Acoustic signal processing systems for intelligent beehive monitoring

Iman Ardekani (1), Soheil Pour (1) and Hamid Sharifzadeh

(1) Unitec Institute of Technology, Auckland
iardekani@unitec.ac.nz

ABSTRACT

Bees, as pollinators and producers of honey and medicinal products, play a crucial role in human life and environmental sustainability. Emerging *Smart Beekeeping* technologies utilise various methodologies in apiology, agricultural science, computer science, and electrical engineering. A significant part of these technologies includes data-driven and intelligent condition monitoring systems that can ideally imitate expert beekeepers. This paper shows that the acoustic signals generated by bees form an efficient and reliable source of knowledge about the beehive and its bee colony. Also, it proposes an acoustic signal processing system for intelligent and data-driven beehive monitoring. The proposed system includes acoustic data acquisition, noise reduction, feature extraction and machine learning techniques for inferential or predictive data analysis. This system can be used for different monitoring purposes; however, this paper focuses on queenless beehive identification. Finally, this paper reports a flexible experimental setup for developing and testing intelligent beehive monitoring systems.

INTRODUCTION

According to the Food and Agriculture Organisation of the United Nations (FAO), more than 35% of the world's food crop production is based on pollination, and more than 90% of the world's commercial pollination services are based on honeybees [1]. Therefore, honey bees affect about one-third of the world's food crop production. According to the data published by the Ministry of Primary Industries in 2021, there are more than 800,000 registered beehives in New Zealand. A significant part of the honey produced in New Zealand is exported overseas. Revenue from honey export in 2021 was 482 Million NZD, about 13% higher from 2020. However, honey is not the only product of the apiculture industry in New Zealand. This industry also generates revenue from beeswax, pollen, propolis, live bees (bulk bees or queen bees) and pollination. Therefore, maintaining the honeybee colony's health and performance is critical to the sustainability of our ecosystem.

A honeybee colony's health and performance depend on various factors; however, the main reasons for the degradation of colonies usually include environmental conditions, queen issues, diseases, and poor beekeeping practice. Beekeeping practice has a crucial role as good practice can minimise the impact of other factors. Conventional beekeeping practice usually involves regular inspections and health assessments that enable beekeepers to identify or predict any problems or upcoming events before they impact the colony. Such inspections and assessments can be conducted by opening the hive, visual observation and manual investigation. However, natural beekeepers avoid disturbing honeybees by opening their hives because it impacts their living conditions by changing the hive's temperature, humidity and carbon dioxide levels or disturbing honeybees' hormone balances. Also, opening hives may cause some bees to be accidentally crushed. Apart from the reasons mentioned above, regular manual inspection is a time-consuming and labour-intensive task that requires a high level of expertise. To overcome the challenges mentioned above, researchers have shown immense interest in developing efficient automated beehive monitoring systems for beekeepers. A relatively comprehensive literature review on this subject can be found in [2]. However, only with the advent of the Internet of Things and Machine Learning technologies did the realisation of such systems become technically, feasible, and commercially attractive.

INTELLIGENT BEEHIVE MONITORING

Several research projects for developing intelligent beehive monitoring systems have been reported; however, their scopes are usually limited. Also, they typically consider only specific applications or techniques. For example, a group of these projects tries to detect queen-less beehives using machine learning algorithms, e.g. [3] and [4]. Another group targets identifying infected bees using computer vision, e.g. [5] and [6]. In this paper, we classify the existing or potential cases of intelligent beehive monitoring systems using two approaches. The first approach considers the questions that the system tries to answer as the classification criteria to classify them into three categories:

- Health monitoring systems, including those for Varroa mite detection [5], infestation level estimation [7], pests detection etc.,
- Activity monitoring systems, including those for monitoring foraging and predicting swarming [8] or absconding, and
- Queen detection systems, including those for queen-less hive identification systems [3],[4].

The second approach considers the type of the sensors used by the system as the classification criteria to classify them into three main categories:

- Simple signal processing systems that instrument various sensors placed inside or outside the beehive to measure simple environmental or physical signals such as temperature, humidity, CO_2 level and hive weight and turn them into *simple data,*

- Acoustic signal processing systems that collect audio samples from the beehive and turn them into *acoustic data*, for example, [3], [4], [8] and [9],
- Machine vision systems that capture photographs or videos inside or outside the hive and turn them into *image data,* for example [5].

