Comparison of Feature Extraction Methods for Sound-Based Classification of Honey Bee Activity

Comparison of Feature Extraction Methods for Sound-based Classification of Honey Bee Activity

Alessandro Terenzi, Nicola Ortolani, Inês Nolasco, Emmanouil Benetos, *Senior Member, IEEE* Stefania Cecchi, *Member, IEEE*

***Abstract*—Honey bees are one of the most important insects on the planet since they play a key role in the pollination services of both cultivated and spontaneous flora. Recent years have seen an increase in bee mortality which points out the necessity of intensive beehive monitoring in order to better understand this phenomenon and try to help these important insects. In this scenario, this work presents an algorithm for sound-based classification of honey bee activity reporting a preliminary comparison between various extracted features used separately as input to a convolutional neural network classifier. In particular, the orphaned colony situation has been considered using a dataset acquired in a real situation. Different experiments with different setups have been carried out in order to test the performance of the proposed system, and the results have confirmed its potentiality.**

***Index Terms*—Convolutional neural networks, feature extraction, continuous wavelet transform, Hilbert-Huang transform, mel frequency cepstrum coefficients, honey bees.**

I. Introduction

HONEY bees are probably one of the most important insects on the planet. Their importance is not only related to the production of honey, beeswax, royal jelly and propolis but mainly to their key role in pollination services [1] for both spontaneous and cultivated flora. The last decades have seen a large increase in bee mortality [2] which has led to serious ecological and economical consequences. The causes of these phenomena can be found in the so called "colony collapse disorder" (CCD), which is a situation characterized by a sudden disappearance of honey bees from the hive [3], [4]. Many bee scientists agree that the decline of honey bee colonies is the result of multiple stressors, acting independently, in combination, or synergistically to impact honey bees' health [5]; in this scenario, it the necessity of an intensive honey bee monitoring process is clear in order to understand the problems and the causes of this mortality. While several different approaches could be used for honey bee monitoring [6]–[8], one of the most promising is based on sound analysis. Indeed, sounds are used by bees to communicate within the colony [9], [10] and an interpretation of these signals could lead to the identification of critical situations. Specifically, these sounds are generated through vibroacoustic signals production that include gross body movements, wing movements, high-frequency muscle contractions without wing movements, and pressure of the thorax against the substrates or another bee [11]–[13]. It has been shown that, under some circumstances, some correlations can be detected between beehive sounds and prediction of events like swarming [14]–[18], the presence of airborne toxic substances in the hive [19], the presence of a young queen inside the hive [20] and the presence or the absence of the queen inside the colony, as reported in the review paper of [13], where also vibrations have been considered.

We need to point out that the presence of the queen bee is a key point for the hive survival. In case the queen bee should die, the hive would not have any chance of surviving because the queen is the only fertile female. If the beekeeper is informed about this event, he/she can act to add another queen bee in the hive. Focusing on this aspect, an innovative approach has been presented in [21] where both convolutional neural networks (CNNs) and support vector machines (SVMs) were used to classify the presence of the queen bee in the hive from audio signals exploiting as features the traditional approach based on mel frequency cepstral coefficients (MFCCs) [22] combined with an innovative feature extraction technique based on the Hilbert-Huang transform (HHT) [23]. The obtained results have shown that the proposed combination is capable of improving the results with SVM classifiers. In this paper, starting from the results of [21], an extended investigation on the performance of other innovative feature extraction techniques exploiting CNNs is presented. In particular, the following novelties are introduced in this work:

- the use of HHT as standalone feature is investigated;
- the use of CNN with HHT as feature is used;
- the introduction of continuous wavelet transform (CWT) and discrete wavelet transform (DWT) [24] as new features for the CNN approach is considered starting from the analysis reported in [25], [26].

Several CNN experiments have been carried out using the same network architecture presented in [21]. The same dataset of [21] has been used with the following differences:

- some changes to the preprocessing section have been performed such as removing the pitch-shift data augmentation and using shorter audio frames;
- the data obtained from the feature extraction has been used as image instead of data matrix [27]–[30] to allow a compact management of the feature coefficients and to have the same number of coefficients for different feature extraction methodologies;

A. Terenzi, N. Ortolani, S.Cecchi are with the Department of Information Engineering, Università Politecnica delle Marche, Italy. Inês Nolasco and Emmanouil Benetos are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, UK.

This work was supported in part by Università Politecnica delle Marche Research Grant (NU-Hive Project). I. Nolasco is supported by an EPSRC DTP studentship (grant ref. EP/R513106/1).

- since the dataset includes recordings from two microphones in a different position inside the hive, data from both microphones has been analyzed independently.

The paper is organized as follows: Section II describes the related work on honey bee monitoring problems, Section III shows the proposed approach based on several feature extraction techniques and a convolutional neural network, with Section III-A describing the feature extraction techniques used for sound analysis, and Section III-B presents the CNN architecture with special attention to the layers and the parameters employed in the system. Section IV introduces the dataset, the experimental setup and the obtained results. Finally, Section V reports conclusions and future work.

II. Related work

In recent years, several works have analyzed the problems of developing effective honey bee monitoring systems exploiting different techniques and analyzing different colony parameters. Several approaches are based on continuous beehive monitoring for the health of honey bees exploiting a wide range of sensors, such as [31]–[34]. While some of these systems are complete acquisition platforms with several sensors, there are other works which are mainly focused on a specific parameter and sensor. In [8] weight variations of the colony have been studied in order to detect events such as swarming and rainfall. In [7] humidity and temperature have been analyzed in order to detect the presence of Varroa destructor inside the colony. In [35] machine learning algorithms have been used to analyze several parameters such as weight and temperature to detect if there are health problems inside the colony.
In [17] the recorded sound inside the colony is analyzed by means of short time Fourier transform (STFT) in order to obtain different spectral parameters (e.g., peak frequency) able to detect the presence of the Varroa mite inside the hive. In [16] STFT spectrograms derived from bee sounds have been used to find in advance the swarming of bees: in particular, authors showed that before this specific event there is an increase in low frequency contributions produced by honey bees. In [36] sound during swarming has been analyzed in combination with weight variations, showing that there is a sound amplitude variation at the same time that the weight drops due to the swarming. Spectral analysis has also been used for studying and monitoring the so called "waggle dance" of bees [10]: in fact, authors explain that there is a correlation between bee dance and the presence of harmonics near 320 Hz in the recorded signals, showing that it is possible to generate signals at which bees react. On [37], authors use accelerometers to acquire the bees sound inside the colony, and then by means of multidimensional FFT and discriminant functions, swarming events have been predicted. In [38], machine learning methods are proposed capable of distinguishing bee buzzing from cricket chirping and ambient noise. In [21], [39] beehive sounds have been used in combination with machine learning methods to develop systems capable of distinguishing different states of a hive.

III. Proposed approach

The proposed approach can be divided in two parts: the feature extraction methods and the use of convolutional neural networks as classifiers. A detailed description of each part is reported in the next sections.

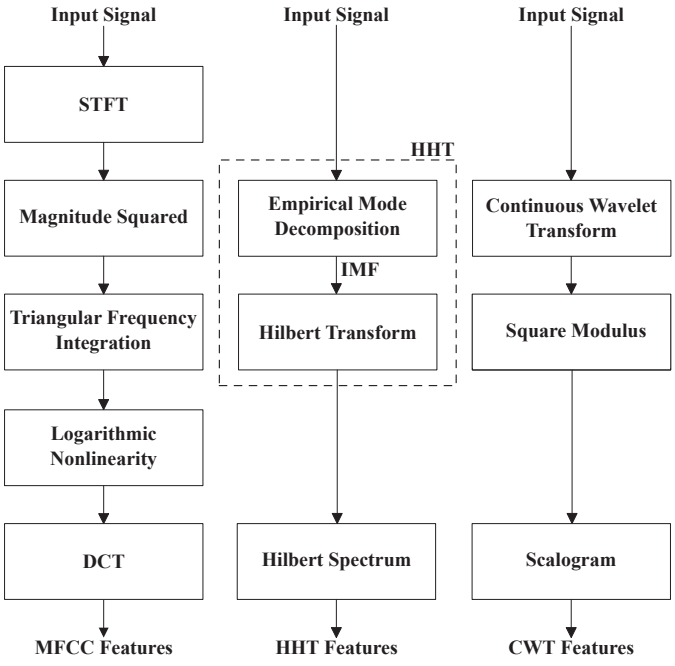


Fig. 1. MFCC, HHT and CWT Feature Extraction procedures.

