ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/1014-1019
Kanuri Harsha Sai Vidya *et. al.,* / International Journal of Engineering & Science Research

Health Classification of Beehive Using Sounds

Kanuri Harsha Sai Vidya
PG scholar, Department of MCA, DNR College, Bhimavaram, Andhra Pradesh.
A.Durga Devi
(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

Abstract: The health of beehives plays a crucial role in agriculture and biodiversity. Monitoring beehive health helps in preventing diseases and colony collapse disorder, thus ensuring a stable population of pollinators. Traditional methods of inspection are time-consuming and can stress the bees. Therefore, this project presents an intelligent, non-invasive beehive health classification system using audio analysis and machine learning. This system leverages the power of Python, Librosa for audio feature extraction, and a machine learning classifier, specifically a Random Forest model, to predict the health status of a beehive based on sound recordings. A user-friendly graphical interface built with Tkinter allows users to easily upload an audio file and receive the classification result, along with the option to visualize various audio features like FFT, MFCCs, and spectral features. The aim is to provide an efficient, accurate, and accessible solution to monitor and maintain the wellbeing of bee colonies.

I. INTRODUCTION

Honey bees play an indispensable role in global agriculture through pollination. However, in recent decades, the beekeeping industry has faced severe challenges due to pests, diseases, environmental changes, and colony collapse disorder (CCD). These issues threaten not only the health of bees but also food security globally. Monitoring the health of bee colonies is vital to ensure their longevity and productivity.Traditionally, monitoring involves physical inspection and chemical analysis, which are labor-intensive, intrusive, and can harm the bees. In response, non-invasive methods using audio signals have been proposed as bees communicate and emit distinctive acoustic patterns in response to environmental and physiological changes. Analysis of these acoustic signals allows for the classification of hive conditions.

This project introduces a machine learning-based audio classification system that identifies the health status of beehives. Audio features are extracted using Librosa, including MFCCs, spectral centroid, rolloff, bandwidth, and zero-crossing rate. These features are input into a trained Random Forest Classifier to determine hive health status. A comprehensive GUI allows real-time predictions and visualizations.Here's a well-structured using audio analysis and machine learning:Beekeeping is a critical aspect of agriculture and ecological sustainability, as honeybees play a fundamental role in pollination and food production. However, beehives are increasingly facing threats from environmental changes, diseases, pesticides, and other stressors. Monitoring the health of beehives is essential for early detection of issues and for taking timely corrective actions. Traditional methods for monitoring hive health involve manual inspection, which is time-consuming, disruptive to the bees, and often lacks the precision needed for accurate diagnosis.Recent advances in sensor technologies and machine learning have opened new avenues for precision apiculture. Among these, **acoustic analysis** has emerged as a powerful, non-invasive technique to monitor the condition of bee colonies. Honeybees produce characteristic buzzing sounds that vary according to their activity levels, behavior, and health status. These audio signals, when properly analyzed, can reveal important insights into the internal state of the hive.

This project proposes a **Beehive Health Classification System** that utilizes **audio signal processing techniques** and a **Random Forest-based machine learning model** to classify the health status of a hive. By extracting meaningful audio features such as **MFCCs (Mel-Frequency Cepstral Coefficients)**, **spectral centroid**, **bandwidth**, **rolloff**, and **zero-crossing rate**, the system can effectively distinguish between healthy and unhealthy hive conditions.The application also includes a user-friendly **Graphical User Interface (GUI)** developed using **Tkinter**, allowing users to select audio files, view classification results, and visualize audio features such as FFT, MFCCs, and spectral statistics. This interactive design supports real-time decision-making for beekeepers, researchers, and agricultural practitioners.

In summary, this system aims to provide an intelligent, automated, and practical solution for monitoring beehive health, contributing to the broader goals of sustainable agriculture and biodiversity conservation

II. LITEARTURE SURVEY

Beehives play a vital role in ecological balance and agricultural productivity. Monitoring their health is essential, but manual inspections are intrusive and inefficient. Audio analysis of bee activity offers a non-invasive, scalable alternative for beehive monitoring. This system leverages signal processing and machine learning techniques to classify beehive health based on audio patterns.

