ASEAN Engineering Journal

Full Paper

MONITORING THE ABSENCE OF QUEEN BEE IN THE HIVE USING DEEP LEARNING AND HILBERT HUANG TRANSFORM

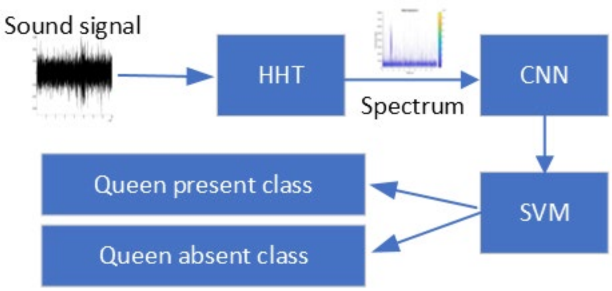
Nghien Nguyen Ba*[a], Phuong Pham Thi Kim[a], Huan Tran Thanh[a], Thang Le Anh[a], Trung Doan Van[a], Thi Thu Hong Phan[b]

[a]Hanoi University of Industry, 298 Cau Dien Street, Minh Khai Ward, Bac Tu Liem District, Hanoi City, Vietnam.
[b]FPT University, Da Nang City, Vietnam.

Article history
Received
04 April 2023
Received in revised form
25 June 2023
Accepted
07 July 2023
Published online
29 February 2024

*Corresponding author
nguyenbanghien_cntt@haui.edu.vn

Graphical abstract



Abstract

In this paper, we present a fusion method to monitor the absence of the queen bee in a hive using a combination of deep learning neural networks, support vector machine (SVM), and Hilbert Huang transform. First, we collect the sound data from the hive in the presence and missing of the queen bee using the Internet of Things system (IoT). Next, we slice the received audio signal into small chunk with a duration of 10 seconds. In the next step, we perform the Hilbert Huang Transform on each chunks to obtain the spectral image of the audio signal with and without the queen bee. Finally, we use the obtained spectral images to train and test the deep learning neural networks model combined with a support vector machine (SVM) to classify the spectral image of the audio signal with and without the queen bee. The test results on the test set achieved a classification accuracy of 98.61%.

Keywords: Hilbert Huang transform, Internet of things, Support vector machine, Deep learning neural networks.

1.0 INTRODUCTION

The queen bee plays a very important role in the colony. In the entire honey bee colony with thousands of bees, there is only one queen bee whose main task is to stabilize the colony through the secretion of pheromones and laying eggs. Beekeepers consider keeping the mother bee in good health to be a crucial responsibility. A hive without a queen means there are no more eggs or larvae, which means there are no young bees to replace the old ones [1]. As a result, the life of the whole colony will be in danger because the worker bees only live about 45 to 60 days. Detection of the nonattendance of the mother bee right time is a key task to be able to find out the cause and have an impact, create a new queen for the bee colony appropriately, thereby maintaining the life of the entire bee colony. In Vietnam, the absence of a queen in a hive is typically determined by various signs, including a significant decrease in the number of bees, the holes of the hive with no eggs, and eggs that do not neatly lie under the honeycomb hole but instead stick to the wall of the hole. Regrettably, inspections of beehives disturb the life cycle of bee colonies. Hence, monitoring bee hives non-invasively by analyzing the sound they generate is a crucial method [2]. Within their colony, honey bees utilize vibrations and sound signals for communication purposes [3-4]. The sounds made by honey bees are a result of various bodily movements, such as rough body motions, movements of the wings, and muscle contractions with high frequency without the involvement of wings, as well as pressing the thorax against substrates or other bees [5-7]. When honey bees experience stressors like the absence of a queen, they produce distinct sounds [8]. Hence, Significant research and development resources have been devoted to audio monitoring of beehives. Several studies in recent years have emphasized the close relationship between certain behaviors of honey bees and the variations in the sounds they produce [9-10]. Specifically, these studies have demonstrated a strong correlation between the amplitudes and frequencies of sounds produced by beehives and certain occurrences, such as swarming [11-14]. The authors in the article [15] have designed hardware to collect the audio signal emitted from the hive through a microphone combined with a bandpass filter which has a frequency range of 20 - 2000 Hz. Next, a Linear Predictive Coding (LPC) algorithm is used to compress the audio signal. Finally, the authors use the Support Vector Machine (SVM) learning model to classify anomalies and normal bee behavior. In the cited paper [16], the researchers developed a system for monitoring hive parameters such as temperature, weight, and sound. Additionally, an image recognition algorithm was employed to identify Varroa mites, then used a laser gun to chase or destroy them. Vladimir Kulyukin and his colleagues in paper [17] collect sounds from honeycombs using microphones and Raspberry. The recorded audio is divided into small segments. These are the samples used to train and test the deep learning neural networks model and four popular machine learning models (logistic regression, k-nearest neighbors, support vector machines, and random forests) to classify the received audio signal. The authors carried out the classification using the direct and spectral images of the audio samples. Experimental results show that the deep learning network model gives better results than the common machine learning model in both cases. Authors in [18] collect six situations of the bee's sound including: 1. A lost Queen, 2. Smoke, 3. Enemy attack, 4. Mites, 5. Lacks pollen, and 6. Normal. They use spectrogram images of the sound in six groups to train the classification models and achieved an accuracy was 99.82% in Logistic Regression. Agnieszka Orlowska and his colleagues in paper [19] compute a summarized spectrogram of the bee's sound signal that is used as the input of a deep convolutional neural network. The experimental results show that their proposal method obtains 96% accuracy on the test set. In [20] Shah Jafor Sadeek Quaderi and his colleagues use deep learning techniques such as Multi Layers Feed Forward Neural Networks, Convolution Neural Networks, Recurrent Neural Networks, and non-deep learning algorithms such as Support Vector Machine, Decision Tree, and Random Forest on the recorded sound to classify bee sound from none beehive noises. The experiment results show that SVM gives the best result in the category of non-deep learning (accuracy 80%) and Multi Layers Feed Forward Neural Networks achieve the highest performance of 100% accuracy. Hien Nguyen Thi et al. have applied Genetic Programming (GP) to recognize bee sounds and non-bee sounds [21]. The experiment results show that the GP with proper parameters setting can get better results than well-known algorithms for the classification bee sounds sample task. In [22] Christos V. Bellos et al. evaluated three methods including Support Vector Machine (SVM), K Nearest Neighbor, and U-Net Convolution Neural Networks for the classification of the audio signal of swarming and non-swarming events. The study shows that the SVM gives the best performance compared to other methods. Jaehoon Kim et al. [23] applied the classical machine learning model and deep learning model for the classification of image spectrum obtained by mel-frequency cepstral coefficients (MFCCs), and a constant-Q transform of bee sound signals. The experiment results show that the VGG-13 model, using MFCCs as input data, achieved the best accuracy (91.93%).

