however, can fall as low as 5000 bees in Winter. The colony is made up of the queen, thousands of sterile female worker bees, a few hundred male drones, eggs, larvae, and pupae. The queen bee is the lifeline of the colony as she is the only female who can lay fertilized eggs. The worker bees carry out all tasks of a colony such as building the comb, foraging, and hive guarding [1]. Generally, while older worker bees work outside the hive, the activities of younger workers take place inside the hive. Unlike worker bees, drones do not do any work, their primary purpose is to fertilize a new queen – although they may play a part in regulating hive temperature [2]. Bee colonies are well structured with daily activities shared between male and female bees as well as between the young and the old [2]. Honeybees are very active during the spring and summer seasons when activities such as swarming takes place and worker bees gather pollen and nectar used to build food reserves in preparation for winter. These colony activities are communicated and coordinated through various audio and vibrational signals.

III.   RELATED PREVIOUS WORK

There have been several previous investigations relating to the sounds produced by bee colonies. The early work of Woods [3] and then Kirchner [4] reported on the sounds made by honeybees under a variety of situations – especially in the period before a swarm . Further studies relying on bee sounds revealed they could be indicative of a swarm being imminent [5], [6]. Alternatively, other authors [7], [8], [9] noted that a rising temperature inside a hive could predict swarms. With regard to the problem of queen lessness in a hive, [10] used a method based on a Fast Fourier Transform (FFT) and Self-Organising Map (SOM) to investigate how the sounds produced by honeybees in hives changed after their queen had been removed, comparing those acoustic data with those produced in nearby "control" hives where the queen was allowed to remain, under identical local environmental conditions. In this present paper, we re- examine the data of the study in [10], providing its analysis using more sophisticated methodologies, including LSTM neural network [11] and Mel Frequency Cepstral Coefficients (MFCC) [12].

IV.   DATASETS AND METHODOLOGIES

A. Data
The data consists of one-minute duration Waveform audio files sampled hourly from the 3rd to the 9th of August 2012 [10]. It comprises both "Queen Present" (QP) and "Queen Absent" (QA) hives recorded in four separate hives. Two of the hives had the Italian sub-species Apis mellifera ligustica

Abstract — Honeybees are of vital importance to both agriculture and ecology, but honeybee populations have been in serious decline over recent years. The queen bee is of crucial importance to the success of a colony. In this paper, we contribute to addressing these problems by employing Long Short-Term Memory (LSTM), Multi-Layer Perceptron (MLP) Neural Networks and Logistic Regression approaches applied to audio data recorded from "queen-absent" and "queen-present" hives to provide a method of prompt detection of a hive lacking a queen. The initial results – particularly from the LSTM - are highly encouraging.

Keywords — Honeybees, queen bee, queenless, bee colony, audio signal, LSTM, MLP, Logistic Regression, FFT, MFCC.

I.   INTRODUCTION

Insects are of vital importance to both agriculture and the wider ecological system, as scavengers to clear up detritus from dead animals and plants, as vital links in the food chain, but most important as pollinators of many flowering plants – especially a very high proportion of food crops. However, many insect species have been in very serious decline over recent decades, due to a wide variety of proposed causes, including pesticides and fungicides, diseases, predators, and parasites, and possibly climate change or even electro-magnetic radiation. Notable amongst insects under serious decline are several species of bees – including the honeybee *Apis Mellifera* – which are essential pollinators of many fruit crops and flower species, in addition to their role as producers of honey. In this paper, we discuss our approach to a contribution towards addressing the decline of honeybees via electronic monitoring of honeybee colonies. We re-examine approaches to determining whether a given colony contains a healthy queen. Queen bees, as normally the only reproductive female in most eusocial bee colonies, are essential for the colony to survive and thrive. The death of the queen, or a serious decline in her health, will usually lead to the colony becoming extinct unless remedial action is taken promptly.

The remainder of this paper is structured as follows. In section II, we give a brief account of the "demography" of a typical honeybee colony. In section III we discuss previous related work on bee monitoring in general, and of detection of "Queen lessness" in particular. In section IV we present our methodology and describe our dataset, whilst in section V we display and discuss our results to date. Finally, in section VI, we present our current conclusions and lay out our plan for our proposed future work on this topic.

II.   HONEYBEE COLONIES

Honeybees are social insects with a successful colony averaging between 30000 and 80000 bees in Summer. This,

and the other two had the Slovenian Apis mellifera carnica. One hive of each species which had a queen throughout was used as control. In total there were 480 one-minute samples, sampled at 44.1kHz, from the four hives used in this study.

