autoencoder on a large dataset of audio signals, the network can learn to extract features that are relevant to the task at hand, while also discarding noise and irrelevant information. They have also been combined with other types of neural networks, such as CNNs or RNNs, to improve performance on more complex tasks.

One common approach for using autoencoders in audio signal identification is to train the network on a reconstruction task, where the input is an audio signal and the output is the reconstructed audio signal. The loss function used during training is typically a measure of the difference between the input and output signals, such as mean squared error (MSE) or mean absolute error. The mean square error is a reconstruction loss of the output produced by the network, obtained for a particular input vector after encoding and decoding stages. Once the network is trained, the compressed representation learned by the encoder can be used as a feature vector for identifying or classifying different types of audio signals. The scheme of such a network is shown in Fig. 7.

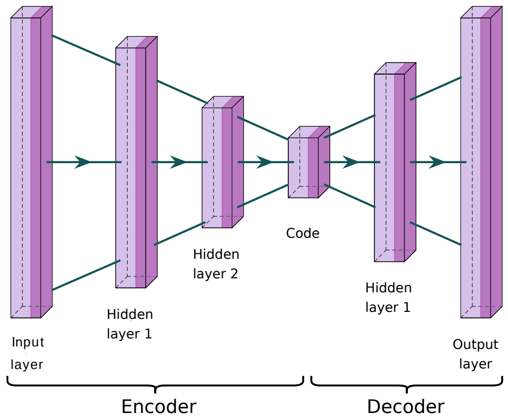

Fig. 7. General schema of autoencoder neural network.

In our simulations we have used neural networks with 1, 2, and 3 activation (ReLu) layers for the encoder. The decoder had always 2 layers: one with ReLu and one with a sigmoid activation function. In the case of power spectrum estimation based on the Burg nad Blackman-Tukey method, we decided for the following neural network settings. For the neural network with 3 activation layers, the number of features for encoder in layer 1 was 64, layer 2 – 32, layer 3 – 16. For the neural network with 2 activation layers, the number of features in layer 1 was 64, and layer 2 – 32. And for the neural network with only 1 activation layer, the number of features in layer 1 was 64. In the case of the usage of the MFCC, we decided to apply only the autoencoder neural network with only 1 activation layer, because the number of cepstral coefficients was relatively small, from 10 to 35 only.

Worker bees are present in the beehive the whole year, while no drones survive the Winter. This is the reason why the detection of a drone can be treated as an anomaly, which occurs most often in the time of the year preceding swarming. The process of anomaly detection using an autoencoder is divided into the following main steps:

- Step 1. Training: in the first step, the autoencoder is trained on the flight sounds of worker bees only.
- Step 2. Testing: in the second step, it is used for a test reconstruction of recordings from both classes: worker bees and drones. Our hypothesis is that the abnormal signals (sounds of drones) will have a higher reconstruction error.
- Step 3. Classification: the last step is the detection of drone signals as anomalies if the reconstruction errors surpass a fixed threshold.

It is worth mentioning that the system takes into account that there are many more worker bees flying in and out the hive in the spring time than drones. The worker bees are extremely busy collecting nectar and pollen, and they generate huge traffic when flying to the hive. Fortunately, a special property of the autoencoder neural network prevents undetectability of less numerous drones, that could be the case for other classifiers. Autoencoder trained on the set containing only worker bees flight recordings generates much larger MSE for recordings of drones. The autoencoder neural network is a type of generative networks, and the reconstruction loss (after encoding and decoding stages) informs of the quality of recreation of the input by the network. An anomaly given as input to such a network, produces a higher loss value. The standard input gives a minimal loss value, related to a deviation between the input audio frames in the training set.

In Fig. 8, we present an exemplary histogram of the MSE in a series of numerical experiments returned by the autoencoder neural network for training (worker bees) and testing (drones) data sets. The vectors of coefficients representing recordings of drone flights produce a higher MSE because they deviate significantly from the signatures of worker bees, in the sense of the frequency components. The threshold is marked by the red dotted line.

