Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Empirical Analysis of Honeybees Acoustics as Biosensors Signals for Swarm Prediction in Beehives

Kainat Iqbal[1], Bayan Alabdullah[2], Naif Al Mudawi[3], Asaad Algarni [4], Ahmad Jalal [5,6], and Jeongmin Park [7]

[1]School of Computing, National University of Computer and Emerging Science, Islamabad, 44000, Pakistan
[2]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[3]Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia
[4]Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha 91911, Saudi Arabia
[5]Faculty of Computing and AI, Air University, E-9, Islamabad, 44000, Pakistan
[6]Department of Computer Science and Engineering, College of Informatics, Korea University, Seoul, 02841, South Korea
[7]Department of Computer Engineering, Tech University of Korea, 237 Sangidaehak-ro, Siheung-si 15073, Gyeonggi-do, Korea

Correspondence: Jeongmin Park (jmpark@tukorea.ac.kr)

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2024-RS-2022-00156326) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation). Also this research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R440), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This research was supported by the Deanship of Scientific Research at Najran University, under the Research Group Funding program grant code (NU/RG/SERC/13/30).

ABSTRACT Honeybees play a vital role in preservation of an healthy environment. Bees not only provide pollination services but also produce honey, beeswax, and royal jelly. Beekeeping has a rich history and substantial economic potential worldwide, but swarming remains a crucial challenge for maintaining profitability. Swarming, a typical colony reproductive process in honeybees, significantly impacts beekeepers profitability by lowering the number of bees in hives and thus effecting honey production. Monitoring of these beehives is therefore of paramount importance to keep an eye on their irregular behavior. Swarm prediction can be done by visually inspecting hives, monitoring temperature, or analyzing acoustic features with machine learning. Acoustic monitoring is instrumental in detecting changes in colony behavior since it overcomes the constraints of visual inspections and is not affected by external factors like temperature. In this paper, we aim to evaluate various state-of-the-art machine learning and deep learning models for swarm prediction by studying wave plot features, Mel Spectrogram, and Melfrequency Cepstral coefficients (MFCC). We use Naive Bayes, K-nearest Neighbors (KNN), and Support Vector Machines (SVM) as machine learning models and Convolution Neural Networks (CNN), Long Short Term Memory (LSTM), and Transformers as deep learning models for comparison purposes. We apply these models on a well-known honey bees audio dataset provided by the NU-hive project and consider classification metrics such as accuracy, precision, recall, and F1 score for the comparative evaluation of our models. Our evaluation demonstrates SVM as the best-performing machine learning algorithm. In particular, SVM with Mel Spectrogram as input data, achieved an accuracy of around 97%. On the other hand, CNN outperformed all the models and achieved an accuracy of 99%, using MFCC features as input data. As a result of these encouraging outcomes, we understand that our results can help the researchers to choose which AI model is more suitable for them to design beehive monitoring systems for accurate identification of abnormal situations in beehives.

INDEX TERMS Acoustics, Audio signals, Audio classification, Bee Swarming, Honey bee, Mel Spectrogram, MFCC.

I. INTRODUCTION

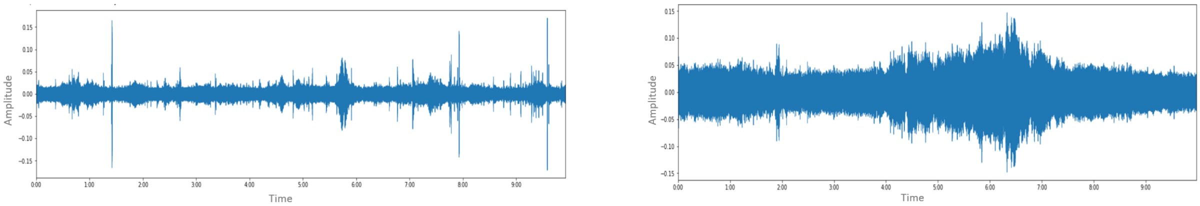
Honeybees play a vital part in the preservation of life and the health of the environment. They are not just a source of royal jelly, beeswax, and honey, but they are also essential pollinators for flowers, fruits, and vegetables. They facilitate plants in the production of seeds and fruits by transferring pollen. Considering how important honey bees are to our ecosystem and the preservation of life, their safety and health are critical. In this regard, the researchers have conceptualized bees as valuable biosensors and designed remote monitoring systems for their improved safety and health [4]. One of the key motivations in designing such a system is the success and advancements made by numerous researchers in the development of wearable and remote technologies for improvement in healthcare, [51] as seen in recent advancements in self-powered moisture detection and wearable biomonitoring systems [9]. It has been observed that a significant portion of the dead and dying colonies have several distinguishing traits, including fast worker bee loss, a notable presence of dead worker bees, and a delayed intrusion of beehive parasites. This syndrome effects the honey bee colonies, known as Colony Collapse Disorder (CCD) [48]. Colony collapse disorder includes bees leaving the colony and failing to return. Although no single cause of CCD has been found, multiple factors have been associated with it, including malnutrition, pesticide usage, viruses, mites, electromagnetic radiation, and genetically modified crops [49]. Usage of various insecticides like neonicotinoids significantly affects the living being of bees too [56]. On a similar note, CCD is related to honey bee swarming as well.

Swarming is a natural colony reproduction process in honeybees. Swarming is among the most sensitive phenomena for beekeepers, and it must be observed and detected quickly. In the past, beekeepers manually managed their hives to detect changes in honey bee behavior. Visual inspection is time-consuming and requires beekeeping skills as well. Secondly, beehives are located far away from the beekeepers, so they are unable to identify the changes that happen in the hives frequently. Many researchers have contributed to the reduction of honey bee colonies in recent years [24], emphasizing the importance of continuous and intensive monitoring to explore factors that may negatively influence the life cycle of honeybees. In this context, the integration of novel materials, such as hierarchical piezoelectric composite films, Maxwell displacement current-induced sensors, and ternary-ordered assembled piezoelectric composites, has been shown to enhance the detection capabilities of wearable and remote monitoring systems [10], [29], [44] and similarly the contribution of [7], [60] is also very important in the domain. Various research studies have been presented over the years which rely on the monitoring of the bee hive's attributes including sound, humidity, temperature, weight, and carbon dioxide [3], [14], [52]. Despite the fact that numerous monitoring systems for beehives have been proposed that rely on various sensors and measured amounts, the most effective techniques are based on sound analysis [38], [45] because the usage of various sensing devices have played a very positive and encouraging role for detecting and processing of audio signals in healthcare [32]. Some researchers have also utilized acoustics data for recognizing human activities, further emphasizing the crucial role of such data [37].

Honeybees employ vibration and sound cues to communicate inside the hive [15], [33]. Natural body movements, wing movements, high-frequency muscular contractions without wing movements, and pushing the thorax onto substrates or another bee are all used by honey bees to generate the sound [19], [20], [23]. When anomalous states like swarming or CCD occur, these behaviors change, resulting in obvious variations in acoustic features including frequency and amplitude. These differences are vital for identifying and understanding the underlying problems that impact the hive, making acoustic monitoring an important tool in beehive health management. Considering the sounds and acoustics data of the honey bees, several research studies have highlighted that honey bee behaviours are closely associated with changes in generated sound [6], [20], [35], [36]. These studies have shown that there is a strong relation between both the amplitudes and frequency of bee hive noises and events such as swarming [14], [64], [65] and the presence of queen bee [35], [36], [40]. Figure 1 illustrates the waveplots of acoustic data captured from a hive with the queen bee present and absent, respectively. The difference in amplitude and frequency over time demonstrates how the presence of the queen bee influences the acoustic signature of the hive. These observations further underscore the importance of acoustic monitoring in beekeeping, particularly in identifying critical events such as swarming or queen bee absence.

Given its large economic contributions, particularly in rural regions, beekeeping has enormous worldwide potential. For example, in Pakistan, beekeeping has a rich history and is recognized as a significant contributor to rural economies, presenting prospects for sustainable development through honey production and associated products [21]. However, the challenge of swarming and its associated challenges to profitability highlight the need for novel monitoring methods. We aim to address this issue by utilizing advances in machine learning and deep learning, as well as contribute to the ongoing development of automated beehive monitoring systems that can boost production and maintain honeybee colonies.

We observe that only a few researchers employ artificial intelligence models for beehive sound classification [36]. Considering the usefulness and effectiveness of artificial intelligence models in various real-world phenomena in this world, the objective of our research is to compare the performance of machine learning and deep learning models for swarm prediction by using audio data features like Wave plots, Mel Spectrogram, and MFCC. We use the Naive Bayes, KNN, and SVM as machine learning models and CNN, LSTM, and Transformer as deep learning models to determine the performance of audio spectrogram for the swarm prediction. In this way, our goal is to analyze and evaluate which machine learning and deep learning models are more suitable for this



(a) Waveplot of Hive Sounds in Presence of the Queen Bee

(b) Waveplot of Hive Sounds in Absence of the Queen Bee

FIGURE 1: Waveplots of Hive Sounds with and without Queen Bee

task, and more suitable to design audio-based automated beehive monitoring systems. We have performed extensive experiments using dataset from NU-hive project which is 47.7 GB in size. We understand that the development of automated beehive monitoring systems can change the beekeeping industry, with benefits including reduced manual inspections, remote monitoring of bee hives, and the ability to quickly detect the events connected to the beehive's natural cycle.

The rest of the paper is organized in the following manner. Section 2 presents the literature review of the domain. In section 3, we discuss the methods used to analyze bee sound classification. Section 4 describes the comparison results of used algorithms, whereas, conclusion is presented in section 5.

II. LITERATURE REVIEW

In this section, we explore the existing work in relevance to our study. The acoustics of honeybees have always been the topic of interest among different scientists. Below, we explore various existing approaches for the monitoring and particularly the prediction of honey bee swarming in the beehive using audio data. On the other hand, the use of machine learning, deep learning, and other contemporary approaches in various domains are also of great usefulness like mining sensory dataset [54] [62] [46] [30] [43], signal analysis [63], image processing [8] [55] [61], cloud computing [57] [31], and others [13] [18].

