Workshops at 18th International Conference on Intelligent Environments (IE2022)
H.H. Alvarez Valera and M. Luštrek (Eds.)

doi:10.3233/AISE220032

Prediction of Honeybee Swarms Using Audio Signals and Convolutional Neural Networks

Stenford RUVINGA[a,1], Gordon HUNTER[a,2], Jean-Christophe NEBEL[a] and Olga DURAN[b]
[a] *School of Computer Science and Mathematics, Kingston University, KT1 2EE, U.K.*
[b] *School of Engineering and the Environment, Kingston University, London SW15 3DW, U.K.*

Abstract. Honeybees are of vital importance to both agriculture and ecology. Unfortunately, their populations have been in serious decline over recent years. Swarms from hives are both of great importance to wider success of a colony and of major significance to beekeepers. In this paper, we contribute to the challenge of predicting when a swarm is going to occur. We have employed a Convolutional Neural Network (CNN) approach applied to audio data recorded from hives. Our initial results are extremely encouraging, since they allow us to distinguish hives which are preparing to swarm from those which are not with high levels of accuracy.

Keywords. Honeybees, queen bee, swarm, bee colony, audio signal, CNN, STFT spectrogram, Mel spectrogram.

1. Introduction

Insects in general, and honeybees in particular, are vital to both the agricultural industry and to the wider ecosystem due to their role in pollinating flowering plants. In the agricultural sector, almonds, apples, blueberries, and many other fruit crops depend on honeybees for their primary method of pollination [1]. However, many pollinator insect populations, especially honeybees, have been in serious decline over recent decades. Whilst various factors have been proposed to explain this – including the use of pesticides, climate change, monoculture agriculture and even the increase in microwave telecommunications (including mobile telephone use) – the true causes are not yet well-understood.

To maintain the health and well-being of honeybee colonies, various approaches for electronic monitoring of beehives have been proposed. Although several commercial systems are available, most of these are not affordable to the typical hobbyist beekeeper or small-scale farmer. However, several researchers have developed low-cost approaches to monitoring bee colonies via measuring the hive's temperature, humidity, and mass, and recording the audio signals made by the bees [2, 3]. Two particularly significant events in the life cycle of a honeybee colony are the loss of the queen and the occurrence of a swarm. The former is critical since the queen is the only reproductive female bee in the colony. The lifespan of a worker bee is typically of the order of just 40 days during

[1] Corresponding Author; E-mail: k1103714@kingston.ac.uk
[2] Corresponding Author; E-mail: G.Hunter@kingston.ac.uk

Summer, so the hive needs a constant supply of new workers over the course of that season. In recent years, acoustic measurements [4,5] have supported claims that experienced beekeepers could tell whether a hive lacked a healthy queen from changes in the sounds made by the bees. These insights have been used by researchers who used machine learning approaches to analyze acoustic signals from beehives to classify them as having a healthy queen present or not [6, 7]. Swarms are also important for the wider success of a bee colony. They normally occur when an old queen, together with a substantial proportion of the workers, leave the hive just before a new queen emerges as a young adult. Beekeepers have reported that they noted changes in the sounds made by the bees in the weeks running up to a swarm leaving a hive [4, 5, 8], and that the temperature within the hive tended to rise over the same period. These observations have been incorporated into swarm prediction systems by various previous authors [9, 10, 11]. This present paper incorporates Machine Learning Techniques, notably Convolutional Neural Networks (CNNs), into a system to predict swarms from beehives based on acoustic signals. As will be seen, this approach proves highly satisfactory.

The remainder of this paper is structured as follows. In section 2, we briefly give an account of the "demography" of a typical honeybee colony. In section 3 we present our methodology and describe our dataset. In section 4 we display our results and in section 5 we discuss our results to date. Finally, in section 6, we present our current conclusions and our proposed future work.

2. Honeybees

Honeybees are social insects which live in large colonies of typically over 30000 individuals in Summer, although that number significantly falls in winter. A colony consists of the queen, female worker bees, male drones, eggs, and larvae. Honeybee colonies are well-organized, with daily activities shared between all members [12, 13]. The queen bee is responsible for laying fertilized eggs while the female worker bees are responsible for building the comb, foraging, and protecting the hive [12, 13]. The drones' main responsibility is to fertilize a new queen and regulate the hive temperature [12, 13]. Young worker bees are responsible for cleaning cells, nursing the brood, feeding bee larvae, and processing incoming pollen and nectar. Honeybee colonies become active during the spring and summer seasons when honey production is at its peak. Notable events such as swarming also take place during this time of the year. Swarming is when a large proportion of the colony moves away from the original hive usually to form a new colony. It is a natural way a honeybee colony uses to reproduce itself. These colony activities are communicated and coordinated through various audio and vibrational signals [4, 5, 8].

