Université du Québec

Institut national de la recherche scientifique

Centre Énergie Matériaux Télécommunications

ACOUSTIC SIGNAL PROCESSING AND MACHINE LEARNING FOR REMOTE AUTOMATED BEEHIVE MONITORING

TRAITEMENT DU SIGNAL ACOUSTIQUE ET APPRENTISSAGE AUTOMATIQUE POUR LA SURVEILLANCE AUTOMATISÉE ET À DISTANCE DES RUCHES

Par

Mahsa Abdollahi

Thèse présentée pour l'obtention du grade de
Philosophiae Doctor (Ph.D.)
en sciences de télécommunications

Jury d'évaluation

Président du jury et examinateur interne	Dr. Shervin Vakili INRS-EMT, Canada
Examinateur externe 1	Dr. Nils Peters Trinity College Dublin, Ireland
Examinateur externe 2	Dr. Sri Krishnan Toronto Metropolitan University, Canada
Directeur de recherche	Dr. Tiago H. Falk INRS-EMT, Canada
Codirecteur de recherche	Dr. Pierre Giovenazzo Laval University, Canada

© Droits réservés de Mahsa Abdollahi, Avril 2026

REMERCIEMENTS

I would like to acknowledge the invaluable support and supervision of Prof. Tiago Falk and Prof. Pierre Giovenazzo. I am also grateful to my colleagues and collaborators from the MuSAE Lab, as well as to research collaborators from other institutes, including Heitor R. Guimarães, Yi Zhu, Evan Henry, and Nico Coallier, for their contributions and insightful discussions.

I would also like to express my deepest gratitude to my family for their unwavering love, encouragement, and constant support throughout this journey. Their patience and belief in me have been a continual source of strength. Finally, I am profoundly thankful to my husband for his understanding, encouragement, and steadfast support during the challenges and milestones of my doctoral studies. His presence and support made this journey both possible and meaningful.

RÉSUMÉ

La surveillance acoustique est apparue comme une approche non invasive prometteuse pour évaluer la santé des colonies d'abeilles mellifères et la dynamique de leurs populations. Cependant, le développement de systèmes fiables et généralisables demeure limité par le manque de jeux de données publics, le bruit environnemental et la sensibilité des caractéristiques acoustiques conventionnelles aux conditions réelles. Cette thèse aborde ces défis par l'introduction d'un jeu de données à grande échelle, une analyse systématique du bruit environnemental, la conception de caractéristiques acoustiques robustes et l'évaluation de modèles d'apprentissage classique et profond.

Premièrement, nous introduisons UrBAN (Urban Beehive Acoustics and PheNotyping), un jeu de données ouvert à grande échelle comprenant plus de 3 000 heures d'enregistrements audio bruts collectés à partir de dix ruches urbaines sur deux années consécutives. Le jeu de données est enrichi par des mesures multimodales issues de capteurs, incluant la température et l'humidité internes de la ruche, les conditions météorologiques externes, ainsi que des annotations phénotypiques collectées manuellement, telles que la taille de la population de la colonie, les niveaux d'infestation par l'acarien Varroa, le statut de la reine et la mortalité hivernale. UrBAN constitue une référence réaliste pour l'évaluation longitudinale et indépendante des ruches des méthodes de surveillance acoustique.

Deuxièmement, nous étudions de manière systématique l'impact de sources importantes de bruit environnemental, incluant le bruit de fond urbain, la parole humaine et les fortes pluies, sur les représentations acoustiques couramment utilisées. Notre analyse révèle que les caractéristiques conventionnelles telles que les spectrogrammes et les Coefficients Cepstraux en Fréquences Mel (CCFM) sont très sensibles au bruit, entraînant une dégradation significative des performances dans des conditions de déploiement réalistes. Ces résultats soulignent la nécessité de représentations robustes au bruit pour la surveillance des ruches en conditions réelles.

Troisièmement, nous proposons et analysons des caractéristiques acoustiques basées sur la modulation, incluant les spectrogrammes de modulation et les tensorgrammes de modulation, qui capturent explicitement les fluctuations temporelles d'énergie associées à l'activité collective des abeilles. Nous montrons que l'utilisation de spectrogrammes de modulation pour améliorer la prédiction de la force des colonies dans des conditions inédites conduit à des taux d'erreur plus faibles et à des corrélations plus élevées, démontrant ainsi leur robustesse accrue pour la surveillance des ruches en conditions naturelles. Ensuite, nous abordons le défi de la détection de l'infestation par l'acarien Varroa dans les colonies d'abeilles. Nous proposons l'utilisation de descripteurs de forme spectrale, extraits à la fois des spectrogrammes conventionnels et des spectrogrammes de modulation, comme caractéristiques informatives et complémentaires pour distinguer les états de ruches saines et infestées. Les résultats démontrent que nos caractéristiques proposées surpassent systématiquement les caractéristiques traditionnelles, telles que les CCFM et les Coefficients Cepstraux en Fréquences Linéaires (CCFL) pour toutes les métriques et tous les classificateurs. Nous réalisons également une analyse comparative avec un modèle CNN, qui montre que les approches d'apprentissage automatique classiques entraînées sur les caractéristiques proposées atteignent non seulement une précision plus élevée, mais nécessitent aussi une complexité computationnelle significativement plus faible.

Enfin, nous évaluons plusieurs architectures d'apprentissage profond, incluant des réseaux neuronaux convolutionnels 2D et 3D ainsi que des modèles convolutionnels-récurrents. Les modèles profonds opérant sur les spectrogrammes de modulation et les tensorgrammes surpassent systématiquement les approches de référence, particulièrement dans les scénarios d'évaluation indépendants des ruches. Les résultats montrent que la modélisation explicite des motifs de modulation spectro-temporels améliore la généralisation entre les ruches et les conditions d'enregistrement.

Dans l'ensemble, ce travail démontre que la combinaison de données réelles à grande échelle, d'analyses tenant compte du bruit, de représentations acoustiques basées sur la modulation et de modèles d'apprentissage profond spatio-temporels permet une surveillance acoustique robuste et généralisable des colonies d'abeilles. Le cadre proposé fait progresser l'état de l'art en apiculture de précision et fournit une base pour l'évaluation évolutive et non invasive de la santé des colonies dans des environnements réels.

Mots-clés Acoustique des ruches, Abeilles domestiques, Spectrogramme de modulation, Infestation par le varroa.

ABSTRACT

Acoustic monitoring has emerged as a promising non-invasive approach for assessing honey bee colony health and population dynamics. However, the development of reliable and generalizable systems remains challenged by limited public datasets, environmental noise, and the sensitivity of conventional acoustic features to real-world conditions. This dissertation addresses these challenges through the introduction of a large-scale dataset, a systematic analysis of the impacts of environmental noise, the design of robust acoustic features, and the evaluation of both classical and deep learning models for remote beehive monitoring.

First, we introduce UrBAN (Urban Beehive Acoustics and PheNotyping), a large-scale, open-access dataset comprising over 3,000 hours of raw audio recordings collected from ten urban beehives over a period of two consecutive years. The dataset is enriched with multimodal sensor measurements, including internal hive temperature and humidity, external weather conditions, and manually collected phenotypic annotations, such as colony population size, Varroa mite infestation levels, queen bee status, and winter mortality. UrBAN provides a realistic benchmark for longitudinal and hive-independent evaluation of acoustic monitoring methods.

Second, we systematically investigate the impact of prominent environmental noise sources including urban background noise, speech, and heavy rainfall on commonly used acoustic representations. Our analysis reveals that conventional features such as spectrograms and mel-frequency cepstral coefficients (MFCCs) are highly sensitive to noise, leading to significant performance degradation under realistic deployment conditions. These findings highlight the necessity of noise-robust representations for real-world beehive monitoring.

Third, we propose and analyze modulation spectrum based acoustic features, including modulation spectrograms and tensorgrams, which explicitly capture spectral-temporal energy fluctuations associated with collective bee activity. We show that by using modulation spectrogram features, improved beehive strength prediction is obtained across unseen conditions (i.e., lower prediction error rates and higher correlations with beekeeper ratings), highlighting their robustness for “in the wild” beehive monitoring.

Next, we address the challenge of detecting Varroa mite infestations in honey bee colonies. We propose the use of spectral shape descriptors, extracted from both the conventional and modulation spectrograms, as informative and complementary features for distinguishing between healthy and infested hives. The results demonstrate that our proposed features consistently outperform traditional features, such as MFCCs and linear frequency cepstral coefficients (LFCCs), across several tested figures-of-merit and classifiers. We compare the accuracy of handcrafted features with that of deep neural network feature representations and show that classical machine-learning approaches trained on the proposed handcrafted features not only achieve higher accuracy, but also require significantly lower computational complexity.

Finally, we evaluate a wide range of deep learning neural network architectures, including 2D and 3D convolutional neural networks and convolutional-recurrent models. Deep models operating on modulation spectrograms and tensorgrams consistently outperform benchmark approaches, particularly in hive-independent evaluation settings. The results show that explicitly modeling spectro-temporal modulation patterns improves generalization across hives and recording conditions.

Overall, this work demonstrates that combining large-scale real-world data, noise-aware analysis, modulation-based feature representations, and spatio-temporal deep learning models enables robust and generalizable acoustic monitoring of honey bee colonies. The proposed framework advances the state-of-the-art in precision apiculture and provides a foundation for scalable, non-invasive colony health assessment in real-world environments.

Keywords Beehive acoustics, Honey bees, Modulation spectrogram, Varroa mite infestation.

TABLE DES MATIÈRES

REMERCIEMENTS ..... **iii**

RÉSUMÉ ..... **v**

ABSTRACT ..... **vii**

TABLE DES MATIÈRES ..... **ix**

LISTE DES FIGURES ..... **xii**

LISTE DES TABLEAUX ..... **xix**

SYNOPSIS **1**

0.1 CHAPITRE 1 : INTRODUCTION 1

0.1.1 Acoustique des abeilles mellifères 1

0.1.2 Défis actuels et problèmes de recherche 3

0.2 CHAPITRE 2 : SURVEILLANCE ACOUSTIQUE DES RUCHES : REVUE DE L'ÉTAT DE L'ART 5

0.2.1 Portée et Méthodologie de la Revue 5

0.2.2 Applications de la Surveillance Acoustique des Abeilles 5

0.2.3 Conception Expérimentale et Pratiques de Collecte de Données 6

0.2.4 Acquisition et Prétraitement du Signal 6

0.2.5 Représentations Audio et Ingénierie des Caractéristiques 7

0.2.6 Stratégies d'Apprentissage Automatique et de Modélisation 7

0.2.7 Protocoles d'Évaluation et Reproductibilité 7

0.2.8 Synthèse et Lacunes Identifiées 8

0.3 CHAPITRE 3 : JEU DE DONNÉES URBAN 8

0.4 CHAPITRE 4 : IMPACT DU BRUIT ENVIRONNEMENTAL ET DU PRÉTRAITEMENT 9

0.4.1 Facteurs de bruit environnemental 9

0.4.2 Détecteurs d'activité vocale pour la surveillance acoustique des ruches 11

0.4.3 Amélioration audio 12

0.4.4 Conclusion 13

0.5 CHAPITRE 5 : CARACTÉRISTIQUES BASÉES SUR LE SPECTROGRAMME DE MODULATION 14

0.5.1 Prédiction de la population des colonies 14

0.5.2 Prédiction non invasive de l'infestation par le varroa dans les colonies d'abeilles ... 16

0.6 Chapitre 6 : Apprentissage profond à partir du spectrogramme de modulation et du tensorgramme ... 19

0.6.1 Représentations basées sur la modulation ... 19

0.6.2 Architectures d'apprentissage profond ... 20

0.6.3 Configuration expérimentale et évaluation ... 20

0.6.4 Résultats et discussion ... 21

0.6.5 Explicabilité et interprétation des modèles ... 21

0.6.6 Conclusion ... 22

0.7 Chapitre 7 : Conclusions et orientations futures de la recherche 22

0.7.1 Conclusion ... 22

0.7.2 Limites de l'étude ... 23

0.7.3 Orientations futures de recherche ... 24

1 INTRODUCTION ... 27

1.1 Honey bee Acoustics ... 27

1.2 Current Challenges and Research Problems ... 29

1.3 Goals and Objectives ... 31

1.4 Thesis Organization ... 32

1.5 Publications ... 33

1.5.1 Publications included in the dissertation ... 33

1.5.2 Other publications ... 34

2 ACOUSTIC BASED BEEHIVE MONITORING : A REVIEW OF THE STATE-OF-THE-ART ... 35

2.1 Preamble ... 35

2.2 Introduction ... 35

2.3 Methods ... 35

2.4 Results and Discussion ... 37

2.4.1 Origin of the selected articles ... 37

2.4.2 Study Goal ... 37

2.4.3 Experiment setup ... 45

2.4.4 Audio Analysis Method ... 48

2.4.5 Reproducibility ... 53

2.5 Summary and Open Challenges ... 55

3 URBAN : URBAN BEEHIVE ACOUSTICS AND PHENOTYPING DATASET 57
3.1 PREAMBLE 57
3.2 INTRODUCTION 57
3.3 METHODS 59
3.3.1 Hive Management and Inspection 59
3.3.2 Varroa Mite Infestation 60
3.3.3 Mortality Rate 62
3.3.4 Sensor Data 63
3.4 DATA RECORDS 64
3.5 SUMMARY 64
4 IMPACT OF ENVIRONMENTAL NOISE AND PREPROCESSING 71
4.1 PREAMBLE 71
4.2 ENVIRONMENTAL NOISE FACTORS 71
4.2.1 Acoustic Feature Extraction 72
4.2.2 Beehive Strength Prediction Model 76
4.2.3 Experiment Setup and Figures-of-Merit 77
4.2.4 Results and Discussion 78
4.2.5 Prediction Model Performance 81
4.3 VOICE ACTIVITY DETECTORS FOR ACOUSTIC BEEHIVE MONITORING 84
4.3.1 Short-time Energy Thresholding 85
4.3.2 WebRTC 85
4.3.3 Convolutional Recurrent Deep Neural Network Based VAD 85
4.3.4 Experimental Results 86
4.4 AUDIO ENHANCEMENT 87
4.4.1 Spectral Amplitude Subtraction 88
4.5 DISCUSSION AND CONCLUDING REMARKS 90
5 MODULATION SPECTROGRAM-BASED FEATURES 93
5.1 PREAMBLE 93
5.2 COLONY POPULATION PREDICTION 93
5.2.1 Proposed Modulation Spectral Features 94
5.2.2 Experimental Setup 98
5.2.3 Results and Discussion 101
5.2.4 Brief Discussion 112

5.3 Non-Invasive Prediction of Varroa Mite Infestation in Honey bee Colonies 112
5.3.1 Proposed Methodology 114
5.3.2 Experimental Results 121
5.4 Conclusion 132
6 DEEP LEARNING WITH MODULATION SPECTROGRAM AND TENSORGRAM 133
6.1 Preamble 133
6.2 Introduction 133
6.3 Methods 134
6.3.1 Modulation Spectrum 134
6.3.2 Discriminative power of the modulation spectrum 134
6.3.3 Deep neural network model architectures 135
6.4 Experimental Setup 138
6.4.1 Dataset Description 138
6.4.2 Benchmark features and noise removal 138
6.4.3 Testing Setup and Figures-of-Merit 138
6.5 Results and Discussion 140
6.5.1 Colony Population Strength Prediction 140
6.5.2 Non-parametric Statistical Analyses 140
6.5.3 Explainability analysis 143
6.5.4 Deep learning vs. Conventional classifiers 145
6.6 Conclusion 147
7 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS 149
7.1 Conclusion 149
7.2 Limitations 149
7.3 Future Research Directions 150
BIBLIOGRAPHIE 153

LISTE DES FIGURES

FIGURE 2.1 COUNTRY OF FIRST AUTHOR AFFILIATION. 39

FIGURE 2.2 STUDY GOAL OF ARTICLES WITH RESPECT TO THE YEAR PUBLISHED. 39

FIGURE 3.1 PHOTOS OF THE (A) ROOFTOP URBAN APIARY IN MONTREAL AND (B) AUDIO RECORDING HARDWARE. 59

FIGURE 3.2 HISTOGRAMS OF THE NUMBER OF FRAMES OF BEES FOR THE YEAR OF (A) 2021 AND (B) 2022 EXPERIMENTS. 60

FIGURE 3.3 BARPLOTS OF THE NUMBER OF FRAMES OF BEES ON DAYS OF INSPECTIONS FOR THE YEAR OF (A) 2021 AND (B) 2022 EXPERIMENTS. 61

FIGURE 3.4 INSULATION USED FOR BEEHIVES OVERWINTERING. 61

FIGURE 3.5 VARROA MITE INFESTATION ON (A) AUGUST 24TH, (B) SEPTEMBER 1ST, AND (C) SEPTEMBER 30TH, 2022. 62

FIGURE 3.6 LOCATION OF THE (A) TEMPERATURE AND HUMIDITY SENSOR PLACED ON TOP OF THE MIDDLE FRAME OF THE FIRST BROOD CHAMBER AND (B) MICROPHONE PLACED NEXT TO IT. 63

FIGURE 3.7 INTERNAL AND EXTERNAL TEMPERATURE AND HUMIDITY DURING (A) A $24\,\mathrm{h}$ PERIOD FOR A SINGLE BEEHIVE ON AUGUST 12TH, 2021, AND B) TOTAL DURATION OF THE EXPERIMENT IN 2021. 65

FIGURE 3.8 AUDIO INTENSITY, AUDIO RMS, AND CORRESPONDING SPECTROGRAM FOR A HIVE WITH WITH 3 FULL BOXES OF BEES (AUGUST 12, 2021). 66

FIGURE 3.9 24H AVERAGE OF RMS VALUE WHEN THE NUMBER FRAMES OF BEES IS BETWEEN (A) 10 AND 20, (B) 20 AND 30, AND (C) 1 AND 30. 67

FIGURE 4.1 RMS CURVES OF A TYPICAL COLONY WITH THREE FULL BOXES OF BEES (BLUE CURVE) AND AN EMPTY BOX (ORANGE) AS MEASURED ON (A) AUGUST 12TH, 2021, AND (B) SEPTEMBER 24TH, 2021. 73

FIGURE 4.2 RMS CURVE AND CORRESPONDING SPECTROGRAM FOR A HIVE WITH 3 FULL BOXES OF BEES (AUGUST 12, 2021). 74

FIGURE 4.3 RMS CURVE AND CORRESPONDING MFCCS FOR A HIVE WITH WITH 3 FULL BOXES OF BEES (AUGUST 12, 2021). 75

FIGURE 4.4 SPECTROGRAMS OF CASES WHERE URBAN NOISES (E.G., TRAIN RAIL SQUEALING SOUNDS) ARE (A) NOT PRESENT AND (B) PRESENT. 78

FIGURE 4.5 COMPARISON OF THE CHANGES IN DISTRIBUTION OF THE (A) FIRST AND (B) SECOND MEL FREQUENCY CEPSTRAL COEFFICIENT WITH AND WITHOUT THE PRESENCE OF TRAIN RAIL SQUEALING SOUNDS. 79

FIGURE 4.6 FIRST MEL FREQUENCY CEPSTRAL COEFFICIENT TIME SERIES WHEN THERE ARE NO TRAINS NEARBY (TOP) VERSUS WHEN THE TRAIN PASSES BY (BOTTOM) BETWEEN THE 4 :00 - 6 :40 TIME RANGE. 79

FIGURE 4.7 COMPARISON OF THE CHANGES IN DISTRIBUTION OF THE DWT DETAIL AND APPROXIMATION COMPONENTS (A) WITHOUT AND (B) WITH THE PRESENCE OF TRAIN RAIL SQUEALING SOUNDS. 80

FIGURE 4.8 SPECTROGRAM COMPARISONS WHEN HUMAN SPEECH IS (A) NOT PRESENT AND (B) PRESENT NEAR THE HIVES. 80

FIGURE 4.9 COMPARISON OF THE CHANGES IN DISTRIBUTION OF THE (A) FIRST AND (B) THIRD MEL FREQUENCY CEPSTRAL COEFFICIENT WITH AND WITHOUT THE PRESENCE OF HUMAN SPEECH. 81

FIGURE 4.10 DWT DETAIL AND APPROXIMATION COMPONENT COMPARISONS WHEN HUMAN SPEECH IS (A) NOT PRESENT AND (B) PRESENT NEAR THE HIVES. 81

FIGURE 4.11 SPECTROGRAM COMPARISONS (A) WITHOUT AND (B) WITH HEAVY RAINFALLS. 82

FIGURE 4.12 COMPARISON OF THE CHANGES IN DISTRIBUTION OF THE (A) FIRST AND (B) SECOND MEL FREQUENCY CEPSTRAL COEFFICIENT WITH AND WITHOUT THE RAIN. 82

FIGURE 4.13 DWT COMPONENT COMPARISONS (A) WITHOUT AND (B) WITH HEAVY RAINFALLS. 82

FIGURE 4.14 A 90-SECOND EXCERPT FROM THE NU-HIVE DATASET, FROM TOP TO BOTTOM, RAW AUDIO SIGNAL IN BLUE AND GROUND TRUTH LABELS IN GREEN, AND SPECTROGRAM. 84

FIGURE 4.15 SPEECH DETECTION ON A 90 SECOND SAMPLE OF THE NU-HIVE DATASET : A) MANUAL LABELS, B) SHORT-TIME ENERGY VAD, C) WEBRTC, D) VAD-CRDNN, AND E) VAD-CRDNN-FT. 87

FIGURE 4.16 BLOCK DIAGRAM OF THE SPECTRAL AMPLITUDE SUBTRACTION. 90

FIGURE 4.17 FROM TOP TO BOTTOM, NOISY AUDIO AMPLITUDE OF A BEEHIVE, THE FILTERED AUDIO, AND THEIR CORRESPONDING SPECTROGRAMS. 90

FIGURE 5.1 BLOCK DIAGRAM OF SIGNAL PROCESSING STEPS INVOLVED IN CALCULATION OF THE MODULATION SPECTROGRAM. 94

FIGURE 5.2 BLOCK DIAGRAM OF THE THREE FEATURE EXTRACTION PROCESSES FROM MODULATION SPECTROGRAM : FLATTENING INTO 160-D MODULATION ENERGY VECTOR (TOP), INTO 28-D BANDED MODULATION ENERGY VALUES (MIDDLE), AND 252-D SPECTRAL DESCRIPTION VALUES (BOTTOM). 95

FIGURE 5.3 MSF FEATURE SET MAP THAT LABELS EACH BIN ONCE THE SPECTROGRAM FLATTENS. 96

FIGURE 5.4 MSAB FEATURE SET MAP LABELING EACH FREQUENCY BAND AFTER AVERAGING ACROSS (A) THE MODULATION FREQUENCY AXIS AND (B) THE ACOUSTIC FREQUENCY AXIS. 97

FIGURE 5.5 MSSD FEATURE SET MAP THAT SHOWS HOW THE EXTRACTED FEATURES ARE NAMED, EXEMPLIFIED HERE WITH JUST ONE OF THE NINE DESCRIPTORS, I.E., THE CENTROID. 97

FIGURE 5.6 PERFORMANCE COMPARISON BETWEEN SHAP AND MRMR FEATURE SELECTION METHODS APPLIED TO FUSED BENCHMARK FEATURES IN A RANDOM-SPLIT SETUP, EVALUATED USING (A) MAE AND RMSE ERROR METRICS, AND (B) CORRELATION. 99

FIGURE 5.7 SHAP VALUE PLOTS DEMONSTRATING THE CONTRIBUTION OF INDIVIDUAL FEATURES TO MODEL PREDICTIONS FOR (A) MFCCS, (B) LFCCS, (C) SPECTRAL DESCRIPTORS, AND (D) THEIR FUSION. FEATURES ARE RANKED BY MEAN ABSOLUTE SHAP VALUE, WITH COLOR DENOTING THE FEATURE VALUE'S MAGNITUDE. 102

FIGURE 5.8 PERFORMANCE COMPARISON OF THE BENCHMARK FEATURES AS A FUNCTION OF THE NUMBER OF TOP-RANKED FEATURES SELECTED IN THE HIVE-INDEPENDENT SETUP, EVALUATED USING (A) MAE, (B) RMSE, AND (C) CORRELATION. 103

FIGURE 5.9 SHAP VALUE PLOTS DEMONSTRATING THE CONTRIBUTION OF INDIVIDUAL FEATURES TO MODEL PREDICTIONS FOR (A) MSF, (B) MSAB, (C) MSSD, AND (D) THEIR FUSION. FEATURES ARE RANKED BY MEAN ABSOLUTE SHAP VALUE, WITH COLOR DENOTING THE FEATURE VALUE'S MAGNITUDE. 104

FIGURE 5.10 A PAIRPLOT ILLUSTRATING RELATIONSHIPS AMONG SELECTED SPECTRAL SHAPE DESCRIPTORS FEATURES FROM THE MSSD, INCLUDING CENTROID, SPREAD, FLATNESS, ENTROPY, AND ROLLOFF. 105

FIGURE 5.11 PERFORMANCE COMPARISON OF THE MODULATION SPECTROGRAM FEATURES AS A FUNCTION OF THE NUMBER OF TOP-RANKED FEATURES SELECTED IN THE HIVE-INDEPENDENT SETUP, EVALUATED USING (A) MAE, (B) RMSE, AND (C) CORRELATION. 105

FIGURE 5.12 SHAP VALUE PLOTS DEMONSTRATING THE CONTRIBUTION OF INDIVIDUAL FEATURES TO MODEL PREDICTIONS FOR FOR A COMBINED SET OF ALL FEATURES 106

FIGURE 5.13 THREE TOP-RANKED FEATURES FROM THE FUSION ALL SET : A) "FREQ_15", B) "FREQ_CENTROID_6", AND C) "MFCCS_2 AVERAGED ON HIVES WITH DIFFERENT NUMBER OF FRAMES ON SEPTEMBER 18TH, 2021. 107

FIGURE 5.14 CRITICAL DIFFERENCE (CD) PLOT ILLUSTRATING THE RESULTS OF THE NEMENYI POST-HOC TEST FOLLOWING THE FRIEDMAN TEST ON THE ROOT MEAN SQUARE ERROR (RMSE) OF TEN FEATURES FOR A) RANDOM-SPLIT AND B) HIVE-INDEPENDENT SCENARIO. THE FEATURES ARE RANKED BY THEIR AVERAGE PERFORMANCE ACROSS THE FIVE FOLDS. FEATURES GROUPED BY A THICK HORIZONTAL BAR EXHIBIT NO STATISTICALLY SIGNIFICANT DIFFERENCE IN RMSE PERFORMANCE ($\alpha = 0.05$). FEATURES WITH LOWER RANKS (TOWARDS 1) DEMONSTRATE SUPERIOR PERFORMANCE. 111

FIGURE 5.15 SPECTROGRAMS OF A A) 15-MINUTE CLEAN, AND B) NOISY BEEHIVE AUDIO SIGNAL. 113

FIGURE 5.16 MODULATION SPECTROGRAMS OF A A) 15-MINUTE CLEAN, AND B) NOISY BEEHIVE AUDIO SIGNAL. 113

FIGURE 5.17 VARROA MITE INFESTATION MEASURED ON (A) AUGUST 24TH, (B) SEPTEMBER 1ST, AND (C) SEPTEMBER 30TH, 2022. 115

FIGURE 5.18 24-H SPECTROGRAMS OF A A) VARROA INFECTED BEEHIVE, AND B) A HEALTHY BEEHIVE ON SEPTEMBER 30TH, 2022. 115

FIGURE 5.19 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, SPECTRAL SHAPE DESCRIPTORS (SSD) 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 MODULATION SPECTRAL SHAPE DESCRIPTORS (MSSD) ARE COMPUTED ACROSS THE MODULATION FREQUENCY AXIS TO CAPTURE SLOWER TEMPORAL MODULATIONS. 117

FIGURE 5.20 MODULATION SPECTROGRAMS OF A) A VARROA INFECTED BEEHIVE, AND B) A HEALTHY BEEHIVE. 118

FIGURE 5.21 PEARSON CORRELATION BETWEEN ALL PAIRS OF SPECTRAL DESCRIPTORS EXTRACTED FROM THE CONVENTIONAL SPECTROGRAM AND THE VARROA LABEL. 122

FIGURE 5.22 A PAIRPLOT ILLUSTRATING THE RELATIONSHIPS AMONG SPECTRAL SHAPE DESCRIPTORS : CENTROID, SPREAD, FLATNESS, ENTROPY, AND ROLLOFF. SAMPLES WITH VARROA LABEL EQUAL TO 0 REPRESENT HEALTHY HIVES, WHILE THOSE WITH VARROA LABEL EQUAL TO 1 INDICATE INFESTATION. 123

FIGURE 5.23 THE TEMPORAL DYNAMICS OF THE CONVENTIONAL SPECTROGRAM A) CENTROID, AND B) FLATNESS MEASURED ACROSS THE ENTIRE DAY ON SEPTEMBER 1, 2022. 124

FIGURE 5.24 A PAIRPLOT ILLUSTRATING THE RELATIONSHIP AMONG CONVENTIONAL AND MODULATION SPECTRAL SHAPE DESCRIPTORS : A) FLUX AND B) ENTROPY. SAMPLES WITH VARROA EQUAL TO 0 REPRESENT HEALTHY HIVES, WHILE THOSE WITH VARROA EQUAL TO 1 INDICATE INFESTATION. 125

FIGURE 5.25 THE TEMPORAL DYNAMICS OF THE MODULATION SPECTROGRAM A) CENTROID, AND B) FLATNESS MEASURED ACROSS THE ENTIRE DAY ON SEPTEMBER 1, 2022. 126

FIGURE 5.26 AGGREGATED CONFUSION MATRICES FOR EACH FEATURE SET EVALUATED ACROSS ALL THREE TEST FOLDS FOR THE SVM MODEL. EACH MATRIX SUMMARIZES THE TOTAL TRUE NEGATIVES (TN), FALSE POSITIVES (FP), FALSE NEGATIVES (FN), AND TRUE POSITIVES (TP) OBTAINED FOR : (A) MFCCS, (B) LFCCS, (C) SSD, (D) MSSD, AND (E) THE PROPOSED SSD+MSSD FEATURE SET. 129

FIGURE 5.27 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. 131

FIGURE 6.1 MODULATION SPECTROGRAMS AVERAGED ON SAMPLES WITH A) $1 \leq$ *fobs* ≤ 10, B) $11 \leq$ *fobs* ≤ 20, AND C) $21 \leq$ *fobs* ≤ 30. 134

FIGURE 6.2 THE NUMBER OF FRAMES OF BEES (*fobs*) IS GROUPED INTO THREE CATEGORIES : CATEGORY 1 CORRESPONDS TO $1 \leq$ *fobs* ≤ 10, CATEGORY 2 TO $11 \leq$ *fobs* ≤ 20, AND CATEGORY 3 TO $21 \leq$ *fobs* ≤ 30. F-RATIO MAPS COMPARING THE FOLLOWING PAIRS OF CATEGORIES : A) CATEGORY 1 VS. CATEGORY 2, B) CATEGORY 2 VS. CATEGORY 3, AND C) CATEGORY 1 VS. CATEGORY 3. 135

FIGURE 6.3 HEATMAPS SHOWING THE NUMBER OF FRAMES OF BEES ESTIMATED DURING VISUAL INSPECTION FOR EACH HIVE THROUGHOUT (A) 2021 AND (B) 2022. VALUES BETWEEN 1-10 SUGGEST ONLY THE BROOD CHAMBER WAS PRESENT, 11-20 BROOD CHAMBER PLUS ONE SUPER, AND 21-30 BROOD CHAMBER PLUS TWO SUPERS. 139

FIGURE 6.4 CRITICAL DIFFERENCE (CD) PLOT ILLUSTRATING THE RESULTS OF THE NEMENYI POST-HOC TEST FOLLOWING THE FRIEDMAN TEST ON THE MAE FOR A) RANDOM-SPLIT AND B) HIVE-INDEPENDENT SCENARIO. THE MODELS AND FEATURES COMBINATION ARE RANKED BY THEIR AVERAGE PERFORMANCE ACROSS THE FIVE FOLDS. FEATURES GROUPED BY A THICK HORIZONTAL BAR EXHIBIT NO STATISTICALLY SIGNIFICANT DIFFERENCE IN PERFORMANCE ($\alpha = 0.05$). FEATURES WITH LOWER RANKS (TOWARDS 1) DEMONSTRATE SUPERIOR PERFORMANCE. 143

FIGURE 6.5 SALIENCY MAPS OF THE TRAINED 2D CNN MODEL AVERAGED OVER SAMPLES FOR EACH COLONY STRENGTH LEVEL : (A) LOW (1–10 FOBS), (B) MEDIUM (11-20 FOBS), AND (C) HIGH (21–30 FOBS). 144

FIGURE 6.6 SALIENCY MAPS OF THE TRAINED 2D CNN MODEL WITH SPATIAL ATTENTION AVERAGED OVER SAMPLES FOR EACH COLONY STRENGTH LEVEL : (A) LOW (1–10 FOBS), (B) MEDIUM (11–20 FOBS), AND (C) HIGH (21–30 FOBS). 145

FIGURE 6.7 GRAD-CAM VISUALIZATIONS FOR REPRESENTATIVE 3D TEST SAMPLES. EACH ROW CORRESPONDS TO A DIFFERENT TIME FRAME (t_1, t_{mid}, t_{max}), AND COLUMNS CORRESPOND TO COLONY STRENGTH LEVELS (LOW, MEDIUM, HIGH). HEATMAPS SHOW HOW THE MODEL'S FOCUS SHIFTS ACROSS THE MODULATION-FREQUENCY SPACE OVER TIME. 146

LISTE DES TABLEAUX

TABLE 2.1 ELIGIBILITY CRITERIA FOR INCLUSION AND EXCLUSION OF ARTICLES. 36

TABLE 2.2 ITEMS EXTRACTED FROM EACH STUDY. 38

TABLE 2.3 LIST OF ARTICLES EXPLORING DIFFERENT STUDY GOALS. 40

TABLE 2.4 BEE SPECIES. 46

TABLE 2.5 MICROPHONE CHARACTERIZATION 47

TABLE 2.6 FEATURE EXTRACTION METHODS EXPLORED. 50

TABLE 2.7 CLASSIFICATION ALGORITHMS. 53

TABLE 2.8 CLASSIFICATION EVALUATION METRICS. 54

TABLE 3.1 COMPARISON OF URBAN DATASET WITH THE OTHER PUBLIC DATASETS. 58

TABLE 3.2 SUMMARY OF THE DATA COLLECTION START AND END DATES FOR EACH YEAR. 60

TABLE 3.3 THE QUANTITY OF RAW AUDIO FOR EACH YEAR OF THE EXPERIMENTS. 64

TABLE 3.4 STRUCTURE OF THE FILES DESCRIBING INSPECTIONS FOR THE YEAR 2021. 68

TABLE 3.5 STRUCTURE OF THE FILES DESCRIBING INSPECTIONS FOR EACH YEAR 2022. 69

TABLE 3.6 STRUCTURE OF THE FILES/FOLDERS OF THE SENSOR DATA (TEMPERATURE AND HUMIDITY), RAW AUDIO RECORDINGS, AND WEATHER INFORMATION. 70

TABLE 4.1 PROPOSED CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE. 77

TABLE 4.2 PERFORMANCE COMPARISON (MAE AND RMSE) FOR BEEHIVE MONITORING SYSTEMS BASED ON THREE FEATURE SETS UNDER RANDOM-SPLIT AND HIVE-INDEPENDENT TESTING SCENARIOS. 83

TABLE 4.3 PERFORMANCE EVALUATION ON NU-HIVE AND URBAN TEST DATASETS. THE VAD-CRDNN-FT-ALL CONDITION EXEMPLIFIES THE CASE WHERE FINE-TUNING WAS DONE USING BOTH DATASETS. 88

TABLE 5.1 HYPERPARAMETER VALUES EXPLORED FOR TUNING THE RANDOM FOREST MODEL. THIS INCLUDES A RANGE OF OPTIONS FOR THE NUMBER OF TREES, FEATURE SELECTION METHODS, TREE DEPTH, AND CRITERIA FOR SPLITTING AND LEAF NODES TO OPTIMIZE MODEL PERFORMANCE. 100

TABLE 5.2 PERFORMANCE COMPARISON OF ERROR METRICS BETWEEN DIFFERENT FEATURE SETS AND DATA PARTITIONING SETUPS, WITH AND WITHOUT SPECTRAL ENHANCEMENT. **BOLD** REPRESENTS THE BEST AND UNDERLINE THE SECOND BEST AVERAGE RESULTS. 108

TABLE 5.3 PERFORMANCE COMPARISON OF CORRELATION AND R^2 BETWEEN DIFFERENT FEATURE SETS AND DATA PARTITIONING SETUPS, WITH AND WITHOUT SPECTRAL ENHANCEMENT. **BOLD** REPRESENTS THE BEST AND UNDERLINE THE SECOND BEST AVERAGE RESULTS. 109

TABLE 5.4 FRIEDMAN TEST RESULTS ACROSS EVALUATION METRICS UNDER RANDOM-SPLIT AND HIVE-INDEPENDENT SCENARIOS. 110

TABLE 5.5 AVERAGE RANKS OF FEATURE SETS ACROSS EVALUATION METRICS UNDER RANDOM-SPLIT AND HIVE-INDEPENDENT SCENARIOS, WITH AND WITHOUT SPECTRAL AMPLITUDE SUBTRACTION PRE-PROCESSING. 111

TABLE 5.6 MODEL CONFIGURATIONS AND HYPERPARAMETER SETTINGS USED IN THE EXPERIMENTS 120

TABLE 5.7 LIGHTWEIGHT MINICNN MODEL ARCHITECTURE 120

TABLE 5.8 COMPARISON OF CLASSIFICATION PERFORMANCE ACROSS DIFFERENT FEATURE SETS AND CLASSIFIERS. ARROWS INDICATE THE DIRECTION OF CHANGE IN PERFORMANCE : ↑ DENOTES IMPROVEMENT, WHILE ↓ INDICATES A DECREASE. 128

TABLE 5.9 COMPARISON OF CLASSIFICATION PERFORMANCE ACROSS DIFFERENT FEATURE SETS USING THE CNN MODEL. 130

TABLE 6.1 DETAILS OF BASELINE CNN MODEL 136

TABLE 6.2 DETAILS OF CNN MODEL WITH SPATIAL ATTENTION 137

TABLE 6.3 DETAILS OF CRDNN MODEL 137

TABLE 6.4 PERFORMANCE COMPARISON OF MAE AND CORRELATION (R) BETWEEN DIFFERENT FEATURE SETS AND DATA PARTITIONING SETUPS, WITH AND WITHOUT SPECTRAL ENHANCEMENT. 141

TABLE 6.5 FRIEDMAN TEST RESULTS ACROSS EVALUATION METRICS MAE AND CORRELATION(R) UNDER RANDOM-SPLIT AND HIVE-INDEPENDENT SCENARIOS. 142

TABLE 6.6 COMPARISON OF AVERAGE RANKS ACROSS EVALUATION METRICS UNDER RANDOM-SPLIT AND HIVE-INDEPENDENT SCENARIOS, WITH AND WITHOUT AUDIO ENHANCEMENT. BOLD VALUES CORRESPOND TO BEST METHODS. 142

TABLE 6.7 PERFORMANCE COMPARISON OF BEST-PERFORMING RANDOM FOREST (RF) CLASSIFIER AND BEST-PERFORMING DEEP LEARNING (DL) METHOD PRESENTED HEREIN IN THE HIVE-INDEPENDENT SCENARIO. 147

SYNOPSIS

0.1 Chapitre 1 : Introduction

0.1.1 Acoustique des abeilles mellifères

Les abeilles mellifères (*Apis mellifera*) jouent un rôle essentiel dans la préservation de l'environnement. Au-delà de la production de miel, de cire d'abeille et de gelée royale, leur activité de pollinisation soutient la biodiversité et la production de nombreuses cultures agricoles majeures, faisant d'elles un véritable « prestataire de services » agroenvironnemental [1]. En effet, près de 75% des cultures produisant fruits et graines destinés à la consommation humaine aux États-Unis dépendent, au moins en partie, des abeilles mellifères pour leur rendement, leur qualité et leur durabilité [2]. Bien que plusieurs espèces d'abeilles contribuent à la pollinisation, les abeilles mellifères sont les principaux pollinisateurs des amandiers, des agrumes, des bleuets et des cucurbitacées [2]. Ainsi, tout changement dans le comportement ou l'état de santé des colonies, ainsi qu'une diminution de leur population, peut avoir des répercussions importantes sur l'industrie alimentaire. C'est pourquoi les apiculteurs procèdent à un suivi régulier de la santé des abeilles et de la vigueur des colonies.

Pendant de nombreuses années, les apiculteurs ont inspecté leurs ruches principalement par des examens visuels, généralement effectués durant la saison de pollinisation et de production de miel. Ces inspections, bien que nécessaires, sont chronophages et peuvent perturber les colonies ; elles sont donc généralement réalisées toutes les deux semaines. Cependant, des changements significatifs peuvent survenir entre deux inspections, pouvant mener à des problèmes de santé graves et à une diminution de la population. De plus, ces dernières années, des pertes importantes de colonies ont été rapportées à l'échelle mondiale [3, 4, 5, 6], et de nombreux chercheurs estiment que divers facteurs de stress en sont responsables [7, 8, 9]. Le développement de systèmes de surveillance continue permettrait une gestion plus efficace des colonies [7, 8, 9] et offrirait la possibilité d'un suivi à long terme des facteurs de stress. À ce titre, la surveillance automatisée des ruches suscite un intérêt croissant au sein des communautés scientifiques et apicoles [10].

Récemment, avec les avancées dans les domaines de l'Internet des objets (IoT), de l'infonuagique et de l'intelligence artificielle, des outils de surveillance automatisée des ruches ont vu le jour, donnant naissance à l'apiculture de précision [11]. Par exemple, plusieurs systèmes ont été proposés basés sur le suivi du poids [10, 12, 13], de la température [14, 15], de l'humidité [15] ou encore du niveau de CO_2[15] à l'intérieur des ruches. Des travaux préliminaires ont également exploré l'analyse acoustique[16, 17, 18] ou l'utilisation de capteurs multisources [10, 12, 15].

Le poids d'une ruche peut refléter la production de miel et la taille de la colonie, avec des variations diurnes significatives. Par exemple, une diminution du poids est souvent observée durant la nuit et tôt le matin, en raison de la consommation de nectar, tandis qu'une augmentation est observée en journée, corrélée à l'activité de butinage [19]. Meikle et al.[13] ont montré, grâce à des balances électroniques, que le poids moyen d'une ruche indique la quantité de nourriture stockée, alors que ses variations journalières reflètent la consommation. D'autres auteurs[10] ont suivi le poids de colonies pendant un an, observant une perte rapide de poids durant l'été (activité de butinage) et une perte plus lente durant l'hiver (alimentation à partir des réserves). Le phénomène d'essaimage peut également être détecté par une baisse soudaine du poids de la ruche.

La température interne d'une ruche est un bon indicateur de la capacité d'adaptation de la colonie et de son état de santé. Il a été démontré que la température influe directement sur la santé des abeilles et du couvain, ainsi que sur la productivité de la ruche [20]. Une thermorégulation efficace permet de réduire la mortalité et d'augmenter la production de miel [21, 22]. De plus, une élévation de température (de $34°C$–$35°C$ à $37°C$–$38°C$) pourrait être le signe annonciateur d'un essaimage [23].

L'humidité relative est également un facteur important pour l'incubation des œufs, la croissance des larves et le développement de la colonie. Elle influe aussi sur les comportements de transport d'eau et d'alimentation [24]. Human et al. [25] ont montré que l'humidité dans les alvéoles de couvain est généralement plus élevée et plus stable que dans les autres parties de la ruche.

La concentration de CO_2 peut quant à elle refléter l'activité métabolique des abeilles. Bien qu'elle soit influencée par la température [26], les abeilles peuvent maintenir des niveaux de CO_2 compris entre 0,1% et 4,25% [27].

Les abeilles communiquent également à l'aide de vibrations et de sons [28], produits par divers moyens : battements d'ailes, contractions musculaires rapides sans mouvement des ailes, ou pression du thorax contre le substrat ou une autre abeille [29]. Parmi les signaux sonores les plus connus figurent le piping, le quacking, le tooting, le hissing et le buzzing. Le « piping » est un signal émis par les jeunes reines pour annoncer leur intention de se battre avec la reine en place. Le « tooting », produit à l'émergence d'une vierge, a une fréquence fondamentale d'environ 400 à 500 Hz [30]. Le « quacking », réponse des autres reines encore enfermées dans leur cellule, se compose de courtes impulsions autour de 350Hz [30]. Les ouvrières peuvent également émettre des signaux similaires au piping [31], parfois même en l'absence de perturbation. Ce signal a une fréquence fondamentale comprise entre 330 et 430 Hz, avec des harmoniques marquées [32], mais peut varier de 100 à 500 Hz lors d'un essaimage [33]. Le « hissing », large bande sonore allant de 300 à 3600 Hz, est émis en cas de stress ou durant l'essaimage [34]. Enfin, le « buzzing », produit par les battements d'ailes à une fréquence de 250–300 Hz [35], est audible à l'oreille humaine.

L'utilisation de microphones placés à l'intérieur des ruches a été proposée comme méthode de surveillance acoustique en temps réel, permettant un suivi non intrusif [36, 37]. Des systèmes por-

tables de surveillance acoustique ont récemment vu le jour, comme l'Apivox Smart Monitor (2015) ou le Bee Health Guru (2019), qui utilisent le microphone d'un téléphone intelligent pour enregistrer des sons dans la ruche, lesquels sont ensuite analysés par intelligence artificielle pour détecter la présence de la reine, des parasites, ou prédire un essaimage [38, 39]. D'autres entreprises, telles que Arnia [40] et Nectar Technologies Inc. [41], ont également développé des solutions multisensorielles intégrant l'analyse acoustique.

Parmi toutes les modalités de surveillance disponibles, l'analyse acoustique s'est révélée particulièrement prometteuse pour détecter l'essaimage, l'absence de reine ou encore estimer la force de la colonie. Néanmoins, la surveillance automatisée des ruches par l'acoustique en est encore à ses débuts, et de nombreuses questions restent en suspens quant à son efficacité. L'objectif de cette thèse est de contribuer à lever certaines de ces limitations et de proposer des solutions innovantes pour un suivi audio à distance, fiable et robuste, des colonies d'abeilles mellifères.

0.1.2 Défis actuels et problèmes de recherche

Bien que des travaux importants aient été réalisés dans certains domaines de la surveillance acoustique des abeilles, plusieurs défis subsistent et nécessitent encore d'être relevés. Cette section présente les plus importants d'entre eux.

1. **Peu d'études sur la prédiction de la force, de la santé et du taux de mortalité des colonies.** Les études portant sur la surveillance de la santé et de la force des colonies d'abeilles à l'aide de l'acoustique demeurent très limitées. Or, la détection de la force d'une ruche est un paramètre clé pour les contrats de pollinisation, le suivi de la production de miel, l'évaluation de la santé de la reine et, de manière plus générale, l'état de santé de la colonie. La détection précoce d'un affaiblissement inattendu d'une ruche pourrait être extrêmement bénéfique pour les apiculteurs. Par ailleurs, le suivi automatisé de la mortalité des colonies, en lien avec divers facteurs environnementaux, pourrait permettre de mieux comprendre l'effet des stress externes sur la survie des abeilles.
2. **Connaissances limitées sur les effets des facteurs environnementaux.** Bien que les applications de surveillance acoustique des ruches soient en plein essor, les systèmes existants reposent majoritairement sur des descripteurs audio classiques issus du traitement de la parole, tels que les Coefficients Cepstraux en Fréquences Mel (CCFM), le spectrogramme ou encore la transformée en ondelettes discrète (DWT), comme l'illustre le tableau 2.6. Or, il est bien établi dans la communauté scientifique du traitement audio que ces paramètres sont fortement sensibles aux conditions environnementales, notamment au bruit ambiant ou à la présence de voix humaines en arrière-plan. Ainsi, la suppression du bruit ou la caractérisation du contexte acoustique (aussi appelée conscience du contexte) s'avèrent essentielles pour concevoir des systèmes robustes et reproductibles [42, 43]. Cette problématique est particulièrement pertinente pour les ruchers urbains, en croissance

constante [44], où des sons urbains forts et variés peuvent interférer avec les enregistrements internes de la ruche. À ce jour, aucune étude ne s'est attaquée directement à cette question, ce qui soulève des inquiétudes quant à la validité et à la généralisabilité des solutions actuelles dans des contextes réels.

3. **Jeux de données limités et peu diversifiés.** Bien que la majorité des études publiées aient démontré les avantages de l'acoustique pour la surveillance des ruches, la plupart d'entre elles reposent sur des enregistrements réalisés sur de très courtes périodes (souvent 24 heures), principalement durant les mois de juin et juillet [45, 46, 47]. Pour qu'un système de surveillance soit fiable, il est impératif d'inclure des données enregistrées sur plusieurs mois et saisons, car le comportement des abeilles varie fortement entre l'été et l'hiver. Un autre problème réside dans le nombre limité de ruches considérées dans chaque étude : les configurations expérimentales les plus courantes ne comportent qu'une à deux ruches, ce qui compromet la généralisabilité des résultats. Par conséquent, les conclusions rapportées dans la littérature peuvent ne pas refléter pleinement le potentiel réel de la surveillance acoustique.

4. **Manque de généralisabilité des modèles entre les ruches.** Enfin, plusieurs travaux ont mis en évidence une faible généralisabilité des modèles existants [48, 49], qui affichent des performances limitées lorsqu'ils sont appliqués à des ruches non vues lors de l'entraînement. En apprentissage automatique, on distingue généralement trois ensembles de données : l'entraînement, la validation et le test. L'ensemble d'entraînement sert à ajuster les paramètres du modèle, tandis que l'ensemble de test permet d'évaluer objectivement les performances finales. L'ensemble de validation, quant à lui, est utilisé pour ajuster les hyperparamètres sans biais. Deux configurations expérimentales principales sont généralement utilisées pour évaluer la généralisabilité d'un modèle : la division aléatoire (« random-split ») et la séparation par ruche (« hive-independent »). Dans la première, l'ensemble des données de toutes les ruches est divisé aléatoirement entre les ensembles d'entraînement, de validation et de test. Dans la seconde, certaines ruches sont exclues de l'entraînement et réservées uniquement au test. Dans tous les cas, les performances observées dans le scénario « hive-independent » sont considérablement dégradées par rapport au cas aléatoire, ce qui met en lumière un problème important de généralisabilité. Cela pourrait s'expliquer par l'utilisation de descripteurs acoustiques inadéquats ou par des biais introduits par la taille limitée des jeux de données, ou encore par la présence d'artéfacts. Le développement de nouvelles représentations audio et de solutions sensibles au contexte est donc essentiel pour améliorer la fiabilité des systèmes de surveillance acoustique automatisés.

Compte tenu de ces limitations, cette recherche doctorale vise à relever ces défis en proposant des solutions innovantes pour le développement de systèmes de surveillance acoustique plus robustes, précis et aptes à être déployés dans des contextes réels.

0.2 Chapitre 2 : Surveillance acoustique des ruches : revue de l'état de l'art

0.2.1 Portée et Méthodologie de la Revue

Ce chapitre présente une revue systématique et critique de la littérature sur la surveillance acoustique des abeilles et des pollinisateurs apparentés. L'objectif principal est de situer cette thèse dans le contexte plus large de la recherche et d'identifier les défis méthodologiques, expérimentaux et pratiques non résolus qui limitent le déploiement en conditions réelles. La revue se concentre sur les études publiées entre janvier 2012 et mai 2025, reflétant la croissance rapide de l'intérêt pour l'apiculture de précision et la surveillance bioacoustique, stimulée par les avancées en apprentissage automatique et en technologies de détection.

Une méthodologie de revue structurée a été adoptée afin d'assurer une couverture exhaustive et reproductible. Les articles de revues et conférences évalués par des pairs, ainsi que certains préprints électroniques, ont été collectés à partir des principales bases de données scientifiques, y compris IEEE Xplore, Scopus, PubMed, Google Scholar et arXiv. Les critères d'inclusion ciblaient les études utilisant des signaux audio pour inférer l'activité des abeilles, la santé de la colonie, le comportement, les stress environnementaux ou l'identité des espèces. Les travaux se concentrant exclusivement sur le développement de matériel ou de logiciels sans analyse acoustique ont été exclus. Chaque article sélectionné a été analysé selon cinq dimensions clés : origine de la publication, objectifs de l'étude, protocoles expérimentaux et d'enregistrement, méthodes d'analyse audio et d'apprentissage automatique, et pratiques de reproductibilité.

0.2.2 Applications de la Surveillance Acoustique des Abeilles

La littérature examinée révèle une communauté de recherche géographiquement diversifiée, la majorité des contributions provenant d'Amérique du Nord, d'Europe et d'Asie de l'Est. Les objectifs de recherche sont très hétérogènes, englobant la détection de l'activité des abeilles, la surveillance de la présence ou de l'absence de la reine, la prédiction de l'essaimage, l'identification des espèces, l'évaluation de la pollinisation et la détection des stress environnementaux. Parmi ces applications, la détection de l'absence de la reine et la prédiction de l'essaimage apparaissent comme les plus étudiées, reflétant leur importance critique pour la survie des colonies et les opérations apicoles.

Dans l'ensemble, la surveillance acoustique a démontré un fort potentiel pour détecter des phénomènes au niveau de la colonie tels que l'essaimage, le vol de ruches et la dynamique de population, ainsi que pour identifier des espèces individuelles et des comportements de pollinisation. Les études impliquant des bourdons et des espèces non-Apis mellifera illustrent davantage la richesse des signatures acoustiques liées à la morphologie, au comportement et à la fonction

écologique. Cependant, les solutions spécifiques à chaque application dominent la littérature, et les cadres généralisables restent limités.

0.2.3 Conception Expérimentale et Pratiques de Collecte de Données

L'échelle et la conception expérimentale varient considérablement dans la littérature. Si certaines études s'appuient sur des enregistrements d'abeilles individuelles ou de petits groupes contrôlés, la majorité utilise entre une et six ruches, avec seulement quelques déploiements à grande échelle impliquant des dizaines de ruches ou domiciles. Notamment, de nombreuses études ne précisent pas le nombre de ruches ou d'individus utilisés, et seule une petite partie évalue explicitement la généralisation indépendante de la ruche, essentielle pour l'applicabilité en conditions réelles.

La collecte de données est fortement dépendante de la saison, la plupart des enregistrements étant acquis pendant la période d'activité maximale des abeilles, au printemps tardif et en été. Plusieurs travaux montrent que les caractéristiques acoustiques varient selon les saisons et l'heure de la journée, motivant certaines études à adopter des stratégies de surveillance sur 24 heures. Néanmoins, les calendriers d'enregistrement, les durées et les conditions environnementales restent très hétérogènes, compliquant la comparaison directe entre les expériences.

0.2.4 Acquisition et Prétraitement du Signal

Les paramètres clés d'acquisition du signal, y compris la fréquence d'échantillonnage et les caractéristiques des microphones, sont souvent peu documentés. Bien que la majeure partie de l'énergie acoustique biologiquement pertinente se situe sous 1 kHz, les fréquences d'échantillonnage varient de quelques kilohertz à 44,1 kHz, et les réponses en fréquence des microphones sont souvent non spécifiées. Cette omission est particulièrement problématique pour l'analyse des basses fréquences, car des microphones avec une sensibilité insuffisante risquent de ne pas capturer des signaux biologiquement significatifs.

Les pratiques de prétraitement audio sont également incohérentes. Bien que certaines études appliquent un filtrage passe-bande ou passe-bas et un sous-échantillonnage pour réduire le bruit et le coût computationnel, la majorité ne décrit aucune étape de prétraitement ni ne fournit d'analyses comparatives entre signaux bruts et traités. Par conséquent, l'impact des choix de prétraitement sur les performances en aval reste mal compris.

0.2.5 Représentations Audio et Ingénierie des Caractéristiques

D'un point de vue analytique, les représentations temps–fréquence dominent le domaine. Les coefficients cepstraux en fréquences mél (CCFM) sont les plus utilisés, suivis des représentations spectrogrammes dérivées de la transformée de Fourier à court terme (STFT). D'autres approches, telles que les transformées en ondelettes, la transformée en Q-constante, la prédiction linéaire et la transformée de Hilbert–Huang, ont montré des résultats prometteurs pour des tâches spécifiques comme la détection de l'absence de la reine et la prédiction de l'essaimage.

La sélection des caractéristiques dépend largement de la tâche, et aucune représentation unique ne surpasse systématiquement les autres selon les applications. Cette diversité souligne la nécessité de représentations équilibrant robustesse, interprétabilité et sensibilité aux motifs acoustiques biologiquement pertinents.

0.2.6 Stratégies d'Apprentissage Automatique et de Modélisation

Une large gamme d'algorithmes d'apprentissage automatique a été explorée. Les modèles classiques—tels que les machines à vecteurs de support, la régression logistique, les forêts aléatoires et les k-plus proches voisins—restent largement utilisés en raison de leur interprétabilité et de leur faible coût computationnel. Parallèlement, les approches d'apprentissage profond telles que les réseaux convolutifs, récurrents, hybrides et basés sur les transformers ont gagné en popularité en exploitant des représentations temps–fréquence de type image ou des embeddings audio appris.

Les études récentes explorent de plus en plus l'apprentissage par transfert, l'apprentissage auto-supervisé et l'apprentissage fédéré, reflétant l'intérêt croissant pour l'efficacité des données, la confidentialité et la scalabilité. Malgré ces avancées, la robustesse des modèles face au bruit, au changement de domaine et à la variabilité saisonnière reste un défi persistant.