A. REVIEW OF TRADITIONAL ML ALGORITHMS

The use of machine learning algorithms for solving various research problems, have been shown promising results till date [53] [26] [58]. For honeybee science, Kulyukin et al. [27] used machine learning to assess the sound of the bee hives. Particularly, the classification of the honey bee sound from background sounds and cricket chirping was the main goal. Six Langstroth bee hives had four microphones installed outside their entrances. From May through July, a sound frame of 30 seconds was captured every hour. Bee hives were put in various places with a variety of background noises. The data was manually classified into three classes, honey bee sound, cricket sound, and background noise. Multiple approaches for the classification of data were tested using the collected dataset. Specifically, various traditional classifiers have been compared with CNN. These traditional classifiers include k-NN (k-Nearest Neighbors), Logistic Regression, linear kernel SVM, Random Forests, and One vs. rest classification. The results suggest that these methods are quite useful, provided some pre-processing is employed to eliminate noisy data. In the same year, Nolasco et al. [35] employed Mel Spectrograms and MFCC, derived from beehive noises using the SVM and CNN models to assess whether the beehive had a queen bee or not. The raw bee audio signals obtained from a beehive are a combination of noises provided by each bee in the colony. These signals are made up of low-frequency audio signals that are dense and continuous. In addition, in the natural beekeeping environment, they can include additional noises known as non-bee sounds, such as human talk, the sound of rain, the automobile engine roaring, and the sounds of wind. The researchers were required to label the original signals based on characteristics taken from pure bee sounds and external noise samples. The labeled acoustic signals were analyzed and subsequently categorized using ML algorithms. In this scenario, SVM surpassed CNN. SVM achieved an AUROC score of 90.1 % using a large receptive field of 30 seconds.

Similarly, [36] used machine learning for the autonomous detection and identification of the presence of a queen bee in a hive using audio data as input. This method used two approaches for the detection of queen bee presence, i.e. SVM and CNN. The importance of feature extraction before classification was demonstrated in this study. The Nu-Hive project data [5] was utilized to analyze sound from two separate colonies under normal and orphaned settings. The Hilbert–Huang Transform and Mel Frequency Cepstral Coefficients were utilized to extract features. They obtained an AUC of roughly 80%. This research, however, did not investigate various methods of feature extraction, ideal parameters, or CNN models with deep layers. The authors reported that the SVM technique outperformed CNN in terms of generalization. On a similar note, the research in [22] developed an audio data pre-processing methodology and classification model specifically for the classification of beehive noises. The main goal of this research was to test the efficacy of a classification model for beehive audio using a variety of machine learning approaches, including MFCCs, mel spectrograms, and CQT. They used five models namely random forest, SVM, shallow CNN, XGBoost, and VGG-13. The extracted features were used as input to these algorithms, and the MFCCs based models outperformed XGBoost, random forest, and SVM. XGBoost performed best with an accuracy of 87.36 percent. Secondly, when the VGG-13 and shallow CNN models were applied with image features as input data, the models based on the MFCCs pre-processing approach performed the best. It was eventually determined that the models based on CNN were more effective in identifying bee sounds after a brief pre-processing procedure.

In [65], the authors analyzed the sound produced by honey bees, using the power spectral density. The goal of the analysis of power spectral is to break down the signal into a series of weighted sinusoids. Frequency content can be determined using this decomposition. The Welch technique was used to determine power spectral density. This approach, also known as a periodogram, separates the signal into many frames and calculates the periodogram for each. The variation in power measurements is then minimized by averaging the periodograms. The goal of this research was to observe how an audio signal's frequency content varies with frequency in the time domain. Another study [40] contributed to the problem of identification of the presence of a queen bee in the hive. They explored Long Short Term Memory (LSTM), Logistic Regression, and MultiLayer Perceptron (MLP) for the detection of hives in the presence and absence of queen bee. They used MFCC as a feature extraction technique and used it as an input to the LSTM model. LSTM achieved an accuracy of 92 % and outperformed MLP and logistic regression. It's worth mentioning that using all 14 features yielded the best accuracy, whereas using only 12 features yielded the worst accuracy. The authors in [59] studied the problem of sound-based swarm detection. MFCCs and LPC were used as features in the data. The open source bee hive project (OSBH) [2] data was used. Two alternative classifier techniques were utilized, one is Gaussian mixture model (GMM) and the other one is the Hidden Markov Model (HMM). They used different classifiers and features, demonstrating that the combination of the MFCC feature extractor and HMM classifier gives the best results.

B. STUDYING DEEP LEARNING-BASED APPROACHES

Deep learning-based techniques have demonstrated decent performance in a variety of domains including speech recognition, and image and video classification. Since the emergence of deep neural networks, audio classification research has progressed from models based on hand-crafted features [41] to end-to-end models that directly translate audio spectrograms to labels [12], [47]. CNNs [28] in particular have been frequently utilized to train representations from raw spectrograms for end-to-end modeling because inductive biases like spatial localization and translation equivariance are seen to be beneficial. Despite this, for many audio processing tasks, CNN's are considered to be effective. However, the downside is that CNN can not capture the long-term context or dependency in audio data. For example, CNN's receptive field can be confined to a short window with a fixed length, hence, in this case, maintenance of long history information is not possible. Recently, the attention-based transformer model has gained extensive success for modeling sequences because of its ability to capture long-range context and have very high training efficiency.

Vaswani et al. [50] presented the Transformer model based approach which eliminates the need for recurrence and convolutions in the encoding stage altogether and rely solely on attention processes to capture the global relationships between input and output variables. As a result, the total architecture becomes more parallelized, and training time is reduced, with favorable outcomes on tasks based on almost every field of artificial intelligence. Although both convolution-based models and attention-based models have their advantages and limitations. As, transformers are good at capturing long-range context, whereas, CNNs capture local context gradually using a local receptive field layer by layer. So, a recent trend is to combine convolution-based models with attention-based models. Typically transformers are used in conjunction with CNN. In [34], the authors used a transformer on the top of CNN for the task of sound event detection to efficiently capture local and global context. Whereas, the authors of [25] used a transformer in combination with CNN for sound events detection task but for weekly labeled data. Moreover, another study [17] combines the transformer with CNN by infusing transformer in each model block of CNN. However, it is uncertain if the use of a CNN is required because neural networks based pure attention models are adequate to get high performance in audio classification tasks [16].

Conclusively, we find several research studies focusing on the acoustics data of honeybees for a wide range of aforementioned purposes. However, a detailed study investigating various machine learning and deep learning models required to evaluate the effectiveness of their application for swarm prediction is required.

III. FEATURE EXTRACTION AND DETAILS OF ALGORITHMS USED FOR SWARM PREDICTION

In this section, we discuss the materials and methods for the classification of bee swarm activity using audio data from the hive. This work is divided into audio pre-processing and audio classification. The audio pre-processing step includes the feature extraction techniques, whereas the audio classification step is further divided into machine learning and deep learning techniques. The machine learning algorithms used in this work are Naive Bayes, K-nearest neighbors (KNN), and Support vector machine (SVM). For deep learning algorithms, Convolution Neural Network (CNN), Long Short Term Memory (LSTM and Transformer Network are used. Figure 2 shows the workflow that breaks down the whole research process into various steps. The first step includes the feature extraction from the audio data using feature extraction techniques. These features are directly given as input to the model for training, evaluation, and prediction on test data for machine learning models, whereas the deep learning models themselves perform feature extraction.

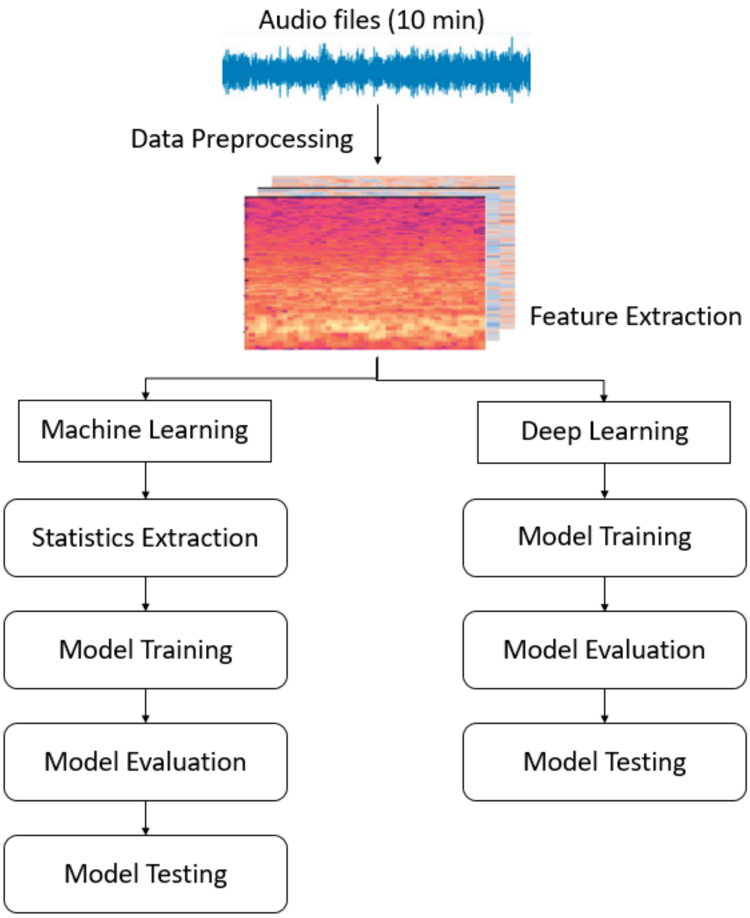


FIGURE 2: Illustration of workflow to perform classification of honeybees acoustics for swarm prediction

A. DATASET USED

In this study, we use an audio dataset collected from the NU-hive project [5]. The main purpose of the NU-hive project is to study the behavior of honeybees with the help of bee monitoring systems. The dataset contains the audio recordings of the sound of bees collected from different beehives. These beehives are located in Europe, North America, and Australia. The dataset contains two classes, one represents the normal activity and the other represents the swarm condition. There are about 576 recordings in the dataset, the duration of each recording is about 10 minutes. Almost half recordings are of the normal activity of honeybees and the rest belong to the sound of honeybees during the swarming. We split the 10-minute recordings into smaller 1-minute chunks to enhance the dataset and improve the model's learning, resulting in an increased number of training samples. Therefore, we were able to better capture the patterns of honeybee activity by obtaining a larger, more granular dataset. The total audio data that we have used has a size of 47.7 GBs.

B. DATA PRE-PROCESSING AND FEATURE EXTRACTION

A significant part of this research is sound analysis which includes audio processing. Audio data cannot be directly given as input to the model because they are not understood by the machine learning models. As a result, it is essential to extract features from audio data. It is critical to pre-process sound data before using it with different machine learning and deep learning algorithms. Since, audio data is multidimensional, with multiple frequencies that change with time, therefore, it is advisable to apply some pre-processing techniques for feature extraction. These feature values represents the signal properties. An audio signal is pre-processed into spectral features, that can be used to represent the variation in energy over frequency and time as an image. These features can easily be applied to image-based deep learning algorithms, energy modulation patterns are learned effectively and different sounds can be identified. In this study, we identify following audio features for our problem.