3. Datasets and Methodologies

3.1. Datasets

The honeybee audio datasets used in this work were provided by Agsenze Ltd., a company which develops "smart" agricultural technology. The audio wav files are organized in 5 categories, according to the number of days before swarming, i.e., 3, 7, 14, 21 and 28. For each such category, audio recordings were made (on the same days) from hives where swarms occurred and control hives where they did not. All the audio files were recorded with a sampling rate of 44.1 kHz during April and May of 2010 and 2011. Table 1 below shows the amount of recorded time for the controlled and swarming hives for each number of days before swarming.

Table 1. Distribution of durations of recording of each category for each period before swarming in minutes.

Days before swarm	No swarm	swarm	Total
3	60	30	90
7	55	30	85
14	50	25	75
21	45	20	65
28	35	20	55

Short-Time Fourier Transform (STFT) spectrograms and Mel-spectrograms [22] were used to train the CNN models. The spectrograms were extracted from 5 second clips of the honeybee audio files using Matlab [14, 15]. The resultant examples from honeybee audio recordings from the control hives were combined to form the "No swarm" class and those from the swarming hives to form the "Swarm" class for each of the 5 periods under investigation. The distribution of samples available for each number of days before the swarm are shown in Table 2.

Table 2. Distribution of the samples (spectrograms and Mel-spectrograms) for each period before swarming. For both the "No swarm" and "Swarm" categiories, whilst, respectively 420 and 240 examples were available for the full 28 day period, more examples were available for the shorter periods of time leading up to the swarm.

Days before swarm	No swarm	swarm	Total
3	720	360	1080
7	660	360	1020
14	600	300	900
21	540	240	780
28	420	240	660

3.2. Methodology

In this paper, we explore the use of STFT spectrograms and Mel spectrograms as input features to a CNN classifier to distinguish between bee audio signals from hives where a swarm occurred and where no swarm occurred.

3.2.1 Convolutional Neural Networks

A convolutional neural network (CNN) is a type of artificial neural network (ANN) widely used to analyse images [16]. Its name derives from the fact that it uses a specialised mathematical operation called a convolution in place of the usual matrix multiplication used in traditional neural networks [17]. For a 2D image input, I, and a kernel K of size m × n, the convolution operation is defined (for the (i, j) pixel) as:

$$C(i,j) = (I * K)(i,j) = \sum_m \sum_n I(n,m)\, K(i-m, j-n) \tag{1}$$

where the 2D output $C(i,j)$ is usually referred to as the feature map. In a CNN all the units in a feature map share the same weights and biases and hence they detect the same features at all possible locations on the input [18]. This reduces the number of free parameters in a model which make it less prone to overfitting. Another important aspect of CNNs is that they use sparse connections which makes them easier and faster to train compared to other networks of comparable size [18].

The convolutional layer computes the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. After the convolution operation, a Rectified Linear Unit (ReLU) activation function is used to increase nonlinearity in the feature map [19]. The ReLU is defined as:

$$g(x) = \max(0, x) \quad (2)$$

The convolutional layer has four hyperparameters, the number of filters c, their spatial extent l, the stride s, and the amount of zero padding p on the input. For an input 'volume' of size w_i x h_i x c, filter of size l x l and stride s, the layer outputs a 'volume' of size w_o x h_o x c where :

$$w_o = \left[\frac{w_i - l + 2p}{s}\right] + 1 \quad \text{and}$$

$$h_o = \left[\frac{h_i - l + 2p}{s}\right] + 1 \quad (3)$$

The pooling layer summarises the outputs of adjacent feature map values in the same filter map. This progressively reduces the size of the feature representation and the number of parameters resulting in faster network computation. Commonly used methods (see Fig. 2) are max pooling, which takes the maximum within a rectangular neighbourhood defined by the filter map, and average pooling, which reports the average output within a rectangular neighbourhood defined by the filter map [18]. No weights are trained in the pooling layer, it only has two hyperparameters the filter spatial extent L and the stride S. It takes the output 'volume' size of the preceding convolutional layer, w_o x h_o x c, as input 'volume' and outputs a 'volume' of size w_p x h_p x c, where:

$$w_p = \left[\frac{w_o - L}{S}\right] + 1 \quad \text{and}$$

$$h_p = \left[\frac{h_o - L}{S}\right] + 1 \quad (4)$$

c is the number of filters; L is the pooling layer spatial extent and S is the stride used in the pooling layer.