A. Feature Extraction Techniques

Six different techniques have been considered to extract information from the recorded sounds. In particular, STFT spectrograms, mel spectrograms, mel frequency cepstral coefficients (MFCCs), Hilbert Huang transform (HHT), continuous wavelet transform (CWT) and discrete wavelet transform (DWT) have been used following the preliminary analysis reported in [25], [26].

The first exploited technique is the **STFT spectrogram** following [22]. It allows a visual representation of the signal, showing its spectral content as it varies with time. The spectrogram is based on short time Fourier transform (STFT), i.e., the signal is first windowed by means of a sliding time window then each frame is frequency transformed through the fast Fourier transform (FFT) with a sample frequency of 32 kHz. For the experiments a Hann window has been used, with a window length of 10 ms with a 50% overlap; the results are visible in Fig. 3. Similar to the STFT spectrogram also **mel spectrograms** have been computed. In this representation, the audio input is first buffered into frames, which are overlapped. A periodic Hamming window is then applied to each frame, and then the frame is converted to a frequency-domain representation by means of the STFT. Each frame of the frequency-domain representation passes through a mel filter bank. The spectral values output from the mel filter bank are summed, and finally the channels are concatenated. The window length

used is of 960 samples, the FFT length has the same value, adjacent frames are overlapped of 640 samples and the filter bank has 32 bands; the obtained results are visible in Fig. 4. Derived from the mel spectrogram, also **mel frequency cepstral coefficients (MFCCs)** have been considered. MFCCs are a signal representation widely used in speech processing [40]. As reported in Fig. 1, it consists of five steps:

1) Compute the short time Fourier transform (STFT) of the input signal.
2) Calculate the squared absolute value of the STFT result to get the magnitude spectrogram of the input.
3) Filter the results of the previous step with a mel-filter bank composed of triangular filters.
4) Apply a logarithmic non-linearity to the filtered signals.
5) Compute the discrete cosine transform (DCT) of the result to obtain the mel-scale cepstral coefficients.

In this specific case, MFCC computation returned 20 coefficients; for the first step, hop length and number of STFT points were set to 512 and 2048 respectively. Fig. 5 shows the result of the described procedure.

Hilbert Huang transform (HHT) [23] was already discussed in [21] with an application limited to the SVM classifier and in combination with MFCCs. The approach presented here is derived from the one presented in [26] but it is capable of working in combination with CNN classifiers. HHT is an algorithm for time-frequency analysis based on empirical mode decomposition (EMD) and the Hilbert transform (HT). The schematic of this extraction procedure can be found in Fig. 1. EMD decomposes the original signal generating a series of basis functions called intrinsic mode functions (IMF). The IMFs are obtained through an adaptive procedure directly from the analyzed signal. Each IMF must satisfy the following properties: (A) The number of extrema and the number of zero-crossings must be either equal or differ at most by one extrema. (B) The mean value of the envelope defined by the local maxima and the local minima is zero.
Once the properties of the IMFs have been defined, for a given $x(n)$ signal in the time domain with $n = 1, \ldots, N$, an iterative procedure for IMF estimation can be derived with the following steps:

1) Identify all local extrema.
2) Connect all the local maxima by a cubic spline.
3) Repeat the procedure to produce the lower envelope.
4) Estimate their mean $m_1(n)$.
5) The first estimation of the IMF can now be written as $h_1(n) = x(n) - m_1(n)$.
6) Repeat the procedure up to k times until the function $h_{1k}(n) = h_{1(k-1)}(n) - m_{1k}(n)$ does not satisfy the IMF properties.
7) Now the first IMF component is equal to $c_1(n) = h_{1k}(n)$.
8) Remove from the original signal the component $c_1(n)$ obtaining the first residue $r_1(n) = x(n) - c_1(n)$.
9) Treat $r_1(n)$ as the new signal for the decomposition procedure.

Now the original signal can be reconstructed as a superimposition of a certain number M of IMF plus a residue $r(n)$ i.e.,

$$x(n) = \sum_{j=1}^{M} c_j(n) + r(n), \tag{1}$$

where $c_j(n)$ is the j-th IMF. Once the signal has been decomposed, the Hilbert transform [41] is applied to each IMF and used for the estimation of the analytic signal $a_j(n)$ as follows:

$$a_j(n) = c_j(n) + j\mathcal{H}\{c_j(n)\} \tag{2}$$

where $\mathcal{H}$ indicates the Hilbert transform with $j = 1, \ldots, M$. Then, Eq. (2) can be expressed in polar coordinates, i.e., $a_j(n) = A_j(n)e^{i\phi_j(n)}$ where $A_j(n)$ is the instantaneous amplitude of the signal from which the instantaneous energy $|A_j(n)|^2$ is extracted, and $\phi_j(n)$ is the phase from which the instantaneous frequency can be derived according to:

$$f_j(n) = \frac{f_s}{2\pi}\big[\phi_j(n+1) - \phi_j(n)\big], \tag{3}$$

where f_s is the sampling frequency. In this way, for each time instant n, the energy $|A_j(n)|^2$ corresponding to a specific frequency $f_j(n)$ is derived from the original signal. Finally, the Hilbert marginal spectrum for the j-th IMF is derived as the sum of the instantaneous energy for each specific frequency according to the following equation:

$$H_j(w) = \begin{cases} \sum\limits_{n \in \{f_j(n) = w\}} |A_j(n)|^2, & \text{if } w = f_j(n) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

with $n = 1...N$ the time instants and w the frequency index.

Thus, the result is a frequency amplitude representation for each Hilbert function which can be represented in a three dimensional map where for each IMF the spectral content is clearly visible. In this work, after several experiments, a five level decomposition has been used. Fig. 6 shows the results of the described procedure.

Another approach proposed here is based on the **wavelet transform (WT)**. Wavelets are mathematical functions that decompose the signal by means of properly amplitude and time shifted basis functions called mother wavelets. Wavelets allow a time-frequency analysis of non stationary signals, and wavelet based approaches have been already applied successfully to animal sound analysis as reported in [42], [43]. For a given signal $x(t)$ its wavelet transform is defined by the following formula:

$$X(s,u) = \frac{1}{\sqrt{\mid s \mid}} \int_{-\infty}^{+\infty} x(t)\psi^* \left(\frac{t-u}{s}\right) dt, \tag{5}$$

where s is the scale index, u the wavelet temporal index, and ψ is the mother wavelet. Variation of scale index is related to the signal frequencies, i.e., a higher scale value means a dilated wavelet which approximates well low frequencies, while a low scale value means a compressed wavelet which is closer to high frequencies. The choice of the mother wavelet is strictly related to the type of signal which has to be analyzed, and a complete description of the most important mother wavelets

can be found in [44]. When a signal is analyzed by means of WT, a graphic time-frequency representation can be derived using the square modulus of the transformed signal, i.e.,

$$S(s,u) = \frac{1}{s}\left|\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-u}{s}\right)dt\right|^2, \qquad (6)$$

where $S(u,s)$ is defined as the *scalogram*. In Eq. (5) and Eq. (6) indices u and s change continuously realizing the **continuous wavelet transform (CWT)**. Scalograms generated by means of CWT are visible in Fig. 8; using non-overlapped frames of one second. The bump wavelet [45] has been selected after several experiments, due to its performance with this type of signals. The number of scale values adopted for a signal representation is chosen automatically from the algorithm itself according to the signal properties and in this application a typical number of coefficients is around 110. On CWT, scale values and time indices change continuously but can be discretized generating the **discrete wavelet transform (DWT)**, limiting the s and u values according to:

$$\begin{cases} s = s_0^{-m}, & \text{with } s_0 > 1,\ m \in \mathbf{Z} \\ u = nu_0 s_0^{-m}, & \text{with } u_0 > 0,\ n \in \mathbf{Z}. \end{cases} \qquad (7)$$