Currently, in Vietnam, only the Vietnam National Academy of Agriculture is researching and building a system to collect sound, humidity, and temperature data from honeycombs to conduct analysis to detect the absence of the queen bee. They publish a paper [24] that uses an advanced technique for tuning hyper-parameters of the machine learning models such as the support vector machine, the random forest, and the logistic regression, and study the new Mel frequency cepstral coefficients (MFCCs) features. The experimental results show that their proposal gives better than several deep learning algorithms for accuracy in classifying the bee buzzing from other ambient noises. Most previous proposals used the Wavelet or Short Fourie Transform (SFT) to transform from the time domain to the frequency domain for producing spectral images of bee's sound. However, these transforms are based on two assumptions that the signal is stationary and linear in the time window. Unfortunately, the bee's sound collected from a bee hive hardly satisfied two of these luxury properties. In addition, researchers applied direct Convolution Neural Networks (CNN) for the classification of the spectral images of the present and absent queen bee sound. The last layer of the CNN is the neural network layer for classification. The literature has proved that the SVM is better than a neural network layer for classification purposes. Therefore, we propose a method for improving the accuracy of the classification of bee's sound from the bee hive when the queen bee is presented and absent. The idea of our proposal is a combination of the Hilbert-Huang transform (HHT), CNN, and SVM. The HHT produces better spectral images from non-station and non-linear bee's sound signals than traditional transforms such as the FFT, and the Wavelet transform. The CNN is used for extracting features from spectral images and then the last layer of CNN is replaced by SVM for classification.

The following is the structure of the paper. The first section is an introduction, the second section presents the method and models that will be applied in the study, including the HHT, CNN, SVM, and our proposal is the combination of the HHT, SVM, and deep learning model for increasing accuracy for classification. The next section is the results and discussion, and the final section is the conclusion.

2.0 METHODOLOGY

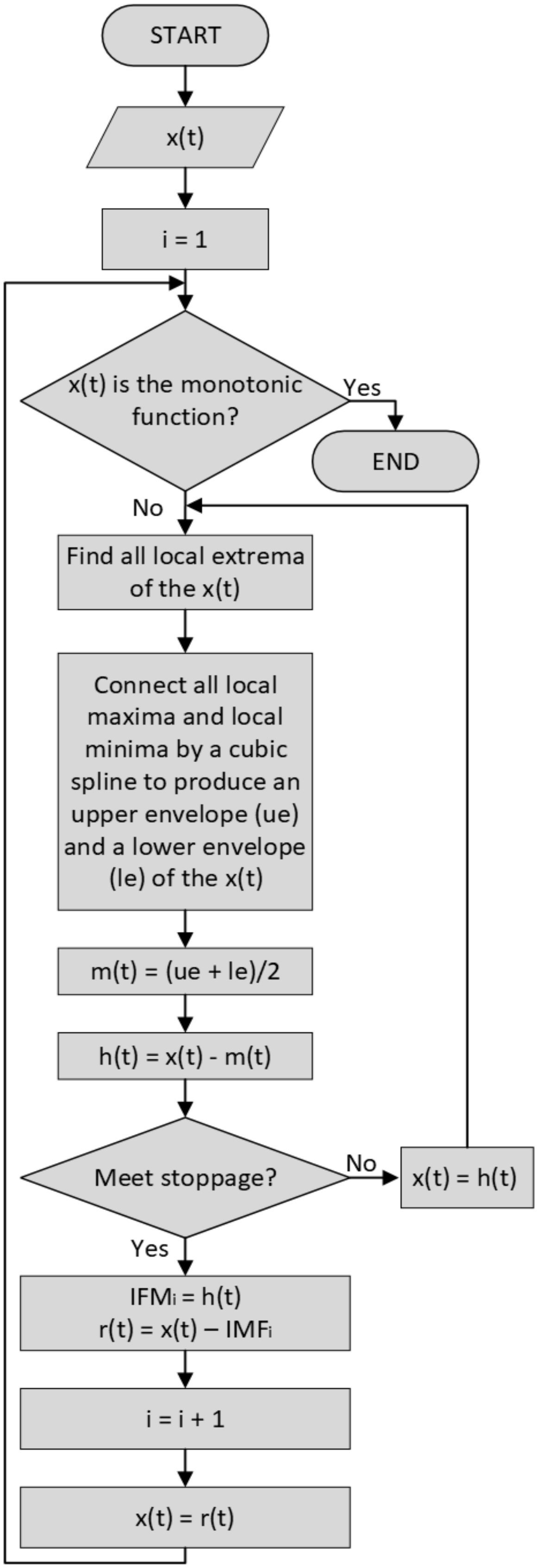
In order to calculate a spectral image of the sound signal the HHT is applied. Dr. Huang first introduced the HHT in 1998 as a method for handling non-stationary and non-linear signals. The HHT includes two phases. The initial phase is the decay of any signal into several components, and a residue is called the Empirical Mode Decomposition (EMD), and the second part is applied the Hilbert transform for each component to calculate the Hilbert spectrum. The EMD is based on a simple assumption that any signal can be decomposed into a finite number of the simple intrinsic modes of oscillations is called Intrinsic Mode Function (IMF) which has two interesting properties [25]:

1. Within the complete dataset, the quantity of extrema and the quantity of zero-crossings must be either equivalent or have a difference of one.

