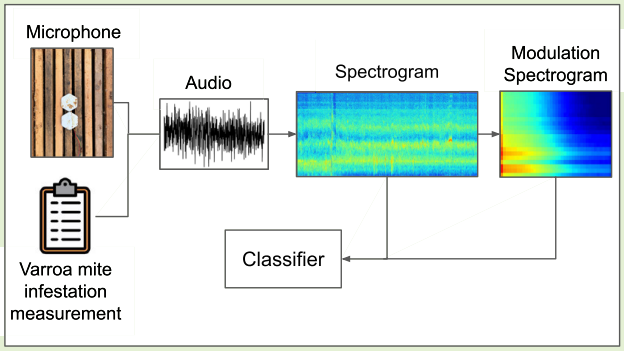
On the Prediction of Varroa Mite Infestations in Honey Bee Colonies via Acoustic Monitoring

Mahsa Abdollahi, Yi Zhu, Heitor R. Guimarães, *Graduate Student Member, IEEE*, Nico Coallier, Ségoléne Maucourt, Pierre Giovenazzo, and Tiago H. Falk, *Fellow, IEEE*

***Abstract*—Honey bees play a vital role in global ecosystems and agriculture through their pollination services. However, the Varroa destructor mite remains one of the most damaging parasites affecting honey bee health, contributing significantly to colony decline. Early detection of Varroa infestation is essential for effective beekeeping management. In this study, we propose a machine learning approach for predicting Varroa mite infestation levels based on the acoustic analysis of hive recordings. We introduce a set of nine spectral shape descriptors (SSDs) computed from conventional and modulation spectrograms to characterize the acoustic environment within hives. We evaluate the predictive power of these features against traditional cepstral features across multiple classifiers, including support vector machines (SVMs), random forests (RFs), and k-nearest neighbors (KNNs). Our results demonstrate that the proposed SSDs significantly improve the classification performance relative to conventional features across numerous figures of merit, particularly in hive-independent test settings. This work highlights the potential of spectral acoustic monitoring combined with the supervised learning as a scalable and noninvasive tool for precision beekeeping and early detection of parasitic threats.**



***Index Terms*— Beehive acoustics, honey bees, modulation spectrogram, Varroa mite infestation.**

I. Introduction

HONEY bees (*Apis mellifera*) are indispensable to both natural ecosystems and global agriculture due to their critical role in pollination. Nearly 75% of the world's crops producing fruits and seeds for human consumption depend, at least in part, on pollinators like honey bees. In addition to supporting biodiversity and food security, honey bees contribute significantly to the economy through honey, beeswax, royal jelly, and other products [1]. However, in recent decades, beekeepers and researchers have reported alarming declines in honey bee populations across the globe, raising serious concerns about the future of pollination-based agriculture and ecosystem resilience [2].

One of the most devastating threats to honey bee health is the *Varroa destructor* mite, an external parasitic mite that has spread to almost every continent where managed honey bees are kept [3]. The Varroa mite feeds on the fat bodies and hemolymph of both developing and adult bees, weakening their immune systems and shortening their lifespan. More critically, it acts as a vector for a range of harmful viruses, particularly the deformed wing virus (DWV), which can lead to severe colony losses. Left unmanaged, high levels of Varroa infestation can result in the collapse of entire colonies within a single season. For these reasons, early detection and accurate prediction of Varroa infestations are central to sustainable beekeeping practices and long-term colony survival.

Traditionally, beekeepers have relied on visual inspection methods such as sugar rolls, alcohol washes, and sticky boards to monitor mite levels. Although these techniques are widely accepted and relatively accurate when performed correctly, they are labor intensive, disruptive to the colony, and typically provide only periodic snapshots of mite infestation. These limitations hinder timely decision-making for integrated pest management strategies and large-scale or commercial

Received 23 January 2026; accepted 13 February 2026. Date of publication 24 February 2026; date of current version 16 April 2026. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Alliance Program (ALLRP) under Grant 548872-19. The associate editor coordinating the review of this article and approving it for publication was Prof. Rui Yuan. *(Corresponding author: Mahsa Abdollahi.)*

Mahsa Abdollahi, Yi Zhu, Heitor R. Guimarães, and Tiago H. Falk are with INRS-EMT, Université du Québec, Montreal, QC H5A 1K6, Canada (e-mail: mahsa.abdollahi@inrs.ca; Yi.Zhu@inrs.ca; Heitor.Guimaraes@inrs.ca; Tiago.Falk@inrs.ca).

Nico Coallier is with Nectar Technologies Inc., Montreal, QC H2V 4C5, Canada (e-mail: nico@nectar.buzz).

Ségoléne Maucourt and Pierre Giovenazzo are with the Department of Biology, Laval University, Québec City, QC G1V 0A6, Canada (e-mail: segolene.maucourt.1@ulaval.ca; pierre.giovenazzo@bio.ulaval.ca).

Digital Object Identifier 10.1109/JSEN.2026.3666127

operations. Furthermore, environmental conditions, beekeeper expertise, and colony variability can introduce significant noise and inconsistency into visual mite monitoring.

In recent years, we have seen a rise in precision beekeeping, driven by advancements in sensor technologies, machine learning, and data-driven modeling. These advances have resulted in novel, noninvasive methods for continuously monitoring honey bee colony health. Techniques, such as acoustic sensing, temperature and humidity tracking, and hive weight measurement, have been employed to infer colony dynamics and well-being [4], [5]. Among these modalities, the acoustic monitoring has demonstrated particularly versatile applications, offering rich information on colony behavior through the sound analysis. More recently, several publicly available honey bee hive acoustic datasets have been introduced, enabling reproducible research and accelerating the development of data-driven approaches for colony health assessment [6], [7], [8].

While the combination of honey bees acoustics and machine learning has shown considerable potential in detecting key colony states, such as swarming detection [9], queenlessness [10], [11], [12], predicting colony strength [13], and winter mortality [14], there remains limited research on their use for predicting Varroa destructor mite infestation levels, despite preliminary evidence showing promising outcomes [15], [16].

Qandour et al. [15], for example, investigated four acoustic features extracted from honey bee colony sounds, including peak frequency, spectral centroid, bandwidth, and root variance frequency to predict Varroa mite infestation using support vector machines (SVMs). Their findings demonstrated that this approach achieved notable effectiveness in distinguishing between infested and healthy beehive conditions. Chen et al. [11] investigated the impact of viruses on the colony sound. Their experiment involved two normal beehives and two beehives in which bees were infected by the DWV, causing seriously infected bees to have deformed wings after metamorphosis and a short life expectancy. In their experiments, the normal colonies showed higher amplitudes between 10 and 1000 Hz in spectrogram analysis. Using Mel-frequency cepstral coefficients (MFCCs) features and different machine learning algorithms, they showed that virus-infected beehives could be reliably detected using beehive acoustics.

Most prior studies on honey bee acoustic monitoring have relied heavily on MFCCs as the primary feature representation [5]. For instance, a recent work used MFCCs to classify bee activity at the hive entrance (worker versus drone) [17]. Another study compared MFCCs to higher level embeddings on multiple bee colony sound datasets, showing that MFCCs remain a competitive baseline for bee-colony sound classification tasks [18]. In yet another contribution, compact MFCC-based patterns were proposed to detect queenlessness: despite aggressive dimensionality reduction, the model achieved very high accuracy (99.3%) in distinguishing queenless versus normal hives [19]. Recent benchmarking of self-supervised audio representations for acoustic beehive monitoring shows that learned embeddings can outperform MFCCs and standard spectrogram features in tasks like bee activity classification, queenlessness detection, and colony strength prediction, while MFCCs remain a competitive baseline. This indicates that self-supervised models capture richer acoustic patterns than traditional cepstral representations [20].