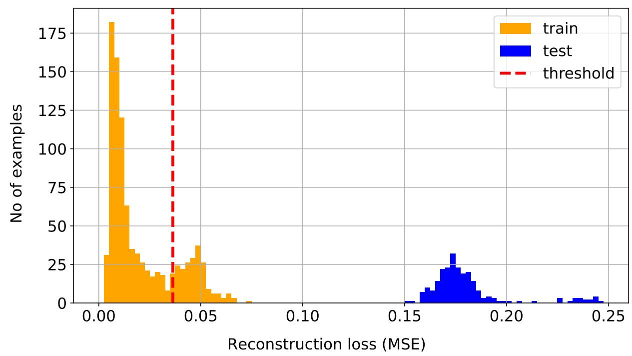

Fig. 8. Exemplary histogram of the MSE loss produced by autoencoder neural network for training and testing data sets.

As we see in Fig. 8, the MSE generated by drones is, on average, significantly greater than the MSE generated for worker bees. In the presented case, both histograms of the MSE are well separated in terms of separability of the corresponding probability densities. The proposed system of detecting drones is in fact detecting only a higher frequency of drones' flights since the system does not count particular individuals but only classifies audio frames of 1 second length. In this paper, we present results of classification recordings of bees into two classes based on a cleared and tagged (unquestionably by worker bees or drone labels) data set. It is absolutely possible that in practice we will encounter numerous situations of huge traffic of bees next to beehive entrance when many worker bees and drones can be present while recording an audio. This problem should be investigated in future research, and we hope that the special property of the autoencoder (giving the reconstruction loss as output) can also make it possible to detect a drone in the presence of many worker bees.

4. Results

The experiment was performed for an autoencoder neural networks with 1, 2, and 3 activation layers, and for three preprocessing methods resulting in three different signal representations in the frequency domain by the following estimates:

- the Burg parametric method of power spectral estimate,
- the Blackman-Tukey nonparametric method of power spectral estimate,
- MFCC calculation.

Two first two methods, Burg and Blackman-Tukey methods, operate on a linear scale in the frequency domain. The third method, MFCC, uses a logarithmic scale. The goal was to investigate whether any approach will present a higher recognition results, considering that the learning process is done by neural networks and the calculated spectral coefficients are not analysed directly, but consist an input of the autoencoder neural networks.

For linear frequency scale, we cut the frequency bandwidth with the step 100 Hz in the range from 100 Hz to 3000 Hz. For the Mel-frequency spectrum scale, we have chosen 10, 15, 20, 25, 30 or 35 cepstral coefficients. The comparison study should determine the most effective method for estimating spectral coefficients, but also the number of cepstral coefficients or frequency bandwidth, depending on the chosen method.

4.1. Statistical evaluation

Our database contained significantly less drone recordings in comparison to the huge number of worker bee recordings. Due to unbalanced data, it is important not only to focus on classification accuracy, but also on other result parameters such as recall and $F1$-score.

The accuracy for the binary classification problem is the proportion of correct predictions, both true positives (TP – number of correct detections) and true negatives (TN – number of correct rejections), to the total number of cases examined:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{4}$$

where FP is false positive, also called false alarm, and FN is false negative.

In addition to accuracy, a recall value was calculated, meaning a sensitivity of the detection test:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{5}$$

We analyzed also a $F1$-score, which is defined as:

$$F1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \tag{6}$$

The $F1$-score takes values in the range $[0, 1]$. The highest possible value of $F1$-score (equal 1) indicates perfect precision and recall of the detection method.

In the series of experiments for different sets of settings, the above parameters: accuracy, recall and $F1$-score are going to indicate the best method for the considered drone detection problem. The ideal method, with zero incorrect classifications for both worker bees and drones, would have all three parameters equal 1. This is, of course, an unrealistic expectation, but the method which obtains the results closest to the value 1, will be considered the best.