2. The mean value of the envelope created by the local maxima and the envelope created by the local minima is zero at all points.

Figure 1 depicts the algorithm for decomposing any given dataset x(t) into Intrinsic Mode Functions (IMFs) using the definition provided above.

Figure 1 The algorithm of the EMD process


START
x(t)
i = 1
x(t) is the monotonic function?
Yes
END
No
Find all local extrema of the x(t)
Connect all local maxima and local minima by a cubic spline to produce an upper envelope (ue) and a lower envelope (le) of the x(t)
m(t) = (ue + le)/2
h(t) = x(t) - m(t)
Meet stoppage?
No
x(t) = h(t)
Yes
IFMi = h(t)
r(t) = x(t) – IMFi
i = i + 1
x(t) = r(t)

Here are the specified stopping criteria: Dr. Huang and his colleagues [25] used the first criterion through the Cauchy-type convergence test that is defined by equation (1).

$$SD_k = \frac{\sum_{t=0}^{T}|h_{k-1}(t)-h_k(t)|^2}{\sum_{t=0}^{T}h_{k-1}(t)^2} \qquad (1)$$

If SDk is less than a predefined value, the current IMF will be found. Another criterion also was used by Dr. Huang and his colleagues [21]. This criterion is used for the zero intersection points and the extreme points. Specifically, for a preselected number C, the shifting process will stop only after C consecutive times when extreme points and zero intersections points remain constant and equal or differ at most by one. The optimal value of C should be chosen in the range of 4 to 8.

The initial data series at time t can be reconstructed by summing up all IMF and residue:

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + r_n \qquad (2)$$

Having obtained the IMFs by using the EMD method we can extract the instantaneous amplitude and frequency of k-th IMF as following steps:

Step 1: Calculate the Hilbert Transform of *$IMF_k(t)$* as

$$H[IMF_k(t)] = IMF_k(t) \times \frac{1}{\pi.t}$$

$$= P \times \int_{-\infty}^{+\infty} \frac{IMF_k(\tau)}{t-\tau} d\tau \qquad (3)$$

where *P* is Cauchy principal value.

Step 2: Form analytic signal *$z_k(t)$* as

$$z_k(t) = IMF_k(t) + j.H[IMF_k(t)] = Re_k + j.Im_k \qquad (4)$$

Step 3: Extract the phase

$$\emptyset_k(t) = \arctan\left(\frac{Im_k}{Re_k}\right) \qquad (5)$$

Step 4: Make *$\Phi_k(t)$* a monotonically increasing function by unwrapping it.

Step 5: Calculate the frequency by differentiating the phase.

$$f_k(t) = \frac{1}{2.\pi}\frac{d\emptyset_k(t)}{dt} \qquad (6)$$

Step 6: Determine the amplitude

$$a_k(t) = \sqrt{Re_k^2 + Im_k^2} \qquad (7)$$

After we calculate the Hilbert Transform for all IMFs, the original signal can be expressed as:

$$x(t) = Re\left\{\sum_{k=1}^{n} a_k(t) \times e^{j\int \omega_k(t)dt}\right\} \qquad (8)$$

where *Re* is the real part of the complex number.

This frequency and time distribution of the amplitude is referred as the Hilbert spectrum, and denoted by H(ω,t).

To extract features from an image, the CNN has been widely used. The base architecture of CNN consists of an input layer, some convolution layers, pooling layers, full connection layers, and an output layer are connected as figure 2. The convolutional layers act as noise filters and edge detectors while the subsequent sampling layers compute local averages that act as dimensionality reduction for the image. These operations also make the CNN network capable of handling distorted, rotated, or scaled input images. Fully Connected Layers normally use the Softmax or the Sigmoid activate function for classification purposes. However, the Softmax function is suitable for multi-class problems such as image classification because it provides a probability distribution over all the classes. While the Sigmoid function is suitable for binary-class problems because it does not provide a probability distribution over all the classes.

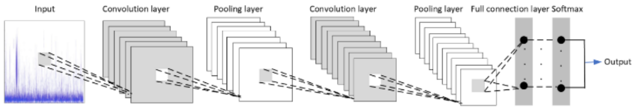


Figure 2 CNN's base architecture

In the field of machine learning, Support Vector Machines (SVM) are commonly and extensively employed for classification tasks. The objective of the SVM is to find the optimal hyper-plane to divide data into two classes and maximize distance between the closest point and the hyper-plane is called margin. Figure 3 sketches a hyper-plane and a margin.