0.2.7 Protocoles d'Évaluation et Reproductibilité

Les pratiques d'évaluation des performances varient considérablement selon les études. La précision est le métrique le plus souvent rapporté, souvent accompagnée de l'AUC, du F1-score, de la précision et du rappel. Cependant, la prévalence du déséquilibre des classes dans les tâches acoustiques liées aux abeilles suggère que la seule précision peut être trompeuse. Les protocoles d'évaluation, les stratégies de validation croisée et les conditions de test sont rapportés de manière incohérente, limitant la comparaison significative entre les approches.

La reproductibilité reste un défi majeur dans le domaine. Seule une fraction des études repose sur des ensembles de données publics, et encore moins partagent le code source. Bien que

certains ensembles de données notables tels que Nu-Hive, OSBH et BUZZ aient permis un benchmarking limité, l'absence de jeux de données standardisés, de pipelines ouverts et de protocoles d'évaluation cohérents continue de freiner les progrès.

0.2.8 Synthèse et Lacunes Identifiées

À travers presque toutes les dimensions examinées—collecte de données, acquisition de signal, représentation des caractéristiques, modélisation et évaluation—une hétérogénéité substantielle et un sous-reporting limitent la robustesse, la généralisation et la reproductibilité. Bien que la surveillance acoustique des abeilles ait donné des résultats prometteurs pour diverses applications, le domaine reste contraint par des pratiques expérimentales fragmentées et un manque de cadres de validation standardisés.

Au-delà du résumé des travaux existants, ce chapitre fournit une synthèse structurée des pratiques expérimentales, des représentations des caractéristiques et des stratégies de modélisation, mettant en évidence des limitations systémiques qui guident directement les choix méthodologiques adoptés dans le reste de cette thèse. Ces lacunes identifiées motivent les contributions méthodologiques développées dans les chapitres suivants, visant à améliorer la robustesse au bruit, renforcer la généralisation entre ruches et saisons, et promouvoir des systèmes de surveillance acoustique reproductibles et évolutifs.

0.3 Chapitre 3 : Jeu de données UrBAN

Ce chapitre présente UrBAN (Urban Beehive Acoustics and PheNotyping), un jeu de données à grande échelle et en accès libre conçu pour soutenir la surveillance acoustique des colonies d'abeilles. Ce jeu de données répond à la rareté des jeux de données audio publics sur les ruches en fournissant plus de 3 000 heures d'enregistrements audio bruts, collectés sur deux années consécutives (2021–2022) à partir de dix ruches installées sur le toit d'un rucher à Montréal, Canada.

UrBAN inclut des données de capteurs multimodaux, telles que la température et l'humidité internes à la ruche enregistrées toutes les 15 minutes, ainsi que des variables météorologiques externes, incluant la température, l'humidité, la vitesse du vent et les précipitations. Des inspections manuelles fournissent des annotations phénotypiques de référence, notamment la population de la colonie (cadres d'abeilles), l'état de la reine, le niveau d'infestation par les acariens Varroa et la mortalité hivernale. Environ 20% des ruches sont mortes après l'hivernage en 2022, soulignant la pertinence du jeu de données pour les études sur la survie des colonies.

Les populations de ruches ont été volontairement variées afin de capturer une diversité de forces coloniales, allant de 6 à 30 cadres d'abeilles, et les enregistrements audio reflètent les

changements correspondants de l'activité acoustique. L'analyse révèle des variations diurnes et saisonnières de l'intensité sonore, qui corrèlent avec la température et l'humidité internes et externes, reflétant l'activité de la colonie et la thermorégulation. La valeur RMS de l'audio et les motifs du spectrogramme montrent une activité accrue lors des périodes de hausse de température et d'humidité faible, cohérente avec le butinage et la dynamique de la colonie.

Comparé aux jeux de données existants, UrBAN offre une couverture temporelle plus étendue, un accès aux enregistrements audio bruts, des mesures de capteurs multimodaux et des enregistrements phénotypiques détaillés, permettant des études longitudinales et le développement de modèles d'apprentissage automatique généralisables. Le jeu de données est accessible publiquement via le Federated Research Data Repository, constituant une ressource pour la recherche reproductible sur la santé des ruches, l'estimation des populations et la résilience des colonies en environnements urbains réels.

0.4 Chapitre 4 : Impact du bruit environnemental et du prétraitement

0.4.1 Facteurs de bruit environnemental

Cette section étudie l'impact du bruit environnemental sur la surveillance acoustique des colonies d'abeilles en utilisant le **jeu de données UrBAN**, qui fournit des enregistrements à long terme provenant de ruches urbaines. Les perturbations environnementales—**bruits urbains, voix humaines et fortes pluies**—peuvent significativement déformer les caractéristiques audio couramment utilisées, compromettant la fiabilité et la robustesse des systèmes automatisés de surveillance. Cette section examine ces effets sur les caractéristiques RMS, spectrogrammes, CCFM et DWT, en soulignant l'importance d'une extraction de caractéristiques contextuellement consciente et résiliente au bruit.

0.4.1.1 Méthodologie

Les enregistrements audio ont été rééchantillonnés à 2048 Hz et découpés en trames de 1 seconde. Les caractéristiques acoustiques suivantes ont été extraites :

- **RMS :** Mesure la puissance du signal et identifie les pics environnementaux causés par le bruit urbain, la parole et la pluie.
- **Spectrogramme :** Représentation temps-fréquence obtenue via STFT (fenêtre de Hann, 125 ms, recouvrement de 50 %) ; utilisée précédemment pour la détection de l'essaimage et de la présence de la reine.
- **CCFM :** Les 12 premiers coefficients plus le coefficient zéro, extraits via préaccentuation, STFT, banque de filtres mel, logarithme et DCT ; sensibles au bruit transitoire et continu.

— **DWT :** Décomposition du signal en composantes d'approximation et de détail (10 niveaux, Daubechies db2), capturant les variations temps-fréquence causées par le bruit environnemental.

Trois réseaux neuronaux convolutionnels (CNN) ont été entraînés pour prédire la force de la ruche (0 à 30 cadres d'abeilles) en utilisant chaque ensemble de caractéristiques. Deux protocoles d'évaluation ont été mis en place :

1. **Random-split :** 25 % des trames pour le test, 20 % pour la validation, le reste pour l'entraînement.
2. **Hive-independent :** Approche leave-some-hives-out, avec sept ruches pour l'entraînement (30 % utilisées pour la validation) et trois ruches inédites pour le test.

Les performances ont été évaluées via l'erreur absolue moyenne (MAE) et l'erreur quadratique moyenne (RMSE).

0.4.1.2 Résultats et discussion

Bruits urbains Les rails de train provoquent des perturbations basses et moyennes fréquences ($\sim$64–250 Hz), affectant RMS, spectrogrammes, CCFM et DWT. Le premier coefficient CCFM augmentait d'environ 50 dB lors du passage des trains, pouvant entraîner des prédictions incorrectes. Les coefficients de détail DWT étaient particulièrement sensibles à ces bruits urbains transitoires.

Voix humaines La parole de l'apiculteur se superpose aux fréquences acoustiques de la ruche (80–250 Hz), modifiant l'énergie du spectrogramme et les distributions CCFM/DWT. Les systèmes basés sur les moyennes CCFM ou DWT étaient vulnérables aux artefacts de parole transitoire, pouvant causer des erreurs de classification.

Pluie intense Les précipitations génèrent un bruit de longue durée, augmentant la puissance audio sur plusieurs bandes de fréquences d'environ 10 dB. Les coefficients CCFM et DWT étaient fortement perturbés pendant plusieurs heures, contrairement aux effets transitoires du bruit urbain ou de la parole, représentant un risque plus important pour la surveillance continue.

Performance des prédictions

— **Random-split (données propres) :** Les modèles basés sur CCFM ont montré la meilleure performance (MAE 0,86), suivis par les modèles spectrogramme et DWT.
— **Random-split (données bruitées) :** CCFM les plus robustes face aux bruits urbains et à la pluie ; spectrogrammes plus performants pour la parole ; DWT les plus affectés.

— **Hive-independent :** Toutes les erreurs augmentent ; spectrogrammes légèrement meilleurs pour les bruits urbains et la parole, CCFM plus performants sous fortes pluies.

Ces résultats montrent que le bruit environnemental dégrade significativement les performances de surveillance, les artefacts de parole provoquant les erreurs les plus importantes lors des tests hive-independent.

0.4.1.3 Conclusion

Le bruit environnemental—bruits urbains, voix humaines et fortes pluies—affecte fortement les caractéristiques acoustiques couramment utilisées (RMS, spectrogramme, CCFM, DWT), modifiant le contenu fréquentiel, les distributions de coefficients et les composantes DWT. Bien que les modèles d'apprentissage profond montrent un potentiel pour la surveillance acoustique des ruches, leurs performances sont très sensibles aux conditions environnementales, soulignant l'importance d'extraire des caractéristiques contextuellement conscientes et résilientes au bruit dans les ruchers urbains.

0.4.2 Détecteurs d'activité vocale pour la surveillance acoustique des ruches

Cette section évalue l'efficacité de trois méthodes de détection d'activité vocale (VAD)—seuil d'énergie à court terme, WebRTC VAD et réseau neuronal profond convolutionnel récurrent (VAD-CRDNN)—pour détecter la parole humaine intégrée dans les enregistrements de ruches. Les expériences ont été menées sur deux jeux de données : UrBAN (interne) et Nu-hive, contenant des enregistrements de ruches avec ou sans reine.

0.4.2.1 Méthodologie

Trois méthodes VAD ont été évaluées, différant par leur complexité, approche et robustesse au bruit de fond dans les enregistrements de ruches.

Seuil d'énergie à court terme Cette méthode classique estime l'énergie des trames courtes et la compare à un seuil prédéfini pour détecter la parole. Ici, l'énergie a été calculée sur l'ensemble du spectre et dans une sous-bande dominante de la parole (300–3000 Hz). Le ratio énergie sous-bande / énergie totale était comparé à un seuil de 0,6 pour classifier les trames. Simple mais très sensible au bruit de fond, comme le bourdonnement des abeilles ou les bruits environnementaux.

WebRTC VAD WebRTC VAD utilise les log-énergies sur six bandes de fréquences (80–4000 Hz) et applique un modèle de mélange gaussien pré-entraîné (GMM) pour classifier les trames

parole/non-parole. Conçu pour les environnements acoustiques bruyants et dynamiques, adapté aux enregistrements de ruches. La version Python a été utilisée avec le mode de filtrage le plus agressif pour réduire les faux positifs. Permet une détection robuste sans entraînement spécifique.

Réseau neuronal profond convolutionnel récurrent (VAD-CRDNN) Architecture CNN-RNN-DNN : deux couches convolutionnelles (VGG, 16 et 32 filtres, noyaux 3x3), deux couches GRU, deux couches pleinement connectées (16 neurones). Entrée : énergies Mel filterbank 40 dimensions. Sortie : probabilités de parole par trame, seuilées pour classification. Un modèle pré-entraîné a été fine-tuné sur les deux jeux de données pendant 30 époques (VAD-CRDNN-FT).

0.4.2.2 Résultats et discussion

WebRTC a surpassé les autres méthodes sur les deux jeux de données, avec la meilleure précision et F1-score. Le seuil d'énergie court terme a eu des difficultés, et VAD-CRDNN a bénéficié du fine-tuning mais restait moins performant que WebRTC. L'inspection visuelle a confirmé que WebRTC capture de manière fiable les segments de parole malgré le bourdonnement et les bruits environnementaux.

0.4.2.3 Conclusion

WebRTC VAD est la méthode la plus adaptée pour détecter la parole humaine dans la surveillance acoustique des ruches, surpassant les méthodes simples et les modèles DNN récents. Le fine-tuning améliore les performances des réseaux profonds mais ne dépasse pas WebRTC, soulignant sa robustesse pour les applications de surveillance automatisée des ruches.

0.4.3 Amélioration audio

Le bruit environnemental peut masquer les sons produits par les abeilles, réduisant la précision des systèmes de surveillance acoustique. Bien que des études antérieures aient abordé la suppression de certains bruits spécifiques (parole humaine), une étape générale et robuste de réduction du bruit reste nécessaire pour les enregistrements réels de ruches.

0.4.3.1 Méthode de soustraction spectrale de l'amplitude

— **Principe :** Les composantes spectrales du bruit sont estimées et soustraites du signal audio original pour isoler les sons des abeilles, améliorant le rapport signal sur bruit tout en conservant les caractéristiques acoustiques biologiquement pertinentes.

— **Estimation du bruit :** Un filtre à moyenne mobile exponentielle (EMA) est utilisé pour estimer le spectre de bruit. Le poids α est adapté trame par trame selon le rapport signal sur bruit a-posteriori (SNR).
— **Ajustement adaptatif :** Le paramètre α est ajusté via une fonction sigmoïde avec un seuil T et une pente a, permettant d'atténuer sélectivement le bruit dans les trames à faible SNR tout en préservant les trames à SNR élevé.
— **Interprétation :**
  — $\text{SNR}_{\text{a-posteriori}} \approx 1$: trame dominée par le bruit
  — $\text{SNR}_{\text{a-posteriori}} > 1$: trame contenant signal et bruit
— **Implémentation et visualisation :** L'algorithme est disponible sur GitHub. Les spectrogrammes avant/après amélioration montrent une réduction nette du bruit dans les bandes de fréquence caractéristiques tout en conservant les structures harmoniques essentielles.

L'analyse qualitative montre une réduction claire du bruit de fond sans supprimer les structures harmoniques caractéristiques. Bien que cette méthode ne supprime pas tous les bruits non stationnaires, elle constitue une étape légère et efficace qui améliore la clarté du signal, facilitant l'extraction de caractéristiques et la modélisation en conditions réelles.

0.4.4 Conclusion

Ce chapitre a examiné la conscience contextuelle dans la surveillance acoustique des ruches, en se concentrant sur l'impact du bruit environnemental réel sur les caractéristiques audio et les modèles d'apprentissage automatique. Les enregistrements UrBAN ont montré que les bruits urbains, la parole humaine et les fortes pluies dégradent significativement les caractéristiques acoustiques (spectrogrammes, CCFM), réduisant la précision des prédictions, surtout pour les évaluations hive-independent.

Le chapitre a également évalué les méthodes VAD (seuil d'énergie, WebRTC, CRDNN), montrant que WebRTC est la plus robuste pour distinguer la parole des sons d'abeilles, soulignant l'importance d'une sélection soigneuse des outils.

Enfin, la soustraction spectrale de l'amplitude a été présentée comme une méthode légère d'amélioration audio qui atténue adaptativement le bruit tout en préservant les sons liés à l'activité des abeilles. Bien qu'elle ne supprime pas tous les bruits, elle améliore la clarté du signal et facilite l'extraction de caractéristiques.

0.5 Chapitre 5 : Caractéristiques basées sur le spectrogramme de modulation

0.5.1 Prédiction de la population des colonies

Des analyses précédentes ont montré que les caractéristiques acoustiques couramment utilisées pour le suivi des ruches, telles que les descripteurs basés sur le spectrogramme, CCFM et les représentations par ondelettes, sont fortement sensibles aux sources de bruit environnemental, notamment les précipitations, le passage de trains à proximité et la parole humaine. Ces perturbations introduisent des distorsions spectrales qui dégradent significativement les performances des modèles dans des conditions d'enregistrement réelles, limitant ainsi l'applicabilité des systèmes de surveillance existants en dehors d'environnements contrôlés.

De plus, les signatures acoustiques varient considérablement d'une ruche à l'autre en raison des différences de population, de structure des ruches et de conditions d'installation, ce qui entraîne une faible capacité de généralisation vers des ruches non observées. La sensibilité au bruit environnemental et la variabilité inter-ruches constituent ainsi des défis majeurs pour le développement de systèmes fiables de surveillance acoustique automatisée.

0.5.1.1 Approche proposée : caractéristiques spectrales de modulation

Afin de relever ces défis, ce chapitre étudie l'utilisation de caractéristiques spectrales de modulation pour la prédiction automatisée de la force des colonies d'abeilles. Contrairement aux représentations temps–fréquence conventionnelles, les caractéristiques de modulation décrivent la dynamique temporelle de l'énergie spectrale en quantifiant les fréquences de modulation associées aux variations d'activité acoustique au cours du temps. Cette représentation fréquence–fréquence permet une meilleure discrimination entre les sons intrinsèques produits par les abeilles et les bruits environnementaux externes.

Bien que les représentations basées sur la modulation aient démontré leur robustesse dans d'autres domaines du traitement du signal, leur application au suivi acoustique des ruches n'avait jusqu'à présent pas été explorée. L'approche proposée repose principalement sur l'ingénierie de caractéristiques plutôt que sur l'apprentissage profond de bout en bout, en raison de la disponibilité limitée de grands ensembles de données acoustiques apicoles annotées et librement accessibles.

0.5.1.2 Extraction des caractéristiques à partir du spectrogramme de modulation

Le spectrogramme de modulation est calculé à l'aide d'un processus de transformée de Fourier à court terme (STFT) en deux étapes. Dans un premier temps, un spectrogramme conventionnel est généré à l'aide d'une fenêtre de 100 ms et d'un pas de 12,5 ms. Une seconde STFT est

ensuite appliquée à l'évolution temporelle de chaque bande spectrale à l'aide d'une fenêtre de 60 s, produisant un tensorgramme de modulation tridimensionnel qui capture la fréquence acoustique, la fréquence de modulation et le temps.

L'agrégation temporelle du tensorgramme de modulation permet d'obtenir un spectrogramme de modulation bidimensionnel, à partir duquel trois ensembles de caractéristiques sont extraits :

— *Modulation Spectrogram Flattened (MSF)* : un vecteur de caractéristiques de dimension 160 obtenu par l'aplatissement du spectrogramme de modulation réduit ;
— *Modulation Spectrogram Averaged Bands (MSAB)* : une représentation compacte de dimension 28 calculée en moyennant l'énergie de modulation sur des bandes prédéfinies de fréquences acoustiques et de modulation ;
— *Modulation Spectrogram Shape Descriptors (MSSD)* : un ensemble de 252 caractéristiques composé de descripteurs de forme spectrale extraits le long des axes de fréquence acoustique et de fréquence de modulation.

Ces ensembles de caractéristiques capturent des informations complémentaires relatives à la distribution de l'énergie de modulation et à la dynamique acoustique structurée associée à l'activité de la colonie.

0.5.1.3 Cadre expérimental

Les caractéristiques de modulation proposées sont évaluées en comparaison avec des caractéristiques de référence largement utilisées, notamment les CCFM, les Coefficients Cepstraux en Fréquences Linéaires (CCFL) et des descripteurs spectraux conventionnels. Afin de réduire l'influence du bruit environnemental, une soustraction de l'amplitude spectrale est appliquée comme étape de prétraitement avant l'extraction des caractéristiques.

Les expériences sont menées selon deux protocoles d'évaluation : une séparation aléatoire des données et une évaluation indépendante par ruche, cette dernière étant essentielle pour évaluer la capacité de généralisation vers des ruches non observées. La sélection des caractéristiques est réalisée à l'aide des méthodes minimum de redondance et maximum de pertinence (mRMR) ainsi que d'un classement basé sur les valeurs SHAP. Un régresseur de forêt aléatoire est utilisé pour prédire la taille de la population des colonies, exprimée en cadres d'abeilles. Les performances des modèles sont évaluées à l'aide de plusieurs métriques, dont l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne en pourcentage (MAPE), le coefficient de corrélation de Pearson et le coefficient de détermination (R^2).

0.5.1.4 Résultats et discussion

L'analyse de l'importance des caractéristiques montre que les caractéristiques spectrales de modulation capturent des motifs informatifs en fréquences acoustiques et de modulation associés à la force des colonies. Parmi les représentations proposées, l'ensemble MSAB présente systématiquement de solides performances prédictives tout en conservant une faible dimensionnalité.

Dans le cadre de l'évaluation par séparation aléatoire, les caractéristiques basées sur les CCFM obtiennent les meilleures performances, ce qui reflète leur efficacité lorsque les conditions d'apprentissage et de test sont similaires. En revanche, dans le cadre de l'évaluation indépendante par ruche, les caractéristiques de référence subissent une dégradation notable des performances, mettant en évidence leur sensibilité à la variabilité inter-ruches.

À l'inverse, les caractéristiques spectrales de modulation, en particulier MSAB, démontrent une robustesse et une capacité de généralisation accrues sur des ruches non observées. La fusion de caractéristiques améliore encore la précision des prédictions, les ensembles combinant des caractéristiques de référence et de modulation offrant les performances les plus stables dans l'ensemble des scénarios d'évaluation.

0.5.1.5 Conclusion

Les résultats confirment que le bruit environnemental et la variabilité spécifique aux ruches limitent fortement la généralisabilité des caractéristiques acoustiques conventionnelles. En mettant l'accent sur la dynamique temporelle et spectrale plutôt que sur le contenu spectral absolu, les caractéristiques spectrales de modulation offrent une robustesse accrue dans des conditions de déploiement réalistes. Leur performance améliorée dans les évaluations indépendantes par ruche soutient leur pertinence pour le développement de systèmes de surveillance acoustique apicole évolutifs et applicables en conditions réelles.

0.5.2 Prédiction non invasive de l'infestation par le varroa dans les colonies d'abeilles

Varroa destructor constitue l'une des menaces les plus critiques pour la survie des colonies d'abeilles à l'échelle mondiale. En se nourrissant des abeilles et en agissant comme vecteur de virus tels que le virus des ailes déformées, l'infestation par le varroa peut rapidement affaiblir les colonies et entraîner leur effondrement en l'absence de traitement opportun. Les pratiques actuelles de surveillance reposent principalement sur des méthodes d'inspection manuelle, qui sont invasives, coûteuses en main-d'œuvre et offrent une information temporelle limitée. Ces contraintes motivent le développement d'approches automatisées et non invasives capables d'évaluer en continu le risque d'infestation et de soutenir une intervention précoce.

Les avancées récentes en apiculture de précision et en apprentissage automatique ont mis en évidence le suivi acoustique comme une modalité prometteuse pour l'évaluation de la santé des colonies. Bien que l'analyse acoustique ait été appliquée avec succès à des tâches telles que la détection de l'essaimage, l'identification de l'absence de reine et l'estimation de la force des colonies, son application à la détection de l'infestation par le varroa reste limitée. Les études existantes reposent principalement sur des caractéristiques cepstrales telles que les CCFM, initialement conçues pour le traitement de la parole, et qui peuvent ne pas capturer de manière optimale la structure spectrale et la dynamique temporelle de l'activité collective des abeilles.

0.5.2.1 Limites des caractéristiques acoustiques existantes

La majorité des travaux antérieurs en surveillance acoustique des ruches utilisent les CCFM comme représentation dominante. Bien que ces caractéristiques aient montré des performances compétitives dans plusieurs tâches de classification, elles mettent l'accent sur des enveloppes spectrales à court terme et une échelle fréquentielle perceptuelle, ce qui peut conduire à négliger des modulations temporelles plus lentes et des motifs rythmiques caractéristiques du comportement à l'échelle de la colonie. Ces limitations sont particulièrement pertinentes pour les changements acoustiques liés au varroa, qui peuvent se manifester par une modification de l'organisation spectrale et une réduction de la variabilité temporelle plutôt que par des événements transitoires isolés.

Afin de fournir une base de comparaison équitable tout en examinant ces limitations, cette étude inclut à la fois les CCFM et les CCFL, permettant une évaluation systématique des représentations perceptuelles et linéaires dans le contexte de la détection du varroa.

0.5.2.2 Représentation des caractéristiques proposée

Ce travail introduit un cadre de fusion de caractéristiques basé sur des descripteurs de forme spectrale extraits à partir de spectrogrammes conventionnels et de spectrogrammes de modulation. Neuf descripteurs — centroïde, étendue, asymétrie, kurtosis, entropie, roll-off, platitude, crête et flux — sont calculés afin de caractériser différents aspects du signal acoustique.

Les descripteurs dérivés du spectrogramme conventionnel capturent la structure spectrale à court terme, notamment l'organisation harmonique, la distribution de l'énergie et le caractère bruité du signal. En parallèle, les spectrogrammes de modulation sont calculés à l'aide d'une seconde STFT appliquée dans le temps, permettant l'analyse de modulations d'amplitude plus lentes et de fluctuations rythmiques. Les descripteurs de forme spectrale extraits le long de l'axe des fréquences de modulation capturent des dynamiques temporelles absentes des caractéristiques basées sur les spectrogrammes standards.

La combinaison de ces deux représentations vise à fournir des informations complémentaires, caractérisant conjointement le contenu fréquentiel instantané et le comportement de modulation à plus long terme associé à l'infestation par le varroa.

0.5.2.3 Cadre expérimental

L'approche proposée est évaluée à l'aide du jeu de données UrBAN, qui comprend des enregistrements acoustiques continus de ruches ainsi que des mesures d'infestation par le varroa obtenues par lavage à l'alcool. La tâche de détection est formulée comme un problème de classification binaire supervisée, où les colonies sont étiquetées selon que le taux d'infestation dépasse ou non un seuil de 3%, conformément aux recommandations apicoles en matière de traitement.

Plusieurs classificateurs d'apprentissage automatique sont évalués, notamment Naive Bayes, forêts aléatoires, gradient boosting, k plus proches voisins, machines à vecteurs de support et perceptron multicouche, ainsi qu'un réseau de neurones convolutionnel léger à titre de comparaison. Afin de garantir une évaluation réaliste des performances, une stratégie de validation croisée indépendante par ruche est employée, consistant à exclure des ruches entières lors des phases de test afin d'éviter toute fuite d'information et de mieux refléter les conditions de déploiement.

Les performances des modèles sont évaluées à l'aide des métriques ROC-AUC, score F1, précision et rappel, offrant une évaluation équilibrée dans un contexte de déséquilibre des classes.

0.5.2.4 Résultats et discussion

L'analyse des descripteurs de forme spectrale révèle des différences acoustiques nettes entre les colonies saines et celles infestées par le varroa. Une infestation accrue est associée à une augmentation de l'entropie spectrale, de la platitude, du centroïde et de l'étendue, indiquant un spectre plus plat, plus bruité et moins structuré harmoniquement. Au niveau de la modulation, les ruches infestées présentent une réduction du flux spectral et un contenu plus faible en fréquences de modulation, reflétant une variabilité temporelle réduite et une dynamique acoustique plus lente.

Dans les expériences de classification, les CCFM et les CCFL présentent un pouvoir discriminant limité, tandis que les descripteurs de forme spectrale proposés surpassent significativement les caractéristiques cepstrales de référence pour l'ensemble des classificateurs. Les caractéristiques dérivées du spectrogramme de modulation améliorent encore les performances, soulignant l'importance des indices de modulation temporelle. La fusion des descripteurs conventionnels et de modulation fournit les meilleurs résultats globaux, confirmant leur complémentarité.

0.5.2.5 Conclusion

Cette section montre que l'infestation par le varroa affecte à la fois la structure spectrale et les motifs de modulation temporelle des sons produits par les colonies d'abeilles. En dépassant les représentations cepstrales et en intégrant des descripteurs spectraux sensibles à la modulation, le cadre proposé offre une robustesse et une capacité de généralisation accrues dans un contexte d'évaluation indépendante par ruche. Ces résultats soutiennent l'utilisation de caractéristiques acoustiques basées sur la modulation comme solution pratique et évolutive pour le suivi non invasif du varroa en apiculture de précision.

0.6 Chapitre 6 : Apprentissage profond à partir du spectrogramme de modulation et du tensorgramme

En s'appuyant sur le chapitre précédent, qui a démontré l'efficacité des caractéristiques de modulation conçues manuellement combinées à des modèles d'apprentissage automatique classiques pour l'estimation de la force des colonies d'abeilles, ce chapitre examine si des architectures d'apprentissage profond peuvent exploiter plus efficacement les représentations de modulation afin d'améliorer les performances prédictives et la capacité de généralisation. Bien que les classificateurs conventionnels aient montré des résultats prometteurs dans des contextes contrôlés, leur aptitude à généraliser à des ruches non observées et à des conditions environnementales variées demeure limitée, en particulier dans des scénarios de déploiement réels.

Des travaux récents en surveillance acoustique des ruches ont montré que les réseaux de neurones convolutionnels (CNN) surpassent les méthodes d'apprentissage automatique traditionnelles lorsqu'ils sont appliqués à des représentations temps–fréquence telles que les spectrogrammes. Parallèlement, les représentations spectrales de modulation ont émergé comme une alternative robuste aux spectrogrammes classiques, car elles encodent explicitement les fluctuations temporelles de l'énergie caractéristiques de l'activité collective des abeilles et présentent une sensibilité réduite au bruit de fond stationnaire. Ce chapitre fait le lien entre ces deux axes de recherche en appliquant des réseaux neuronaux profonds directement aux spectrogrammes de modulation et aux tensorgrammes de modulation.

0.6.1 Représentations basées sur la modulation

Le spectrogramme de modulation et son extension tridimensionnelle, le tensorgramme de modulation, sont utilisés pour représenter les signaux acoustiques des ruches en fonction de la fréquence acoustique, de la fréquence de modulation et du temps. Ces représentations sont obtenues en appliquant une seconde transformée de Fourier à court terme à l'évolution temporelle

de chaque bande spectrale, permettant ainsi l'extraction de motifs rythmiques et de modulations temporelles qui ne sont pas capturés par les spectrogrammes conventionnels.

L'analyse des spectrogrammes de modulation moyennés pour différents niveaux de force de colonie révèle des variations systématiques de la distribution de l'énergie de modulation à mesure que la population de la ruche augmente. Une analyse basée sur le critère de Fisher confirme en outre que certaines régions spécifiques en fréquence acoustique et en fréquence de modulation — notamment autour de 150–200 Hz et 500–600 Hz — présentent un fort pouvoir discriminant entre les catégories de colonies de faible, moyenne et forte intensité, soutenant ainsi la pertinence des représentations de modulation comme entrées pour le suivi automatisé des ruches.

0.6.2 Architectures d'apprentissage profond

Plusieurs architectures de réseaux neuronaux profonds sont évaluées pour modéliser les représentations basées sur la modulation. Celles-ci incluent des CNN 2D opérant sur les spectrogrammes de modulation, des CNN intégrant des mécanismes d'attention spatiale, des réseaux convolutionnels récurrents profonds (CRDNN) incorporant des unités récurrentes de type GRU afin de capturer des dépendances temporelles à plus long terme, ainsi que des variantes 3D de CNN et de CRDNN conçues pour traiter les tensorgrammes de modulation complets.

Les architectures 3D modélisent explicitement l'évolution conjointe de la fréquence, de la fréquence de modulation et du temps, permettant l'extraction de caractéristiques spatio–spectro–temporelles riches. Des mécanismes d'attention sont intégrés afin de renforcer les régions de modulation les plus informatives tout en atténuant les composantes moins pertinentes, améliorant ainsi la robustesse face au bruit de fond et aux artefacts spécifiques aux ruches.

0.6.3 Configuration expérimentale et évaluation

Les méthodes proposées sont évaluées à l'aide du jeu de données UrBAN, qui comprend plus de 3 000 heures d'enregistrements audio collectés à partir de neuf ruches urbaines surveillées sur une période de deux ans. La force des colonies est définie par le nombre de cadres d'abeilles (*frames of bees*, *fob*) obtenu lors d'inspections visuelles et utilisé comme variable cible dans un cadre de régression.

Les performances sont évaluées selon deux protocoles : une séparation aléatoire des données et une évaluation indépendante par ruche, cette dernière reflétant plus fidèlement les conditions de déploiement réel avec des ruches non observées lors de l'apprentissage. L'erreur absolue moyenne (MAE) et le coefficient de corrélation de Pearson sont utilisés comme métriques d'évaluation. Des caractéristiques de référence, incluant les spectrogrammes et les CCFM, sont

considérées à des fins de comparaison, avec et sans application de la soustraction d'amplitude spectrale pour la réduction du bruit.

0.6.4 Résultats et discussion

Dans les deux cadres d'évaluation, les représentations basées sur la modulation surpassent les spectrogrammes et les CCFM utilisés comme références. Dans le scénario de séparation aléatoire, les modèles profonds exploitant les spectrogrammes et les tensorgrammes de modulation atteignent de faibles erreurs de prédiction et des corrélations élevées, confirmant que la force des colonies est principalement encodée dans les motifs de modulation temporelle plutôt que dans le contenu spectral statique.

Dans le cadre de l'évaluation indépendante par ruche, l'ensemble des modèles subit une dégradation des performances ; toutefois, les modèles d'apprentissage profond basés sur la modulation conservent des corrélations significativement plus élevées et des erreurs plus faibles que les approches de référence, indiquant une capacité de généralisation supérieure. L'architecture CRDNN-3D obtient systématiquement les meilleures performances, mettant en évidence l'intérêt de combiner un traitement convolutionnel tridimensionnel avec une modélisation récurrente des dynamiques temporelles.

Des analyses statistiques à l'aide des tests de Friedman et de Nemenyi confirment que les différences de performance observées sont statistiquement significatives pour l'ensemble des métriques et des conditions expérimentales.

0.6.5 Explicabilité et interprétation des modèles

Des analyses d'explicabilité basées sur des cartes de saillance et la méthode Grad-CAM révèlent que les modèles d'apprentissage profond se concentrent sur des régions structurées en fréquence de modulation plutôt que sur des composantes spectrales isolées. À mesure que la force de la colonie augmente, les modèles s'appuient progressivement sur des motifs de modulation plus larges et plus stables, en particulier dans la bande de fréquences de modulation comprise entre 10 et 25 Hz.

L'analyse Grad-CAM en trois dimensions montre que les colonies de forte intensité présentent des activations cohérentes et continues dans le temps, tandis que les colonies de faible intensité affichent des motifs plus intermittents et localisés. Ces observations indiquent que les modèles exploitent conjointement la structure spectrale et son évolution temporelle, renforçant l'interprétabilité des représentations basées sur la modulation pour le suivi des colonies.

0.6.6 Conclusion

Les résultats montrent que les modèles d'apprentissage profond opérant sur des spectrogrammes et des tensorgrammes de modulation constituent une solution robuste et généralisable pour l'estimation non invasive de la force des colonies d'abeilles. En modélisant explicitement les motifs de modulation temporelle et en exploitant des architectures neuronales spatio–temporelles, l'approche proposée surpasse à la fois les méthodes d'apprentissage automatique traditionnelles et les modèles profonds fondés sur des caractéristiques acoustiques conventionnelles. Ces résultats soulignent le potentiel des systèmes d'apprentissage profond sensibles à la modulation pour des applications évolutives et réalistes en apiculture de précision.

0.7 Chapitre 7 : Conclusions et orientations futures de la recherche

0.7.1 Conclusion

Dans cette dissertation, nous avons développé des systèmes de surveillance acoustique généralisables pour les colonies d'abeilles domestiques. Les principales contributions de ce travail peuvent être résumées comme suit :

Facteurs environnementaux : Nous avons étudié de manière systématique l'impact des principales sources de bruit environnemental — sons urbains, parole humaine en arrière-plan et fortes précipitations — sur des représentations acoustiques couramment utilisées, notamment les spectrogrammes, les CCFM et les coefficients en ondelettes discrètes (DWT). Notre analyse met en évidence la manière dont différentes conditions de bruit affectent la robustesse des caractéristiques acoustiques et les performances des modèles dans des scénarios de déploiement réels.

Jeu de données : Nous avons introduit le jeu de données *UrBAN*, un ensemble de données acoustiques de ruches à grande échelle et en accès libre, comprenant plus de 3000 heures d'enregistrements audio bruts collectés auprès de dix ruches sur une période de deux ans. Ce jeu de données est annoté avec des estimations de la population des colonies ainsi que des mesures d'infestation par le varroa, permettant une évaluation complète des méthodes de surveillance acoustique dans des conditions réalistes.

Ensembles de caractéristiques : Nous avons proposé de nouvelles représentations acoustiques basées sur le spectrogramme de modulation et réalisé une analyse approfondie de leur efficacité pour la prédiction de la population des colonies d'abeilles. De plus, nous avons introduit des descripteurs de forme spectrale extraits à la fois des spectrogrammes conventionnels et des spectrogrammes de modulation, démontrant leur pertinence pour la détection des infestations par le varroa.

Approches de modélisation : Nous avons étudié l'utilisation des spectrogrammes de modulation et des tensorgrammes comme entrées pour des algorithmes d'apprentissage automatique classiques ainsi que pour des architectures d'apprentissage profond. Les résultats montrent que ces représentations permettent des prédictions précises et computationnellement efficaces de la force des colonies d'abeilles, tout en assurant une bonne robustesse face à la variabilité entre ruches et conditions d'enregistrement.

0.7.2 Limites de l'étude

Précision du jeu de données et fiabilité des étiquettes : Malgré les résultats prometteurs obtenus, certaines limites doivent être reconnues. L'évaluation phénotypique de la force des colonies peut présenter une variabilité de mesure en raison du caractère subjectif du classement, des contraintes environnementales et des limitations opérationnelles lors des inspections. Cette variabilité peut introduire du bruit dans les étiquettes de référence, ce qui peut affecter les performances prédictives ainsi que l'interprétabilité des modèles.

Données manquantes et incomplètes : Le jeu de données contient également des observations manquantes pour certaines colonies et certaines périodes temporelles. Ces lacunes peuvent avoir influencé la stabilité de l'apprentissage et les résultats d'évaluation, particulièrement si la mécanisme de manque de données n'est pas complètement aléatoire. Bien que des stratégies de prétraitement et de validation croisée aient été appliquées, la couverture temporelle incomplète limite inévitablement la robustesse des représentations apprises et peut contribuer à la variabilité des performances expérimentales.

Validité externe : Les résultats sont également limités par les conditions environnementales spécifiques, les localisations géographiques, les configurations des ruches et les caractéristiques des capteurs utilisés lors de la collecte des données. Les signatures acoustiques des colonies d'abeilles peuvent varier en fonction du climat, de la génétique des colonies, du matériau des ruches ou du positionnement des microphones. Par conséquent, la généralisation du cadre proposé à d'autres contextes apicoles nécessite une validation supplémentaire.

De manière plus générale, ce travail souligne le rôle central de la qualité des données phénotypiques en apiculture de précision. L'efficacité des modèles d'intelligence artificielle dépend fondamentalement de la fiabilité et de la pertinence biologique des données d'apprentissage. Un alignement étroit entre les signaux captés par les capteurs et les mesures phénotypiques réelles au niveau des colonies est essentiel pour développer des systèmes de surveillance robustes et déployables. Les travaux futurs devraient donc privilégier des protocoles de classification standardisés ainsi qu'une meilleure complétude des données.

0.7.3 Orientations futures de recherche

Malgré les résultats prometteurs obtenus dans cette thèse, plusieurs axes de recherche restent ouverts afin d'améliorer les systèmes de surveillance acoustique des ruches.

Déploiement en périphérie et surveillance en temps réel : Les travaux futurs devraient explorer le déploiement des modèles proposés sur des dispositifs embarqués situés directement dans ou à proximité des ruches. L'informatique en périphérie permet un traitement local des signaux acoustiques, réduisant ainsi le besoin de transmission continue des données vers des serveurs centralisés. Ceci est particulièrement pertinent dans les environnements apicoles isolés où la connectivité peut être limitée ou instable. En effectuant l'inférence localement, le système peut fournir une surveillance en temps réel de l'activité des colonies et des événements liés à leur santé, tout en minimisant la latence et l'utilisation de la bande passante. De plus, ce type de déploiement impose des contraintes en termes de ressources computationnelles et de consommation énergétique, ce qui motive le développement de modèles légers et efficaces. L'intégration de l'analyse acoustique avec des systèmes embarqués basse consommation constitue une étape clé vers des solutions de surveillance des ruches autonomes et à grande échelle.

Apprentissage auto-supervisé et modèles fondamentaux : Les travaux futurs devraient explorer l'utilisation de l'apprentissage auto-supervisé et des modèles fondamentaux audio afin d'exploiter de grands volumes d'enregistrements de ruches non annotés. Des approches contrastives ou basées sur la prédiction masquée pourraient permettre d'apprendre des représentations acoustiques robustes et transférables, capables de généraliser à travers différentes ruches, saisons et conditions d'enregistrement, réduisant ainsi la dépendance aux annotations manuelles coûteuses.

Représentation des caractéristiques et dynamiques temporelles : Bien que les coefficients delta et delta-delta des CCFM aient été explorés dans des expériences préliminaires, ils n'ont pas apporté d'améliorations significatives. Cela peut s'expliquer par leur mode de calcul conventionnel, reposant sur des fenêtres temporelles courtes (généralement de ± 2 ou ± 3 trames). Comme démontré dans les chapitres ultérieurs, des contextes temporels plus longs sont souvent nécessaires pour capturer des motifs spectro-temporels pertinents dans les signaux acoustiques des ruches. Les travaux futurs devraient donc étudier l'utilisation de fenêtres temporelles étendues pour le calcul des caractéristiques delta et delta-delta, ou proposer des formulations alternatives mieux adaptées à la capture des dynamiques à long terme.

Enregistrement audio multi-localisation : Alors que le jeu de données actuel repose sur un seul microphone par ruche, de futurs déploiements pourraient bénéficier de l'installation de microphones dans plusieurs compartiments (par exemple, le corps de ruche et les hausses). Cela permettrait une localisation plus précise de l'activité acoustique et offrirait aux modèles la possibilité de capturer des variations spatiales liées à la force de la colonie, à la distribution du couvain et au comportement de butinage.

Réseaux de microphones : L'extension du dispositif de captation à des réseaux de microphones au sein de la ruche pourrait fournir des informations spatiales sur les sources sonores. De telles configurations permettraient d'analyser des motifs acoustiques directionnels et d'améliorer la robustesse face au bruit environnemental, notamment en milieu urbain ou en présence de vents forts.

Formation de faisceaux et traitement spatial du signal : Avec la disponibilité d'enregistrements multi-microphones, des techniques avancées de formation de faisceaux et de filtrage spatial pourraient être utilisées pour renforcer les sons produits par les abeilles tout en atténuant les sources de bruit externes. L'intégration de ces techniques avec des modèles d'apprentissage automatique pourrait améliorer davantage la précision de détection des dynamiques de population et des événements liés à la santé des colonies.

Dans l'ensemble, ces perspectives visent à améliorer la robustesse, la scalabilité et l'interprétabilité biologique des systèmes de surveillance acoustique, afin de les rapprocher d'un déploiement fiable en conditions réelles dans le cadre de l'apiculture de précision.

1 INTRODUCTION

1.1 Honey bee Acoustics

Honey bees (*Apis mellifera*) are vital to environmental welfare. Beyond producing honey, beeswax, and royal jelly, their pollination activity supports biodiversity and major agricultural crops, making them an important agro-environmental "service provider" [1]. In fact, close to 75 percent of the crops in the United States of America producing fruits and seeds for human consumption depend, at least in part, on honey bees for sustained production, yield, and quality [2]. While all types of bees contribute to pollination, honey bees are the main pollinators of almonds, citrus fruits, blueberries and curcubits [2]. Therefore, any change in bee colony behavior and health, as well as drops in population sizes can have drastic effects on the food industry. To this end, continuous monitoring of honey bee health and colony strength is performed by beekeepers.

For many years, beekeepers have been inspecting their beehives mainly by visual examinations. These inspections are usually performed during pollination services and honey production season. As inspections are disruptive for the colonies and are also time-consuming for the beekeepers, they are usually only performed every two weeks. Between inspections, however, changes may occur in a colony that could lead to important health problems that impact on population size.

In the last decade, important colony losses have been reported worldwide [3, 4, 5, 6]. Scientists believe that multiple stressors are responsible for this phenomenon [7, 8, 9]. Development of continuous monitoring systems would greatly help the management of honey bee colonies [7, 8, 9] and allow for the long-term monitoring of stressors. As such, automated beehive monitoring has become a topic of great interest to the beekeeping and honey bee research communities [10].

Recently, with advances in Internet of Things (IoT), cloud computing, and machine learning tools, automated beehive monitoring tools have emerged resulting in the field of "precision beekeeping" [11]. For example, systems based on beehive weight [10, 12, 13], internal temperature [14, 15], humidity [15], and CO_2 levels [15] have been proposed. Initial attempts using beehive acoustics [16, 17, 18] or multisensor approaches [10, 12, 15] have also emerged.

The weight of a beehive can be used as an indicator of honey production and bee population size, with colony weight changes being reported throughout the day. For example, weight decreases have been reported during the night and early morning due to nectar consumption, while it increases with foraging activity during the daytime [19]. The authors in [13] monitored the weight of four colonies using electronic scales and showed that the average weight indicates the amount of stored food (during nectar flow), while the changes could be related to the daily consumption of food. Authors in [10] monitored colony weight for a year and described honey bee foraging activity to collect pollen and nectar during the summer (sharp decrease in weight) while they do not leave

the hive in winter and use the stored food to feed themselves (weight slowly decreasing). Moreover, swarming activity could be tracked based on the weight change of the hive resulting in a sharp decrease.

Temperature stability and regulation in a beehive can indicate the colony's adaptive response and health state. It has been found that while the internal temperature affects the health of the bees and brood, the productivity of the hive is also strongly affected by internal hive conditions [20]. Proper thermoregulation inside the hive, for example, can help decrease mortality rates and increase honey production by reducing the internal food consumption [21, 22]. Moreover, it has been found that a rise in temperature (from $34°C$–$35°C$ to $37°C$–$38°C$) could be a signature of swarming [23].

Relative humidity (RH) is another important factor for egg incubation, larvae growth, colony development, as well as bee behavior, where changes in water availability and larvae feeding have been reported as a function of ambient humidity, hive temperature, and nectar moisture [24]. The authors in [25] used temperature and RH sensors to monitor external conditions, as well as conditions within the brood combs and the nectar combs. Their results showed that RH in the brood combs was the highest and had less daily fluctuations. CO_2 measurement can be used to monitor the metabolism in bees. While it has been shown that the level of CO_2 in the bee colonies correlates with changes in temperature [26], bees are able to adjust internal hiveCO_2 concentration and maintain 0.1–4.25% [27].

Moreover, bees are known to communicate within the colony via vibration and sound [28] produced using gross body movements, wing movements, high-frequency muscle contractions without wing movements, and pressing the thorax against the comb substrates or another bee [29]. Commonly-generated queen sound signals include piping, quacking, tooting [50], hissing and buzzing. Piping, for example, is emitted by young queens to signal readiness for battle with another queen. The sound made as soon as they emerge is called tooting [30], which has been shown to have a fundamental frequency between 400 Hz during the day of emergence, to around 500 Hz two to four days post-emergence. When a virgin queen emerges, several other mature queens may still be in their cells. These typically reply with a sound called quacking. Quacking is described as a series of short pulses (with duration shorter than 0.2 s) having a fundamental frequency of around 350 Hz [30]. Worker bees are also able to produce acoustic signals that are similar to piping. While it was believed that piping in worker bees only happened with lack of a queen or disturbance of the hive, this signal has been observed in colonies that are neither disturbed nor without a queen [31]. When produced by undisturbed foragers, this signal is emitted once every few seconds and has a fundamental frequency of 330–430 Hz with an evident harmonic behaviour [32]. The piping produced by worker bees before or during swarming is shown to have greater frequency variability, ranging from 100 to 500 Hz or beyond, and each pipe is a pulse that has a duration of around one second [33]. Worker piping happens before or during hissing, which is an acoustic signal produced by worker bees when they are distressed or during swarms [34]. Having a large fundamental frequency ranging from 300 to 3600 Hz, the hissing signal has a broadband and noisy structure.

In addition, bees produce buzzing sounds by beating their wings with a frequency of 250 to 300 Hz [35]; hence, they can be audible to the human ear.

Having microphones placed inside the hives has been proposed as a method for real-time precision beekeeping and hive health monitoring with little disturbance to the hives [36, 37]. Recent advances in computational power and hardware miniaturization have enabled portable bee acoustics monitoring systems. For example, the Apivox Smart Monitor, released in 2015, uses the microphone of an Android smartphone to monitor hive sounds to detect swarming and brood, among other factors [38]. More recently, the Bee Health Guru (released in 2019) [39] also proposes the use of a smartphone to record sound inside a hive in 30-second- to 1-minute-long tracks. These sound snippets are then processed by artificial intelligence to determine whether the hive has a queen, has a Varroa mite infestation, or the presence of small hive beetles. Other companies, such as Arnia [40] and Nectar Technologies Inc. [41], have developed multi-sensor solutions to monitor hive health and strength, queen bee presence, and parasite presence, including relying on acoustic data recorded via microphones placed inside the hive.

Among the different modalities used for monitoring honey bee colonies, acoustics has shown to provide a more diverse practical use, including the detection of swarming, queen absence, and hive strength prediction, as well as complementary information to temperature and humidity sensors. Automated acoustics-based beehive monitoring, however, is still in its infancy and many questions about its effectiveness remain. The goal of this thesis is to further probe into the existing limitations of remote audio-based beehive monitoring and provide solutions to overcome these issues. The ultimate goal is to develop tools that allow for remote beehive monitoring in a robust and reliable way.

1.2 Current Challenges and Research Problems

Although significant research has been done in some of the areas of the bee acoustic monitoring, there are still some challenges that need to be addressed. In this section, the most important ones are presented.

1. **Limited studies in predicting beehive strength/health.** The number of studies related to beehive health and strength monitoring using hive acoustics is very limited. Detecting beehive strength is an important factor that is needed for pollination contracts, to monitor honey production, to assess presence of the queen bee, as well as the overall health of a hive. Detection of unexpected shortages in hive strength could be extremely beneficial for beekeepers. Automated monitoring of colony status alongside numerous environmental factors could also allow for the generation of insights around stressors and their impact on honey bee colony survival.

2. **Limited knowledge on the effects of environmental factors.** Despite the burgeoning of acoustics-based beehive monitoring applications, existing tools have relied on conventional audio features developed for speech processing, namely mel-frequency cepstral coefficients (MFCCs), the spectrogram, and features based on the discrete wavelet transform (DWT). It is widely known within the audio processing community, that such parameters are extremely sensitive to environmental factors, such as ambient noise and background speakers. As such, noise suppression and/or characterization of the background noise (known as context-awareness) have been crucial for the development of accurate and replicable systems (e.g., [42, 43]). This should be particularly crucial for urban hives, which are on the rise [44], where various loud urban sounds could interfere with the internal hive sound recordings. To date, no works have addressed this issue, raising concerns on the validity and generalizability of the existing solutions for real-world applications.

3. **Small and non-diverse datasets.** While most published studies have shown the benefits of using acoustics for hive monitoring, the majority have relied on data collected over a short period of time (e.g., 24h), usually in the months of June and July [45, 46, 47]. In order to have a reliable monitoring system, we need audio from different seasons and months, as the bee behaviour changes from summer to winter. Another problem is the limited number of beehives in each study. The most common studies include one to two beehives, which could lead to low generalizability of the monitoring system. As such, published insights may not be indicative of the potential of acoustics based monitoring.

4. **Lack of beehive generalizability.** Lastly, the results seen in the literature have suggested that existing models suffer from poor generalizability [48, 49] and achieve poor accuracy on unseen beehives. In the process of building a machine learning model, there are three different sets of datasets : training, validating, and testing. While the training samples are used to fit the model, the testing samples are used to provide an unbiased evaluation of a final model fit on the training dataset, once the model is completely trained. In the model development phase, the validation set is used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters.

   In order to test model generalizability, usually two different experimental setups have been considered : “random-split” and “hive-independent”. In the former, the entire dataset from all the hives is randomly divided into the three disjoint parts. With the latter, a leave-some-hives-out setup is used, separating training and testing hives and keeping the test hives completely unseen to the trained model. In all cases, the model performance in the “hive-independent” scenario is degraded significantly compared to the “random-split” case, thus signaling poor generalizability to unseen hives. This may be due to suboptimal features that are not truly indicative of the hive parameter being monitored (e.g., strength), to biases in the model that could arise from limited amounts of data, or with the presence of artifacts in the test data. As such, building new features and context-aware solutions will be crucial for the success of automated audio-based beehive monitoring.

Given these limitations, this proposed doctoral research aims to tackle these issues and propose solutions that will enable the development of more robust and accurate acoustics-based monitoring systems with greater potential to be applied in real-world settings.

1.3 Goals and Objectives

The main goals and hypotheses of this doctoral research are described as follows :

1. **Collect large amounts of beehive acoustic data.** A fundamental objective of this research was to collect a large-scale and longitudinal dataset representative of real-world beehive acoustic conditions. To this end, we instrumented multiple hives located in an urban apiary in the Montreal region of Québec, Canada, and continuously recorded audio over a two-year period. The resulting dataset contains thousands of hours of raw acoustic data alongside measurements of colony strength and Varroa mite infestation levels. This extensive collection enables the study of long-term trends, seasonal variability, and the impact of environmental factors on acoustic monitoring systems. The availability of such a dataset is essential for developing and validating machine learning models that can generalize beyond controlled laboratory conditions.
2. **Quantify the impact of environmental factors on existing audio features and develop new tools to enhance them.** As mentioned previously, the most common audio features being used today in monitoring systems include the spectrogram and MFCCs. These features have been widely used in speech processing and have been shown to be extremely sensitive to environmental factors, such as background noise and room reverberation effects. As such, it is hypothesized that beehive monitoring systems based on these features will also suffer from these environmental factors. In the case of hives, these could include rain, wind, surrounding sounds (e.g., tractor noises) and even beekeepers speaking near the hives. To achieve this goal, we have instrumented beehives in an urban apiary and have monitored the collected audio signals over a period of several months. We found three main sources of ambient noise that caused severe detrimental effects on the performance of existing tools. These included : i) train rail squealing, ii) human speech, and iii) heavy rain hitting the hives. While the audio segments affected by heavy rain and train rail squealing sounds could potentially be detected based on weather stations information and/or train schedules, for example, detection of human speech could be more challenging. As such, one of the objectives of this doctoral project is to develop a system to detect environmental noises to ensure that such auditory segments are not included in the machine learning models, thus minimizing their impact on system accuracy. Results of these experiments with this aspect are reported in Chapter 4.
3. **Development of noise-robust features.** Existing monitoring tools have relied on time-frequency signal representations that are known to be sensitive to environmental factors.

We hypothesize that alternate signal representations may provide additional cues about beehive characteristics and provide a more robust view against environmental nuances. To this end, we use the so-called modulation spectral signal representation, which captures the rate-of-change of frequency components, and has been shown to separate signal from noise. This spectro-temporal characterization has been shown to be useful in the development of environment-robust speech tools [51]. As such, we will explore the development of new features and compare them to the existing benchmark ones in the presence of the three noise factors mentioned above. Results of these experiments are reported in Chapter 5.

4. **Varroa mite infestation detection.** Another objective of this doctoral research is to investigate whether acoustic signals can be used as a non-invasive indicator of Varroa destructor infestation levels. Varroa mites remain one of the primary threats to honey bee health worldwide, yet current assessment methods rely on manual inspections that are labor-intensive and infrequent. We hypothesize that infestation alters the internal acoustic activity of the colony through changes in bee behavior, stress levels, and population dynamics, which can be captured using machine learning models. To test this hypothesis, we developed predictive models using classical machine learning algorithms trained on long-term acoustic recordings paired with ground-truth Varroa counts. Special attention was given to feature interpretability and robustness under noisy field conditions.

1.4 Thesis Organization

This dissertation is organized into seven chapters, structured to progressively address the research objectives from data collection and analysis to feature development, modeling, and final conclusions.

Chapter 2 – Acoustic based beehive monitoring : a review of the state-of-the-art This chapter presents a comprehensive review of acoustics-based beehive monitoring systems. It discusses existing signal processing and machine learning approaches used for honey bee colony assessment, including commonly adopted audio features such as spectrograms, MFCCs, and wavelet-based representations. The limitations of these methods in real-world noisy environments are highlighted, along with recent advances in context-aware bioacoustic monitoring and deep learning applications.

Chapter 3 – Urban Dataset This chapter introduces the UrBAN dataset collected as part of this doctoral research. The data acquisition setup, recording protocol, and ground-truth measurements for colony strength and Varroa mite infestation are described in detail.

Chapter 4 – Impact of Environmental Noise and Preprocessing This chapter investigates the impact of environmental noise factors on conventional acoustic features and quantifies their detrimental effects on remote beehive monitoring performance. We explore the use of off-the-shelf voice activity detection (VAD) methods to identify and exclude noise-dominated audio segments,

as well as a preprocessing framework based on spectral amplitude subtraction to enhance beehive acoustic signal quality prior to feature extraction.

Chapter 5 – Modulation Spectrogram-Based Features This chapter proposes the use of modulation spectrogram representations and novel spectral descriptors for beehive monitoring. The developed features are evaluated for two main tasks : honey bee colony strength estimation and Varroa mite infestation prediction. Their performance is compared against traditional benchmark features to demonstrate their robustness and discriminative power.

Chapter 6 – Deep Learning with Modulation Spectrograms and Tensorgrams This chapter explores the use of deep learning models with modulation spectrogram and tensorgram representations as inputs. Convolutional and hybrid neural network architectures are developed and compared with classical machine learning approaches to assess whether learned representations can further improve predictive performance relative to handcrafted features.

Chapter 7 – Conclusions and Future Research Directions The final chapter summarizes the main contributions of the dissertation and discusses the implications of the findings for real-world beehive monitoring. Limitations of the current work are outlined and potential directions for future research are proposed, including self-supervised learning, microphone arrays, and long-term deployment strategies.

1.5 Publications

1.5.1 Publications included in the dissertation

1. **Mahsa Abdollahi**, Pierre Giovenazzo, and Tiago H Falk. Automated beehive acoustics monitoring : A comprehensive review of the literature and recommendations for future work. Applied Sciences, 12(8) :3920, 2022.