While these studies demonstrate the effectiveness of MFCCs for a range of beekeeping applications, they also highlight an important limitation: MFCCs were originally designed for human speech perception and primarily emphasize short-term spectral envelopes on a perceptually motivated frequency scale. As a result, they may not optimally capture the distinctive spectral structure and temporal modulation patterns that characterize collective bee activity and stress-related hive dynamics. Moreover, most existing approaches focus on either short-term spectral information or highly compact representations, potentially overlooking slower rhythmic fluctuations and modulation cues that are critical for characterizing Varroa-related acoustic changes.

To enable a fair comparison with the existing work while addressing these limitations, we adopt MFCCs as a benchmark and additionally include linear-frequency cepstral coefficients (LFCCs), which remove the perceptual warping and provide a complementary linear spectral representation. This allows us to systematically assess the impact of frequency scaling on the classification performance and to highlight the shortcomings of cepstral features when applied to hive-acoustic monitoring.

Building on this, we propose a novel feature-fusion framework to predict Varroa mite infestation using new acoustic features extracted from honey bee colony recordings. Specifically, we compute nine spectral shape descriptors (SSDs)—centroid, spread, skewness, kurtosis, entropy, roll-off, flatness, crest, and flux—derived from both conventional and modulation spectrograms. While the conventional spectrogram captures short-term energy distributions across frequency bands, making it effective for detecting variations in wingbeat harmonics, buzzing intensity, and other activity-related patterns, it does not fully encode the rhythmic fluctuations and slow amplitude modulations inherent in bee sounds. The modulation spectrogram complements this by representing the temporal dynamics of spectral changes, such as the rate and strength of energy modulations across frequency bands, thereby providing a richer characterization of acoustic patterns associated with Varroa infestation.

Modulation-based representations have shown strong performance across various acoustic monitoring tasks, particularly in speech quality measurement and enhancement and environmental sound analysis [21], [22], [23], [24]. More recently, modulation spectrograms have also been applied in the context of beehive monitoring. In a study focused on predicting hive population dynamics [25], features derived from the modulation spectrogram outperformed traditional benchmark features such as MFCCs, highlighting the effectiveness of temporal modulation cues in characterizing complex bioacoustic signals.

Here, we frame the Varroa detection problem as a supervised binary classification task, determining whether a colony requires Varroa treatment based on a critical threshold. We evaluate a range of machine learning models, including Naïve Bayes (NBs) [26], random forests (RFs) [27], gradient boosting (GB) [28], k-nearest neighbors (KNNs) [29],

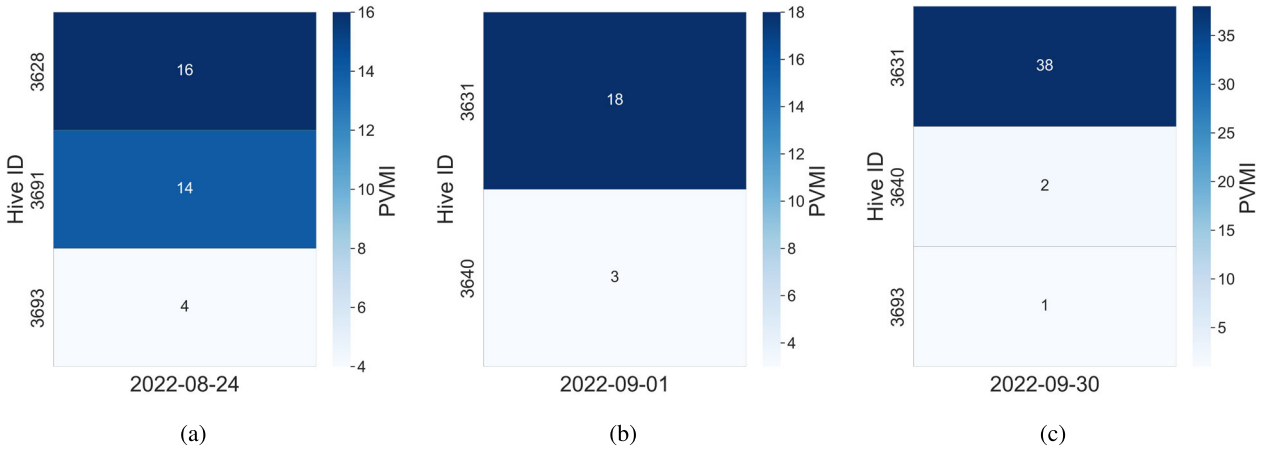


Fig. 1. PVMI measured on (a) August 24, (b) September 1, and (c) September 30, 2022.

SVMs [30], a multilayer perceptron (MLP) [31], [32], and a compact convolutional neural network (MiniCNN).

An important aspect of our experimental methodology is the consideration of hive-level independence in model training and evaluation. That is, rather than relying on random training–testing splits that may leak information between samples from the same hive, we adopt a strategy where entire hives are held out as testing sets. This more closely reflects real-world deployment scenarios, where a model trained on the historical data from some hives is used to predict the unseen data from different hives.

II. Methods and Materials

In this section, we present the dataset used, features extracted, testing setup, and figures of merit used in our study.

A. Dataset

We use the UrBAN audio dataset [7], which includes measurements of beehive strength and Varroa mite infestation across ten beehives over a period of two years. The audio data consists of 15-min audio segments recorded every 30 min at a sampling rate of 16 kHz. This dataset was collected from real-world urban beehives and, as reported in [7] and [13], inherently contains environmental noises such as traffic, wind, rain, and human activity, reflecting the acoustic variability encountered in practical deployments. The Varroa mite infestations were measured using the alcohol wash method [33] at different times of the year. In this method, approximately 300 adult bees are collected from the brood frames of each hive and washed in alcohol to dislodge the mites, which are then counted. The amount of Varroa infestation is reported as the number of mites per 100 bees, providing a standardized metric for comparison across colonies and time points. In this article, we refer to this metric as the percentage of Varroa mite infestation (PVMI). Fig. 1, for example, shows the heatmaps indicating the PVMI values in each measurement for different hives.

B. Feature Extraction

1) Proposed: We propose nine SSDs—centroid, spread, skewness, kurtosis, entropy, roll-off, flatness, crest, and flux—computed from both the conventional and the modulation spectrograms, where each captures a distinct aspect of the audio signal's spectral properties. While the conventional spectrogram-based descriptors quantify the short-term frequency structure of the signal (e.g., brightness, tonal balance, and noisiness), the descriptors computed from the modulation spectrogram, specifically across the modulation frequency axis, capture the temporal dynamics of spectral energy fluctuations. This enables the analysis of rhythmic and periodic patterns embedded in different frequency bands, offering a complementary perspective on the signal variability over time.

Fig. 2 depicts the block diagram showing the signal processing steps involved in the computation of the conventional and modulation spectrograms. The process begins by computing the short-time Fourier transform (STFT) of the raw hive audio signal to obtain a time–frequency representation. The magnitude spectrogram ($|STFT|$) is then extracted by taking the absolute value of the complex STFT, which forms the basis for both sets of features.

From the $|STFT|$, two parallel paths are followed.

1) *SSDs:* The magnitude spectrogram is used directly to compute nine SSDs: centroid, spread, skewness, kurtosis, entropy, roll-off, flatness, crest, and flux to capture the short-term frequency structure of the hive acoustics [34].
2) *Modulation SSDs (MSSDs):* In the second path, the magnitude spectrogram undergoes a second STFT operation (i.e., a modulation STFT) applied across the time axis of each frequency bin. This results in a modulation spectrogram, from which the same nine descriptors are computed along the modulation frequency axis. These modulation features capture slower temporal dynamics and rhythmic energy fluctuations that are relevant to colony behavior. Interested readers can refer to [23] for a more in-depth description of how the modulation spectrogram is computed.