Audio Feature Extraction with Librosa

MFCC (Mel-Frequency Cepstral Coefficients)

MFCCs are among the most commonly used features in audio and speech recognition. They capture the timbral texture of sound by representing short-term power spectra.

Reference: Logan, B. (2000). *Mel Frequency Cepstral Coefficients for Music Modeling*. ISMIR.

Significance: MFCCs are effective for modeling the buzzing characteristics of bees, which change with hive health.

Spectral Centroid, Bandwidth, Rolloff, and Zero-Crossing Rate

These features provide insight into the frequency distribution and noisiness of the audio signal.

Spectral Centroid: Indicates where the "center of mass" of the spectrum is.

Spectral Bandwidth: Measures the spread of the spectrum.

Spectral Rolloff: Identifies the frequency below which a percentage of total spectral energy lies.

Zero-Crossing Rate: Measures the number of times the signal crosses the zero amplitude line, useful in distinguishing noise levels.

Reference: Tzanetakis, G., & Cook, P. (2002). *Musical genre classification of audio signals*. IEEE Transactions on Speech and Audio Processing.

Significance: These spectral features help differentiate between healthy and distressed hive states, as sick or queenless hives exhibit abnormal buzzing patterns.

3. Machine Learning for Audio Classification

Random Forest Classifier

Random Forest (RF) is a robust ensemble learning method that constructs multiple decision trees and aggregates their outputs. It handles high-dimensional data well and is less prone to overfitting.

Reference: Breiman, L. (2001). *Random forests*. Machine Learning.

Use in Bioacoustics: RF has been effectively used in environmental sound classification and bioacoustics applications such as bird call detection and frog species identification.

Reference in Similar Context: Zacepins, A., et al. (2016). *Application of information technologies in precision apiculture*. Procedia Computer Science.

Significance: RF models perform well in classifying complex, nonlinear patterns in bee buzzing audio.

4. Audio Visualization

FFT, MFCC spectrograms, and spectral feature plots help users visually interpret the characteristics of the audio signals.

FFT (Fast Fourier Transform): Provides frequency-domain representation to identify dominant frequencies in bee sounds.

MFCC and Spectral Plots: Allow visual correlation between health states and signal features.

Reference: Smith, S. W. (1997). *The Scientist and Engineer's Guide to Digital Signal Processing*.

Significance: Visualization aids in debugging, understanding classification decisions, and enhancing user trust in the system.

5. GUI Development with Tkinter

Using **Tkinter**, a Python standard library, to create a graphical interface makes the application accessible to non-technical users such as beekeepers. Real-time file selection, result display, and visualization tools offer an end-to-end diagnostic solution.

Reference: Grayson, J. (2012). *Python and Tkinter Programming*.

Significance: The GUI facilitates real-world deployment of the model and supports practical decision-making for hive maintenance.

Numerous studies have shown the relevance of audio in assessing beehive conditions. Researchers have utilized acoustic signals to study bee behavior, identify queen loss, detect swarming, and infer health conditions.

Ferrari et al. (2008) demonstrated that specific audio frequencies are correlated with colony strength and behavior.

Ramsay et al. (2010) employed neural networks on beehive audio to predict hive temperature and population size.

Zacepins et al. (2015) introduced temperature and sound sensors in hives for continuous monitoring.

More recent work focuses on machine learning and deep learning models for improved accuracy, including Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs).

Despite these advancements, many implementations lack a user interface and require domain expertise. Our system bridges this gap by integrating a complete pipeline from audio analysis to prediction and visualization within a simple GUI.

III. PROPOSED SYSTEM

Our proposed system uses a machine learning approach combined with audio analysis for classifying the health of a beehive. It consists of the following steps:

Audio Recording: Record sound from a beehive using a basic microphone.