2. **Mahsa Abdollahi**, Evan Henry, Pierre Giovenazzo, and Tiago H Falk. The importance of context awareness in acoustics-based automated beehive monitoring. Applied Sciences, 13(1) :195, 2022.

3. **Mahsa Abdollahi**, Nico Coallier, Pierre Giovenazzo, and Tiago H Falk. Performance comparison of voice activity detectors for acoustic beehive monitoring. In 2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pages 320–323. IEEE, 2023.

4. **Mahsa Abdollahi**, Yi Zhu, Heitor R Guimarães, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H Falk. Urban : Urban beehive acoustics and phenotyping dataset. Scientific Data, 12(1) :536, 2025.

5. **Mahsa Abdollahi**, Yi Zhu, Heitor R. Guimarães, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H. Falk. Audio modulation spectral features for improved honey bee colony population prediction. IEEE Sensors Journal, pages 1–1, 2025.
6. **Mahsa Abdollahi**, Yi Zhu, Heitor R. Guimarães, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H. Falk. On the Prediction of Varroa Mite Infestations in Honey bee Colonies via Acoustic Monitoring. Submitted to IEEE Sensors Journal.
7. **Mahsa Abdollahi**, Yi Zhu, Heitor R. Guimarães, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H. Falk. Improved Monitoring of Honey bee Colony Strength via Audio IoT Sensors, Modulation Tensorgrams and Recurrent Neural Networks. To be submitted to IEEE Sensors Journal.

1.5.2 Other publications

1. Guimarães, Heitor R., Abhishek Tiwari, **Mahsa Abdollahi**, Anderson R. Avila, and Tiago H. Falk. "BioME : A Resource-Efficient Bioacoustic Foundational Model for IoT Applications." arXiv preprint arXiv :2602.09970 (2026).
2. Heitor R. Guimarães, **Mahsa Abdollahi**, Yi Zhu, Ségolène Maucourt, Nico Coallier, Pierre Giovenazzo, and Tiago H. Falk, "Benchmarking Self-Supervised Audio Representations for IoT-Enabled Acoustic Beehive Monitoring," in IEEE Internet of Things Journal, vol. 12, no. 21, pp. 45000-45010, 1 Nov.1, 2025.
3. Zhu, Yi, **Mahsa Abdollahi**, Ségolène Maucourt, Nico Coallier, Heitor R. Guimarães, Pierre Giovenazzo, and Tiago H. Falk. "MSPB : a longitudinal multi-sensor dataset with phenotypic trait measurements from honey bees." Nature Scientific Data 11, no. 1 (2024) : 860.
4. Zhu, Yi, **Mahsa Abdollahi**, Ségolène Maucourt, Nico Coallier, Heitor R. Guimarães, Pierre Giovenazzo, and Tiago H. Falk. "Early prediction of honey bee hive winter survivability using multimodal sensor data." In 2023 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), pp. 657-662. IEEE, 2023.
5. Guimarães, Heitor R., **Mahsa Abdollahi**, Yi Zhu, Ségolène Maucourt, Nico Coallier, Pierre Giovenazzo, and Tiago H. Falk. "Adapting Self-Supervised Features for Background Speech Detection in Beehive Audio Recordings." In 2023 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), pp. 663-667. IEEE, 2023.

2 ACOUSTIC BASED BEEHIVE MONITORING : A REVIEW OF THE STATE-OF-THE-ART

2.1 Preamble

This chapter is compiled from material extracted from the following peer-reviewed article : "Automated Beehive Acoustics Monitoring : A Comprehensive Review of the Literature and Recommendations for Future Work [52]"

2.2 Introduction

Building upon the motivations and challenges outlined in Chapter 1, this chapter provides a systematic review of the literature on acoustics-based honey bee monitoring. Over the past decade, increasing interest in precision apiculture has led to the development of numerous audio-driven approaches for assessing colony health, behaviour, and environmental conditions. However, these studies vary considerably in their objectives, data acquisition protocols, feature extraction strategies, and evaluation methodologies, making it difficult to draw consistent conclusions regarding best practices and remaining research gaps.

The purpose of this chapter is to critically analyze and synthesize the state-of-the-art in bee acoustics research. We examine articles published between January 2012 and May 2025 and organize them according to five key dimensions : the origin and context of the studies, their primary research goals, experimental and recording setups, audio analysis and machine learning methodologies, and the degree of reproducibility and data availability. Through this structured analysis, we identify common trends, methodological strengths, and persistent limitations that have shaped the field.

2.3 Methods

English journal and conference papers, as well as electronic preprints, published between January 2012 and May 2025, were chosen as the target of this review. PubMed, Google Scholar, Scopus, IEEE Xplore, and arXiv were queried to collect an initial list of papers containing specific search terms in their title or abstract. Additional papers were identified by scanning the reference sections of these papers. The search terms used included :

1. Apis Mellifera
2. Apis Cerana

TABLE 2.1 : Eligibility criteria for inclusion and exclusion of articles.

Inclusion – Studies using audio to :
Detect bees
Predict hive strength
Detect swarming
Detect queen bee missing
Detect mite attack
Identify bee species
Measure pollination efficacy
Detect environmental effects
Detect arrival and departure of bees
Exclusion – Studies focusing on :
Software development of hive monitoring
Hardware development of hive monitoring

3. Apis Dorsata
4. Apis Bombus
5. Bumblebee
6. Bee
7. Hive
8. Honey bee
9. Apiculture
10. Sound
11. Audio
12. Acoustics
13. ZigBee

These search terms were combined using the following rule : (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9) AND (10 OR 11 OR 12) AND NOT 13. The last rule was included in order to avoid erroneous hits with papers focusing on ZigBee, a wireless transmission technology used with IoT devices and not related to the review. Since the Google Scholar search engine only allows searches of full text or titles, and not titles and abstracts, the query was performed using the flag 'allintitle' to search titles only. On arXiv, Scopus, and PubMed, on the other hand, both abstracts and titles were queried. With IEEExplore, all metadata was queried. From the resulting hits, the inclusion and exclusion criteria presented in Table 2.1 were used.

To assess the eligibility of the selected papers, the titles were read first. If the title did not clearly indicate whether the inclusion and exclusion criteria were met, the abstract was read as well. Finally, when reading the full text during the data collection process, papers that did not meet the review criteria were also rejected. For the included papers, Table 2.2 lists the items that were extracted and used for further analyses. The first category covers the origin of the article, i.e., whether it was published as a journal manuscript, in a conference proceedings, or as a non-peer

reviewed preprint, the first author's affiliation and country, as well as the venue in which the paper was published. These details provide a quick overview of the types of publication available, as well as which countries are working most actively in the field. It is important to highlight that we have chosen to include non peer-reviewed articles, such as those available on arXiv electronic preprints, as there is an increasing trend in publishing pre-prints, especially within the artificial intelligence community, prior to submission to major conferences and/or journals. This also allows for inclusion of a more diverse pool of papers from all corners of the world. Henceforth, such non peer-reviewed papers will be represented by an asterisk.

2.4 Results and Discussion

2.4.1 Origin of the selected articles

Our search methodology returned 59 journal papers, 42 conference papers and 3 pre-prints that met the inclusion criteria. Next, we looked at the location of the first author's affiliation to get a sense of the geographical distribution of research on bee acoustics analysis. Figure 2.1 summarizes the distribution of countries of the included articles. As can be seen, most contributions came from the USA, UK, Italy and Poland, and China. In the figure, countries listed as "others" group all countries in which two or less paper was included; representative countries in this category include Canada, France, Australia, New Zealand, Belgium, India, Latvia, Ireland, Republic of Korea, The Bahamas, Uruguay, and Germany.

2.4.2 Study Goal

The study goals of the included papers can be summarized by Figure 2.3, as a function of publication year. As can be seen, the most common goals in using beehive acoustics are to detect queen bees and swarming events. A handful of articles have also explored the influence of external factors on beehive acoustics, including pollution, chemicals exposure, smoke, microwave electromagnetic field, 2.4 GHz radio frequency exposure, and solar eclipses, to name a few. Moreover, as can be seen from Figure 2.2, the number of papers increased substantially starting in 2018, coinciding with the period in which open datasets started becoming available. The recent drop in publication numbers from 2020 onwards could be due to the limited number of available open datasets, hence suggesting that new and more diverse datasets may be needed to help advance the research.

Table 2.3 provides a more detailed description of the aspects explored in each paper. In this section, we start with the studies that focused on honey bees and proceed with studies focusing on other bee species, such as *Bombus*, *Halictus*, and *Megachile*.

TABLE 2.2 : Items extracted from each study.

Category	Data item	Description
Origin of article	Type of publication	Whether the study was published as a journal article, a conference paper or in an electronic preprint repository.
	Venue	Publishing venue, such as the name of a journal or conference.
	Country of first author affiliation	Location of the affiliated university, institute or research body of the first author.
Study rationale	Study goal	Application or aim of the article.
Experiment setup	Species of bee	Exact name of the bee specie.
	Number of beehives or bees	Exact number of beehives or single bees.
	Time of the recordings	Month and hour of the day for audio recordings.
	Sampling rate of audio files	Sampling rate of audio files
	Microphone information	Resolution and frequency response of the microphone.
	Audio preprocessing	Steps applied to the raw data to prepare it for use by the architecture or for feature extraction.
Audio analysis method	Features	Features extracted from audio files for visualization or a machine learning framework.
	Machine learning algorithm	Name of Machine Learning algorithm if a machine learning framework is used.
	Performance metrics	Metrics used in the study to report performance (e.g. accuracy, f1-score, etc) if a machine learning framework is used.
Reproducibility	Dataset	Whether the data used for the experiment comes from private recordings or from a publicly available dataset.
	Code	Whether the code used for the experiment is available online or not, and if so, where.

2.4.2.1 Honey bees

As a first step in honey bee analytics, one must first discriminate bee sounds from non-bee sounds, such as wind, rain, and urban sounds. Therefore, one of the common goals in the included

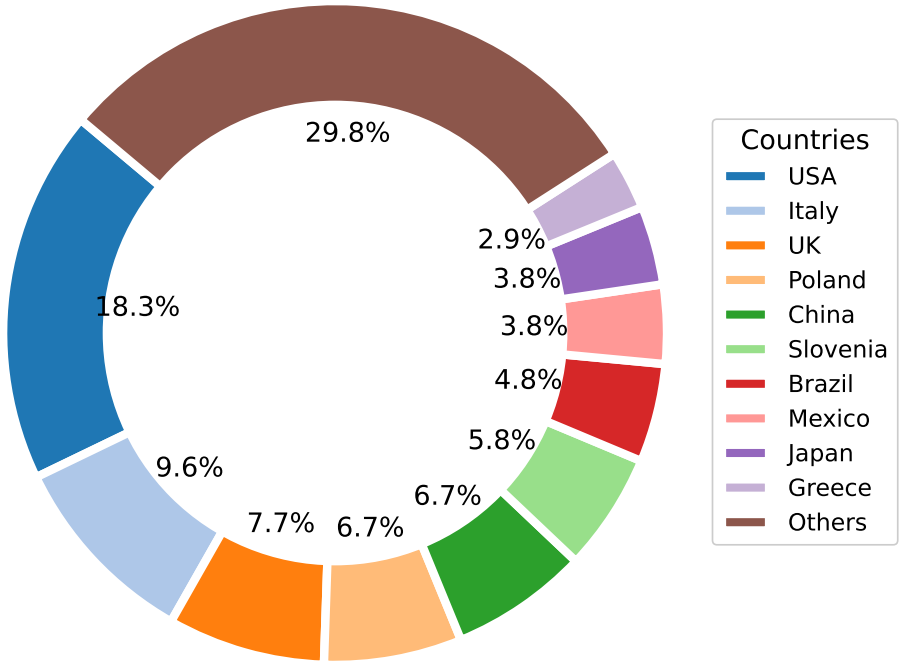


FIGURE 2.1 : Country of first author affiliation.

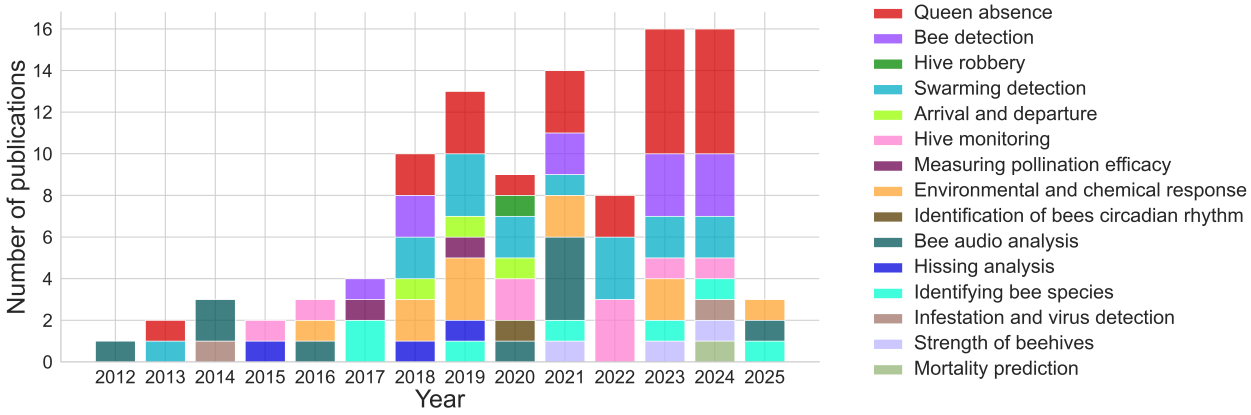


FIGURE 2.2 : Study goal of articles with respect to the year published.

papers is to develop bee activity detection algorithms for (i) classification of bee and non-bee sound as a preprocessing step [61, 55], and potentially use the processed audio for detecting swarming and queen bee absence, and (ii) detection of bee foraging for pollen or nectar [17]. The authors in [55], for example, used a machine learning framework to discriminate bee buzzing from cricket chirping and ambient noise. In [17], in turn, a bee detection system based on the

TABLE 2.3 : List of articles exploring different study goals.

Study Goal	Articles
Bee detection	[17, 53, 54, 55, 56, 57] [58, 59, 60] [61]
Strength of beehives	[62, 63, 64]
Queen absence	[16, 18, 65, 48, 49, 66, 67] [68, 47, 37, 69, 70, 71, 72, 73, 74] [75, 76, 77, 78, 79, 80, 81, 82] [83]
Swarming detection	[84, 85, 86, 87, 88, 89, 90, 91, 92] [93, 94, 95, 96, 97, 98, 99]
Identifying bee species	[100, 101, 102, 103, 104]
Hive robbery	[105]
Bee arrival and departure	[106, 107, 108]
Hive monitoring	[45, 109, 110, 111, 112, 113, 114, 82, 115]
Infestation and virus detection	[34, 82, 116]
Identification of bees circadian rhythm	[117]
Measuring bee pollination efficacy	[118, 119]
Detecting environmental pollution and chemicals	[120, 121, 122, 123, 124, 125, 126]
Smoke reaction	[67]
Effect of electromagnetic exposure	[46, 127]
Impacts of a total solar eclipse on flight behavior	[128]
Hissing analysis	[129, 130, 131]
Bee audio analysis	[132, 133, 134, 135, 136, 137, 138, 139, 140]
Mortality prediction	[141]

harmonic characteristic of buzzes recorded using microphones placed in the orchard trees was designed for pollination mapping and to evaluate pollen production.

A recent study has shown that hive acoustics could be useful to measure hive strength [62], defined as the number of frames within a hive that were at least 60% covered by bees, disease type, and disease severity. While the disease types were not mentioned in the study, four disease severity levels were considered, namely, none, low, moderate, and severe. The designed system is a semi-supervised generative-prediction network (GPN) that takes advantage of environmental variables such as temperature, humidity, and air pressure, alongside labelled and unlabelled audio. The average Huber loss [142] for predicting number of full frames and disease severity were around 0.0166 and 0.0331, respectively, while the accuracy of disease type was around 78%.

Queen bees are essential for colony survival and thrivingness. The death of the queen, or a serious decline in her health, will usually lead to the colony producing a new queen, unless a new queen is introduced by the beekeeper. As such, one of the most active research topics in bee audio analysis has been the detection of a queen bee within the hive. Articles that aimed to detect queen bee presence mostly relied on frequency domain features. While the work in [65, 47] aimed at reporting differences between a hive with a queen and a queenless hive, the work in [16, 48, 66] relied on machine learning algorithms to classify between these two states (presence/absence).

Another common goal has been swarming detection. During swarming, a colony produces a new queen, and the old queen leaves her colony with approximately half of the worker bee population to start a new colony. Swarming is part of the reproduction cycle of the colony and typically occurs when resources are plentiful. After mating, the younger queen takes over the colony that is left behind and is tasked with growing it. With swarming, the result will be two smaller colonies that will use most of their energy for colony growth, rather than honey harvesting. If weather conditions are not suitable (e.g., close to winter), this can lead to colony deaths. Moreover, if not well controlled by the beekeeper, swarming can also have a significant economical impact on honey production. The work in [84, 85, 86, 87, 88, 89, 90, 91] showed that swarming prediction could be predicted using a remote and automated acoustics-based system. Using power spectral density, it was shown that the power increases by around 200 Hz when approaching the swarming period [84, 87].

Furthermore, it is known that honey bees will collect nectar or honey from any source they can find, and that includes a weak, or poorly guarded, colony. The process of invading another hive to steal the stored honey is referred to as “robbing”, and it can occur anytime during the year, but it is most evident in the late summer or early fall, especially during a nectar dearth [143]. Authors in [105] showed that hive robbing could be detected via video and audio recordings of a beehive. A Beemon system [144] was placed right above the entrance of the hive with a camera facing down to record the videos and a microphone was placed inside the hive. To estimate the traffic using videos, a technique based on object tracking and neural networks called Beevee [145] was used to count the number of bees entering or leaving the hive. Their analysis on hive traffic video and inside hive audio showed that before robbery, the target hive had become weak and an easy target.

Other important factors contributing to hive health can include the presence of parasites [34] and exposure to environmental pollutants [120, 121], to name a few. In fact, recent research has shown that honey bees can be used as environmental bioindicators for heavy metals and polycyclic aromatic hydrocarbons [146]. The work in [34], for example, showed that honey bee hive acoustics could be used to detect a Varroa mite infestation, whereas the work in [147, 148] showed that hive acoustics could be used to detect certain atmospheric pollutants. Using four features, namely, spectral centroid, peak frequency, frequency standard deviation, and bandwidth (range of predominant frequencies with the threshold of -9 dB), coupled with a machine learning framework could

lead to Varroa mite detection [34] with an accuracy of 81%. In [121], in turn, changes in acoustic behaviour were monitored and classified using machine learning models, when hives were exposed to acetone, trichloromethane, glutaric dialdehyde, and ethyl ether, and compared to the control condition of blank air. Using MFCCs (mel-frequency cepstral coefficients, a widely-used feature in audio analytics) as features and principal component analysis (PCA) to reduce the dimensionality to two, clear differences in the beehive sounds in the morning, relative to the afternoon, were seen in the control group. Smoke is also known to generate a defensive response in bees and is used by beekeepers when manually inspecting hives. The work in [67] showed that smoke can have a direct effect on hive acoustics with several harmonics showing up in the low-frequency spectrum (0–500 Hz).

In addition to the above factors, researchers have also explored the short-term impact that microwave electromagnetic fields [127] and radio frequency electromagnetic radiation [46] have had on hive acoustics. No significant changes were reported; in [46], the same was also true for temperature and humidity sensor data. Designing a hive monitoring system with multiple sensors, in addition to audio, has also been popularized recently [45, 111]. The work in [45], for example, also included measures of temperature and humidity inside the hive, bee combs' weight, and number of entries to the hive to monitor the colony. The long-term monitoring of the colony showed dramatic weight drops, swarming detected in the audio spectrogram analysis with frequency around 400 Hz, as well as entrance counts and temperature.

Lastly, hive acoustic analysis has also been used as a means to better understand bee behaviour. For example, in [131], sound analysis was used to monitor the learning of worker bees (*Apis mellifera*). In this study, bee hives were presented with two odours, where one of them was also presented alongside an electric shock. While bees did not produce any different sound upon odour presentation only, they responded to the electric shock with a strong hissing response defined as the production of loud sounds or vibrations by simultaneous wing movements of several honey bees and emit short (0.5–1.0 s) bursts of broadband-frequency sounds [149]. After learning, many bees were shown to hiss when presented with the odour that was previously paired with the electric shock, even though the shock was no longer being presented, with fewer bees hissing upon presentation of other odours. While hissing is known to be an aposematic signal [149], the authors in [130] showed that Japanese honey bees (*Apis cerana japonica*) also hissed when there was no obvious threat near the colony. In a later study by the same authors [129], they explored the relationship of hissing onset and offset times with sunrise and sunset times, as well as environmental factors. They showed that honey bees hiss daily during daytime, most frequently at dawn, and the hissing onset and offset times were close to sunrise and sunset times, respectively.

The authors in [141] highlight the potential of integrating various sensor modalities—such as temperature, humidity, and acoustic signals—to anticipate winter mortality in honey bee colonies. By leveraging machine learning techniques, the authors demonstrate that early-season data can

provide valuable insights into colony health and survivability, offering a proactive approach to hive management.

2.4.2.2 *Bombus* and Other Bee Species

Bumble bees' different acoustics behaviour in the colony and in the field could contain information that aids our understanding and ability to monitor these important pollinators. While some researchers investigated the acoustics in the field for flight [53, 100, 128, 119] and buzz-pollination analysis [101], others used in-colony acoustics for foraging activity [106, 107], defence [150] (reaction to mammalian breath and CO2-enriched air), dominance [138] (fundamental frequencies between 300–400 Hz), and thermoregulation [151] (broadband acoustic signal with fundamental frequencies between 180–200 Hz) monitoring [138]. The relationship between "buzz-pollination" and "floral sonication" has been investigated in measures of duration, frequency, and amplitude in [132]. The results showed that airborne recordings of sonication vibrations in *B. impatiens* characterize duration and frequency faithfully while acoustic amplitude (relative peak amplitude and sound pressure level) are not correlated with vibrational amplitude.

Detecting different species of bees is another research area that has increased interest recently. This information is useful in both agricultural production and ecological research, as it helps to understand and protect the biodiversity of these important pollinators and their habitat and also evaluate the pollination services. Authors in [100], for example, designed a system to distinguish the flight sounds of three pollinating bee species in Japan (i.e., *Apis mellifera, Bombus ardens, Tetralonia nipponensis*) and one hornet species (i.e., Japanese yellow hornet, *Vespa simillima xanthoptera*, a predator of honey bees in Japan. Results showed clear differences in the spectra of these sounds, in MFCCs and in fundamental frequency; classification was performed by a support vector machine (SVM) classifier. Using acoustic properties such as peak amplitude, peak frequency, and duration to identify different species of bumblebees during "buzz-pollination" or "floral sonication" (bees vibrating flowers) has been the aim of some studies [101, 139, 152]. The results showed that body shape or wing size can be correlated with fundamental frequency [101], and also duration of the buzzes was shown to differ with body size (larger bees producing shorter buzzes) in six different species of bumblebees (*B. hortorum, B. lapidarius, B. pascuorum, B. pratorum, B. terrestris, and B. lucorum*) [139]. Moreover, using amplitude, frequency, and duration of flight and "floral sonication" sounds of 15 pollinating bee species (see Table 2.4) on *Solanum lycopersicum* flowers showed that the sonication sounds are more relevant to bee species recognition as the machine learning algorithms reached an accuracy of 73.4% on sonication and 64% flight sounds, respectively [152].

In fact, detection of foraging activity has also been explored to provide insight into pollination service [107, 108]. The work in [108], for example, used microphones installed in front of the colony entrance of three different species of bumblebees (*B. pascuorum, B. humilis and B. hypnorum*)

to detect arrival and departure of the bees. The daily histogram of departure on each day showed that there are different foraging habits in different species. *B. pascuorum* were mostly active around noon and in the afternoon, while less active in the morning. On the other hand, *B. hypnorum* were more active after 3 p.m. Another interesting observation was that light rain in the afternoon made *B. humilis* workers stay inside, but it did not affect *B. hypnorum*. Moreover, inspired by computational auditory scene analysis (CASA), the work in [53] used focal templates to recognize audio segments which contained buzzes. To discriminate between arrivals and departures, an algorithm based on the maximum absolute difference between two consecutive peaks of the smoothed signal envelope were used. Classification of arrival and departures reached sensitivity and precision levels of 95%.

In [107], in turn, the foraging activity of the *B. sylvicola* and *B. pascuorum* was investigated, and a system based on the time it takes for the maximum peak amplitude of a buzz to be achieved was proposed to classify arrival and departure. Peaks closer to the beginning of a buzz were indicative of an arrival, otherwise a departure was inferred; an overall accuracy around 96% was achieved. In addition to foraging activity detection, the authors of [118] also proposed the use of spectral contents of acoustics to predict pollination efficacy of the bumblebee (*B. Terrestris*) and wild bees in the field; similar results were found in [119] for *B. balteatus* and *B. sylvicola*.

The work in [117], in turn, relied on audio to identify the colony's circadian rhythm. One-second-length audio were recorded every 15 minutes for 9 days to investigate the existence of circadian rhythm. Using the root mean square (RMS) signal level as bee buzz or loudness, RMS growth in the early hours of the morning and repetitive cycles were noticed. Time range between 11 p.m. and 4 a.m. was concluded as the bees' deep night, as bees have repetitive behaviour during which colony activity decreases the most. As an attempt to complement this study, the authors in [137] investigated sound indices features to identify the bee day as a convenient time for bee colony characterization. They showed that by using temperature, time, and spectral entropy it is possible to find colony characteristics which are the hive-specific hours when the sound of a particular colony is most valuable. Using a temperature filtering step, they rejected recordings collected during weather conditions that are not favourable to bees, and by comparing each hive's spectral entropy daily trends, 11 :00 a.m. to 4 :00 p.m. for the fall and winter seasons were considered as a bee day.

Lastly, in [128], the authors investigated the effect of the 2017 total solar eclipse, which caused a drop in sunlight by three orders of magnitude and temperature by 10–15 ℃ (21 August 2017, which traversed over 5000 km from coast to coast of North America). Hive acoustics were used to monitor flight patterns of *Apis mellifera, bombus, Megachile,* and *Halictus* across the Pacific Coast, Rocky Mountain, and Midwest regions of the United States. It was found that buzzing rates changed as a function of light intensity (and not temperature drop), with the bees ceasing flying during periods of complete darkness.

2.4.3 Experiment setup

2.4.3.1 Bee species

Western honey bees (*Apis mellifera*) are the most common specie of bees explored in the included studies. Not all of the studies mentioned the exact subspecies of *A. mellifera* used in their investigation. Notwithstanding, Table 2.4 shows the bee species and subspecies studies in the respective articles. In the case of detecting arrival and departure of bees and measuring pollination efficacy, the genus Bombus was mostly present [106, 107, 119]. As shown in [139], buzzing vibrations exhibit species-specific characteristics at least in genus Bombus. Therefore, it is vital to report the specie name under experiment in order to better understand the relationship between buzz production and specie.

2.4.3.2 Number of beehives

The studies included in this review used different number of beehives. Some studies (11.5%) used the audio of a limited number of bees for a specific purpose. In [131], the authors used the audio of 104 bees (forager bees which were caught at feeders placed nearby the hives) to investigate hissing patterns in *Apis Mellifera*. Also, in [108], 60 bees from 3 species of bumblebees, *Bombus pascuorum*, *B. humilis* and *B. hypnorum*, were used to monitor daily foraging activity and distinguish between their arrivals and departures. The authors in [139], studied similar variables using groups (n=105) of *B. hortorum, B. lapidarius*, *B. pascuorum*, *B. pratorum*, *B. terrestris*, and *B. lucorum* to determine the similarity of vibrations among different species of bees and examine the extent to which vibrations generated from buzz pollination differed from those produced in a defensive context. While the majority (49.5 %) of the studies did not mention the number of beehives, the most common setups in the 51.5% of the remaining studies included between 1 and 6 beehives. A limited number of studies used a larger number of beehives [62]* [106, 107, 97]. For example, in [106, 107], 46 wooden domiciles were deployed to collect data from *B. spp* in 2018 and 2019. Out of 46 domiciles, only 6 and 11 were occupied by different bumble bee species in 2018 and 2019, respectively. Moreover, the work in [62]* recorded audio of 26 beehives which provided a very large number of audio samples and “hive-independent” models (predicting the beehive state in hives different than the ones where the model was trained on) were used to predict hive strength.

2.4.3.3 Time of the recordings

Most of the studies reported that their data collections took place in the months of June [84, 106, 107, 88, 134, 95] and July [106, 107, 88, 134, 135, 119, 92], with a few extending to August [106, 47, 69] and September [139, 154]. Some studies reported that acoustic changes could be

TABLE 2.4 : Bee Species.

Bee species	Articles
Apis mellifera	[84, 87, 110, 120, 121, 88, 117, 111] [89, 46, 68, 122, 131, 34] [62]
A. cerana	[45, 126]
A. mellifera ligustica	[16, 105, 65, 48, 49, 55, 67] [47, 135, 92, 69, 63][83]
A. cerana japonica	[129, 130]
A. mellifera carnica	[66, 55, 127, 47, 37, 69, 71, 73]
Augochloropsis brachycephala	[152]
Amegilla murrayensis	[153]
Bombus spp.	[132, 128, 102]
B. impatiens	[132]
B. ardens	[100]
B. latreille	[128]
B. balteatus	[119]
B. terrestris	[118, 101, 139, 154, 104]
B. pascuorum	[107, 101, 139, 108]
B. sylvicola	[119, 106, 107]
B. morio and *B. atratus*	[152]
B. humilis and *B. hypnorum*	[101, 108]
B. argillaceus, B. jonellus, B. ruderarius, B. sylvarum	[101]
B. hortorum, B. lapidarius, B. pratorum, , B. lucorum	[101, 139]
B. nevadensis, B. frigidus, B. flavifrons, B. mixtus	[106]
Centris tarsata and *Centris trigonoides*	[152]
Eulaema nigrita	[152]
Exomalopsis analis and *Exomalospsis minor*	[152]
Halictus	[128, 136]
Megachile	[128, 136]
Melipona bicolor and *Melipona quadrifasciata*	[152]
Osmia brevis (oligolectic)	[134]
Pseudaugochlora graminea	[152]
Slovenian apis mellifera carnica	[16]
T. nipponensis and *V. s. xanthoptera*	[100]
Xylocopa nigrocincta and *Xylocopa suspecta*	[152]

seen according to the time of the day [122, 117]. By exploring different acoustic properties during the morning, midday, afternoon, and night, authors in [122] showed that audio frequency changes slowly through the day. Measuring data over a 24h period was a common setup in many studies [45, 129, 66, 46, 130]. On the other hand, some studies focused on recording audio only when the bees are the most active. For example, 3pm-6pm was used for the purpose of swarming detection [84, 88], while arrival and departure activity was monitored between 7am-5pm and 8am-6pm in [106, 107] and in [108], respectively.

TABLE 2.5 : Microphone characterization

Microphone frequency range (Hz)	Articles
6–20,000 Hz	[91]
15–20,000 Hz	[55]
20–20,000 Hz	[66]
30–20,000 Hz	[126]
40–20,000 Hz	[131]
50–10,000 Hz	[119]
70–16,000 Hz	[135]
100–15,000 Hz	[46, 67]

2.4.3.4 Sampling rate of the audio signals

Sampling rate has a direct effect on the quality of the recorded audio, on storage requirements, and on computational resources needed. It was found that almost 43.1% of the articles did not mention their audio sampling rate, while 14.7% and 7.8% mentioned using 44.1 kHz and 22.05 kHz, respectively. As various studies have emphasized that acoustic events of interest lie typically between 100 -1000 Hz [155, 122], some studies relied on downsampling from 22.05 kHz to 4 kHz [53, 66, 37], or even 3 kHz [117], to reduce storage requirements while still keeping useful information for analysis.

2.4.3.5 Microphone Characteristics

One important aspect in audio analysis is the description of microphone characteristics, in special the frequency range that the microphone can cover. Table 2.5 shows the frequency ranges of the microphones reported in articles that provided this information. As can be seen, the great majority did not report this information (61.4%). While the upper end of the spectrum clearly covers the regions of interest for beehive acoustics, of greater importance is the lower end of the spectrum. For example, lower frequencies, around 100 Hz, have been linked to fundamental frequencies of some bee species such as *B. hortorum*, *B. argillaceus* and *B. pascuorum* [101, 106]. Detection of such characteristics, with a microphone with a frequency resolution starting at higher than 100 Hz, for example, would not be possible. As such, accurate characterization of the used microphones should be detailed in future works to facilitate replication.

2.4.3.6 Audio Pre-processing

Most of the studies reviewed did not mention any pre-processing steps for the audio analysis, thus it is not clear how background noise was processed and/or removed, or if downsampling was applied. Nevertheless, five manuscripts mentioned using a filtering step to remove noise, as

well as high frequencies in the audio files. Different filters were used, including a Butterworth, Chebyshev, Elliptic, and a Bessel filter [102]. In [121], a 10th order Butterworth band-pass filter with edge frequencies set at 50 Hz and 1500 Hz was used to remove background noise (e.g., bird sounds, workers' voices, and car noises). The work in [34], in turn, relied on a 7th order Butterworth low-Pass filter with a cut-off frequency at 4 kHz for feature extraction for infestation detection. In [100], the goal was of bee type identification, and a 12 kHz low-pass filter was applied to eliminate only very high frequency sounds, such as machinery. All of these studies, however, did not report results before (i.e., with raw signals) pre-processing, thus it is not clear how much advantage resulted from this signal processing step or if removing high frequencies could be detrimental to some other classification tasks not explored in the related papers. Nonetheless, the gains achieved in storage requirements and real-time processing capability could be well worth the downsampling schemes.

2.4.4 Audio Analysis Method

We divided the bee acoustic audio analysis methods into two categories : research that used machine learning algorithms and those that did not. For the first category, we look at the feature extraction and classification methods used in each articles. For the second category, only the features used and/or methods of visualization are reported. Table 2.6 lists the different features that have been extracted from the audio signals, some of them serving as input to different classification algorithms, and more details are provided below.

2.4.4.1 Features

Mel-frequency cepstral coefficients (MFCCs) [156] are the most common set of features used in studies that exploited a machine learning framework. MFCCs mimic some of the processing performed by the human hearing system, such as the nonlinear perceptual mel-frequency scale of pitch. These features have been widely used on speech processing applications, including speech/speaker/emotion recognition, to name a few. Around 30% of studies analyzed MFCCs to extract information for bee detection, queen absence and swarming detection, bee species identification, as well as environmental effects, with the three first coefficients showing the most discrimination. The authors in [18] used improved MFCCs (IMFCCs) [157], which invert the Mel filterbank to emphasize higher-frequency components that are under-represented in standard MFCCs. In conventional MFCC extraction, the Mel scale allocates a higher density of filters below approximately 1–2 kHz and progressively fewer filters above 3–4 kHz, effectively biasing the representation toward low-frequency content. IMFCCs address this limitation by increasing spectral resolution in the higher-frequency range (e.g., above 3 kHz for a 16 kHz sampling rate), enabling the capture of additional discriminative information. However, the effectiveness of IMFCCs depends on the sam-

pling rate, since high-frequency content is bounded by the Nyquist frequency; thus, their benefit is more pronounced when sufficiently high sampling rates are used.

In [88], the short-time Fourier transform (STFT) calculated with filterbank and overlap-add methods was used to detect swarming events. The STFT was calculated using 128, 256, 512 and 1024 bins to investigate the impact of window width. The best time localization and frequency resolution achieved was with 512 bins, thus inline with other studies [48, 55]. In [47], in turn, the performance of the fast Fourier transform (FFT), the S-transform (S-T), and the STFT were compared for the task of queenless hive detection. The S-T was shown to be very effective in identifying time-varying events, such as the “warbles” mentioned in [158] and which are indicative of a queenless hive. STFT based features were the second most used in the reviewed works.

In [45], in turn, a spectrogram with a frequency range of 0-512 Hz and resolution of 8 Hz was used to visualize bee behaviour before and during swarming. It was found that higher energy in the 400–500 Hz band could be an indicator of swarming activity. In [48], the authors used MFCCs, mel-scaled spectrogram, and the time aggregated time-frequency spectrogram produced by Hilbert-Huang transform (HHT) [159] (a time-frequency signal analysis method based on empirical mode decomposition) with different classifiers to detect queenless hives in “hive-dependent” and “hive-independent” scenarios. The results indicated that all tested features could discriminate queenless hives, with the fusion of HHT-based features and MFCCs showing to improve accuracy for the “hive-independent” setup, thus better suitable for applications that generalize to unseen hives. In comparison with the spectrogram, the HHT better captured the harmonic information, due to its higher resolution, relative to the MFCCs, hence the observed gains. In a similar vein, the constant-Q transform (CQT), a time-frequency transform widely used in music processing, was used for bee detection in [54]. Experiments with a bee detection system showed that MFCCs could better distinguish bee sounds from noise, as compared to the mel-scaled spectrogram and the CQT.

Another powerful time-frequency analysis method is the wavelet transform (WT). In [65], the discrete wavelet transform (DWT) was tested for queenless beehive detection. Results indicated that a DWT with a 10-level dyadic filterbank resulted in a clear distinction between queenless and queenright hives. The work in [49] showed that Continuous Wavelet Transform (CWT) outperformed other time-frequency features such as MFCCs, spectrogram, HHT and DWT in queenless beehive detection with ”hive-independent” setup. The work in [89] also briefly discussed the use of the DWT for swarming detection and compared it with the spectrogram; improved results were attributed to the increased frequency resolution provided by the DWT. Moreover, they showed that CWT had a good signal representation with a higher computational cost.

Linear predictive coding (LPC), motivated by the speech/audio compression field [160], has also been used as a feature extraction method for queen presence detection and swarming detection [68, 91]. The work in [91], for example, compared LPC features with MFCCs and showed that MFCCs were slightly better for swarming detection (82.2% vs 80.4% accuracy). Power spectral density (PSD) was also explored in [84] to look for spectral power changes in certain frequency

TABLE 2.6 : Feature extraction methods explored.

Feature	Articles
Fast Fourier transform and s-transform	[47]
Spectrogram (STFT)	[17, 45, 109, 49, 88, 89] [55, 122, 131, 34, 134, 47, 94, 95]
Mel-scaled Spectrogram	[48, 49, 54, 55, 61, 57, 93, 94]
Mel-frequency cepstral coefficients	[16, 48, 86, 49, 120, 121, 54, 117, 66, 100, 89] [101, 102, 152, 90, 91, 37, 61, 56, 70, 58, 71, 95]
Improved Mel-frequency cepstral coefficients	[18]
Discrete wavelet transform	[65, 49, 89]
Continuous wavelet transform	[49, 89]
Hilbert-Huang transform	[65, 48, 49, 89]
Linear predictive coding	[68, 91]
Power spectral density	[84, 110]
constant-Q transform	[54, 56, 70]
Temporal features	[111, 132, 128, 46]
Spectral features	[106, 34, 139, 70, 95]
Computational auditory scene Analysis	[107, 53, 119]
Sound indices	[120, 137]

ranges a few days prior to swarming; experiments showed that increases in spectral power around the 200 Hz region was an indication of swarming. Alternately, other studies have explored changes in audio signal duration and amplitude, as well as changes in spectral characteristics. While the work in [128], for example, explored changes in buzz duration and amplitude due to solar eclipse, the work in [132], showed changes in duration, amplitude and fundamental frequency during floral sonication vibration (a behaviour in which certain bees generate vibrations by contracting their flight muscles to extract collected pollen). In turn, the authors in [34] showed that peak frequency and relative peak amplitude could be used to detect varroa mite infestation, whereas in [139], these measures were used to discriminate between pollination from defense buzzes in bumblebees. Computational auditory scene analysis (CASA) tools (see e.g., [161]) were also used in [53] and [119] to detect the buzzing of bees from environmental recordings and to measure pollination efficacy by detecting bee buzzes in the environment, respectively. Lastly, the work in [120] proposed new sound indices, such as the acoustic complexity index, acoustic diversity index, acoustic evenness index, and the bioacoustic index to characterize changes in beehive acoustics when exposed to certain chemicals.

2.4.4.2 Machine learning algorithm

A typical machine learning framework encompasses signal measurement, pre-processing, feature extraction, and, lastly, classification. In the area of bee acoustic analyses, many different classifiers have been explored. Table 2.7 summarizes the different classification algorithms used. As can be seen, the most common classifier is a support vector machines (SVM) [162], a kernel-based method that projects data into higher dimensions in which a hyperplane can separate the classes. Authors in [118] used an SVM to predict the pollination efficacy of bumblebees. The performance of the SVM using different input features (e.g., MFCCs, HHT, mel-scaled spectrogram), on the other hand, was explored for queenless hive detection in [48]. The impact of different kernels, including linear, quadratic, and radial basis function, was compared in [121] against Random Forest classifier [163] and K-nearest neighbors method for detecting common air pollution chemicals such as Acetone, Trichloromethane, Glutaric dialdehyde and ethyl ether; the SVM with radial basis function as the kernel showed the best accuracy. To detect parasite infestation, authors in [34] compared linear discriminant analysis (LDA) with SVM and found that while LDA outperformed the SVM in most cases, the SVM had lower computational cost and complexity. Logistic regression has also been explored intensively, especially for queenless hive detection [66, 16, 37].

Decision trees have been widely used for classification as they are easy to interpret [164]. Decision tree has been used to classify bee specie [101, 102, 152]. A performance comparison between decision tree and other classifiers such as SVM, Logistic regression and random forest showed that they are not usually the most accurate one. Naive Bayes has also been used in classifying bee specie [101, 102]. Authors in [102] showed that Random Forest, Logistic Regression and K-nearest neighbors outperformed Naive Bayes in detecting bee species.

The abovementioned methods have been termed “classical” or “conventional” machine learning paradigms, to separate them from the more recent and burgeoning deep-learning neural network based methods. The work in [16], for example, proposed the use of so-called Long Short-Term Memory (LSTM) recurrent neural networks (RNN) for queen bee presence detection. A comparison between an LSTM, a multi-layer perceptron (MLP) neural network, and logistic regression was made and showed the power of the LSTM for the task at hand. The MLP was again explored in [18] for queen bee presence detection and showed to outperform classical SVM and KNN classifiers. It was also tested for swarming detection in [90, 91] and compared to traditional methods, such as Gaussian mixture models (GMM) and hidden Markov models (HMM), showing its superiority.

Recently, convolutional neural networks (CNN) have gained popularity, especially within computer vision tasks. To make them directly applicable to bee acoustics analysis, researchers have relied on image-like inputs, such as spectrograms, mel-scaled spectrogram, or other 2-dimensional time-frequency representations of the audio. Authors in [48], for example, relied on the spectrogram as input to a CNN with 4 convolution layers to detect queenless hives. The challenge of training “hive-independent” models was emphasized and an SVM was shown to better generalize

to unseen hives. Notwithstanding, state-of-the-art accuracy was achieved with the CNN for the "hive-dependent" scenario. CNN architectures were also explored in [54] for bee detection with results outperforming SVM, Random Forest, and XGBoost (a decision tree-based ensemble learning algorithm [165] that uses a gradient boosting framework) classifiers. In a similar vein, the work in [55] compared the performance of 2-D (spectrogram input) versus 1-D (raw audio input) CNNs and concluded that the latter resulted in more robust decisions, especially when test data differed from the training data by beehive, location, time, and bee race.

Transfer learning has emerged as a practical approach, where models pre-trained on large-scale audio datasets (e.g., environmental or speech audio) are fine-tuned on limited beehive recordings to improve classification or anomaly detection accuracy. The authors in [166] investigate the effectiveness of deep learning models in automatically classifying beehive audio recordings. The authors evaluate several architectures, including convolutional neural networks (CNNs) and pre-trained audio transformers, on tasks such as identifying normal hive activity versus stress conditions. Their results demonstrate that deep learning models—especially those using transfer learning from large-scale audio datasets—can achieve high classification accuracy and outperform traditional machine learning approaches.

In parallel, self-supervised learning offers a promising solution for harnessing the vast amounts of unlabeled beehive audio by training models to learn meaningful representations through tasks such as masked audio reconstruction or contrastive learning. These representations can then be fine-tuned for downstream tasks like swarm detection or population estimation. Authors in [62] developed a semi-supervised audio representation learning model to assess beehive strength, effectively utilizing unlabeled data to model both low-level audio features and temporal dynamics. Authors in [104] investigate the application of transformer-based neural networks for classifying bee species based on their buzzing sounds. The researchers implemented and evaluated several models, including the Audio Spectrogram Transformer (AST), Self-Supervised AST (SSAST), and Masked Autoencoding AST (MAE-AST). These transformer models were compared against traditional convolutional neural networks like EfficientNet V2 [167] and Pre-trained Audio Neural Networks (PANNs) [168]. The transformer architectures, particularly when combined with pre-training and data augmentation techniques, demonstrated superior performance in accurately identifying bee species from acoustic data.

Moreover, federated learning presents an innovative path forward for collaborative model training across geographically distributed hives or apiaries while preserving data privacy. This decentralized approach allows beekeepers to contribute to global model improvement without exposing sensitive or proprietary recordings, fostering scalable and privacy-aware acoustic monitoring solutions. Authors in [60] explored different deep learning models, namely DenseNet121, ResNet50, and InceptionV3 fine-tuned for detecting the presence or absence of the bees. Moreover, they evaluated the federated learning approach with a VGG16 model on a local dataset that shared the updated model with the central server. The global model achieved an accuracy of 65%.

TABLE 2.7 : Classification algorithms.

Classifier	Articles
Linear discriminant analysis	[34]
Logistic regression	[16, 66, 152, 37, 56]
Support vector machines	[118, 48, 121, 54, 117, 100, 55] [68, 101, 102, 34, 152, 62, 56, 70, 63]
K-nearest neighbors	[121, 56, 125, 126, 116]
Naive Bayes	[101, 102]
Decision Tree	[101, 102, 152, 125, 126]
Random Forest	[120, 121, 54, 101, 102, 152, 56, 70, 63, 125, 99, 82]
XGBoost	[54, 126]
Multi-layer perceptron	[16, 18, 90, 63][62]
Convolutional neural networks	[48, 49, 54, 55, 57, 72][62]
Long Short-Term Memory	[16, 113, 77]
Transfer Learning	[73, 166, 60]
Self-Supervised Learning	[104]
Federated Learning	[60]
Gaussian mixture models	[86, 91]
Hidden Markov models	[86, 90, 91, 98]

2.4.4.3 Performance metrics

In classification tasks, different metrics can be used as figures of merit, with some showing improved robustness to unbalanced datasets. Table 2.8 lists the metrics used in the papers that focused on machine learning based classification. As can be seen, the most common metric is accuracy, followed by area under the curve (AUC). The majority of the other metrics rely on the classification confusion matrix, and includes e.g., sensitivity, F1-score, and recall. For cases with class imbalance, measures such as F1-score and AUC are most appropriate for binary classification task.

2.4.5 Reproducibility

Reproducibility is an important part of science : reproducible results can help move research fields forward, especially in new and interdisciplinary fields, such as beehive acoustics analysis. Here, we evaluate reproducibilty by using two key criteria : the availability of recorded data and the availability of written code.

TABLE 2.8 : Classification evaluation metrics.

Evaluation metric	Articles
Confusion matrix	[107, 72, 140, 126]
Accuracy	[16, 17, 18, 86, 120, 121, 54, 117] [100, 55, 101, 34, 152, 90, 91, 108, 62, 83, 56, 57]
Area under the curve	[48, 49, 66, 102, 58, 71] [61]
Sensitivity	[53, 70, 78]
Selectivity	[53]
Precision	[49, 56, 57]
Recall	[106, 49, 56, 57]
F1-score	[106, 54, 49, 152, 56, 57]

2.4.5.1 Data availability

From the review articles, three main sources of public bee audio data was identified. The first source is the so-called Nu-hive project [67] data, which consisted of audio for queen state identification, smoke reaction, and bee detection. The public release of this dataset is related to queen state identification from two different beehives. Authors in [65, 48, 49][83]* used the Nu-hive data for queen absence detection, whereas [61]* used part of this data for bee detection. The second source is the OSBH project [169] audio data, acquired through a citizen science initiative where members of the general public were asked to record sounds from their personal beehives along with some metadata. Given its nature, recordings from the OSBH project present great diversity, including different recording devices, different environments where the hives were placed, and even different positions for the microphones inside the hive. This diversity, while making the dataset valuable for evaluation in real-life conditions, also makes it extremely challenging for machine learning applications. Authors in [86, 90, 91] used OSBH data for swarming detection, whereas those in [54][61]* used it for bee detection.

Lastly, the third public source is a dataset called BUZZ, a bee detection dataset including audio of bee buzzes, ambient noise, and cricket sounds [55]. The data is available from six different hives, which have been each separated into different disjoint train, test, and validation subsets. In the first partition (termed BUZZ1), the train and test samples are separated from the validation set by beehive and location. The second partition (BUZZ2), in turn, separates the train and test samples by beehive and location, while the validation sets are by beehive, location, time (2017 vs. 2018), and bee race (*A. Mellifera ligustica* vs. *A. mellifera carnica*). Authors in [55] created and used this datasets for bee detection using CNNs.

2.4.5.2 Code availability

Sharing code is an integral part of enabling repoducible research, as well as to allow future researchers to quickly and reliably build upon existing knowledge. Of the 57 reviewed articles, only two [48, 55] shared their code.

2.5 Summary and Open Challenges

This chapter presented a comprehensive and critical review of the experimental setups, signal processing pipelines, feature representations, machine learning methodologies, and evaluation practices employed in acoustics-based monitoring of honey bees and related pollinators. By systematically analyzing more than a decade of research, the chapter highlighted both the significant progress achieved and the persistent challenges that limit the deployment of robust and generalizable acoustic monitoring systems in real-world apicultural settings.

From an experimental perspective, the review revealed substantial heterogeneity across studies in terms of bee species, number of monitored hives, recording schedules, and sensing configurations. While most works focus on the western honey bee (*Apis mellifera*), species- and subspecies-level reporting remains inconsistent, despite mounting evidence that acoustic signatures are strongly species-dependent. Moreover, the predominance of small-scale experiments—often involving fewer than six hives—and the frequent omission of hive counts hinder meaningful cross-study comparison and limit conclusions regarding scalability and generalization. Seasonal and diurnal effects are widely acknowledged, yet recording protocols remain largely ad hoc, further contributing to variability across datasets.

At the signal acquisition level, key parameters such as sampling rate and microphone frequency response are often underreported. This lack of standardization is particularly problematic given the biological relevance of low-frequency components in bee acoustics. Similarly, audio pre-processing steps are inconsistently applied and rarely evaluated quantitatively, leaving open questions regarding the trade-offs between noise suppression, information loss, and computational efficiency.

From an analysis standpoint, the literature demonstrates a clear dominance of time–frequency representations, with MFCCs and spectrogram-based features serving as the backbone of most machine learning pipelines. While alternative representations—such as wavelet transforms, Hilbert–Huang transform, and constant-Q transform—have shown promise for specific tasks, feature selection remains highly task-dependent, and no single representation consistently outperforms others across applications. Classical machine learning models continue to be widely adopted due to their interpretability and modest data requirements, whereas deep learning approaches increasingly achieve state-of-the-art performance, particularly when leveraging transfer learning or self-supervised pre-training. Nonetheless, several studies indicate that deep models may struggle with

hive-independent generalization, underscoring the need for evaluation protocols that better reflect deployment conditions.

Performance evaluation practices further expose a lack of methodological consistency. Accuracy remains the most commonly reported metric, despite widespread class imbalance across bee acoustics tasks. More robust metrics such as AUC and F1-score are used inconsistently, and standardized evaluation frameworks are largely absent.

Finally, reproducibility emerges as a major bottleneck for progress in the field. Only a small fraction of studies rely on publicly available datasets, and even fewer share source code. Although notable datasets such as Nu-Hive, OSBH, and BUZZ have enabled limited benchmarking, differences in data collection protocols and partitioning strategies complicate fair comparison across methods.

Overall, this chapter underscores the need for more standardized experimental designs, transparent reporting of signal acquisition parameters, rigorous evaluation of preprocessing and feature extraction choices, and stronger emphasis on reproducibility and hive-independent generalization. These gaps directly motivate the methodological choices adopted in the subsequent chapters of this thesis, which aim to develop robust, interpretable, and scalable machine learning frameworks for acoustic monitoring of honey bee colonies under realistic and noisy conditions.

3 URBAN : URBAN BEEHIVE ACOUSTICS AND PHENOTYPING DATASET

3.1 Preamble

This chapter is compiled from material extracted from the following peer-reviewed article : "UrBAN : Urban Beehive Acoustics and PheNotyping Dataset [170]"

3.2 Introduction

Open access data on beehive management is crucial for fostering research output and advancements in the field, especially with the recent advances seen in artificial intelligence (AI) and machine learning (ML). Unfortunately, the beekeeping community faces a significant scarcity of open access data, hindering the progress of research and innovation. Access to comprehensive datasets on beehive dynamics, including factors such as colony health, behavior, and environmental influences, empowers researchers to conduct in-depth analyses and develop impactful solutions. While there are some public bee audio database available, such as Buzz [171], Nu-hive [172], and OSBH [61], they are mostly focused on bee and queen bee detection and are limited in sample size. The 'To bee or not to bee' dataset [61] is a combination of Nu-hive and OSBH audio samples, which were labeled to detect bee buzzing sounds.

Recently, the MSPB dataset was released [173]; this is a longitudinal multi-sensor dataset with phenotypic measurements. It was released to show that audio signals can be used for detection of winter survivability and population estimation. The dataset, however, does not provide access to the raw audio signals and makes available only pre-processed parameters extracted from the audio signal (e.g., overall hive audio power). Table 3.1 lists the available beehive acoustic datasets along with some relevant statistics. As can be seen, existing datasets are limited in the number of raw audio samples available, as well as in the number of days covered. Therefore, to address these limitations and support advancements in beehive monitoring, we present here a new beehive acoustics dataset.

In this chapter, we introduce UrBAN (Urban Beehive Acoustics and PheNotyping), a dataset that includes over 3000 hours of raw audio collected from beehives during a period of two years. The main focus of the data was on colony population prediction. The population of a colony can be estimated using the number of frames in the hive that are covered by least 70% [175] by bees. There have been studies showing that the number of frames of bees can be predicted using various features extracted from the raw audio within a machine learning framework [176, 173, 125]. Authors

TABLE 3.1 : Comparison of UrBAN dataset with the other public datasets.

Name	# Hives	Labels	Modality	Size	Availability
NU-Hive [67]	2	Queenright; queenless	Raw audio	$96\,\mathrm{h}$	Public
			Temperature	-	Not available
			Humidity	-	Not available
BUZZ [171]	6	Bee buzzing; cricket; noise	Raw audio	$7\,\mathrm{h}$	Upon requests
OSBH [169, 61]	6	Queenrigh; tqueenless	Raw audio	$140\,\mathrm{min}$	Subset available
Too bee or not to bee [61]	8	Bee buzzing; no bee buzzing	Raw audio	$12\,\mathrm{h}$	Public
MSPB [173]	53	Population; Honey; Queen conditions; Hygienic behavior; Winter mortality	Hand crafted audio features	Quarter-hourly during 365 day	Public
			Temperature		
			Humidity		
Smart Bee Colony Monitor [174]	4	Queenright; queenless	Raw audio	$118\,\mathrm{h}$	Public
			Temperature	Hourly during 40 day	
			Humidity		
			Pressure		
UrBAN (ours)	10	Population; Queen conditions; Winter mortality	Raw audio	$3171\,\mathrm{h}$	Public
			Temperature	Quarter-hourly during 135 day	
			Humidity		

in [173] showed that features, such as audio power in specific frequency bands, could be used to predict hive population.

The UrBAN dataset was gathered over the period spanning 2021 to 2022, originating from observations made across a network of ten beehives, part of an urban apiary, positioned in the rooftop of a building located in Montreal, Québec, Canada. Various parameters, such as audio recordings, temperature measurements, and relative humidity readings were collected. Notably, the dataset encompasses a broad spectrum of critical metrics pertaining to hive inspections, including assessments of honey bee colony population dynamics, evaluations of queen-related conditions, and records of overall health status. Specific health indicators, such as *Varroa* mite infestation rates and assessments of winter survivability, were also cataloged, providing invaluable insights into the factors influencing hive health and resilience.



(a) (b)

FIGURE 3.1 : **Photos of the (a) rooftop urban apiary in Montreal and (b) audio recording hardware.**

3.3 Methods

Urban Apiary

The rooftop apiary situated in Montréal, Canada (45.5253° N, 73.6123° W), comprised ten active honey bee hives. These hives were placed on wooden pallets (2 hives per pallet) in a row facing south east. Figure 3.1a shows the apiary placement. Each hive consisted of one brood chamber and one to two honey supers, housed within 10-frame standard Langstroth boxes, with a maximum of three boxes per hive. All hives originated from four-frame nucs, which were acquired and installed during the month of May in the year 2021.

Experienced beekeepers deliberately maintained varied hive populations to capture diverse data points, aligning with the primary objective of predicting beehive strength, which correlates with population size. At the beginning of the data collection, each hive contained a distinct number of (full) frames of bees with a minimum of six frames of bees in hives with a single brood box to a maximum of 20 frames in hives with both a brood box and a honey super. As colony populations expanded over time, additional honey supers were introduced. Consequently, within our apiary, the maximum configuration comprised three boxes and up to 30 frames of bees.

3.3.1 Hive Management and Inspection

The hives were manually inspected roughly every two weeks to measure the strength of the hives (i.e., the number of frames of bees), to verify the presence of a laying queen, as well as to

TABLE 3.2 : Summary of the data collection start and end dates for each year.