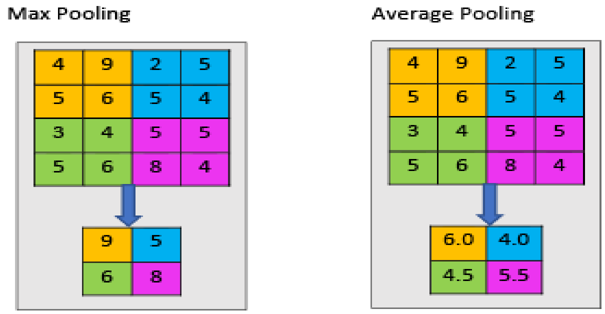

Figure 1. An illustration of the two most commonly used pooling methods with $L = 2$ and $S = 2$.

The fully connected layer is just a regular multi-layer perceptron artificial neural network. It takes the unrolled or flattened volume from the last convolutional or pooling layer as its input.

3.2.2. STFT Spectrograms

A spectrogram is a time-frequency transformation which converts a one-dimensional sequence x[n] into a two-dimensional function of a discrete-time and a discrete frequency [20]. The STFT spectrogram is obtained by applying a STFT to the signal. The STFT is basically a discrete Fourier Transform applied to equal, usually overlapping, portions of a finite length signal. For a signal x[n] and window $w[n]$, the STFT is defined as:

$$X[m,k] = \sum_{n=0}^{N-1} \mathrm{x}[n]\, w[n-m] e^{-\frac{2\pi i k n}{N}} \tag{5}$$

for $0 \leq m, k \leq N - 1$, where N is the number of points used to compute the STFT.

The spectrogram S is then generated by computing the squared magnitude of the STFT of the signal x[n]

$$S\,(m,k) = |X(m,k)|^{\,2} \tag{6}$$

Note that spectrograms are usually represented as 2D images with frequency on the y-axis and time on the x-axis, with the color or intensity representing the magnitude of $X[m, k]$.

3.2.3. Mel Spectrograms

A Mel spectrogram is a representation of an audio signal obtained by using the Mel scale frequency. The spectrum from the STFT described above is warped along its frequency axis f (in Hz) into the Mel-scale using triangular overlapping windows [21] using the formula:

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{7}$$

where f denotes the normal frequency in Hz, and f_{mel} denotes the corresponding Mel frequency [22]. The resultant Mel frequencies are then filtered using the formula below.

$$Y[m] = \sum_{k=1}^{N} W_m[k]|X(m,k)|^2 \tag{8}$$

for $0 \leq k \leq N$ and $0 \leq k \leq M$, where k is the STFT bin number, m is the Mel-filter bank number, M is the total number of triangular Mel weighting filters and $W_m[k]$ is the weight given to the k^{th} energy spectrum bin contributing to the m^{th} output band.

3.3 CNN Hive Swarm Status Classification

To evaluate the classification performances, the Accuracy defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

is calculated for each classification task where TP = number of true positives, FP = number of false positives, FN = number of false negatives, FN = number of true negatives, and the total number of hive status examples is n = $TP + TN + FP + FN$.

For constructing the CNN, we used the Keras API [23]. Our network had 2 convolutional layers, 2 pooling layers and 2 fully connected layers (see Figure 2 for the architectural details). For training the models we used a binary cross-entropy loss function, max pooling, batch size of 64 and an ADAM [24] optimiser. For all the experiments we used 80% of each class for training and 20% for validation. The input features, STFT spectrograms and Mel-spectrograms, were of size 256 x 256. Due to data limitations, we employed a 5-fold cross validation to train the models for all the periods under investigation.

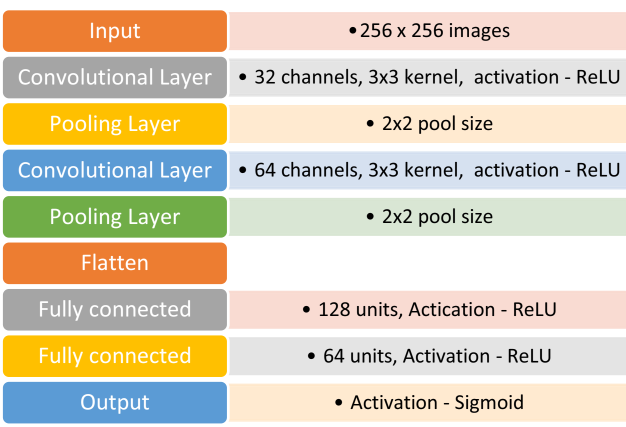

Figure 2. The CNN network architecture used for both sets of features.