Table 1 The status of each hive for the 5 recording days.

Honeybee	5/8	6/8	7/8	8/8	9/8
Italian	QA	QA	QA	QA	QA
Italian (Control)	QP	QP	QP	QP	QP
Slovenia	QA	QA	QA	QA	QA
Slovenia (Control)	QP	QP	QP	QP	QP

Key: QP – Queen present, QA – Queen absent

B. Methodologies

In this work, we proposed the use of MFCCs as input features in an LSTM classifier to discriminate between a hive with a queen present and that without a queen.

i) Mel Frequency Cepstral Coefficients

MFCC are features extracted from audio signals which are motivated by observations in human speech recognition [13]. They are proposed in this work due to their success reported in other audio related areas such as music genre classification [14] and heart sounds recognition [15] and automatic speaker recognition [16]. Their success is attributed to their good representation of the continually relevant aspects of the short-term audio spectrum [17]. MFCC features are uncorrelated hence there is less redundant information. Usually, the first few MFCCs features are used as they represent most of the information in a signal. The MFCC feature extraction process is summarised in figure 1. A more detailed description of this 7-step process follows.

1) Pre-Emphasis
The signal is filtered to generate energy in a high frequency that was previously compressed during the audio generating process.

2) Framing
The sound signal is partitioned into overlapping frames of equal length about 20–40ms duration [18]. This enables the capturing of signal characteristics while reducing the non-stationarity effects of the signal.

3) Windowing

Windowing is then applied to individual consecutive frames to prevent discontinuity of the signals generated by the framing process. The most commonly used windows are the Hamming and the Hanning windows. The two windows are defined as $w[n]$:

$$w[n] = (1 - \omega) - \omega * cos\left(\frac{2\pi n}{L-1}\right) \qquad (1)$$

for $0 \leq n \leq L – 1$, where L is the window width, $\omega = 0.5$ for the Hanning window and $\omega = 0.46164$ for the Hamming window [19].

4) Fast Fourier Transform

The FFT is an efficient algorithm that computes the Discrete Fourier Transform of a signal (DFT), converting the framed and windowed signal from the time domain to the frequency domain. For a signal x[n] = x(t = n/f$_s$) sampled at regular discrete time periods, where f$_s$ is the sampling frequency and n = 1, 2, 3, … the DFT of the signal X is defined as:

$$X[k] = \sum_{k=0}^{N-1} x[n]e^{-2\pi ikn/N} \qquad (2)$$

For $0 \leq k \leq N – 1$, $i^2 = -1$ and N = number of frames

5) Mel Filter Bank
In Mel Filter banks, the spectrum from the FFT is warped along its frequency axis f (in Hz) into the Mel-scale using triangular overlapping windows [19] using the equation:

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

where f denotes the physical frequency in Hz, and f_{mel} denotes the perceived frequency [19]. The resultant Mel frequencies are then filtered using the equation below.

$$Y[m] = \sum_{k=1}^{N} W_m[k]|X[k]|^2$$

For $0 \leq k \leq N$ and $0 \leq m \leq M$

Where k is the FFT bin number and m is Mel-filter bank number

6) Logarithms Transformation
The powers of the Mel frequencies are then transformed using logarithms to reduce the dynamic range [19].

7) Discrete Cosine Transform (DCT)
Finally, the DCT is taken on the logarithmic outputs from above. This is results in decorrelated MFCC features.

$$c(n) = \sum_{m=0}^{M-1} \log_{10}(s\,(m)\cos\left(\frac{\pi n(m-0.5)}{M}\right))$$

For $0 \leq n \leq C – 1$ and $0 \leq m \leq M - 1$

where c(n) are the cepstral coefficients and C is the number of MFCCs and M the number of filter banks.

Figure 1 Showing the MFCC feature extraction process. FFT is the Fast Fourier Transform, whilst DCT is Discrete Cosine Transform
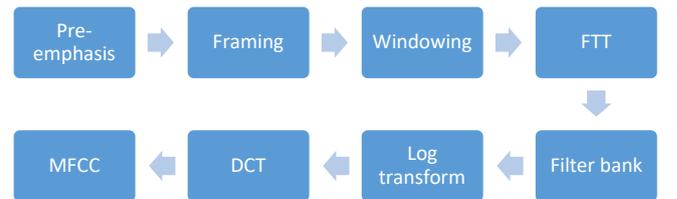
ii) Long Short-Term Memory