Regardless of their data types, all existing intelligent beehive monitoring systems usually apply different data analysis techniques, such as Machine Learning (ML) techniques, to extract useful information from the data. This information can then be used for automated decision-making. Today digital technology makes the use of Machine Learning techniques technically feasible and commercially attractive; thus, recently-proposed intelligent beehive monitoring systems are usually based on ML techniques. More specifically, they are based on Classification ML models because the questions they try to answer can be formulated as a classification problem. For example, queen-less hive detection systems try to answer the question of whether there is a queen in the hive or not. The answer could be either Yes or No, so the problem is a binary classification problem. Many ML models have been proposed for various classification problems, including but not limmited to Logistic Regression, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbours (KNN) models and Convolutional Neural Networks (CNN). These models can be trained using supervised learning algorithms, the ML theory's most straightforward learning algorithms. The main challenge here is creating an appropriate dataset for training the ML models using supervised learning algorithms. In this case, a good training dataset should be

- Collected carefully to cover various conditions,
- Labelled by expert beekeepers so that supervised learning algorithms can use it efficiently, and
- Noise-free or with a high Signal-to-Noise Ratio.

DATA SELECTION

In the ML theory, Data Selection is the process of determining reliable and relevant data types and sources for a given application. This section discusses why selecting acoustic data for intelligent beehive monitoring is advantageous based on three comparison criteria: Collectability, Simplicity, and Universality. Table 1 summarises the proposed comparison of the intelligent beehive monitoring systems. According to this table, the systems using acoustic data are preferred in practice. The detail of this comparison is discussed below.

Table 1. Data Selection Criteria

	Environmental	Acoustic	Image
Collectability	✓✓✓	✓✓	✓
Simplicity	✓✓✓	✓✓	✓
Universality	✓	✓✓	✓✓✓

A. Collectability

Environmental data such as temperature, humidity and CO_2 level data can be collected using simple digital or analogue sensors placed inside or outside hives. Also, environmental data is less noise-sensitive for the application considered in this paper.

Technically, collecting acoustic data from hives is more challenging for at least three reasons. Firstly, acoustic data is usually time-variant and non-stationary; hence, the data should be collected using appropriate signal sampling and conditioning techniques. Secondly, raw acoustic data is usually corrupted by noise as hives are always placed in open spaces. Thirdly, acoustic data collection requires several microphones, which can be expensive and damaged by beeswax if not installed securely. Therefore, collecting quality acoustic data is more challenging compared to the previous case. However, collecting quality image data is even more difficult because, in addition to the difficulties mentioned above, it requires control over background light, distance and orientation of the bees from the camera etc.

B. Simplicity

Environmental data, such as temperature or humidity data, is usually low-dimensional and changes slowly compared to the other types of data introduced in the previous section. Hence, the intelligent beehive monitoring systems that use environmental data are rather simple. Acoustic data collected from a hive is usually time-variant and non-stationary; thus, it should be presented using multi-dimensional models, such as time series or discrete Fourier series. This causes more complexity in the design and implementation of the system compared to the previous case. Furthermore, there are usually many acoustic noise sources around hives; thus, the collected acoustic data is highly corrupted by noise. Therefore, intelligent beehive monitoring systems that use acoustic data should include a noise reduction algorithm for preparing reliable training data. Working with image data is even more complicated because image data is high dimensional and requires advanced filtering and adjustment techniques. Also, extracting appropriate features from image data is challenging.

C. Universality

The universality of the data refers to its ability to cover all the questions that intelligent beehive monitoring systems can potentially answer. As discussed above, these questions are related to three main categories: health monitoring, activity monitoring and queen detection. Temperature and humidity data are only helpful in detecting absconding activity. CO_2 gas level data can be used for limited health monitoring. For example, it can be used for Varroa mite infection detection, but only when the infection level exceeds a high level. Acoustic data is appropriate for health monitoring and queen detection. Also, it is appropriate for detecting many activities in the hive. However, it is not helpful for monitoring some activities such as foraging or detecting pollen-bearing bees. Image data is the most appropriate because it can cover all three categories if the data is high quality.

Therefore, considering the above three criteria, this paper proposes to use acoustic data for beehives' condition monitoring systems.

Intelligent Beehive Monitoring using Acoustic Data

The block diagram of the proposed system is demonstrated in Figure 1. This system includes three main computational steps. In the first step, the input signals are enhanced using Independent Component Analysis (ICA) method. This method is usually used in signal processing for separating a signal into its additive components based on two assumptions. The first assumption is that, at most, one component is non-Gaussian, and the second assumption is that the components are statistically independent. In the second step, a feature vector is extracted by calculating Mel-Frequency Cepstrum Coefficients (MFCCs). In audio processing, Mel-Frequency Cepstrum (MFC) is a modelling technique for representing the power spectrum of a short audio signal. The coefficients of this model are usually referred to as the MFC Coefficients or MFCCs. Finally, in the third step, the extracted features are fed to a Multi-Layer Perceptron (MLP) neural network. The implementation of these three steps is discussed below.