For the spectrogram computation, a window length of 100 ms and a hop length of 12.5 ms were used to capture short-term spectral dynamics. Fig. 3, shows the 24-h spectrograms of a Varroa-infected beehive with a PVMI value of

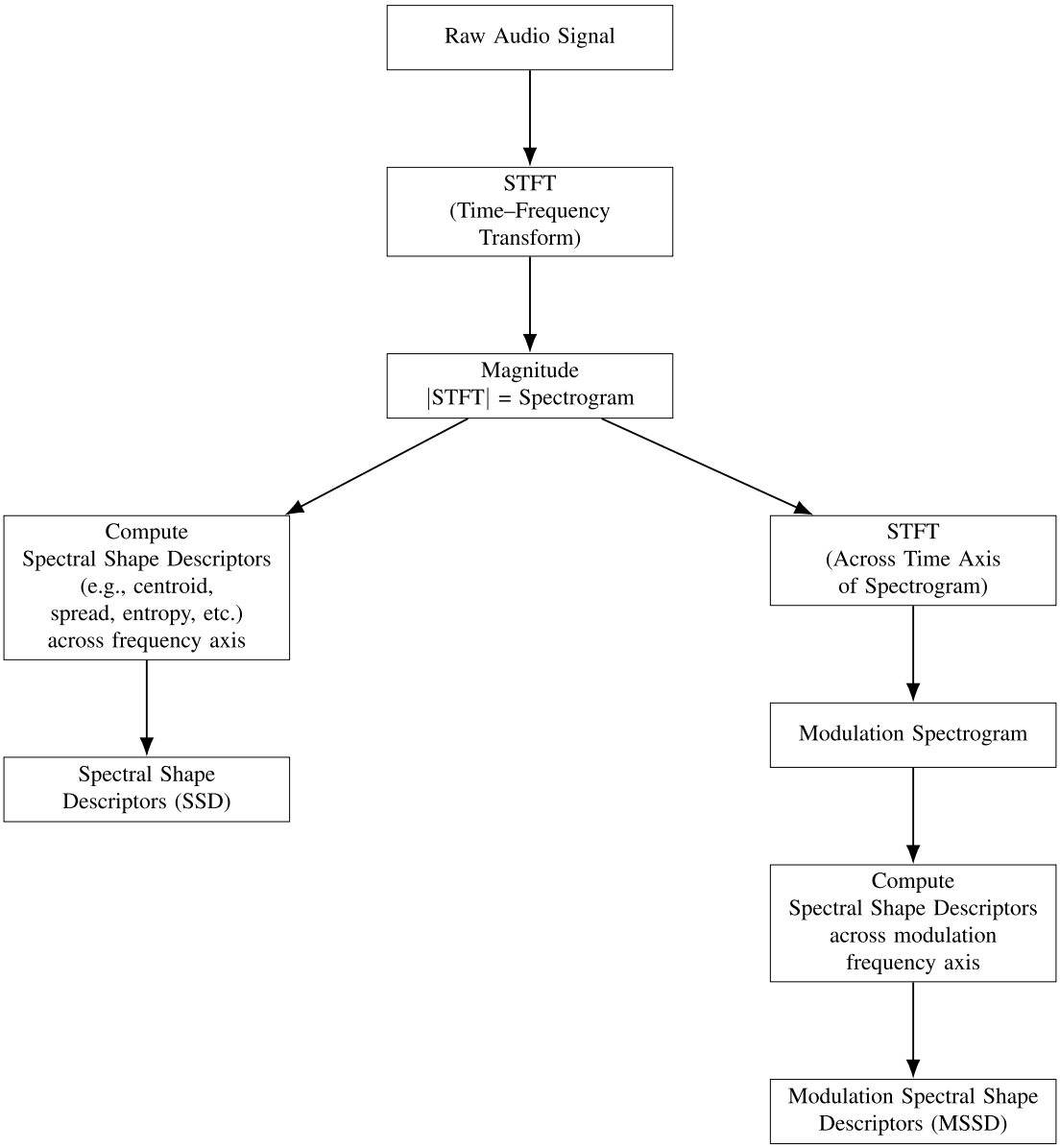


Fig. 2. Block diagram of the proposed acoustic feature extraction pipeline. The raw hive audio is first transformed into a time–frequency representation via STFT, followed by magnitude computation to obtain the spectrogram. From this, SSDs are computed across the frequency axis to capture short-term spectral structure. In parallel, a second STFT is applied across the time axis of the spectrogram to generate a modulation spectrogram, from which MSSDs are computed across the modulation frequency axis to capture the slower temporal modulations.

38 and a noninfested beehive on September 30, 2022. As the spectrograms show, the infected beehive shows lower intensity in specific bands of frequency (between 100 and 500 Hz), thus corroborating the findings in [11].

To generate the modulation spectrogram, a second STFT was applied to the temporal evolution of each frequency bin's magnitude using a 60-s window without overlap, producing a 3-D time–frequency–modulation representation. The 60-s window length was chosen to capture the slower amplitude modulations present in honey bee colony sounds, such as rhythmic wingbeat patterns and collective buzzing activity. Shorter windows primarily capture faster fluctuations, potentially missing these longer timescale modulation patterns, while longer windows reduce temporal resolution and may smooth out relevant dynamics. The modulation spectrogram was then averaged across time frames to obtain a 2-D time-averaged representation suitable for feature extraction. To reduce dimensionality while preserving relevant information, adjacent frequency and modulation bins were averaged, resulting in a final resolution of 50 Hz for the acoustic frequency axis and 5 Hz for the modulation frequency axis.

Fig. 4(a) and (b) shows the modulation spectrograms of Varroa-infected beehive and a healthy beehive, respectively. As can be seen, the infected beehive shows not only lower intensity in specific conventional frequency bands (frequency between 100 and 500 Hz) but also slower spectral variability over time, with infested hives showing the highest amount of energy in modulation frequencies between 0 and 10 Hz, while healthy hives show spectral dynamics up to nearly 20 Hz.

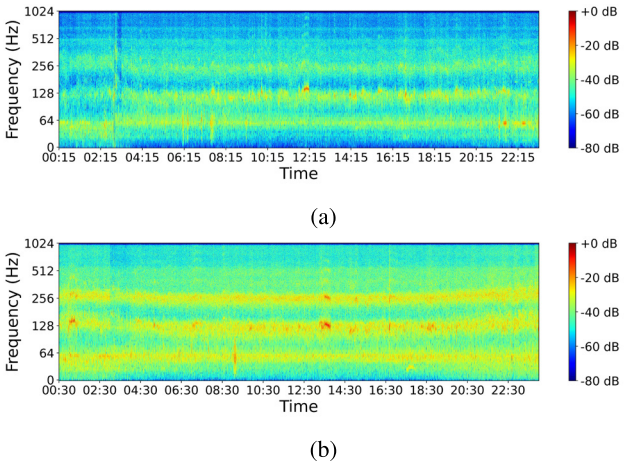


Fig. 3. 24-h spectrograms of (a) Varroa-infected beehive and (b) healthy beehive measured on September 30, 2022.