LSTM is a type of recurrent neural networks (RNN). Although unlike traditional networks, RNN have loops that enable them to retain sequential information, they are not able to learn long-term dependencies as error signals flowing backwards in time" tend to either blow up or vanish [20]. The LSTM network overcomes this problem by using an architecture which enforces constant error flow through each repeating cell [20]. The repeating module has four interacting layers, forget gate layer, update/input gate layer, a *tanh* layer, and output layer. In the forget gate layer, a sigmoid function is used to determine the information to be thrown away from the cell state. In the input gate layer, another sigmoid function is used to decide the values to be updated in the cell state. In a *tanh*

layer, a vector of new possible values that could be added to the state is created. Lastly the output layer uses a sigmoid function to decide the cell state output values [12].

Figure 2. The LSTM repeating cell which contains four interacting layers [12, 21]
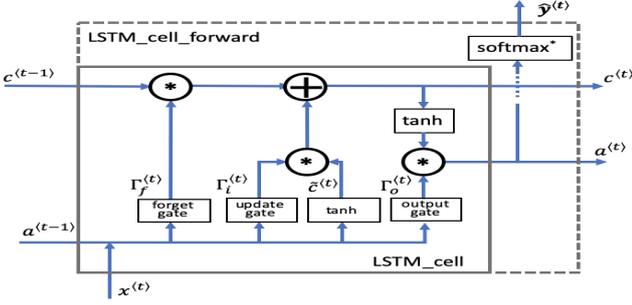
y_t is the y estimate at time t and x_t is the input sequence value at time t.

For an input sequence $x = (x_1, \ldots, x_T)$ the LSTM network outputs $y = (y_1, \ldots, y_T)$ by implementing the following equations.

$$i_t = \sigma(W_{ix}x_t + W_{ia}a_{t-1} + b_i) \qquad (1)$$

$$f_t = \sigma(W_{fx}x_t + W_{fa}a_{t-1} + b_f) \qquad (2)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{ca}a_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{ox}x_t + W_{om}a_{t-1} + b_o) \qquad (4)$$

$$a_t = o_t \odot h(c_t) \qquad (5)$$

$$y_t = \phi(W_{ya}a_t + b_y) \qquad (6)$$

where the W terms are the weight matrices, b terms are the bias vectors, σ is the logistic sigmoid function, and i, f, o and c are the input, "forget" gate, output gates and cell activation vectors respectively, a is the cell output activation vector, $\odot$ is the element-wise product of the vectors, g and h are generally *tanh* cell input and output activation functions and ϕ is the output activation function, sigmoid in this paper [22].

Two standard classifiers, Logistic regression and a MLP neural network, were used for baseline comparisons with the LSTM model performance.

V. RESULTS

i) Preliminary Analysis

We began by exploring the original audio samples using plots of the bee audio samples. The plots reveal clear difference between the two signals in the time domain, where the queen present signal shows less time variability and a smaller amplitude standard deviation than the queen absent signal.

Figure 3 Plots of randomly selected 1 second clips for queenless (red) and queen present (blue) audio signals.
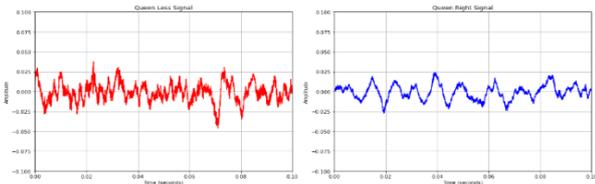
The bee acoustic signals also appear to have distinctive spectral characteristics. The queen absent signal displays higher energy levels overall. Moreover, the energy

distributions are also different when comparing morning and afternoon signals. This suggests that hive status can be classified acoustically for hive monitoring.

Figure 4 Acoustic Spectra for queen absent (left) and queen present (right) hives. The upper graphs show data recorded in the morning, whilst the lower ones display afternoon data.
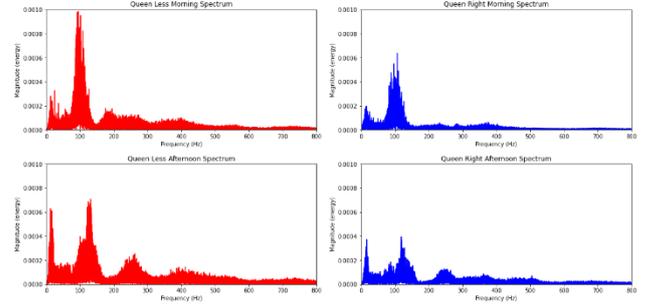
From the audio data, 14 features, 13 MFCCs and log energy and were computed in MATLAB [23]. The resultant MFCCs and log energy from control hives were combined to form the queen present class and those from the none control hives were combined to form the queen absent class. The mean for each Mel coefficient and the log energy was computed for healthy and unhealthy hives data. Then an ANOVA test for each of the MFCCs and log energy averages for the classes were conducted. They indicated that there is a statistically significantly different (p=0.001) at 5% level of significance between the means of the "healthy" and "unhealthy" classes for all MFCCs and the log energy. This confirmed the initial observation illustrated in Figure 3 and 4 suggesting that the two hive classes are distinguishable.