Year	Inspections	Varroa mite measurment	Audio recording	Temperature/Humidity
2021	June 22th - October 19th	-	August 11th - October 31st	June 19th - October 31st
2022	July 11th - September 7th	August 24th, September 1st, and September 30th	February 1st - October 31st	-

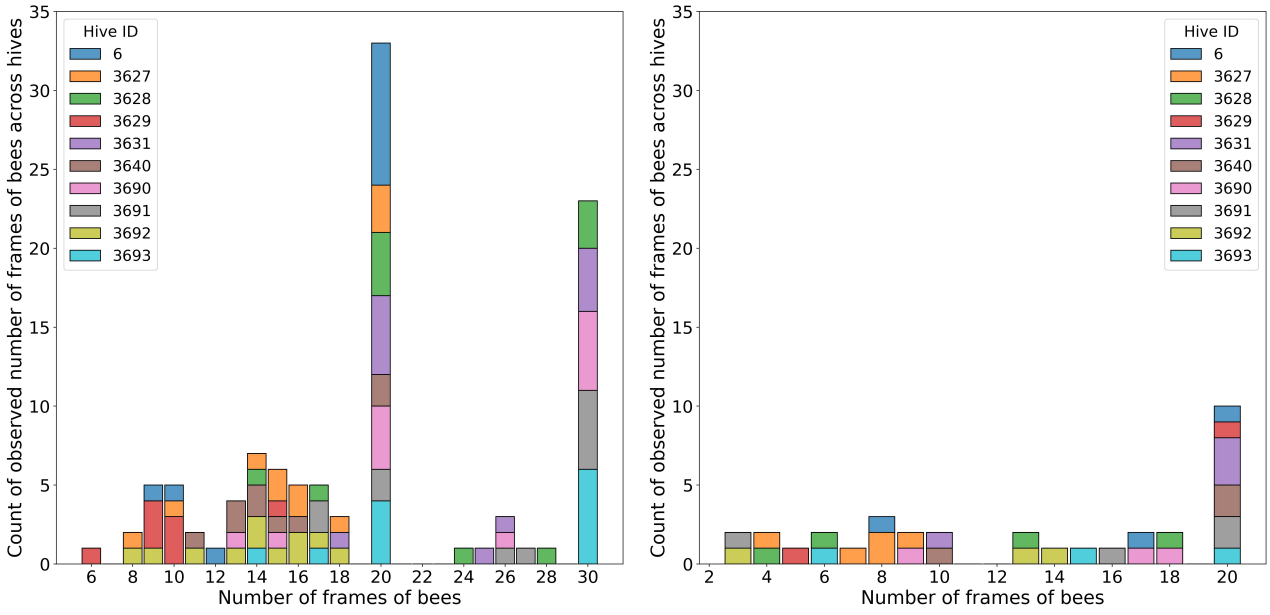


FIGURE 3.2 : Histograms of the number of frames of bees for the year of (a) 2021 and (b) 2022 experiments.

report any additional observations related to the colony activity. The start and date of these inspections are listed in Table 3.2 for each year. Figure 3.2 shows the histogram of the number of frames of bees observed during multiple inspections for each year of the experiments. Each histogram indicates the count of observed frames of bees for all of the beehives during the experiments. Moreover, Figure 3.3 shows the number of frames of bees for each hive during inspections. The beehive colonies were wintered outdoors with the aid of insulation, a strategic approach aimed at enhancing their survivability and well-being during the colder months, as shown in Figure 3.4. Insulation provides an additional layer of protection against harsh environmental conditions, helping to maintain stable temperatures within the hive and reduce heat loss.

3.3.2 Varroa Mite Infestation

Varroa mite infestation poses a significant threat to the health and well-being of honey bee colonies, making it a matter of utmost importance for beekeepers and researchers alike. In the

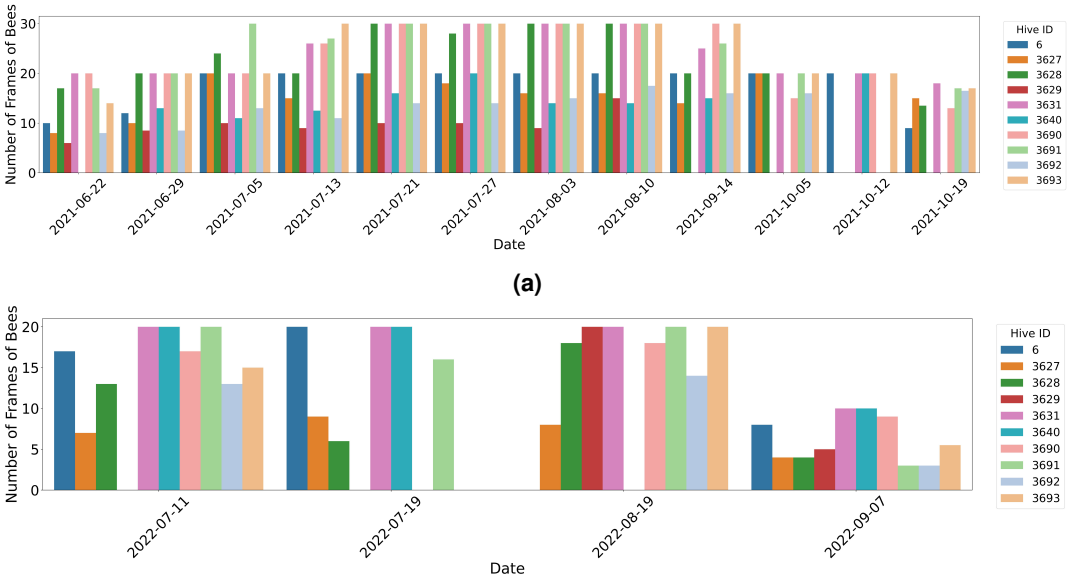


FIGURE 3.3 : Barplots of the number of frames of bees on days of inspections for the year of (a) 2021 and (b) 2022 experiments.



FIGURE 3.4 : Insulation used for beehives overwintering.

2022 data collection, the beekeepers measured the varroa mite infestation using the alcohol wash method [177] in some of the beehives. The alcohol wash method is a common technique for measuring Varroa mite infestations in honey bee colonies. It involves collecting a sample of about 300 bees from the brood nest, submerging them in isopropyl alcohol (70% or higher), and gently shaking the container to dislodge the mites. The mixture is then strained, and the mites are counted to estimate the infestation rate. This method, while resulting in the loss of the sampled bees, provides an accurate assessment of mite levels, helping beekeepers make informed decisions about mite control and ensuring colony health. Figure 3.5 indicates the amount of Varroa mite infestation in each measurement.

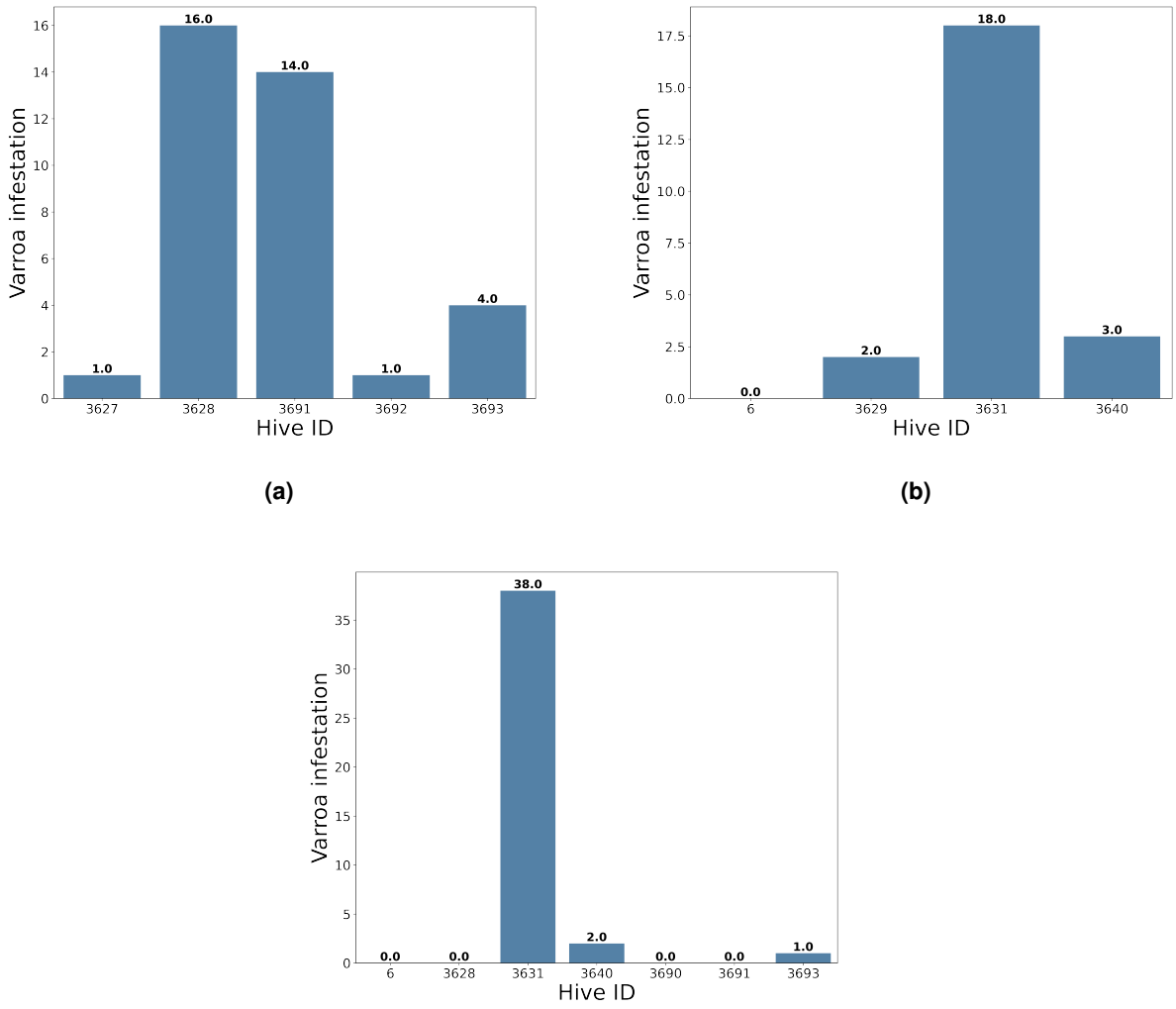


FIGURE 3.5 : Varroa mite infestation on (a) August 24th, (b) September 1st, and (c) September 30th, 2022.

3.3.3 Mortality Rate

Winter mortality is one of the primary causes of beehive losses globally and is a significant concern for both beekeepers and researchers [178]. As temperatures drop and resources become scarce, honey bee colonies face numerous challenges that can impact their survival. Factors such as disease prevalence, mite infestations, inadequate food stores, and harsh environmental conditions all contribute to increased mortality rates among bee colonies during the winter [179]. Understanding and monitoring these mortality rates is crucial for assessing the health of bee populations and implementing strategies to mitigate losses. During the data collection, 20% of the beehives died after overwintering in 2022.



(a) (b)

FIGURE 3.6 : Location of the (a) temperature and humidity sensor placed on top of the middle frame of the first brood chamber and (b) microphone placed next to it.

3.3.4 Sensor Data

A multimodal sensor, positioned atop the central frame within the bottom brood box, facilitated continuous monitoring of internal hive temperature and humidity (SHTC1, Beecon, Nectar Technologies Inc, Canada). This sensor operates within a relative humidity range of 0% to 100% RH, with an accuracy of ±3% RH. Its temperature range spans from -40℃ to +125℃, offering a measurement accuracy of ±0.3℃. Additionally, an accompanying microphone, depicted in Figure 3.6b, was installed adjacent to the sensor. The microphones were installed inside a custom-designed enclosure, which served to minimize direct contact with the bees and reduce the likelihood of propolization. The setup utilized a MEMS microphone (SPK0641HT4H-1) with a frequency range of $20\,\mathrm{Hz}$ to $20\,\mathrm{kHz}$. This omnidirectional PDM microphone operates at a voltage range of 1.6 V to 3.6 V and has a sensitivity of -26 dB ±1 dB @ 94 dB SPL. The multi-modal data is comprised of the average temperature and humidity readings every 15 minutes, and a 15-minute audio segment every 30 minutes with a sampling rate of $48\,\mathrm{kHz}$. To minimize storage usage, every audio file was downsampled to $16\,\mathrm{kHz}$.

Moreover, local external temperatures, humidity, and rainfall amounts levels were obtained from the Environment and Climate Change Canada website [1]. A representative example of a $24\,\mathrm{h}$ snapshot of the changes in internal/external temperature and humidity levels for a single beehive, as well as a 4-month period average of all beehives is shown in Figure 3.7a and b, respectively. The $24\,\mathrm{h}$ snapshot (3.7a) is for a strong and healthy colony in August with one brood chamber and 2 honey supers with a total of 30 frames of bees (covered with at least 70% of bees).

1. https://climate.weather.gc.ca/

TABLE 3.3 : The quantity of raw audio for each year of the experiments.

Year	Type	Total quantity - size	Quantity during months of inspections
2021	Raw audio	1752 h - 185 Gb	1466 h - 155 Gb
2022	Raw audio	12491 h - 1315 Gb	1705 h - 180 Gb

Figure 3.8 illustrates the intensity of the audio, the smoothed root mean square (RMS) power, and its corresponding spectrogram. It is evident from the figure that the audio power experiences a rise throughout the day, particularly during periods characterized by rising external temperatures and declining humidity levels. This observation suggests increased foraging activity and thermohygrometric regulation within the colony.

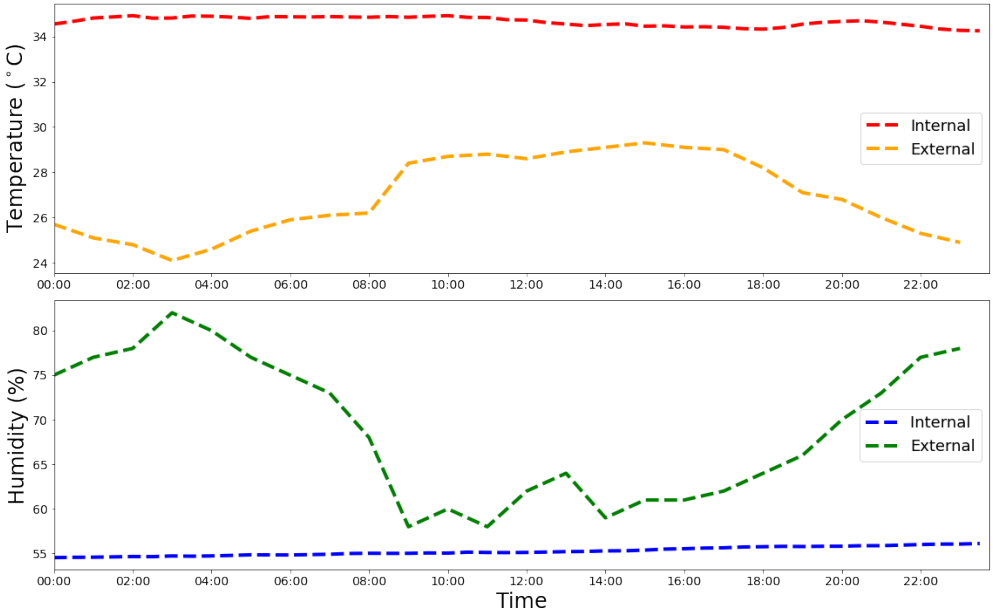
On the importance of beehive size and its effect on audio, Figure 3.9 shows the $24\,\mathrm{h}$ bar-plots for different number of frames of bees. Each of these plots show the average RMS value with specified frames of bees. Similar to Figure 3.8, a rising trend during the day can be seen. Table 3.3 provides an overview of the quantity and size of raw audio collected for each year. It consists of two columns detailing the total duration in hour and size in gigabytes (GB) of recordings, with one column encompassing all recordings regardless of inspection periods, and the other focusing solely on recordings made during the inspection periods specified in Table 3.2.

3.4 Data Records

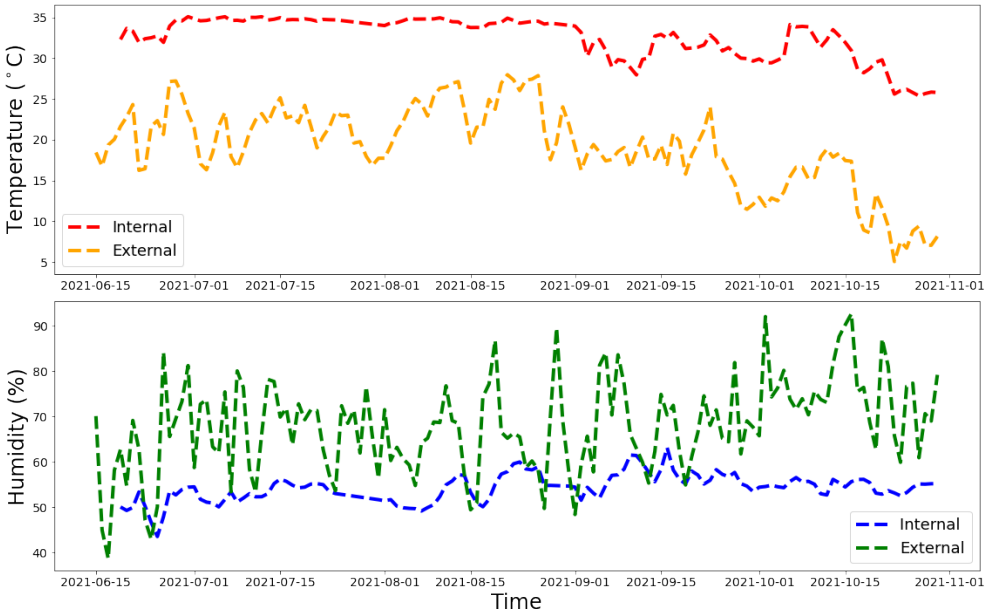
Tables 3.4, 3.5, and 3.6 provide a comprehensive overview of the inspections (labels) and sensor data for years 2021-2022. The inspection files called `inspections_2021.csv` and `inspections_2022.csv` contain details such as the count of frames of bees, presence of varroa mite infestation, queen status, and mortality rates stored as `.csv` files. Each beehive is identified by a unique hive ID, enabling discrimination between them. In Table 3.6, the details about recording of internal temperature, humidity, raw audio, and also weather information such as external temperature, humidity, and amount of precipitation are described. The raw audio recordings are `wav` files stored in a compressed format for easy download. The sensor data and weather information comprise two `.csv` files. The UrBAN dataset is made fully publicly available at the Federated Research Data Repository [180] (`https://doi.org/10.20383/103.0972`).

3.5 Summary

In this chapter, we introduced the UrBAN dataset [180], a large-scale, open-access audio dataset designed to support research in acoustic-based beehive monitoring. The dataset comprises more than 3,000 hours of raw audio recordings collected over two consecutive years from ten urban beehives, alongside external weather measurements. In addition to continuous sensor data,



(a)



(b)

FIGURE 3.7 : Internal and external temperature and humidity during (a) a $24\,\mathrm{h}$ period for a single beehive on August 12th, 2021, and b) total duration of the experiment in 2021.

the dataset includes detailed hive inspection records capturing colony population dynamics, queen status, Varroa mite infestation levels, and winter mortality outcomes.

Compared to existing public beehive acoustic datasets, UrBAN offers substantially greater temporal coverage, access to raw audio signals, and richer phenotypic annotations, enabling both short-term and longitudinal analyses. By making this dataset publicly available, we aim to facilitate reproducible research, support the development of generalizable machine learning models,

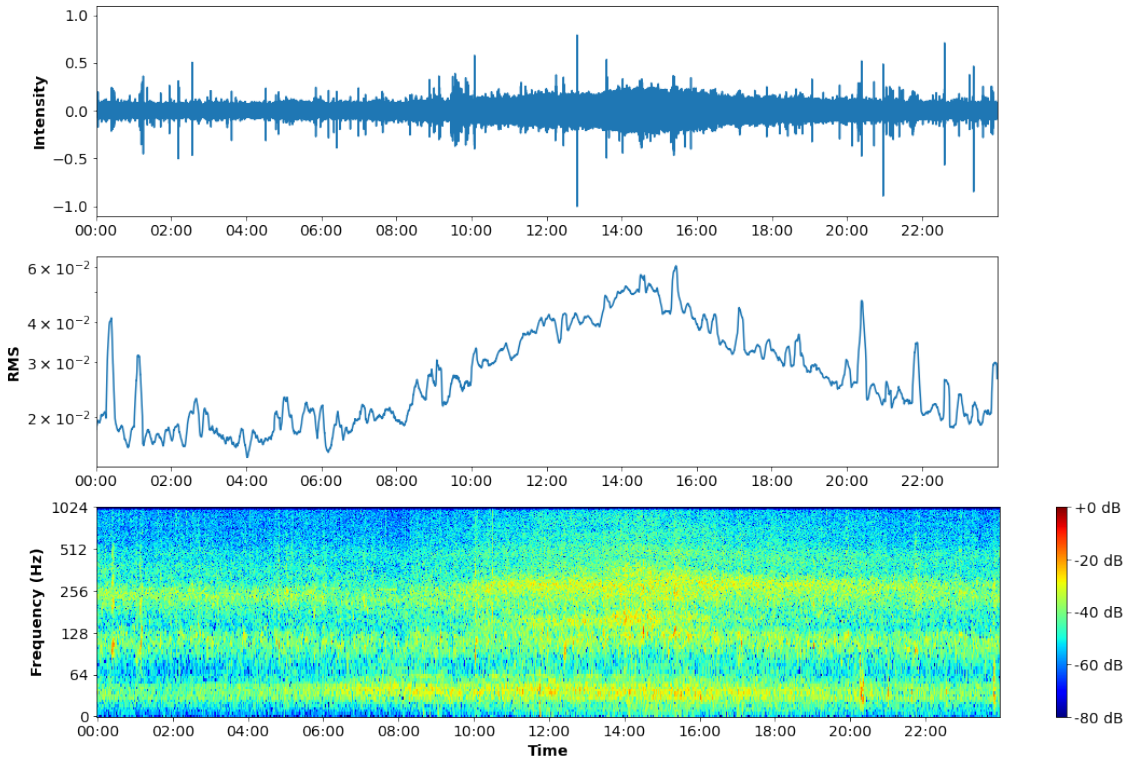


FIGURE 3.8 : Audio intensity, audio RMS, and corresponding spectrogram for a hive with with 3 full boxes of bees (August 12, 2021).

and advance data-driven approaches for monitoring honey bee colony health and resilience in real-world environments.

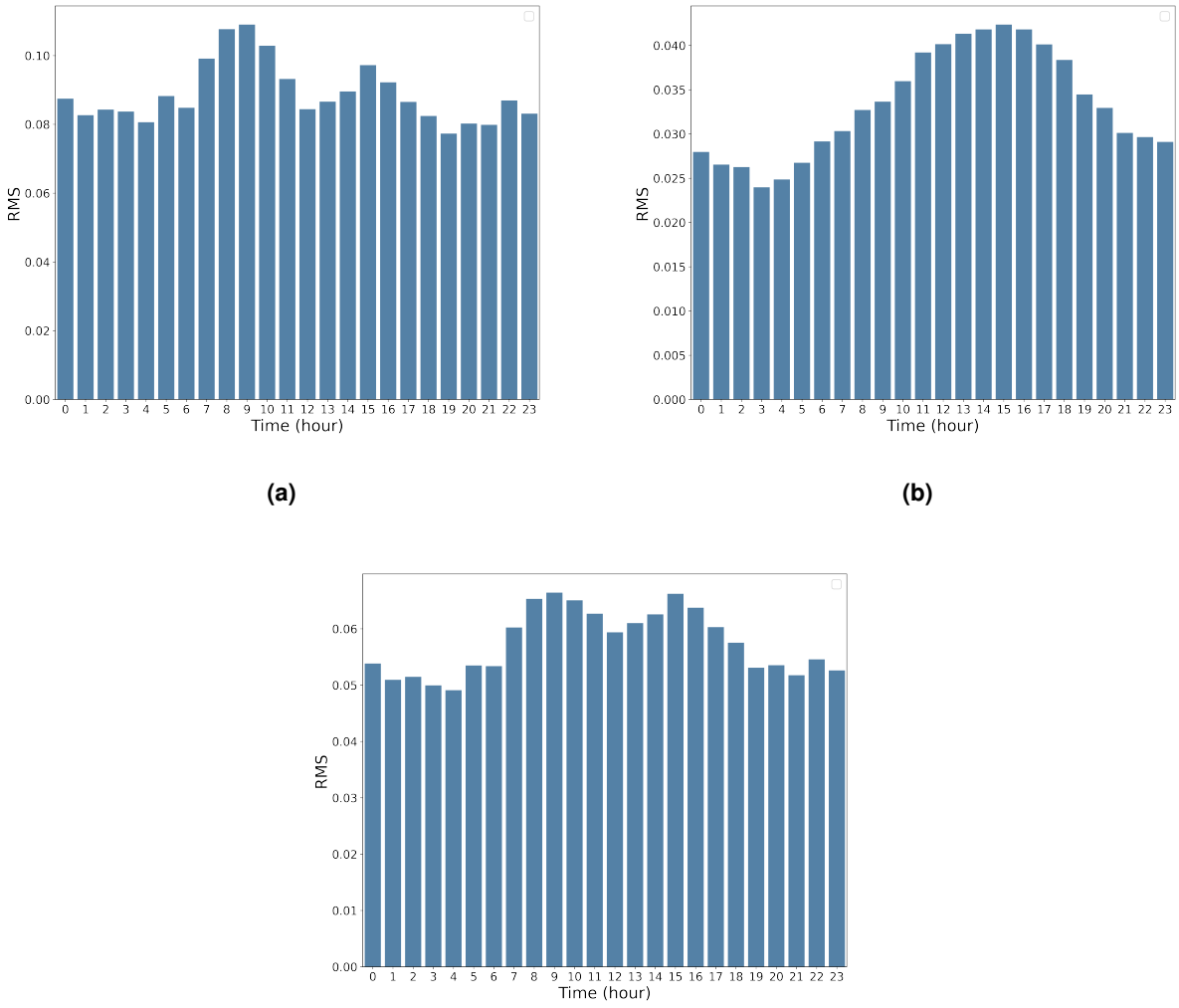


FIGURE 3.9 : 24h average of RMS value when the number frames of bees is between (a) 10 and 20, (b) 20 and 30, and (c) 1 and 30.

TABLE 3.4 : Structure of the files describing inspections for the year 2021.

Year	File name	Columns	Description
2021	`inspections_2021.csv`	Date	Time stamp (YYYY-MM-DD)
		Tag number	ID unique to each hive
		Colony size	The number of boxes for each beehive
		Fob 1st	The number of frames of bees in the first box
		Fob 2nd	The number of frames of bees in the second box
		Fob 3rd	The number of frames of bees in the third box
		FoBrood	The number of frames of brood
		Frames of honey	The number of frames of honey
		Queen status	QR/QNS (queen seen or not seen)
		Open	Time stamp indicating opening the box for inspections (HH :MM)
		Close	Time stamp indicating closing the box after inspections (HH :MM)
		Note	Additional observation such as beehive being weak or aggressive

TABLE 3.5 : Structure of the files describing inspections for each year 2022.

Year	File name	Columns	Description
2022	`inspections_2022.csv`	Date	Time stamp (YYYY-MM-DD HH :MM :SS)
		Tag number	ID unique to each hive
		Category	Hive grading, hive status, frames of bees, varroa, treatment, feeding, custom practice, queen management, hive issues
		Action detail	Detail of each category. Hive grading : 'strong', 'medium', 'weak', 'pulled honey super', 'size - 1d'; Hive status : 'queenright', 'queenless', 'deadout'; frames of bees : the number of frames of bees; Varroa : the varroa mite measurement; Treatment : 'mite away'; Feeding : 'sugar'; Custom practice : 'add entrance reducer', 'supering', 'added bee escape', 'added trash bag (feeder trick)'; Queen management : 'potential breeder'; Hive issues : 'chalk brood'
		Queen status	Queenright/queenless
		Is alive	0/1 (Zero indicates a dead hive)
		Report notes	Additional observation such as beehive being weak or aggressive

TABLE 3.6 : Structure of the files/folders of the sensor data (temperature and humidity), raw audio recordings, and weather information.

Year	File/folder name	Columns	Description
2021	`sensor_2021.csv`	Date	Time stamp (YYYY-MM-DD HH :MM :SS)
		Tag number	ID unique to each hive
		Temperature	Internal temperature in degree Celsius
		Humidity	Internal humidity in percentage
	`audio_2021`	-	Audio files names : `"DD-MM-YYYY_HHhMM_HIVE_Tag.wav"` where 'Tag' is the hive ID number.
2022	`audio_2022`	-	Audio files names : `"DD-MM-YYYY_HHhMM_HIVE_Tag.wav"` where 'Tag' is the hive ID number.
2021 - 2022	`weather_2021_2022.csv`	Date/Time (LST)	Time stamp (YYYY-MM-DD HH)
		Temp (°C)	External temperature in degree Celsius
		Rel Hum (%)	External humidity in percentage
		Wind Spd (km/h)	The speed of wind
		Precip. Amount (mm)	The amount of precipitation

4 IMPACT OF ENVIRONMENTAL NOISE AND PREPROCESSING

4.1 Preamble

This chapter is compiled from material extracted from the following peer-reviewed articles :

1. "The Importance of Context Awareness in Acoustics-Based Automated Beehive Monitoring [176]" (section 4.2).
2. "Performance Comparison of Voice Activity Detectors for Acoustic Beehive Monitoring [181]" (section 4.3).
3. "UrBAN : Urban Beehive Acoustics and PheNotyping Dataset [170]" (section 4.4).

4.2 Environmental Noise Factors

In the previous chapter, we introduced the UrBAN dataset, a large-scale, long-term collection of beehive acoustic recordings acquired under real-world conditions. The dataset captures substantial variability arising from environmental factors, recording conditions, and hive-specific characteristics, making it well suited for studying the robustness and generalizability of acoustic monitoring methods.

Despite the rapid growth of acoustics-based beehive monitoring applications, as reviewed in Chapter 2, most existing systems continue to rely on conventional audio features originally developed for speech and general audio processing. These include root mean square power (RMS), mel-frequency cepstral coefficients (MFCCs) [156], time–frequency representations such as the spectrogram, and features derived from the discrete wavelet transform (DWT). While these features have proven effective in controlled conditions, it is well established in the audio signal processing literature that they are highly sensitive to environmental disturbances, including ambient noise, background speech, and other non-hive acoustic events.

Consequently, noise suppression strategies and/or explicit characterization of background acoustic conditions—commonly referred to as context-awareness—are essential for developing accurate, robust, and replicable monitoring systems (e.g., [42, 43]). This challenge is particularly pronounced for urban beehives, which are becoming increasingly prevalent [44]. In such environments, diverse and often high-energy urban sounds (e.g., traffic, construction activity, human speech) can significantly interfere with internal hive recordings, thereby degrading feature reliability and model performance.

Here, we systematically analyze several prominent environmental noise sources commonly encountered in real-world beehive monitoring scenarios and quantify their impact on widely used

benchmark acoustic features. By evaluating how different noise conditions affect feature robustness, this analysis highlights the limitations of conventional representations and motivates the need for more noise-resilient and context-aware feature extraction strategies, which are explored in subsequent chapters.

4.2.1 Acoustic Feature Extraction

In this section, we describe the four features that have been commonly reported to be used for audio-based precision beekeeping, namely : RMS, spectrogram, MFCCs and DWT. Before feature extraction, audio recordings are resampled from 48 kHz to 2048 Hz, as the literature suggests that the majority of the bee sounds are produced below 1000 Hz [52].

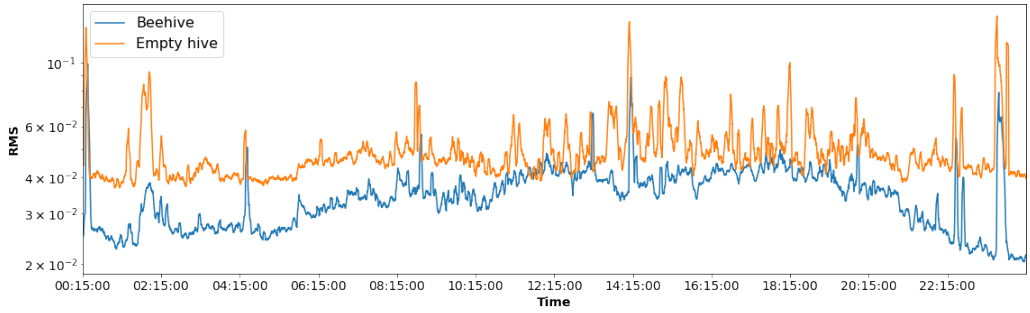
4.2.1.1 Root Mean Square Power

The moving RMS power is calculated based on equation 4.1,

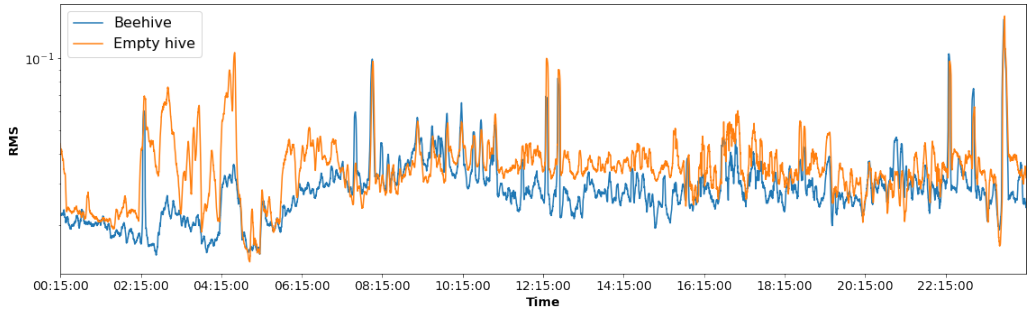
$$y_{RMS}(i) = \sqrt{\frac{1}{N}\sum_{k=i}^{i+N} x(k)^2}, \qquad i = 0, 1, 2, ..., L - N - 1 \tag{4.1}$$

where L is the signal length and N the windowed segment length. In the speech processing community, such windowed segment is referred to as a 'frame'. We use the same notation here, but not to be confused with the frames within the beehive. As such, the term signal frame will be used henceforth. In this experiment, we set the signal frame length to 1 second and use 50% overlapping shifts.

Having access to the acoustics of an empty hive allows us to identify certain environmental factors that would be present in both the empty and occupied hives. As shown in Figure 4.1a, the empty hive has a fairly constant value with noise-related fluctuations around this value, whereas the occupied hive has an increase in audio power during the day, which stabilizes around the afternoon and later drops in the evening. Close comparison between the two curves, however, shows events that are common in both boxes, thus signalling environmental factors that could affect the monitoring system. For example, after close investigation it was found that the abrupt peaks seen in both curves were train rail squealing from a passenger train that ran behind the building (1 km distance) in which the apiary was located or car horns from nearby traffic. The more slower peaks, in turn, were related to beekeepers voices being recorded by the microphone when they spoke near the hive. Another environmental factor that appeared throughout the recordings were the noises of raindrops hitting the hive boxes on rainy days. A segment of data showing this factor is shown in Fig. 4.1b. Given these three prominent environment noises, this paper investigated the



(a)



(b)

FIGURE 4.1 : RMS curves of a typical colony with three full boxes of bees (blue curve) and an empty box (orange) as measured on (a) August 12th, 2021, and (b) September 24th, 2021.

impact that they have on three other popular audio features and an automated monitoring system based on them.

4.2.1.2 Spectrogram

The spectrogram is a time-frequency representation of the audio signal that has been widely used for swarming and queen absence detection [49, 88]. The spectrogram is commonly measured with a short time Fourier transform (STFT), where the signal is divided to equal-length segments by a moving window and then each segment is transformed using the fast Fourier transform (FFT). In this study, a Hann window with a length of 125 ms with a 50% overlap was used. Figure 5.18 shows the spectrogram of a 24-hour recording of the beehive previously shown in Figure 4.1a. In the top figure, the $\times's$ depict the instances in which environmental noises were detected. From the bottom spectrogram plot, their influence on the spectrogram can be seen as sudden bursts of energy below 64 Hz, thus coinciding with important hive sounds that were present throughout the day.

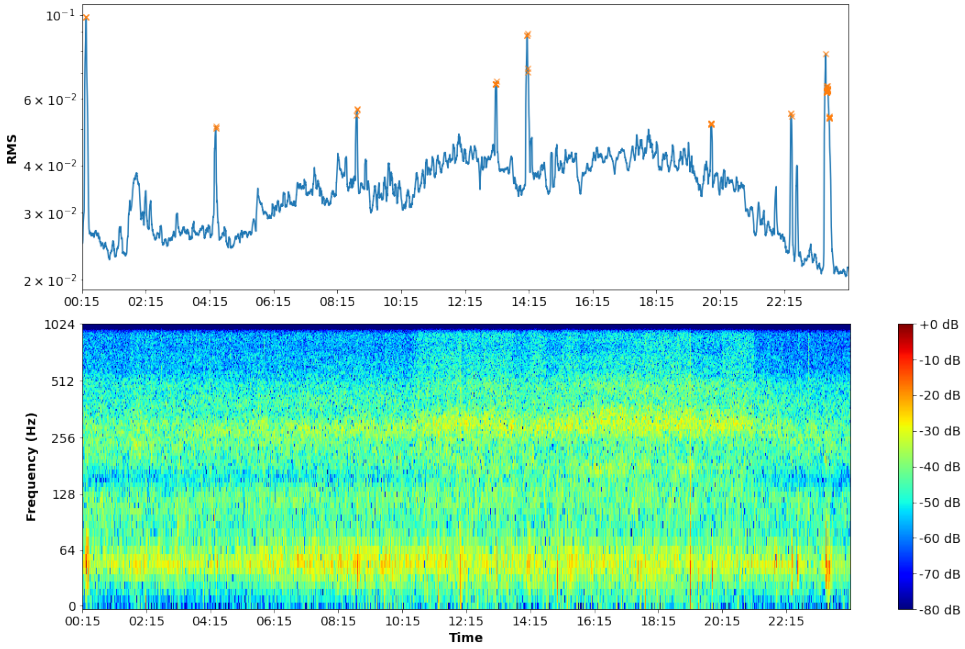


FIGURE 4.2 : RMS curve and corresponding spectrogram for a hive with 3 full boxes of bees (August 12, 2021).

4.2.1.3 Mel-Frequency Cepstral Coefficients

MFCCs have been widely used in many speech applications as they model the human cochlear processing steps by including a mel-scale frequency mapping prior to cepstrum processing [156]. For precision beekeeping, these features have also shown to be popular, appearing in roughly 30% of the papers reviewed in [52]. For example, the first three MFCC coefficients have been widely used for bee, queen presence, and swarming detection. To calculate MFCCs, the five following steps are taken :

1. Apply a pre-emphasis filter to enhance high frequencies.
2. Compute the STFT of pre-emphasized signal and its power spectrum.
3. Apply the mel filterbank composed of triangular filters simulating cochlear processing.
4. Apply the logarithm operation.
5. Compute the discrete cosine transform (DCT) to extract the mel frequency cepstral coefficients.

In this experiment, a Hann window with a length of 125 ms with a 50% overlap is used in step 2. A total of 12 coefficients are extracted in step 5, plus the zeroth coefficient which is used as a measure of signal power. Figure 4.3 shows the 12 MFCCs, plus the zeroth coefficient, as computed over the 24h duration of the previous audio signal. In Section 4.2.4, more details will be provided on the effects of ambient noise on the MFCCs.

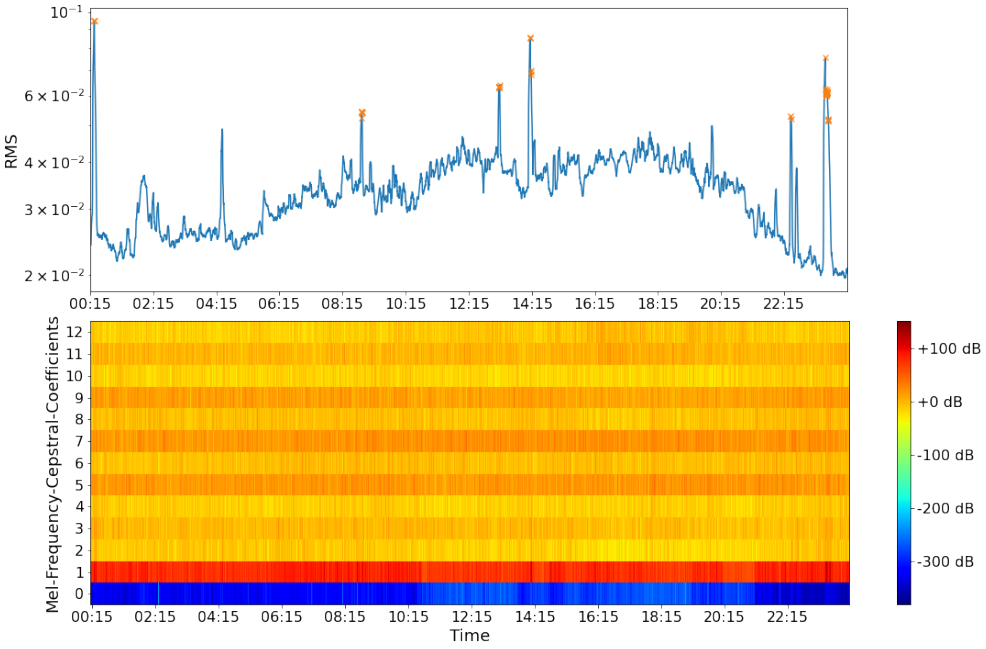


FIGURE 4.3 : RMS curve and corresponding MFCCs for a hive with with 3 full boxes of bees (August 12, 2021).

4.2.1.4 Discrete Wavelet Transform

Like the spectrogram, the DWT is a time-frequency representation that is widely used in time series analysis and has shown promising results for beehive acoustics monitoring applications [65, 49]. DWTs rely on so-called wavelets to represent signals as a superposition of short finite-length waves that can be scaled and shifted to capture changes in time–frequency space. The DWT is the discrete form of continuous wavelet transform (CWT) described by equation 4.2 :

$$C(a,b) = \int_{-\infty}^{\infty} x(t)\psi^*_{a,b}(t)dt, \tag{4.2}$$

where $x(t)$ is the audio signal, ψ^* is the complex conjugate of the mother wavelet ψ and a, b the translation and scale, respectively. The mother wavelet is given by :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{b}}\psi(\frac{t-a}{b}). \tag{4.3}$$

In the DWT, the translation and scale values have the following conditions :

$$b = 2^{-s} \qquad a = 2^{-s}l, \tag{4.4}$$

where l and s are discrete values.

Combining these conditions with the assumption of discretization of the signal results in the DWT shown in equation 4.5 :

$$W(l,s) = 2^{\frac{s}{2}} \sum_{n} x(n)\psi(2^s n - 1); \qquad n = 0, 1, ..., L. \tag{4.5}$$

In this study, the audio is decomposed to an “approximation” and a “detail” component using a filtering implementation of the DWT [182]. In this case, a low-pass (G) and a high-pass (H) filter are used to separate the approximate (A(k)) and detail (D(k)) components of the signal, respectively :

$$A(k) = \sum_{n} x(n)h(2k - n), \text{and} \tag{4.6}$$

$$D(k) = \sum_{n} x(n)g(2k - n). \tag{4.7}$$

Here, we used the second order Daubechies wavelet [183] (’db2’) and 10 levels of decomposition based on combinations reported in the literature.

With these common features extracted for all hives during three months, we use the empty hive and rainfall information to detect environmental sounds of interest and explore the effects they have on these features.

4.2.2 Beehive Strength Prediction Model

To quantify the impact of the three environmental factors on beehive monitoring systems, three hive strength prediction models are developed, one based on each of the three feature sets described in Section 4.2.1. The goal of this monitoring system is to predict the number of frames of bees at any given point in time, thus serving as an indicator of the hive strength. In our case, a maximum of three boxes were included per hive, each box containing 10 frames. As such, the model predicts a value between 0 and 30, where zero indicates an empty hive and 30 indicates that all three boxes are full.

Following recent insights from chapter 2, a state-of-the-art convolutional deep neural network classifier was used. The three different two-dimensional time-frequency representations served as input to the classifier. Table 4.1 shows the network specifications. Each convolution layer has L2 regularization with a weight decay of 0.001. All layers use a Rectified Linear Unit (ReLU) as an activation function, except the output layer which uses a linear function. The best learning rate was found to be 0.001 and the batch size was set to 128. Lastly, mean-absolute-error (MAE) was used as the loss function to train the network.

TABLE 4.1 : Proposed convolutional neural network architecture.

Layer	Description
Convolution	128 filters, kernel size = 3×3 , strides = 1×1
Convolution	128 filters, kernel size = 3×3 , strides = 1×1
Batch normalization	momentum = 0.99, gamma = 1.0, epsilon = 0.001
Max-pooling	pool size = 3×3, strides= 2×2
Convolution	64 filters, kernel size = 3×3 , strides = 1×1
Batch normalization	momentum = 0.99, gamma = 1.0, epsilon = 0.001
Convolution	64 filters, kernel size = 3×3 , strides = 1×1
Batch normalization	momentum = 0.99, gamma = 1.0, epsilon = 0.001
Max-pooling	pool size = 3×3 , strides = 2×2
Dropout	0.25 rate
Dense	128 units
Batch normalization	momentum = 0.99, gamma = 1.0, epsilon = 0.001
Dropout	0.5 rate
Dense	1 unit

4.2.3 Experiment Setup and Figures-of-Merit

Audio data was recorded in 15-minute segments. These are divided into multiple 1-second signal frames, which were then used for feature extraction that served as input to the classifier. Following suggestions from [48, 49], two different experimental setups are considered : “random-split” and "hive-independent". In the former, the entire dataset from the 10 hives is randomly divided into three parts : 25% is for testing, 20% for validation and the rest is reserved for training. In "hive-independent" scenario, data from seven hives was used to train the models (of which 30% of the training data was set aside for validation), while the remaining three unseen hives were used for testing.

For model evaluation, two figures-of-merit are used : mean-absolute-error (MAE) and the root-mean-square-error (RMSE) obtained from the estimates and computed as follows, respectively, for a sample size of M :

$$MAE = \frac{\sum_{n=1}^{M} |y_i - \hat{y}_i|}{M}, \tag{4.8}$$

$$RMSE = \sqrt{\frac{\sum_{n=1}^{M} (y_i - \hat{y}_i)^2}{M}}, \tag{4.9}$$

where y is the observed value and $\hat{y}$ the prediction.

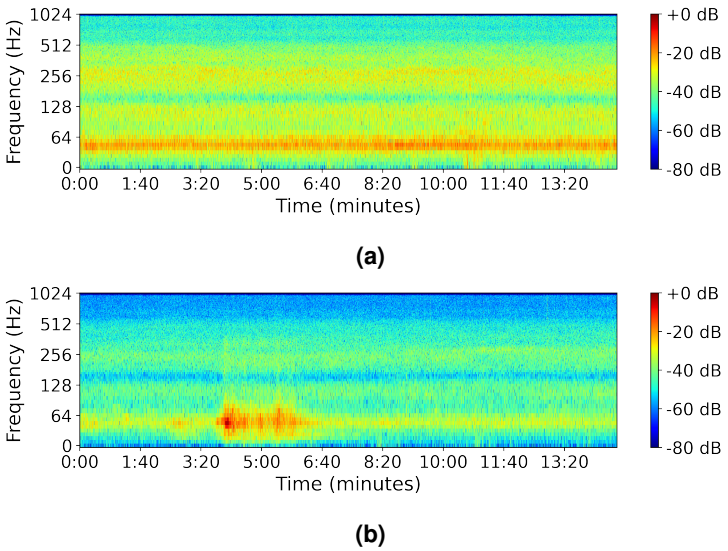


FIGURE 4.4 : Spectrograms of cases where urban noises (e.g., train rail squealing sounds) are (a) not present and (b) present.

4.2.4 Results and Discussion

4.2.4.1 Urban Sound Effects

As urban apiaries burgeon, common ambient sounds, such as car horns and train rail squealing, will become nuances for audio-based hive monitoring applications. As such, it is crucial to understand the impact that these sounds can have on the extracted audio features. For our study, sounds of a nearby train passing by was a common noise that contaminated the audio recordings. Figure 4.4, for example, shows the spectrogram measured from a 15-minute audio recording when the train was not passing by (subplot a), as well as when the train passed by (subplot b). As can be seen, the train rail squealing caused a significant disturbance in the audio around 64 Hz and higher frequencies such as 250 Hz which is crucial for beehive dynamics.

Next, we explore the effects that such urban sounds may have on the MFCCs. Figure 4.5 shows the histograms of the distribution of the first two coefficients (subplots and and b, respectively) across time instances in which the train sound was present versus absent. As can be seen, the train rail squealing sound has a substantial effect on the distribution of the MFCCs; in the case of the first coefficient, it increases the value by around 50 db. Such effects may result in erroneous decisions by automated systems that rely on the first three mel-frequency cepstral coefficients, such as the systems described in [16, 48, 49]. Figure 4.6 depicts the temporal changes seen in the 1st MFCC when the train passes by (see subplot b around the 5-minute time). For systems that rely on average MFCCs over time, such increases will likely result in errors.

Lastly, Fig. 4.7 (a) and (b) depict the DWT detail (D1-D10) and approximation (A10) components extracted from the clean and noisy segments, respectively. As can be seen, the detail coefficients



(a) (b)

FIGURE 4.5 : Comparison of the changes in distribution of the (a) first and (b) second mel frequency cepstral coefficient with and without the presence of train rail squealing sounds.

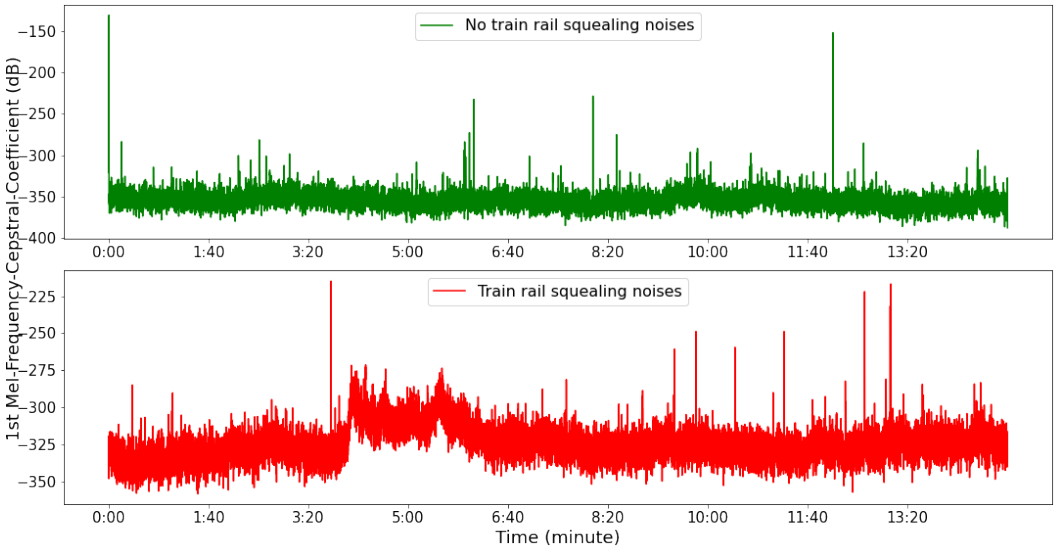


FIGURE 4.6 : First mel frequency cepstral coefficient time series when there are no trains nearby (top) versus when the train passes by (bottom) between the 4 :00 - 6 :40 time range.

are particularly affected by this urban noise source, and would likely result in errors for systems that rely on DWTs, such as the ones described in [65, 49].

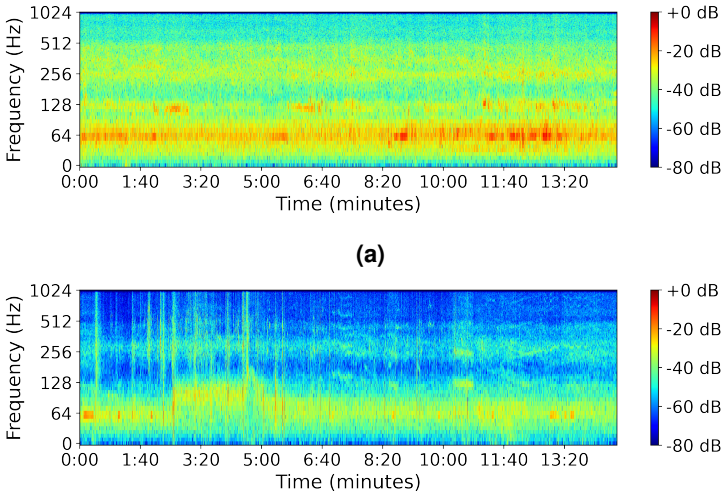
4.2.4.2 Speech Artifacts

Speech of people talking near the beehives is another common noise source when it comes to audio-based monitoring systems. Male speech, for example, is known to have fundamental frequencies ranging from 80-150 Hz, whereas female speech from 160-250 Hz. These are typical ranges in which beehive acoustics are monitored at, thus system performance could be severely compromised. The spectrograms in Figure 4.8b, for example, show scenarios in which speech is not present (plot a) near a particular hive and when beekeepers were talking near the same hive



(a) (b)

FIGURE 4.7 : Comparison of the changes in distribution of the DWT detail and approximation components (a) without and (b) with the presence of train rail squealing sounds.



(a)

(b)

FIGURE 4.8 : Spectrogram comparisons when human speech is (a) not present and (b) present near the hives.

(plot b). Next, the histogram of the distribution of the 1st and 3rd MFCCs shown in Figure 4.9 shows the impact due to speech artifacts. Similar changes in the DWT detail and approximation components can be seen in Figure 4.10.

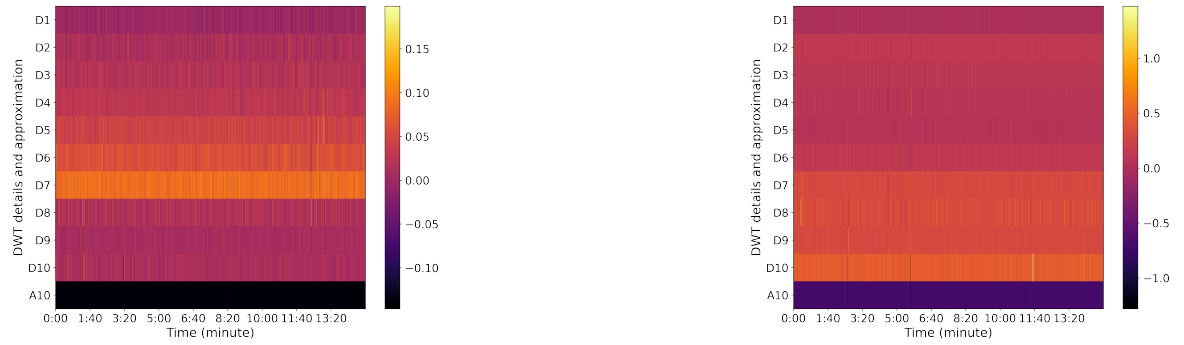
4.2.4.3 Acoustic Effects of Heavy Rain

Lastly, the sound of heavy rainfall hitting the hive boxes can also cause sufficient changes in the recorded audio to make monitoring systems fail. In our experiment, we relied on weather information to pinpoint periods of the day in which precipitation amounts greater than 9 mm were present. Figure 4.11 shows a comparison between the spectrograms of a hive during a typical night (plot a) and the same hive under heavy rainfall (plot b). As can be seen, rainfall sounds can severely affect multiple frequency ranges of the hive, increasing the recorded sound in some



(a) (b)

FIGURE 4.9 : Comparison of the changes in distribution of the (a) first and (b) third mel frequency cepstral coefficient with and without the presence of human speech.



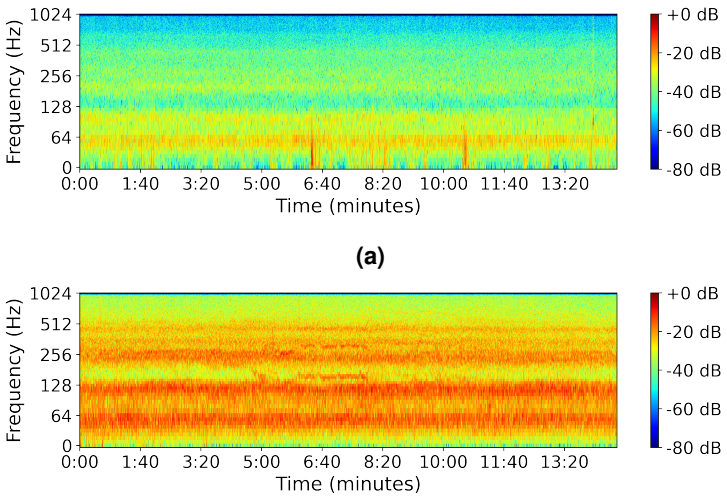
(a) (b)

FIGURE 4.10 : DWT detail and approximation component comparisons when human speech is (a) not present and (b) present near the hives.

bands by as much as 10 dB, thus is likely to deteriorate monitoring performance. The histogram of the distribution of the 1st and 2nd MFCC shown in Figure 4.12 shows the impact that rainfall can have on these parameters. Lastly, Figure 4.13 shows the changes in the DWT components with and without rain artifacts. Unlike the urban and speech effects, which were transient and lasted a few seconds, rain sound can be much longer, potentially contaminating several hours of data, thus severely affecting automated monitoring system.