1) Waveplot

Waveplot shows the loudness of audio at a given time. It is used to plot the waveform of the audio signal, where the x-axis represents the time and the y-axis represents the amplitude.

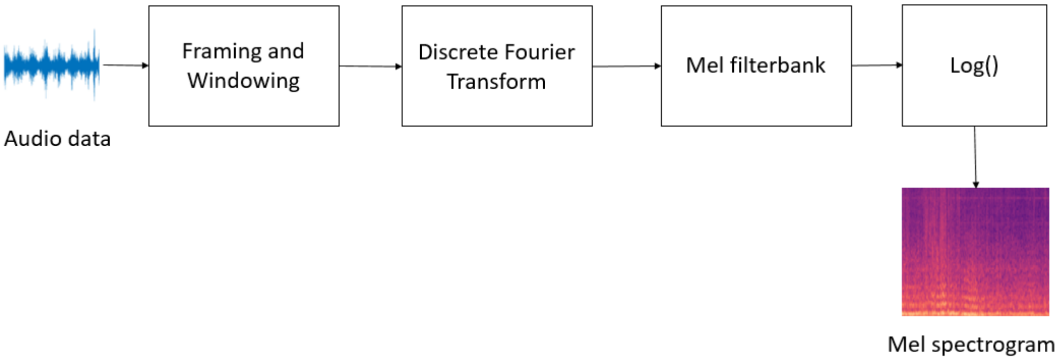


FIGURE 3: Steps to transform raw audio to Mel Spectrogram

2) Mel Spectrogram

The feature value of a Mel spectrogram is based on images that describe sound. It is mostly employed in the field of acoustic analysis [42]. Firstly, the audio stream is separated into frame-by-frame pieces, with the spectrum for each component calculated separately. The time-domain audio signal is then re-represented in the domain of frequency. The Mel scale spectrum refers to the spectrum to which the Mel scale has been applied. Figure 3 shows the steps to transform raw audio into a Mel spectrogram.

It is a combination of a waveform that visually depicts the change in the amplitude over time with a spectrum that depicts the same change in the amplitude over the frequency. Furthermore, it also indicates the color amplitude difference.

3) Mel Frequency Cepstral Coefficients (MFCC)

MFCCs are the most commonly used technique for sound processing [11]. The cepstral analysis is used to extract them from the Mel scale spectrum. The cepstral analysis takes the spectrum and extracts unique sound values. It employs the inverse of fast Fourier transform and logarithmic transformation to derive the coefficients after binding the spectrum into a constant frequency range. The correlation induced by the filter banks overlapping is separated using this method, which results in a diagonal covariance matrix. The coefficients containing a considerable data are left at the end. This ensures that the final MFCCs are robust against fast signal shifts. Figure 4 shows the steps to transform raw audio to MFCC.

C. DETAILS OF MACHINE LEARNING ALGORITHMS USED

We use Naive Bayes, KNN, and SVM as machine learning models to determine the performance of audio spectrogram for the prediction of bee hive states like swarming. In this section, we now present brief details of these models.

1) Naive Bayes

Naive Bayes is a simple conditional probability-based machine learning model which is used for the classification task. This classification algorithm is based on the Bayes rule. The probability model of the Naive Bayes classifier is shown in equation 1

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \tag{1}$$

Naive Bayes can characterize the likelihood of an event depending on previous knowledge of the circumstances of the event.Naive Bayes was selected as a baseline model because of its simplicity and computational performance in classification tasks, making it ideal for quick prototyping in the early stages of the research. Moreover, its probabilistic approach improves interpret-ability by helping in the identification of key audio features of swarm activity and providing a helpful comparison to more complex models.

2) K-Nearest Neighbors

K-nearest neighbor (KNN) is a supervised learning model that uses proximity to generate classifications about the grouping of certain data points. k in the KNN algorithm represents the number of nearest neighbors selected to cast a 'vote'. Various values for k can produce different classification results for the same example item.

It may be used for both classification and regression tasks, although, it is most usually used as a classification technique, with the assumption that similar points may be found close together. For the prediction of a correct class of the testing data, KNN calculates the distance of testing data from all the training points. k points closest to the testing data are selected then. It evaluates if the testing data belongs to one of the class of k training data and selects the highest probability class and is a simplified form of a Naive Bayes classifier. KNN algorithm, unlike the Naive Bayes classifier, does not need the use of probabilities.

KNN was chosen because of its instance-based learning technique, which is effective for detecting local patterns in audio data. It also offers a non-parametric approach to analyze the relationship between acoustic features and swarm behavior. Its instance-based learning approach allows it to explore complex decision boundaries in swarm prediction while taking advantage of the high dimensionality of the dataset.

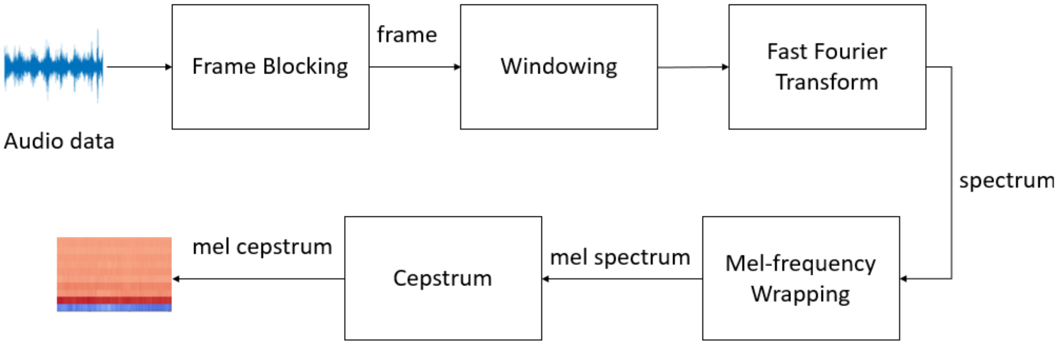


FIGURE 4: Steps to transform raw audio to MFCC

3) Support Vector Machine

SVM is a supervised machine learning algorithm that can handle both classification and regression issues, whereas it is primarily used for classification tasks. Each data item is shown as a point on n-dimensional space whereas, the number of features is depicted as n here. Each feature value is represented as a coordinate value in the space. Classification is then performed by identifying the best hyperplane that differentiates between the class labels. As many hyper-planes can help distinguish the class labels. The objective is to the find the best possible hyperplane that has the greatest distance between the points plotted in the space for each class and has the greatest margin between them. Hyperplane with maximum margin helps to classify the data points with more accuracy. The hyper-planes are the markers or decision boundaries which classify the data points. The size of the hyperplane is dependent on the number of features, n. Hyperplane must be a line when the number of features is $n = 2$. When n increases, the number of dimensions also increases, for instance, if $n = 3$, the hyperplane becomes 2 dimensional, and so on.

SVM can classify both types of data (linear and nonlinear). It can easily classify linear data by creating a linear hyperplane. To classify nonlinear data, SVM kernel trick is the solution, i.e. it changes the input space from low dimension to high dimensional data. For this purpose, it transforms the linearly nonseparable problem into a separable problem. Before deciding how to partition the data based on the labels specified, it performs several complex data transformations. It is well-known for its ability to handle high-dimensional spaces and is most effective when classes are separated. It works well with audio data because distinct patterns in the sound spectrum can be identified for swarm prediction.

D. DETAILS OF DEEP LEARNING ALGORITHMS USED

We use CNN, LSTM, and the Transformer Network as representative deep learning models for swarm prediction.

1) Convolution Neural Network

Convolution Neural Network (CNN) is a Deep Learning algorithm that accepts an image as an input, assigns a value to different objects of the image, and distinguishes between them. The values assigned are biases and learnable weights. CNNs work by convolving input with learnable kernels. A 1-dimensional temporal or a 2-dimensional time frequency convolution is widely used for spectral input features, however, for raw waveform inputs, a time-domain 1-dimensional convolution is commonly utilized. Multiple feature maps (channels) are commonly computed using a convolutional layer, each from its kernel. The learned feature maps can be down-sampled by adding pooling layers on top of the convolution layers. A CNN typically consists of a sequence of convolution layers separated by pooling layers, topped with one or more dense layers. To get a fully-convolution network (FCN) for sequence tagging, the dense layers might be deleted. The architecture of a CNN determines its receptive field (the number of samples or spectra used to compute a prediction). It can be raised by stacking additional layers or utilizing larger kernels. Reaching an appropriate receptive field size, especially for raw waveform inputs with a high sample rate, may result in a large number of CNN parameters and considerable computing complexity. Figure 5 shows the architecture of CNN with multiple hidden layers.

CNN was chosen because of its powerful ability to extract spatial hierarchies of characteristics from audio data when converted to formats like Mel spectrograms and MFCC. The success of CNNs in image recognition tasks can be applied to audio data, where the visual representation of acoustic features enables effective pattern detection.

2) Long Short Term Memory

Long Short Term Memory (LSTM) network is a kind of recurrent neural network (RNN) that can learn order dependencies in sequential prediction tasks. LSTMs are complex deep learning models and are more complex than sequential RNNs that allow for the storage of information. It is capable of dealing with RNN's vanishing gradient problem. Consider an example if we are viewing a video and recalling the previous scene, or we are reading a book and remembering the events that occurred in the previous chapter. Similarly, RNNs recall earlier knowledge and utilize it to process the present

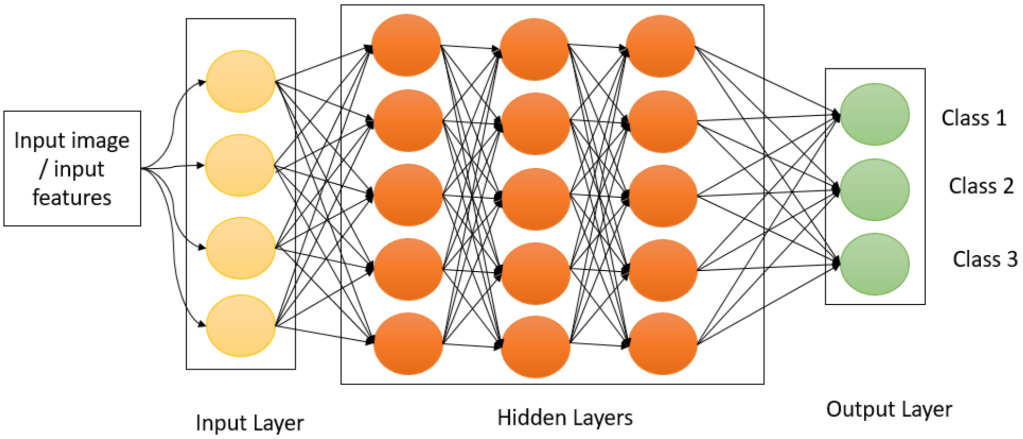


FIGURE 5: A simplified illustration of CNN architecture for classification task

input. The vanishing gradient problem prevents RNNs from remembering long-term sequences. Long term dependence concerns are expressly avoided in designing LSTM. At a high level, LSTM works similarly to RNN. The LSTM architecture is divided into three sections, each with its function. The first component determines whether the content from the previous timestamp should be remembered or deleted. The second component of LSTM learns the new information from input data. Lastly, the third component provides updated information to the subsequent time stamp from the previous time stamp. These three components of LSTM are referred as gates. The Forget gate is the initial component, followed by the Input gate, and finally by the Output gate.