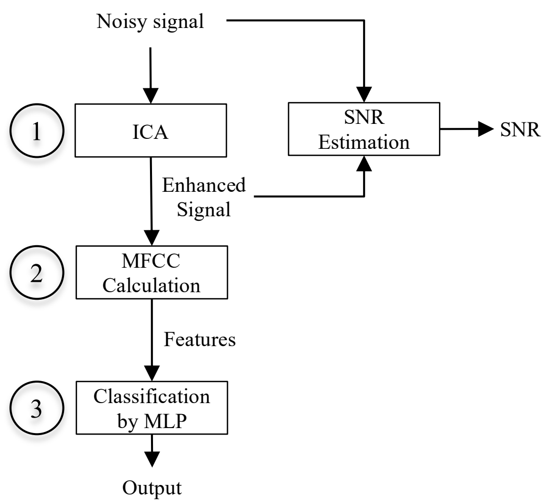

Figure 1: Block diagram of the proposed system

The original noisy signals are enhanced by using the ICA method, resulting in two signals: the primary signal and noise. Here, the signal whose power is lower is identified as the noise, and the other signal is considered the primary signal, called the enhanced signal. The enhanced signal is then used for further processing. Before calculating the MFCCs, the enhanced signal is split into short 40-milli-second frames with 25% overlapping. The power spectrum of each frame is then calculated separately. The results are then passed through a Mel filter bank, consisting of 26 filters. Each filter is modelled by an impulse response with 257 weights. The energy of the signals generated by the filter bank is calculated, resulting in a set of 26 energy values for each frame. In the next step, the logarithm of all energy values is calculated. Finally, the Cepstral Coefficients are computed by applying the Discrete Cosine Transform (DCT) to the values calculated using the logarithm function. The outputs include 26 values that are referred to as the MFCCs. Repeating the above operation for all frames results in the data required for training and testing the classification ML algorithm. This data is divided into training and testing datasets.

For this research, a Multi-Layer Perceptron (MLP) neural network is used for the classification task. The proposed structure includes two hidden layers of 6 and 3 neurons with the hyperbolic tangent activation function. Since the feature vectors are 20-dimensional, the input layer consists of 20 nodes. Also, the output layers consist of a single neuron, as the classification task is binary. After training the MLP model, the testing dataset is used to evaluate its confusion matrix and accuracy.

Case Study: Queenless State Identification

In this section, the proposed intelligent beehive monitoring system is used to identify the presence of the queen in a beehive. The raw data consists of recordings from beehives with and without queens. The sampling frequency of all recordings is 16k Hz. The raw data is labelled by the beehive status, for example, “1” for the “Queenless” state and 0 for the normal state. The labelled data is used to create 20 datasets of 1000 input frames (with an equal number of labels in each dataset).

In the first scenario, the signal enhancement task (ICA method) is bypassed; therefore, the noisy signals are directly used for the feature extraction step. In this case, the MLP model’s accuracies (one for each dataset) are calculated and visualised using the black boxplot sown in Figure 2. The average value and standard deviation of the accuracies are 81% and 5%, respectively. In the second scenario, the signal enhancement task is completed before extracting the features. In this case, the accuracies obtained for the available datasets are visualised using the red boxplot shown in Figure 2. The average value and standard deviation of the accuracies are 90% and 2%, respectively. Therefore, enhancing the signal using the ICA algorithm has improved the accuracy of the classification algorithm by about 9% on average.