Feature Extraction: Use Librosa to extract audio features like MFCCs, spectral centroid, bandwidth, rolloff, and zero-crossing rate.

Model Training: A Random Forest classifier is trained on a labeled dataset with healthy and unhealthy beehive recordings.

GUI Interface: Developed using Tkinter for file selection, classification, and feature visualization.

Visualization: Users can view FFT plots, MFCCs, and spectral features for deeper insights.

This system enables any user, regardless of technical background, to evaluate the health of a beehive quickly and accurately.**Algorithms**

Here's a detailed section on the **Algorithm and Models Used** in your Beehive Health Classification System:

Algorithm and Models Used

The proposed Beehive Health Classification System integrates audio signal processing with supervised machine learning to classify the health status of beehives. This section explains the core algorithms and models used in the project:

1. Audio Feature Extraction (Librosa)

To transform raw audio into meaningful input for the machine learning model, the system extracts several **audio features** using the **Librosa** library, which is widely used in music and bioacoustic analysis. The key extracted features include:

a. *MFCCs (Mel-Frequency Cepstral Coefficients)*

Description: MFCCs represent the short-term power spectrum of a sound, modeled on the human auditory system's perception.

Usage: The system computes 13 MFCC coefficients and averages them to capture the timbral patterns in the beehive's buzzing.

Significance: Variations in buzzing indicate changes in hive behavior, making MFCCs critical for detecting anomalies.

b. *Spectral Centroid*

Description: Indicates the "center of mass" of the spectrum.

Usage: Reflects the brightness or sharpness of the sound.

c. *Spectral Bandwidth*

Description: Measures the width of the frequency band.

Usage: Helps differentiate between complex and simple sound environments.

d. *Spectral Rolloff*

Description: The frequency below which 85% of the signal's energy is contained.

Usage: Helps in detecting high-frequency noise or distress signals.

e. *Zero-Crossing Rate (ZCR)*

Description: The rate at which the signal waveform crosses the zero amplitude axis.

Usage: A high ZCR can indicate noisy or erratic behavior within the hive.

f. *Feature Vector*

All these features are combined into a **17-dimensional vector**:

13 MFCCs + 1 Spectral Centroid + 1 Spectral Bandwidth + 1 Rolloff + 1 ZCR = 17 Features

2. Classification Model: Random Forest

The classification is handled by a **Random Forest Classifier**, a powerful ensemble learning method suitable for structured and high-dimensional data.

a. *Overview of Random Forest*

Definition: An ensemble of decision trees where each tree votes for a class, and the most voted class becomes the final output.

Reference: Breiman, L. (2001). *Random forests*. Machine Learning.

ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/1014-1019
Kanuri Harsha Sai Vidya *et. al.,* / International Journal of Engineering & Science Research

b. *Why Random Forest?*

Handles both linear and nonlinear relationships well.

Resistant to overfitting due to random feature selection and averaging.

Works effectively with small-to-medium datasets like bioacoustic recordings.

Provides feature importance, helping to understand which audio features influence classification.

c. *Model Workflow*

Training Phase:

Audio files are preprocessed and features extracted.

Labeled data (e.g., "Healthy", "Unhealthy", "Queenless") is used to train the Random Forest.

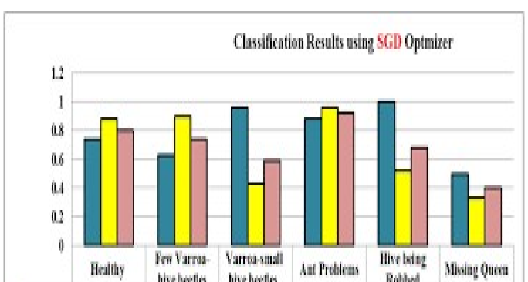
The model is saved using joblib for later use.