LSTM is more like simple RNN except for the long term dependency. It contains hidden state from the previous time stamp and the current time stamp as well. As shown in Figure 6, H(t-1) represents the previous time stamp whereas, H(t) represents the current time stamp. Cell state is also part of the LSTM, which is represented by the C(t-1) previous and C(t) current time stamp. The hidden state in the LSTM is termed as short-term memory, whereas long-term memory is the cell state.

Long-range dependencies can be captured by LSTMs, which makes them very useful for time-series data. Bee sounds have temporal dynamics that can be used to anticipate swarms, and since audio data is sequential, LSTMs are a good fit for modeling these dynamics.

3) Transformer Network

A transformer is a deep learning algorithm that employs the self-attention process to assign distinct weights to each data input fragment. Transformers, like RNNs, are meant to analyze data, such as natural language, with applications in text summarization and text translation. Unlike RNNs, however, transformers process the full input at once. Any place in the input sequence is given context by the attention mechanism. If the incoming data is a natural language phrase, for example, the transformer does not need to parse each word individually. This allows for greater parallelization than RNNs, resulting in faster training.

The Transformer design is based on an encoder-decoder structure; however, it generates an output without using recurrence or convolutions. In a nutshell, the encoder's function is to encode an input sequence into a series of continuous representations that are then fed into the decoder on the Transformer's left half. The encoder output is combined with the decoder output from the previous time step by the decoder on the right half of the architecture to form an output sequence. Figure 7 shows the architecture of the Transformer model.

The attention mechanism in a transformer is a critical element of it. The value of other tokens in an input for the encoding of a specific token is represented by the attention mechanism. In a machine translation model, for example, the attention mechanism allows the transformer to convert terms like ”it” into a gender-appropriate French or Spanish word by paying attention to all relevant words in the original text. Importantly, the transformer's attention mechanism allows it to focus on specific words to the left and right of the current word to determine how to translate it. The launch of the transformer-based model has almost replaced the usage of RNN and LSTM neural networks because of their lower accuracy.

The Transformer model, well-known for its attention mechanism, provides a more sophisticated technique to capture long-range dependencies in sequential data without relying on recurrence. This model was chosen to investigate its utility in audio classification tasks, particularly in dealing with the complicated temporal correlations in bee sounds.

IV. EXPERIMENTAL EVALUATION

In this section, we present the empirical analysis of our research. We performed on the experiments on Google Colab. Below, we present detailed experimental evaluation for the three machine learning algorithms i.e. Naive Bayes, K-Nearest Neighbors, Support Vector Machines and three deep learning algorithms Convolution Neural Network, Long Short

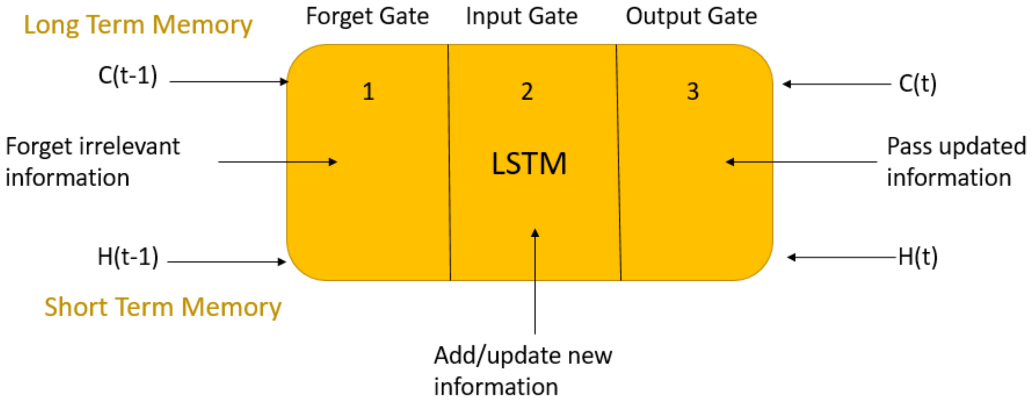


FIGURE 6: A simple illustration of LSTM architecture

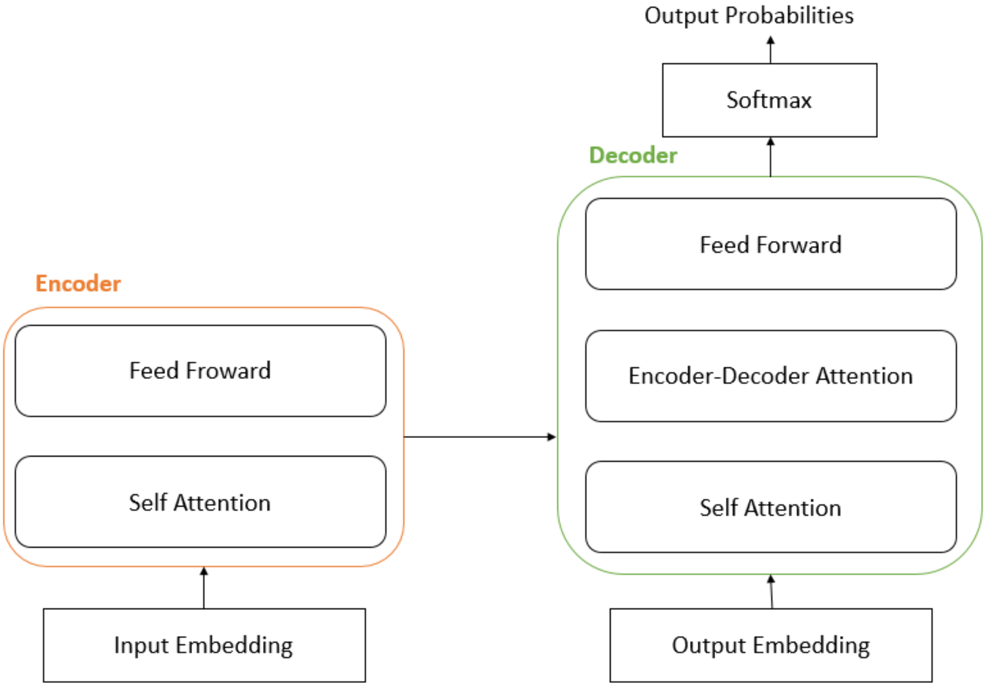


FIGURE 7: A simple illustration of Transformer architecture

Term Memory and Transformer Network.

A. EXPERIMENTAL SETUP

In this study, we performed many experiments using the features extracted from the audio data of honey bees. We focus on three important acoustics features of Waveplots, Mel Spectrogram, and MFCC which are not considered in existing studies for bee swarm prediction. In addition to this, we also compute sum of the amplitude of the waveplot as our fourth feature. We have used a Python package, known as librosa, for sound analysis and feature extraction. Librosa implements a wide variety of audio features and hence serves as a foundation for the development of audio classification tasks [1].

We developed all the machine learning and deep learning models using a variety of parameters. To generalize diverse patterns of the data or to have improved prediction, the machine learning algorithm uses different parameters for learning rate and requires problem-specific tuning of weights. These parameters are known as hyper-parameters and they must be tuned to obtain better result out of the model. We used Scikit-learn library in Python [39] for machine learning algorithms. It provides a package for the hyper-parameter tuning of the models known as Grid search, which takes a sample of parameters performs the exhaustive search on all the combinations of parameters, and returns those parameters which produce the best results. Finally, for the evaluation of models, we used two evaluation techniques i.e. train/test split and k-fold cross validation.

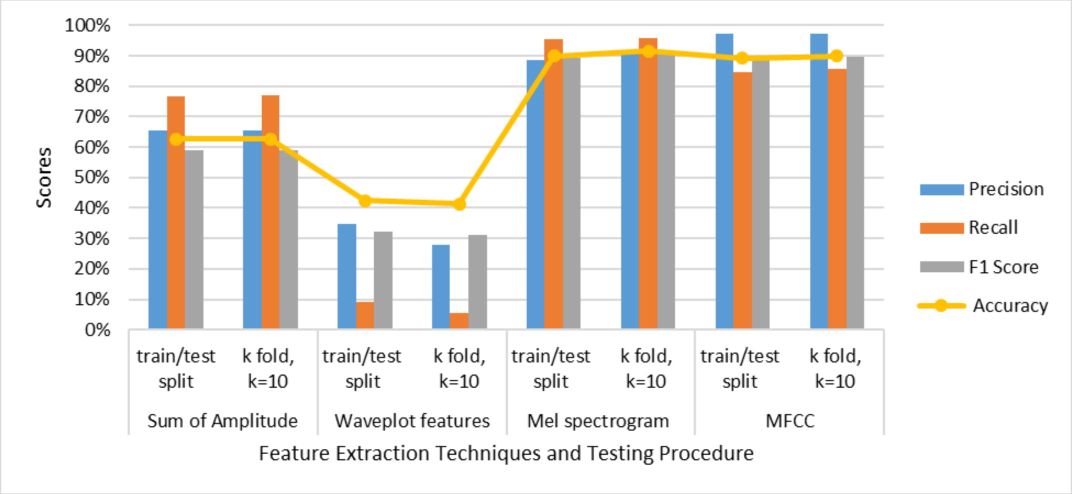


FIGURE 8: Performance Comparison of Accuracy, Precision, Recall and F1 Score for KNN