ii) Hive Status Classification

Two target class labels defined above were used to classify the hives status. From the 14 features extracted, 4 combinations of 13 MFCCs and the log energy features were used as features to train three classifiers, a logistic regression MLP and LSTM in our experiments. These are 1) dataset 1 with 13 MFCCs plus the log energy 2) dataset 2 with 13 MFCCs 3) dataset 3 with MFCCs minus the 1st coefficient plus the log energy and 4) dataset 4 with MFCCs minus both the 1st coefficient and the log energy. These combinations were investigated due to their success in other related audio classification tasks [16].

For each of these combinations, a Logistic regression, MLP and an LSTM classifier were applied. We used *Scikit learn* for both the Logistic regression [24] and the MLP classifiers [25]. The LSTM architecture had 100 units, an ADAM optimizer, binary cross-entropy loss function on a Keras [26] framework. 70:30 train – test split of the beehives data was used for training and testing all models. The training set data were normalized so that each coefficient and the log energy had zero mean and unit variance to reduce the effects of undue influence of larger values on the model training weights. To evaluate the model performance the model accuracy was obtained for each classification task. The predictive accuracy of a classifier is defined as:

$Accuracy = \frac{TP+TN}{N}$ were - TP = true positive and TN = true negatives, are the correctly predicted hive status samples for the two classes and N is the total number of hive status samples.

Whatever the feature combination, the LSTM model performed either similarly or better than the other models achieving the best accuracy of 0.92, see Table 2. While the MLP delivered a best accuracy of 0.90, the Logistic regression model could only reach 0.87. One should note that the best accuracy results were achieved when all 14 features were used, while the worst accuracies correspond to the usage of only 12 features.

Table 2: Accuracy values (with perfect = 1.00) across the four datasets considered, as described in Section V B)

Data	Number of Features	Logistic model	MLP model	LSTM model
Dataset 1	14	0.87	0.9	0.92
Dataset 2	13	0.85	0.9	0.9
Dataset 3	13	0.87	0.9	0.91
Dataset 4	12	0.81	0.85	0.87

Figure 5: shows a confusion matrix of the best LSTM model.

Dataset 1 LSTM Confusion Matrix
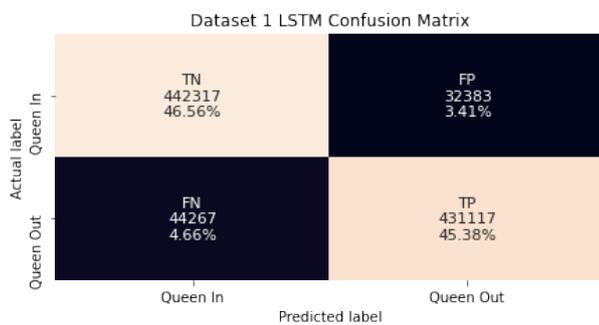
VI. CONCLUSION

Our work adds evidence supporting the value of acoustics in monitoring beehives. The best model high accuracy of 92% discriminating between Queen Less and Queen present hives, make acoustic monitoring a useful tool for beekeepers to remotely monitor the status of their hives. In future, we plan to test the current models on data from different hives and extend this work to the use of Mel spectrograms as features on a Convolution Neural Network in a bid to improve the results. In addition, we intend to investigate the use of audio monitoring to predict swarms.

REFERENCES

[1] M. L. Winston (1987) "Temperate and Tropical Honey Bees" in *The Biology of the Honey Bee,* Harvard University Press; (11 April 1991 Edition)

[2] B.R, Johnson (2009) "Division of labor in honeybees: form, function, and proximate mechanisms", Behav Ecol Sociobiol. 2010 Jan;64(3):305-316. doi: 10.1007/s00265-009-0874-7. Epub 2009 Nov 10. PMID: 20119486; PMCID: PMC2810364.