The DWT has already been used for feature extraction and animal sound analysis [42], [46]. One of the main advantages of the DWT is its reduced computational cost due to the fact that the transform can be easily implemented with a three structure filter bank as the one reported in Fig. 2. The filter bank implements a multi-resolution analysis of the signal, i.e., the lower frequencies are analyzed with a higher resolution. Each branch of the filter bank is composed by a high-pass filter $H(z)$ and a low-pass filter $G(z)$ and two decimators. Each level generates two distinct outputs: the detailed coefficients (i.e., the output of the high-pass filter), and the approximation coefficients (i.e., the output of the low-pass filter). The detailed coefficients are stored, while the approximation coefficients are used as input for the next decomposition level. Since the taps for the low-pass and high-pass filter are generated from the mother wavelet [47], the output of the filter bank corresponds to the output of the discrete wavelet transform. Due to non-uniform decimation, each wavelet coefficient vector has a different length. In order to obtain a more clear graphic representation, each DWT coefficient has been properly interpolated replicating its samples and thus obtaining a set of uniform length vectors with the same length of the input signal. This operation has been tested through several experiments verifying the impact of the interpolation on the system classification. The selected interpolation values guarantee the best classification results without data alteration. For this work, a ten level filter bank has been used obtaining eleven coefficients (i.e., ten detailed coefficients and one approximation coefficient). For the mother wavelet many different wavelet families have been tested and then the discrete Meyer [48] has been selected due to its results with the analyzed signals. For both continuous and discrete wavelet transform, the estimated coefficients have been first rescaled within the range between 0 and 1.

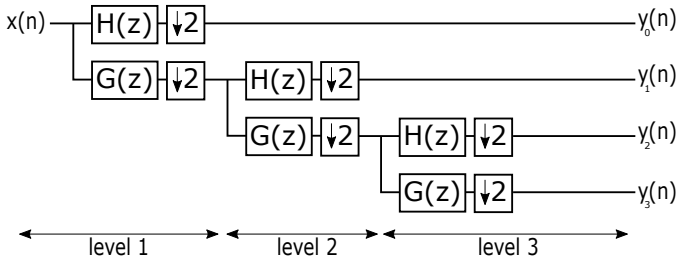


Fig. 2. A 3-Level asymmetric dyadic analysis filter Bank. $G(z)$ and $H(z)$ are respectively the low-pass and the high-pass filters derived from the mother wavelet.

B. Network Architecture

The network architecture, visible in Figure 9, was derived from the one presented in [21]. Table I illustrates the details of it, with a focus on the output sizes and the parameter values used in each layer. The output dimensions refer to an input obtained considering an RGB image as input with a dimension of 256x256 pixels. As it can be seen from Table I, there are four convolutional layers composed of 16 filters, each one followed by a max-pool layer and a dropout layer; the first two convolutional layers have a kernel size of 3×3 and a max-pool size of 3×3 with a 2×2 stride, whereas the last two convolutional layers have a 3×1 kernel size and a pool size of the same dimensions (2×2). The last block, instead, has three dense layers composed of 256, 32, 1 units respectively. To regularize the training procedure, dropout layers with a rate of 0.5 were used after each max-pooling layers and after the first two dense layers. A leaky rectifier linear unit (LReLU) [49] was employed for all convolutional and dense layers except the last one, which uses a sigmoid function for the classification. The network training was performed by optimizing the binary cross-entropy between predictions and targets; root mean square propagation (RMSProp) [50] was used as optimizer with a learning rate of 0.001 and the batch size was set to 145. The training was stopped when the validation loss did not decrease for more than five epochs. As regards the initialization of the network weights, a uniform distribution with zero biases [51] was used for both the convolutional and dense layers.

IV. Algorithms evaluation

A. NU-Hive Dataset

Data used in this work came from the NU-Hive project [52]. NU-Hive is a project which acquires continuously several different parameters from three different hives of *Apis mellifera L.* located inside the University campus; the acquired parameters include weight, temperature, humidity, CO_2 and sound. More details about the acquisition platform can be found in [36]. For this work, the same data of [21] has been used, focusing on the orphaned colony situations and the dataset is publicly available in [53]. In particular, the sound used for the classifier was recorded with a Behringer UCA22 sound card at a sampling rate of 32 kHz and ADMP401 MEMS microphones. The sound came from a total of two different hives, these were both recorded in a normal condition

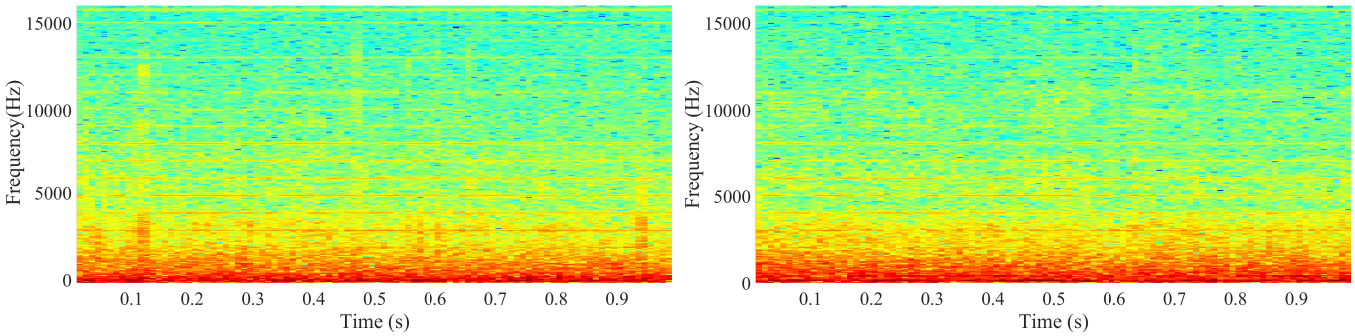


Fig. 3. STFT spectrograms considering (a) normal colony, and (b) an orphaned colony.

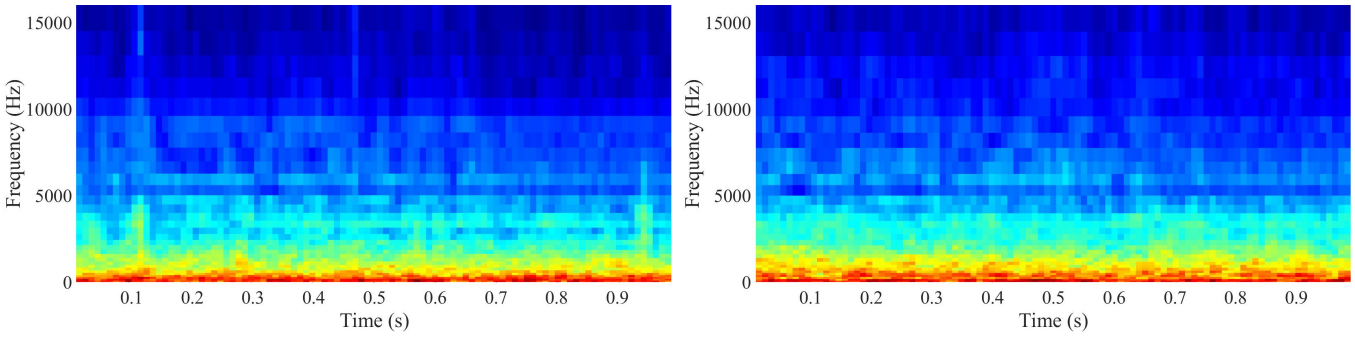


Fig. 4. Mel spectrograms considering (a) normal colony, and (b) an orphaned colony.