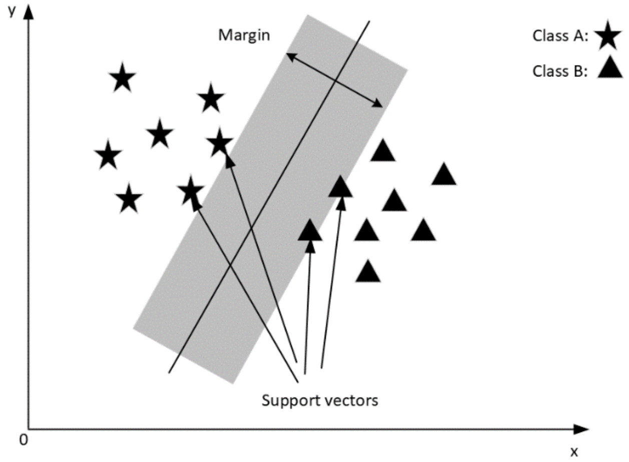


Figure 3 Sketch principle of the SVM

Assume, the training set consists of n sample *X = {(x1, y1), (x2, y2), …, (xn, yn)}*, where xi is a vector in d dimension space, and *yi* {-1, 1} is a set of the label. A hyper-plane that divides X into two regions is $w.x + b = 0$. The goal of the SVM algorithm is to find w and b to maximize the margin. To find such a tube (w, b), we solve the following optimization problem:

$$\min_{w,b,\xi}\left\{\frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\xi_i\right\} \tag{9}$$

to satisfy condition (10):

$$\mathbf{\Omega}:\begin{cases} (w,b,\xi)\,\epsilon\, R^d\ \times R\ \times\ R_+^n \\ y_i(\langle w,x_i\rangle + b)\ \geq 1 -\ \xi_i, \forall\, 1\ \leq i\ \leq n \end{cases} \tag{10}$$

where ξ_i are the slack variables added to loosen the classification condition, and C is an adjustment parameter of the model. It controls the trade-off between maximizing the margin and minimizing the training error term. Instead of solving the above problem, we often consider its dual problem as follows:

$$\min_{\alpha}\frac{1}{2}\alpha^T H\alpha - \vec{1}\alpha \tag{11}$$

with constraints:

$$\Delta:\begin{cases} y^T\alpha = 0 \\ 0\ \leq\ \alpha_i \leq C, i = 1, \dots n \end{cases} \tag{12}$$

where *y = (y1, y2, …, yn)*, $\vec{1}$ is an unit vector, and *H* is a symmetrical matrix is defined by:

$$H_{i,j} = y_i y_j\langle\phi(x_i),\phi(x_j)\rangle = y_i y_j K(x_i,x_j) \tag{13}$$

where $\langle .,. \rangle$ is a scalar product, and $\phi(.)$ is a mapping from the input space to the feature space with a higher number of dimensions to handle the case where the data is not linearly separable. The function $K(.)$ is called the kernel function and is defined as:

$$K(x,y) = \langle\phi(x),\phi(y)\rangle \tag{14}$$

The authors propose to use the combination of HHT, CNN and SVM to classify the sounds received from the hive with and without a queen. The principle of the proposed method is illustrated in Figure 4. Firstly, we collect bee sounds from the bee hive by using an IoT system depicted in Figure 6. Next, the sound signal files are sliced into small chunks. Then, each chunk is applied HHT to obtain a spectral image. The spectral images are divided into two sets including a training set (80%) and a test set (20%). These images are used for training and testing the CNN. Finally, the feature of the spectral images are input of the SVM for classification bee's sound with and without the queen bee.

The effectiveness of the CNN model depends mainly on the architecture of the network. In this article, we will propose a convolutional neural network architecture used as feature extraction. The goal is to use a network that is not too complex but just enough for feature extraction and dimensionality reduction of the image. The proposed model is depicted in Figure 5 which includes five main blocks with five convolutional layers. The convolution layer has the recommended number of filters 32, 64, 128, 256, and 512, respectively.

In each convolution layer, a ReLU nonlinear activation function of form f(x) = max(0,x) will be used to convert negative values to zero. Next, Batch Normalization (BN) helps to avoid the phenomenon that the values fall into the saturation range after passing through the nonlinear activation function. This can help reduce over-fitting.

A max pooling layer with size 2x2 is placed at the end of each block to extract useful information, remove noisy information as well as reduce the dimensionality of input image data to decrease the training time of the model.

The data after passing through the five main blocks will be flattened to put into classification. Two fully connected layers (FC) are used. The first FC layer has 512 nodes. The output of this layer is also passed to the ReLU and BN layers before being sent to the final FC layer for classification. The output layer has a size the same as the number of classes in the training set and uses the Softmax activate function.

Pseudocode of the proposed model is illustrated as following.

```
Model for classification bee's sound
Input: Bee's sound files
Output: model for classification bee's sound
imageSet = {}
featureSet = {}
1.  Slice sound file to small chunk of ten seconds
2.  Foreach chunk do
    2.1. Apply EMD process to obtain set of IMF
    2.2. Foreach IMF do
        2.2.1. Calculate Hilbert transform according to Eq. 2
        2.2.2. Determine instantaneous amplitude and
               frequency based on Eq. 3 to 7.
    2.3. Obtain Hilbert spectrum by Eq. 8
    2.4. Append spectral image to imageSet
3.  Divide imageSet into 5 folds
4.  Foreach fold k in 5 folds do
    4.1. Create a Sequential model according to proposed
         architecture depicted in Figure 5.
        4.2.  Train model on remaining 4 folds
        4.3.  Evaluate model performance on fold k and
              remaining 4 folds.
        4.4.  Store the trained model, and performance index
5.    **Foreach** img in imageSet **do**
        5.1.  Calculate output of the FC(512,ReLU) layer of the
              trained model with the img input as a feature
              vector.
        5.2.  Append feature vector into featureSet
6.    Divide featureSet into 5 folds
7.    **Foreach** fold k in 5 folds **do**
        7.1.  Assign C with a constant value such as 0.1 and
              choose a kernel function like radial basic function.
        **7.2.  Repeat**
              **For** all (xi, yi), (xj,yj) in 4 remaining folds **do**
                    Optimize $\alpha_i$ and $\alpha_j$ according to Eq. 11 to 13.
              **End for**
        **Until** $\alpha$ no change or other constrain criteria met
        7.3.  Obtain support vectors ($\alpha_i$>0)
        7.4.  Evaluate the trained SVM model performance on
              fold k and remaining 4 folds.
        7.5.  Store the SVM model and performance index for
              fold k.
```