4.2.5 Prediction Model Performance

Table 4.2 presents the prediction model performance when only clean test samples are available, as well as when the noisy samples are included. As can be seen, in the “random-split” scenario, the system based on MFCCs outperformed the two other systems and achieved an MAE of 0.86 for the clean samples. When the environment factors were included, the DWT-based system was the most affected. Overall, MFCC-based systems achieved lower MAE under urban



(a)

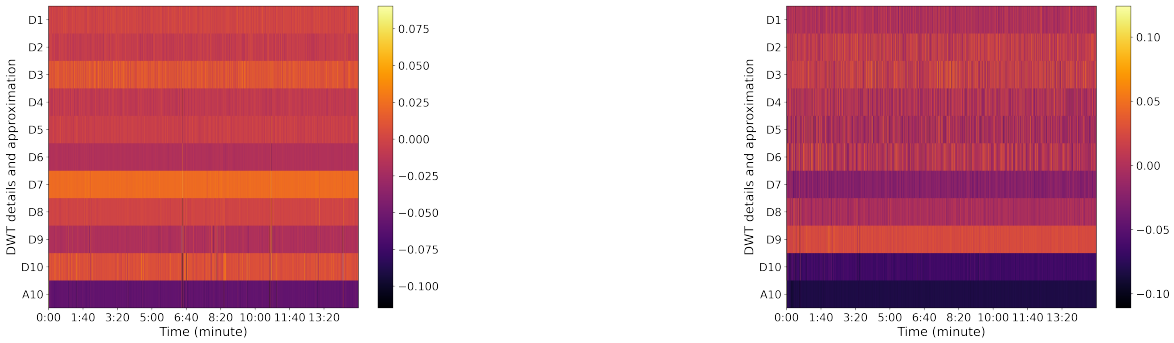
(b)

FIGURE 4.11 : Spectrogram comparisons (a) without and (b) with heavy rainfalls.



(a) (b)

FIGURE 4.12 : Comparison of the changes in distribution of the (a) first and (b) second mel frequency cepstral coefficient with and without the rain.



(a) (b)

FIGURE 4.13 : DWT component comparisons (a) without and (b) with heavy rainfalls.

TABLE 4.2 : Performance comparison (MAE and RMSE) for beehive monitoring systems based on three feature sets under random-split and hive-independent testing scenarios.

Features	Random-Split							
	Clean		Urban sound		Speech artifacts		Heavy rain	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Spectrogram	1.03	1.93	1.46	2.13	**1.23**	2.43	2.88	4.53
MFCCs	**0.86**	2.34	**1.17**	2.72	1.86	3.93	**1.00**	2.52
DWT	1.42	4.43	4.84	6.52	1.92	6.12	6.47	9.01
	Hive-Independent							
Spectrogram	**4.28**	5.17	**6.37**	6.97	**7.98**	8.02	7.20	7.39
MFCCs	4.91	5.73	6.93	7.86	8.77	9.23	**6.95**	7.75
DWT	4.94	5.33	6.62	8.39	9.07	9.42	8.91	9.10

sound and heavy rain conditions, while the spectrogram based system achieved lower MAE when speech was present. For the "hive-independent" setting, the three feature sets achieved similar results under clean conditions. The errors, however, are substantially larger than those achieved in the random-split case, suggesting great variability between hives. As previously, noise severely degraded performance. In this case, the spectrogram features achieved lower MAE for urban sounds and speech, whereas MFCCs performed better for heavy rain. In this setting, speech artifacts caused the most harm to monitoring systems. Taken together, these results highlight the potential of acoustics used in a machine learning framework and also, the importance of context awareness for beehive monitoring to ensure that such environmental factors are taken into account during remote monitoring.

In addition, delta and double-delta MFCCs were also evaluated; however, the obtained results did not justify the increased computational cost and feature dimensionality, as performance either remained similar or decreased. This observation is consistent with prior works reporting limited benefits of delta-based features in similar settings [184].

In this section, we showed the impact that three prominent noise sources (urban sounds, background speech, and heavy rain) had on three widely-used features (spectrogram, MFCCs, and DWT). These impacts include significant changes in important bands of frequency in spectrogram, changes in the MFCCs distribution, especially in the first two coefficients, and lastly, distortion in the details and approximation DWT coefficients. These features were then used to train three separate hive strength classifiers. The results indicated the potential of audio used in a deep learning model for predicting beehives population and health state as the system based on MFCCs achieved an MAE of 0.86 for the clean samples. Tests using clean audio segments and noisy segments showed the sensitivity of the systems to environmental noise. Next, we explore if off-the-shelf voice activity detectors (VADs) could be used to mitigate the effect of human speech on hive monitoring performance.

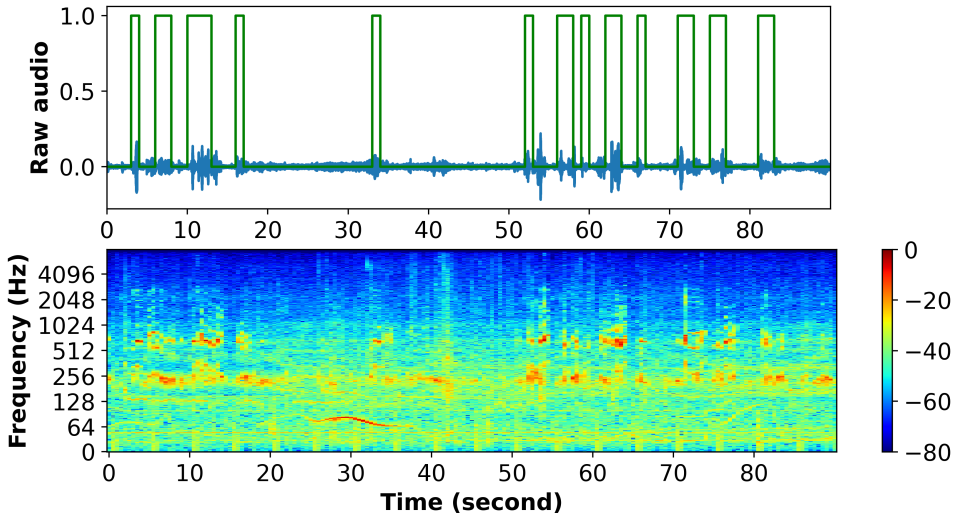


FIGURE 4.14 : A 90-second excerpt from the Nu-hive dataset, from top to bottom, raw audio signal in blue and ground truth labels in green, and spectrogram.

4.3 Voice Activity Detectors for Acoustic Beehive Monitoring

Speech detection in noise has been a prominent area of study in telecommunications, and voice activity detectors (VADs) have been developed, ranging from simple energy threshold based systems [185] to more recent deep neural network systems [186], as well as spanning applications in wireless communications to voice over internet protocols [187]. While VADs have shown to be accurate in applications involving ambient noises, their accuracy has yet to be quantified if speech data is collected together with bee buzzing sounds. Here, we fill this gap and explore if existing VADs could be a useful ally for automated beehive monitoring systems.

In order to test VAD accuracy across datasets, we run experiments on our in-house dataset (UrBAN [180]) and a second dataset obtained via the Nu-hive open source project [67]. This data is comprised of audio recordings with a sampling rate of 32 kHz and obtained from two beehives. These two beehive located in Ancona, Italy were monitored during two days, one day where the queen bee was present and one other day without the queen bee (i.e., the queen is dead). From these two datasets, a subset was selected and manually labelled to find the beginning and end of different noise events. Figure 4.14, for example, shows a 90-second excerpt from the Nu-hive dataset with the detected noise events shown via the green curve.

In this section, we describe three VAD based methods used in this study, including energy-based, a standard method used with the web real-time communication (WebRTC) project, and a more recent method based on deep neural networks. We opted for the energy-based technique mainly because of its simplicity, chose WebRTC for its ability to handle noisy situations, and considered a deep neural networks approach due to its capacity to adapt to diverse acoustic environments and offer transfer learning capabilities.

4.3.1 Short-time Energy Thresholding

One of the simplest and most classical methods of detecting voice activity is via short-time energy thresholding [188]. In essence, the assumption is that when speech is active, the energy levels will surpass a noise floor and this can be detected via simple thresholding rule. First, the energy of the acoustic signal (i.e., a frame) is computed across all frequencies. Next, the energy is computed over a specific frequency subband known to contain mostly speech; in our experiments, this was set to frequencies between $300 - 3000$ Hz. The speech-to-total energy ratio is then taken and compared to a pre-defined threshold. In our experiments, this threshold was set to 0.6, thus energy ratios above this value were considered to be speech and those below to be silence or noise. While energy-based VAD algorithms are simple and low in complexity, they can be sensitive to background noise, thus may be affected by bee buzzing sounds inside the hive.

4.3.2 WebRTC

WebRTC is an open-source project aimed at creating a common set of protocols for real-time communications in web browsers and mobile applications. Voice activity detection is one of the main components of the WebRTC stack [187] and relies on the analysis of log energies across six frequency bands (between 80Hz-4000Hz) with a pre-trained Gaussian mixture model (GMM) classifier. The WebRTC VAD has been designed to work in noisy and variable acoustic environments, thus it is hypothesized that it will perform well under beehive audio recordings. Moreover, it has been widely used in domains such as voice and video communication, where its consistent efficacy is demonstrated, underscores its proficiency in handling real-world noise challenges. In this study, the Python implementation of the WebRTC VAD is used [189]. This implementation has an option to adjust the level of aggressiveness for filtering non-speech audio by using an integer ranging from zero to three. A value of three (representing the most aggressive mode) is chosen in this study. More details about the WebRTC VAD can be found in [189].

4.3.3 Convolutional Recurrent Deep Neural Network Based VAD

Lastly, a pre-trained model from the SpeechBrain toolkit was tested, namely the VAD-CRDNN [190], which relies on a convolutional recurrent deep neural network (CRDNN) architecture which combines a convolution neural network (CNN), a recurrent neural network (RNN) and a fully connected network. The CNN block includes 2 CNN layers based on VGG blocks with 16 and 32 neurons, respectively, and kernel size of $(3, 3)$. The RNN block is a 2-layer Gated Recurrent Unit (GRU) while the DNN block includes 2 layers with 16 neurons in each. This model was trained on the Libriparty dataset [190], which includes several simulated acoustic scenes with speech and noisy sequences periodically active (alone or simultaneously). The features that are input to the neural network are

standard mel filterbank energies calculated using 40 mel filters. The output of the model are posterior frame-level/segment level probabilities where a threshold can be applied to classify segments as speech and non-speech. One advantage of a pre-trained model and that it can be further fine-tuned on different tasks. Here, we explore fine-tuning the pre-trained VAD-CRDNN with audio data taken from both datasets; fine-tuning was achieved by training for 30 epochs. The fine-tuned model will be termed VAD-CRDNN-FT. The binary cross entropy is used for loss function.

4.3.4 Experimental Results

In this section, the performance of each VAD method is verified on the two test sets. To evaluate the models, four evaluation metrics are considered : accuracy, true positive rate (TPR), true negative rate (TNR), and F1-score. TPR and TNR are given by :

$$TPR = \frac{TP}{TP + FN}; \quad TNR = \frac{TN}{TN + FP}. \tag{4.10}$$

Here, we consider speech segments as “positive” and non-speech segments as “negative”. Therefore, true positives (TP) are correctly predicted speech segments, while true negatives (TN) are correctly predicted non-speech parts. In turn, false positives (FP) are non-speech samples wrongly predicted as speech, and false negatives (FN) are speech samples wrongly predicted as non-speech. With these values, the accuracy and F1-scores are given by :

$$Accuracy = \frac{TP + TN}{P + N}; \quad F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}, \tag{4.11}$$

where $P + N$ is the summation of all positive and negative samples, thus total number of samples in the analysis.

Table 5.8 presents the performance of each method, where Nu-hive results are listed in Panel A and U on Panel B. As can be seen, WebRTC outperformed the other methods across both datasets. On the Nu-hive data, it achieved an accuracy and F1 score of 0.688 and 0.695, respectively, whereas for Nectar it was 0.760 and 0.807, respectively. Moreover, while the simple energy thresholding based method performed well on detecting non-speech segments, it performed poorly on detecting speech samples. In the case of VAD-CRDNN, fine-tuning the model showed to improve TPR and F1-score, but reduced TNR. Fine-tuning with data from the two datasets, further improved detection of speech and improved F1 score. Figure 4.15 depicts an excerpt comprised of bee buzzing and beekeeper background speech with manually labeled speech segments (plot a), as well as speech segments detected by the energy thresholding method (b), WebRTC (c), VAD-CRDNN (d), and VAD-CRDNN-FT (e).

Overall, it can be seen that WebRTC provides the best overall speech detection accuracy independent of dataset. By analyzing the signal in both the time and frequency domains, WebRTC

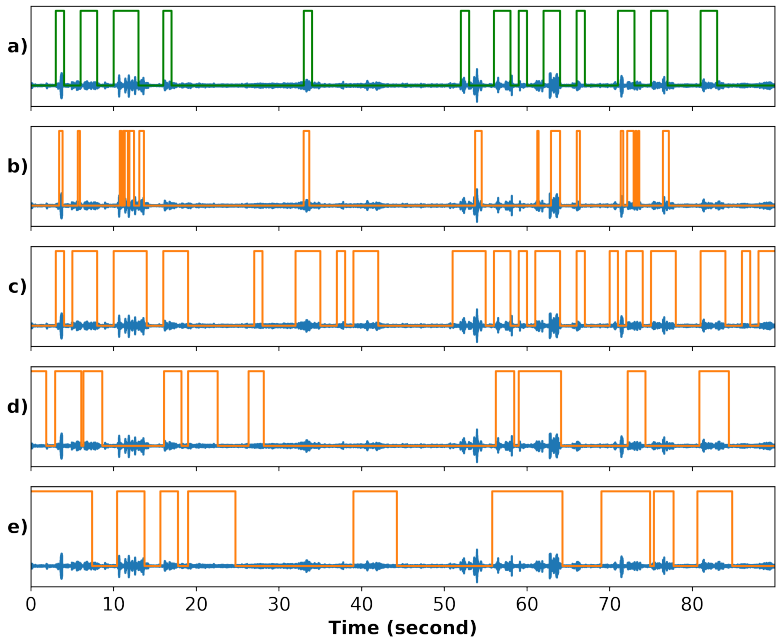


FIGURE 4.15 : Speech detection on a 90 second sample of the Nu-hive dataset : a) manual labels, b) short-time energy VAD, c) WebRTC, d) VAD-CRDNN, and e) VAD-CRDNN-FT.

captures spectral characteristics that are indicative of speech, even when the signal is masked by noise. As expected, the simple energy based VAD does not perform well for beehive acoustics monitoring. Despite the high accuracy of recent DNN-based systems for telecommunications applications (e.g., [191]), for automated beehive monitoring, such architectures were not very effective. While fine tuning showed some improvement, results were still substantially lower than what could be achieved with the WebRTC VAD. Next, we explore the effects that audio enhancement algorithms can have on remote beehive monitoring performance.

4.4 Audio Enhancement

Removing environmental noise from beehive audio signals is crucial for enhancing the effectiveness of precision beekeeping monitoring. As shown in the previous sections, environmental noise can obscure the sounds produced by bees, making it difficult to accurately detect and analyze important behaviors and events within the hive. Moreover, it could reduce the accuracy of the monitoring system significantly. While some studies explored methods to remove noises, including detecting human speech [192, 193], as in the previous section, there is still a need for a general noise removal preprocessing step.

TABLE 4.3 : Performance evaluation on Nu-hive and UrBAN test datasets. The VAD-CRDNN-FT-All condition exemplifies the case where fine-tuning was done using both datasets.

Panel A : Nu-hive.

Speech detection methods	**Nu-hive**			
	Accuracy	**TPR**	**TNR**	**F1**
Short-time energy	0.558	0.129	0.987	0.226
WebRTC	**0.688**	**0.709**	**0.668**	**0.695**
VAD-CRDNN	0.625	0.442	0.809	0.541
VAD-CRDNN-FT	0.615	0.661	0.570	0.632

Panel B : UrBAN

Speech detection methods	**UrBAN**			
	Accuracy	TPR	TNR	F1
Short-time energy	0.502	0.005	1.00	0.001
WebRTC	**0.760**	**1.00**	**0.521**	**0.807**
VAD-CRDNN	0.631	0.230	1.00	0.417
VAD-CRDNN-FT	0.632	0.260	1.00	0.418
VAD-CRDNN-FT-All	0.633	0.520	1.00	0.559

4.4.1 Spectral Amplitude Subtraction

Spectral amplitude subtraction is a classical technique used in audio processing to enhance the quality of audio recordings [194]. As shown in the block diagram in Figure 4.16, it involves subtracting the spectral components of noise or unwanted signals from the original audio signal to reduce background noise and improve signal clarity. By identifying the spectral profile of noise or interference and subtracting it from the audio signal, spectral amplitude subtraction helps isolate the desired sound, resulting in cleaner and more intelligible audio recordings. Figure 4.16 shows the diagram of the spectral subtraction algorithm, where $y(n)$ and $\hat{x}(n)$ are the noisy and cleaned signals, respectively. As the figure indicates, after framing the audio and calculating the fast Fourier transform (FFT), the noise spectral profile needs to be estimated and eventually subtracted from the noisy signal, following the equation :

$$|\hat{X}_m(\omega)| = |Y_m(\omega)| - |D_m(\omega)|. \tag{4.12}$$

Here, we used an exponential moving average (EMA) filter to estimate the noise.

$$d[n] = (1-\alpha)y[n] + \alpha d[n-1], \tag{4.13}$$

where $d[n]$ and α are the noise signal and the weighting factor or smoothing parameter, respectively. In order to only update α in noisy frames, an adaptive algorithm is used [195]. Using this method, α will be a function of a-posteriori signal-to-noise ratio (denoted as $\text{SNR}_{\text{a-posteriori}}$) and is calculated for each frame, i.e.,

$$\alpha(m) = \frac{1}{1 + exp{-}a(SNR_{a-posteriori}(m) - T)}, \tag{4.14}$$

where the parameter a represents the steepness of the sigmoid function's slope and T is a threshold value for the $\text{SNR}_{\text{a-posteriori}}$. The parameter T determines the point at which the smoothing parameter α transitions between its extremes in response to varying $\text{SNR}_{\text{a-posteriori}}$ values. A larger T causes the adaptive algorithm to identify a frame as noisy only if the $\text{SNR}_{\text{a-posteriori}}$ is considerably low, thereby reducing α. Conversely, a smaller T makes the algorithm more sensitive to detecting noise. The parameter T must be tuned based on the characteristics of the input signals and the noise environment.

The $\text{SNR}_{\text{a-posteriori}}$ is estimated using the power spectrum of the noisy signal and the mean noise power spectrum over the past N frames, i.e., :

$$\text{SNR}_{\text{a-posteriori}}(m) = \frac{|Y_m(\omega)|^2}{\frac{1}{N}\sum_{k=m-N}^{m-1}|D_k(\omega)|^2}, \tag{4.15}$$

where $|Y_m(\omega)|^2$ is the power spectrum of the noisy signal in the current frame, and the denominator represents the average noise power spectrum estimate over the last N frames, effectively smoothing out noise variations. The estimation of the $\text{SNR}_{\text{a-posteriori}}$ provides an indication of how much noise is present in each frame. The $\text{SNR}_{\text{a-posteriori}}$ can be interpreted as follows :

— When $\text{SNR}_{\text{a-posteriori}} \approx 1$, the frame predominantly contains noise.
— For $\text{SNR}_{\text{a-posteriori}} > 1$, the frame contains both speech and noise.

The scripts for the spectral subtraction algorithm are made available in our Github repository (`https://github.com/mahsa-abdollahi/UrBAN`) to facilitate study replication. Figure 4.17 depicts the audio signal collected from a hive before and after enhancement. The spectrograms illustrate distinct energy bands at approximately 64 Hz, 128 Hz, and 256 Hz both before and after spectral amplitude subtraction. In the noisy signal's spectrogram, noise is shown to overlap with frequency bands characteristic of bee audio. Following spectral amplitude subtraction, the processed spectrogram exhibits a reduction in noise while retaining the essential frequency components of the bee audio. Experiments in the subsequent chapters will compare performance with and without spectral subtraction to gauge it benefits for remote beehive monitoring.

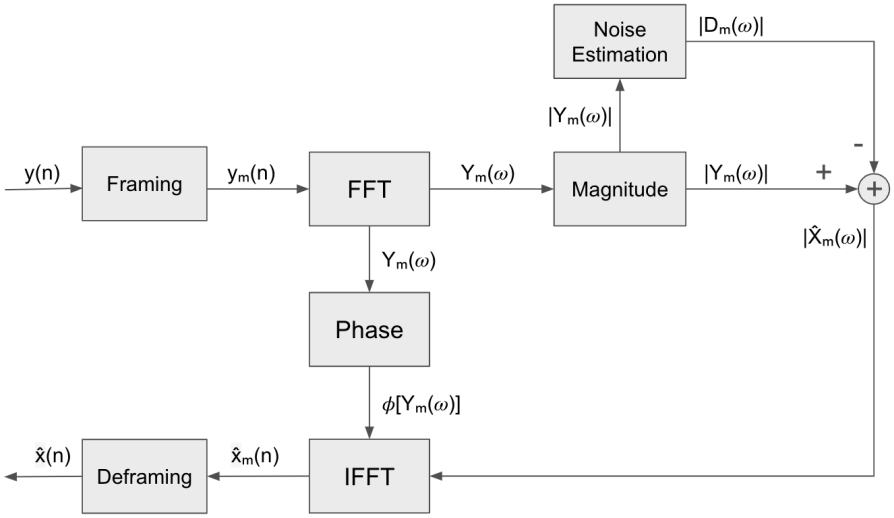


FIGURE 4.16 : Block diagram of the spectral amplitude subtraction.

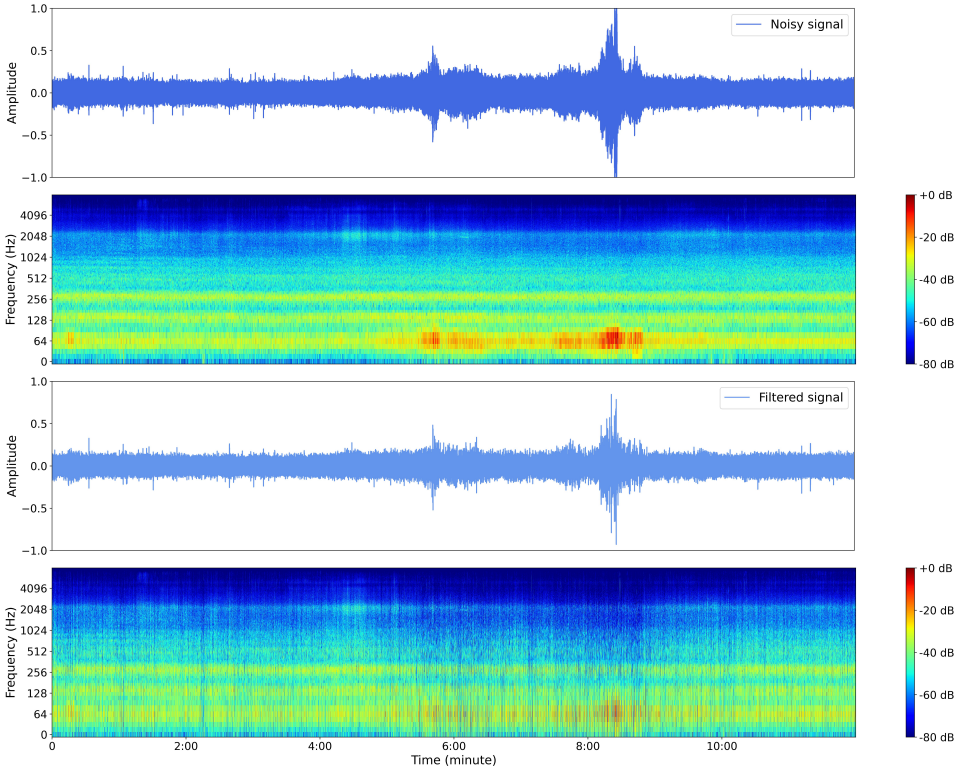


FIGURE 4.17 : From top to bottom, noisy audio amplitude of a beehive, the filtered audio, and their corresponding spectrograms.

4.5 Discussion and concluding remarks

This chapter investigated the role of context awareness in acoustics-based beehive monitoring systems, with a particular focus on understanding and mitigating the effects of real-world environmental noise. Using long-term field recordings from the UrBAN dataset and complementary

datasets, we analyzed how common environmental disturbances influence acoustic features, machine learning models, and preprocessing strategies used in automated hive monitoring.

First, we demonstrated that prominent environmental noise sources, namely urban sounds (e.g., train rail squealing and traffic), human speech, and heavy rainfall have a substantial and often detrimental impact on widely used acoustic representations, including spectrograms, MFCCs, and DWT-based features. These noise sources introduce distortions in frequency bands critical for beehive acoustics, alter feature distributions, and contaminate time–frequency structures. Experimental results showed that such effects directly translate into degraded performance of hive strength prediction models, particularly in hive-independent evaluation settings where robustness and generalization are essential. These findings emphasize that feature sensitivity to environmental context remains a critical bottleneck in real-world deployments.

Second, we explored the use of voice activity detection as a means of identifying and potentially mitigating speech-related artifacts in beehive recordings. By evaluating three representative VAD approaches energy-based thresholding, WebRTC, and a deep learning–based CRDNN, we showed that classical VAD methods developed for telecommunications can behave very differently in the presence of continuous bee buzzing. Among the evaluated methods, WebRTC consistently achieved the best balance between speech detection accuracy and robustness across datasets. In contrast, both simple energy-based methods and deep neural network–based VADs struggled to reliably distinguish speech from bee-generated sounds, even after fine-tuning. These results suggest that careful selection and validation of auxiliary tools such as VADs is necessary before integrating them into beehive monitoring pipelines.

Finally, we propose the incorporation of spectral amplitude subtraction as a lightweight audio enhancement technique to reduce background noise prior to feature extraction. By employing an adaptive noise estimation strategy based on the a-posteriori SNR, the proposed approach selectively attenuates noise-dominated frames while preserving the dominant harmonic components associated with bee activity. While spectral subtraction alone cannot fully address non-stationary or highly overlapping noise sources, it provides an efficient and practical enhancement stage that can improve the reliability of downstream feature extraction and modeling. Future experiments in subsequent chapters will show the quantitative effects of the spectral subtraction method on the tasks at hand.

Taken together, the results presented in this chapter highlight the necessity of incorporating context-awareness into acoustics-based beehive monitoring systems. Environmental noise is not a marginal nuisance but a defining characteristic of real-world deployments, particularly in urban apiaries. The analyses in this chapter motivate the development of more noise-resilient feature representations and modeling strategies, which move beyond conventional speech-inspired descriptors. These insights directly inform the subsequent chapters, where more robust spectral and modulation-based features are proposed to better capture the intrinsic dynamics of beehive acoustics while remaining resilient to environmental variability.

5 MODULATION SPECTROGRAM-BASED FEATURES

5.1 Preamble

This chapter is compiled from material extracted from the following peer-reviewed articles :

1. "Audio Modulation Spectral Features for Improved Honey bee Colony Population Prediction [196]" (section 5.2).
2. "Non-Invasive Prediction of Varroa Mite Infestation in Honey bee Colonies via Machine Learning and Acoustic Features [197]" (section 5.3).

5.2 Colony Population Prediction

In the previous chapter, we examined the impact of several environmental noise sources such as nearby train rail squealing, beekeeper speech, and rainfall on commonly used acoustic features, and demonstrated the significant degradation in performance observed for spectral, MFCC, and wavelet-based representations under noisy conditions. These findings highlight the limitations of conventional features in real-world beehive monitoring scenarios and underscore the need for more robust feature representations that are less sensitive to environmental interference. Developing such features is essential to improve model generalization and ensure reliable performance under unseen and variable acoustic conditions.

Another challenge in automated beehive monitoring is that of poor generalizability of the models to unseen hives [72, 48, 49]. As such, developing hive-independent models that function reliably across hives has become a pressing issue, ensuring that models can account for differences in bee populations and hive structures. As can be seen, addressing both environmental noise sensitivity and hive variability has become an essential component of beehive monitoring technologies, facilitating early detection of anomalies and supporting proactive beekeeping practices.

To address these issues, we propose the use of so-called modulation spectral features for honey bee colony population prediction, an automated monitoring task that has gained popularity in recent years (e.g., [176, 173, 125]). Modulation spectral features quantify the rate-of-change of the acoustic spectra over time, thus can be seen as a frequency-frequency representation of the acoustic signal. Modulation spectral features provide greater discrimination of the naturally occurring temporal dynamics of the buzzing spectra, relative to the dynamics from environmental noise and artifacts, as well as hive structures. The modulation spectral signal representation has shown improved robustness to noise across modalities, ranging from speech, electrocardiogram, to electroencephalogram signals, just to name a few (see e.g., [198]). Here, they are explored for the first time for acoustic beehive monitoring.

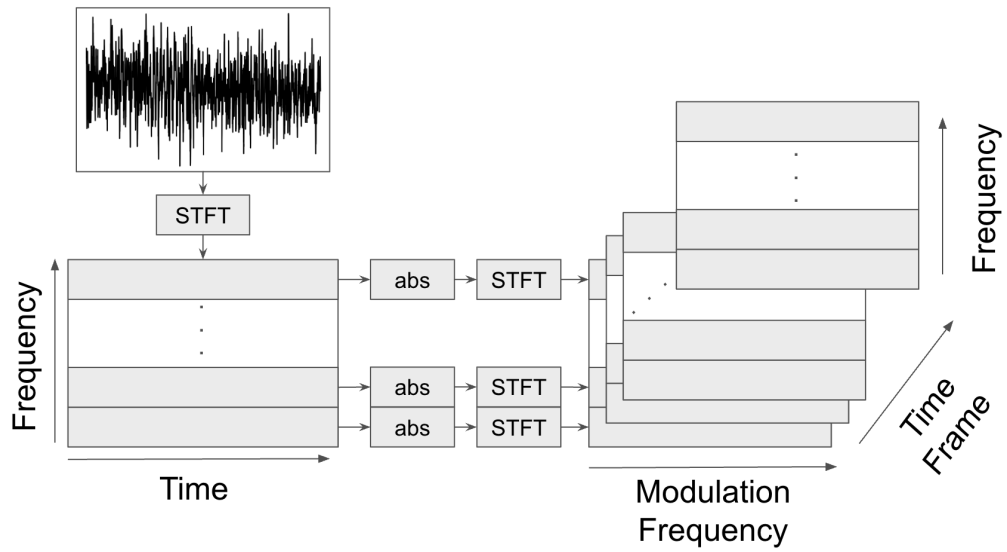


FIGURE 5.1 : Block diagram of signal processing steps involved in calculation of the modulation spectrogram.

5.2.1 Proposed Modulation Spectral Features

The traditional spectrogram offers insights into the variation of frequency components over time. However, it often falls short in classification tasks with noisy data due to the overlap of noise and signal in both time and frequency domains. The modulation spectrogram, on the other hand, captures the rate of change of these frequency components, effectively better distinguishing between signal and noise dynamics. This makes the modulation spectral representation a more suitable option for real-world applications.

The modulation spectrogram is an advanced signal processing tool that captures the dynamic characteristics of frequency components over time. Figure 5.1 depicts the block diagram showing the signal processing steps involved in the computation of the modulation spectrogram. For a complete mathematical formulation of the representation, the interested reader is referred to [199]. Here, only a brief overview is provided.

As can be seen from the figure, the first step is to compute a time-frequency representation of the acoustic signal. Here, a short-time Fourier transform (STFT) with a 100 ms window length and a 12.5 ms hop length is used, resulting in the widely-used spectrogram. This process divides the signal into overlapping segments and applies the Fourier transform to each segment to obtain the time-frequency representation.

Next, for each frequency bin, the spectral magnitude is taken, followed by another STFT with a 60 s window length and no overlap operating over the spectral magnitude time series. The result is a 3-dimensional tensor where the y-axis corresponds to the conventional acoustic frequency of the spectrogram and the x-axis to the so-called modulation frequency, which measures the rate of change of each spectral bin. Since a second STFT was used, the third dimension (z-axis) corresponds to time, where each $freq \times modulationfrequency$ modulation spectrogram is computed per

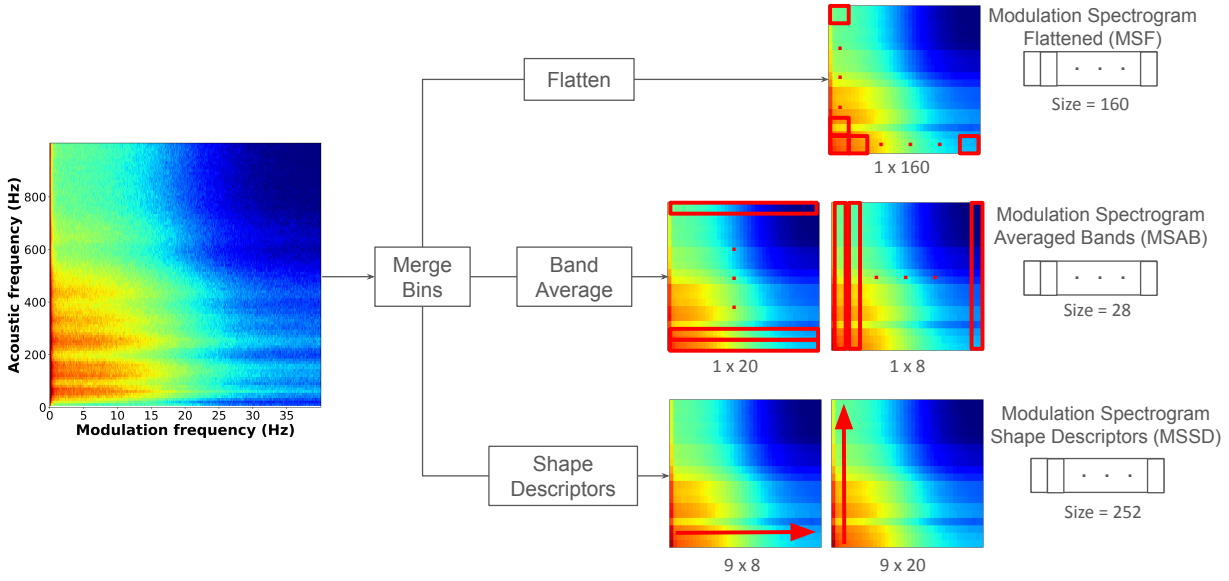


FIGURE 5.2 : Block diagram of the three feature extraction processes from modulation spectrogram : flattening into 160-D modulation energy vector (top), into 28-D banded modulation energy values (middle), and 252-D spectral description values (bottom).

STFT frame. This 3-dimensional $freq \times modulationfrequency \times time$ representation is referred to as a modulation tensorgram.

In the past, various methods of processing the modulation tensorgram have been proposed. For example, in [200], the tensorgram was processed directly by a deep neural network for speech-based COVID-19 detection, while in [201] it was used for speech quality measurement. In turn, in [202] and [203], different temporal averaging methods were used for speech emotion recognition and room acoustics characterization, respectively. Temporal averaging results in just an average 2-dimensional $freq \times modulationfrequency$ representation referred to as a modulation spectrogram, which conveys the available modulation energy in each $freq \times modulationfrequency$ bin.

Since acoustic beehive monitoring applications are still emerging, very few open-access datasets exist today, thus limiting the use of more complex deep neural networks. To this end, we propose here to use feature engineering from the modulation spectrogram as a first step towards environment-robust beehive monitoring. Figure 5.2 depicts three separate methods used to extract features/patterns from the modulation spectrogram computed by simple time-averaging of the computed modulation tensorgram. In all cases, we first merge the acoustic frequency and modulation frequency bins into a more manageable number, chosen empirically to be 20 and 8, respectively. This reduction achieves a target resolution of 50 Hz for the acoustic frequency and a finer 5 Hz resolution for the modulation frequency.

The three feature sets are then derived as follows :

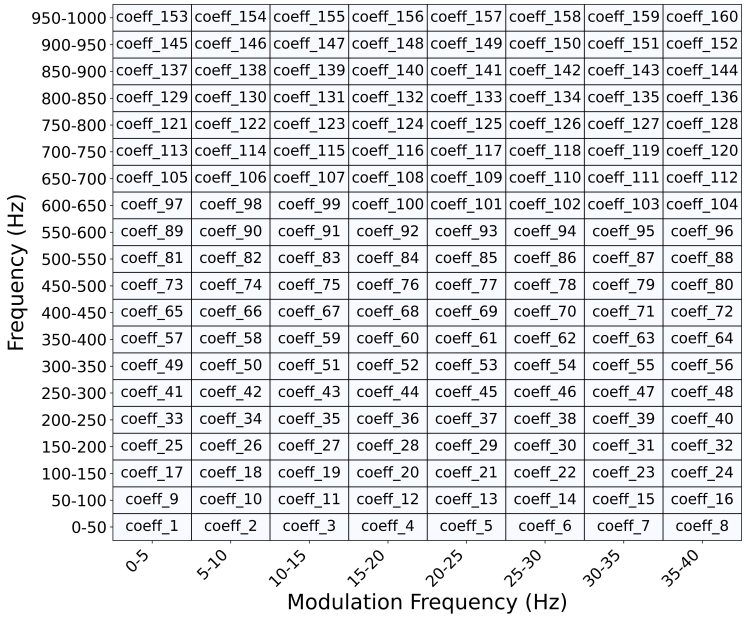


FIGURE 5.3 : MSF feature set map that labels each bin once the spectrogram flattens.

— **Modulation Spectrogram Flattened (MSF)** : This feature set is obtained by flattening the 20×8 modulation spectrogram into a single 160-dimensional vector of modulation energies. Figure 5.3 shows a map that labels each bin once the spectrogram flattens.

— **Modulation Spectrogram Averaged Bands (MSAB)** : For this feature set, we compute the average energy across each acoustic frequency band (resulting in 20 values) and across each modulation frequency band (resulting in 8 values). The 28 values are then appended into one 28-dimensional vector of band-averaged modulation energies. Figures 5.4 (a) and (b) describe the map that labels each frequency band after averaging across modulation and acoustic frequencies, respectively.

— **Modulation Spectrogram Shape Descriptors (MSSD)** : This set involves calculating nine spectral shape descriptors along both the acoustic frequency (resulting in $9 \times 20 = 180$ values) and modulation frequency bands (resulting in $9 \times 8 = 72$ values). The nine shape descriptors include the centroid, spread, skewness, kurtosis, entropy, rolloff, flatness, crest, and flux [204]. The 252 features are then appended into one vector of shape descriptors. Figure 5.5 illustrates the extracted features from both axes, exemplified here by the spectral centroid.

In the next section, we explore the benefits of using these proposed features for improved automated beehive monitoring.

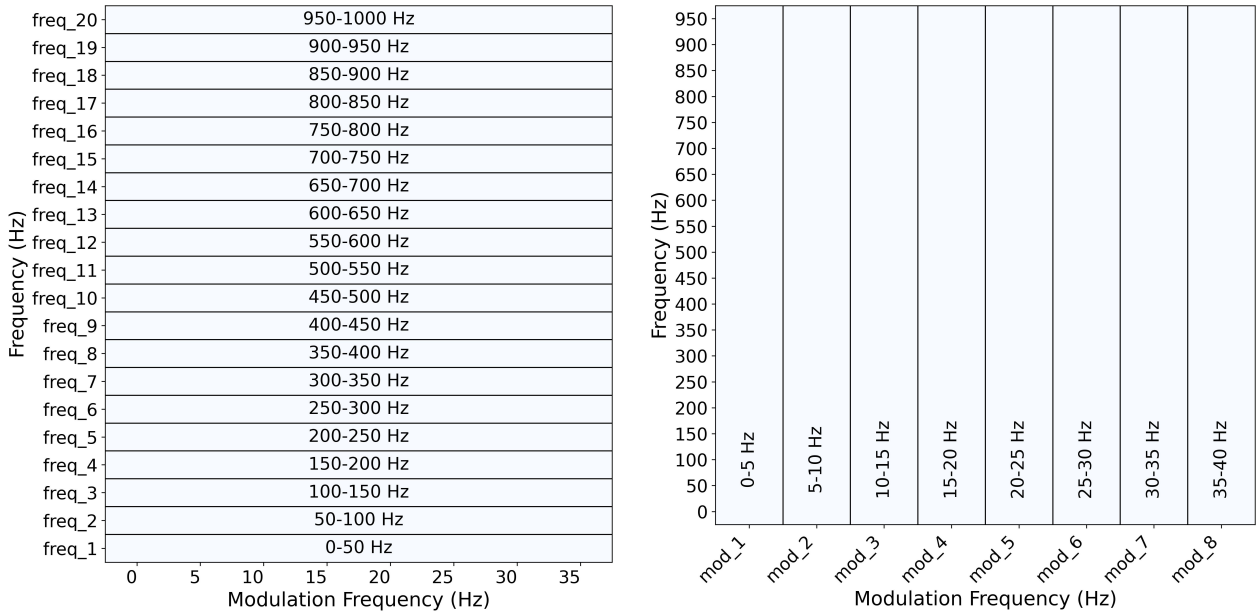


FIGURE 5.4 : MSAB feature set map labeling each frequency band after averaging across (a) the modulation frequency axis and (b) the acoustic frequency axis.

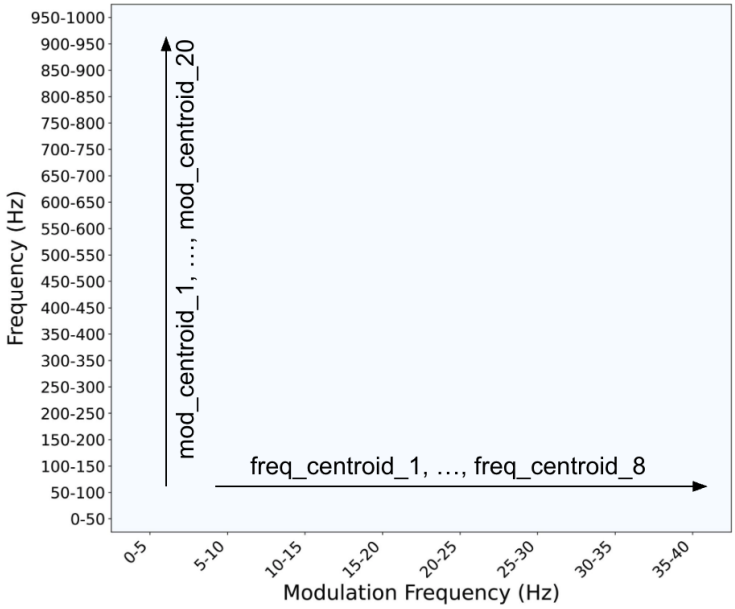


FIGURE 5.5 : MSSD feature set map that shows how the extracted features are named, exemplified here with just one of the nine descriptors, i.e., the centroid.

5.2.2 Experimental Setup

5.2.2.1 Benchmark features and noise removal

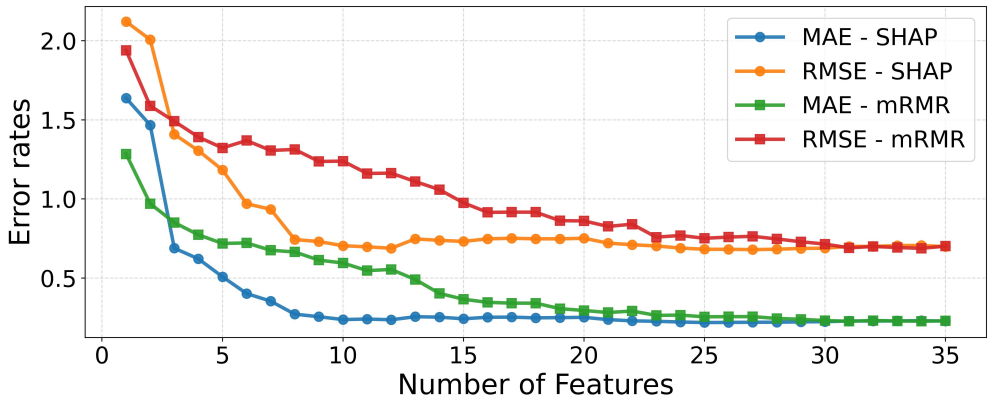
To assess the effectiveness of the proposed features, three benchmark features are also used for comparison. As mentioned previously, MFCCs simulate human auditory processing through mel-scale frequency mapping and have been widely used for precision beekeeping [52]. Here, we extract 12 coefficients, plus the zeroth coefficient as a measure of signal power, using 26 mel filters. Linear-frequency cepstral coefficients (LFCCs) are also extracted via linear filters for comparison. Lastly, we use the same nine descriptors (centroid, spread, skewness, kurtosis, entropy, rolloff, flatness, crest, and flux), but computed from the spectrogram. The spectral amplitude subtraction method is used for noise reduction, as described in chapter 3.

5.2.2.2 Feature selection, regression task, and testing setup

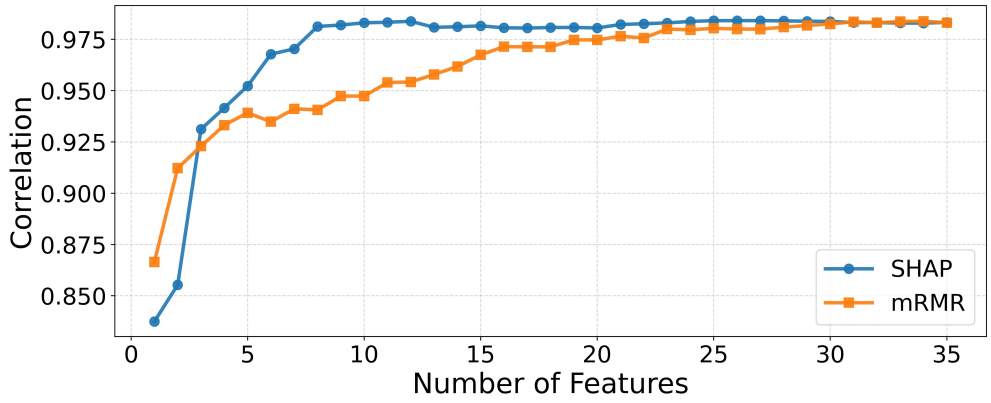
Building on recommendations from previous studies, we explored two distinct experimental configurations, namely “random-split" and “hive-independent". The complete audio dataset spanning 18 (one of the beehives had a problem in audio recording) hives from the years 2021 and 2022 is used for the experiment. In the random-split approach, a nested 5-fold cross-validation was used to ensure robust evaluation of the models. This cross-validation process ensures that every data point is included in the test set exactly once, reducing the risk of overfitting and providing a more reliable estimate of model performance. Conversely, in the hive-independent setup, the 5 folds are selected hive-independently for both validation and testing.

To address the curse of dimensionality and improve model generalization, implementing a feature selection step is critical. This process helps eliminate redundant or irrelevant features, thus minimizing overfitting and enhancing model capability to discern significant patterns in the data. We examined two feature selection methods, including the minimum redundancy maximum relevance (mRMR) [205] and SHAP (SHapley Additive exPlanations) [206]. SHAP works by quantifying the contribution of each feature to the model’s predictions. For each sample, SHAP computes a value that reflects how much a feature increases or decreases the prediction compared to a baseline. By aggregating these values across all samples, typically by taking the mean absolute SHAP value, features can be ranked according to their overall importance. The most influential features are then selected, as they consistently contribute the most to the model’s output, while less important features can be discarded.

To ensure that feature selection does not bias the model evaluation, a subset of the dataset, comprising four beehives, was exclusively reserved for this purpose. Once the top-ranked features are found, a random forest (RF) regressor was used to predict the number of frames of bees, a value ranging from 0-30. Table 5.1 provides more details on the hyperparameters explored for the



(a)



(b)

FIGURE 5.6 : Performance comparison between SHAP and mRMR feature selection methods applied to fused benchmark features in a random-split setup, evaluated using (a) MAE and RMSE error metrics, and (b) correlation.

model tuning. RF models were trained for each benchmark feature set, as well as each proposed feature set.

Figure 5.6 presents the performance comparison of two feature selection methods applied to a fusion of benchmark features derived from the feature selection subset of dataset. The error metrics, i.e., mean absolute error (MAE), root-mean-square error (RMSE), and correlation, consistently improve as the number of selected features increases, indicating that incorporating more informative features enhances model performance. The optimal feature count is determined by identifying the earliest point at which these error metrics are minimized. Notably, features selected using SHAP consistently reach optimal values much quicker than with mRMR, thus requiring a smaller number of features. This can be important for IoT monitoring, ensuring lower computational costs and longer battery life.

TABLE 5.1 : Hyperparameter values explored for tuning the random forest model. This includes a range of options for the number of trees, feature selection methods, tree depth, and criteria for splitting and leaf nodes to optimize model performance.

Hyperparameter	Values	Description
`n_estimators`	100, 200, 300, 400, 500	Number of trees.
`max_features`	'auto', 'sqrt', 'log2'	Number of features.
`max_depth`	10, 20, 30, 40, 50	Maximum depth.
`min_samples_split`	2, 5, 10, 15, 20	Minimum number of samples required to split an internal node.
`min_samples_leaf`	1, 2, 4, 8	Minimum number of samples required to be at a leaf node.
`bootstrap`	True, False	Use of bootstrap samples.

5.2.2.3 Figures-of-Merit

For evaluation of the regression models trained on the benchmark and the proposed feature sets, five figures-of-merit are utilized : mean absolute error (MAE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), Pearson correlation, and R-squared (R^2). These metrics are computed as follows for a sample size of L :

$$MAE = \frac{\sum_{n=1}^{L} |y_i - \hat{y}_i|}{L}, \tag{5.1}$$

$$RMSE = \sqrt{\frac{\sum_{n=1}^{L} (y_i - \hat{y}_i)^2}{L}}, \tag{5.2}$$

$$MAPE = \frac{100}{L} \sum_{i=1}^{L} \left| \frac{y_i - \hat{y}i}{yi} \right|, \tag{5.3}$$

$$r = \frac{\sum_{n=1}^{L} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{n=1}^{L} (y_i - \bar{y})^2 \sum_{n=1}^{L} (\hat{y}_i - \bar{\hat{y}})^2}}, \tag{5.4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{L} (y_i - \hat{y}i)^2}{\sum_{i=1}^{L} (y_i - \bar{y})^2}, \tag{5.5}$$

where y_i represents the observed values, $\hat{y}_i$ denotes the predicted values, $\bar{y}$ is the mean of the observed values, and $\bar{\hat{y}}$ is the mean of the predicted values. These metrics provide comprehensive insights into the models' accuracy and predictive performance. An ideal/perfect regressor will have MAE and RMSE close to zero and a correlation close to unity. Metrics are evaluated across the five folds to gauge their significance against a random RF classifier. In the random baseline classifier, MFCCs are used, but the frame-of-bees values are randomized between 0 and 30 during training.

5.2.3 Results and Discussion

5.2.3.1 Feature importance

Figure 5.7 shows SHAP value plots, which illustrate the feature importance calculated for each of the benchmarks features. Features are ordered by their mean absolute SHAP value, placing the most influential features at the top. The color (red to blue) represents feature values, with red indicating higher values and blue indicating lower values. Top features show more spread across the x-axis, while least important ones cluster near zero with minimal impact on the model. For MFCCs and LFCCs, the first coefficients (mfccs_0 and lfccs_0) represent the overall energy of the recorded audio signal, and as expected, these have a significant impact on the model output. As for the spectral descriptor features, the flux, spread, and entropy stand out.

Figure 5.8 shows the performance of benchmark features based on the selected number of top features in the hive-independent setup. The optimal number of features based on minimizing error rates and maximizing correlation is between 5-8 for all benchmark feature sets.

Next, we explore the usefulness of the proposed feature sets. Figure 5.9 illustrates the SHAP-based feature importance for the modulation spectrogram features. In Figure 5.9a, top-ranked features such as 'coeff_10' and 'coeff_90' correspond to frequency bins of 50–100 Hz and 550–600 Hz, respectively, with modulation frequencies between 5–10 Hz, as detailed in Figure 5.3. Additionally, features such as 'coeff_126' and 'coeff_69' align with frequency bins of 750–800 Hz and 400–450 Hz, coupled with modulation frequencies of 25–30 Hz and 20–25 Hz, respectively.

Next, Figure 5.9b presents the SHAP-based feature importance for the MSAB features. Notably, the top-ranked features (i.e., 'freq_9', 'freq_10', and 'freq_4') exhibit a great impact on the model, as evidenced by their extensive spread along the x-axis. These features correspond to specific bands in the modulation spectrogram as described in Figure 5.4, namely : 'freq_9' aligns with the 400–450 Hz frequency band, 'freq_10' with 450–500 Hz, and 'freq_4' with 150–200 Hz.

Figure 5.9c shows the feature rank for the MSSD features. As can be seen, the modulation spectrogram flatness, entropy and spread have the highest impact on the model output. Figure 5.10 illustrates a pairwise comparison of selected spectral shape descriptors features derived from the MSSD, stratified by the number of bee frames. To ensure balanced representation across catego-

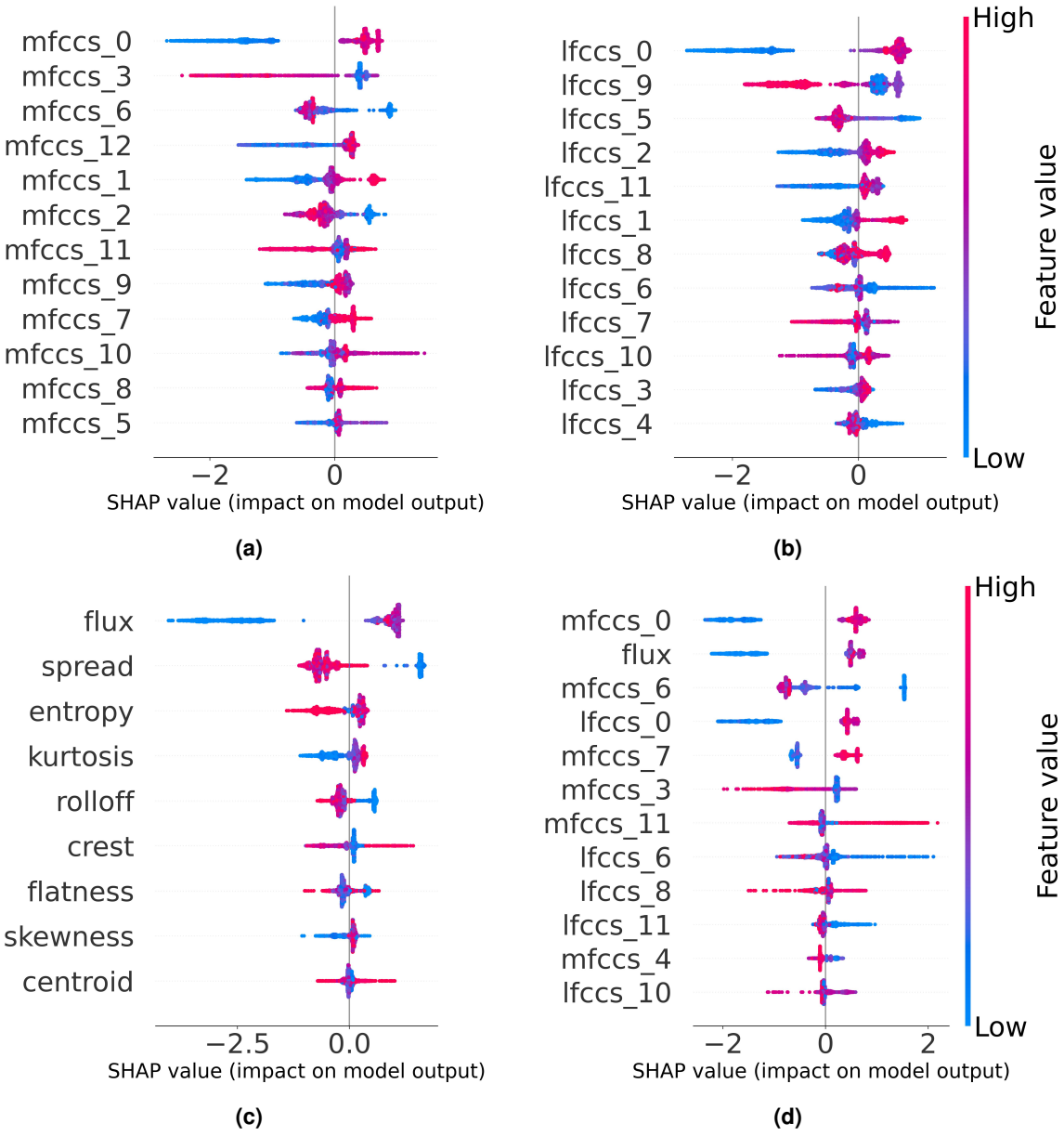
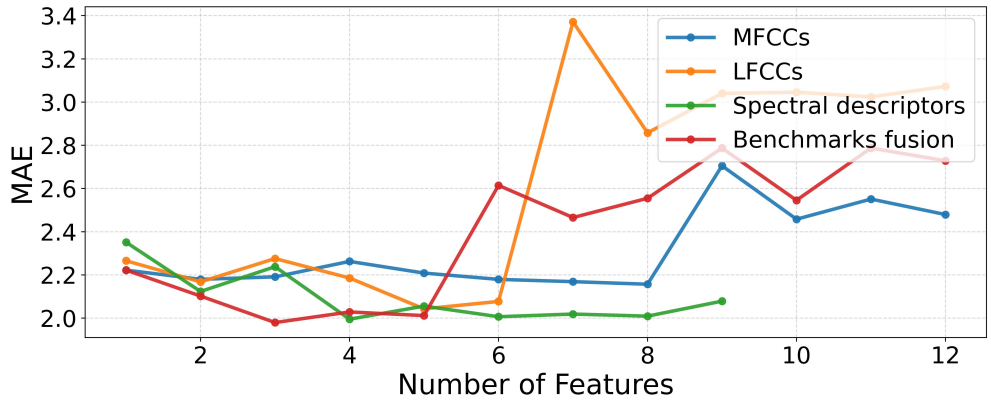


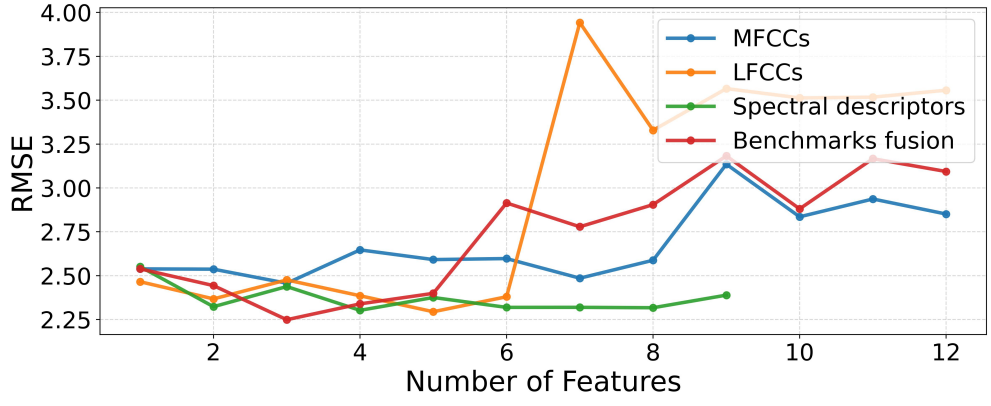
FIGURE 5.7 : SHAP value plots demonstrating the contribution of individual features to model predictions for (a) MFCCs, (b) LFCCs, (c) spectral descriptors, and (d) their fusion. Features are ranked by mean absolute SHAP value, with color denoting the feature value's magnitude.

ries, the dataset was filtered to have an equal number of samples randomly selected from each category. The pair plot displays scatter plots for each pair of the selected features, while the diagonal plots present kernel density estimates, offering insights into the distribution of individual features. Samples from the beehives with lower number of frames of bees show higher values in spectral spread, flatness, and rolloff.