For the better understanding of the obtained results, we also apply a weighted confusion matrix in the form presented in Fig. 9. In standard confusion matrix, there are simple counts: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), showing numbers of cases classified correctly or wrongly to both classes. Due to highly unbalanced data set, with many more worker bee audio recordings (10 000 samples) compared to only 1700 drone samples, the presented values were accordingly divided and presented in percent points. That way of presenting results, shows the actual percentage of correctly classified or misclassified worker bees and drones.

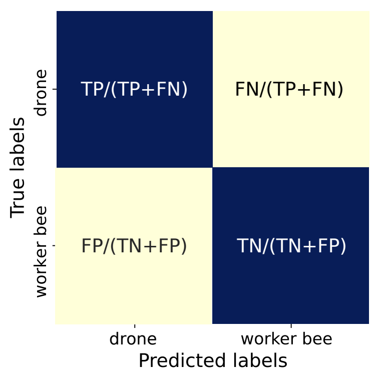

Fig. 9. Weighted confusion matrix.

4.2. Drone detection results

The classification of honey bees into two classes: worker bees and drones, based on audio recordings of the sound generated during their flight was performed for a data set of 10 000 audio samples for workers and 1700 for drones. The training of the autoencoder was performed with the use of 80 % of the honey bee flight recordings. The rest (20 %, 2000 samples) was used for testing, together with all records of drone flights (1700 samples). The experiment was carried out in 100 iterations. In each iteration, the autoencoder was trained and tested. The training stage gives full calculation of all parameters of the neural network. At this stage also a threshold dividing standard class (worker bees) and anomaly (drones) was derived from the formula:

$$\text{threshold} = \text{mean}(\text{MSE}_{\text{train}}) + \text{std}(\text{MSE}_{\text{train}}), \qquad (7)$$

where mean is a mean value and std is a standard deviation of mean square errors, which are estimated using only outputs of the worker bee recordings during a training stage. In that way, the autoencoder was trained and ready for testing. The cases for which the mean square error was smaller than the calculated threshold were classified to the worker bees class and the cases for which was bigger to the drone class:

$$\text{class}(x) = \begin{cases} \text{worker class}, & \text{if } \text{MSE}_{\text{test}}(x) \leq \text{threshold}, \\ \text{drone class}, & \text{if } \text{MSE}_{\text{test}}(x) > \text{threshold}. \end{cases} \qquad (8)$$

In the end all cases were compared with the true labels, which led to obtaining statistical indicators presented in this section.

The resulting accuracy for the Burg and Blackman-Tukey methods is shown in Fig. 10. The recall values for the methods are placed in Fig. 11. For five of the six methods: NN_2_Burg, NN_3_Burg, NN_1_B_T, NN_2_B_T, and NN_3_B_T, the obtained values of accuracy and recall, after reaching a certain bandwidth, remain in almost constant value ranges, different for each specific method. Surprisingly, the only exception is a neural network with one activation layer for input in the form of Burg estimates

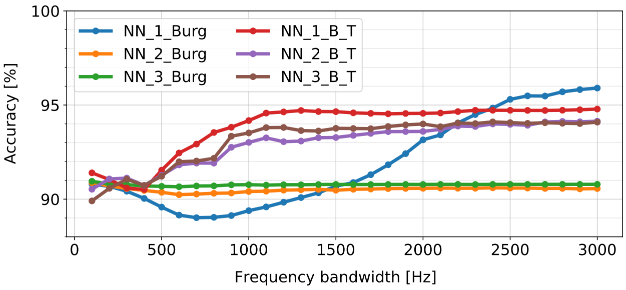

Fig. 10. Accuracy for neural networks with 1, 2, and 3 activation layers, and power spectrum estimation by Burg and Blackman-Tukey method.