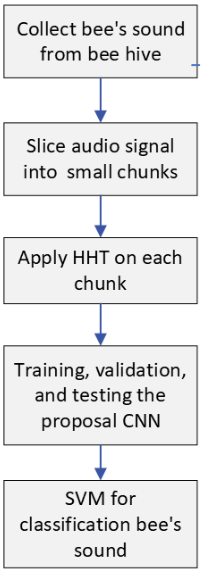

Collect bee's sound from bee hive
Slice audio signal into small chunks
Apply HHT on each chunk
Training, validation, and testing the proposal CNN
SVM for classification bee's sound

Figure 4 Illustration of our proposal

3.0 RESULTS AND DISCUSSION

We used the device depicted in Figure 6 to collect sound signals from a bee hive with and without a queen bee. We perform audio monitoring of beehives with and without queen bees from 8am to 10 am and 11 am to 2 pm on September 8, 2022, with a sampling frequency of 16,000 HZ to obtain two audio files of length 2 hours and 4 hours, respectively. Two sound file are sliced into small chunks with a duration of 10 seconds. Apply HHT for each chunk to obtain a spectral image. Table 1 summarizes datum are collected and pre-processed.

Table 1 Summary of the input data

Class	Length of sound file (hour)	Sampling frequency (HZ)	Number of chunks	Samples	Size (pixel)
Queen bees present	2	16,000	715	715	128 x 128 x 3
Queen bees absent	4	16,000	1450	1450	128 x 128 x 3

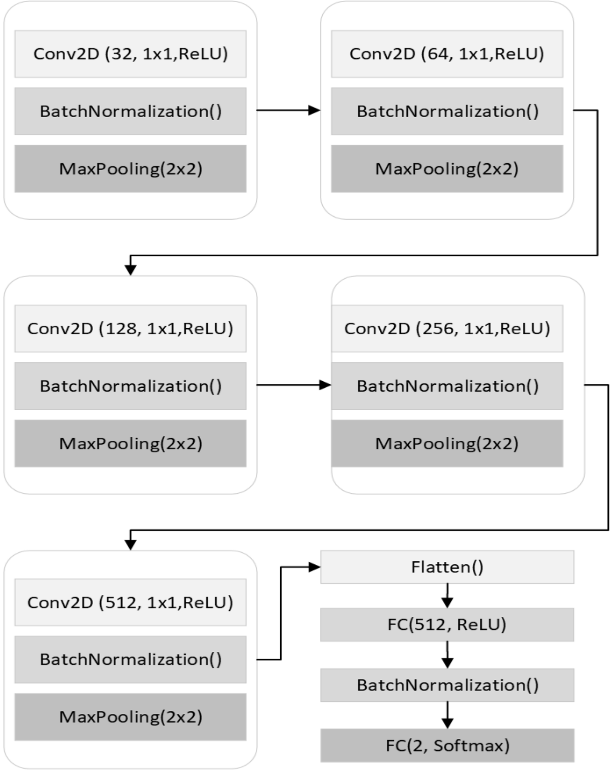


Figure 5 The proposal convolution neural networks architecture

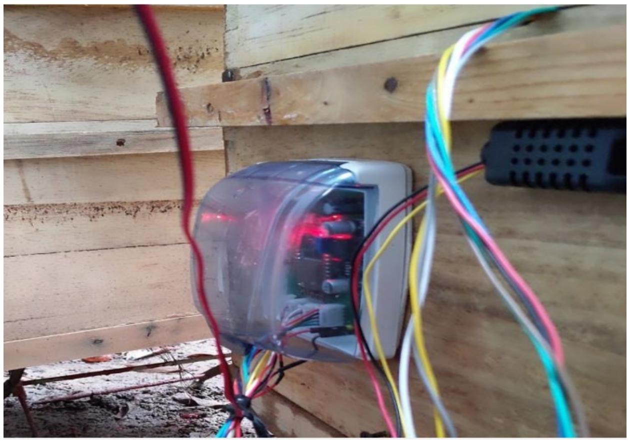


Figure 6 The IOT system for collecting the bee’s sound

To improve performance for the classification problems, we propose a fusion of CNN and SVM. Firstly, CNN plays the feature extraction role. Then, feature vectors obtained from the CNN are the input of the SVM. This combination is expected to take advantage of both models as CNN performs feature extraction very efficiently on the image, while SVM has good classification accuracy if the input data is efficiently preprocessed. The combination of the two models is illustrated in Figure 7.

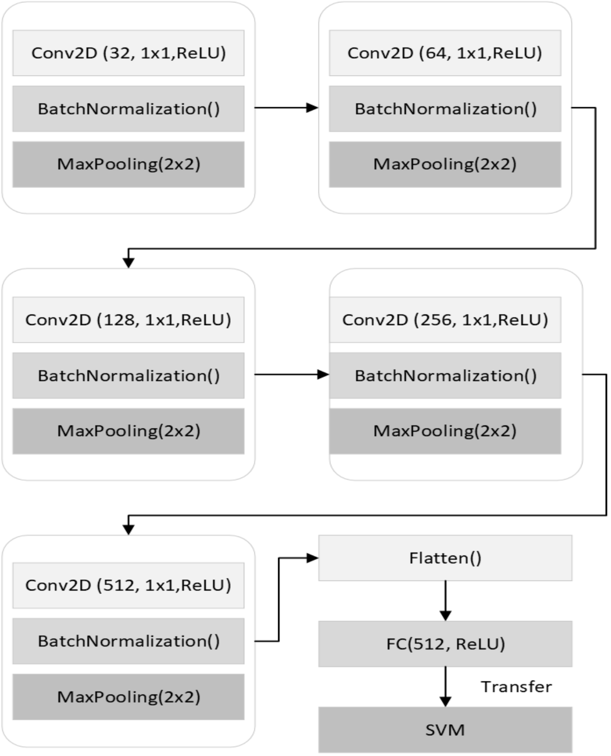


Figure 7 Fusion of the CNN and the SVM models

Figures 8, 9, and 10 illustrate the sound signal, IMFs, and spectral image of the given chunk.