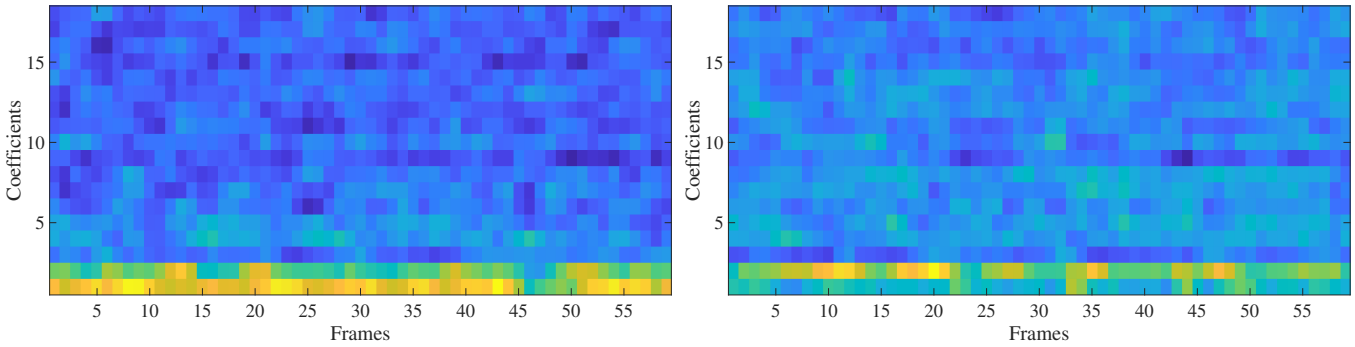


Fig. 5. Mel frequency cepstrum coefficients feature extractions considering (a) normal colony, and (b) an orphaned colony.

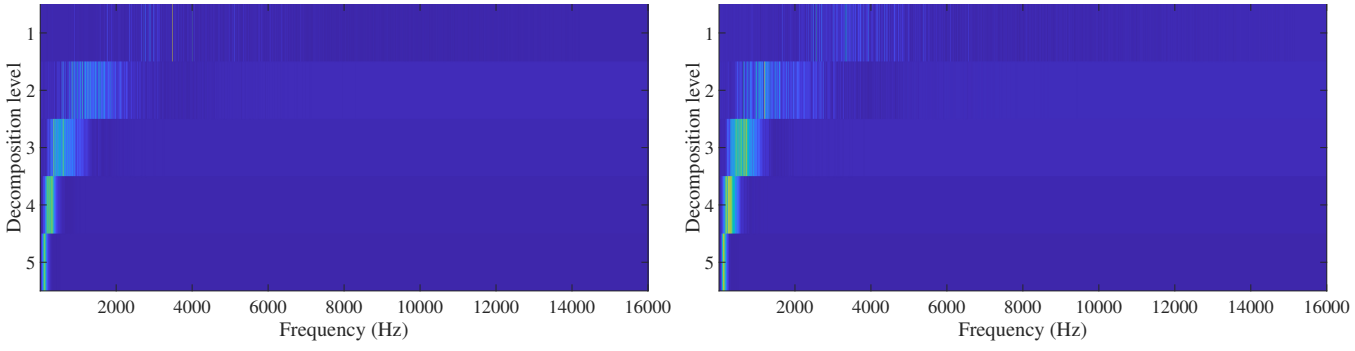


Fig. 6. Hilbert-Huang transform feature extractions considering (a) normal colony, and (b) an orphaned colony.

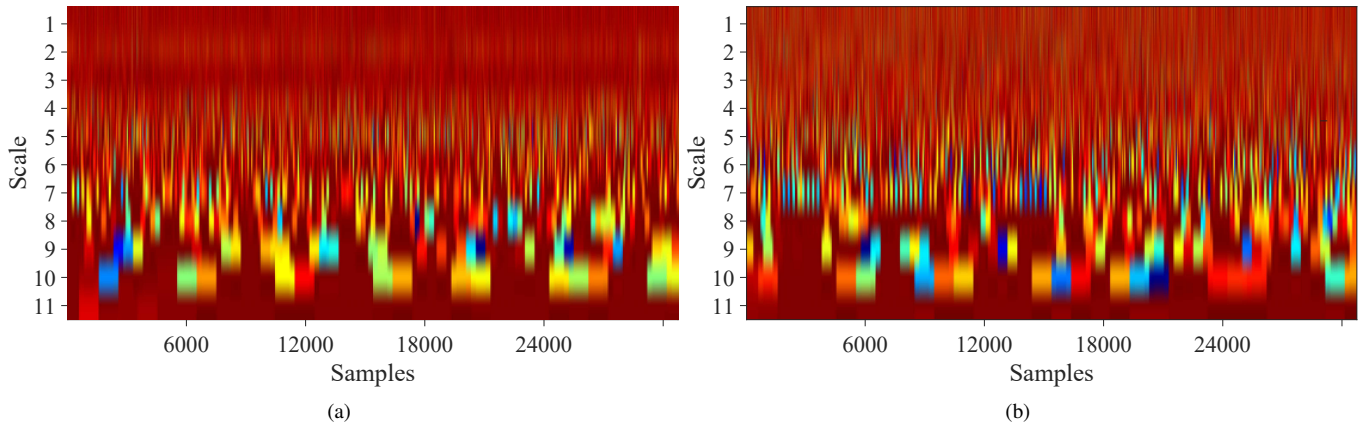


Fig. 7. Discrete wavelet transform feature extractions considering (a) normal colony, and (b) an orphaned colony.

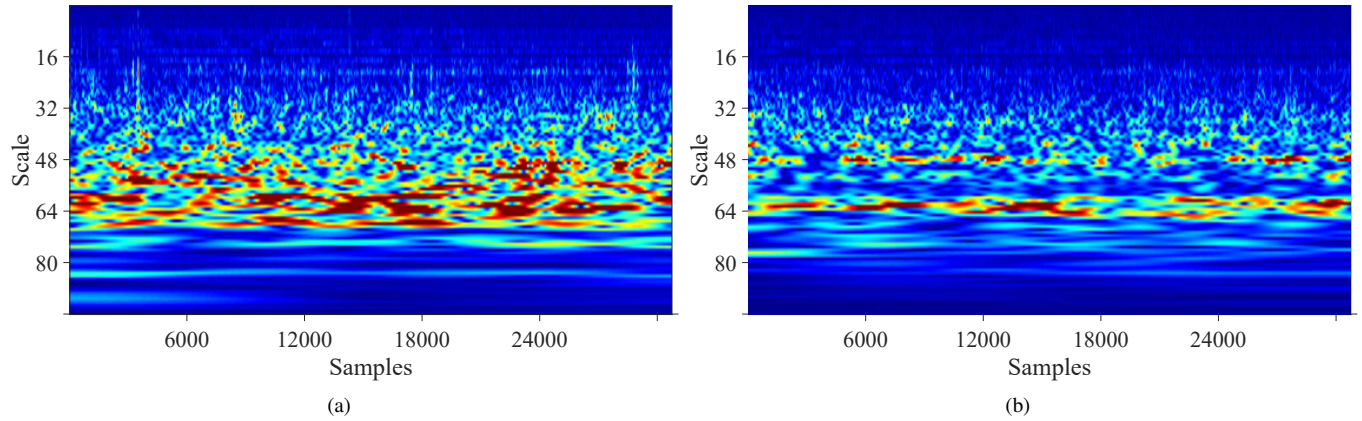


Fig. 8. Continuous wavelet transform feature extraction considering (a)a normal colony, and (b) an orphaned colony.