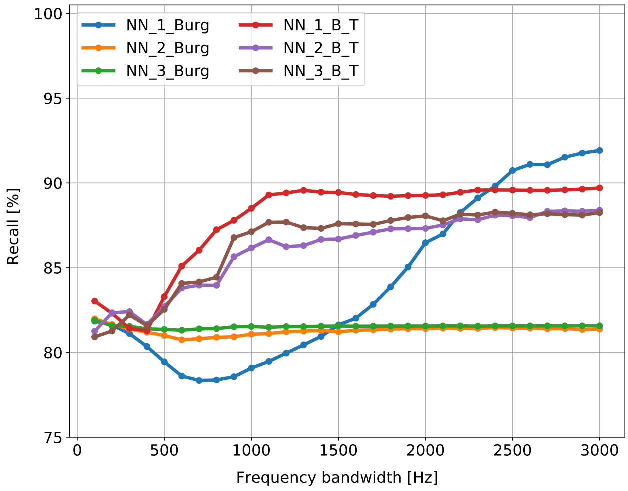

Fig. 11. Recall for neural networks with 1, 2, and 3 activation layers, and power spectrum estimation by Burg and Blackman-Tukey method.

(NN_1_Burg), for which the accuracy and recall values increase significantly as the frequency band is extended. The method has reached the highest evaluation factor values from all analyzed methods for the frequency bandwidth 3000 Hz.

The results for the MFCC are presented in Fig. 12. Both accuracy and recall values are presented as a function of the number of cepstral coefficients on the Mel-frequency scale. The experiment was carried out only for a neural network with 1 activation layer due to the small number of input coefficients, it is 10, 15, 20, 25, 30 or 35 MFCCs. It turns out that for this method (NN_1_Mel), the highest accuracy and recall are reached for only 10 cepstral coefficients and for the 15 and more coefficients they fall slightly.

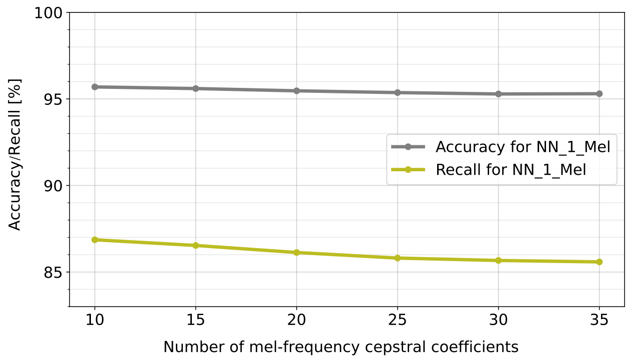

Fig. 12. Accuracy and recall for MFCC as input of neural network with 1 activation layer.

In Figs. 13a–c, we present the confusion matrices of the best results for all the three preprocessing methods: MFCC (NN_1_Mel 10 MFCC), Blackman-Tukey (NN_1_B_T 3000 Hz), and Burg (NN_1_Burg 3000 Hz), respectively. The results were normalized by dividing by the size of the drone class in the up-