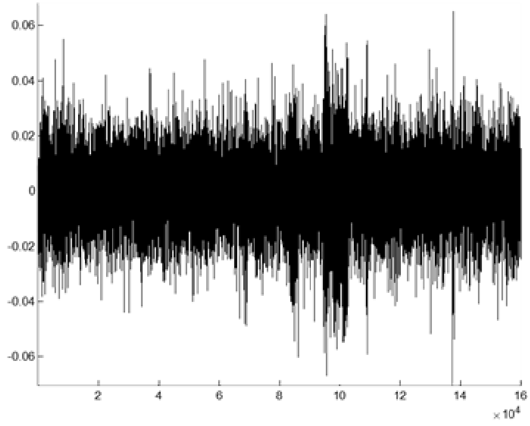


Figure 8 The sound signal of the given chunk

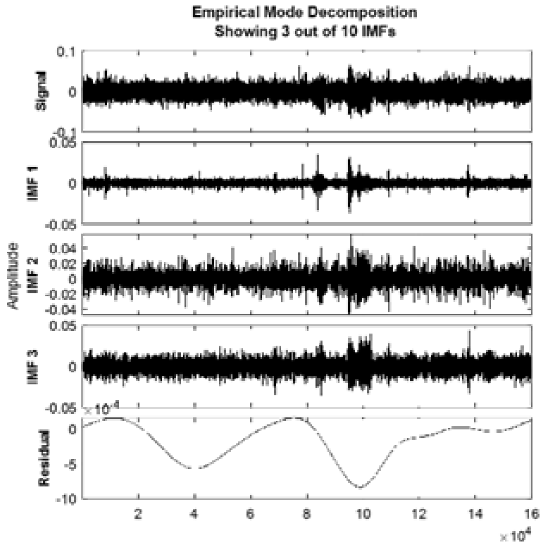


Figure 9 The first 3 IMFs and the residue of the given chunk

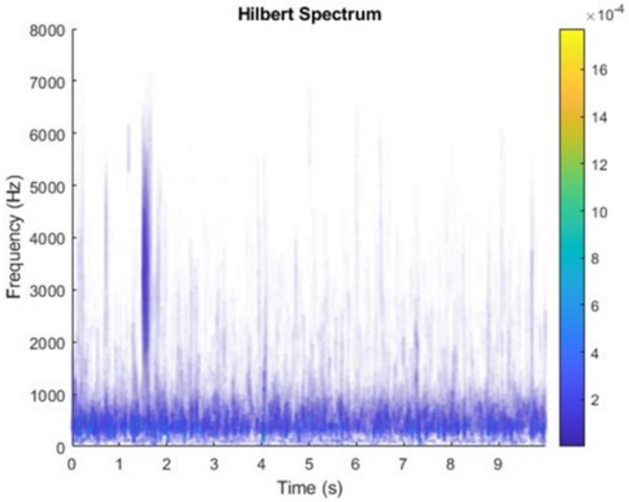


Figure 10 Hilbert spectrum of the given chunk

Our proposal CNN is trained by using Google Colab with the Python program language. This experiment uses a 5-fold cross-validation procedure to calculate the average result of each model with each training data set to give the most reliable classification results. Because the number of images of each class in the dataset is not balanced so the ClassWeight method is used to give a higher weight to the class with fewer data. The learning rate of the model is initialized to 1e-3. If the model fails to show improvement after 6 epochs, the learning rate will be decreased by a factor of 5 using the ReduceLROnPlateau module's learning rate schedule. Simultaneously, the EarlyStopping method is implemented alongside ReduceLROnPlateau to terminate the training process if the model continues to lack improvement even after 3 consecutive reductions in the learning rate. The parameters of the SVM model are specified as follows: The adjustment parameter C is selected in the set {10-1, 1, 10} respectively. The kernel function used is Radial Basic Function (RBF) with parameter σ selected in the set {10-6, 10-5, 10-4, 10-3, 10-2, 10-1} respectively. The model is set to train 200 epoch. But by using the EarlyStopping method so the actual maximum epochs in the training phase for five training folds is 137. Figures 11 and 12 illustrate accuracy and error during the training phase. The accuracy of each fold is 82.60%, 79.84%, 84.67%, 81.30%, and 80.66%, respectively for the proposed CNN model. The fusion of the proposed CNN and SVM model increases the accuracy of each fold to 98.45%, 98.59%, 98.61%, 98.37%, and 98.47%, respectively. The SVC Hyperparameters are C = 10, and σ = 0.001.

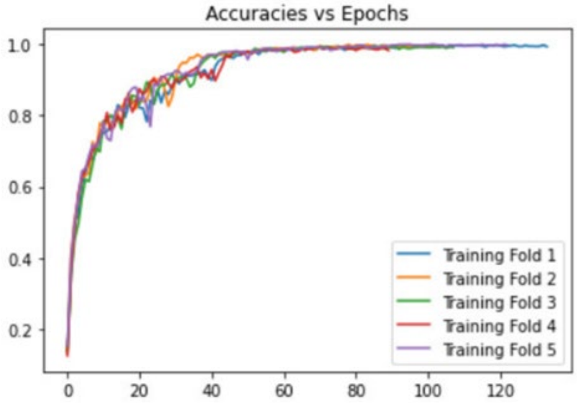


Figure 11 The CNN classification accuracy over epoch

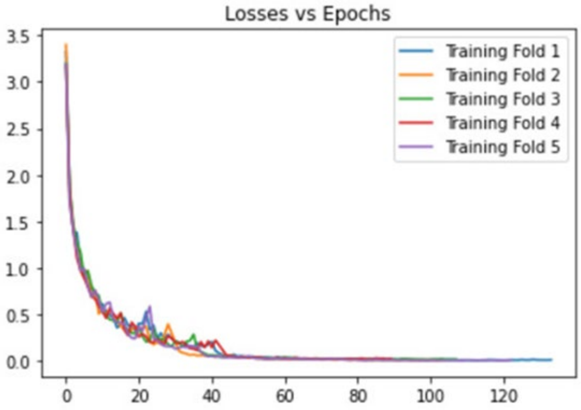


Figure 12 The CNN's error over epochs

The combination of the SVM and CNN model is experimented with hyperparameters of the SVM that are listed in Table 2.