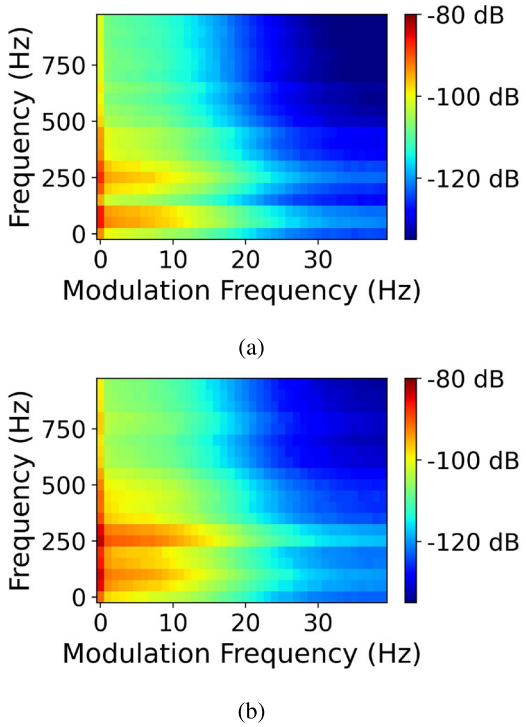


Fig. 4. Modulation spectrograms of (a) Varroa-infected beehive and (b) healthy beehive.

Together, it is expected that the conventional and modulation spectrograms will result in features that are not only compact but also complementary to each other. In our experiments, each 15-min audio segment is treated as a separate sample for analysis. To increase the sample size and capture temporal variability, we include the acoustic data not only from the day on which the Varroa mite inspection was performed (as illustrated in Fig. 1) but also from the preceding day.

2) Benchmarks: To evaluate the performance of the proposed features, two established feature sets are included as benchmarks for comparison. MFCCs approximate human auditory perception by applying a Mel-scale frequency transformation and have been extensively employed in precision beekeeping applications [5]. In this study, we extract

TABLE I
MODEL CONFIGURATIONS AND HYPERPARAMETER SETTINGS USED IN THE EXPERIMENTS

Classifier	Hyperparameters
Random Forest (RF)	n_estimators = 50, max_depth = 5, min_samples_split = 2
Naive Bayes (NB)	*No hyperparameters tuned*
Gradient Boosting (GB)	n_estimators = 50, max_depth = 3, learning_rate = 0.1
K-Nearest Neighbors (KNN)	n_neighbors = 5, weights = uniform
Multilayer Perceptron (MLP)	hidden_layer_sizes = (50,), alpha = 0.0001, max_iter = 500
Support Vector Machine (SVM)	C = 1, kernel = rbf, gamma = scale

12 MFCCs along with the zeroth coefficient, which captures signal energy, using 26 Mel-scale filters. In addition, LFCCs are extracted using a bank of linear filters to serve as a complementary baseline. For comparability with MFCCs, we extract 13 LFCCs (including the zeroth coefficient) using the same number of filters (26) but distributed linearly. This provides an alternative representation that emphasizes the uniform frequency resolution across the spectrum, in contrast to the perceptual warping of the Mel scale.

C. Classification

In this study, we frame the problem as a supervised binary classification task of determining whether a colony requires Varroa treatment. Specifically, we introduce a new variable, *Varroa label*, defined as follows:

$$\text{Varroa label} = \begin{cases} 0, & \text{if PVMI } < 3\% \text{ (Do not trigger treatment)} \\ 1, & \text{if PVMI } \geq 3\% \text{ (Trigger treatment)}. \end{cases}$$

This binary labeling allows the model to learn a clear distinction between colonies that require intervention (1) and those that do not (0). The threshold for applying treatment in apicultural management typically ranges between 2% and 5%, as colonies exceeding this infestation rate are at significant risk of rapid Varroa population growth, colony weakening, and potential collapse if left untreated [35], [36], [37]. In this study, when categorizing colonies based on the need for treatment, we adopted the same criteria. To remain conservative and ensure early intervention, we selected the lower bound of 3% as the cutoff for triggering treatment. This choice is further justified by the timing of the PVMI measurements, which were obtained in late summer—a period during which the cited guidelines specifically recommend a 3% infestation threshold.

Different classifiers are explored, including NB, RF, KNN, SVM, and MLP. For each of these classifiers, their hyperparameters are fine-tuned using a grid search. The fine-tuned hyperparameters are shown in Table I. To ensure robust evaluation of the models, a nested threefold cross validation is

TABLE II
LIGHTWEIGHT MINICNN MODEL ARCHITECTURE

Layer	Details
Conv2D	8 filters, 3×3, ReLU, same padding
MaxPool2D	2×2 pool
Conv2D	16 filters, 3×3, ReLU, same padding
MaxPool2D	2×2 pool
Flatten	–
Dense	16 units, ReLU, L2 regularization
Dense	1 unit, sigmoid activation

used ensuring hive-independence. This cross-validation process ensures that every data point is included in the testing set exactly once, reducing the risk of overfitting and providing a more reliable estimate of model performance. The choice of threefold is motivated by the limited number of available hives (seven in total): more folds would result in testing sets with only a single hive, producing unstable estimates, whereas fewer folds (e.g., two) would not sufficiently capture interhive variability and would leave too few hives for training. In our design, each fold contains approximately 2–3 hives for testing and 4–5 for training/validation, with folds constructed to have a roughly balanced number of samples. This provides an effective compromise among training diversity, test reliability, and computational feasibility.

In this work, we adopt a lightweight convolutional neural network primarily to provide a comparison with traditional machine learning models. This choice is motivated by the characteristics of bioacoustic datasets, which often exhibit substantial variability and limited sample sizes. While deep architectures with millions of parameters may easily overfit under such conditions, a compact CNN with only a few thousand parameters offers a practical and computationally efficient alternative. Importantly, this lightweight architecture enables the model to learn task-specific representations directly from the spectral and modulation spectrogram inputs, potentially capturing discriminative patterns beyond handcrafted features, while remaining comparable in complexity to classical approaches. Table II summarizes the architecture of the lightweight MiniCNN used in this study. The model consists of two convolutional layers with small kernel sizes and a modest number of filters, followed by max-pooling operations to reduce dimensionality. A fully connected layer with L2 regularization is included to promote stable learning, and the final output layer uses a sigmoid activation for binary classification.

D. Figures of Merit

To evaluate the performance of our classification models in detecting Varroa mite infestation, we employ several standard figures of merit: receiver operating characteristic area under the curve (ROC-AUC), $F1$-score, precision, and recall. These metrics are particularly useful in binary classification tasks, especially under class imbalance conditions.

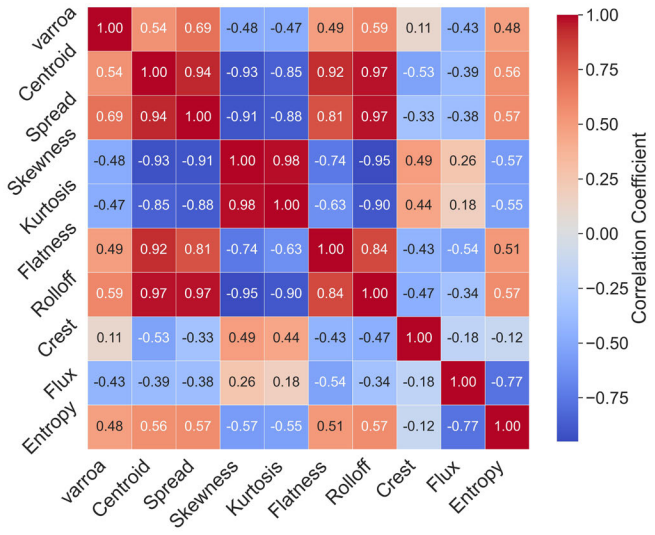


Fig. 5. Pearson correlation between all pairs of spectral descriptors extracted from the conventional spectrogram and the Varroa label.

The ROC-AUC metric measures the ability of the classifier to distinguish between classes and is computed as the area under the ROC curve, which plots the true-positive rate (TPR) against the false positive rate (FPR) at various threshold settings. A value closer to 1 indicates better performance.