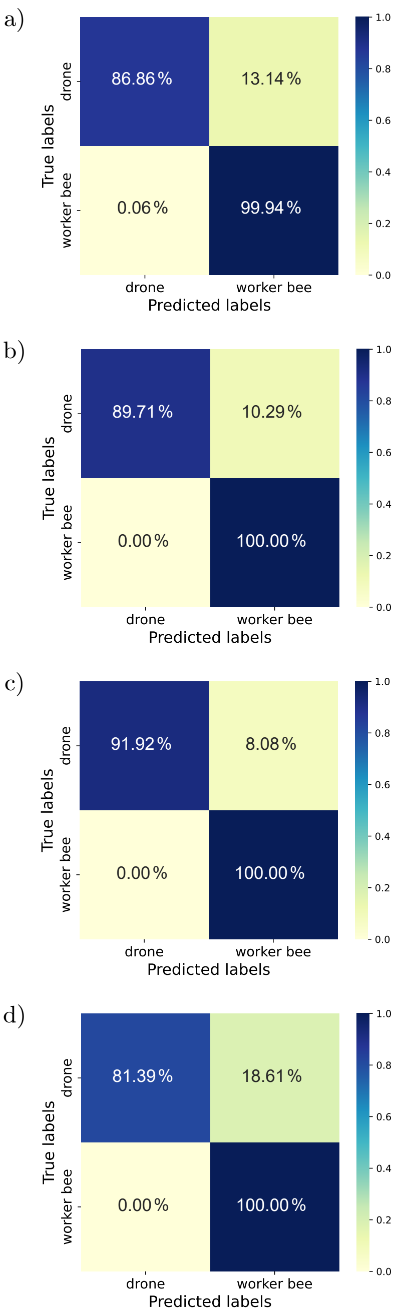

Fig. 13. Weighted confusion matrices, presented in percent points: a) NN_1_Mel 10 MFCC; b) NN_1_B_T 3000 Hz; c) NN_1_Burg 3000 Hz; d) NN_2_Burg 3000 Hz.

per row of the confusion matrix and by the size of the worker bee class in the lower row. The worst result for the same 3000 Hz bandwidth was obtained for the Burg method and the neural network with 2 activation layers (NN_2_Burg 3000 Hz). We place the confusion matrix for that case in Fig. 13d.

In all cases with preprocessing based on Blackman-Tukey or Burg methods, the true negative rate (TNR) is 100 % and the false positive rate (FPR) is 0 %, which means that all signals from the worker bee class were correctly labeled as worker bees by neural networks. On the other hand, for all cases using MFCC, the false positive rate has values higher than zero, meaning that the probability of a false alarm is also non-zero. Our method allows for an early detection of the swarming mood of honey bees, thanks to an analysis of the number of occurrences of drones flying in close proximity to the entrance of a beehive. Drones represent at most around 15 % of the population in beehives at the top, that is, during the late spring, and the worker bees are significantly more numerous at the same time – around 85 % of the population. That is why a higher false alarm rate can lead to many more false classifications of worker bees to the drone class. And as a consequence, the system would start the swarming alarm the whole year, except in winter, when worker bees are not active and stay inside the beehive.

4.3. The best results

The best results, taking into account the highest values of the three statistical factors: the accuracy of drone recognition, the recall value (sensitivity of the method) and the $F1$-score, were obtained for the Burg power spectrum estimate method with the widest considered frequency band of 3000 Hz and neural network with one activation layer (marked as NN_1_Burg 3000 Hz – see Table 1). Accuracy reached 95.90 %, recall 91.92 %, and $F1$-score 96.11 %. The same method with a slightly narrower frequency band (NN_1_Burg 2900 Hz and NN_1_Burg 2800 Hz) achieved comparatively excellent results. Next in order, methods based on MFCC and Mel-frequency scale (such as, e.g., NN_1_Mel 10 MFCC) have given the accuracy at a very similar level, but recall recorded a decrease of around 5 % in all cases. Similarly, $F1$-score for MFCC method drops for more than 2 % compared to the Burg method.

However, it is worth noting that the unrivaled Burg method requires recordings with a bandwidth of 2500 Hz–3000 Hz to achieve the high classification results – see Table 1. If this would be a hardware limitation for some reasons, then it is better to use the

Table 1. Ten the best accuracy results and the corresponding recall and $F1$-score values.

	Method	Accuracy	Recall	$F1$-score
1	NN_1_Burg 3000 Hz	0.959010	0.919165	0.961152
2	NN_1_Burg 2900 Hz	0.958210	0.917626	0.960442
3	NN_1_Burg 2800 Hz	0.957031	0.915206	0.959327
4	NN_1_Mel 10 MFCC	0.956934	0.868622	0.938068
5	NN_1_Mel 15 MFCC	0.955990	0.865345	0.936714
6	NN_1_Burg 2600 Hz	0.954863	0.910953	0.957374
7	NN_1_Burg 2700 Hz	0.954778	0.910762	0.957287
8	NN_1_Mel 20 MFCC	0.954680	0.861279	0.934943
9	NN_1_Mel 25 MFCC	0.953653	0.858056	0.933549
10	NN_1_Burg 2500 Hz	0.953010	0.907349	0.955725

MFCC-based method, at the cost of lowering the recall parameter and $F1$-score, and thus making more incorrect drone detections.