4. Results

As Table 3 shows, for both types of spectrograms, the CNN models give good mean accuracies for their task of discriminating between the two hive swarm statuses. The models for predicting swarming 7 days before the event proved to be the most accurate (and with lowest standard deviation) for both features: 99.80% accuracy for Mel and 99.51% for STFT spectrograms. On the other hand, predictions 28 days before the swarming period were the worst in terms of accuracy (and had the highest standard deviations): 89.55% and 90.91% for Mel and for STFT spectrograms, respectively, and the most variation of the mean accuracies. In general, as might be expected, there is a downward trend in performance (accuracy) from the period 3 days to that of 28 days before a swarm, with significant drops in mean accuracy between 14 and 21 days, and between 21 and 28 days, before the swarm. Neither set of features performed markedly better than the other: Mel-spectrograms performed slightly better on the periods closer to the swarm, namely 3 and 7 days, while STFT spectrograms proved a little better on data longer (i.e., 14, 21 or 28 days) before the swarm.

Table 3. Accuracy (%) statistics for the trained CNN classification models

	SFTT Spectrograms		**Mel Spectrograms**	
Days before swarm	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**
3	98.52	0.90	99.63	0.74
7	99.51	0.44	99.80	0.39
14	99.00	0.78	98.61	1.81
21	97.98	0.89	95.12	2.48
28	90.91	5.57	89.55	3.57

5. Discussion

Our work adds evidence supporting the value of acoustics in monitoring beehives and, in particular, time-based prediction of hives swarming. The CNN models for both STFT and Mel spectrograms achieved a high accuracy for discriminating between hives where a swarm occurred and those where no swarm occurred. One notable observation is that our approach gives very good accuracy (> 89%) as much as 28 days before a hive swarms. This is quite remarkable in that it corresponds to a period quite some time before the old queen had laid the eggs which would develop into new queens – which occurs around 16 days before a swarm [13]. While there is no difference genetically between an egg or larva which will develop into a queen bee from one which will develop into a worker, the key distinction is that queen larvae are fed a much richer diet by existing adult workers. Our results suggest that the workers already in a hive decide to rear new queens well before the required eggs are laid by the current queen. Critically, we have also shown that a swarming event can be detected at least 4 weeks before it occurs, which should give beekeepers enough time to prepare for it and put in place measures that can safeguard the welfare of their honeybees.

A curious observation is that our models perform better predicting whether a swarm will occur 7 days before the date than 3 days before. Whilst we do not claim to fully understand the reason for this, we note that the balance of the 3 day dataset is a little different to that of the 7 day dataset, with the former having 33% of the data "swarm" cases, whereas the latter has 35% of the data swarm cases. Alternatively, the larger 3 day dataset has larger standard deviations for both categories and hence may contain more "outlier" examples for swarm and/or non-swarm cases, which may contribute to skewing the results. These hypotheses may warrant further investigation in the future.

6. Conclusions and Future Work

In this work, we showed that a CNN network using both spectrograms and Mel spectrograms can be used to detect the eminent occurrence of a swarm at least 28 days before. The best results achieved by both sets of features were for 7 days before a swarm occurred. The predictive accuracies for both sets of features gradually decrease with greater time before the swarm. Our results show that acoustic monitoring of hives can be a useful tool for beekeepers to remotely monitor and predict the swarming status of their hives. In the future, we aim to combine information from both acoustic and hive temperature signals with the objective of obtaining an exceptionally reliable prediction of when a hive is going to swarm. We also hope to compare data on European honeybees with those from African bees, which appear to be more resilient to the problems affecting their European cousins.

7. Acknowledgements

The data were provided by Agsenze Ltd. We are grateful to Dr. Heather Sanders of Agsenze for making this available, to apidologist Dr. Michael Adjaloo and to beekeepers Colm Treacy and Stewart Westsmith for providing valuable insights into the life of honeybees. Stenford Ruvinga is supported by a Postgraduate Research Studentship from Kingston University. This work is in part funded by Innovate U.K. project grant 45568.