The $F1$-score is the harmonic mean of precision and recall, providing a single metric that balances both false positives and false negatives

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Precision measures the proportion of true-positive predictions among all positive predictions

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}.$$

Recall, also known as sensitivity or true-positive rate, quantifies the proportion of actual positive cases that are correctly identified

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}.$$

Together, these metrics offer a comprehensive assessment of classification performance, highlighting the model's ability to correctly identify infestation cases while minimizing false alarms.

III. RESULTS AND DISCUSSION

In this section, we present the experimental results and discuss them in light of the existing literature.

A. Feature Analysis

1) Conventional Spectrogram: Fig. 5 displays the Pearson correlation coefficients between all pairs of spectral descriptors of the conventional spectrogram and the Varroa label. As can be seen, a positive correlation is observed between the Varroa label and some of the features, such as centroid, spread, flatness, roll-off, and entropy, suggesting that higher levels

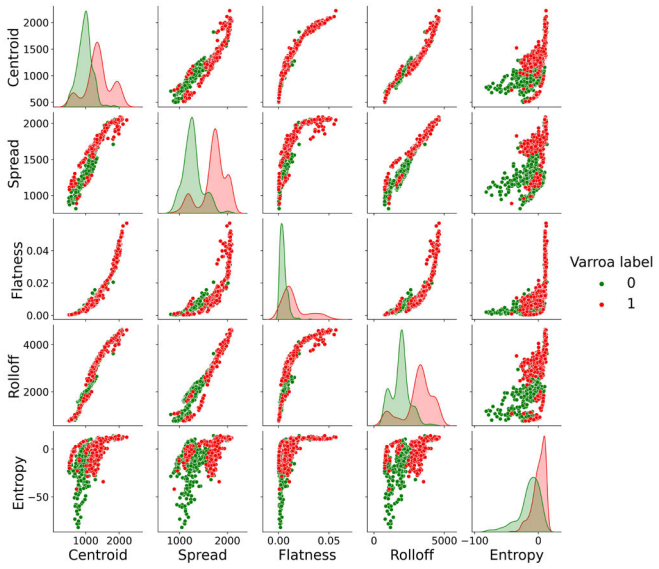


Fig. 6. Pairplot illustrating the relationships among SSDs: centroid, spread, flatness, entropy, and roll-off. Samples with a Varroa label equal to 0 represent healthy hives, while those with a Varroa label equal to 1 indicate infestation.

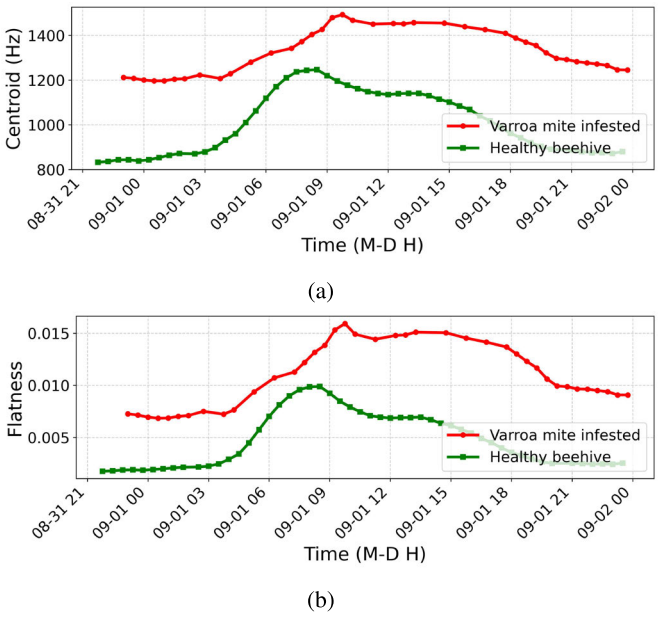


Fig. 7. Temporal dynamics of the conventional spectrogram (a) centroid and (b) flatness measured across the entire day on September 1, 2022.

of Varroa infestation are associated with acoustic spectra that become more "noisy" and less structured.

In particular, the positive correlation with entropy indicates that as mite infestation increases, the hive acoustic spectrum becomes more disordered and information-rich, losing its harmonic structure and resembling a more random distribution of energy. Similarly, the positive correlation with flatness suggests that the spectrum tends to flatten under higher infestation levels, meaning the distinction between dominant and nondominant frequencies diminishes. The correlations with centroid, spread, and roll-off imply that energy shifts toward higher frequencies and becomes more broadly distributed, which could reflect increased agitation or stress-induced changes in honey bee buzzing patterns. Altogether, these descriptors provide evidence that Varroa infestation influences the short-term frequency structure of hive sounds, making them less tonal and more broadband in nature. On the other hand, negative correlations are observed with skewness, kurtosis, and flux. The reduction in skewness and kurtosis with increasing Varroa levels suggests that the spectral distribution becomes less asymmetric and less peaked, indicating a move away from well-defined tonal components toward flatter, broadband spectra. The negative correlation with flux implies that the temporal variability of the spectrum decreases with higher infestation, meaning that the hive sound becomes more stationary and less dynamic.

Fig. 6 illustrates a pairwise comparison of SSDs features, stratified by the Varroa infestation label. To ensure the balanced representation across categories, the dataset was filtered to have an equal number of samples randomly selected from each category. Samples from the beehives with Varroa mite infestation show higher values in spectral centroid, spread, roll-off, and entropy.

To visualize the temporal dynamics of the two proposed SSDs, we plotted their smoothed values over a 1-day period (September 1, 2022) for two representative hives: one exhibiting a Varroa mite infestation (Hive ID: 3631) and one healthy hive (Hive ID: 6), as shown in Fig. 7. Each curve represents an eight-point moving average for a specific feature, allowing us to observe trends and fluctuations over time. As shown in Fig. 7(a) and (b) for the conventional spectral features, distinct patterns emerge between the infested and healthy hives, with features demonstrating elevated values in the presence of Varroa, highlighting their potential for infestation discrimination.

2) Modulation Spectrogram: Fig. 8, in turn, presents joint plots comparing the flux and entropy features extracted from both the spectrogram and the modulation spectrogram. These visualizations highlight the relationship between corresponding spectral and modulation-based descriptors, offering insight into how each varies with respect to Varroa infestation status. The joint plots reveal distinct patterns associated with the Varroa infestation. Entropy features show that healthy hives tend to exhibit more consistent and less dispersed modulation entropy values, suggesting reduced acoustic variability. Similarly, flux features indicate that Varroa-positive samples are concentrated in regions of lower spectral and modulation flux, reflecting less dynamic frequency changes.

Fig. 9 shows the temporal dynamics of the two proposed SSDs of the modulation spectrogram over a 1-day period (September 1, 2022) for two representative hives: one exhibiting a Varroa mite infestation (Hive ID: 3631) and one healthy hive (Hive ID: 6). While the spectral centroid shows a fairly stable separation between healthy and infested hives in Fig. 7(a), the modulation centroid and flatness shown in Fig. 9(a) and (b) highlight a more dynamic divergence over time, with the gap between the two conditions becoming more pronounced at specific periods of the day. Taken together, these results suggest that Varroa infestation affects both the instantaneous frequency structure (spectral centroid) and the

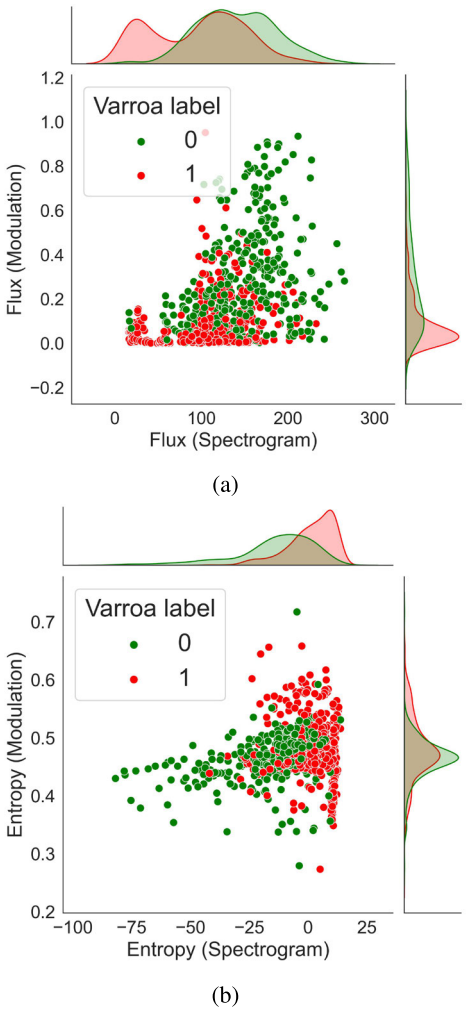


Fig. 8. Pairplot illustrating the relationship between conventional and MSSDs (a) flux and (b) entropy. Samples with Varroa equal to 0 represent healthy hives, while those with Varroa equal to 1 indicate infestation.