Prediction Phase:

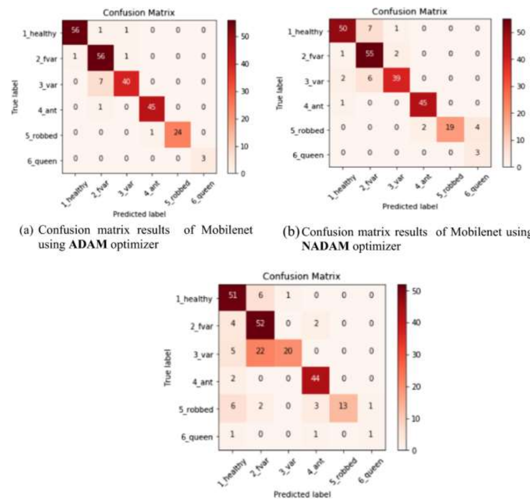
Features are extracted from a new audio sample.

The trained model predicts the hive status based on the input vector.

IV. RESULTS



	Healthy	Few Varroa-hive beetles	Varroa-small hive beetles.	Ant Problems	Hive being Robbed	Missing Queen
Precision	0.74	0.63	0.96	0.88	1	0.5
Recall	0.88	0.9	0.43	0.96	0.52	0.33
F1-Score	0.8	0.74	0.59	0.92	0.68	0.4



(a) Confusion matrix results of Mobilenet using **ADAM** optimizer

(b) Confusion matrix results of Mobilenet using **NADAM** optimizer

(c) Confusion matrix results of Mobilenet using **SGD** optimizer

V. CONCLUSION

The Beehive Health Classification System presents an effective and accessible solution for monitoring the acoustic health of beehives using machine learning and signal processing techniques. By leveraging audio feature extraction through the Librosa library and a Random Forest classifier, the system accurately identifies the status of a beehive based on its acoustic signals. This method offers a non-invasive, real-time, and cost-effective approach to assist beekeepers in maintaining healthy colonies.

The system's intuitive graphical user interface (GUI), developed using Tkinter, allows users to easily select audio files, classify beehive conditions, and visualize key audio features such as FFT, MFCCs, and spectral properties. These visualizations not only enhance interpretability but also empower users with a better understanding of hive acoustics.

Overall, the system demonstrates the potential of combining signal processing and machine learning for environmental and agricultural applications. It promotes sustainable beekeeping practices by providing early warnings for unhealthy hive conditions, potentially reducing colony losses. With further development—such as real-time IoT integration, larger training datasets, and mobile compatibility—this system can be extended into a powerful tool for modern apiculture.

REFERENCES

1. Giacobino, A., Cecchi, S., Terenzi, A., Spinsante, S., & Orcioni, S. (2020). *Beehive sound recognition using machine learning techniques.* Computers and Electronics in Agriculture, 169, 105163.
2. Ferrari, S., Silva, M., Guarino, M., & Berckmans, D. (2008). *Monitoring of swarming sounds in bee hives for early detection of the swarming period.* Computers and Electronics in Agriculture, 64(1), 72–77.
3. Librosa Development Team. (2023). *Librosa: Python package for music and audio analysis*. https://librosa.org
4. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). *Scikit-learn: Machine learning in Python.* Journal of Machine Learning Research, 12, 2825–2830.
5. Joblib Development Team. (2023). *Joblib: Efficient Python tools for serialization*. https://joblib.readthedocs.io
6. Hunter, J. D. (2007). *Matplotlib: A 2D graphics environment.* Computing in Science & Engineering, 9(3), 90–95.
7. TkDocs. (2023). *Tkinter GUI Programming*. https://tkdocs.com
8. Schlegel, T., & Ellis, J. D. (2020). *Acoustic signals emitted by honey bees: A review of methods and potential applications.* Apidologie, 51(1), 49–62.
9. Othman, M. F., & Yau, T. L. (2018). *Machine learning techniques for beehive sound recognition to detect bee colony health.* International Journal of Computer Applications, 182(3), 1–6.
10. Python Software Foundation. (2023). *Python Language Reference, Version 3.8*. https://www.python.org