References

[1] United Nations "We all depend on the survival of bees in Ukraine" [Internet] United Nations; [Accessed 8 March 2022]. Available from: https://ukraine.un.org/en/112086-we-all-depend-survival-bees

[2] Kridi DS, de Carvalho CG, Gomes DG. "Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection". Computers and Electronics in Agriculture. 2016, September; 1, 127:221-35.

[3] Hunter G, Howard D, Duran O. "A low-cost multi-modal sensor network for the monitoring of honeybee colonies/hives". Proceedings of ISAPEP 2018, Amsterdam, Netherlands, IOS Press BV; e2018. pp. 69-78. doi:10.3233/978-1-61499-874-7-69

[4] Boys R [Internet] "Listen to the Bees", available on-line at https://beedata.com.mirror.hiveeyes.org/data2/listen/listenbees.htm [Last accessed 7 March 2022]

[5] Kirchner WH. "Acoustical communication in honeybees". Apidologie. 1993;24(3):297-307.

[6] Howard D, Duran O, Hunter G, Stebel K. "Signal processing the acoustics of honeybees (Apis Mellifera) to identify the 'Queenless' state in hives", Proceedings of the Institute of Acoustics 2013;35(1):290-297.

[7] Ruvinga S, Hunter GJA, Duran O, Nebel J-C. "Use of LSTM Networks to Identify "Queenlessness" in Honeybee Hives from Audio Signals". 2021, Proceedings of 17th International Conference on Intelligent Environments (IE). 2021 June; doi : 10.1109/IE51775.2021.9486575

[8] Seeley TD, Tautz J. "Worker piping in honey bee swarms and its role in preparing for liftoff". Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology. 2001 October 1;187(8):667–76.4(1):72-7

[9] Ferrari S, Silva M, Guarino M, Berckmans D. "Monitoring of swarming sounds in bee hives for early detection of the swarming period". Computers and electronics in agriculture. 2008 November 1;6

[10] Zacepins A, Kviesis A, Stalidzans E, Liepniece M, Meitalovs J. "Remote detection of the swarming of honey bee colonies by single-point temperature monitoring". Biosystems Engineering. 2016 Aug;76–80.

[11] Hunter G, Howard D, Gauvreau S, Duran O, Busquets R. "Processing of multi-modal environmental signals recorded from a 'smart' beehive". 2019 Proceedings of the Institute of Acoustics, 41(1):339-348

[12] Johnson BR. "Division of labor in honeybees: form, function, and proximate mechanisms". Behavioral Ecology and Sociobiology. 2010 Jan; 64(3):305-16.

[13] Winston ML. "Temperate and Tropical Honey Bees" in *The Biology of the Honey Bee,* Harvard University Press; 1991

[14] Mathworks : Spectrograms [Internet]. Spectrogram using short-time Fourier transform - MATLAB. [Accessed 7 March 2022]. Available from: https://www.mathworks.com/help/signal/ref/spectrogram.html

[15] Mathworks : Mel spectrogram [Internet]. melSpectrogram - MATLAB. [Accessed 7 March 2022]. Available from: https://uk.mathworks.com/help/audio/ref/melspectrogram.html

[16] Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT press; 2016 Nov 10. 326 p.

[17] Piczak KJ. "Environmental sound classification with convolutional neural networks". In 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP) 2015, 17 September (pp. 1-6).

[18] Krizhevsky A, Sutskever I, Hinton GE. "Imagenet classification with deep convolutional neural networks". Advances in Neural Information Processing Systems (NIPS). 2012;25.

[19] Phung VH , Rhee EJ. "A Deep Learning Approach for Classification of Cloud Image Patches on Small Datasets". Journal of Information and Communication Convergence Engineering [Internet]. 2018. September ; 3 (3) Available from http://dx.doi.org/10.6109/jicce.2018.16.3.173

[20] Wyse L. "Audio spectrogram representations for processing with convolutional neural networks". arXiv preprint arXiv:1706.09559. 29 June 2017.

[21] Rao KS, Manjunath KE. *Speech recognition using articulatory and excitation source features*. Springer; 2017, 11 January.

[22] Rabiner L, Schafer R. *Theory and Applications of Digital Speech Processing*. New Jersey: Prentice Hall Press; 2010.

[23] Keras documentation: "Conv2d layer" [Internet]. Keras. [Accessed 8 March 2022]. Available from: https://keras.io/api/layers/convolution_layers/convolution2d/

[24] Kingma DP, Ba J. "ADAM: A method for stochastic optimization". arXiv preprint arXiv:1412.6980. 2014, 22 December.