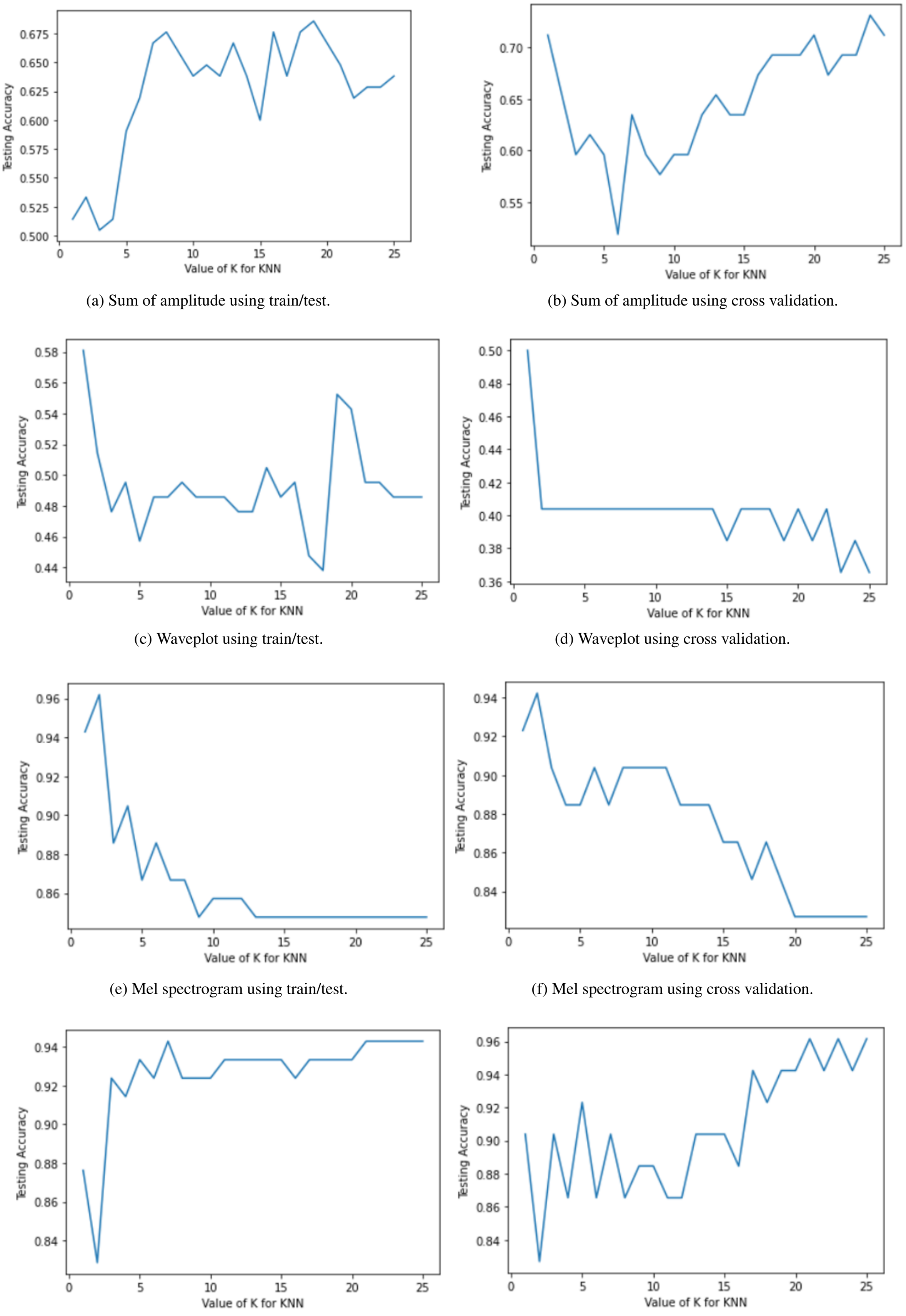
B. EXPERIMENTAL EVALUATION USING MACHINE LEARNING ALGORITHMS

In this section, we present our evaluation using the three machine learning algorithms used i.e. kNN, Naive Bayes, and SVM. As stated above, we trained the models using scikit-learn package. Scikit-learn is a machine learning Python library that supports a variety of machine learning (supervised and unsupervised) algorithms.

1) Results and Discussion for Swarm Prediction using K-Nearest Neighbors, kNN

In the empirical evaluation of kNN, the value of k for each experiment is in the range of 1 to 26. For each value of k, audio data is provided as input to the model in the form of waveplot features, sum of amplitude features, mel spectrogram, and mfcc. For the train/test split technique, each experiment runs 5 times, and classification accuracy is averaged to get the final accuracy. Figure 8 shows the performance comparison of Accuracy, Precision, Recall, and F1 score for each experiment. It has been observed that kNN provides higher accuracy with Mel spectrogram and MFCC features as input, compared to Waveplots and Sum of amplitude of waveplot features. The Mel spectrogram and MFCC features achieve almost 90% accuracy in both train/test split and k-fold cross-validation approaches, outperforming waveplot features, which have a score of roughly 40%. Similarly, the Precision, Recall, and F1 scores continually indicate higher scores for the Mel spectrogram and MFCC, demonstrating their efficiency. Waveplot and Sum of Amplitude features, on the other hand, perform significantly low, with waveplot features scoring less than 50% across all metrics, implying that they are less effective at capturing the key characteristics of bee sound data than Mel spectrogram and MFCC. The reason is that the former pair of features capture more relevant information in the acoustics compared to the latter ones. In the beehives, the acoustics of all the bees are compared to the bees signaling that they swarm or not swarm. Moreover, there is also an external noise i.e. wind, atmosphere, etc which also needs to be taken care of. In this regard, both the features are capable of making use of the right kind of information. In addition to that, Mel spectrogram provides a frequency representation of the signal over time whereas MFCC tells the spectral features, hence they help the model to better distinguish between classes to predict. Moreover, both are more robust to noise and variations in the acoustics, hence clearly segregating between the noise of the bees from the outside noise of the hives.

In Figure 9, we show detailed results for the testing accuracy using train/test split and k fold cross validation with the different values of k for kNN, where we side-by-side demonstrate the results for the two types of approaches. We observe an interesting set of results i.e. the accuracy using Sum of amplitude of Waveplot and MFCC increases for both types of evaluations, with an increase in value of k for kNN however, it goes the other way around for Waveplot and Mel spectrogram features. Generally, the train/test approach is simpler and is computed efficiently. However, k-fold cross validation is often preferred because it uses multiple train/test splits and then averages the results, hence providing better accuracy. On comparing the accuracy drop in the Mel spectrogram but the increase in MFCC, we notice that MFCC features are low dimensional compared to the Mel spectrogram. Hence, when an increase in the value of k in kNN happens, kNN finds it hard to manage high dimensional features of the Mel spectrogram, thus providing lesser accuracy, because KNN in essence is prone to the curse of dimensionality problem. On a similar note, sum of amplitude features is lower-dimensional compared to Waveplot features, hence, we find an increase in accuracy in Figure 9 (a) and (b) for sum of amplitude whereas drop in (c) and (d) for Waveplot. However, Waveplot is more sensitive to noise compared to the sum of amplitude, making KNN yield lesser accuracy.



(a) Sum of amplitude using train/test.

(b) Sum of amplitude using cross validation.

(c) Waveplot using train/test.

(d) Waveplot using cross validation.

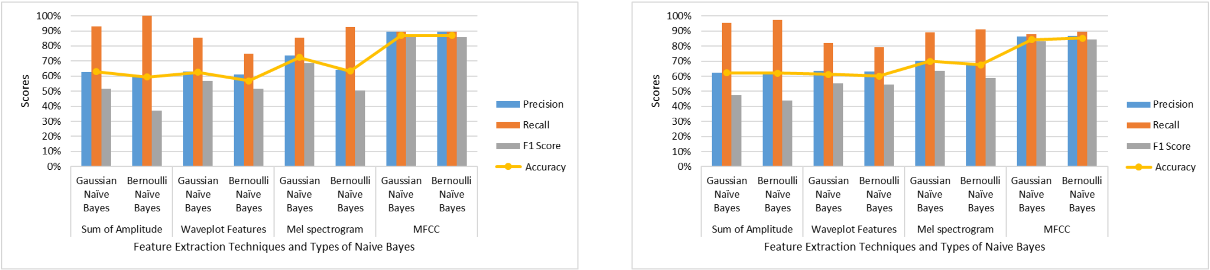
(e) Mel spectrogram using train/test.

(f) Mel spectrogram using cross validation.

(g) MFCC using train/test.

(h) MFCC using cross validation.

FIGURE 9: Comparison of testing accuracy with different values of k in KNN utilizing train/test and k fold cross validation approaches



(a) Train/test split with 5 runs.

(b) K-fold cross validation, having $k = 10$.

FIGURE 10: Performance Comparison of Accuracy, Precision, Recall and F1 Score for Naive Bayes

2) Results and Discussion for Swarm Prediction using Naive Bayes

We have implemented Naive Bayes for all the aforementioned extracted features. We run each experiment 5 times for train/test split and average the accuracy at the end. We implemented two types of Naive Bayes i.e. Gaussian Naive Bayes (GNB) and Bernoulli Naive Bayes (BNB) with all the feature extraction techniques and evaluation techniques as well. Figure 10 shows the performance comparison of Accuracy, Precision, Recall, and F1 Score for each type. It has been observed that the model does not perform well using waveplot features and the sum of amplitude technique, as all the performance metrics are staying below 60% for GNB and BNB. Whereas, the accuracy with precision, recall, and F1 score improves in the case of Mel Spectrogram and MFCC. In particular MFCC feature is serving to provide higher scores for all performance metrics surpassing 80% than rest of the features, for both types of model evaluation techniques. This pattern remains consistent for both evaluation methods: train/test split (Figure 10(a)) and k-fold cross-validation (Figure 10(b)). Mel Spectrogram and MFCC features better capture the relevant information from the acoustics of the bees, hence, providing higher accuracy.

It is observed that GNB outperforms BNB due to the presence of continuous features, which align better with the former algorithm. The situation with MFCC is interesting as BNB shows comparable accuracy in 10(a) and slightly superior performance in 10(b). It is interesting to observe that BNB is competing well and offering a level of accuracy that is not significantly different. One of the reasons is that although the values of all the features are continuous in nature, being an ideal fit for GNB, their exact distribution might not be Gaussian in nature. In addition to that, the features independence aspect of Naive Bayes has a major impact during the calculations. When the features are independent of each other, then BNB provides comparable accuracy to GNB.