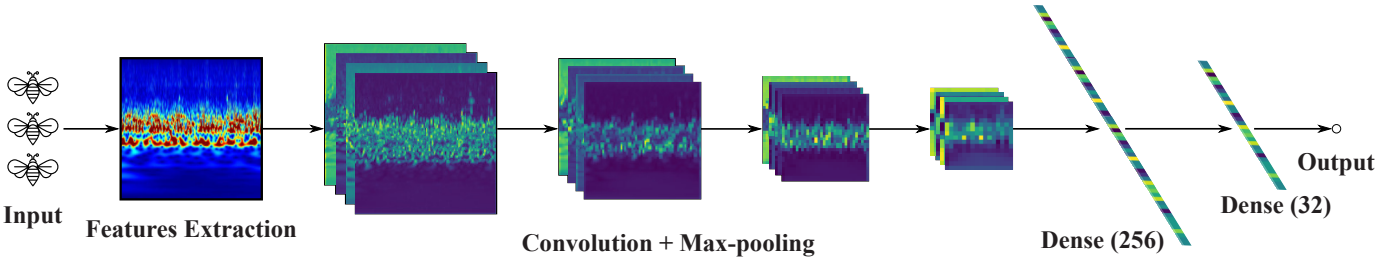


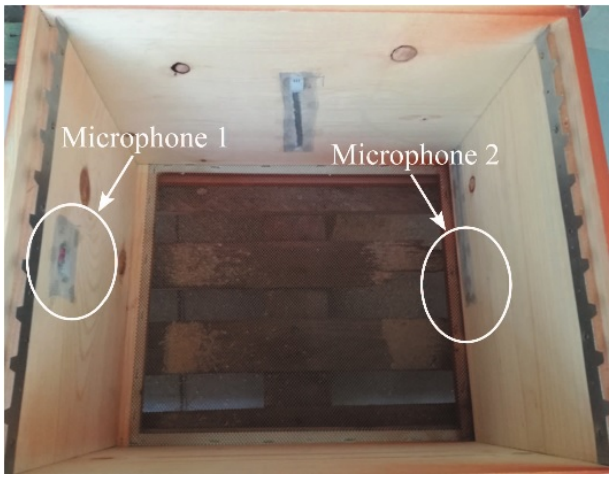
Fig. 9. Schematic of the proposed convolutional neural network, containing four convolution layers each followed by max-pooling layers. The last three layers are dense layers, with 256, 32, 1 units respectively.

(i.e., the queen bee is present) and in an orphaned situation (i.e., the queen bee is dead). Four days of continuous recording have been used, with a total amount of 576 files of ten minutes duration each (96 hours). The number of orphaned and normal colony recordings are almost of the same length (i.e., the same number of files). Each recording consists of two separated tracks which belong to two different microphones placed inside the colony. The microphones have been placed according to entomologists' indications in order to achieve the best results: in particular, microphone 1 is placed near the hive entrance and microphone 2 is placed near the bees brood in the center back of the hive, where the brood cluster is firstly initialized (as visible in Figure 10(a)). Furthermore, to avoid propolization of the microphones, they have been positioned in a groove covered by a particular type of fabric. Figure 10(b) shows the microphones' status inside the hive after the installation. Since the position of the two microphones has been chosen in order to monitor different parts of the colony (i.e., the activity near the brood and the activity near the colony entrance), data from both microphones has been analyzed independently, comparing the results in order to see if there is a difference between the two recordings.

Before applying any feature extraction techniques, each ten minute audio file was segmented into 1 second slices in order to increase the number of training/testing samples (345 600 files in total). The frame length of one second has been chosen according to several experiments with different frame length, moreover this frame length was already used in other work

TABLE I
PROPOSED NETWORK ARCHITECTURE.

Layer	Output	Description
input	(256, 256, 3)	RGB Image
convolution	(256, 256, 16)	3×3 kernel size, 16 filters
max-pooling	(128, 128, 16)	3×3 pool size, 2×2 strides
dropout	(128, 128, 16)	0.5 rate
convolution	(128, 128, 16)	3×3 kernel size, 16 filters
max-pooling	(64, 64, 16)	3×3 pool size, 2×2 strides
dropout	(64, 64, 16)	0.5 rate
convolution	(64, 64, 16)	3×1 kernel size, 16 filters
max-pooling	(32, 32, 16)	3×1 pool size, 2×2 strides
dropout	(32, 32, 16)	0.5 rate
convolution	(32, 32, 16)	3×1 kernel size, 16 filters
max-pooling	(16, 16, 16)	3×1 pool size, 2×2 strides
dropout	(16, 16, 16)	0.5 rate
dense	(256)	256 units
dropout	(256)	0.5 rate
dense	(32)	32 units
dropout	(32)	0.5 rate
dense	(1)	1 unit



(a)

(b)

Fig. 10. (a) Microphone positions inside the hive, and (b) detail of the microphone after three months from the first installation.

[54].The data augmentation method used in [21] was not employed here since according to [10], [12], [16] most of the information is carried by the signal frequencies, and a pitch shifting in these signals could be misleading. Finally, images were derived from each of the one second slices and then normalized dividing each pixel by the maximum permitted value (i.e., 255), in order to have values between 0 and 1. The exact same procedure was applied for training/validation/test datasets. The choice of using images as input for the neural network is related to the different dimensionality of each approach as also used and reported in [27]–[30]. The use of images allows a simpler data management using the exact same neural network and normalization procedure for each approach tested. In this case ".jpg" format has been considered without data compression guaranteeing a reliable data representation. Furthermore, several experiments have also been carried out in order to find the best colormap to represent the data without loss of information.

B. Experimental Setup

In order to evaluate the performance of the proposed feature extraction methods, two different experimental setups have been considered, following [21]. In the first scenario, called "random-split" setup, the entire dataset was divided in three parts: in particular, 5% of the whole dataset was reserved for the testing set and the remaining 95% was randomly split in half for the training and the validation sets. In the second scenario, called "hive-independent" setup, the data was divided according to which hive it belongs: in this way, the network can be trained using only data coming from hive 1, and tested on independent data coming from hive 2. In particular, the validation set was obtained by randomly selecting 10% of training data of hive 1. This second scenario is more interesting especially for a real-world application scenario, since the system needs to have good generalization properties toward unseen hives.

Following the evaluation method used in [21], the performance of the various systems has been evaluated using the area under the curve (AUC) score [55], which is a metric that describes the performance of a binary classifier across all possible classification thresholds. In particular, a model whose predictions are 100% correct has an AUC score of 1.0; on the contrary, a model whose predictions are completely wrong has an AUC score of 0.0. Precision, recall and F1 measure, which are evaluation metrics widely used in sound recognition tasks [56], [57], have also been computed. They are defined respectively as:

$$P_l = \frac{tp_l}{tp_l + fp_l} \tag{8}$$

$$R_l = \frac{tp_l}{tp_l + fn_l} \tag{9}$$

$$F1_l = \frac{2P_lR_l}{P_l + R_l} \tag{10}$$

where tp_l is the number of the true positives, fp_l the number of false positives and fn_l the number of false negatives for each of the two considered labels l (i.e., queen-bee presence

or absence). In other words, precision quantifies how many of the predicted results are relevant, while recall measures how many truly relevant results are returned. F1 measure quantifies instead the overall performance of the network and it is computed as the harmonic mean of precision and recall.

The experiments were run two times each, changing the random seed for the data split. Averaged results are reported in the next section.

C. Results

Qualitative evaluations

In the next section, the approaches adopted will be discussed from a qualitative point of view. For the sake of brevity the image discussed came from only one of the two colonies, however similar results have been obtained also for the recordings of the other hive. Fig. 3 shows the STFT spectrograms obtained from the analyzed data. Comparing the normal situation with the orphaned bees some differences are visible especially in the higher frequency regions. In particular, some lines are evident in Fig. 3(a) that are not present in Fig. 3(b), lower frequencies show also a different energy spread, in particular the orphaned colony seems to have more energy and a more uniform energy distribution.

Taking into consideration mel spectrograms of Fig. 4, the same differences between the normal colony on Fig. 4(a) and orphaned 4(b) of STFT spectrograms are visible. However, the non-uniform scale resolution of mel spectrograms allows a clearer analysis of the different harmonic content of the two signals.