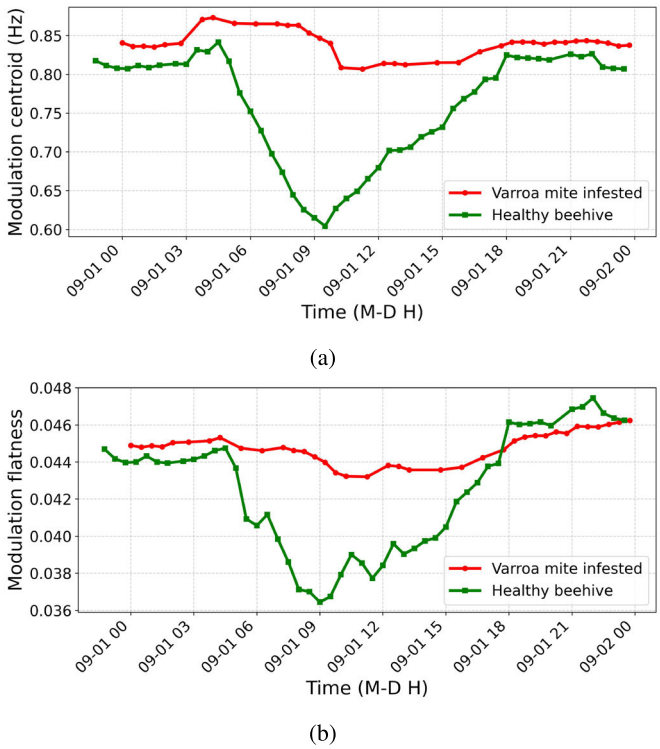


Fig. 9. Temporal dynamics of the modulation spectrogram (a) centroid and (b) flatness measured across the entire day on September 1, 2022.

temporal modulation dynamics (modulation centroid), but in different ways.

B. Classification Performance

Table III presents a comparative evaluation of the classification performance across different benchmarks and proposed feature sets, as well as tested classifiers. Each metric is an average across the threefold. As can be seen, the benchmark MFCCs and LFCCs showed modest performance across all classifiers, with SVM and MLP generally outperforming others. For instance, MFCCs with SVM reached an ROC-AUC of 0.571. LFCCs performed slightly better overall, with SVM reaching a ROC-AUC of 0.767. Notwithstanding, these results were consistently lower than those achieved with the proposed SSD and/or MSSD features.

The SSD set, derived from spectrograms, significantly outperformed MFCCs and LFCCs. Notably, SVM with SSD features achieved an ROC-AUC of 0.871 and an $F1$-score of 0.832, with high precision (0.867), indicating strong discriminative power. This underscores the value of capturing spectral shape characteristics, such as centroid and spread, in highlighting signal structure. Next, the MSSD features, computed from the modulation spectrogram along the modulation frequency axis, provided a complementary perspective by emphasizing temporal modulations in spectral content.

Finally, to test the complementarity of the two feature sets, we tested their fusion. As can be seen, the concatenation of SSD and MSSD led to the highest overall performance across all classifiers, confirming the complementarity of the two feature types, corroborating the patterns seen in Figs. 7–9. The best results were obtained using SVM, which achieved a ROC-AUC of 0.874, an $F1$-score of 0.859, a precision of 0.863, and a recall of 0.862. Several models showed clear improvements in ROC-AUC with the fused feature set. For instance, RF increased from 0.678 to 0.756 and KNN from 0.758 to 0.800. Notably, SVM achieved the best performance overall, improving from 0.871 (SSD) to 0.874 (SSD+MSSD).

To complement the averaged performance metrics ($F1$, AUC, precision, and recall), we also report the aggregated confusion matrix computed across all three test folds, as shown in Fig. 10. Rather than presenting a single fold's results which may be sensitive to data partitioning, we sum the confusion matrices from each fold to obtain a more stable and representative view of the model's overall classification behavior. This aggregated matrix reflects the total counts of true positives, true negatives, false positives, and false negatives observed across all testing sets, providing a clearer indication of the model's global performance and error patterns. In this analysis, we report the confusion matrix for the SVM classifier, as it was the strongest performing traditional machine learning model in our experiments.

TABLE III
COMPARISON OF CLASSIFICATION PERFORMANCE ACROSS DIFFERENT FEATURE SETS AND CLASSIFIERS. ARROWS INDICATE THE DIRECTION OF CHANGE IN PERFORMANCE: ↑ DENOTES IMPROVEMENT, WHILE ↓ INDICATES A DECREASE

Feature	Classifier	ROC-AUC	F1	Precision	Recall
MFCCs	RF	0.736	0.416	0.991	0.281
	GB	0.614	0.403	0.592	0.329
	NB	0.263	0.108	0.198	0.090
	KNN	0.457	0.392	0.753	0.353
	MLP	0.493	0.213	0.515	0.145
	SVM	0.571	0.057	0.057	0.057
LFCCs	RF	0.347	0.520	0.531	0.558
	GB	0.438	0.418	0.394	0.458
	NB	0.356	0.164	0.581	0.103
	KNN	0.576	0.545	0.665	0.585
	MLP	0.530	0.568	0.563	0.655
	SVM	0.707	0.534	0.662	0.540
SSD	RF	0.678	0.673	0.717	0.644
	GB	0.700	0.655	0.731	0.603
	NB	0.841	0.807	0.839	0.793
	KNN	0.758	0.784	0.836	0.743
	MLP	0.780	0.710	0.752	0.763
	SVM	0.871	0.832	0.867	0.811
MSSD	RF	0.744	0.633	0.682	0.654
	GB	0.738	0.653	0.689	0.671
	NB	0.678	0.449	0.572	0.450
	KNN	0.726	0.555	0.660	0.561
	MLP	0.810	0.616	0.704	0.631
	SVM	0.739	0.628	0.721	0.638
SSD+MSSD	RF	↑ 0.756	0.715	0.760	0.686
	GB	↑ 0.755	0.677	0.739	0.634
	NB	↑ 0.872	0.818	0.860	0.787
	KNN	↑ 0.800	0.827	0.845	0.812
	MLP	↓ 0.713	0.674	0.715	0.644
	SVM	↑ 0.874	0.859	0.863	0.862