3) Results and Discussion for Swarm Prediction using Support Vector Machines

For the experiments using SVM, the model requires some hyperparameters which needs to be tuned. As explained above, we used Grid search techniques to achieve optimal parameters. We find C = 0.1, kernel = ’poly’, degree = 1, and gamma = 1, as the best hyperparameters using Grid search, from the following range of parameters:

C = [0.1,1,100,1000]
kernel = [’rbf’, ’poly’, ’sigmoid’, ’linear’]
degree = [1,2,3,4,5,6]
gamma = [1, 0.1, 0.01, 0.001, 0.0001]

Figure 11 shows the performance comparison of Accuracy, Precision, Recall, and F1 Score for each experiment using various acoustics features. We observe that SVM gives the best classification accuracy with the mel spectrogram feature. Furthermore, SVM obtains good Precision and Recall with the Mel spectrogram, resulting in an overall strong F1 Score, demonstrating its efficacy in managing both false positives and false negatives in classification. Mel spectrograms possess higher dimensionality than MFCC, and SVM performs well with high-dimensional data. Mel spectrograms show the frequency spectrum throughout the duration. They retain more information about the frequency content of the signal across time. Hence signals contain more information, making SVM do better classification. On a similar note, the variation in the acoustics of the bees is for swarm and non-swarm, is sometimes quite complex, primarily because of various factors like changes in environment due to hives being parks in different locations after a certain time among others. So hence, there is a variation in the sounds inside and outside the hives as well. So for such complex tasks, Mel spectrograms incorporate more relevant features as compared to MFCCs. The performance of SVMs gets better when it comes across non-trivial decision boundaries in high-dimensional data. In this regard, richer information is provided by the Mel spectrogram as compared to MFCC, hence, we get superior performance using them. Moreover, Mel spectrograms are more sensitive to noise, clearly differentiating between the acoustics of the bees and the noise outside the beehives.

C. EXPERIMENTAL EVALUATION USING DEEP LEARNING ALGORITHMS

In this section, we present our experimental evaluation using CNN, LSTM, and the Transformer models. We have implemented these algorithms using Keras TensorFlow library in Python, in multiple convolution layers. For the hyperparameter tuning of the model, learning rate of 0.0001 and 0.00001 is used and the ’adam’ function as an optimizer.

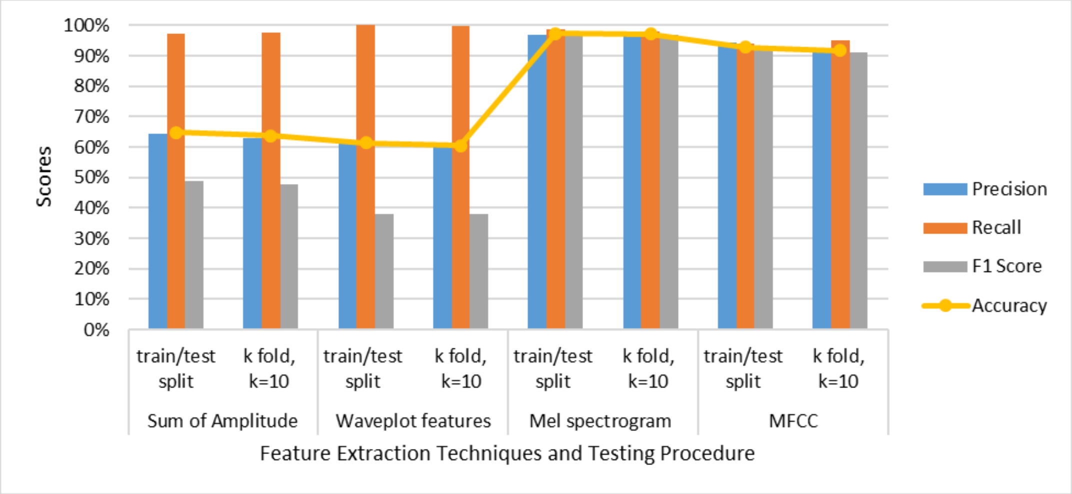


FIGURE 11: Performance Comparison of Accuracy, Precision, Recall and F1 Score for SVM

Each experiment is tested using the two values of epochs, i.e. 20 and 50. For the experiments using the Transformer model, we executed the algorithm for 100 epochs. We used only Mel Spectrogram and MFCC as features of bees acoustics, because of the reason that both the other features (sum of amplitude and waveplot) does not contain much useful contents to serve as a motivating factor for highly accurate classification, as witnessed when they are used in machine learning algorithms above.

1) Results and Discussion for Swarm Prediction using Convolution Neural Network

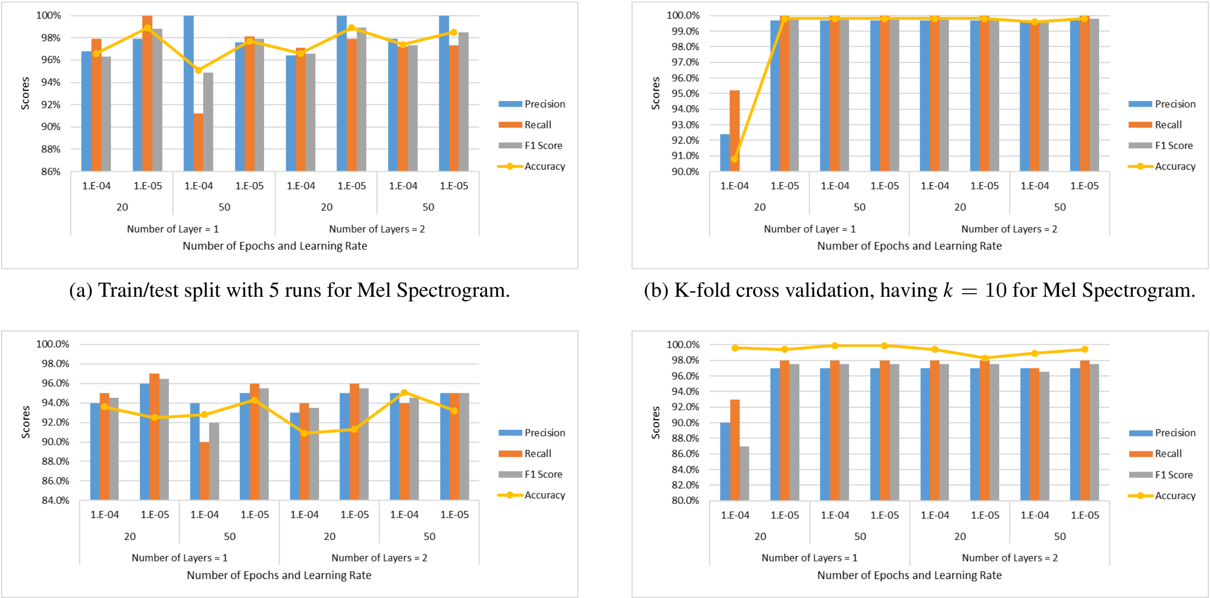
Figure 12 shows the performance comparison of Accuracy, Precision, Recall, and F1 Score for the CNN model with different learning rates and number of epochs. Overall, we find that the CNN provides higher accuracy for classification, for both of the features i.e. Mel Spectrogram and MFCC. Specifically, it has been observed that the Mel Spectrogram yields higher accuracy in the results. Utilizing Mel Spectrograms is preferable as they provide more details of the audio data in order to capture the entire frequency spectrum. Mel Spectrogram features are more consistent and provide higher performance metrics than MFCC, reaching almost 100% accuracy in most cases. Particularly, with the Mel Spectrogram, we achieve high precision, recall, and F1 score, showing that the model excels at classifying between different classes while maintaining a balance between precision and recall. As discussed above, Mel Spectrogram incorporates temporal instincts of the acoustics, making it more useful. MFCC requires a pre-processing step and they transform the higher dimensional frequency information into lower dimensional feature space, hence, some information is lost. Therefore, we find slightly lesser prediction accuracy from the MFCC feature.

2) Results and Discussion for Swarm Prediction using LSTM

Figure 13 shows the performance comparison of Accuracy, Precision, Recall, and F1 Score for LSTM with different number of layers, learning rate and number of epochs. It is evident that LSTM performs reasonably well but does not achieve the same high levels of accuracy as the CNN model. More precisely, in Figure 13(a), the LSTM faces challenges in maintaining high scores across all the metrics, with Precision and Recall showing significant variability. Likewise, figure 13(b) indicates that LSTM shows lower Precision, Recall, and F1 scores when utilizing MFCC features, especially with increased learning rates and reduced epochs, in contrast to Mel Spectrogram features. Additionally, the results of K-fold cross-validation, as displayed in Figures 13(c) and 13(d), also suggest that the performance of the LSTM model varies more noticeably when compared to the CNN model, especially when there are changes in learning rates and the number of epochs. Therefore, it has been observed that CNN performs better than LSTM with all the configurations. LSTM directly extracts features from the raw input dataset, leading to higher accuracy for both features. Using MFCC results in slightly lower accuracy due to the need for pre-processing, leading to potential loss of information. On a similar note, we observe that for most of the evaluation with different hyperparameters, we find the accuracy of LSTM with Mel spectrograms is better. This is because Mel spectrograms illustrate frequency changes over time, making them well-suited for LSTM, which is a useful model for sequence modeling and classifying datasets with temporal features. So using Mel spectrograms, LSTM is good at understanding the relationships among the frequency patterns.

3) Results and Discussion for Swarm Prediction using Transformer

We present the performance comparison of Accuracy, Precision, Recall, and F1 Score for the Transformer model with different learning rates in Figure 14. Figure 14(a) shows



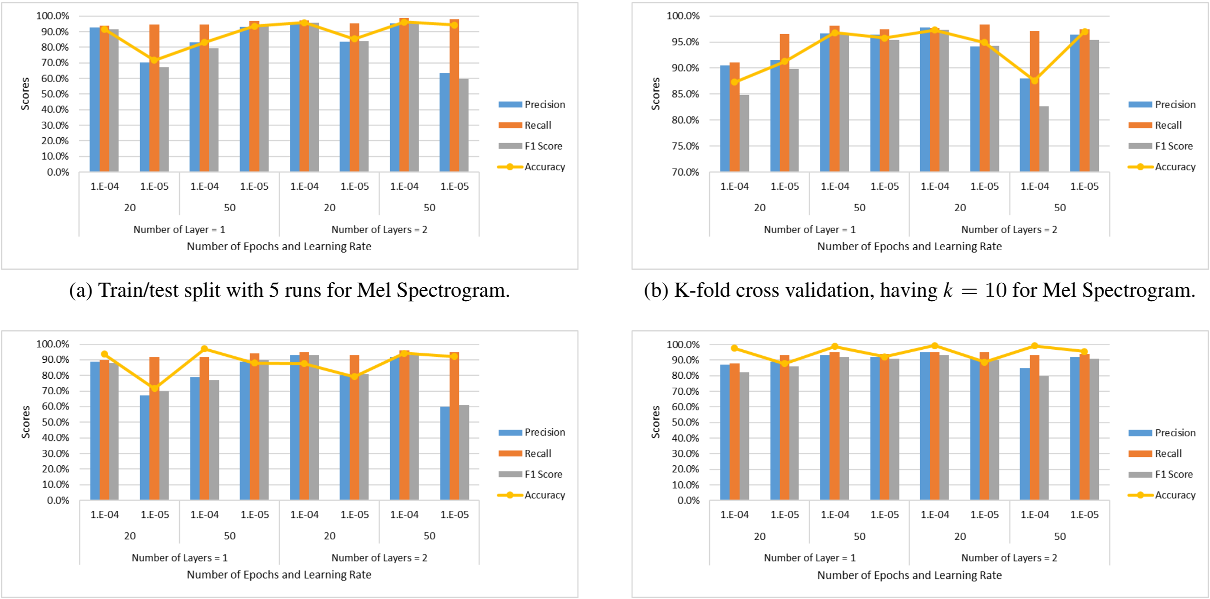
(a) Train/test split with 5 runs for Mel Spectrogram.