Fig. 5 shows the MFCC results. According to Section III-A the first two coefficients have been discharged. Comparing the two figures, the first two coefficients (the third and the fourth considering the two discharged) show the major difference. In particular on Fig. 5(a), with the normal colony recording, the first coefficient has more energy, while for the orphaned colony on Fig. 5(b) it is the second one that is the most energetic. Fig. 6 shows the obtained results for the HHT-based approach in terms of decomposition level and frequency range. Comparing the two situations, the harmonic content is different since the orphaned colony on Fig. 6(b) shows more harmonics at the higher and lower frequencies in contrast with the normal colony on Fig. 6(a), that has more energy in the middle frequencies, i.e., second level of IMF. With respect to other techniques, this approach gives a better estimation of the fundamental frequencies of the original signal. The shifting process used for empirical mode decomposition [23] removes riding waves and smooths uneven amplitudes, leaving only the mode of oscillation contained in the original signal. In comparison with STFT spectrograms and mel spectrograms, the HHT shows differences in the harmonic content with a better resolution, while with respect to MFCC and wavelet transform, the HHT is more clear in representing the harmonics actually included in the analyzed signal.

Focusing on the wavelet transform, both discrete and continuous cases have been considered and applied. Fig. 8 shows the results for the CWT case. Fig. 8(a) shows the WT-analysis from a normal colony, while Fig. 8(b) shows scalograms from an orphaned colony, and it is evident that we have obtained different spectral contents in the two situations. The orphaned colony figure shows a behaviour with less and more limited frequencies, while the normal situation shows more harmonics. Finally the DWT approach shows different harmonic content, in particular the differences between normal colony on Fig. 7(a) and the orphaned one on Fig. 7(b) are more visible on medium-high frequencies, i.e., on lower scale coefficients, where the normal colony seems to have less energy on the first and third coefficient.

Objective evaluation

The six feature extraction techniques have been evaluated comparing the AUC scores obtained by analyzing microphone 1 and 2 individually. As can be seen from Table II, for the random split case the STFT spectrograms have the highest AUC score, even if the CWT and the mel spectrograms are quite similar; for the hive-independent setup, on the contrary, the CWT has the highest score on channel 1, while on channel 2 seems to perform better the mel spectrograms. As regards the HHT, for both the random split and the hive-independent setups the performance is worse than the other two techniques, indicating that this particular method is not well suited for this classification task; this approach probably should work better with other colony health status (i.e., the swarming as reported in [25]). Since the hive-independent setup, as said, is the most interesting for a real-world implementation, CWT as a feature extraction method applied to honey bees is definitely to be taken into consideration. Another interesting thing that can be seen from Table II is that, for the random split experiments, all six approaches have shown better performance on channel 2, while for the hive-independent this is not true. This aspect should be taken into consideration for those situations in which the installation of two microphones within the hives is not feasible. Furthermore, we have to point out that the microphone positions are very important because the hive modifies its dimension during different seasons and years. When the colony is young or during the winter, the brood and the bees' activities are focused at the center of the hive, and microphone 1 is more close to the center. Microphone 2 is positioned inside the hive but near the landing pad, so it gives important information when the colony is sufficiently large to have extended the brood inside the hive and during the spring and summer when the bees go outside for the pollen. It is probably that for depending on the season, the two microphone positions could be different importance. This aspect will be investigated in future works. To clarify the results reported in Table II, Figure 11 shows the receiver operating characteristic (ROC) curve for microphone 1 and considering the discrete wavelet transform algorithm. It is evident that the random split approach obtains better results than hive independent approach. Similar results have been achieved for the other algorithms but for the sake of brevity they have been omitted.

Focusing on precision, recall and F1, the scores are presented in Table III for the orphaned colony and in Table IV for the normal colony. The results for the random split

experiment are positive for almost all the techniques proposed, again as for the AUC scores the best performance is reached by STFT spectrograms and CWT. It is interesting to underline that comparing the two tables for the the hive independent experiment the network tends to classify most of the images on the normal colony class. However, while STFT spectrograms and mel spectrograms seem to suffer more from this problem, the approaches based on HHT and wavelets seem to have less issues.

TABLE II
AUC COMPARISON BETWEEN THE SIX FEATURE EXTRACTION TECHNIQUES.

	AUC			
	Random Split		Hive-Independent	
	Mic.1	Mic.2	Mic.1	Mic.2
STFT spectrograms	**0.9967**	**0.9997**	0.3247	0.7082
Mel spectrograms	0.9866	0.9950	0.4429	**0.7858**
MFCCs	0.9757	0.9906	0.5060	0.6555
HHT	0.9195	0.9514	0.6301	0.6290
DWT	0.9189	0.9480	0.7589	0.4892
CWT	0.9814	0.9968	**0.7648**	0.5937

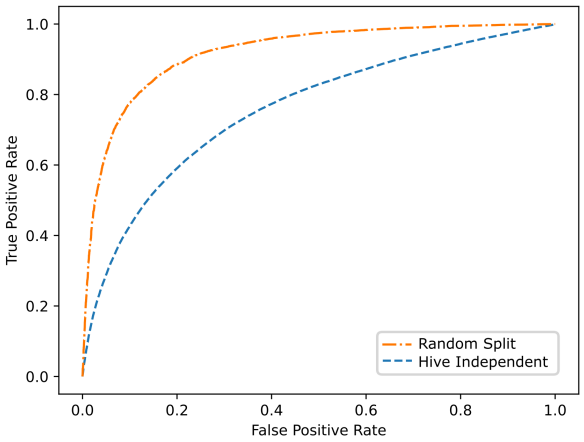


Fig. 11. ROC curves for random split and hive independent experiments, considering microphone 1 and discrete wavelet transform as algorithm.

V. CONCLUSIONS AND FUTURE WORK

In this paper, a comparison of different feature extraction techniques for sound-based classification of honey bee activity has been presented. In particular, starting from well-known approaches of the state of the art (i.e., STFT spectrograms, mel spectrograms, MFCCs), HHT, DWT and CWT techniques have been used exploiting a convolutional neural network for classification of beehive activity.

Two experimental setups have been considered, the first one with a random split of the data and the other one considering a hive-dependent split. Different metrics have been used to evaluate the overall performance of the network with the different approaches and with the different microphones. Focusing on the AUC results, the best performance is achieved by the STFT spectrograms and CWT on random split experiments, where microphone 2 seems also to have the best performance. On the hive independent case, CWT and mel spectrograms have the best performance, in particular on channel 1 the CWT and on channel 2 the mel spectrograms. In comparison with the random split case some techniques perform better on channel 1 and other on channel 2, so even if a difference seems present, it is not clear which microphone has the best positioning. Focusing on precision, recall and F1 metrics, the results show that the network seems to overestimate the normal colony situation, but the innovative approaches seem to have less problems. Some of the errors in the classification could be related to the fact that the dataset used in this work has not been checked manually and some recordings could have unwanted sound such as car noises. Future works will consider an improvement to the acquisition system including an external microphone to detect unwanted noise, and a manual check of the dataset similar to the one reported in [58].

Future work will also deeper analyze the performance of the proposed algorithms using a bigger dataset (the NU-Hive project is still acquiring data until now) and considering a multi-channel classifier and a late fusion or ensemble approach in order to improve the classification results. The focus will be also on the hive independent case which is the most important for these applications. Different microphone positions could be also considered, in order to evaluate the sound in different colony locations also with relation to the hive dimensions and the different activities of the bees during seasons. The proposed approach will be also used to classify other important beehive states such as swarming detection, pest presence or dangerous situations. Regarding the implementation, the final system should be able to acquire and analyze the data in real-time in order to give an immediate alert to the beekeeper. Furthermore, the influence of ambient noise on the obtained results will be investigated. Finally, the classification algorithm will be also implemented in a low cost board (e.g., a Raspberry Pi) in order to allow a direct evaluation of the colony health status.

REFERENCES

[1] H. Hoshiba and M. Sasaki, "Perspectives of multi-modal contribution of honeybee resources to our life," *Entomological research*, vol. 38, pp. S15–S21, 2008.

[2] A.-M. Klein, B. E. Vaissière, J. H. Cane, I. Steffan-Dewenter, S. A. Cunningham, C. Kremen, and T. Tscharntke, "Importance of pollinators in changing landscapes for world crops," *Proceedings of the Royal Society B: Biological Sciences*, vol. 274, no. 1608, pp. 303–313, 2007. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2006.3721

[3] J. Faucon, L. Mathieu, M. Ribière, A. Martel, P. Drajnudel, S. Zeggane, and et al., "Honey bee winter mortality in france in 1999 and 2000," *Bee World*, vol. 83, pp. 14–23, 2002.