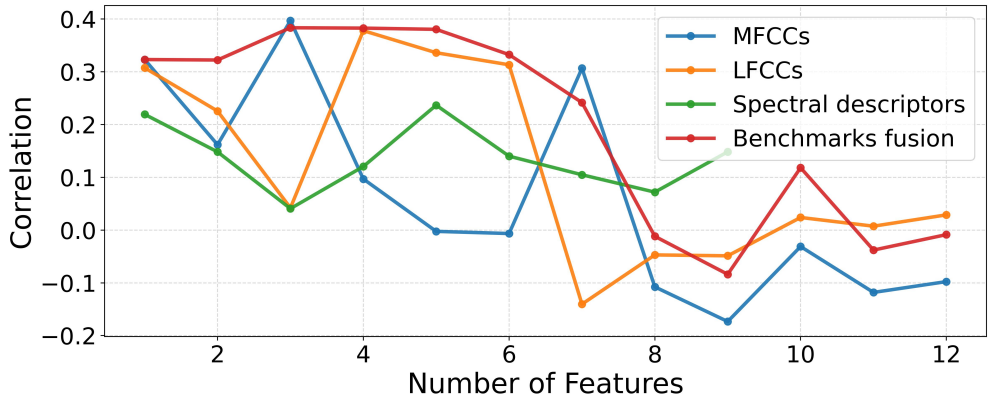
The SHAP feature importance calculated on the fused feature set (i.e., a fusion of all modulation spectrogram features) is shown in Figure 5.9d. As can be seen, the top features are mostly selected from the MSAB feature set, including "freq_10" and "mod_5".



(a)



(b)



(c)

FIGURE 5.8 : Performance comparison of the benchmark features as a function of the number of top-ranked features selected in the hive-independent setup, evaluated using (a) MAE, (b) RMSE, and (c) correlation.

Figure 5.11 illustrates the performance of the modulation spectral features as a function of the number of top-ranked features selected using SHAP in the hive-independent setup. Similar to the benchmark features, figures-of-merit tend to stabilize after a certain number of features. In this case, that number was around 10-30 features, depending on the metric. A visual comparison between Figures 5.8 and 5.11 shows that modulation spectral features achieve lower error rates and

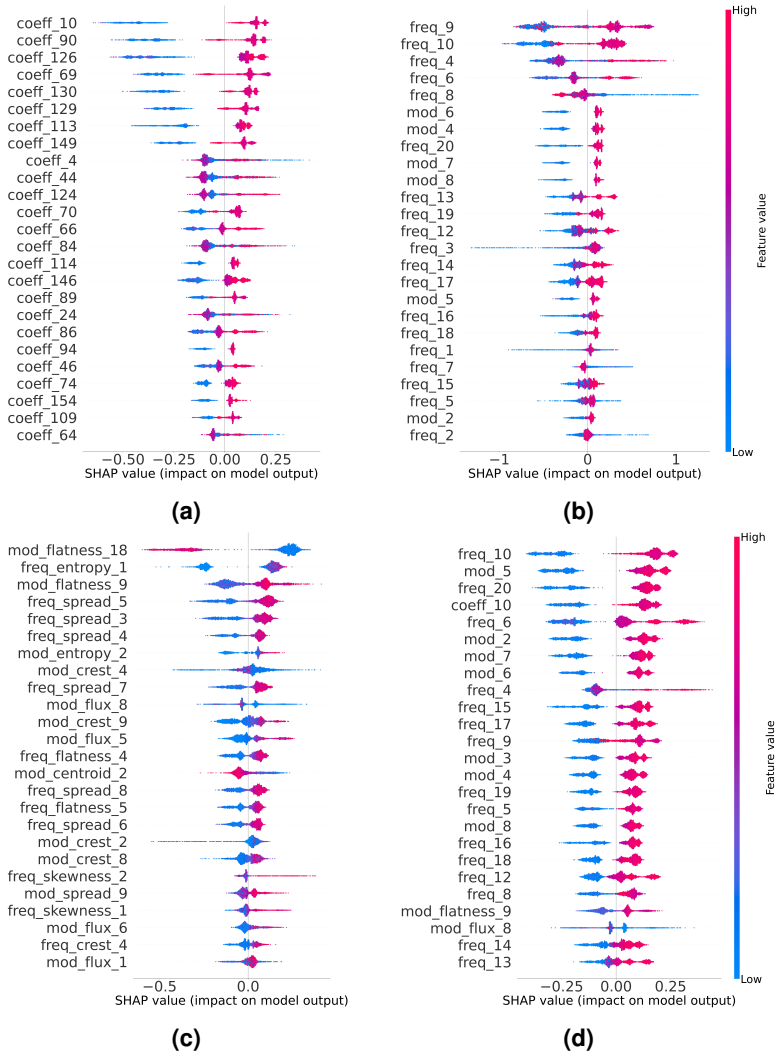


FIGURE 5.9 : SHAP value plots demonstrating the contribution of individual features to model predictions for (a) MSF, (b) MSAB, (c) MSSD, and (d) their fusion. Features are ranked by mean absolute SHAP value, with color denoting the feature value's magnitude.

higher correlation coefficients relative to the benchmark features, indicating enhanced predictive performance.

Next, we explore the complementarity of the proposed and benchmark features. To this end, we combine them into a fused feature set and perform feature ranking on this fused set; top features are shown in Figure 5.12. Notably, several top-ranked features originate from the MSAB set, including 'freq_4', 'freq_15', and 'freq_12'. Additionally, features from the MFCCs and MSSD sets are prominently represented among the top 10, highlighting the complementarity of the different feature types toward behive strength predictions. To further explore these differences, Figure 5.13 shows the daily variation of three of the top-10 features, namely, "freq_15", "freq_centroid_6", and "mfccs_2" across two different beehive strength conditions. These features were averaged across multiple hives, ensuring equal representation in each category, to depict their daily variation. The plots reveal distinct diurnal patterns, effectively differentiating between samples with fewer than

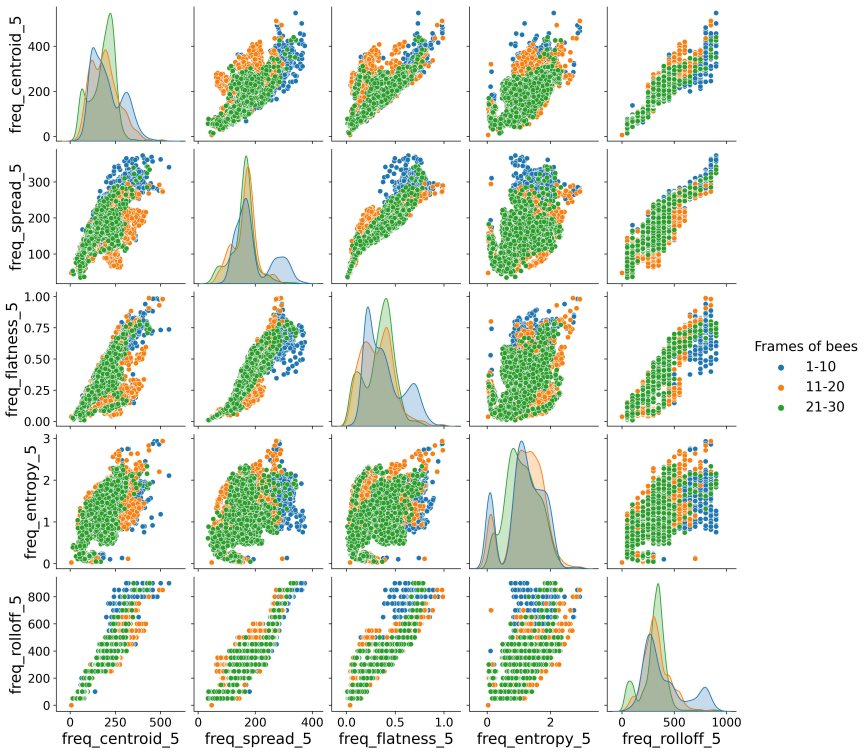


FIGURE 5.10 : A pairplot illustrating relationships among selected spectral shape descriptors features from the MSSD, including centroid, spread, flatness, entropy, and rolloff.

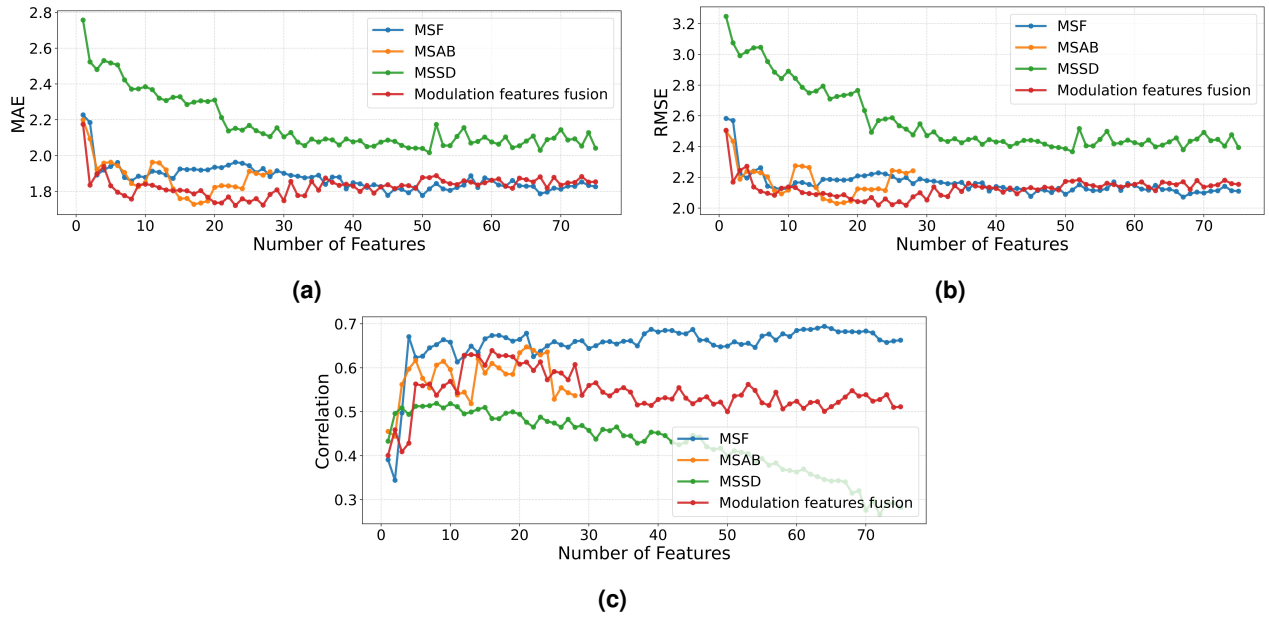


FIGURE 5.11 : Performance comparison of the modulation spectrogram features as a function of the number of top-ranked features selected in the hive-independent setup, evaluated using (a) MAE, (b) RMSE, and (c) correlation.

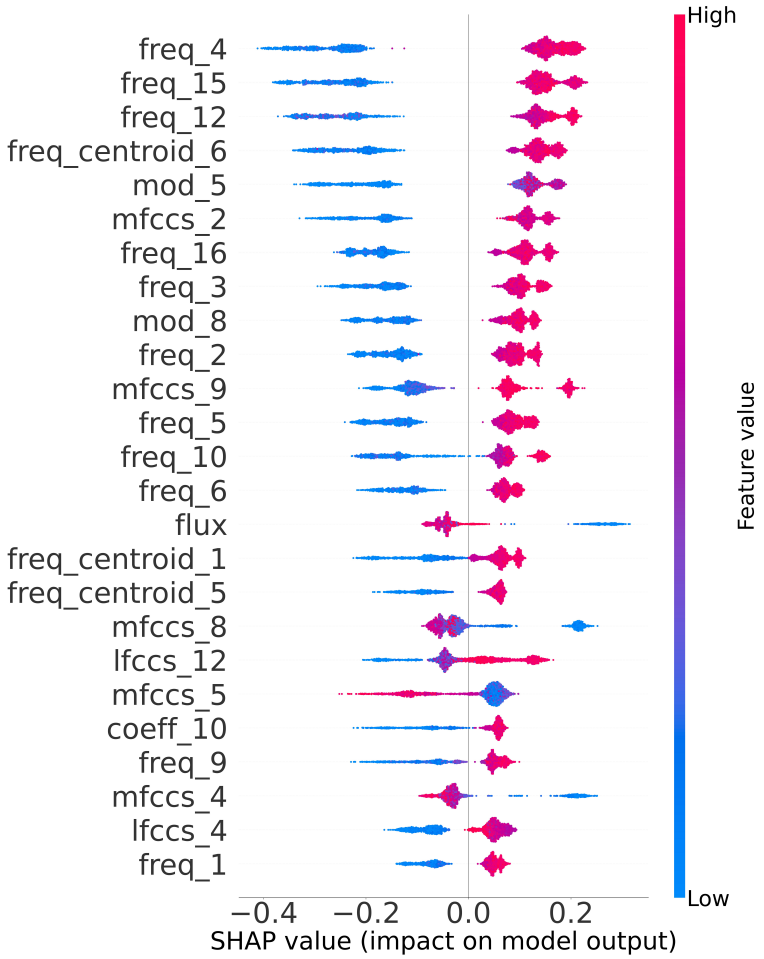
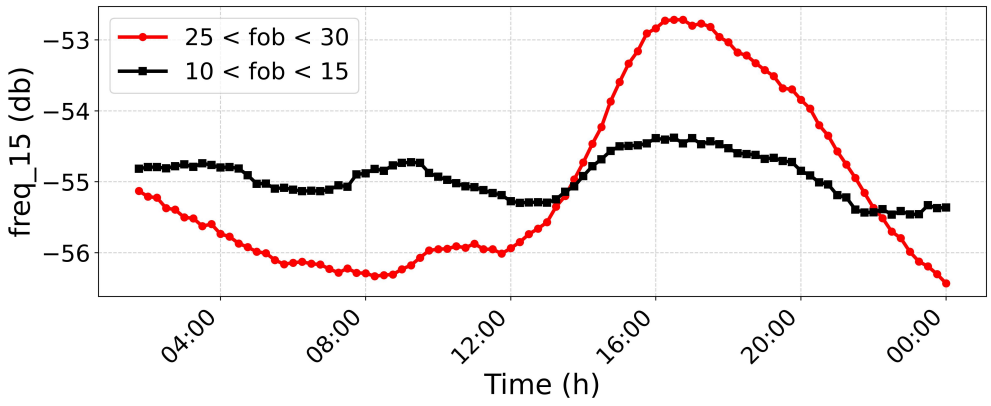


FIGURE 5.12 : SHAP value plots demonstrating the contribution of individual features to model predictions for for a combined set of all features

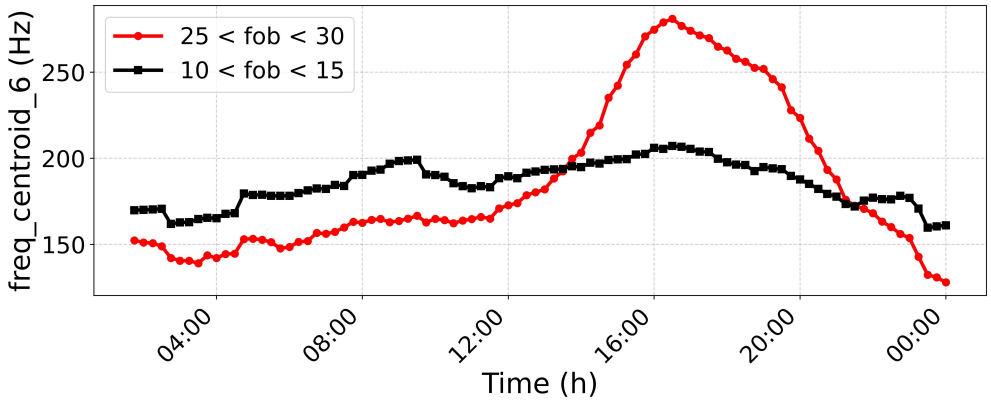
15 frames of bees and those exceeding 25 frames of bees. Notably, the modulation spectrogram-derived features, "freq_15" and "freq_centroid_6", exhibit higher energy levels and frequency in samples with more than 25 frames of bees. In contrast, "mfccs_2" demonstrates a different trend, underscoring the complementary nature of these features in capturing variations in bee activity throughout the day.

5.2.3.2 Colony Population Strength Prediction

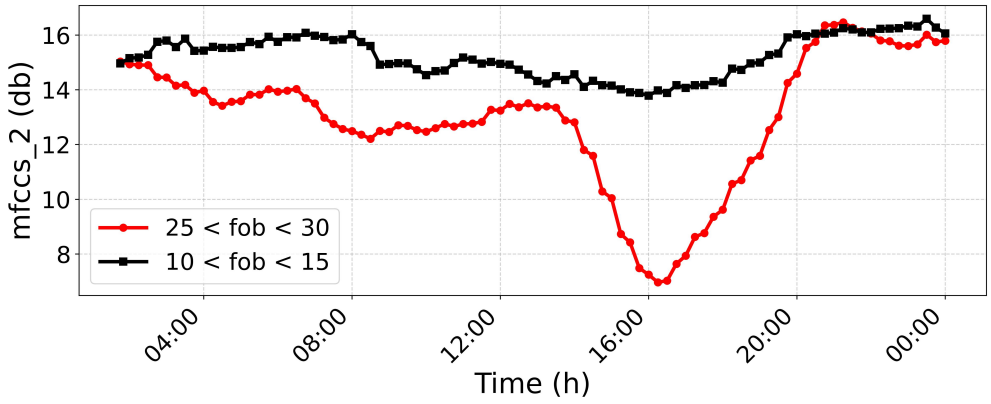
Once the best number of features is selected for each feature set, the RF regressor is then trained and tested on the remaining 14 unseen beehives. Tables 5.2 and 5.3 present the performance achieved by the RF regressor for each feature set, as well as the random classifier, both before and after enhancement under the two different dataset split scenarios for the error-based and correlation/R^2 figures of merit, respectively. Each metric is presented as an average $\pm$ standard deviation across the 5 folds.



(a)



(b)



(c)

FIGURE 5.13 : Three top-ranked features from the fusion all set : a) "freq_15", b) "freq_centroid_6", and c) "mfccs_2 averaged on hives with different number of frames on September 18th, 2021.

As can be seen, for the Random-Split condition, all metrics suggest that features with spectral amplitude subtraction consistently perform better than their counterparts without pre-processing. In particular :

TABLE 5.2 : Performance comparison of error metrics between different feature sets and data partitioning setups, with and without spectral enhancement. Bold represents the best and underline the second best average results.

Features	Random-Split					
	No pre-processing			Spectral amplitude subtraction		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Random baseline	3.98 ± 0.03	4.98 ± 0.04	22.34 ± 0.28	3.94 ± 0.04	5.01 ± 0.08	22.42 ± 0.34
LFCCs	0.56 ± 0.02	1.47 ± 0.03	4.03± 0.08	0.51 ± 0.02	1.39 ±0.06	3.73 ± 0.20
MFCCs	**0.34** ± 0.03	**1.07** ± 0.09	**2.53** ± 0.23	**0.33** ± 0.02	**1.05** ± 0.06	**2.47** ± 0.20
Spectral descriptors	1.68 ± 0.03	2.95 ± 0.06	10.73 ± 0.15	1.58 ± 0.07	2.82 ± 0.09	10.05 ± 0.41
MSF	1.76 ± 0.05	3.20 ± 0.08	10.94 ± 0.35	1.74 ± 0.04	3.17 ± 0.06	10.88 ± 0.11
MSAB	1.24 ± 0.02	2.93 ± 0.01	8.02 ± 0.38	1.19 ± 0.06	2.48 ± 0.09	7.89 ± 0.36
MSSD	1.70 ± 0.07	2.97 ± 0.12	11.26 ± 0.63	1.67 ± 0.05	2.96 ± 0.07	11.07 ± 0.53
Benchmarks fusion	0.49 ± 0.03	1.34 ± 0.08	3.65 ± 0.27	0.45 ± 0.02	1.13 ± 0.18	3.40 ± 0.23
Modulation fusion	1.68 ± 0.04	2.94 ± 0.05	10.96 ± 0.22	1.65 ± 0.05	2.92 ± 0.08	10.92 ± 0.47
Fusion all	0.47 ± 0.02	1.30 ± 0.02	3.57 ± 0.23	0.46 ± 0.04	1.25 ± 0.08	3.55 ± 0.35
	Hive-Independent					
Random baseline	4.27 ± 1.86	5.17 ± 1.83	26.19 ± 4.41	4.23 ± 1.88	5.14 ± 1.83	26.15 ± 4.58
LFCCs	4.06 ± 1.51	4.91 ± 1.36	24.59 ± 5.01	4.01 ± 1.37	4.84 ± 1.25	23.63 ± 3.39
MFCCs	3.71 ± 0.58	4.75 ± 0.49	25.65 ± 8.00	3.61 ± 0.68	4.54 ± 0.63	25.00 ± 8.70
Spectral descriptors	3.67 ± 1.09	4.77 ± 1.01	23.81 ± 6.05	3.55 ± 0.95	4.44 ± 0.96	22.30 ± 8.46
MSF	3.75 ± 1.08	4.81 ± 1.05	23.42 ± 6.70	3.71 ± 0.94	4.73 ± 0.91	23.31 ± 7.19
MSAB	**3.48** ± 0.85	**4.36** ± 0.83	**22.63** ± 6.51	**3.41** ± 1.03	**4.30** ± 1.02	**22.03** ± 5.74
MSSD	3.53 ± 0.94	4.43 ± 0.96	22.72 ± 6.91	3.42 ± 1.13	4.33 ± 1.12	22.21 ± 7.25
Benchmarks fusion	3.81 ± 0.91	4.61 ± 0.77	24.76 ± 4.94	3.73 ± 1.10	4.56 ± 0.95	24.09 ± 6.56
Modulation fusion	3.50 ± 0.98	4.49 ± 0.87	22.82 ± 4.81	3.48 ± 1.13	4.46 ± 1.11	22.42 ± 7.23
Fusion all	3.58 ± 0.96	4.47 ± 0.72	22.94 ± 4.24	3.53 ± 0.92	4.42 ± 0.90	22.41 ± 7.19

— MFCCs outperform all other feature sets with MAE and RMSE of 0.33 and 1.05, respectively. This corroborates previous findings that MFCCs are effective features for beehive monitoring tasks.

— LFCCs also show reliable performance (MAE = 0.51), achieving lower MAE and RMSE values and higher correlation than other spectral descriptors and modulation spectrogram features.

— Spectral descriptors show moderate effectiveness, especially with pre-processing, where they outperform MSF, though still lagging behind MFCCs, LFCCs, MSAB, and MSSD.

— Within the modulation spectrogram features, MSAB outperforms MSF, and MSSD across all metrics.

In the hive-independent evaluation scenario, model performance was not as significantly better as the random baselines, underscoring the substantial variability in acoustic characteristics across different hives. This variability suggests that models trained on data from specific hives struggle to generalize to unseen hives, highlighting the potential need for an initial calibration phase to account for inter-hive differences.

Despite these challenges, certain feature sets demonstrated improved generalization capabilities. Notably, the MSAB features achieved superior performance among individual feature sets. Without spectral enhancement, MSAB features yielded an MAE of 3.48 ± 0.85, RMSE of 4.36 ±

TABLE 5.3 : Performance comparison of correlation and R^2 between different feature sets and data partitioning setups, with and without spectral enhancement. Bold represents the best and underline the second best average results.

Features	Random-Split			
	No pre-processing		Spectral amplitude subtraction	
	Correlation	R^2	Correlation	R^2
Random baseline	0.87 ± 0.00	0.51 ± 0.01	0.88 ± 0.00	0.54 ± 0.00
LFCCs	0.97 ± 0.00	0.95 ± 0.00	0.98 ± 0.00	0.96 ± 0.00
MFCCs	**0.98** ± 0.00	**0.97** ± 0.00	**0.98** ± 0.00	**0.98** ± 0.00
Spectral descriptors	0.91 ± 0.00	0.82 ± 0.00	0.92 ± 0.00	0.84 ± 0.00
MSF	0.89 ± 0.00	0.79 ± 0.00	0.90 ± 0.00	0.80 ± 0.00
MSAB	0.91 ± 0.00	0.86 ± 0.00	0.93 ± 0.00	0.87 ± 0.00
MSSD	0.88 ± 0.00	0.79 ± 0.00	0.89 ± 0.00	0.81 ± 0.00
Benchmarks fusion	**0.98** ± 0.00	**0.96** ± 0.00	**0.98** ± 0.00	**0.97** ± 0.00
Modulation fusion	0.88 ± 0.01	0.77 ± 0.01	0.90 ± 0.00	0.81 ± 0.00
Fusion all	0.98 ± 0.00	0.96 ± 0.00	0.98 ± 0.00	0.96 ± 0.00
	Hive-Independent			
Random baseline	0.52 ± 0.20	-1.06 ± 3.16	0.52 ± 0.20	-1.05 ± 3.16
LFCCs	0.60 ± 0.20	-0.71 ± 1.79	0.62 ± 0.18	-0.64 ± 1.66
MFCCs	0.66 ± 0.05	0.15 ± 0.89	0.68 ± 0.06	0.20 ± 0.95
Spectral descriptors	0.62 ± 0.14	0.16 ± 0.41	0.63 ± 0.12	0.25 ± 0.40
MSF	0.68 ± 0.18	0.18 ± 0.43	0.69 ± 0.21	0.21 ± 0.56
MSAB	**0.72** ± 0.22	**0.34** ± 0.36	**0.74** ± 0.15	**0.37** ± 0.44
MSSD	0.69 ± 0.19	0.26 ± 0.14	0.71 ± 0.18	0.29 ± 0.21
Benchmarks fusion	0.64 ± 0.15	0.12 ± 0.51	0.65 ± 0.20	0.16 ± 0.48
Modulation fusion	0.70 ± 0.16	0.27 ± 0.37	0.71 ± 0.19	0.31 ± 0.46
Fusion all	0.69 ± 0.23	0.24 ± 0.67	0.71 ± 0.28	0.27 ±0.71

0.83, and MAPE of 22.63 ± 6.51. With spectral enhancement, these metrics improved to an MAE of 3.41 ± 1.03, RMSE of 4.30 ± 1.02, and MAPE of 22.03 ± 5.74. These results indicate that MSAB features contribute significantly to reducing prediction errors in unseen hives.

When considering feature fusion strategies, combining all features (fusion all) yielded the best overall performance. Without spectral enhancement, this approach resulted in an MAE of 3.58 ± 0.96, RMSE of 4.47 ± 0.72, and MAPE of 22.94 ± 4.24. With spectral enhancement, the performance further improved to an MAE of 3.53 ± 0.92, RMSE of 4.42 ± 0.90, and MAPE of 22.41 ± 7.19. These results underscore the advantage of integrating diverse feature sets to enhance model robustness and accuracy in hive-independent scenarios.

In contrast, traditional benchmark features such as LFCCs and MFCCs exhibited limited effectiveness in the hive-independent context. For instance, LFCCs without spectral enhancement achieved an MAE of 4.06 ± 1.51 and MAPE of 24.59 ± 5.01, while MFCCs yielded an MAE of 3.71 ± 0.58 and MAPE of 25.65 ± 8.00. These metrics indicate that benchmark features may not generalize well across different hives, highlighting the importance of incorporating modulation-based features for improved performance.

TABLE 5.4 : Friedman test results across evaluation metrics under random-split and hive-independent scenarios.

Metric	Random-Split			
	No pre-processing		Spectral amplitude subtraction	
	Friedman χ^2	p-value	Friedman χ^2	p-value
MAE	42.55	2.60e-06	42.56	2.59e-06
RMSE	43.56	1.70e-06	42.94	2.20e-06
MAPE	42.81	2.32e-06	43.08	2.08e-06
Correlation	42.48	2.68e-06	42.29	2.89e-06
R^2	43.17	2.00e-06	43.13	1.98e-06
	Hive-Independent			
MAE	18.42	0.031	26.89	1.45e-03
RMSE	32.64	1.54e-04	41.27	4.45e-06
MAPE	35.75	4.37e-05	14.71	0.099
Correlation	38.30	1.50e-05	40.00	7.57e-06
R^2	17.74	0.038	33.73	9.95e-05

In terms of correlation, modulation spectrogram features demonstrate notable improvements, as observed during the feature selection process and depicted in Figure 5.11. Specifically, the MSAB feature set achieves the highest average correlation of 0.73, indicating its strong predictive capability.

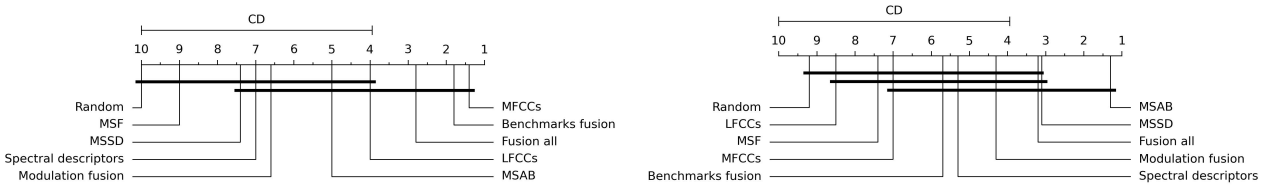
5.2.3.3 Friedman and Nemenyi Test

In this analysis, we applied the Friedman test [207, 208] across the four scenarios and five evaluation metrics to assess differences among the feature sets, and subsequently performed post-hoc pairwise comparisons using the Nemenyi test [209], which determines whether the observed differences between feature sets are statistically significant. Table 5.4 reports the results of the Friedman test comparisons for all evaluation metrics under both random-split and hive-independent scenarios. Across almost all metrics, the Friedman test statistics are significant ($p < 0.05$), indicating notable differences among the methods. The critical chi-square value for the Friedman test with $k = 10$ is 16.91 at a significance level of 0.05. Within the Friedman framework, each method is ranked in every fold and for each metric, and the average ranks across folds are reported in Table 5.5. For error-based metrics (MAE, RMSE, MAPE), lower ranks indicate better performance, whereas for predictive metrics (Correlation, R^2), higher ranks correspond to better outcomes. MSAB and MSSD achieve particularly competitive positions, especially under the hive-independent setting, suggesting that they capture meaningful information beyond standard spectral and cepstral features.

Figure 5.14 illustrates the Critical Difference (CD) diagrams of the RMSE metric for random-split and hive-independent scenarios. The CD plots provide a visual interpretation of the post-hoc

TABLE 5.5 : Average ranks of feature sets across evaluation metrics under random-split and hive-independent scenarios, with and without spectral amplitude subtraction pre-processing.

Features	Random-Split									
	No pre-processing					Spectral amplitude subtraction				
	MAE ↓	RMSE ↓	MAPE ↓	Correlation ↑	R^2 ↑	MAE ↓	RMSE ↓	MAPE ↓	Correlation ↑	R^2 ↑
Random baseline	10.0	10.0	10.0	2.0	1.0	10.0	10.0	10.0	2.0	1.0
LFCCs	4.0	4.0	4.0	7.1	7.1	3.8	3.8	3.8	7.8	7.4
MFCCs	**1.0**	**1.4**	**1.0**	**9.3**	**9.6**	**1.0**	**1.6**	**1.0**	**8.9**	**9.5**
Spectral descriptors	7.6	7.0	6.8	4.7	4.7	6.8	6.7	6.2	4.8	4.9
MSF	7.9	9.0	7.6	2.0	3.0	8.8	9.0	7.6	1.8	2.6
MSAB	5.0	5.0	5.0	5.9	6.0	5.2	5.0	5.0	6.0	6.0
MSSD	7.8	7.4	8.2	2.4	2.9	7.0	7.5	8.0	2.9	3.0
Benchmarks fusion	2.6	1.8	2.0	9.2	9.2	2.6	2.1	2.4	9.1	9.0
Modulation fusion	6.7	6.6	7.4	3.7	3.4	7.2	6.8	8.2	3.5	3.5
Fusion all	2.4	2.8	3.0	8.4	8.1	2.6	2.5	2.8	8.2	8.1
	Hive-Independent									
Random baseline	9.3	9.2	9.4	1.0	5.6	9.5	10.0	8.0	1.0	1.0
LFCCs	6.8	8.5	6.6	3.4	5.2	7.2	8.8	7.6	2.6	2.0
MFCCs	4.9	7.0	9.2	6.1	3.5	5.2	6.0	6.4	5.8	4.9
Spectral descriptors	6.0	5.3	4.9	2.9	4.5	6.8	4.1	5.0	2.6	7.3
MSF	6.4	7.4	5.2	5.5	3.0	7.2	8.2	5.7	5.4	5.6
MSAB	**3.0**	**1.3**	**1.9**	**10.0**	**8.8**	**2.0**	**1.7**	**4.5**	**9.4**	**8.9**
MSSD	4.3	3.1	2.6	7.3	6.5	4.0	2.2	5.7	8.2	7.4
Benchmarks fusion	6.9	5.7	8.0	3.7	3.7	6.5	6.6	6.2	4.2	3.6
Modulation fusion	3.8	4.3	3.4	8.7	6.6	3.8	4.5	3.2	8.0	7.7
Fusion all	3.6	3.2	3.8	6.4	7.6	2.8	2.9	2.7	7.8	6.6



(a) **(b)**

FIGURE 5.14 : Critical Difference (CD) plot illustrating the results of the Nemenyi post-hoc test following the Friedman test on the Root Mean Square Error (RMSE) of ten features for a) random-split and b) hive-independent scenario. The features are ranked by their average performance across the five folds. Features grouped by a thick horizontal bar exhibit no statistically significant difference in RMSE performance ($\alpha = 0.05$). Features with lower ranks (towards 1) demonstrate superior performance.

comparisons : methods connected by a horizontal line are not significantly different, whereas those separated beyond the critical distance exhibit statistically distinct performance. Under the random-split scenario, MFCCs achieved the best average rank and were significantly better than features such as LFCCs and MSF as reflected by their clear separation in the CD diagram. In contrast, under the hive-independent scenario, MSAB (and occasionally MSSD) clustered separately from baseline features, indicating significantly stronger performance in cross-colony generalization.

5.2.4 Brief Discussion

The identified top features, including ‘freq_4’ (150–200 Hz), ‘freq_9’ (400–450 Hz), and ’‘freq_10’ (450–500 Hz) from MSAB align well with established frequency ranges associated with key honey bee behaviors and colony states, especially related to hive strength. For instance, the 150–200 Hz band corresponds to the wingbeat frequencies of individual bees, typically ranging from 150 to 300 Hz, with a prominent peak around 200 Hz [17]. Moreover, specific queen bee vocalizations correspond to these frequency bands. Tooting, produced by emerged virgin queens, typically starts around 400 Hz and can increase to over 500 Hz as the queen matures. Quacking, emitted by queens still confined within their cells, generally occurs around 350 Hz [140]. These signals play crucial roles in colony dynamics, such as coordinating swarming behavior and establishing dominance hierarchies among queens.

To highlight the noise robustness of the proposed modulation spectrum, Figure 5.15 and 5.16 show the spectrogram and the modulation spectrogram, respectively, of a clean recording from a sunny day and noisy recording from a rainy day. As can be seen, while the rain-generated noise impacts most of the bands of interests in the spectrogram, in the modulation spectrogram some regions of interest remain less impacted. Representative examples include features ’mod_6’, ’mod_7,’ selected from MSAB with modulation frequency of 25-30 Hz and 30-35 Hz, respectively, and ’mod_flatness_18 and ’freq_spread_5,’ selected from MSSD with frequency of 850-900 Hz, and modulation frequency of 25-30 Hz, respectively. These regions of the modulation spectrogram convey useful information for beehive strength characterization and are less affected by environmental noise sources.

Overall, while modulation spectral analysis for beehive acoustics is in its infancy, our initial findings suggest that incorporating modulation features can enhance performance, both in terms of environmental robustness, as well as generalizability across unseen hives. The alignment of our top features with established literature on hive acoustics indicates that modulation spectral analysis may offer a promising complementary alternative, potentially improving resilience to noise and other environmental factors, showing increased robustness for “in the wild” beehive monitoring.

5.3 Non-Invasive Prediction of Varroa Mite Infestation in Honey bee Colonies

One of the most devastating threats to honey bee health is the Varroa destructor mite, an external parasitic mite that has spread to nearly every continent. 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 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.

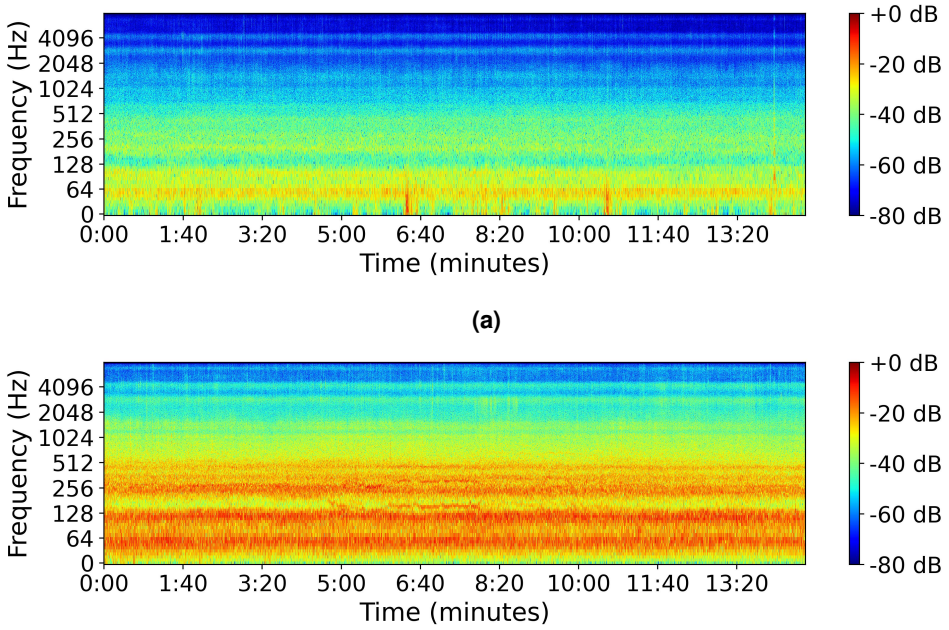


FIGURE 5.15 : Spectrograms of a a) 15-minute clean, and b) noisy beehive audio signal.

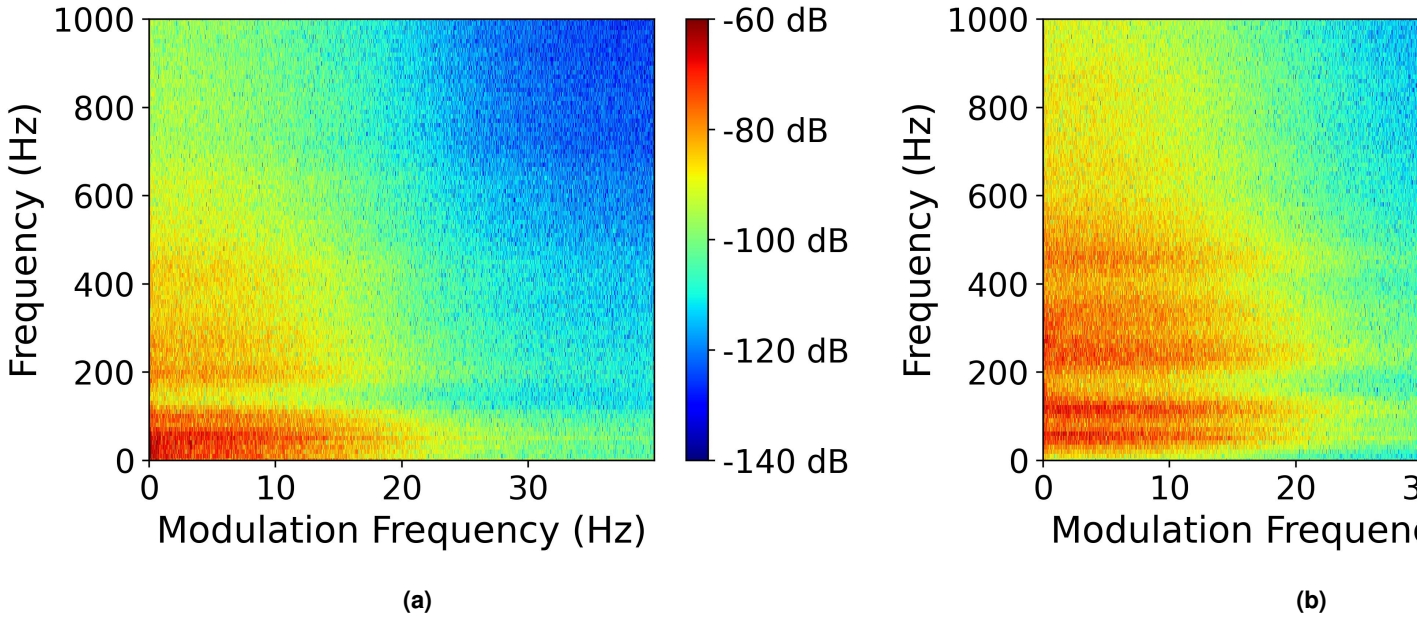


FIGURE 5.16 : Modulation spectrograms of a a) 15-minute clean, and b) noisy beehive audio signal.

For these reasons, early detection and accurate prediction of Varroa infestations are central to sustainable beekeeping practices and long-term colony survival.

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—centroid, spread, skewness, kurtosis, entropy, rolloff, 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.

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 Naive Bayes (NB) [210], Random Forests (RF) [211], Gradient Boosting (GB) [212], k-Nearest Neighbors (KNN) [213], Support Vector Machines (SVM) [162], a Multilayer Perceptron (MLP) [214, 215], 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 train-test splits that may leak information between samples from the same hive, we adopt a strategy where entire hives are held out as test sets. This more closely reflects real-world deployment scenarios, where a model trained on historical data from some hives is used to predict unseen data from different hives.

5.3.1 Proposed Methodology

5.3.1.1 Dataset

In this study, we use the UrBAN audio datatset [170, 180]. This dataset includes different measurement such as beehive strength and Varroa mite infestation. The varroa mite infestations have been measured using alcohol wash method [177] in some of the beehives. Figure 5.17 shows the heatmaps indicating the amount of Varroa mite infestation in each measurement. Figure 5.18, shows the 24-hour spectrograms of a Varroa-infected beehive with Percentage of Varroa Mite Infestation (PVMI) value of 38 and a non-infested beehive on September 30, 2022. As the spectrograms show, the infected beehive shows lower intensity in specific frequency bands (between 100 and 500 Hz), thus corroborating the findings in [116].

5.3.1.2 Feature Extraction

Proposed We propose nine spectral shape descriptors – centroid, spread, skewness, kurtosis, entropy, rolloff, flatness, crest, and flux – computed from both the conventional and the modula-

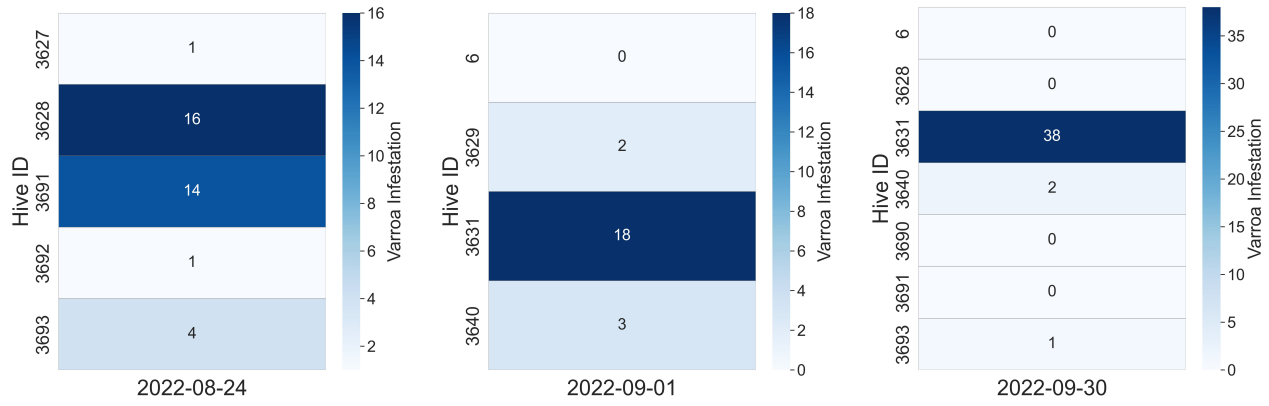


FIGURE 5.17 : Varroa mite infestation measured on (a) August 24th, (b) September 1st, and (c) September 30th, 2022.

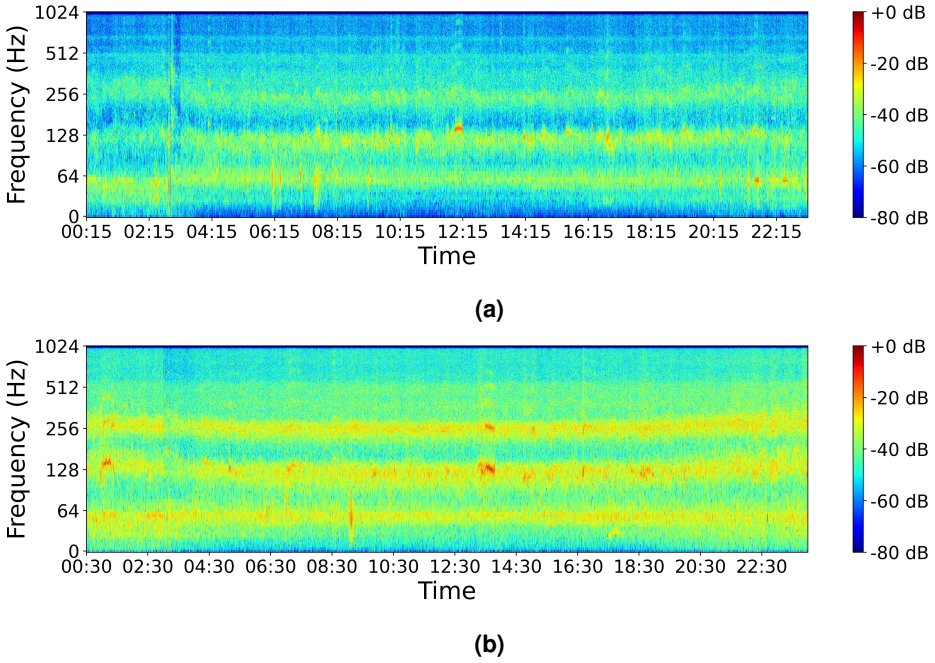


FIGURE 5.18 : 24-h spectrograms of a a) Varroa infected beehive, and b) a healthy beehive on September 30th, 2022.

tion 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 signal variability over time.

Figure 5.19 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. Spectral Shape Descriptors (SSD) : The magnitude spectrogram is used directly to compute nine spectral shape descriptors : centroid, spread, skewness, kurtosis, entropy, rolloff, flatness, crest, and flux to capture the short-term frequency structure of the hive acoustics [216].
2. Modulation Spectral Shape Descriptors (MSSD) : 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 [198] 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. To generate the modulation spectrogram, a second STFT was applied to the temporal evolution of each frequency bin's magnitude using a 60-second window without overlap, producing a 3D time–frequency–modulation representation. The 60-second 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 2D 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.

Figures 5.20 (a) and (b) show 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.

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-minute audio segment is treated as a separate sample for analysis. To increase the sample size

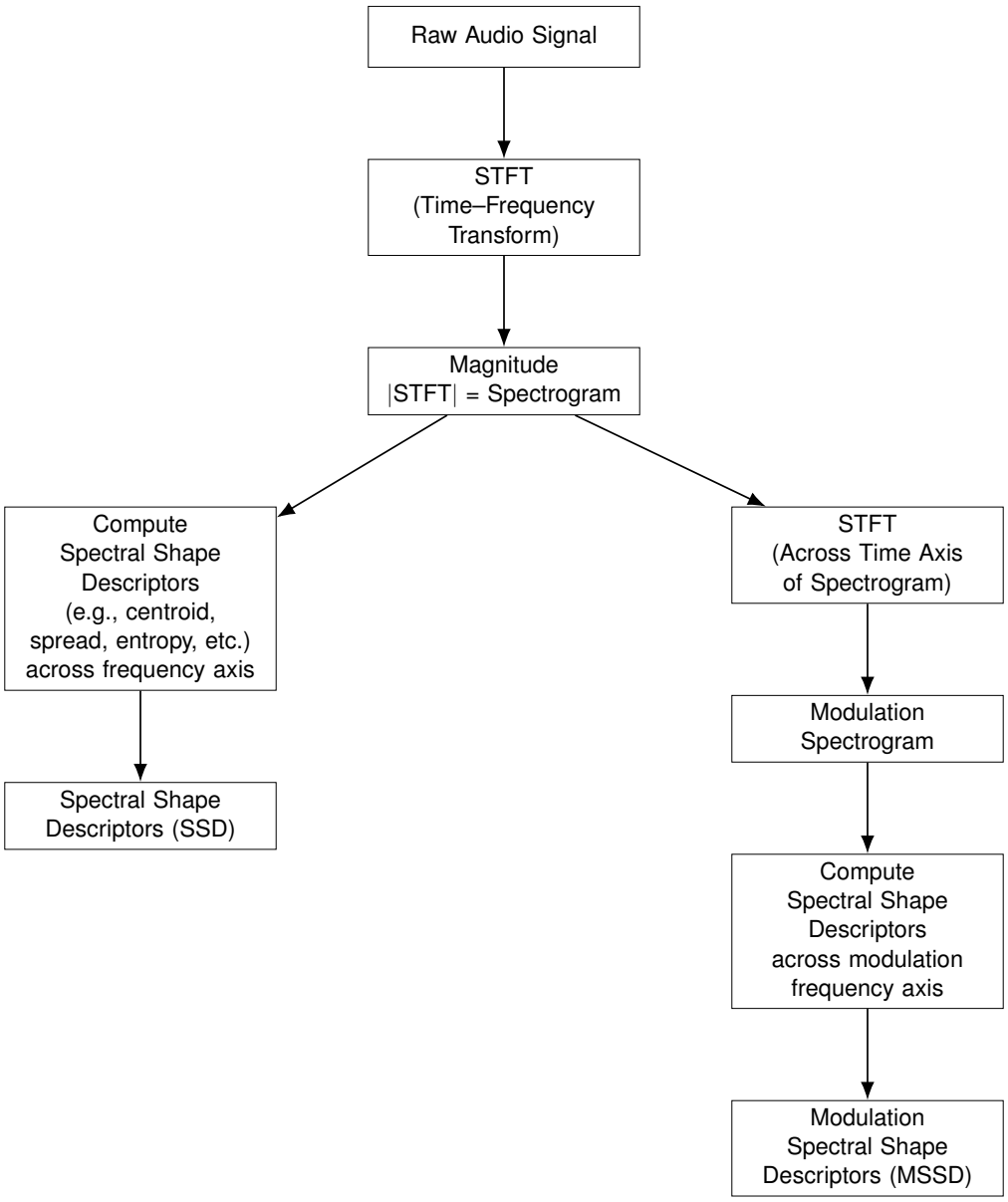


FIGURE 5.19 : Block diagram of the proposed acoustic feature extraction pipeline. The raw hive audio is first transformed into a time–frequency representation via STFT, followed by magnitude computation to obtain the spectrogram. From this, spectral shape descriptors (SSD) are computed across the frequency 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 modulation spectral shape descriptors (MSSD) are computed across the modulation frequency axis to capture slower temporal modulations.

and capture temporal variability, we include acoustic data not only from the day on which the Varroa mite inspection was performed (as illustrated in Figure 5.17), but also from the preceding day.

Benchmarks To evaluate the performance of the proposed features, two established feature sets are included as benchmarks for comparison. Mel-frequency cepstral coefficients (MFCCs) approximate human auditory perception by applying a mel-scale frequency transformation and have been extensively employed in precision beekeeping applications [52]. In this study, we extract 12 MFCCs

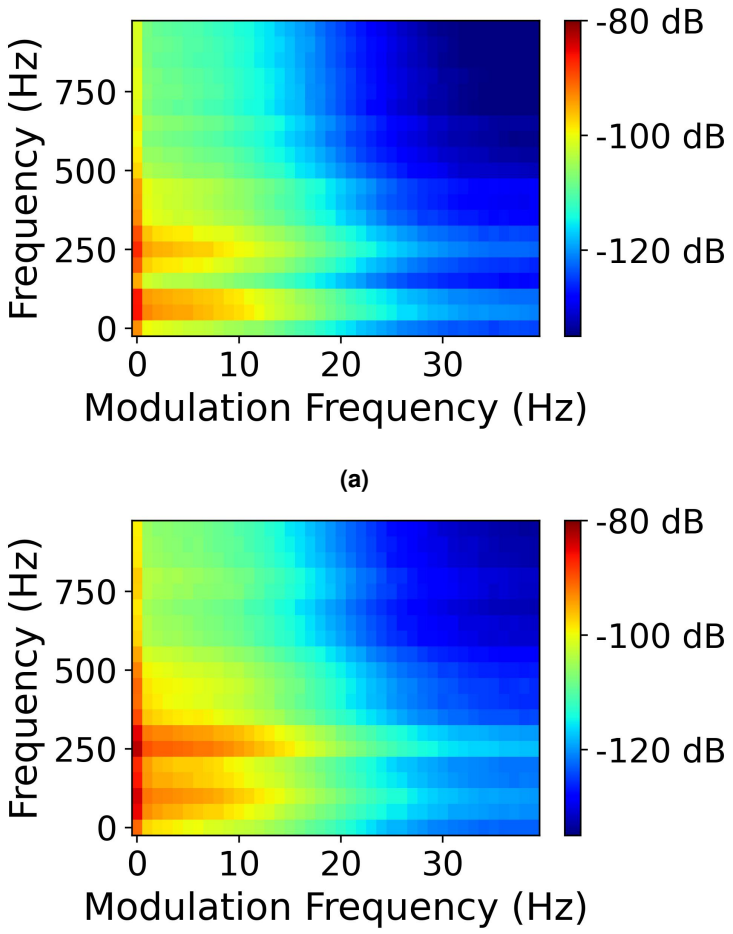


FIGURE 5.20 : Modulation spectrograms of a) a Varroa infected beehive, and b) a healthy beehive.

along with the zeroth coefficient, which captures signal energy, using 26 mel-scale filters. Additionally, linear-frequency cepstral coefficients (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 uniform frequency resolution across the spectrum, in contrast to the perceptual warping of the mel scale.

5.3.1.3 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 :

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

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 2–5%, as colonies exceeding this infestation rate are at significant risk of rapid Varroa population growth, colony weakening, and potential collapse if left untreated [217, 218, 219]. 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 5.6. To ensure robust evaluation of the models, a nested 3-fold cross-validation is used ensuring hive-independence. This cross-validation process ensures that every data point is included in the test set exactly once, reducing the risk of overfitting and providing a more reliable estimate of model performance. The choice of three folds is motivated by the limited number of available hives (seven in total) : more folds would result in test sets with only a single hive, producing unstable estimates, whereas fewer folds (e.g., two) would not sufficiently capture inter-hive 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 between 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 5.7 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. The binary cross entropy is used as loss function.

TABLE 5.6 : 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

The learning rate is initialized at 0.0001 and is dynamically adjusted using a Reduce-on-Plateau scheduling strategy, where the learning rate is reduced when the validation loss stops improving.