(b) K-fold cross validation, having $k = 10$ for Mel Spectrogram.

(c) Train/test split with 5 runs for MFCC.

(d) K-fold cross validation, having $k = 10$ for MFCC.

FIGURE 12: Performance Comparison of Accuracy, Precision, Recall and F1 Score for CNN



(a) Train/test split with 5 runs for Mel Spectrogram.

(b) K-fold cross validation, having $k = 10$ for Mel Spectrogram.

(c) Train/test split with 5 runs for MFCC.

(d) K-fold cross validation, having $k = 10$ for MFCC.

FIGURE 13: Performance Comparison of Accuracy, Precision, Recall and F1 Score for LSTM

the results of experiments utilizing the Mel Spectrogram, it is clear that the Transformer model consistently achieves high accuracy with different learning rates, showing minimal changes in Precision and Recall. Yet, the Transformer model provides higher accuracy overall when using MFCC features as input and a learning rate of 0.0001, as indicated in 14(b). This configuration also ensures that precision, recall, and F1 score remain high, especially during k-fold cross-validation for k=10. The results show that although both Mel Spectrogram and MFCC are effective, the Transformer model performs particularly well with MFCC features at a learning rate of 0.0001.

Transformer models excel in leveraging the contextual information present in temporal data such as Mel spectrograms. This allows them to obtain valuable data from the audio signals. For some of the experiments, we observe that the results of using MFCC are better. One of the reasons is that MFCCs are carefully generated pre-processed signals, so sometimes, they hold more useful information than Mel spectrograms. Secondly, transformer models struggle due to higher dimensionality. MFCC have low dimensional features, hence, serving as a better case for the transformers.

V. CONCLUSION

In this research, we have focused on an important phenomenon in honeybee farming, called bee swarming. Bee swarming is critical to honey production and hence requires special attention. Using bee acoustics for swarm prediction is an area of research that has never gotten much focus in the past. Considering this research gap, we have demonstrated the efficacy of different machine learning and deep learning models to identify beehive states using audio data recorded inside beehives. This study takes advantage of bee buzzing signals in the form of audio data, which is a useful tool for estimating the status of a bee colony. We have provided a systematic data science approach, beginning with data gathering and feature extraction, followed by swarm prediction by using the spectral properties of the bee acoustics to serve as input to the models. We used four different audio features, named Sum of amplitude of wave plot, waveplot, Mel Spectrogram, and MFCC, along with various hyperparameters for the models. Extensive experiments using machine learning (kNN, Naive Bayes, and SVM) and deep learning models (CNN, LSTM, and Transformer) are conducted to evaluate and identify the optimal models. The results demonstrate that by utilizing these models and a suitable pre-processing strategy, acoustics analysis of beehives may effectively identify between different states of the hives, and anomalies in the state of beehives might be detected early. We understand that the application and usefulness of our research are quite high, as our findings can aid in the development of automated beehive monitoring systems.

REFERENCES

[1] Librosa. Accessed 01 April 2024.

[2] Open source beehives project. Accessed 01 April 2024.

[3] Antonio Rafael Braga, Danielo G Gomes, Richard Rogers, Edgar E Hassler, Breno M Freitas, and Joseph A Cazier. A method for mining combined data from in-hive sensors, weather and apiary inspections to forecast the health status of honey bee colonies. *Computers and Electronics in Agriculture*, 169:105161, 2020.

[4] Jerry J Bromenshenk, Colin B Henderson, Robert A Seccomb, Phillip M Welch, Scott E Debnam, and David R Firth. Bees as biosensors: chemosensory ability, honey bee monitoring systems, and emergent sensor technologies derived from the pollinator syndrome. *Biosensors*, 5(4):678–711, 2015.

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

[6] T. Cejrowski, J. Szymanski, H. Mora, and D. Gil. Detection of the bee queen presence using sound analysis. In *Asian Conference on Intelligent Information and Database Systems*, pages 297–306. Springer, March 2018.

[7] Chunxu Chen, Guangzhong Xie, Jing Dai, Weixiong Li, Yulin Cai, Jing Li, Qiuping Zhang, Huiling Tai, Yadong Jiang, and Yuanjie Su. Integrated core-shell structured smart textiles for active no2 concentration and pressure monitoring. *Nano Energy*, 116:108788, 2023.

[8] Jing Chen, Ying Song, Daping Li, Xianxuan Lin, Sihang Zhou, and Wenqiang Xu. Specular removal of industrial metal objects without changing lighting configuration. *IEEE Transactions on Industrial Informatics*, 2023.

[9] Jing Dai, Guangzhong Xie, Chunxu Chen, Yulin Liu, Huiling Tai, Yadong Jiang, and Yuanjie Su. Hierarchical piezoelectric composite film for self-powered moisture detection and wearable biomonitoring. *Applied Physics Letters*, 124(5), 2024.

[10] Jing Dai, Guangzhong Xie, Chunxu Chen, Yulin Liu, Huiling Tai, Yadong Jiang, and Yuanjie Su. Hierarchical piezoelectric composite film for self-powered moisture detection and wearable biomonitoring. *Applied Physics Letters*, 124(5), 2024.

[11] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan. Heart sound classification based on improved mfcc features and convolutional recurrent neural networks. *Neural Networks*, 130:22–32, 2020.

[12] S. Dieleman and B. Schrauwen. End-to-end learning for music audio. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6964–6968. IEEE, May 2014.

[13] Ye Ding, Wenyi Zhang, Xibo Zhou, Qing Liao, Qiong Luo, and Lionel M Ni. Fraudtrip: Taxi fraudulent trip detection from corresponding trajectories. *IEEE Internet of Things Journal*, 8(16):12505–12517, 2020.

[14] S. Ferrari, M. Silva, M. Guarino, and D. Berckmans. Monitoring of swarming sounds in bee hives for early detection of the swarming period. *Computers and Electronics in Agriculture*, 64(1):72–77, 2008.

[15] Hubert Frings and Franklin Little. Reactions of honey bees in the hive to simple sounds. *Science*, 125(3238):122–122, 1957.

[16] Y. Gong, Y.A. Chung, and J. Glass. Ast: Audio spectrogram transformer. *arXiv preprint arXiv:2104.01778*, 2021.

[17] A. Gulati, J. Qin, C.C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang. Conformer: Convolution-augmented transformer for speech recognition. *arXiv preprint arXiv:2005.08100*, 2020.

[18] Xuyan Hou, Linbo Xin, Yulei Fu, Zhonglai Na, Guowei Gao, Yuhui Liu, Qingzhang Xu, Pingting Zhao, Gongzhuo Yan, Yilin Su, et al. A self-powered biomimetic mouse whisker sensor (bmws) aiming at terrestrial and space objects perception. *Nano Energy*, 118:109034, 2023.

[19] Michael Hrncir, Friedrich G Barth, and Jurgen Tautz. 32 vibratory and airborne-sound signals in bee communication (hymenoptera). *Insect sounds and communication: physiology, behaviour, ecology, and evolution*, page 421, 2005.

[20] James H Hunt and F-J Richard. Intracolony vibroacoustic communication in social insects. *Insectes Sociaux*, 60:403–417, 2013.

[21] Khalid Khan. Beekeeping in pakistan (history, potential, and current status). 2020.

[22] J. Kim, J. Oh, and T.Y. Heo. Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and grad-cam. *Mathematical Problems in Engineering*, 2021.

[23] WH Kirchner. Acoustical communication in honeybees. *Apidologie*, 24(3):297–307, 1993.

[24] Alexandra-Maria Klein, Bernard E Vaissière, James H Cane, Ingolf Steffan-Dewenter, Saul A Cunningham, Claire Kremen, and Teja Tscharntke. Importance of pollinators in changing landscapes for world crops. *Proceedings of the royal society B: biological sciences*, 274(1608):303–313, 2007.

[25] Q. Kong, Y. Xu, W. Wang, and M.D. Plumbley. Sound event detection of weakly labelled data with cnn-transformer and automatic threshold optimization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2450–2460, 2020.

[26] Hadi Kordestani, Chunwei Zhang, and Ali Arab. An investigation into the application of acceleration responses' trendline for bridge damage detection using quadratic regression. *Sensors*, 24(2):410, 2024.

[27] V. Kulyukin, S. Mukherjee, and P. Amlathe. Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples. *Applied Sciences*, 8(9):1573, 2018.

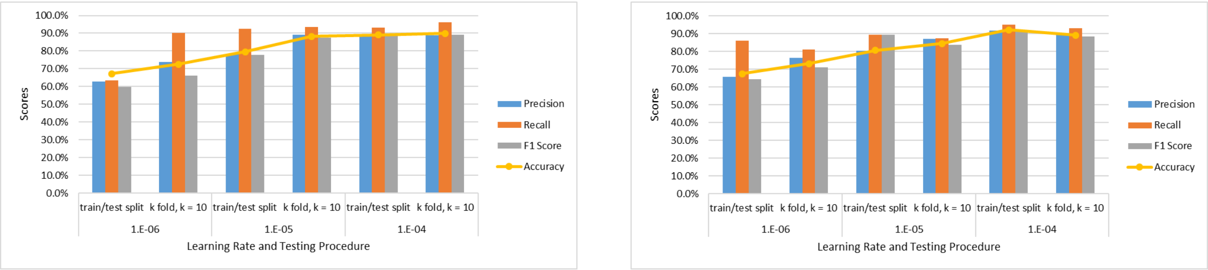
[28] Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10), 1995.

[29] Yi Li, Weixiong Li, Ziyang Jin, Xiaolan Luo, Guangzhong Xie, Huiling Tai, Yadong Jiang, Yajie Yang, and Yuanjie Su. Ternary ordered assembled piezoelectric composite for self-powered ammonia detection. *Nano Energy*, 122:109291, 2024.