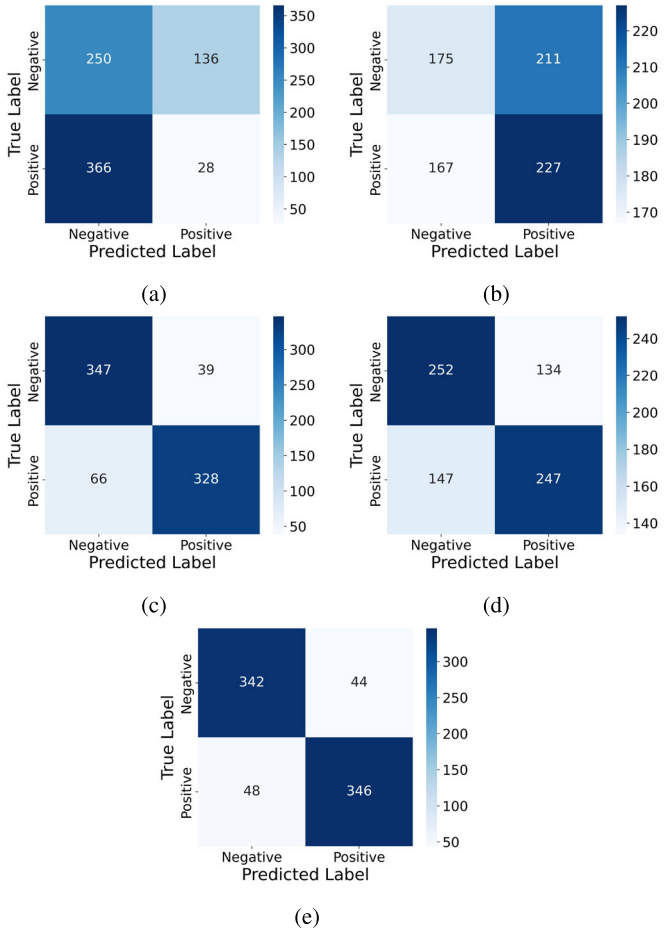


Fig. 10. Aggregated confusion matrices for each feature set evaluated across all three test folds. Each matrix summarizes the total true negatives (TNs), false positives (FPs), false negatives (FNs), and true positives (TPs) obtained for (a) MFCCs, (b) LFCCs, (c) SSD, (d) MSSD, and (e) proposed SSD+MSSD feature set.

Benchmark features (MFCCs and LFCCs) exhibit substantial confusion between classes, with MFCCs showing a high number of false negatives and LFCCs producing many false positives. In contrast, our proposed SSD and MSSD features demonstrate much stronger discriminative power, with SSD in particular yielding a balanced reduction in both error types. The combined SSD+MSSD feature set achieves the most favorable outcome overall, with the highest true-positive count and the lowest number of misclassifications.

C. Comparative Performance of CNN and Traditional Machine Learning Models

Table IV shows a comparative evaluation of the classification performance across different features. The results show that the benchmarks such as MFCCs, LFCCs, and spectrogram inputs yield relatively modest performance, with ROC-AUC values close to 0.55–0.60. In contrast, the modulation spectrogram achieves substantially higher accuracy across all metrics—particularly recall—highlighting its superior ability to capture the temporal–spectral patterns associated with Varroa infestation.

When comparing the CNN results with the traditional machine learning models, it becomes evident that classical approaches remain highly effective for this task while being far less computationally demanding. Models trained on the handcrafted SSD and MSSD features consistently outperform the CNN across most metrics, with several classifiers—such as SVM, KNN, and NB—achieving ROC-AUC values above 0.85, surpassing the best CNN configuration. These results indicate that the spectral shape and modulation descriptors capture the discriminative acoustic characteristics sufficiently well that even lightweight classifiers can exploit them efficiently. In contrast, the CNN requires substantially more computation, longer training time, and more data to reach competitive performance, yet still does not exceed the strongest traditional ML models. Overall, this comparison underscores that carefully designed feature representations combined with

TABLE IV
COMPARISON OF CLASSIFICATION PERFORMANCE ACROSS DIFFERENT FEATURE SETS USING THE CNN MODEL

Feature	ROC-AUC	F1	Precision	Recall
MFCCs	0.593	0.573	0.549	0.648
LFCCs	0.546	0.498	0.563	0.498
Spectrogram	0.543	0.528	0.551	0.541
Modulation Spectrogram	0.843	0.754	0.715	0.869

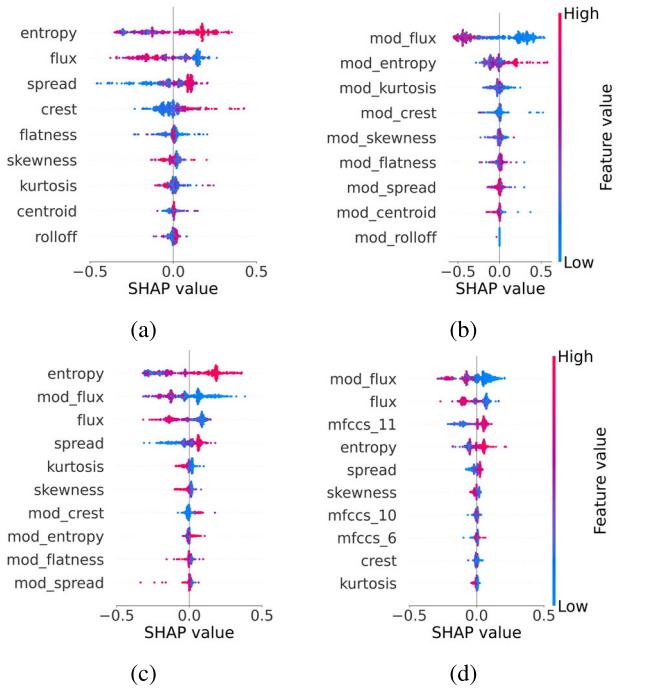


Fig. 11. SHAP summary plots illustrating the contribution of individual features to the model's predictions for (a) SSD, (b) MSSD, (c) SSD+MSSD, and (d) SSD+MSSD+MFCCs. Features are ranked by their mean absolute SHAP value, indicating overall importance, while color encodes the relative feature value (low to high) and its directional impact on the model output.

classical machine learning techniques offer a more accurate, interpretable, and computationally efficient solution than deep learning for this specific hive-acoustic classification problem.

D. Feature Contribution Analysis Using SHAP

To quantify the individual and complementary contributions of SSD and MSSD, we conducted a SHapley Additive exPlanations (SHAP)-based [38] ablation analysis. While Section III reported classification performance for different feature sets, the goal of this analysis is to understand why each feature group improves the model and to what extent the SSD and MSSD components contribute to the final prediction. For this interpretability study, we used an RF classifier. RF was selected because tree-based models are fully compatible with SHAP's TreeExplainer, which provides exact and computationally efficient SHAP value calculations.

Fig. 11 shows SHAP value plots for SSD, MSSD, SSD+MSSD, and the combined feature set including MFCCs. In all plots, features are ranked by mean absolute SHAP value, with color indicating feature magnitude (blue = low and pink = high).

Among SSD features, entropy, flux, and spread are the most influential predictors. High-entropy values strongly drive the model toward predicting infestation, reflecting increased spectral disorder in infested colonies. In MSSD, modulation flux and modulation entropy dominate the SHAP rankings, with high modulation entropy consistently increasing the likelihood of infestation.

When combining the spectrogram and modulation spectrogram descriptors, spectral entropy and modulation flux remain the strongest contributors, highlighting the importance of temporal variability as a marker of Varroa presence. Adding MFCCs to this combined set, the modulation flux and spectral entropy still emerge as the top features. While some MFCC features appear in the top 10, most of the leading predictors come from the spectrogram and modulation spectrogram descriptors.