[4] B. Oldroyd, "What's killing american honey bees?" *PLOS Biol*, vol. 5, 2007.

[5] D. Van Engelsdorp, J. J. Hayes, R. Underwood, and P. J.S., "A survey of honey bee colony losses in the united states, fall 2008 to spring 2009," *J Apic Res.*, vol. 49, pp. 7–14, 2010.

[6] W. Meikle, N. Holst, G. Mercadier, F. Derouané, and R. James, "Using balances linked to dataloggers to monitor honey bee colonies," *Journal of Apicultural Research*, vol. 45, no. 1, pp. 39–41, 2006.

[7] C. Hou, B. Li, S. Deng, and Q. Diao, "Effects of varroa destructor on temperature and humidity conditions and expression of energy metabolism genes in infested honeybee colonies," *Genet. Mol. Res*, vol. 15, 2016.

TABLE III
PRECISION, RECALL, AND F1 MEASURE COMPARISON FOR THE NO-QUEEN-BEE CLASS.

No-queen-bee class												
	Random split						Hive-Independent					
	Mic. 1			Mic. 2			Mic. 1			Mic. 2		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
STFT spectrograms	**0.9864**	0.9439	**0.9647**	**0.9967**	**0.9928**	**0.9948**	0.3025	0.2378	0.2663	0.6163	**0.8719**	**0.7221**
Mel spectrograms	0.8811	**0.9799**	0.9279	0.9902	0.9904	0.9699	0.3243	0.2085	0.2538	**0.7043**	0.7272	0.7155
MFCCs	0.9429	0.8990	0.9204	0.9620	0.9572	0.9596	0.4857	0.3553	0.4104	0.6248	0.2254	0.3313
HHT	0.8570	0.7892	0.8218	0.8736	0.9014	0.8873	0.6015	0.4724	0.5292	0.6012	0.5875	0.5821
DWT	0.8046	0.8967	0.8481	0.8256	0.9353	0.8770	0.7155	**0.6445**	**0.6781**	0.4856	0.2680	0.3454
CWT	0.9525	0.9020	0.9266	0.9700	0.9795	0.9747	**0.7353**	0.4911	0.5889	0.5540	0.5238	0.5385

TABLE IV
PRECISION, RECALL, AND F1 MEASURE COMPARISON FOR THE QUEEN-BEE CLASS.

Queen-bee class												
	Random split						Hive-Independent					
	Mic. 1			Mic. 2			Mic. 1			Mic. 2		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
STFT spectrograms	0.9462	**0.9869**	**0.9662**	**0.9929**	**0.9968**	**0.9948**	0.3694	0.4488	0.4052	**0.7791**	0.4543	0.5739
Mel spectrograms	**0.9774**	0.8678	0.9193	0.9524	0.9906	0.9711	0.4145	0.5633	0.4776	0.7165	0.6930	**0.7046**
MFCCs	0.9035	0.9455	0.9240	0.9574	0.9621	0.9598	0.4897	0.6218	0.5479	0.5260	**0.8640**	0.6539
HHT	0.8025	0.8668	0.8334	0.8968	0.8680	0.8822	0.5638	0.6854	0.6187	0.5875	0.6239	0.6052
DWT	0.8830	0.7817	0.8293	0.9254	0.8024	0.8595	**0.6751**	0.7424	**0.7071**	0.4927	0.7146	0.5832
CWT	0.9069	0.9551	0.9304	0.9793	0.9696	0.9744	0.6165	**0.8223**	0.7046	0.5462	0.5760	0.5607

[8] W. G. Meikle, N. Holst, T. Colin, M. Weiss, M. J. Carroll, Q. S. McFrederick, and A. B. Barron, "Using within-day hive weight changes to measure environmental effects on honey bee colonies," *PloS one*, vol. 13, no. 5, 2018.

[9] H. Frings and F. Little, "Reactions of honey bees in the hive to simple sounds," *Science*, pp. 122–125, 1957.

[10] A. Michelsen, W. H. Kirchner, and M. Lindauer, "Sound and vibrational signals in the dance language of the honeybee, apis mellifera," *Behavioral Ecology and Sociobiology*, vol. 18, no. 3, pp. 207–212, Jan. 1986.

[11] W. H. Kirchner, "Acoustical communication in honeybees," *Apidologie*, vol. 24, no. 3, pp. 297–307, 1993.

[12] M. Hrncir, F. G. Barth, and J. Tautz, "Vibratory and airborne sound-signals in bee communication," in *In Insect Sounds and Communication: Physiology, Behaviour, Ecology, and Evolution*, S. Drosopoulos and M. Claridge, Eds. CRC Press, 2006, pp. 421–436.

[13] J. H. Hunt and F. J. Richard, "Intracolony vibroacoustic communication in social insects," *Insectes Sociaux*, vol. 60, pp. 403–417, 2013.

[14] D. Dietlein, "A method for remote monitoring of activity of honeybee colonies by sound analysis," *Journal of Apicultural Research*, vol. 24, no. 2, pp. 176–183, 1985.

[15] S. Ferrari, M. Silva, M. Guarino, and D. Berckmans, "Monitoring of swarming sounds in bee hives for prevention of honey loss," in *International Workshop on Smart Sensors in Livestock Monitoring*, Sep. 2006.

[16] ——, "Monitoring of swarming sounds in beehives for early detection of the swarming period," *Computers and Electronics in Agriculture*, vol. 65, pp. 72–77, 2008.

[17] A. Qandour, I. Ahmad, D. Habibi, and M. Leppard, "Remote beehive monitoring using acoustic signals," *Acoustics Australia / Australian Acoustical Society*, vol. 42, no. 3, pp. 204–209, Dec. 2014.

[18] M. Bencsik, J. Bencsik, M. Baxter, A. Lucian, J. Romieu, and M. Millet, "Identification of the honey bee swarming process by analysing the time course of hive vibrations," *Computers and electronics in agriculture*, vol. 76, no. 1, pp. 44–50, 2011.

[19] J. J. Bromenshenk, "Honey bee acoustic recording and analysis system for monitoring hive health," 2007, uS Patent 7549907.

[20] H. Eren and et al., "Electronic sensing and identification of queen bees in honeybee colonies," in *IEEE Instrumentation and Measurement Technology Conference*, May 19-21 1997.

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

[22] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. Prentice Hall International Inc., 1999.

[23] N. Huang, "Introduction to the Hilbert-Huang transform and its related mathematical problems," in *Hilbert-Huang Transform and Its Applications*. World Scientific, Sep. 2005, pp. 1–26.

[24] I. Daubechies, *Ten Lectures on Wavelets*. USA: Society for Industrial and Applied Mathematics, 1992.

[25] S. Cecchi, A. Terenzi, S. Orcioni, and F. Piazza, "Analysis of the sound emitted by honey bees in a beehive," in *Audio Engineering Society Convention 147*. Audio Engineering Society, 2019.

[26] A. Terenzi, S. Cecchi, S. Orcioni, and F. Piazza, "Features extraction applied to the analysis of the sounds emitted by honey bees in a beehive," in *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 2019, pp. 03–08.

[27] A. Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey, and P. Tiwari, "Sound classification using convolutional neural network and tensor deep stacking network," *IEEE Access*, vol. 7, pp. 7717–7727, 2019.

[28] S. Amiriparian, M. Gerczuk, S. Ottl, N. Cummins, M. Freitag, S. Pugachevskiy, A. Baird, and B. W. Schuller, "Snore sound classification using image-based deep spectrum features." in *INTERSPEECH*, vol. 434, 2017, pp. 3512–3516.

[29] Z. Mushtaq, S.-F. Su, and Q.-V. Tran, "Spectral images based environmental sound classification using cnn with meaningful data augmentation," *Applied Acoustics*, vol. 172, p. 107581, 2020.