[30] Fang Liu, Xinhang Zhao, Zihao Zhu, Zhongping Zhai, and Yongbin Liu. Dual-microphone active noise cancellation paved with doppler assimilation for tads. *Mechanical Systems and Signal Processing*, 184:109727, 2023.

[31] Qi Liu, Hui Yuan, Raouf Hamzaoui, Honglei Su, Junhui Hou, and Huan Yang. Reduced reference perceptual quality model with application to rate control for video-based point cloud compression. *IEEE Transactions on Image Processing*, 30:6623–6636, 2021.

[32] Norma Mallegni, Giovanna Molinari, Claudio Ricci, Andrea Lazzeri, Davide La Rosa, Antonino Crivello, and Mario Milazzo. Sensing devices for detecting and processing acoustic signals in healthcare. *Biosensors*, 12(10):835, 2022.



(a) Experiments using Mel Spectrogram.

(b) Experiments using MFCC.

FIGURE 14: Performance Comparison of Accuracy, Precision, Recall and F1 Score for Transformers Model

[33] Axel Michelsen, Wolfgang H Kirchner, and Martin Lindauer. Sound and vibrational signals in the dance language of the honeybee, apis mellifera. *Behavioral ecology and sociobiology*, 18:207–212, 1986.

[34] K. Miyazaki, T. Komatsu, T. Hayashi, S. Watanabe, T. Toda, and K. Takeda. Convolution augmented transformer for semi-supervised sound event detection. In *Proc. Workshop Detection Classification Acoust. Scenes Events (DCASE)*, pages 100–104, June 2020.

[35] I. Nolasco and E. Benetos. To bee or not to bee: Investigating machine learning approaches for beehive sound recognition. *arXiv preprint arXiv:1811.06016*, 2018.

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

[37] Stavros Ntalampiras and Ilyas Potamitis. Transfer learning for improved audio-based human activity recognition. *Biosensors*, 8(3):60, 2018.

[38] Stavros Ntalampiras, Ilyas Potamitis, and Nikos Fakotakis. Acoustic detection of human activities in natural environments. *Journal of the Audio Engineering Society*, 60(9):686–695, 2012.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[40] S. Ruvinga, G.J. Hunter, O. Duran, and J.C. Nebel. Use of lstm networks to identify "queenlessness" in honeybee hives from audio signals. In *2021 17th International Conference on Intelligent Environments (IE)*, pages 1–4. IEEE, June 2021.

[41] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, and M. Mortillaro. The interspeech 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 2013.

[42] J. Shen, R. Pang, R.J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. Skerrv-Ryan, and R.A. Saurous. Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 4779–4783. IEEE, April 2018.

[43] Yifei Shi, Junhua Xi, Dewen Hu, Zhiping Cai, and Kai Xu. Raymvsnet++: learning ray-based 1d implicit fields for accurate multi-view stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[44] Yuanjie Su, Shuo Chen, Bohao Liu, Haijun Lu, Xiaolan Luo, Chunxu Chen, Weixiong Li, Yin Long, Huiling Tai, Guangzhong Xie, et al. Maxwell displacement current induced wireless self-powered gas sensor array. *Materials Today Physics*, 30:100951, 2023.

[45] A. Terenzi, S. Cecchi, and S. Spinsante. On the importance of the sound emitted by honey bee hives. *Veterinary Sciences*, 7(4):168, 2020.

[46] Wei Tian, Yangqing Zhao, Rui Hou, Mianxiong Dong, Kaoru Ota, Deze Zeng, and Junmin Zhang. A centralized control-based clustering scheme for energy efficiency in underwater acoustic sensor networks. *IEEE Transactions on Green Communications and Networking*, 7(2):668–679, 2023.

[47] G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M.A. Nicolaou, B. Schuller, and S. Zafeiriou. Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5200–5204. IEEE, March 2016.

[48] Dennis VanEngelsdorp, Jay D Evans, Claude Saegerman, Chris Mullin, Eric Haubruge, Bach Kim Nguyen, Maryann Frazier, Jim Frazier, Diana Cox-Foster, Yanping Chen, et al. Colony collapse disorder: a descriptive study. *PloS one*, 4(8):e6481, 2009.

[49] Dennis Vanengelsdorp, Kirsten S Traynor, Michael Andree, Elinor M Lichtenberg, Yanping Chen, Claude Saegerman, and Diana L Cox-Foster. Colony collapse disorder (ccd) and bee age impact honey bee pathophysiology. *PLoS One*, 12(7):e0179535, 2017.

[50] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[51] Diana Vitazkova, Erik Foltan, Helena Kosnacova, Michal Micjan, Martin Donoval, Anton Kuzma, Martin Kopani, and Erik Vavrinsky. Advances in respiratory monitoring: A comprehensive review of wearable and remote technologies. *Biosensors*, 14(2):90, 2024.

[52] S. Voskarides, L. Josserand, J.P. Martin, C. Novales, and D. Micheletto. Electronic bee–hive (e–ruche) project. In *Proceedings of the Conference VIVUS-Environmentalism, Agriculture, Horticulture, Food Production and Processing Knowledge and Experience for New Entrepreneurial Opportunities*, pages 24–25, Naklo, Slovenia, April 2013.

[53] GuiPing Wang, JianXi Yang, and Ren Li. Imbalanced svm-based anomaly detection algorithm for imbalanced training datasets. *Etri Journal*, 39(5):621–631, 2017.

[54] Zhaobao Wang, Ran Ma, Bingjing Chen, Xiaotong Yu, Xue Wang, Xinyun Zuo, Bo Liang, and Jianming Yang. A transcription factor-based bacterial biosensor system and its application for on-site detection of explosives. *Biosensors and Bioelectronics*, 244:115805, 2024.

[55] Haitao Xu, Qiang Li, and Jing Chen. Highlight removal from a single grayscale image using attentive gan. *Applied Artificial Intelligence*, 36(1):1988441, 2022.

[56] Mang Xu, Liza Portier, Toine Bovee, Ying Zhao, Yirong Guo, and Jeroen Peters. Neonicotinoid microsphere immunosensing for profiling applications in honeybees and bee-related matrices. *Biosensors*, 12(10):792, 2022.

[57] Jianxi Yang, Hao Li, Junzhi Zou, Shixin Jiang, Ren Li, and Xinlong Liu. Concrete crack segmentation based on uav-enabled edge computing. *Neurocomputing*, 485:233–241, 2022.

[58] Li Yu, Jieliang Zhao, Zhiyun Ma, Wenzhong Wang, Shaoze Yan, Yue Jin, and Yu Fang. Experimental verification on steering flight of honeybee by electrical stimulation. *Cyborg and Bionic Systems*, 2022.

[59] A. Zgank. Bee swarm activity acoustic classification for an iot-based farm service. *Sensors*, 20(1):21, 2019.

[60] Qiuping Zhang, Guangzhong Xie, Manyi Duan, Yutong Liu, Yulin Cai, Ming Xu, Kang Zhao, Huiling Tai, Yadong Jiang, and Yuanjie Su. Zinc oxide nanorods for light-activated gas sensing and photocatalytic applications. *ACS Applied Nano Materials*, 6(19):17445–17456, 2023.

[61] Wei Zhang, Fulong Zhu, Shenghuai Wang, Pengyan Lu, and Xin Wu. An accurate method to calibrate shadow moiré measurement sensitivity. *Measurement Science and Technology*, 30(12):125021, 2019.

[62] Guoqing Zhou, Yi Tang, Wenxi Zhang, Weiguang Liu, Yue Jiang, Ertao Gao, Qiang Zhu, and Yuhang Bai. Shadow detection on high-resolution digital orthophoto map (dom) using semantic matching. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[63] Guoqing Zhou, Chao Xu, Haotian Zhang, Xiang Zhou, Dawei Zhao, Gongbei Wu, Jinchun Lin, Zhexian Liu, Jiazhi Yang, Xueqin Nong, et al. Pmt gain self-adjustment system for high-accuracy echo signal detection. *International Journal of Remote Sensing*, 43(19-24):7213–7235, 2022.

[64] A. Zlatkova, B. Gerazov, D. Tashkovski, and Z. Kokolanski. Analysis of parameters in algorithms for signal processing for swarming of honeybees. In *2020 28th Telecommunications Forum (TELFOR)*, pages 1–4. IEEE, November 2020.

[65] A. Zlatkova, Z. Kokolanski, and D. Tashkovski. Honeybees swarming detection approach by sound signal processing. In *2020 XXIX International Scientific Conference Electronics (ET)*, pages 1–3. IEEE, September 2020.

FIRST KAINAT IQBAL KAINAT IQBAL completed her MS in Data Science from School of Computing, National University of Computing and Emerging Sciences, Islamabad, Pakistan. She has research interests in Honey Bee Science, Computer Vision, Deep Learning, and Data Analytics.

SECOND BAYAN ALABDULLAH BAYAN ABDULLAH received the Ph.D. degree in informatics from the University of Sussex, Brighton, U.K., in May 2022. She is an Assistant Professor with the Department of Information System, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University. She teaches several courses with the Information System Department, such as data governance, system security, and database system. Her research interests include machine learning, data science, and privacy and security..

THIRD NAIF AL MUDAWI NAIF AL MUDAWI works in Department of Computer Science and Information system, Najran University. He holds a PhD from the Collage of Engineering and Informatics at University of Sussex in Brighton, UK with distinction with honours in a delicate specialization in Adopting of cloud computing in online system for public organisation in 2018. He graduated from the Australian La Trobe University with a master's degree in computer science in (2011).

FOURTH ASAAD ALGARNI ASAAD ALGARNI is working as Assistant Professor at the Department of Computer Sciences in the College of Computing and Information Technology, Northern Borders University, Kingdom of Saudi Arabia. He holds a PhD in Software Engineering from North Dakota State University, USA. His research interests revolve around Software Engineering, Computer Vision applications and Machine Learning.

FIFTH AND SIXTH AHMAD JALAL AHMAD JALAL received his Ph.D. degree from the Department of Biomedical Engineering, Kyung Hee University, Republic of Korea. He was a Postdoctoral Research Fellow with POSTECH. He is currently an Associate Professor with the Department of Computer Science and Engineering, Air University, Pakistan. His research interests include multimedia contents, artificial intelligence, and machine learning.

SEVENTH JEONGMIN PARK JEONGMIN PARK received the Ph.D. degree from the College of Information and Communication Engineering, Sungkyunkwan University, South Korea, in 2009. He is currently an Associate Professor with the Department of Computer Engineering, Tech University of Korea, South Korea. Before joining the Tech University of Korea, in 2014, he was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI) and a Research Professor with Sungkyunkwan University. His research interests include high reliable autonomic computing mechanism and human oriented interaction systems.

• • •