Table 2 The accuracy classification of each model

C	σ	Training set (%)	Test set (%)
0.1	0.001	98.59	98.35
1.0	0.001	99.79	98.41
10	**0.001**	**100.00**	**98.61**
10	1e-06	95.64	95.62
10	1e-05	98.53	98.54
10	1e-04	99.58	98,56
10	0.01	100.00	98.32
10	0.1	100.00	23.21
10	1	100.00	12.18

.

To evaluate the performance of the proposed model with other models. The authors tested two more well-known models for the image recognition problem, VGG16, and MobileNetV2, which were pre-trained on the ImageNet dataset. Table 3 summarizes the experiment results.

Table 3 The accuracy classification of each model

Model	Training set (%)	Test set (%)
The proposed CNN	99.76	84.67
The fusion proposed CNN and SVM	**100.00**	**98.61**
VGG16	99.63	83.36
MobileNetV2	98.27	84.12

.

Through the results, it can be seen that the proposed CNN model achieved the same accuracy classification as the VGG16 and MobileNetV2 models. The fusion of the proposed CNN and SVM achieves the best classification performance specifically, 100% on the training set and 98.61% on the test set.

4.0 CONCLUSION

In this article, authors present the combination of the HHT, CNN, and SVM for classification sound recorded from the bee hive with and without a queen. The HTT can apply to non-station and non-linear signals. The HHT contains two phases. The first phase is EMD to decompose any signal into a set of IMFs and a residue. The second phase is applied the Hilbert transform for each IMF to calculate the Hilbert spectrum. This spectrum image presents hidden phenomena of the signal. The CNN is good at the extraction feature of the image. The base structure of the CNN consists of convolutional layers, pooling layers, and full connection layers. The convolutional layers act as noise filters and edge detectors while the subsequent sampling layers compute local averages that act as dimensionality reduction for the image. Fully connected layers normally use the Softmax function for classification purposes. Softmax function is suitable for multi-class problems such as image classification because it provides a probability distribution over all the classes. The architecture of the CNN consists of five convolution layers with kernels 32, 64, 128, 256, and 512, respectively. This choice is based on empirical observations, it has been observed that models with these filter numbers tend to generalize well across different datasets and tasks. In addition, many deep learning frameworks, and hardware accelerators are optimized for power-of-two dimensions, and gradually increasing the number of filters in deeper layers allows the network to capture increasingly complex features and patterns.

The SVM is better than a neural network layer for classification purposes. The parameter C is selected from a range of 1 to 10. Increasing the value of C enhances the classification accuracy for the training data, but there is a risk of overfitting. Similarly, the parameters σ of the Radial Basis Function (RBF) typically fall within the range of 3/k to 6/k, where k represents the number of inputs. Increasing the value of σ improves the classification accuracy for the training data, but it also increases the likelihood of overfitting. Combining the advantages of the listed method and models above, we propose the fusion method of the HHT, CNN, and SVM for classification sound obtained from a bee hive with and without the queen. The proposed solution offers several advantages compared to conventional approaches. It utilizes the Hilbert-Huang Transform (HHT) to generate a spectral image from the audio signal obtained from the honeycomb, which is advantageous for both non-linear and non-stationary signals. In contrast, conventional studies typically rely on the Short-Time Fourier Transform (SFT) or wavelet transform, which are suitable only for linear and stable signals. Consequently, using these conventional transformations on non-linear and non-stationary signals may result in the loss of crucial hidden features within the signal.

Moreover, the proposed method employs a custom-built CNN network architecture instead of utilizing existing CNN networks like those used in conventional studies. This approach reduces complexity and significantly shortens model training times, which can be lengthy with complex architectures commonly used in conventional approaches.

Lastly, the proposed solution employs a CNN network to extract features from spectral images, which are then fed into an SVM model for classification. This two-step process aims to achieve higher accuracy compared to conventional studies that directly use the output layer of the CNN network for classification.

The experimental results show that the proposed method achieves better performance compared to two well-known VGG16 and MobileNetV2 models.

Acknowledgment

This research was funded by the Hanoi University of Industry with a research project titled “Monitoring the absence of Queen bee in hive using Hilbert Huang transform and AI”, Grant Number: 23-2022-RD/HĐ-ĐHCN and the Vietnam national research project titled “Study and application on industry 4.0 technologies in the management of honey bee production for export and national consumption,” Grant Number: KC4.0-20/19-25.

References

[1] Ratnieks, F.L.: 1993. Egg-laying, egg-removal, and ovary development by workers in queen right honey bee colonies. *Behavioral Ecology and Sociobiology* 32(3): 191–198. DOI: https://doi.org/10.1007/BF00173777

[2] S. Ntalampiras, I. Potamitis, and N. Fakotakis: 2012, Acoustic detection of human activities in natural environments. *Journal of the Audio Engineering Society* 60(9): 686–695.