[30] Z. Ren, N. Cummins, V. Pandit, J. Han, K. Qian, and B. Schuller, "Learning image-based representations for heart sound classification," in *Proceedings of the 2018 International Conference on Digital Health*, 2018, pp. 143–147.

[31] F. E. Murphy, M. Magno, P. Whelan, and E. P. Vici, "b+ wsn: Smart beehive for agriculture, environmental, and honey bee health monitoring—preliminary results and analysis," in *2015 IEEE Sensors Applications Symposium (SAS)*. IEEE, 2015, pp. 1–6.

[32] F. E. Murphy, M. Magno, K. O'Leary, L.and Troy, P. Whelan, and E. M. Popovici, "Big brother for bees (3b)—energy neutral platform for remote monitoring of beehive imagery and sound," in *2015 6th International Workshop on Advances in Sensors and Interfaces (IWASI)*. IEEE, 2015, pp. 106–111.

[33] L. Chazette, M. Becker, and H. Szczerbicka, "Basic algorithms for bee hive monitoring and laser-based mite control," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2016, pp. 1–8.

[34] S. Gil-Lebrero, F. J. Quiles-Latorre, M. Ortiz-López, V. Sánchez-Ruiz, V. Gámiz-López, and J. J. Luna-Rodríguez, "Honey bee colonies remote monitoring system," *Sensors*, vol. 17, no. 1, p. 55, 2017.

[35] A. R. Braga, D. G. Gomes, R. Rogers, E. E. Hassler, B. M. Freitas, and J. A. Cazier, "A method for mining combined data from in-hive sensors, weather and apiary inspections to forecast the health status of honey bee colonies," *Computers and Electronics in Agriculture*, vol. 169, p. 105161, 2020.

[36] S. Cecchi, S. Spinsante, A. Terenzi, and S. Orcioni, "A smart sensor-based measurement system for advanced bee hive monitoring," *Sensors*, vol. 20, no. 9, p. 2726, 2020.

[37] M.-T. Ramsey, M. Bencsik, M. I. Newton, M. Reyes, M. Pioz, D. Crauser, N. S. Delso, and Y. Le Conte, "the prediction of swarming in honeybee colonies using vibrational spectra," *Scientific reports*, vol. 10, no. 1, pp. 1–17, 2020.

[38] V. Kulyukin, S. Mukherjee, and P. Amlathe, "Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples," *Applied Sciences*, vol. 8, p. 1573, 09 2018.

[39] A. Zgank, "Bee swarm activity acoustic classification for an iot-based farm service," *Sensors*, vol. 20, no. 1, p. 21, 2020.

[40] R. Serizel, V. Bisot, S. Essid, and G. Richard, "Acoustic features for environmental sound analysis," in *Computational Analysis of Sound Scenes and Events*, T. Virtanen, M. D. Plumbley, and D. Ellis, Eds. Springer, 2018, pp. 71–101.

[41] L. Marple, "Computing the discrete-time "analytic" signal via fft," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2600–2603, Sept 1999.

[42] M. Lubis, S. Pujiyati, T. Hestirianoto, and W. P.D., "Haar wavelet method to spectral analysis continuous wavelet transform 1d using whistle sound to position of dolphins (tursiops aduncus)," *Journal of Applied & Computational Mathematics*, vol. 5, no. 2, pp. 1–8, 2016, doi: 10.4172/2168-9679.1000305.

[43] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani, "Asymptotic wavelet and gabor analysis: Extraction of instantaneous frequencies," *Information Theory, IEEE Transactions on*, vol. 38, pp. 644 – 664, 04 1992.

[44] I. Daubechies, "Ten lectures on wavelets," vol. 61. Siam, 1992.

[45] S. C. Olhede and A. T. Walden, "Generalized morse wavelets," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2661–2670, 2002.

[46] A. Selin, J. Turunen, and J. T. Tanttu, "Wavelets in recognition of bird sounds," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 051806, Dec 2006.

[47] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[48] Y. Meyer, *Ondelettes et opérateurs*, ser. Actualités Mathématiques. Hermann, 1991, no. v. 3.

[49] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[50] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[51] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. PMLR, May 2010, pp. 249–256.

[52] S. Cecchi, A. Terenzi, S. Orcioni, P. Riolo, S. Ruschioni, and N. Isidoro, "A preliminary study of sounds emitted by honey bees in a beehive," in *Audio Engineering Society Convention 144*, May 2018.

[53] "Nuhive project database." [Online]. Available: https://zenodo.org/record/2667806.YOw-i-gzaUl

[54] T. Cejrowski, J. Szymański, H. Mora, and D. Gil, "Detection of the bee queen presence using sound analysis," in *Intelligent Information and Database Systems*, N. T. Nguyen, D. H. Hoang, T.-P. Hong, H. Pham, and B. Trawiński, Eds. Springer International Publishing, 2018, pp. 297–306.

[55] A. Mesaros, T. Heittola, and D. Ellis, "Datasets and evaluation," in *Computational Analysis of Sound Scenes and Events*, T. Virtanen, M. D. Plumbley, and D. Ellis, Eds. Springer, 2018, pp. 147–179.

[56] J. J. Bosch, J. Janer, F. Fuhrmann, and P. Herrera, "A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals." in *ISMIR*, 2012, pp. 559–564.

[57] F. Fuhrmann and P. Herrera, "Polyphonic instrument recognition for exploring semantic similarities in music," in *Proc. of 13th Int. Conference on Digital Audio Effects DAFx10*, 2010, pp. 1–8.

[58] I. Nolasco and E. Benetos, "To bee or not to bee: Investigating machine learning approaches for beehive sound recognition," in *2018 Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2018, pp. 133–137.

Alessandro Terenzi was born in Senigallia, Italy, in 1991. He received the Laurea degree (with honors) in electronic engineering in July 2016 at the Polytechnic University of Marche (Italy), and the Ph.D degree in Electronic engineering on March 2021. He is now a Post Doc Researcher at DII (Department of Information Engineering) at the same university. His current research interests are in the area of digital signal processing, including nonlinear audio system and audio processing. Dr. Terenzi is a member of the Audio Engineering Society (AES).

Nicola Ortolani was born in Recanati, Italy, in 1992. He received the Master's Degree (with honors) in electronic engineering from the Università Politecnica delle Marche (Ancona, Italy) in 2019. From 2019 to 2020 he worked as a research fellow at the Department of Information Engineering (DII, Università Politecnica delle Marche). His main research activities are in the area of digital signal processing, with special focus on audio processing and audio beamforming.

Inês Nolasco started her a PhD in 2020 at the Centre for digital music at Queen Mary university of London. Her research is concerned with automatic identification of individual animals through their vocalisations. Has research interests in sound event classification and bioacoustics.

Emmanouil Benetos (Senior Member, IEEE) received the Ph.D. degree in electronic engineering from Queen Mary University of London, U.K., in 2012. He is Senior Lecturer with the School of Electronic Engineering and Computer Science, Queen Mary University of London, and Turing Fellow with The Alan Turing Institute. From 2013 to 2015, he was University Research Fellow with the Department of Computer Science, City, University of London. He has authored or co-authored more than 140 peer-reviewed papers spanning several topics in audio and music signal processing. His research focuses on signal processing and machine learning for audio and music signal analysis, as well as applications to music information retrieval, sound scene analysis, and computational musicology.

Stefania Cecchi was born in Amandola (Italy) in 1979. She received the Laurea degree (with honors) in electronic engineering from the University of Ancona (now University Politecnica delle Marche, Italy) in 2004 and the Ph.D. degree in electronic engineering from the University Politecnica delle Marche (Ancona, Italy) in 2007. She was a Post Doc Researcher at DII (Department of Information Engineering) at the same university above, from February 2008 to October 2015 and she was Assistant Professor from November 2015 to October 2018. She is Associate Professor at the same department since November 2018. She is the author or coauthor of several international papers. Her current research interests are in the area of digital signal processing, including adaptive DSP algorithms and circuits, speech, and audio processing. Dr. Cecchi is a member of the Audio Engineering Society (AES), Institute of Electrical and Electronics Engineers (IEEE), Italian Acoustical Association (AIA).