5. Conclusion

In this article, we have investigated the possibility of building an early swarming alert system for beekeepers, based on the detection of a larger number of drones flying at the entrance to a beehive. The system applies neural networks of autoencoder type, which must be previously trained on the basis of a signal database, containing worker bees and drones flight sound recordings, preferably registered in the environment where the system will be installed.

The preliminary study focused on finding the best signal processing methods and settings for the assumed task. We have compared three signal preprocessing methods, producing frequency-domain coefficients representing the recorded signals. They are: the Burg parametric method, the Blackman-Tukey nonparametric method, and the MFCC method. The power spectral or cepstral coefficients were the input of the autoencoder neural network. The detection was performed by the three settings of the encoder-decoder neural network pairs: with various (1, 2 or 3) number of activation layers for the encoder and with fixed 2 activation layers for the decoder.

The results obtained show that the configuration of the autoencoder neural network with only 1 activation layer has given the highest accuracy and recall values for all preprocessing methods. The best results have been received for the Burg parametric method of power spectrum estimation in a linear frequency scale and the frequency bandwidth of 3000 Hz (NN_1_Burg 3000 Hz). The accuracy of drone detection is 95.90 %, the recall (sensitivity) – 91.92 %, $F1$-score – 96.12 %, and false alarm rate equals 0.00 %. Cutting the bandwidth (to 2900 Hz, 2800 Hz, ..., and so on) has gradually decreased the accuracy of the drone detection.

The method using MFCC and the mel-frequency scale was found to give slightly worse results than the Burg preprocessing method with accuracy 95.69 %, the recall (sensitivity) – 86.86 %, $F1$-score – 93.81 %, and false alarm rate equals 0.06 % (for the case NN_1_Mel 10 MFCC). The accuracy level stays close to the accuracy for the best Burg method, but recall drops for more than 5 %, $F1$-score for more than 2 %, and the false alarm rate increases. The probability of a false alarm for the MFCC method is non-zero in all investigated cases, meaning that worker bees can be classified by the neural network as drones. Contrary to the Burg and Blackman-Tukey preprocessing methods, for which the probability of a false alarm is always zero.

Considering that the aim of the proposed method is an early detection of swarming mood of honey bees, based on more frequent observations of drones close to a beehive entrance, it is important that the worker bees, which are more numerous in a swarm (around 85 % in late spring) than the drones (around 15 % in late spring), are not mistaken with the drones. This would increase the drone detection rate and falsely alarm beekeepers of a possible start of the swarming mood. In the future, a further study on the behavior of drone bees should be conducted. In particular, the correlation between the frequency of drone observation in relation to other bees and the state of the swarm should be investigated.

Acknowledgments

The authors would like to express their great appreciation to Dr. Bartłomiej Golenko from the Wroclaw University of Science and Technology for his valuable and constructive suggestions during the planning and development of this research work.

References

1. Blackman R.B., Tukey J.W. (1958), The measurement of power spectra from the point of view of communications engineering – Part I, *Bell System Technical Journal*, **37**(1): 185–282, doi: 10.1002/j.1538-7305.1958.tb03874.x.
2. Cejrowski T., Szymański J., 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.
3. Cooley J.W. (1997), The re-discovery of the fast Fourier transform algorithm, *Microchimica Acta*, **3**: 33–45, doi: 10.1007/BF01201681.
4. Davis S., Mermelstein P. (1980), Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, [in:] *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **28**(4): 357–366, doi: 10.1109/TASSP.1980.1163420.
5. Gourisaria M.K., Agrawal R., Sahni M., Singh P.K. (2024), Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques, *Discover Internet of Things*, **4**: 1, doi: 10.1007/s43926-023-00049-y.
6. Gradišek A., Slapničar G., Sorn J., Luštrek M., Gams M., Grad J. (2017), Predicting species identity of bumblebees through analysis of flight buzzing sounds, *Bioacoustics*, **26**(1): 63–76, doi: 10.1080/09524622.2016.1190946.
7. Hadjur H., Ammar D., Lefèvre L. (2022), Toward an intelligent and efficient beehive: A survey of precision beekeeping systems and services, *Computers and Electronics in Agriculture*, **192**: 106604, doi: 10.1016/j.compag.2021.106604.
8. Hinton G.E., Salakhutdinov R.R. (2006), Reducing the dimensionality of data with neural networks,

Science, **313**(5786): 504–507, doi: 10.1126/science.1127647.