IV. Conclusion

In this study, we addressed the challenge of detecting a Varroa mite infestation in honey bee colonies using remote acoustic analysis. We proposed the use of SSDs, extracted from both the conventional and modulation spectrograms, as informative and complementary features for distinguishing between healthy and infested hive states. To ensure generalizability, all models were evaluated in a hive-independent setting using various classifiers. The results demonstrate that SSD features consistently outperform traditional features, such as MFCCs and LFCCs across all metrics and classifiers. Furthermore, the fusion of the proposed modulation-based descriptors with the SSD features resulted in the highest overall performance. A comparative analysis with a CNN model further revealed that classical machine learning approaches trained on the proposed handcrafted features not only achieve higher accuracy but also require significantly lower computational complexity. Overall, these results highlight the benefit of capturing details about both the spectral structure and the temporal modulation patterns in the beehive acoustics signal for accurate Varroa mite infestation detection.

Acknowledgment

The authors acknowledge Nectar Technologies Inc. for their support with data collection. The authors would like to thank Evan Henry for his valuable contributions to the dataset collection in the field.

References

[1] FAO, Apimondia, CAAS, IZSLT, *Good Beekeeping Practices for Sustainable Apiculture*, Food and Agriculture Organization, Rome, Italy, 2021.

[2] A. Gray et al., "Honey bee colony winter loss rates for 35 countries participating in the COLOSS survey for winter 2018–2019, and the effects of a new queen on the risk of colony winter loss," *J. Apicultural Res.*, vol. 59, no. 5, pp. 744–751, Oct. 2020.

[3] K. S. Traynor et al., "Varroa destructor: A complex parasite, crippling honey bees worldwide," *Trends Parasitol.*, vol. 36, no. 7, pp. 592–606, Jul. 2020.

[4] M. Alleri et al., "Recent developments on precision beekeeping: A systematic literature review," *J. Agricult. Food Res.*, vol. 14, Dec. 2023, Art. no. 100726.

[5] M. Abdollahi, P. Giovenazzo, and T. H. Falk, "Automated beehive acoustics monitoring: A comprehensive review of the literature and recommendations for future work," *Appl. Sci.*, vol. 12, no. 8, p. 3920, Apr. 2022.

[6] 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 Eng. Soc. Conv.*, vol. 144, pp. 1–24, May 2018.

[7] M. Abdollahi et al., "UrBAN: Urban beehive acoustics and PheNotyping dataset," *Sci. Data*, vol. 12, no. 1, p. 536, Mar. 2025.

[8] Y. Zhu et al., "MSPB: A longitudinal multi-sensor dataset with phenotypic trait measurements from honey bees," *Sci. Data*, vol. 11, no. 1, p. 860, Aug. 2024.

[9] P. Janetzky, M. Schaller, A. Krause, and A. Hotho, "Swarming detection in smart beehives using auto encoders for audio data," in *Proc. 30th Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jun. 2023, pp. 1–5.

[10] A. Terenzi, N. Ortolani, I. Nolasco, E. Benetos, and S. Cecchi, "Comparison of feature extraction methods for sound-based classification of honey bee activity," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 112–122, 2022.

[11] S.-H. Chen et al., "A machine learning-based multiclass classification model for bee colony anomaly identification using an IoT-based audio monitoring system with an edge computing framework," *Expert Syst. Appl.*, vol. 255, Dec. 2024, Art. no. 124898.

[12] L. Barbisan, G. Turvani, and F. Riente, "A machine learning approach for queen bee detection through remote audio sensing to safeguard honeybee colonies," *IEEE Trans. AgriFood Electron.*, vol. 2, no. 2, pp. 236–243, Sep. 2024.

[13] M. Abdollahi, E. Henry, P. Giovenazzo, and T. H. Falk, "The importance of context awareness in acoustics-based automated beehive monitoring," *Appl. Sci.*, vol. 13, no. 1, p. 195, Dec. 2022.

[14] Y. Zhu et al., "Early prediction of honeybee hive winter survivability using multi-modal sensor data," in *Proc. IEEE Int. Workshop Metrology Agricult. Forestry (MetroAgriFor)*, Nov. 2023, pp. 657–662.

[15] A. Qandour, I. Ahmad, D. Habibi, and M. Leppard, "Remote beehive monitoring using acoustic signals," *Acoust. Aust.*, vol. 42, no. 3, p. 205, 2014.

[16] P. Mekha, N. Teeyasuksaet, T. Sompowloy, and K. Osathanunkul, "Honey bee sound classification using spectrogram image features," in *Proc. Joint Int. Conf. Digit. Arts, Media Technol. ECTI Northern Conf. Electr., Electron., Comput. Telecommun. Eng. (ECTI DAMT NCON)*, Jan. 2022, pp. 205–209.

[17] U. Libal and P. Biernacki, "MFCC selection by LASSO for honey bee classification," *Appl. Sci.*, vol. 14, no. 2, p. 913, Jan. 2024.

[18] N. Di, M. Z. Sharif, Z. Hu, R. Xue, and B. Yu, "Applicability of VGGish embedding in bee colony monitoring: Comparison with MFCC in colony sound classification," *PeerJ*, vol. 11, Jan. 2023, Art. no. e14696.

[19] W. Huang, W. Yang, Z. Luo, J. Qi, T. Zhang, and X. Kong, "Sound-based bee colony state analysis using compact MFCC patterns," in *Proc. IEEE Int. Symp. Parallel Distrib. Process. with Appl. (ISPA)*, Oct. 2024, pp. 2164–2169.

[20] H. R. Guimarães et al., "Benchmarking self-supervised audio representations for IoT-enabled acoustic beehive monitoring," *IEEE Internet Things J.*, vol. 12, no. 21, pp. 45000–45010, Nov. 2025.

[21] T. H. Falk, S. Stadler, W. B. Kleijn, and W.-Y. Chan, "Noise suppression based on extending a speech-dominated modulation band," in *Proc. Interspeech*, Aug. 2007, pp. 970–973.

[22] B. Cauchi, K. Siedenburg, J. F. Santos, T. H. Falk, S. Doclo, and S. Goetze, "Non-intrusive speech quality prediction using modulation energies and LSTM-network," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 7, pp. 1151–1163, Jul. 2019.

[23] A. Tiwari, R. Cassani, S. Kshirsagar, D. P. Tobon, Y. Zhu, and T. H. Falk, "Modulation spectral signal representation for quality measurement and enhancement of wearable device data: A technical note," *Sensors*, vol. 22, no. 12, p. 4579, Jun. 2022.

[24] P. Singh, M. Sahidullah, and G. Saha, "Modulation spectral features for speech emotion recognition using deep neural networks," *Speech Commun.*, vol. 146, pp. 53–69, Jan. 2023.

[25] M. Abdollahi et al., "Audio modulation spectral features for improved honeybee colony population prediction," *IEEE Sensors J.*, vol. 25, no. 24, pp. 44378–44391, Dec. 2025.

[26] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," 2013, *arXiv:1302.4964*.

[27] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.

[28] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[29] E. Fix, *Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties*. Wright-Patterson AFB, OH, USA: USAF School of Aviation Medicine, 1951.

[30] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sep. 1995.

[31] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[32] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[33] V. Dietemann et al., "Standard methods for varroa research," *J. Apicultural Res.*, vol. 52, no. 1, pp. 1–54, Jan. 2013.

[34] G. Peeters et al., "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," *CUIDADO Ist Project Rep.*, vol. 54, pp. 1–25, Apr. 2004.

[35] H. B. H. Coalition, "Tools for Varroa management: A guide to effective Varroa sampling & control," Honey Bee Health Coalition, CO, USA, 2022, pp. 1–37.

[36] Virginia Tech Entomology. (2021). *Varroa Mite Sampling Methods*. [Online]. Available: https://www.ento.vt.edu/the-bee-group-at-vt/beekeeping/mites2.html

[37] Ontario Beekeepers' Association. (2017). *Treatment Recommendations*. Accessed: Aug. 18, 2025. [Online]. Available: https://www.ontariobee.com/sites/ontariobee.com/files/Treatment%20Recommendations%20201 7-11-02.pdf

[38] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.