[3] H. Frings and F. Little: 1957. Reactions of honey bees in the hive to simple sounds. *Science* 125(3238): 122–125. DOI: 10.1126/science.125.3238.122

[4] A. Michelsen, W. H. Kirchner, and M. Lindauer: 1986. Sound and vibrational signals in the dance language of the honeybee, apis mellifera. *Behavioral Ecology and Sociobiology* 18(3): 207–212. DOI: https://doi.org/10.1007/BF00290824

[5] W. H. Kirchner: 1993. Acoustical communication in honeybees. *Apidologie* 24(3): 297–307. DOI: https://doi.org/10.1051/apido:19930309

[6] M. Hrncir, F. G. Barth, and J. Tautz: 2006. Vibratory and airborne sound-signals in bee communication. CRC Press: 421–436.

[7] J. H. Hunt and F. J. Richard: 2013. Intracolony vibroacoustic communication in social insects. *Insectes Sociaux* 60: 403–417. DOI: https://doi.org/10.1007/s00040-013-0311-9

[8] Bromenschenk, J., Henderson, C., Seccomb, R., Rice, S., Etter, R.: 2007. Honey Bee Acoustic Recording and Analysis System for Monitoring Hive Health. *U.S. Patent 7549907B2*

[9] T. Cejrowski, J. Szymanski, H. Mora, and D. Gil: 2018. Detection ́ of the bee queen presence using sound analysis. *Intelligent Information and Database Systems* 10752: 297–306. DOI: https://doi.org/10.1007/978-3-319-75420-8_28

[10] A. Robles, T. Saucedo-Anaya, E. Gonzlez-Ramrez, and C. Galvn Tejada: 2017. Frequency analysis of honey bee buzz for automatic recognition of health status: A preliminary study. *Research in Computing Science* 142(1): 89–98. DOI: https://doi.org/10.13053/rcs-142-1-9

[11] D.G. Dietlein: 1985. A method for remote monitoring of activity of honeybee colonies by sound analysis. *Journal of Apicultural Research* 24(2): 176-183. DOI: https://doi.org/10.1080/00218839.1985.11100668

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

[13] S. Ferrari, M. Silva, M. Guarino, and D. Berckmans: 2008. Monitoring of swarming sounds in beehives for early detection of the swarming period. *Computers and Electronics in Agriculture* 64(1): 72–77. DOI: https://doi.org/10.1016/j.compag.2008.05.010

[14] A. Qandour, I. Ahmad, D. Habibi, and M. Leppard: 2014. Remote beehive monitoring using acoustic signals. *Acoustics Australia / Australian Acoustical Society* 42(3): 204–209.

[15] Tymoteusz Cejrowski, Julian Szymański, Higinio Mora, David Gil: 2018. Detection of the Bee Queen Presence Using Sound Analysis. *Intelligent Information and Database Systems* 10752: 297-306. DOI: https://doi.org/10.1007/978-3-319-75420-8_28

[16] Larissa Chazette, Matthias Becker, Helena Szczerbicka: 2016. Basic algorithms for bee hive monitoring and laser-based mite control. 2016 *IEEE Symposium Series on Computational Intelligence*: 1-8. DOI: 10.1109/SSCI.2016.7850001

[17] Vladimir Kulyukin, Sarbajit Mukherjee, Prakhar Amlathe: 2018, Toward Audio Beehive Monitoring: Deep Learning vs. Standard Machine Learning in Classifying Beehive Audio Samples. *Applied Sciences* 8(9): 1573. DOI: https://doi.org/10.3390/app8091573

[18] P. Mekha, N. Teeyasuksaet, T. Sompowloy and K. Osathanunkul: 2022. Honey Bee Sound Classification Using Spectrogram Image Features. *2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering*: 205-209. DOI: https://doi.org/10.1109/ECTIDAMTNCON53731.2022.9720352

[19] Agnieszka Orlowska, Dominique Fourier, Jean-Paul Gavini, Dominique Cassou-Ribehart: 2021. Honey Bee Queen Presence Detection from Audio Field Recordings using Summarized Spectrogram and Convolutional Neural Networks. *21st International Conference on Intelligent Systems Design and Applications*: 83–92. DOI https://doi.org/ff10.1007/978-3- 030-96308-8_8ff. ffhal-03439646.

[20] Shah Jafor Sadeek Quaderi, Sadia Afrin Labonno, Sadia Mostafa, Shamim Akhter: 2022. Identify the beehive sound using deep learning. *International Journal of Computer Science & Information Technology* 14(4): 13-29. DOI: https://doi.org/10.5121/ijcsit.2022.14402.

[21] Hien Nguyen Thi, Thi Thu Hong Phan, Cao Truong Tran: 2023. Genetic Programming for Bee Audio Classification. *8th International Conference on Intelligent Information Technology*: 246–250. DOI: https://doi.org/10.1145/3591569.3591612.

[22] Kiromitis I. Dimitrios, Christos V. Bellos, Konstantinos A. Stefanou, Georgios S. Stergios, Ioannis Andrikos, Thomas Katsantas, Sotirios Kontogiannis: 2022. Performance Evaluation of Classification Algorithms to Detect Bee Swarming Events Using Sound. *Signals* 3(4): 807-822. DOI: https://doi.org/10.3390/signals3040048.

[23] Jaehoon Kim, Jeongkyu Oh, Tae-Young Heo: 2021. Acoustic Scene Classification and Visualization of Beehive Sounds Using Machine Learning Algorithms and Grad-CAM. *Mathematical Problems in Engineering* 2021: 1-13. DOI: https://doi.org/10.1155/2021/5594498.

[24] Thi Thu Hong Phan, Dong Nguyen Doan, Du Nguyen Huu, Hanh Nguyen Van, Thai Pham Hong: 2022. Investigation on new Mel frequency cepstral coefficients features and hyper-parameters tuning technique for bee sound recognition. *Application of soft computing* 27: 5873–5892. DOI: https://doi.org/10.1007/s00500-022-07596-6

[25] N. E. Huang, S. P. Shen: 2005. Hilbert-Huang Transform and Its Application. *Interdisciplinary Mathematical Sciences* 5: 1-324. DOI: https://doi.org/10.1142/5862