TABLE 5.7 : 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

5.3.1.4 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.

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.

5.3.2 Experimental Results

5.3.2.1 Features Analysis

Conventional Spectrogram Figure 5.21 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, rolloff, and entropy, suggesting that higher levels 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 non-dominant frequencies diminishes. The correlations with centroid, spread, and rolloff 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

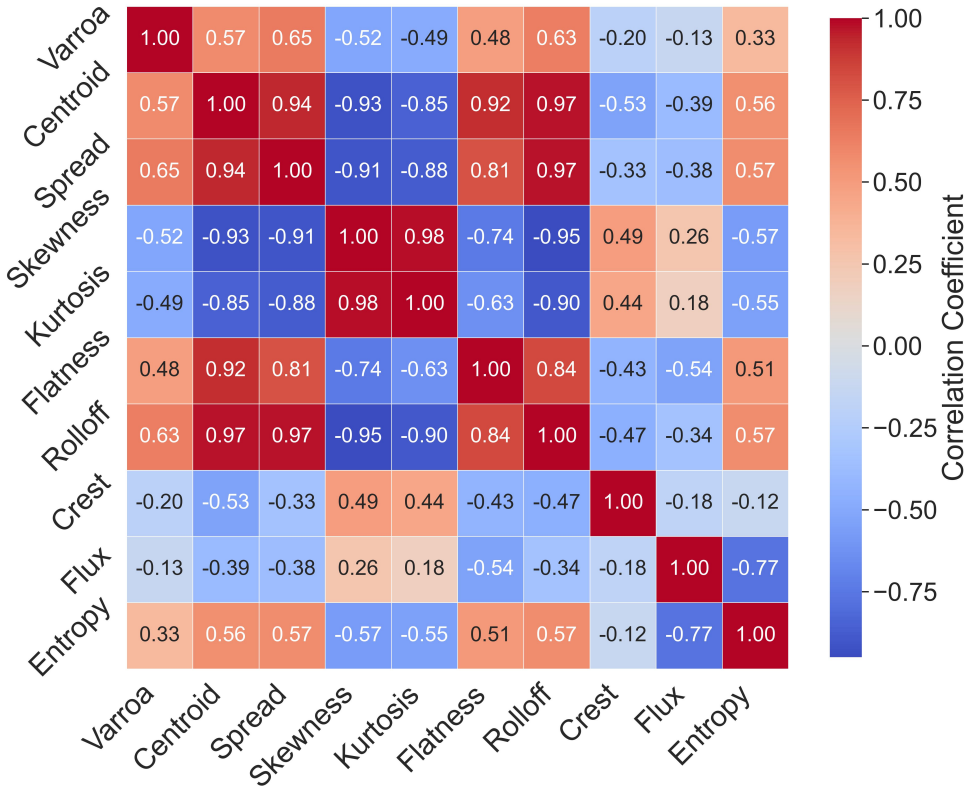


FIGURE 5.21 : Pearson correlation between all pairs of spectral descriptors extracted from the conventional spectrogram and the Varroa label.

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.

Figure 5.22 illustrates a pairwise comparison of spectral shape descriptors features, stratified by the Varroa infestation label. To ensure 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, rolloff, and entropy.

To visualize the temporal dynamics of the two proposed spectral shape descriptors, we plotted their smoothed values over a 1-day period (September 1, 2022) for two representative hives :

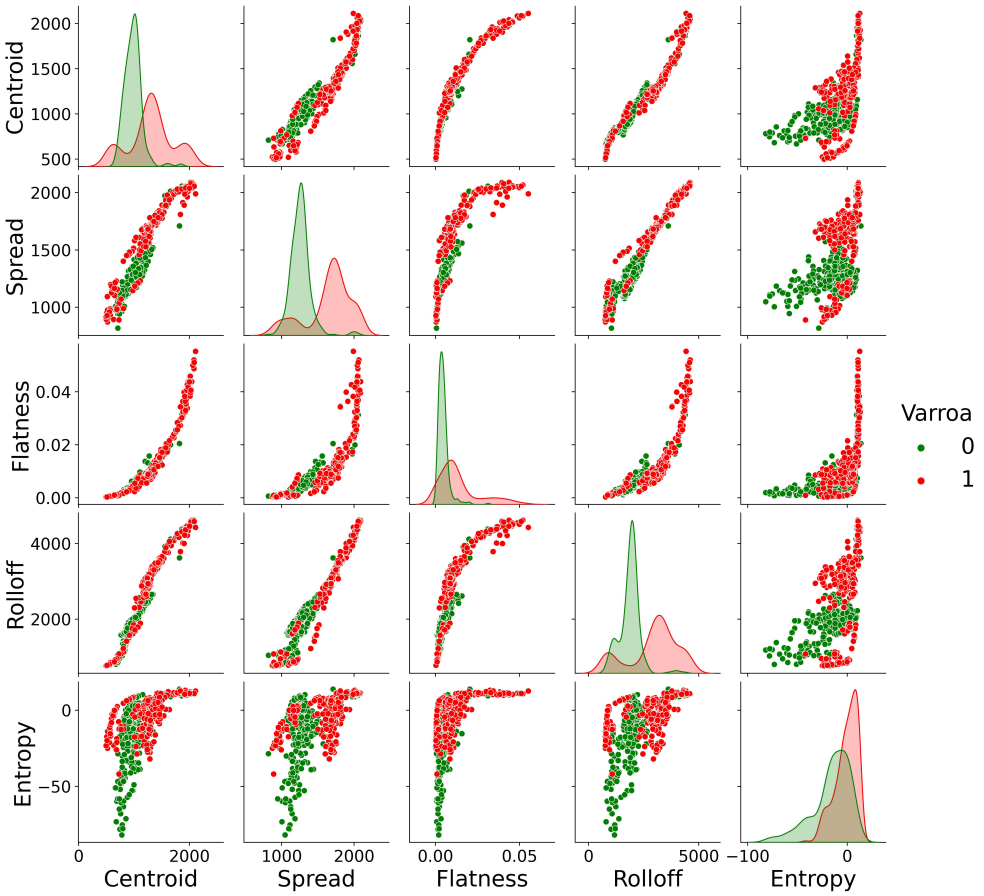
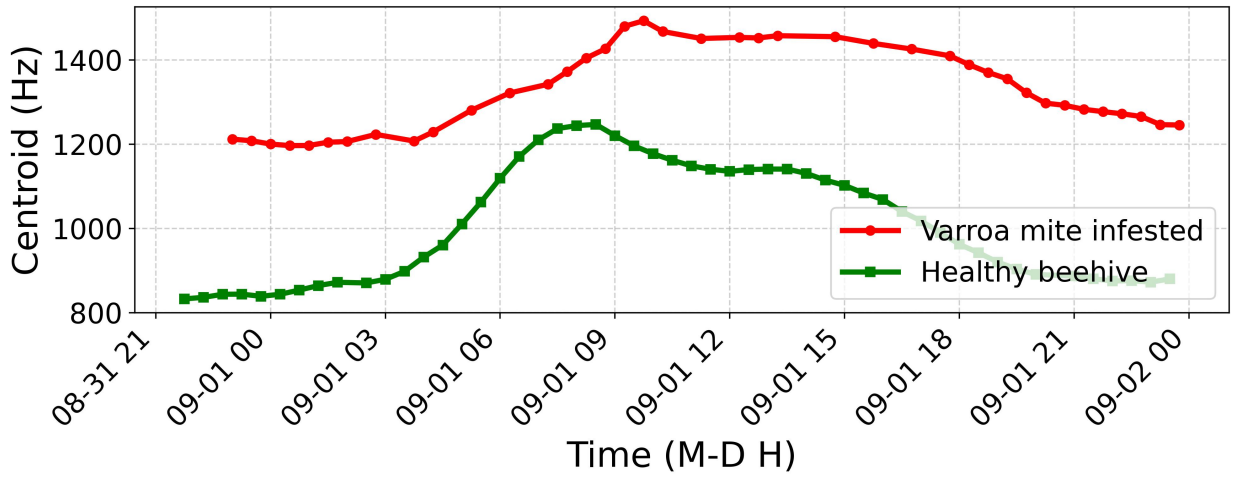


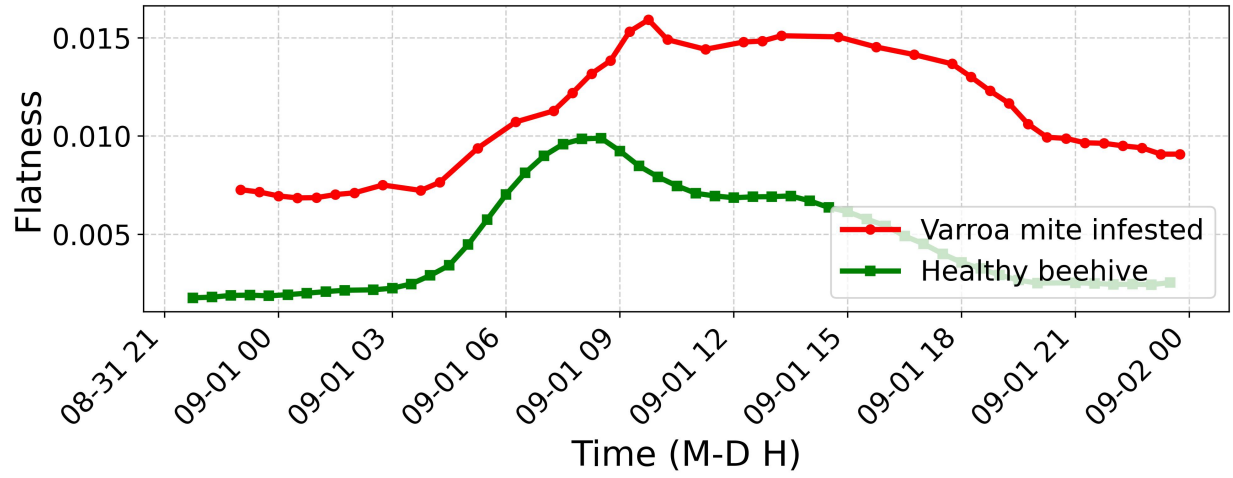
FIGURE 5.22 : A pairplot illustrating the relationships among spectral shape descriptors : centroid, spread, flatness, entropy, and rolloff. Samples with Varroa label equal to 0 represent healthy hives, while those with Varroa label equal to 1 indicate infestation.

one exhibiting a Varroa mite infestation (Hive ID : 3631) and one healthy hive (Hive ID : 6) in Figure 5.23. Each curve represents an 8-point moving average for a specific feature, allowing us to observe trends and fluctuations over time. As shown in the Figures 5.23a and 5.23b 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.

Modulation Spectrogram Figure 5.24, 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 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.



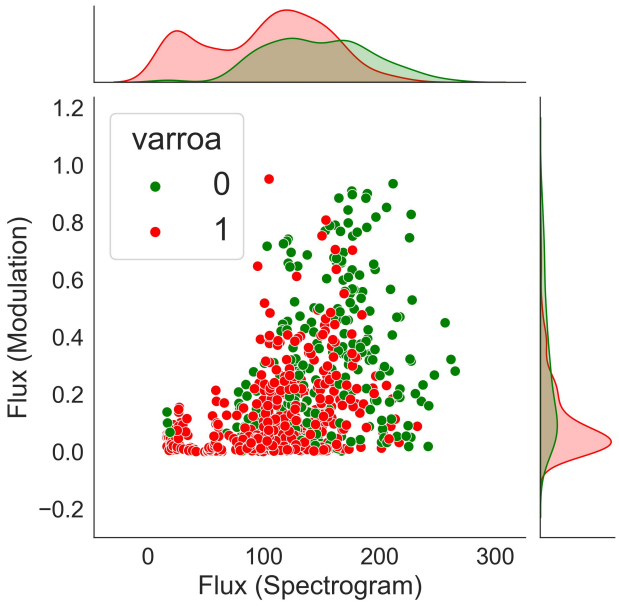
(a)



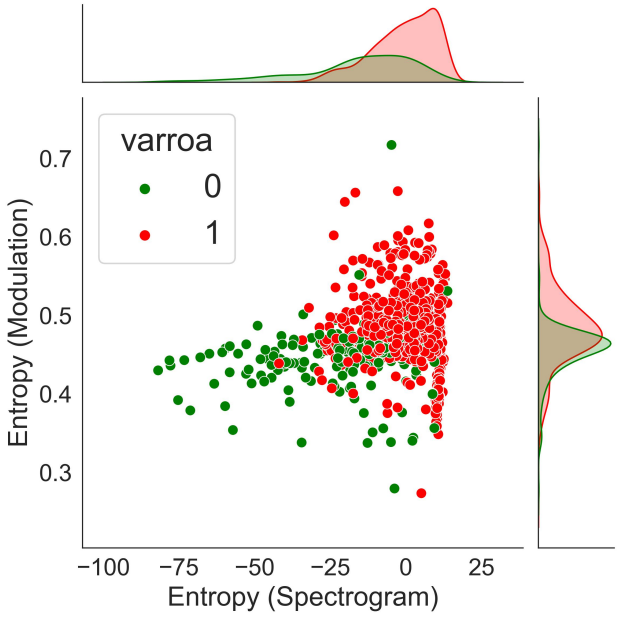
(b)

FIGURE 5.23 : The temporal dynamics of the conventional spectrogram a) centroid, and b) flatness measured across the entire day on September 1, 2022.

Figure 5.25 shows the temporal dynamics of the two proposed spectral shape descriptors 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 Figure 5.23a, the modulation centroid and flatness shown in Figures 5.25a and 5.25b 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 temporal modulation dynamics (modulation centroid), but in different ways.



(a)

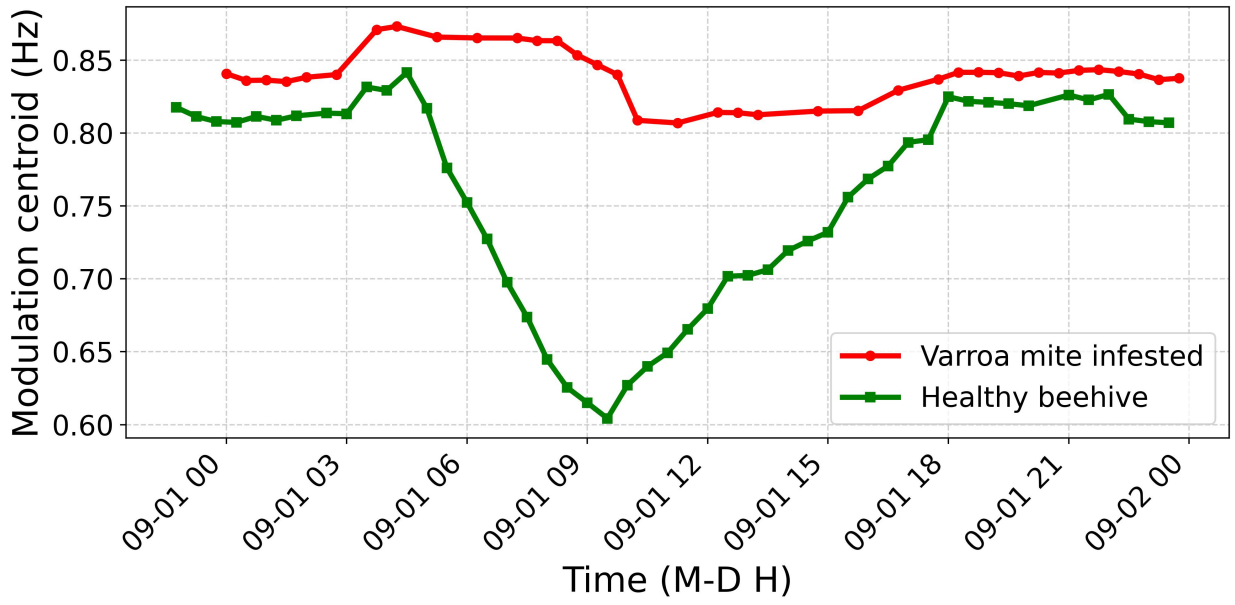


(b)

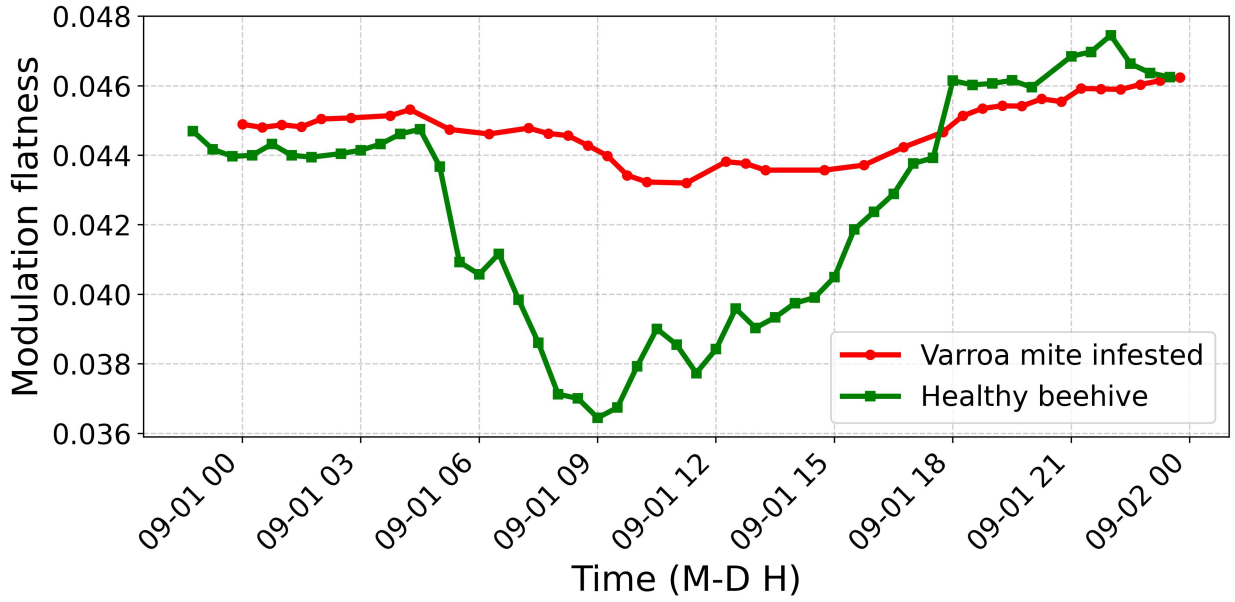
FIGURE 5.24 : A pairplot illustrating the relationship among conventional and modulation spectral shape descriptors : a) flux and b) entropy. Samples with Varroa equal to 0 represent healthy hives, while those with Varroa equal to 1 indicate infestation.

5.3.2.2 Classification Performance

Table 5.8 presents a comparative evaluation of the classification performance across the different benchmark and proposed feature sets, as well as tested classifiers. Each metric is an average across the 3 folds. As can be seen, the benchmark MFCCs and LFCCs showed modest performance across all classifiers, with SVM and MLP generally outperforming others. For ins-



(a)



(b)

FIGURE 5.25 : The temporal dynamics of the modulation spectrogram a) centroid, and b) flatness measured across the entire day on September 1, 2022.

tance, MFCCs with SVM reached a 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 a ROC-AUC of 0.871 and F1-score of 0.832, with high precision (0.867), indicating strong discriminative power. This underscores the value of capturing spec-

tral 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.

Lastly, 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 Figures 5.24, 5.23, and 5.25. The best results were obtained using SVM, which achieved a ROC-AUC of 0.874, a F1-score of 0.859, precision of 0.863, and recall of 0.862. Several models showed clear improvements in ROC-AUC with the fused feature set. For instance, Random Forest increased from 0.678 to 0.756, 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, shown in Figure 5.26. 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 behaviour. This aggregated matrix reflects the total counts of true positives, true negatives, false positives, and false negatives observed across all test 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.

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.

5.3.2.3 Comparative Performance of CNN and Traditional Machine-Learning Models

Table 5.9 presents a comparative evaluation of classification performance across different feature representations. In particular, MFCCs, LFCCs, and spectrograms, each representing two-dimensional time–frequency features, are used as inputs to the CNN model. 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.

TABLE 5.8 : 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

Moreover, a clear difference can be observed in the performance of features between the two experimental setups. For example, when used as two-dimensional time–frequency representations with a CNN model (Table 5.9), MFCCs achieve moderate performance (ROC-AUC = 0.593, F1 = 0.573). In contrast, when MFCCs are averaged over time and used as one-dimensional features with traditional classifiers (Table 5.8), the performance varies significantly across models, with the best ROC-AUC reaching 0.736 using Random Forest, but generally lower and less consistent F1-scores. These findings suggest that preserving the temporal structure of MFCCs and leveraging convolutional architectures provides more stable and meaningful representations compared to

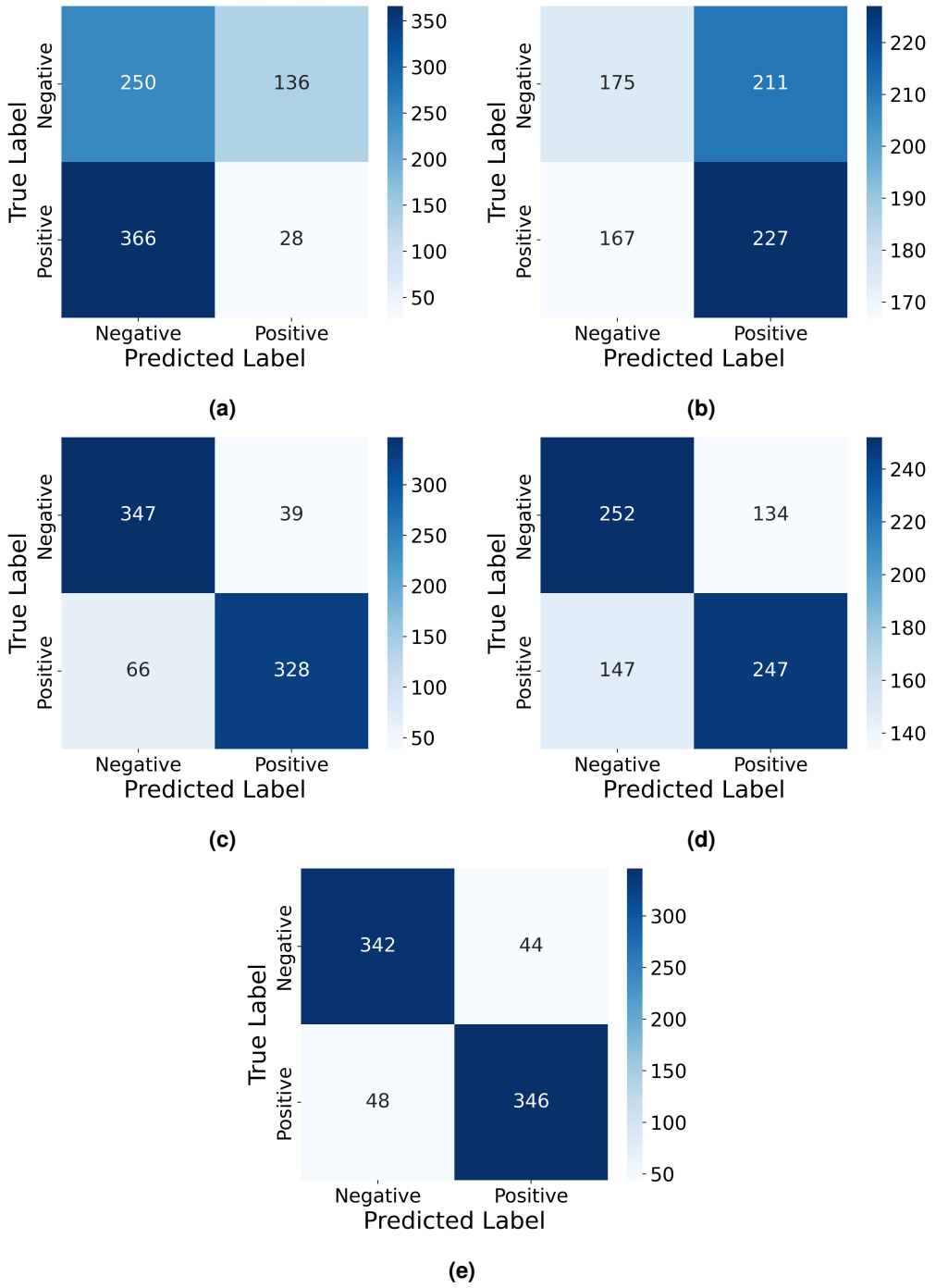


FIGURE 5.26 : Aggregated confusion matrices for each feature set evaluated across all three test folds for the SVM model. Each matrix summarizes the total true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP) obtained for : (a) MFCCs, (b) LFCCs, (c) SSD, (d) MSSD, and (e) the proposed SSD+MSSD feature set.

time-averaged features. Temporal averaging likely removes important dynamic information, limiting the ability of classical models to capture discriminative acoustic patterns.

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 com-

TABLE 5.9 : 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

putationally 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 Naïve Bayes—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 suggests that, under the present experimental conditions and given the relatively lightweight CNN architecture considered, classical machine-learning techniques combined with carefully designed feature representations can offer competitive or superior performance, along with improved interpretability and computational efficiency. However, broader conclusions regarding deep learning would require evaluation of more advanced architectures.

5.3.2.4 Feature Contribution Analysis Using SHAP

To quantify the individual and complementary contributions of SSD and MSSD, we conducted a SHAP-based (SHapley Additive exPlanations) [206] ablation analysis. While the previous section 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 a 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.

Figures 5.27 show 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, 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.

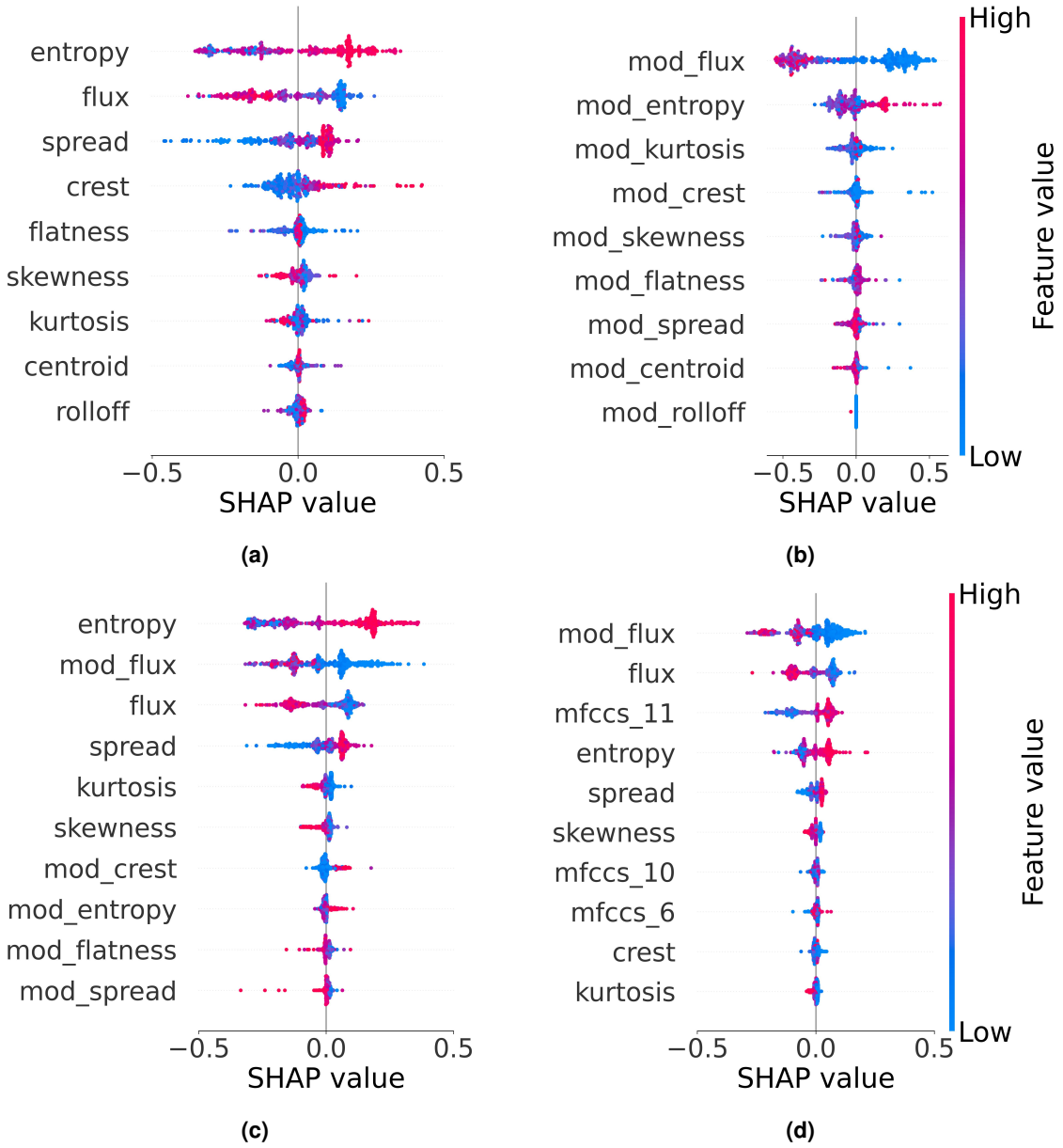


FIGURE 5.27 : SHAP summary plots illustrating the contribution of individual features to the model's predictions for (a) SSD, (b) MSSD, (c) SSD+MSSD, and (d) SSD+MSSD+MFCCs. Features are ranked by their mean absolute SHAP value, indicating overall importance, while color encodes the relative feature value (low to high) and its directional impact on the model output.

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, 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. 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.

5.4 Conclusion

In this chapter, we introduced the use of modulation spectrogram features for improved beehive strength prediction and Varroa mite detection across unseen conditions. Features extracted from the modulation spectrogram showed lower error rates and higher correlations, showing their increased robustness for “in the wild” beehive monitoring. For beehive strength prediction, two different feature selection mRMR and SHAP were used. Once the top features were selected, a RF regressor was trained and tested. In the task of detecting a Varroa mite infestation, spectral shape descriptors, extracted from both the conventional and modulation spectrograms were proposed. 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.

6 DEEP LEARNING WITH MODULATION SPECTROGRAM AND TENSORGRAM

6.1 Preamble

This chapter is based on material from the following manuscript submitted to a peer-reviewed journal : "Improved Monitoring of Honey bee Colony Strength via Audio IoT Sensors, Modulation Tensorgrams and Recurrent Neural Networks [220]"

6.2 Introduction

In the previous chapter, we introduced the modulation spectrogram and proposed several feature sets derived from it for predicting honey bee colony strength and detecting Varroa mite infestation. We demonstrated that these features, when combined with classical machine learning models such as Random Forests and Support Vector Machines, achieve strong predictive performance. Building on these findings, this chapter explores the use of deep learning models to automatically learn discriminative representations from modulation-based inputs and further improve prediction accuracy.

While traditional hive monitoring tools relied on conventional machine learning classifiers [52], recently, there has been a rise on the use of deep neural networks, specifically convolutional neural networks (CNNs). CNNs were originally developed for image-like inputs, thus are well suited for spectrograms or mel-scaled spectrograms. The work in [48] employed a CNN with four convolutional layers to detect queenless hives using the spectrogram as input. In another study, CNNs were successfully applied to detect honey bee piping signals, a key indicator of colony reproductive dynamics [140]. Similarly, the work in [54] showed that CNN-based classifiers outperformed traditional machine learning models across several beehive monitoring tasks.

In parallel to these developments, the so-called modulation spectral features emerged to better capture the rhythmic and temporal patterns in beehive acoustics. These features highlight modulations in signal energy across time and frequency, making them well-suited for identifying the repetitive vibratory patterns typical of bee colonies and separating these patterns from unwanted noise sources and artifacts. In prior work, we proposed to handcraft some features from the modulation spectrum and we showed that such features were complementary to spectrogram-based ones using traditional machine learning classifiers, such as random forests [196]. Here, we build on this foundational work and explore the use of the raw modulation spectrogram as input to a CNNs and a convolutional recurrent deep neural networks (CRDNNs). This approach allows us to exploit the spatial and temporal richness of the modulation representations, as well as their

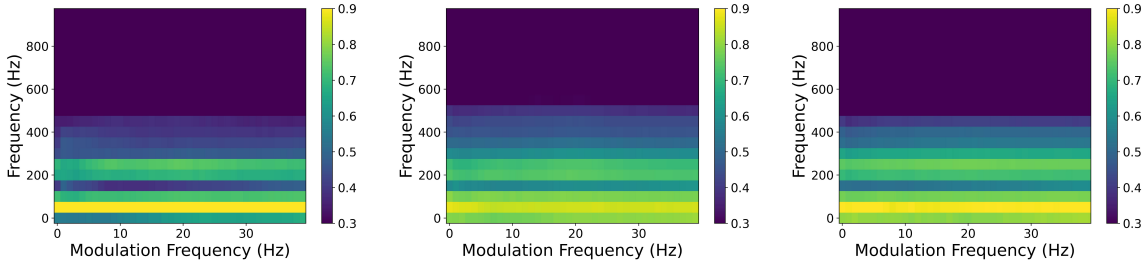


FIGURE 6.1 : Modulation spectrograms averaged on samples with a) $1 \leq$ ***fobs*** ≤ 10**, b)** $11 \leq$ ***fobs*** ≤ 20**, and c)** $21 \leq$ ***fobs*** ≤ 30**.**

noise-robustness properties, while leveraging the modeling power of deep neural networks. Our experiments show the proposed method generalizing to unseen conditions more efficiently than several benchmark methods, thus helping address a critical issue in remote beehive monitoring - generalizability across unseen environments [72, 48].

6.3 Methods

6.3.1 Modulation Spectrum

In many acoustic classification tasks, traditional spectrograms struggle in noisy environments. This limitation arises because both signal and noise are distributed across the same time-frequency space, making separation difficult. To address this limitation, the modulation spectrogram and its time-resolved counterpart, the modulation tensorgram, have been proposed [198]. Figure 5.1 depicts the signal processing steps involved in the computation of the modulation tensorgram.

To illustrate some representative modulation spectrograms, Figure 6.1 (a)-(c) shows an average representation for beehives of varying strengths, namely with frames of bees (fob) ranging from $1 \leq$ *fobs* ≤ 10, $11 \leq$ *fobs* ≤ 20, and $21 \leq$ *fobs* ≤ 30, respectively. For these plots, all audio recordings were taken between 8-11pm.

6.3.2 Discriminative power of the modulation spectrum

To measure the discriminative power of the modulation spectrogram, we compute the Fisher ratio (F-ratio) between different hive strength groups. The F-ratio quantifies how well a given feature distinguishes between two classes by comparing two types of variance : the between-group variance, which captures the separation between group means, and the within-group variance, which reflects variability within each group. A higher F-ratio value indicates greater discriminability.

To compute the F-ratio, the number of frames of bees (*fobs*) was divided into three population categories : Category 1 includes $1 \leq$ *fobs* ≤ 10, Category 2 includes $11 \leq$ *fobs* ≤ 20, and Ca-

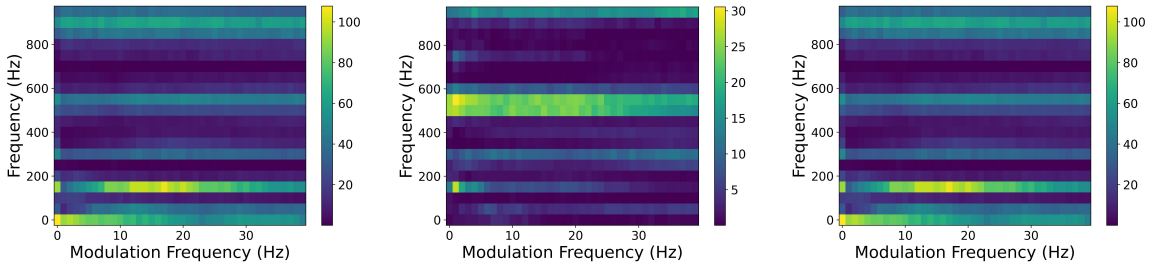


FIGURE 6.2 : The number of frames of bees (*fobs*) is grouped into three categories : Category 1 corresponds to $1 \leq$ *fobs* ≤ 10, Category 2 to $11 \leq$ *fobs* ≤ 20, and Category 3 to $21 \leq$ *fobs* ≤ 30. F-ratio maps comparing the following pairs of categories : a) Category 1 vs. Category 2, b) Category 2 vs. Category 3, and c) Category 1 vs. Category 3.

tegory 3 includes $21 \leq$ *fobs* ≤ 30. These three categories correspond to the conditions shown in Figures 6.1 (a)-(c). Figure 6.2 presents the resulting F-ratio maps for the following pairwise comparisons : (a) Category 1 vs. Category 2, (b) Category 2 vs. Category 3, and (c) Category 1 vs. Category 3. These maps highlight the modulation–spectral regions where the differences in energy are most pronounced across varying levels of bee activity. In particular, the lower-frequency regions around 150–200 Hz, as well as the band near 500–600 Hz, show the strongest discriminative power, indicating that changes in these modulation components should be closely associated with increases in colony activity and could be leveraged for automated beehive strength monitoring tasks.

6.3.3 Deep neural network model architectures

To model the acoustic patterns of beehive audio, we implemented several deep learning architectures capable of learning from 2D and 3D input representations. These models are optimized to handle spectrograms, modulation spectrograms, and modulation tensorgrams. The architectures tested include a baseline CNN, a CNN with spatial attention, a convolutional-recurrent deep neural network (CRDNN), and a 3D variant for modulation tensorgrams. All models were trained using the RMSE as the loss function. An initial learning rate of 0.0001 was employed, together with an adaptive learning rate strategy to improve convergence during training. More details on these architectures and benchmark systems are given next.

6.3.3.1 Baseline CNN

The baseline convolutional model consists of three convolutional layers using 3×3 kernels and ReLU activation. Each block is followed by batch normalization and pooling, with L1 regularization applied to the deeper layers to reduce overfitting. The final convolutional features are flattened and passed through a dense layer with 64 units, followed by dropout, and finally mapped to a scalar regression output. Details about this architecture are given in Table 6.1.

TABLE 6.1 : Details of baseline CNN model

Layer	Details
Conv2D	32 filters, 3×3, ReLU, same padding
BatchNorm	–
MaxPool2D	2×2 pool
Conv2D	64 filters, 3×3, ReLU, L1 reg, same padding
BatchNorm	–
MaxPool2D	2×2 pool
Conv2D	128 filters, 3×3, ReLU, L1 reg, same padding
BatchNorm	–
Dropout	Rate 0.25
Flatten	–
Dense	64 units, ReLU
BatchNorm	–
Dropout	Rate 0.5
Dense	1 unit, linear activation

6.3.3.2 CNN with Spatial Attention

To enhance model focus on salient acoustic regions, spatial attention modules were integrated into the CNN pipeline (see Table 6.2 for more details). Attention maps are computed using a 7×7 convolution after each convolutional layer, emphasizing relevant feature regions. This attention-guided feature enhancement improves robustness to irrelevant background noise, while L2 regularization and dropout are employed to reduce overfitting.

6.3.3.3 CRDNN (Convolutional-Recurrent-DNN)

To capture both local and long-range temporal dependencies in the acoustic data, we designed a CRDNN architecture (see Table 6.3). The model begins with three convolutional layers, each followed by batch normalization and pooling. After flattening, the output is reshaped into a temporal sequence suitable for a Gated Recurrent Unit (GRU) processing. A two-layer GRU extracts sequential patterns, followed by fully connected layers for final prediction. This hybrid structure combines the strengths of CNNs for spatial and GRUs for temporal modeling.

6.3.3.4 3D variants

Lastly, to leverage the full dimensionality of the modulation tensorgram representations, we designed 3D variants of our convolutional and attention-based models. Unlike the 2D models, which operate on standard time-frequency spectrograms, the 3D models process volumetric inputs that

TABLE 6.2 : Details of CNN model with spatial attention

Layer	Details
Conv2D	32 filters, 3×3, ReLU, same padding
BatchNorm	–
Spatial Attn.	7x7 Conv2D, sigmoid
MaxPool2D	2×2 pool
Conv2D	64 filters, 3×3, ReLU, L2 reg
BatchNorm	–
Spatial Attn.	7x7 Conv2D, sigmoid
MaxPool2D	2×2 pool
Conv2D	128 filters, 3×3, ReLU, L2 reg
BatchNorm	–
Spatial Attn.	7x7 Conv2D, sigmoid
Dropout	Rate 0.25
Flatten	–
Dense	64 units, ReLU
BatchNorm	–
Dropout	Rate 0.5
Dense	1 unit, linear activation

TABLE 6.3 : Details of CRDNN model

Layer	Details
Conv2D	32 filters, 3×3, ReLU, same padding
BatchNorm	–
MaxPool2D	2×2 pool
Conv2D	64 filters, 3×3, ReLU, L1 reg, same padding
BatchNorm	–
MaxPool2D	2×2 pool
Conv2D	128 filters, 3×3, ReLU, L1 reg, same padding
BatchNorm	–
Dropout	Rate 0.25
Flatten	–
Reshape	To $(\text{seq_len}, 128)$
GRU	32 units, return seq = True
Dropout	Rate 0.5
GRU	64 units, return seq = False
Dense	256 units, ReLU
BatchNorm	–
Dropout	Rate 0.5
Dense	128 units, ReLU
BatchNorm	–
Dropout	Rate 0.5
Dense	1 unit, linear activation

capture time, frequency, and modulation frequency simultaneously. This three-dimensional structure enables the models to extract joint spatio-spectro-temporal features, making them well-suited for tasks involving modulation-based acoustic representations. The 3D CNN with attention integrates spatial attention mechanisms across the 3D input, enhancing relevant regions within the tensor while suppressing less informative ones. Each convolutional block includes batch normalization, 3D max-pooling, and dropout for regularization, followed by dense layers for regression output. By capturing patterns across all axes of the tensorgram, these models are able to exploit modulation structures more effectively, improving the robustness and generalizability of acoustic monitoring tasks.

6.4 Experimental Setup

6.4.1 Dataset Description

To test the efficacy of our proposed method, we rely on the UrBAN dataset [170] which was described in chapter 3. Figure 6.3 shows the number of frames of bees (fob) for each beehive in 2021 (top plot) and 2022 (bottom). Values between 1-10 suggest only the brood chamber was present, while 11-20 the brood chamber plus one super, and 21-30 the brood chamber plus two supers.

6.4.2 Benchmark features and noise removal

To establish a baseline for evaluating the performance of the proposed features, we incorporated two widely used acoustic representations into our analysis : MFCCs and spectrogram.

Notwithstanding, as mentioned previously, these features have been shown to be sensitive to environmental noise. As the dataset used herein was collected “in the wild,” it is crucial that some noise suppression be performed to ensure reliable prediction from the benchmark systems. Here, we employ a classical spectral amplitude subtraction method for noise reduction, as described in chapter 3.

6.4.3 Testing Setup and Figures-of-Merit

The experimental setup follows the configurations described in Chapter 5, Section 5.2.2, specifically the random-split and hive-independent strategies. Model training, validation, and testing were conducted using five-fold nested cross-validation to ensure robust performance evaluation.

As our task is a regression aimed at predicting the strength of a hive (given by its fob value), two conventional figures-of-merit are utilized : mean absolute error (MAE) and Pearson correlation

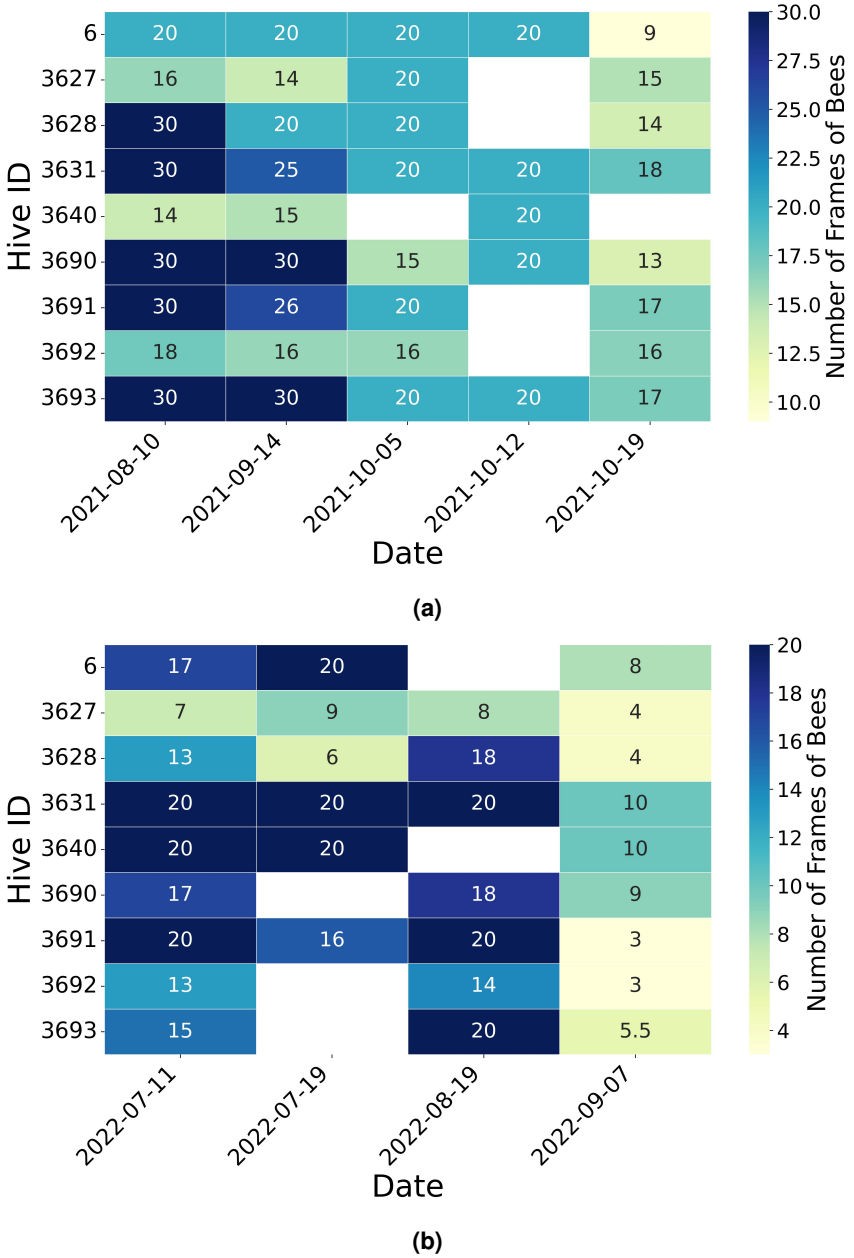


FIGURE 6.3 : Heatmaps showing the number of frames of bees estimated during visual inspection for each hive throughout (a) 2021 and (b) 2022. Values between 1-10 suggest only the brood chamber was present, 11-20 brood chamber plus one super, and 21-30 brood chamber plus two supers.

(r) between the real and predicted fob values. An ideal/perfect regressor will have MAE close to zero and a correlation close to unity.

6.5 Results and Discussion

6.5.1 Colony Population Strength Prediction

Table 6.4 presents the performance achieved by the different benchmark and proposed methods. As can be seen, in the random-split scenario, while the spectrogram and MFCC inputs yield MAE values around 3.5 and correlation $r \approx 0.76$, the modulation-based methods are capable of reducing the MAE to around 1.5 and elevate the correlation to $r \approx 0.95$. This underscores the hypothesis that bee acoustic activity is best characterized by temporal modulations (i.e., the rate and structure of changes in the dominant frequencies) rather than the static spectral envelope or cepstral coefficients. Moreover, the traditional spectrogram and MFCC benchmarks showed substantial gains in performance once audio enhancement was applied. The gains were not as substantial with the proposed method, corroborating the noise-robustness properties of the modulation spectrogram and highlighting its importance for in-the-wild hive monitoring. Overall, for this setting, the CRDNN-3D model utilizing the modulation tensorgram as input resulted in the best performance.

For the more strict hive-independent setting, the performance is shown to drop for all models and features. Notwithstanding, overall, the modulation spectrogram and tensorgram inputs retain their superior performance, achieving correlations greater than 0.7, while the spectrogram and MFCC benchmarks remain around 0.50. This resilience suggests that the underlying frequency-modulation patterns are more robust and generalizable indicators of bee activity than the raw spectral content, which is likely dominated by hive-specific artifacts.

In both settings, the 3D CNN and CRDNN models, designed to process the time-frequency-modulation cube, generally outperformed their 2D counterparts. The CRDNN-3D achieved the lowest MAE (1.01 and 3.31) in both scenarios, indicating that the combination of 3D convolutional layers and a Recurrent Neural Network (RNN) component is best suited to capture the complex spatiotemporal evolution of the modulation patterns over time. This architectural advantage is maximized when the input explicitly contains the time axis, as with the modulation tensorgram input. Across both 2D and 3D inputs, the CRDNN variants consistently show lower MAE and higher r than the pure CNN models, suggesting that the recurrent component effectively models the temporal dependencies within the audio signal's feature representation.

6.5.2 Non-parametric Statistical Analyses

To statistically evaluate the differences in performance seen amongst the models, we employed the Friedman test [207] across all four experimental scenarios (random-split and hive-independent, with and without audio enhancement) and two evaluation metrics. Next, we used the Nemenyi post-hoc test [209] to conduct pairwise comparisons and identify which specific differences were statistically significant. The results of the Friedman test are summarized in Table 6.5 and showed

TABLE 6.4 : Performance comparison of MAE and correlation (r) between different feature sets and data partitioning setups, with and without spectral enhancement.

Feature	Model	Random-Split			
		No pre-processing		Spectral amplitude subtraction	
		MAE	r	MAE	r
Spectrogram	CNN-2D	3.70 ± 0.17	0.76 ± 0.03	1.71 ± 0.13	0.83 ± 0.02
	CNN-att-2D	3.68 ± 0.08	0.75 ± 0.02	2.17 ± 0.09	0.76 ± 0.01
	CRDNN-2D	3.49 ± 0.24	0.78 ± 0.02	1.77 ± 0.05	0.81 ± 0.01
MFCCs	CNN-2D	3.62 ± 0.13	0.76 ± 0.02	1.85 ± 0.25	0.81 ± 0.03
	CNN-att-2D	3.72 ± 0.09	0.75 ± 0.01	2.15 ± 0.17	0.77 ± 0.01
	CRDNN-2D	3.37 ± 0.05	0.78 ± 0.03	1.78 ± 0.10	0.82 ± 0.01
Modulation spectrogram	CNN-2D	1.54 ± 0.14	0.95 ± 0.01	1.14 ± 0.06	0.96 ± 0.01
	CNN-att-2D	1.52 ± 0.16	0.96 ± 0.03	1.39 ± 0.06	0.97 ± 0.03
Modulation tensorgram	CNN-3D	1.61 ± 0.10	0.94 ± 0.02	1.36 ± 0.09	0.96 ± 0.01
	CNN-att-3D	1.45 ± 0.05	0.96 ± 0.02	1.23 ± 0.06	0.96 ± 0.03
	CRDNN-3D	1.24 ± 0.03	0.95 ± 0.01	1.01 ± 0.08	0.97 ± 0.01
		Hive-Independent			
Spectrogram	CNN-2D	4.87 ± 1.47	0.47 ± 0.22	4.74 ± 1.46	0.52 ± 0.21
	CNN-att-2D	4.60 ± 1.61	0.46 ± 0.21	4.50 ± 1.59	0.51 ± 0.20
	CRDNN-2D	4.55 ± 1.62	0.56 ± 0.20	4.22 ± 1.16	0.61 ± 0.20
MFCCs	CNN-2D	4.95 ± 1.37	0.50 ± 0.22	4.65 ± 1.32	0.53 ± 0.22
	CNN-att-2D	4.99 ± 1.45	0.47 ± 0.24	4.97 ± 1.46	0.50 ± 0.22
	CRDNN-2D	4.42 ± 1.68	0.52 ± 0.25	4.38 ± 1.68	0.54 ± 0.20
Modulation spectrogram	CNN-2D	3.48 ± 0.90	0.69 ± 0.13	3.41 ± 0.87	0.74 ± 0.11
	CNN-att-2D	3.66 ± 1.09	0.72 ± 0.09	3.56 ± 1.05	0.76 ± 0.08
Modulation tensorgram	CNN-3D	3.84 ± 1.21	0.65 ± 0.22	3.69 ± 1.60	0.69 ± 0.21
	CNN-att-3D	3.76 ± 1.56	0.71 ± 0.15	3.67 ± 1.55	0.76 ± 0.16
	CRDNN-3D	3.42 ± 1.37	0.72 ± 0.20	3.31 ± 1.36	0.78 ± 0.17

statistically significant differences ($p < 0.05$) among the feature set and models performances across evaluation metrics. The critical Chi-square value for this test (with $k = 11$) was determined to be 18.30 at $\alpha = 0.05$. Next, Table 6.6 reports the average ranks derived from the Friedman procedure. For error metrics such as MAE, a lower rank signifies superior performance, while for predictive metrics such as Correlation, a higher rank is better. Notably, the CRDNN-3D achieved particularly strong average ranks, especially when tested under the challenging hive-independent scenario.

Figures 6.4 (a)-(b) present the Critical Difference (CD) diagrams for the MAE metric under both the random-split and hive-independent evaluation setups, respectively. These diagrams summarize the post-hoc statistical comparisons : methods linked by a horizontal bar do not differ significantly, while those separated by more than the critical distance show statistically significant differences. As can be seen, in the random-split scenario, Mod-Tgram(tensorgram)-CRDNN-3D obtained the top average rank and outperformed models such as MFCCs-CNN-2D and Spec-CNN-att-2D, evidenced by the clear separation in the CD diagram. Conversely, in the hive-independent scenario,

TABLE 6.5 : Friedman test results across evaluation metrics MAE and correlation(r) under random-split and hive-independent scenarios.

Metric	Random-Split			
	No pre-processing		Spectral amplitude subtraction	
	Friedman χ^2	p-value	Friedman χ^2	p-value
MAE	46.32	1.25e-06	47.30	8.31e-07
r	46.98	9.48e-07	47.44	7.82e-07
	Hive-Independent			
MAE	32.78	2e-04	35.90	8.75e-05
r	37.71	4.24e-05	40.32	1.48e-05

TABLE 6.6 : Comparison of average ranks across evaluation metrics under random-split and hive-independent scenarios, with and without audio enhancement. Bold values correspond to best methods.

Feature	Model	Random-Split			
		No pre-processing		Spectral amplitude subtraction	
		MAE ↓	r ↑	MAE ↓	r ↑
Spectrogram	CNN-2D	6.6	4.0	6.6	5.5
	CNN-att-2D	10.7	2.8	10.7	1.3
	CRDNN-2D	7.9	3.6	7.9	3.7
MFCCs	CNN-2D	8.4	2.0	8.4	3.9
	CNN-att-2D	10.0	2.8	10.0	1.7
	CRDNN-2D	7.4	5.8	7.4	4.9
Modulation spectrogram	CNN-2D	1.8	7.9	1.8	9.9
	CNN-att-2D	4.8	9.4	4.8	8.16
Modulation tensorgram	CNN-3D	4.2	7.9	4.2	7.9
	CNN-att-3D	2.8	9.9	2.8	9.0
	CRDNN-3D	**1.4**	**9.9**	**1.4**	**10.1**
		Hive-Independent			
Spectrogram	CNN-2D	9.3	2.5	9.1	3.4
	CNN-att-2D	7.6	2.6	7.6	3.1
	CRDNN-2D	7.2	6.0	6.3	6.4
MFCCs	CNN-2D	9.2	4.1	8.8	3.0
	CNN-att-2D	9.1	3.0	10.1	2.4
	CRDNN-2D	6.6	3.8	7.4	3.2
Modulation spectrogram	CNN-2D	3.2	8.2	3.2	9.1
	CNN-att-2D	3.8	8.7	3.3	8.6
Modulation tensorgram	CNN-3D	5.1	8.1	5.0	8.2
	CNN-att-3D	2.9	9.7	3.2	9.4
	CRDNN-3D	**2.0**	**9.3**	**2.0**	**9.2**

Mod-Tgram-CRDNN-3D grouped distinctly from MFCCs-att-CNN-2D, highlighting its superior ability to generalize across hives.

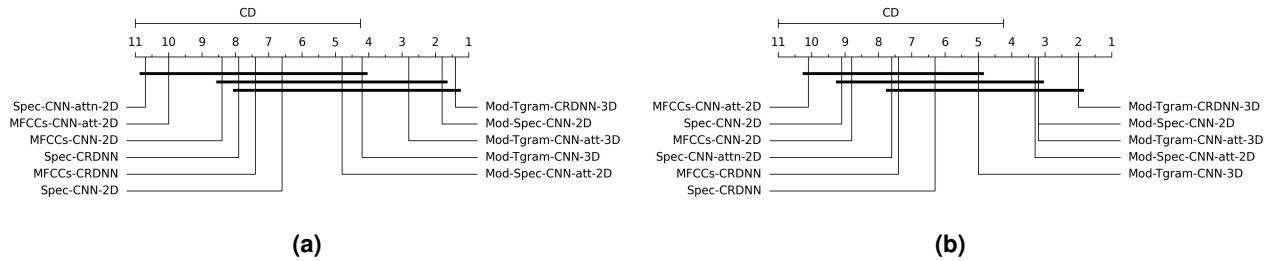


FIGURE 6.4 : Critical Difference (CD) plot illustrating the results of the Nemenyi post-hoc test following the Friedman test on the MAE for a) random-split and b) hive-independent scenario. The models and features combination are ranked by their average performance across the five folds. Features grouped by a thick horizontal bar exhibit no statistically significant difference in performance ($\alpha = 0.05$). Features with lower ranks (towards 1) demonstrate superior performance.