[3] R. Boys (1999) "Listen to the Bees", available on-line at https://beedata.com.mirror.hiveeyes.org/data2/listen/listenbees.htm [Last accessed 18 October 2020]

[4] W.H. Kirchner (1993) "Acoustical Communication in Honeybees", Apidologie, Vol. 24, pp 297-307

[5] T.D. Seeley and J. Tautz (2001) "Worker Piping in Honeybee Swarms and its Role in Preparing for Liftoff", Journal of Comparative Physiology A, Vol. 187, pp 667 - 676

[6] S. Ferrari, M. Silva, M. Guarino & D. Berckmans (2008) "Monitoring of swarming sounds in bee hives for early detection of the swarming period", Computers and Electronics in Agriculture, Vol. 64, pp 72-77

[7] D.S. Kridi et al (2016) "Applications of Wireless Sensor Networks for Beehive Monitoring and In-Hive Thermal Patterns Detection", Computers and Electronics in Agriculture, Vol. 127, pp 221 – 235

[8] A. Zacepins et al (2016) "Remote Detection of the Swarming of Honey Bee Colonies by Single-Point Temperature Monitoring", Biosystems Engineering, Vol. 148, pp 76 – 80

[9] G. Hunter, D. Howard, S. Gauvreau, O. Duran & R. Busquets (2019) "Processing of Multi-Modal Envronmental Signals from a 'Smart' Beehive", Proceedings of the Institute of Acoustics, Vol 41, Part 1, pp 337 - 348

[10] D. Howard, O. Duran, G. Hunter and S. Stebel (2013) "Signal Processing the Acoustics of Honeybees (Apis Mellifera) to Identify the 'Queenless' State in Hives", Proceedings of the Institute of Acoustics, Vol. 35, Part 1, pp 290 – 297

[11] http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[12] F. Zheng, G. Zhang and Z. Song, "Comparison of Different Implementations of MFCC". Available from: https://www.researchgate.net/publication/220584952_Comparison_of_Different_Implementations_of_MFCC [accessed Oct 28 2020].

[13] S. Davis and P. Mermelstein (1980) "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357-366, August 1980, doi: 10.1109/TASSP.1980.1163420.

[14] Gursimran Kour, Neha Mehan (2015), "Music Genre Classification using MFCC, SVM and BPNN" International Journal of Computer Applications (0975 – 8887) Volume 112 – No. 6, February 2015

[15] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang & H. Fan (2020), "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks", Neural Networks, Volume 130, Pages 22-32, https://doi.org/10.1016/j.neunet.2020.06.015.

[16] T. Ganchev, N. Fakotakis, and G. Kokkinakis (2005), "Comparative evaluation of various MFCC implementations on the speaker verification task Archived 2011-07-17 at the Wayback Machine," in *10th International Conference on Speech and Computer (SPECOM 2005),* Vol. 1, pp. 191–194.

[17] A. Mohamed (2014). "Deep Neural Network acoustic models for ASR", PhD thesis, University of Toronto, Canada.

[18] K. Paliwal, J. Lyons & K. Wojcicki (2011) "Preference for 20-40 ms window duration in speech analysis", Proceedings of 4th International Conference on Signal Processing and Communication Systems (ICSPCS'2010) 1 - 4. 10.1109/ICSPCS.2010.5709770

[19] https://www.inf.ed.ac.uk/teaching/courses/asr/2012-13/asr02-signal-4up.pdf

[20] S. Hochreiter, J. Schmidhuber (1997), "Long Short-term Memory", December 1997, Neural Computation, 9 (8) : 1735-80, doi: 10.1162/neco.1997.9.8.1735

[21] A. Ng, K. Katanforoosh & Y.B.Y. Mouri (n.d.). *Sequence models* [MOOC]. Coursera : https://www.coursera.org/learn/nlp-sequence-models?

[22] H. Sak, A. Senior, and F. Beaufays, (2014) "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling" In *Proceedings of INTERSPEECH*, 338-342.

[23] https://uk.mathworks.com/help/audio/ref/mfcc.html

[24] https://scikitlearn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

[25] https://scikitlearn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

[26] https://keras.io/api/layers/recurrent_layers/lstm/

[27] J.W. Picone (1993) "Signal modeling techniques in speech recognition", Proc. IEEE 81, 1215–1247

[28] P. D. Wasserman and T. Schwartz (1988) "Neural networks. II. What are they and why is everybody so interested in them now?," in IEEE Expert, vol. 3, no. 1, pp. 10-15, doi: 10.1109/64.209