9. Kawakita S., Ichikawa K. (2019), Automated classification of bees and hornet using acoustic analysis of their flight sounds, *Apidologie*, **50**: 71–79, doi: 10.1007/s13592-018-0619-6.

10. Kay S.M. (1988), *Modern Spectral Estimation: Theory and Application*, Englewood Cliffs, Prentice Hall.

11. Kim J., Oh J., Heo T.-Y. (2021), Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and Grad-CAM, *Mathematical Problems in Engineering* **2021**: 5594498, doi: 10.1155/2021/5594498.

12. Kulyukin V., Mukherjee S., Amlathe P. (2018), Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples, *Applied Sciences*, **8**(9): 1573, doi: 10.3390/app8091573.

13. Libal U., Biernacki P. (2024), MFCC selection by LASSO for honey bee classification, *Applied Sciences*, **14**(2): 913, doi: 10.3390/app14020913.

14. Nolasco I., Terenzi A., Cecchi S., Orcioni S., Bear H.L., Benetos E. (2018), Audio-based identification of beehive states, [in:] *ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 8256–8260, doi: 10.1109/ICASSP.2019.8682981.

15. Orfanidis S.J. (1995), *Introduction to Signal Processing*, Prentice-Hall.

16. Ostrowska W. (1980), *Beekeeping* [in Polish: *Gospodarka Pasieczna*], Państwowe Wydawnictwo Rolnicze i Leśne, Warszawa.

17. Peng R., Ardekani I., Sharifzadeh H. (2020), An acoustic signal processing system for identification of queen-less beehives, [in:] *Proceedings of the 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (*APSIPA ASC*), pp. 57–63.

18. Ribeiro A.P., da Silva N.F.F., Mesquita F.N., Araújo P.D.C.S., Rosa T.C., Mesquita-Neto J.N. (2021), Machine learning approach for automatic recognition of tomato-pollinating bees based on their buzzing-sounds, *PLOS Computational Biology*, **17**(9): e1009426, doi: 10.1371/journal.pcbi.1009426.

19. Ruvinga S., Hunter G.J., Duran O., Nebel J. (2021), 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, doi: 10.1109/IE51775.2021.9486575.

20. Sarma M.S., Fuchs S., Werber C., Tautz J. (2002), Worker piping triggers hissing for coordinated colony defence in the dwarf honeybee Apis florea, *Zoology*, **105**(3): 215–223, doi: 10.1078/0944-2006-00064.

21. Seeley T.D., Tautz J. (2001), Worker piping in honey bee swarms and its role in preparing for liftoff, *Journal of Comparative Physiology A*, **187**: 667–676, doi: 10.1007/s00359-001-0243-0.

22. Soares B.S., Luz J.S., de Macêdo V.F., Silva R.R.V., de Araújo F.H.D., 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.

23. Wright W.D. (1913), *The Honey Bee*, Albany, J.B. Lyon Company.

24. Yegnanarayana B., Prasanna S.M., Zachariah J.M., Gupta C.S. (2005), Combining evidence from source, suprasegmental and spectral features for a fixed-text speaker verification system, *IEEE Transactions on Speech and Audio Processing*, **13**(4): 575–582, doi: 10.1109/TSA.2005.848892.

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

26. Zlatkova A., Gerazov B., Tashkovski D., Kokolanski Z. (2020), Analysis of parameters in algorithms for signal processing for swarming of honeybees, [in:] *2020 28th Telecommunications Forum* (*TELFOR*), pp. 1–4, doi: 10.1109/TELFOR51502.2020.9306562.