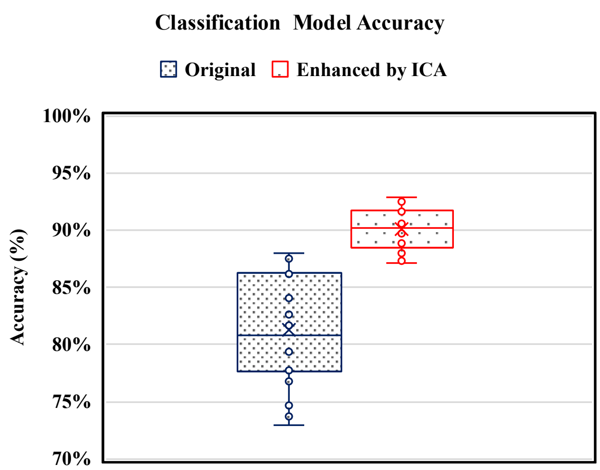

Figure 2: Accuracy of the proposed system for queenless state identification

Experimental Setup for Data Collection

This paper is based on a pilot study of a larger-scale research project to develop a comprehensive intelligent beehive condition monitoring system that uses different types of data, including acoustic and image data. The experimental setup designed and partially implemented for this research consists of two computing platforms, as demonstrated in Figure 4. The first platform is a Raspberry Pi 4 Model B single-board computer (1.5 GHz 64-bit quad-core CPU and 4 GB RAM) dedicated to collecting acoustic and image data. This board is connected to two cameras and four microphones. Two microphones (Omnidirectional 3.5mm 3-Pole: TRS) are installed on two different frames inside the hive, and two microphones are installed above the hive lid. The cameras (Resolution: 5MP 2592×1944, Frame Rates 30 fps@1080 p, M12 Lens, Adjustable and Interchangeable, Focus and Angle Enhancement) are installed on the right and left sides of the hive entrance to take photos and short videos of the bees entering or exiting periodically (from outside). The second platform, dedicated to collecting data on the hive's temperature, humidity and CO_2 gas level, is an Arduino MKR WiFi 1010 microcontroller board with Opla IoT Carrier. Also, a small weather station (WS-2902C WiFi Smart Weather Station) for monitoring metrological conditions outside the beehive is placed a few meters away from the hive in open space. The system is powered by city electricity; however, a 180W solar power system that can generate the electricity required for the experimental setup is available. The system is able to collect different types of data periodically and transfer them to a remote host or personal cloud storage using a WiFi connection. The data is also partially labelled by an expert beekeeper. For example, the beekeeper labels the data as "no queen" upon noticing that the queen is dead. The host computer uses different ML models, including classification and clustering models, to develop different algorithms for comprehensive intelligent beehive monitoring. The most time-consuming and challenging part is collecting quality labelled data for training ML models.

Figure 3: Schematic diagram of the experimental setup

CONCLUSION

Acoustic signals emitted by honeybees form a reliable data source for automatic condition monitoring of beehives. A practical signal processing system that can monitor a beehive from acoustic data must include signal enhancement, feature extraction and machine learning techniques. For example, in this paper, we propose that the ICA method can be used as the signal enhancement technique, MFCCs can be used as the features of the acoustic signals, and CNN models can be used as machine learning techniques. This combination led us to an automatic beehive monitoring system that can identify the queen's presence in the bee colony with an accuracy of 90%.

ACKNOWLEDGMENTS

We would like to thank Associate Professor Marcus Williams and his team at the Research & Enterprise Office, Unitec Institute of Technology, for funding and supporting this project. Also, we would like to thank postgraduate students Sumit Biswas, Manjit Kaur, Rui Peng, and Hangyu Wang for contributing to this project.

REFERENCES

[1]. A. Klein, B. Vaissiere, J. Cane, I. Steffan-Deventer, S. 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, 2007, pp. 303–313.

[2]. M. Abdollahi, P. Giovinazzo, T. Falk, "Automated beehive acoustics monitoring: A comprehensive review of the literature and recommendations for future work," Applied Sciences, vol. 12, issue 8, 2022, pp 3920–3942.

[3]. D. Howard, O. Duran, G. Hunter, K. Stebel, "Signal processing the acoustics of honeybees (APIS MELLIFERA) to identify the queenless state in hives," Proceedings of the Institute of Acoustics, vol. 35, 2013, pp.290–297.

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

[5]. M. Kaur, I. Ardekani, H. Sharifzadeh and S. Varastehpour, "A CNN-based identification of honeybees' infection using augmentation," The 2nd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2022, pp. 1–6

[6]. R. Tashakkori and A. Ghadiri, "Image processing for honey bee hive health monitoring," IEEE Southeast Con 2015, pp. 1–7.

[7]. A. Qandour, I. Ahmad, D. Habibi, Daryoush and M. Leppard, "Remote beehive monitoring using acoustic signals," Acoustics Australia / Australian Acoustical Society. Vol. 42, 2014, pp. 204–209.

[8]. S. Ferrari, M. Silva, M. Guarino, D. Berckmans, "Monitoring of swarming sounds in bee hives for early detection of the swarming period," Computers and Electronics in Agriculture, vol. 64, no. 1, 2008, pp. 72–77.

[9]. 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," 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019, pp. 03–08.