6.5.3 Explainability analysis

To interpret the spatial and temporal relevance of the input modulation spectrograms, we generated saliency maps for the trained 2D CNN and attention-based models. Saliency maps highlight the input regions that most strongly influence the model's predictions by computing the gradient of the output with respect to each feature. In our case, these maps reveal which frequency and modulation bands are most discriminative for different levels of bee activity. To reduce noise and capture consistent activation patterns, we averaged the saliency maps across multiple samples within each colony strength category (low, medium, and high). The resulting maps indicate that specific modulation frequency regions contribute more strongly as colony strength increases, suggesting that the model relies on structured acoustic patterns rather than isolated features. Figures 6.5 and 6.6 show the saliency maps for the three strength levels for the 2D CNN and 2D CNN with spatial attention models, respectively.

The saliency maps for both the standard 2D CNN and the 2D CNN with spatial attention confirm that the models primarily rely on specific modulation-frequency patterns to predict bee activity levels, suggesting a reliance on structured acoustic patterns. For the standard 2D CNN (Figure 6.5), increasing colony strength correlates with a progressive expansion and upward shift of the relevant frequency region, suggesting that higher harmonics or broader spectral features become relevant. In contrast, the 2D CNN with spatial attention (Figure 6.6) maintains a consistently narrower, more horizontal high-saliency band, primarily focused around 200 Hz, which expands across the modulation frequency axis as activity increases. This suggests the attention mechanism is effective at isolating the single most discriminative fundamental frequency band and focusing on its temporal variation (modulation), whereas the standard CNN integrates a broader range of adjacent frequencies. Notably, both models show a peak relevance in the 10–25 Hz modulation frequency range across the medium and high strength categories, indicating that these specific rates of temporal fluctuation are crucial discriminative features.

Next, for explainability of the 3D CNN model, we used the gradient-weighted class activation mapping (Grad-CAM) method [221]. Grad-CAM leverages intermediate feature activations to gene-

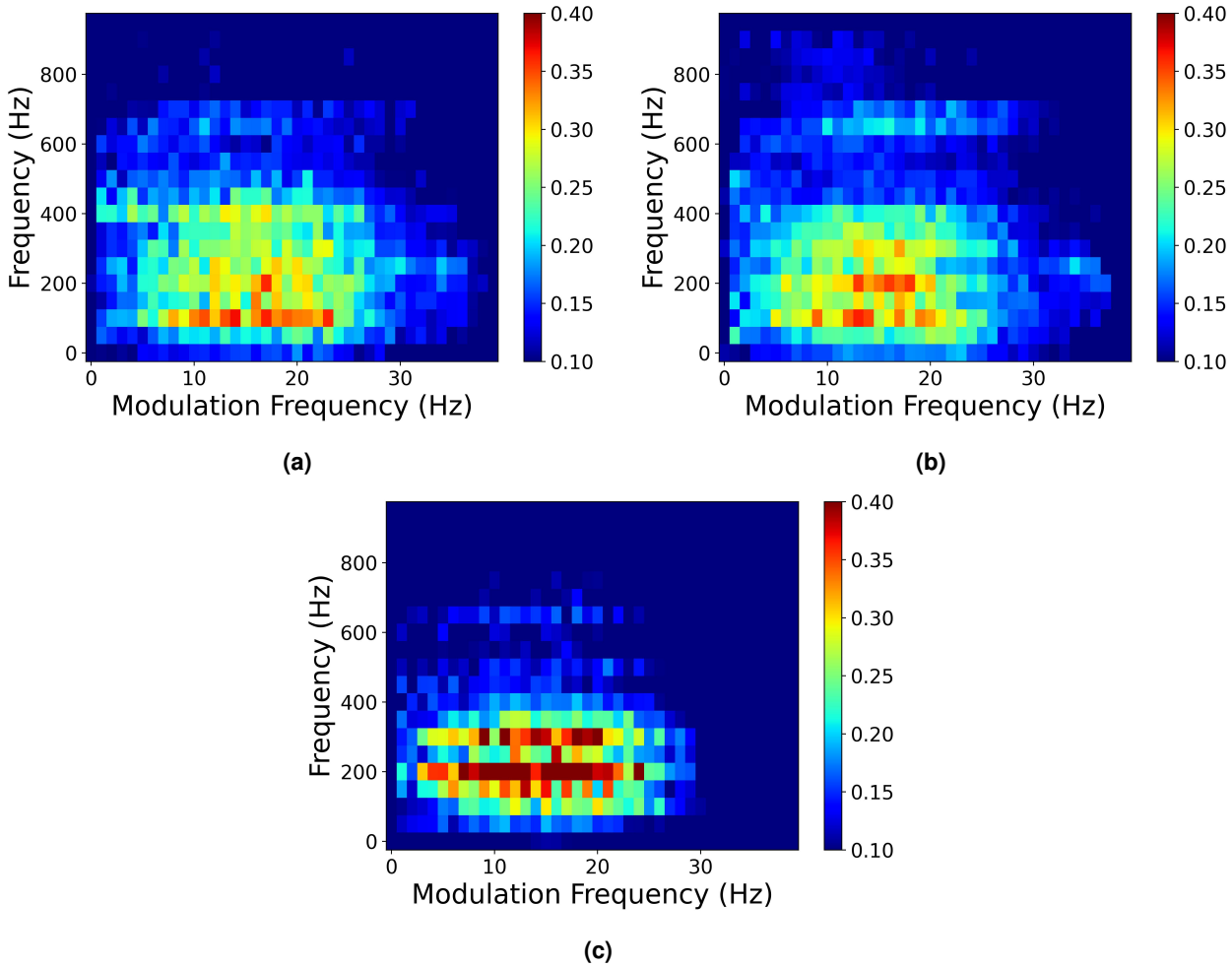


FIGURE 6.5 : Saliency maps of the trained 2D CNN model averaged over samples for each colony strength level : (a) low (1–10 fobs), (b) medium (11-20 fobs), and (c) high (21–30 fobs).

rate interpretable maps across spatial and temporal dimensions. By averaging these maps across samples, we identified the regions most relevant to the model's predictions, revealing how specific frequency–modulation patterns evolve with colony strength. This approach allowed us to visualize and compare differences in the 2D and 3D architectures.

Rather than focusing solely on a single aggregated explanation, we examined several representative frames including the first, middle, and last frames to capture the model's response to temporal variations in the modulation spectrogram. We found that high-activity samples exhibit more consistent and spatially distributed activation across frames, while low-activity samples display more localized and intermittent patterns. These observations suggest that the 3D models leverage both spectral structure and temporal evolution to estimate colony strength levels. Figure 6.7 provides a detailed visualization of these temporal dynamics. For each strength level (left column=low, middle=medium, right=high), the figure shows the three Grad-CAM heatmaps extracted from three different time frames (top row=first, middle=middle, bottom=last frame).

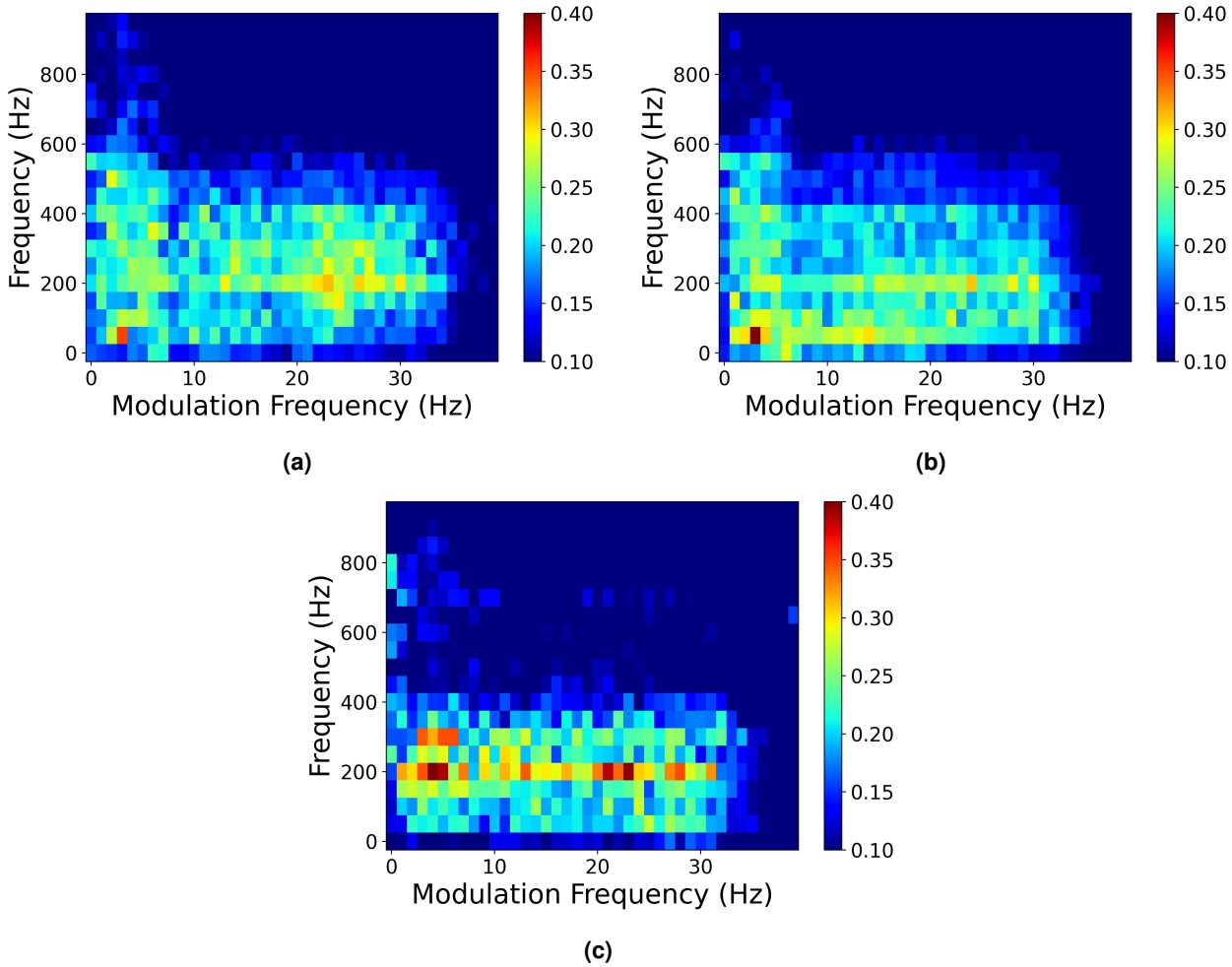


FIGURE 6.6 : Saliency maps of the trained 2D CNN model with spatial attention averaged over samples for each colony strength level : (a) low (1–10 fobs), (b) medium (11–20 fobs), and (c) high (21–30 fobs).

As can be seen, for low strength colonies (1–10 fobs), the model's focus is generally localized to low frequencies (below 400 Hz) and very low modulation frequencies (below 10 Hz), becoming notably intermittent and contracting severely at the middle frame (t_{mid}). In contrast, high strength colony (21–30 fobs) samples show a consistent and broadly distributed activation across all frames, leveraging low-to-mid frequencies (up to 550 Hz) and spanning the entire range of modulation frequencies (up to 35 Hz), which suggests the model uses a stable, widespread feature set. The medium strength class (11–20 fobs) exhibits a dynamic pattern, shifting from a broad focus in t_1 to a localized one in t_{mid}, and then expanding again in t_{max}. These observations confirm that the 3D architecture effectively leverages both spectral structure (Frequency) and its temporal evolution (Modulation Frequency over time) to differentiate colony strength levels.

6.5.4 Deep learning vs. Conventional classifiers

As a last exploration, we wish to compare the results obtained with the deep learning methods described herein with those of our earlier work that relied on a conventional random forest (RF)

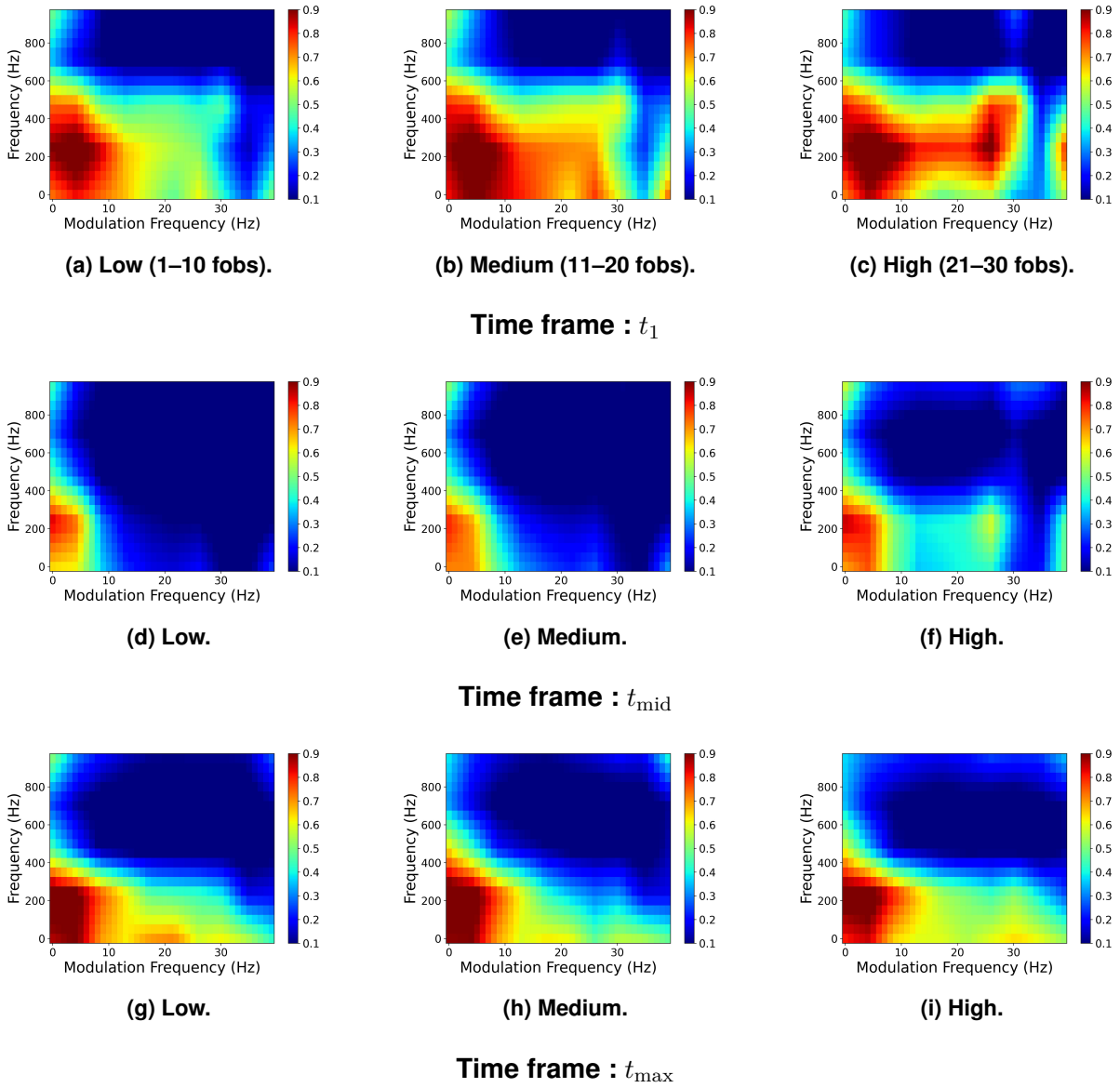


FIGURE 6.7 : Grad-CAM visualizations for representative 3D test samples. Each row corresponds to a different time frame (t_1, t_{mid}, t_{max}), and columns correspond to colony strength levels (low, medium, high). Heatmaps show how the model's focus shifts across the modulation-frequency space over time.

classifier as discussed in chapter 5, Table 5.2. Table 6.7 summarizes the results obtained with the best RF model and ours. As can be seen, the proposed method improves accuracy across both figures-of-merit in the more stringent hive-independent scenario. Particularly, the proposed CRDNN-3D model is able to take advantage of the modulation tensorgram 3D input to better characterize the temporal changes of the bee buzzing sounds, something that the RF classifier with handcrafted features is not able to achieve.

TABLE 6.7 : **Performance comparison of best-performing random forest (RF) classifier and best-performing deep learning (DL) method presented herein in the hive-independent scenario.**

Metric	RF Best (MSAB)	DL Best (CRDNN-3D)
MAE	3.41 ± 1.03	3.31 ± 1.36
r	0.74 ± 0.15	0.78 ± 0.17

6.6 Conclusion

In this chapter, we proposed the use of modulation spectrograms and tensorgrams, derived from beehive acoustic recordings, as input to deep learning architectures for improved honey bee colony strength prediction. With a recent open dataset of raw audio beehive recordings, we compared the performance of two- and three-dimensional CNNs and their attention-enhanced variants, as well as convolutional recurrent CNNs. Results show that the proposed modulation tensorgrams and modulation spectrograms provide richer information for model learning, leading to improved and more generalizable prediction accuracy compared to conventional audio features. These results suggest that modulation spectral signal processing is a viable tool for automated in-the-wild beehive monitoring.

7 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

7.1 Conclusion

In this dissertation, we focused on developing generalizable acoustic monitoring systems for honey bee colonies. The main contributions of this work can be summarized as follows :

Environmental factors : We systematically investigated the impact of prominent environmental noise sources—urban sounds, background speech, and heavy rainfall-on commonly used acoustic representations, including spectrograms, MFCCs, and discrete wavelet transform (DWT) features. Our analysis highlights how different noise conditions affect feature robustness and model performance in real-world deployment scenarios.

Dataset : We introduced the *UrBAN* dataset, a large-scale, open-access beehive acoustics dataset comprising over 3,000 hours of raw audio recordings collected from ten beehives over a two-year period. The dataset is annotated with colony population estimates and Varroa mite infestation measurements, enabling comprehensive evaluation of acoustic-based monitoring methods under realistic conditions.

Feature sets : We proposed novel audio feature representations based on the modulation spectrogram and conducted an in-depth analysis of their effectiveness for beehive population prediction. In addition, we introduced spectral shape descriptors extracted from both conventional spectrograms and modulation spectrograms, demonstrating their suitability for Varroa mite infestation detection.

Modeling approaches : We investigated the use of modulation spectrograms and tensorgram representations as inputs to both classical machine-learning algorithms and deep neural network architectures. Our results show that these representations enable accurate and computationally efficient prediction of honey bee colony strength, supporting robust performance across diverse hives and recording conditions.

7.2 Limitations

Dataset Precision and Label Reliability : Despite the promising results, several limitations should be acknowledged. A primary constraint concerns the precision of the UrBan dataset used for modeling. Phenotypic colony strength assessments may involve measurement variability due to subjective grading, environmental factors, and operational constraints during inspections. Such variability introduces potential noise in the ground-truth labels, which may limit predictive performance and affect model interpretability.

Missing and Incomplete Data : The dataset also contains missing observations across certain colonies and time periods. These gaps may have influenced training stability and evaluation outcomes, particularly if the missingness is not random. Although appropriate preprocessing and cross-validation strategies were applied, incomplete temporal coverage inevitably constrains the robustness of the learned representations and may contribute to performance variability across experiments.

External Validity : The findings are further limited by the specific environmental conditions, geographic locations, hive configurations, and sensor setups under which the data were collected. Acoustic characteristics of honeybee colonies may vary due to climate, genetics, hive materials, or microphone placement. Consequently, the generalizability of the proposed framework to other apicultural contexts requires further validation.

More broadly, this work highlights the central role of phenotypic data quality in precision apiculture. The effectiveness of artificial intelligence models fundamentally depends on the reliability and biological relevance of the training data. Strong alignment between sensor-derived signals and accurately measured colony-level ground truth is essential for developing robust and deployable monitoring systems. Future efforts should therefore prioritize standardized grading protocols and improved data completeness.

7.3 Future Research Directions

Despite the promising results obtained in this dissertation, several research directions remain open for further improving acoustic-based beehive monitoring systems.

Edge deployment and real-time monitoring : Future work should explore the deployment of the proposed models on edge devices located directly within or near the beehive. Edge computing enables on-device processing of acoustic signals, reducing the need for continuous data transmission to centralized servers. This is particularly important in remote apiary environments where connectivity may be limited or unreliable. By performing inference locally, the system can provide real-time monitoring of colony activity and health-related events while minimizing latency and bandwidth usage. Furthermore, edge deployment introduces constraints in terms of computational resources and energy consumption, motivating the development of lightweight and efficient models. Integrating acoustic analysis with low-power embedded systems represents a key step toward scalable and autonomous beehive monitoring solutions.

Self-supervised and foundation models : Future work should investigate the use of self-supervised learning and foundation audio models to leverage large volumes of unlabeled beehive recordings. Contrastive and masked-prediction approaches could enable the learning of robust and transferable acoustic representations that generalize across hives, seasons, and recording conditions, reducing the dependence on costly manual annotations.

Feature representation and temporal dynamics : While delta and double-delta MFCCs were explored in preliminary experiments, they did not provide significant performance improvements. This may be attributed to the conventional computation of these features, which relies on short temporal windows (typically spanning ± 2 or ± 3 frames). As demonstrated in later chapters, longer temporal contexts are often required to capture meaningful spectro-temporal patterns in beehive acoustics. Therefore, future work should investigate the use of extended temporal windows in the computation of delta and double-delta features, or alternative formulations that better capture long-term dynamics.

Multi-location audio recording : While the current dataset relies on a single microphone per hive, future deployments may benefit from placing microphones in multiple boxes (e.g., brood chamber and honey supers). This would enable more precise localization of acoustic activity and allow models to capture spatial variations related to colony strength, brood distribution, and foraging behavior.

Microphone arrays : Extending the sensing setup to include microphone arrays within the hive could provide spatial information about sound sources. Such configurations would allow the analysis of directional acoustic patterns and improve robustness to environmental noise, particularly in urban or high-wind environments.

Beamforming and spatial signal processing : With the availability of multi-microphone recordings, advanced beamforming and spatial filtering techniques could be employed to enhance bee-generated sounds while suppressing external noise sources. Integrating beamforming with machine learning models may further improve detection accuracy for colony population dynamics and health-related events.

Overall, these directions aim to improve the robustness, scalability, and biological interpretability of acoustic monitoring systems, bringing them closer to reliable real-world deployment in precision apiculture.

BIBLIOGRAPHIE

[1] FAO, Apimondia, CAAS, IZSLT. Good beekeeping practices for sustainable apiculture, 2021.

[2] Yuwei Liao, Anya McGuirk, Byron Biggs, A. Chaudhuri, Allen Langlois, and Vince Deters. Noninvasive beehive monitoring through acoustic data using sasa® event stream processing and sas® viya®. 2020.

[3] Kelly Kulhanek, Nathalie Steinhauer, Karen Rennich, Dewey M Caron, Ramesh R Sagili, Jeff S Pettis, James D Ellis, Michael E Wilson, James T Wilkes, David R Tarpy, et al. A national survey of managed honey bee 2015–2016 annual colony losses in the usa. *Journal of Apicultural Research*, 56(4) :328–340, 2017.

[4] Robert Brodschneider, Alison Gray, Noureddine Adjlane, Alexis Ballis, Valters Brusbardis, Jean-Daniel Charrière, Robert Chlebo, Mary F Coffey, Bjørn Dahle, Dirk C de Graaf, et al. Multi-country loss rates of honey bee colonies during winter 2016/2017 from the coloss survey. *Journal of Apicultural Research*, 57(3) :452–457, 2018.

[5] Alison Gray, Robert Brodschneider, Noureddine Adjlane, Alexis Ballis, Valters Brusbardis, Jean-Daniel Charrière, Robert Chlebo, Mary F. Coffey, Bram Cornelissen, Cristina Amaro da Costa, et al. Loss rates of honey bee colonies during winter 2017/18 in 36 countries participating in the coloss survey, including effects of forage sources. *Journal of Apicultural Research*, 58(4) :479–485, 2019.

[6] Alison Gray, Noureddine Adjlane, Alireza Arab, Alexis Ballis, Valters Brusbardis, Jean-Daniel Charrière, Robert Chlebo, Mary F Coffey, Bram Cornelissen, Cristina Amaro da Costa, 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. *Journal of Apicultural Research*, 59(5) :744–751, 2020.

[7] Benjamin P Oldroyd. What's killing american honey bees? *PLoS biology*, 5(6) :e168, 2007.

[8] Claudio Porrini, Franco Mutinelli, Laura Bortolotti, Anna Granato, Lynn Laurenson, Katherine Roberts, Albino Gallina, Nicholas Silvester, Piotr Medrzycki, Teresa Renzi, et al. The status of honey bee health in italy : Results from the nationwide bee monitoring network. *PloS one*, 11(5) :e0155411, 2016.

[9] Zoran Stanimirović, Uroš Glavinić, Marko Ristanić, Nevenka Aleksić, Nemanja Jovanović, Branislav Vejnović, and Jevrosima Stevanović. Looking for the causes of and solutions to the issue of honey bee colony losses. *Acta Veterinaria*, 69(1) :1–31, 2019.

[10] Stefania Cecchi, Susanna Spinsante, Alessandro Terenzi, and Simone Orcioni. A smart sensor-based measurement system for advanced bee hive monitoring. *Sensors*, 20(9), 2020.

[11] Hugo Hadjur, Doreid Ammar, and Laurent Lefèvre. Toward an intelligent and efficient beehive : A survey of precision beekeeping systems and services. *Computers and Electronics in Agriculture*, 192 :106604, 2022.

[12] 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.

[13] William G Meikle, Brian G Rector, Guy Mercadier, and Niels Holst. Within-day variation in continuous hive weight data as a measure of honey bee colony activity. *Apidologie*, 39(6) :694–707, 2008.

[14] Egils Stalidzans and Almars Berzonis. Temperature changes above the upper hive body reveal the annual development periods of honey bee colonies. *Computers and electronics in agriculture*, 90 :1–6, 2013.

[15] Fiona Edwards-Murphy, Michele Magno, Pádraig M Whelan, John O'Halloran, and Emanuel M Popovici. b+ wsn : Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring. *Computers and Electronics in Agriculture*, 124 :211–219, 2016.

[16] Stenford Ruvinga, Gordon JA Hunter, Olga Duran, and Jean-Christophe 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, 2021.

[17] Stanislas Dubois, Julien Choveton-Caillat, Wedji Kane, Terii Gilbert, Mohcine Nfaoui, Mohamed El Boudali, Marwane Rezzouki, and Guillaume Ferré. Bee detection for fruit cultivation. In *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2021.

[18] Rui Peng, Iman Ardekani, and Hamid Sharifzadeh. An acoustic signal processing system for identification of queen-less beehives. In *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 57–63. IEEE, 2020.

[19] JI Hambleton. The quantitative and qualitative effect of weather upon colony weight changes. *Journal of Economic Entomology*, 18(3) :447–448, 1925.

[20] U Cetin. The effects of temperature changes to bee losts. *Uludag Bee J*, 4(4) :171–174, 2004.

[21] Thomas Seeley and Bernd Heinrich. Regulation of temperature in the nests of social insects. 1981.

[22] Thomas D Seeley. Honeybee ecology. In *Honeybee Ecology*. Princeton University Press, 2014.

[23] Aleksejs Zacepins, Armands Kviesis, Egils Stalidzans, Marta Liepniece, and Jurijs Meitalovs. Remote detection of the swarming of honey bee colonies by single-point temperature monitoring. *Biosystems engineering*, 148 :76–80, 2016.

[24] HF Abou-Shaara, AA Owayss, YY Ibrahim, and NK Basuny. A review of impacts of temperature and relative humidity on various activities of honey bees. *Insectes sociaux*, 64(4) :455–463, 2017.

[25] Hannelie Human, Sue W Nicolson, and Vincent Dietemann. Do honeybees, apis mellifera scutellata, regulate humidity in their nest? *Naturwissenschaften*, 93(8) :397–401, 2006.

[26] Kenneth A Nagy and John N Stallone. Temperature maintenance and co/sub 2/concentration in a swarm cluster of honey bees, apis mellifera. *Comp. Biochem. Physiol., A;(United Kingdom)*, 55, 1976.

[27] Thomas D Seeley. Atmospheric carbon dioxide regulation in honey-bee (apis mellifera) colonies. *Journal of insect physiology*, 20(11) :2301–2305, 1974.

[28] 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(3) :207–212, 1986.

[29] JH Hunt and F-J Richard. Intracolony vibroacoustic communication in social insects. *Insectes Sociaux*, 60(4) :403–417, 2013.

[30] Axel Michelsen, Wolfgang H Kirchner, Bent Bach Andersen, and Martin Lindauer. The tooting and quacking vibration signals of honeybee queens : a quantitative analysis. *Journal of Comparative Physiology A*, 158(5) :605–611, 1986.

[31] Corinna Thom, David C Gilley, and Jürgen Tautz. Worker piping in honey bees (apis mellifera) : the behavior of piping nectar foragers. *Behavioral Ecology and Sociobiology*, 53(4) :199–205, 2003.

[32] SC Pratt, Susanne Kühnholz, Thomas D Seeley, and Anja Weidenmüller. Worker piping associated with foraging in undisturbed queenright colonies of honey bees. *Apidologie*, 27(1) :13–20, 1996.

[33] Thomas D Seeley and Jürgen Tautz. Worker piping in honey bee swarms and its role in preparing for liftoff. *Journal of Comparative Physiology A*, 187(8) :667–676, 2001.

[34] Amro Qandour, Iftekhar Ahmad, Daryoush Habibi, and Mark Leppard. Remote beehive monitoring using acoustic signals. 2014.

[35] Axel Michelsen. Signals and flexibility in the dance communication of honeybees. *Journal of Comparative Physiology A*, 189(3) :165–174, 2003.

[36] Michael-Thomas Ramsey, Martin Bencsik, Michael Ian Newton, Maritza Reyes, Maryline Pioz, Didier Crauser, Noa Simon Delso, and Yves Le Conte. The prediction of swarming in honeybee colonies using vibrational spectra. *Scientific reports*, 10(1) :1–17, 2020.

[37] Antonio Robles-Guerrero, Tonatiuh Saucedo-Anaya, Efrén González-Ramérez, and Carlos Eric Galván-Tejada. Frequency analysis of honey bee buzz for automatic recognition of health status : A preliminary study. *Res. Comput. Sci.*, 142 :89–98, 2017.

[38] Apivox. `https://apivox-smart-monitor.weebly.com/`. Accessed : 2021-10-31.

[39] Bee health guru. `https://www.beehealth.guru/`. Accessed : 2021-10-31.

[40] Arnia. `https://www.arnia.co/`. Accessed : 2021-10-31.

[41] Nectar. `https://www.nectar.buzz/`. Accessed : 2024-04-15.

[42] Daniele Barchiesi, Dimitrios Giannoulis, Dan Stowell, and Mark D Plumbley. Acoustic scene classification : Classifying environments from the sounds they produce. *IEEE Signal Processing Magazine*, 32(3) :16–34, 2015.

[43] Amr Gaballah, Abhishek Tiwari, Shrikanth Narayanan, and Tiago H Falk. Context-aware speech stress detection in hospital workers using bi-lstm classifiers. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8348–8352. IEEE, 2021.

[44] Erik Stange. Optimizing urban beekeeping. In *Achieving sustainable urban agriculture*, pages 331–352. Burleigh Dodds Science Publishing, 2020.

[45] Wei Hong, Baohua Xu, Xuepeng Chi, Xuepei Cui, Yinfa Yan, and Tongyang Li. Long-term and extensive monitoring for bee colonies based on internet of things. *IEEE Internet of Things Journal*, 7(8) :7148–7155, 2020.

[46] Evan Henry, Viacheslav Adamchuk, Trevor Stanhope, Christopher Buddle, and Nathaniel Rindlaub. Precision apiculture : Development of a wireless sensor network for honeybee hives. *Computers and electronics in agriculture*, 156 :138–144, 2019.

[47] Gordon Hunter. Signal processing the acoustics of honeybees (apis mellifera) to identify the "queenless" state in hives. 2013.

[48] Inês Nolasco, Alessandro Terenzi, Stefania Cecchi, Simone Orcioni, Helen L Bear, and Emmanouil Benetos. Audio-based identification of beehive states. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8256–8260. IEEE, 2019.

[49] Alessandro Terenzi, Nicola Ortolani, Ines Nolasco, Emmanouil Benetos, and Stefania Cecchi. Comparison of feature extraction methods for sound-based classification of honey bee activity. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021.

[50] A closer look : piping, tooting, quacking , available online :. `https://www.beeculture.com/a-closer-look-piping-tooting-quacking/`. Accessed : 2021-10-31.

[51] Tiago H Falk and Wai-Yip Chan. Modulation spectral features for robust far-field speaker identification. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(1) :90–100, 2009.

[52] Mahsa Abdollahi, Pierre Giovenazzo, and Tiago H Falk. Automated beehive acoustics monitoring : A comprehensive review of the literature and recommendations for future work. *Applied Sciences*, 12(8) :3920, 2022.

[53] David Heise, Nicole Miller-Struttmann, Candace Galen, and Johannes Schul. Acoustic detection of bees in the field using casa with focal templates. In *2017 IEEE Sensors Applications Symposium (SAS)*, pages 1–5. IEEE, 2017.

[54] Jaehoon Kim, Jeongkyu Oh, and Tae-Young Heo. Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and grad-cam. *Mathematical Problems in Engineering*, 2021, 2021.

[55] Vladimir Kulyukin, Sarbajit Mukherjee, and Prakhar Amlathe. Toward audio beehive monitoring : Deep learning vs. standard machine learning in classifying beehive audio samples. *Applied Sciences*, 8(9) :1573, 2018.

[56] Furqan Rustam, Muhammad Zahid Sharif, Wajdi Aljedaani, Ernesto Lee, and Imran Ashraf. Bee detection in bee hives using selective features from acoustic data. *Multimedia Tools and Applications*, 83(8) :23269–23296, 2024.

[57] Thu Huong Truong, Huu Du Nguyen, Thi Quynh Anh Mai, Hoang Long Nguyen, Tran Nhat Minh Dang, and Thi-Thu-Hong Phan. A deep learning-based approach for bee sound identification. *Ecological Informatics*, 78 :102274, 2023.

[58] Thi-Thu-Hong Phan, Dong Nguyen-Doan, Du Nguyen-Huu, Hanh Nguyen-Van, and Thai Pham-Hong. Investigation on new mel frequency cepstral coefficients features and hyper-parameters tuning technique for bee sound recognition. *Soft Computing*, 27(9) :5873–5892, 2023.

[59] Andrew Liang. Developing a multimodal system for bee object detection and health assessment. *IEEE Access*, 12 :158703–158713, 2024.

[60] Luca Borgianni, Md Sabbir Ahmed, Davide Adami, and Stefano Giordano. Spectrogram based bee sound analysis with dnns : a step toward federated learning approach. In *2023 4th International Symposium on the Internet of Sounds*, pages 1–8. IEEE, 2023.

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

[62] Tony *Zhang, Szymon Zmyslony, Sergei Nozdrenkov, Matthew Smith, and Brandon Hopkins. Semi-supervised audio representation learning for modeling beehive strengths. *arXiv preprint arXiv :2105.10536*, 2021.

[63] Jederson S Luz, Myllena C de Oliveira, Fábia de M Pereira, Flávio HD de Araújo, and Deborah MV Magalhães. Cepstral and deep features for apis mellifera hive strength classification. *Journal of Internet Services and Applications*, 15(1) :548–560, 2024.

[64] Yash Mahajan, Deep Mehta, Joel Miranda, Ron Pinto, and Vandana Patil. Neuralbee-a beehive health monitoring system. In *2023 International Conference on Communication System, Computing and IT Applications (CSCITA)*, pages 84–89. IEEE, 2023.

[65] Alessandro Terenzi, Stefania Cecchi, Simone Orcioni, and Francesco Piazza. Features extraction applied to the analysis of the sounds emitted by honey bees in a beehive. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 03–08. IEEE, 2019.

[66] Antonio Robles-Guerrero, Tonatiuh Saucedo-Anaya, Efrén González-Ramírez, and José Ismael De la Rosa-Vargas. Analysis of a multiclass classification problem by lasso logistic regression and singular value decomposition to identify sound patterns in queenless bee colonies. *Computers and Electronics in Agriculture*, 159 :69–74, 2019.

[67] Stefania Cecchi, Alessandro Terenzi, Simone Orcioni, Paola Riolo, Sara Ruschioni, and Nunzio Isidoro. A preliminary study of sounds emitted by honey bees in a beehive. In *Audio Engineering Society Convention 144*. Audio Engineering Society, 2018.

[68] Tymoteusz Cejrowski, Julian Szymański, Higinio Mora, and David Gil. Detection of the bee queen presence using sound analysis. In *Asian Conference on Intelligent Information and Database Systems*, pages 297–306. Springer, 2018.

[69] Stenford Ruvinga, Gordon Hunter, Olga Duran, and Jean-Christophe Nebel. Identifying queenlessness in honeybee hives from audio signals using machine learning. *Electronics*, 12(7) :1627, 2023.

[70] Bianca Sousa Soares, Jederson Sousa Luz, Valderlândia Francisca de Macêdo, Romuere Rodrigues Veloso e Silva, Flávio Henrique Duarte de Araújo, and Deborah Maria Vieira Magalhães. Mfcc-based descriptor for bee queen presence detection. *Expert Systems with Applications*, 201 :117104, 2022.

[71] Antonio Robles-Guerrero, Tonatiuh Saucedo-Anaya, Carlos A Guerrero-Mendez, Salvador Gómez-Jiménez, and David J Navarro-Solís. Comparative study of machine learning models for bee colony acoustic pattern classification on low computational resources. *Sensors*, 23(1) :460, 2023.

[72] Augustin Bricout, Philippe Leleux, Pascal Acco, Christophe Escriba, Jean-Yves Fourniols, Georges Soto-Romero, and Rémi Floquet. Bee together : Joining bee audio datasets for hive extrapolation in ai-based monitoring. *Sensors*, 24(18) :6067, 2024.

[73] Antonio Robles-Guerrero, Salvador Gómez-Jiménez, Tonatiuh Saucedo-Anaya, Daniela López-Betancur, David Navarro-Solís, and Carlos Guerrero-Méndez. Convolutional neural networks for real time classification of beehive acoustic patterns on constrained devices. *Sensors*, 24(19) :6384, 2024.

[74] Dimitrios Kanelis, Vasilios Liolios, Fotini Papadopoulou, Maria-Anna Rodopoulou, Dimitrios Kampelopoulos, Kostas Siozios, and Chrysoula Tananaki. Decoding the behavior of a queenless colony using sound signals. *Biology*, 12(11) :1392, 2023.

[75] Luca Barbisan, Giovanna Turvani, and Riente Fabrizio. Audio-based identification of queen bee presence inside beehives. In *2023 IEEE Conference on AgriFood Electronics (CAFE)*, pages 70–74. IEEE, 2023.

[76] Alexander D Maralit, Alexel A Imperial, Rinoa T Cayangyang, Jose B Tan, Roselyn A Maaño, Rodrigo C Belleza, Pedro Jose C De Castro, and David Eric S Oreta. Queenbuzz : A cnn-based architecture for sound processing of queenless beehive towards european apis mellifera bee colonies' survivability. In *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, pages 691–696. IEEE, 2023.

[77] Michaela Sakova, Patrik Jurik, Pavol Galajda, and Miroslav Sokol. Bee hive acoustic monitoring and processing using convolutional neural network and machine learning. In *2024 34th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pages 1–5. IEEE, 2024.

[78] Weihai Huang, Weize Yang, Zhicong Luo, Jun Qi, Tingting Zhang, and Xiangzeng Kong. Sound-based bee colony state analysis using compact mfcc patterns. In *2024 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA)*, pages 2164–2169. IEEE, 2024.

[79] Luca Barbisan, Giovanna Turvani, and Fabrizio Riente. A machine learning approach for queen bee detection through remote audio sensing to safeguard honeybee colonies. *IEEE Transactions on AgriFood Electronics*, 2024.

[80] Dimitrios Kampelopoulos, Ioannis Sofianidis, Chrysoula Tananaki, Konstantina Tsiapali, Spyridon Nikolaidis, and Kostas Siozios. Analyzing the beehive's sound to monitor the presence of the queen bee. In *2022 Panhellenic Conference on Electronics & Telecommunications (PACET)*, pages 1–4. IEEE, 2022.

[81] Andrea De Simone, Luca Barbisan, Giovanna Turvani, and Fabrizio Riente. Iot-based bee colony health monitoring : A focus on energy impact and audio feature extraction. In *2024 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, pages 289–294. IEEE, 2024.

[82] Panuwat Mekha, Nutnicha Teeyasuksaet, Tanapong Sompowloy, and Khukrit Osathanunkul. Honey bee sound classification using spectrogram image features. In *2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, pages 205–209. IEEE, 2022.

[83] Bernardo B *Gatto, Juan G Colonna, Eulanda M dos Santos, Alessandro L Koerich, and Kazuhiro Fukui. Discriminative singular spectrum classifier with applications on bioacoustic signal recognition. *arXiv preprint arXiv :2103.10166*, 2021.

[84] Aleksandra Zlatkova, Zivko Kokolanski, and Dimitar Tashkovski. Honeybees swarming detection approach by sound signal processing. In *2020 XXIX International Scientific Conference Electronics (ET)*, pages 1–3. IEEE, 2020.

[85] Grzegorz Krzywoszyja, Ryszard Rybski, and Grzegorz Andrzejewski. Bee swarm detection based on comparison of estimated distributions samples of sound. *IEEE Transactions on Instrumentation and Measurement*, 68(10) :3776–3784, 2018.

[86] Andrej Žgank. Acoustic monitoring and classification of bee swarm activity using mfcc feature extraction and hmm acoustic modeling. In *2018 ELEKTRO*, pages 1–4. IEEE, 2018.

[87] NIRMAL ANAND, V BHARATH RAJ, MS Ullas, and ASHUTOSH SRIVASTAVA. Swarm detection and beehive monitoring system using auditory and microclimatic analysis. In *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, pages 1–4. IEEE, 2018.

[88] Aleksandra Zlatkova, Branislav Gerazov, Dimitar Tashkovski, and Zivko Kokolanski. Analysis of parameters in algorithms for signal processing for swarming of honeybees. In *2020 28th Telecommunications Forum (TELFOR)*, pages 1–4. IEEE, 2020.

[89] Stefania Cecchi, Alessandro Terenzi, Simone Orcioni, and Francesco Piazza. Analysis of the sound emitted by honey bees in a beehive. In *Audio Engineering Society Convention 147*. Audio Engineering Society, 2019.

[90] Andrej Zgank. Iot-based bee swarm activity acoustic classification using deep neural networks. *Sensors*, 21(3) :676, 2021.

[91] Andrej Zgank. Bee swarm activity acoustic classification for an iot-based farm service. *Sensors*, 20(1) :21, 2020.

[92] Loren Hord and Errol Shook. Determining honey bee behaviors from audio analysis. *National Science Foundation Research Experience for Teachers. Appalachian State Univ*, 2013.

[93] Kiromitis I Dimitrios, Christos V Bellos, Konstantinos A Stefanou, Georgios S Stergios, Ioannis Andrikos, Thomas Katsantas, and Sotirios Kontogiannis. Performance evaluation of classification algorithms to detect bee swarming events using sound. *Signals*, 3(4) :807–822, 2022.

[94] Stenford Ruvinga, Gordon Hunter, Jean-Christophe Nebel, and Olga Duran. Prediction of honeybee swarms using audio signals and convolutional neural networks. In *Workshops at 18th International Conference on Intelligent Environments (IE2022)*, pages 146–154. IOS Press, 2022.

[95] Dániel T Várkonyi, Dóra T Bányai, and Annamária R Varkonyi-Koczy. Investigating traditional machine learning models and the utility of audio features for lightweight swarming prediction in beehives. *Acta Polytechnica Hungarica*, 21(10) :283–299, 2024.

[96] Kainat Iqbal, Bayan Alabdullah, Naif Al Mudawi, Asaad Algarni, Ahmad Jalal, and Jeongmin Park. Empirical analysis of honeybees acoustics as biosensors signals for swarm prediction in beehives. *Ieee Access*, 2024.

[97] Christopher Campell, R Mitchell Parry, and Rahman Tashakkori. Non-invasive spectral-based swarm detection. In *SoutheastCon 2023*, pages 253–260. IEEE, 2023.

[98] Andrej Zgank. Reduced mfcc feature extraction dimension for acoustic classification of bee swarm activity. In *2022 ELEKTRO (ELEKTRO)*, pages 1–4. IEEE, 2022.

[99] Pascal Janetzky, Melanie Schaller, Anna Krause, and Andreas Hotho. Swarming detection in smart beehives using auto encoders for audio data. In *2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 1–5. IEEE, 2023.

[100] Satoshi Kawakita and Kotaro Ichikawa. Automated classification of bees and hornet using acoustic analysis of their flight sounds. *Apidologie*, 50(1) :71–79, 2019.

[101] Anton Gradišek, Gašper Slapničar, Jure Šorn, Mitja Luštrek, Matjaž Gams, and Janez Grad. Predicting species identity of bumblebees through analysis of flight buzzing sounds. *Bioacoustics*, 26(1) :63–76, 2017.

[102] Martin Gjoreski, Borut Budna, Anton Gradišek, and Matjaž Gams. Jsi sound–a machine-learning tool in orange for simple biosound classification. 2017.

[103] Urszula Libal and Pawel Biernacki. Non-intrusive system for honeybee recognition based on audio signals and maximum likelihood classification by autoencoder. *Sensors*, 24(16) :5389, 2024.

[104] Alef Iury Siqueira Ferreira, Nádia Felix Felipe da Silva, Fernanda Neiva Mesquita, Thierson Couto Rosa, Stephen L Buchmann, and José Neiva Mesquita-Neto. Transformer models improve the acoustic recognition of buzz-pollinating bee species. *Ecological Informatics*, page 103010, 2025.

[105] Rahman Tashakkori, Gurney B Buchanan, and Luke M Craig. Analyses of audio and video recordings for detecting a honey bee hive robbery. In *2020 SoutheastCon*, pages 1–6. IEEE, 2020.

[106] David Heise, Zachary Miller, Mackenzie Wallace, and Candace Galen. Bumble bee traffic monitoring using acoustics. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE, 2020.

[107] David Heise, Zachary Miller, Ellie Harrison, Anton Gradišek, Janez Grad, and Candace Galen. Acoustically tracking the comings and goings of bumblebees. In *2019 IEEE Sensors Applications Symposium (SAS)*, pages 1–6. IEEE, 2019.

[108] Anton Gradišek, N Cheron, D Heise, C Galen, and J Grad. Monitoring bumblebee daily activities using microphones. In *Proceedings of the 21st Annual International Multiconference Information Society–IS 2018*, pages 5–8, 2018.

[109] A ZACEPINS et al. Beekeeping in the future—smart apiary management. 2016. 808-812 p. *Disponıvel em :< https ://ieeexplore. ieee. org/document/7501207/>. Acesso em*, 10, 2018.

[110] Fiona Edwards Murphy, Bruno Srbinovski, Michele Magno, Emanuel M Popovici, and Pádraig M Whelan. An automatic, wireless audio recording node for analysis of beehives. In *2015 26th Irish Signals and Systems Conference (ISSC)*, pages 1–6. IEEE, 2015.

[111] Agbotiname Lucky Imoize, Samuel Damilola Odeyemi, and John Adetunji Adebisi. Development of a low-cost wireless bee-hive temperature and sound monitoring system. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, 8(3) :476–485, 2020.

[112] Dániel Tamás Várkonyi, José Luis Seixas Junior, and Tomáš Horváth. Dynamic noise filtering for multi-class classification of beehive audio data. *Expert Systems with Applications*, 213 :118850, 2023.

[113] Kartikeye Dangwal, Iman Ardekani, Soheil Varastehpour, and Abdolhossein Sarrafzadeh. Bioacoustic data augmentation techniques for data-driven apiculture. In *2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS)*, pages 110–115. IEEE, 2024.

[114] Benjamen Underwood and Rahman Tashakkori. Detecting anomalies in honey bee hives using their audio recordings. In *SoutheastCon 2022*, pages 173–177. IEEE, 2022.

[115] Tymoteusz Cejrowski and Julian Szymański. Detection of anomalies in bee colony using transitioning state and contrastive autoencoders. *Computers and Electronics in Agriculture*, 200 :107207, 2022.

[116] Sheng-Hao Chen, Jen-Cheng Wang, Hung-Jen Lin, Mu-Hwa Lee, An-Chi Liu, Yueh-Lung Wu, Pei-Shou Hsu, En-Cheng Yang, and Joe-Air Jiang. 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 Systems with Applications*, 255 :124898, 2024.

[117] Tymoteusz Cejrowski, Julian Szymański, and Doina Logofătu. Buzz-based recognition of the honeybee colony circadian rhythm. *Computers and Electronics in Agriculture*, 175 :105586, 2020.

[118] Sander Van Goethem, Stijn Verwulgen, Frank Goethijn, and Jan Steckel. An iot solution for measuring bee pollination efficacy. In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 837–841. IEEE, 2019.

[119] Nicole E Miller-Struttmann, David Heise, Johannes Schul, Jennifer C Geib, and Candace Galen. Flight of the bumble bee : Buzzes predict pollination services. *PloS one*, 12(6) :e0179273, 2017.

[120] Muhammad Zahid Sharif, Fernando Wario, Nayan Di, Renjie Xue, and Fanglin Liu. Soundscape indices : New features for classifying beehive audio samples. *Sociobiology*, 67(4) :566–571, 2020.

[121] Yangguang Zhao, Guoqing Deng, Long Zhang, Nayan Di, Xueli Jiang, and Zhigang Li. Based investigate of beehive sound to detect air pollutants by machine learning. *Ecological Informatics*, 61 :101246, 2021.

[122] Nicolás Pérez, Florencia Jesús, Cecilia Pérez, Silvina Niell, Alejandro Draper, Nicolás Obrusnik, Pablo Zinemanas, Yamandú Mendoza Spina, Leonidas Carrasco Letelier, and Pablo Monzón. Continuous monitoring of beehives' sound for environmental pollution control. *Ecological engineering*, 90 :326–330, 2016.

[123] Gordon Hunter, Donald Howard, Sylvain Gauvreau, Olga Duran, and Rosa Busquets. Processing of multi-modal environmental signals recorded from a" smart" beehive. *Proceedings of the Institute of Acoustics*, 41(1) :339–348, 2019.

[124] Baizhong Yu, Xinqiu Huang, Muhammad Zahid Sharif, Xueli Jiang, Nayan Di, and Fanglin Liu. A matter of the beehive sound : Can honey bees alert the pollution out of their hives? *Environmental Science and Pollution Research*, 30(6) :16266–16276, 2023.

[125] Nayan Di, Muhammad Zahid Sharif, Zongwen Hu, Renjie Xue, and Baizhong Yu. Applicability of vggish embedding in bee colony monitoring : Comparison with mfcc in colony sound classification. *PeerJ*, 11 :e14696, 2023.

[126] Nayan Di, Chunjing Zhu, Zongwen Hu, Muhammad Zahid Sharif, Baizhong Yu, and Fanglin Liu. Honeybee colony soundscapes : Decoding distance-based cues and environmental stressors. *Ecotoxicology and Environmental Safety*, 297 :118241, 2025.

[127] Bartłomiej Zubrzak, Paweł Bieńkowski, Paweł Cała, Przemysław Płaskota, Krzysztof Rudno-Rudziński, and Piotr Nowakowski. Thermal and acoustic changes in bee colony due to exposure to microwave electromagnetic field–preliminary research. *Przegląd Elektrotechniczny*, 94(12), 2018.

[128] Candace Galen, Zachary Miller, Austin Lynn, Michael Axe, Samuel Holden, Levi Storks, Eddie Ramirez, Emilia Asante, David Heise, Susan Kephart, et al. Pollination on the dark side : Acoustic monitoring reveals impacts of a total solar eclipse on flight behavior and activity schedule of foraging bees. *Annals of the Entomological Society of America*, 112(1) :20–26, 2019.

[129] Satoshi Kawakita, Kotaro Ichikawa, Fumio Sakamoto, and Kazuyuki Moriya. Sound recordings of apis cerana japonica colonies over 24 h reveal unique daily hissing patterns. *Apidologie*, 50(2) :204–214, 2019.

[130] Satoshi Kawakita, Kotaro Ichikawa, Fumio Sakamoto, and Kazuyuki Moriya. Hissing of a. cerana japonica is not only a direct aposematic response but also a frequent behavior during daytime. *Insectes Sociaux*, 65(2) :331–337, 2018.

[131] Henja-Niniane Wehmann, David Gustav, Nicholas H Kirkerud, and C Giovanni Galizia. The sound and the fury—bees hiss when expecting danger. *PloS one*, 10(3) :e0118708, 2015.

[132] Paul A De Luca, Noah Giebink, Andrew C Mason, Dan Papaj, and Stephen L Buchmann. How well do acoustic recordings characterize properties of bee (anthophila) floral sonication vibrations? *Bioacoustics*, 29(1) :1–14, 2020.

[133] Vladimir A Kulyukin and Sai Kiran Reka. Toward sustainable electronic beehive monitoring : Algorithms for omnidirectional bee counting from images and harmonic analysis of buzzing signals. *Engineering Letters*, 24(3), 2016.

[134] James H Cane. The oligolectic bee osmia brevis sonicates penstemon flowers for pollen : a newly documented behavior for the megachilidae. *Apidologie*, 45(6) :678–684, 2014.

[135] Thomas Schlegel, P Kirk Visscher, and Thomas D Seeley. Beeping and piping : characterization of two mechano-acoustic signals used by honey bees in swarming. *Naturwissenschaften*, 99(12) :1067–1071, 2012.

[136] Mandeep Tayal and Rupesh Kariyat. Examining the role of buzzing time and acoustics on pollen extraction of solanum elaeagnifolium. *Plants*, 10(12) :2592, 2021.

[137] Tymoteusz Cejrowski and Julian Szymański. Buzz-based honeybee colony fingerprint. *Computers and Electronics in Agriculture*, 191 :106489, 2021.

[138] Zachary J Miller. What's the buzz about? progress and potential of acoustic monitoring technologies for investigating bumble bees. *IEEE Instrumentation & Measurement Magazine*, 24(7) :21–29, 2021.

[139] Paul A De Luca, Darryl A Cox, and Mario Vallejo-Marín. Comparison of pollination and defensive buzzes in bumblebees indicates species-specific and context-dependent vibrations. *Naturwissenschaften*, 101(4) :331–338, 2014.

[140] Christopher Campell, R Mitchell Parry, Rahman Tashakkori, Alex Somer, Logan Richardson, and Aedan Simons-Rudolph. Honey bee piping detection utilizing convolutional neural networks. *IEEE Sensors Journal*, 2025.

[141] Yi Zhu, Mahsa Abdollahi, Ségolène Maucourt, Nico Coallier, Heitor R. Guimarães, Pierre Giovenazzo, and Tiago H. Falk. Early prediction of honeybee hive winter survivability using multi-modal sensor data. In *2023 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, pages –. IEEE, 2023.

[142] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

[143] Damir Šekulja, Hermann Pechhacker, and Elisabeth Licek. Drifting behavior of honey bees (apis mellifera carnica pollman, 1879) in the epidemiology of american foulbrood. *Zbornik Veleučilišta u Rijeci*, 2(1) :345–358, 2014.

[144] Rahman Tashakkori, Nathaniel P Hernandez, Ahmad Ghadiri, Aleksander P Ratzloff, and Michael B Crawford. A honeybee hive monitoring system : From surveillance cameras to raspberry pis. In *SoutheastCon 2017*, pages 1–7. IEEE, 2017.

[145] David J Kale, Rahman Tashakkori, and R Mitchell Parry. *Automated beehive surveillance using computer vision*. IEEE, 2015.

[146] Émilie Grenier, Pierre Giovenazzo, Carl Julien, and Isabelle Goupil-Sormany. Honeybees as a biomonitoring species to assess environmental airborne pollution in different socioeconomic city districts. *Environmental Monitoring and Assessment*, 193(11) :1–14, 2021.

[147] F Como, E Carnesecchi, S Volani, JL Dorne, J Richardson, A Bassan, M Pavan, and E Benfenati. Predicting acute contact toxicity of pesticides in honeybees (apis mellifera) through a k-nearest neighbor model. *Chemosphere*, 166 :438–444, 2017.

[148] Hugh M Robertson and Kevin W Wanner. The chemoreceptor superfamily in the honey bee, apis mellifera : expansion of the odorant, but not gustatory, receptor family. *Genome research*, 16(11) :1395–1403, 2006.

[149] Stefan Fuchs and Jürgen Tautz. Colony defence and natural enemies. In *Honeybees of Asia*, pages 369–395. Springer, 2011.

[150] WH Kirchner and J Röschard. Hissing in bumblebees : an interspecific defence signal. *Insectes sociaux*, 46(3) :239–243, 1999.

[151] J Heidelbach, H Böhm, and WH Kirchner. Sound and vibration signals in a bumble bee colony (bombus terrestris). *Zoology*, 101(suppl I) :82, 1998.

[152] Alison Pereira Ribeiro, Nádia Felix Felipe da Silva, Fernanda Neiva Mesquita, Priscila de Cássia Souza Araújo, Thierson Couto Rosa, and José Neiva Mesquita-Neto. Machine learning approach for automatic recognition of tomato-pollinating bees based on their buzzing-sounds. *PLoS computational biology*, 17(9) :e1009426, 2021.

[153] Callin M Switzer, Katja Hogendoorn, Sridhar Ravi, and Stacey A Combes. Shakers and head bangers : differences in sonication behavior between australian amegilla murrayensis (blue-banded bees) and north american bombus impatiens (bumblebees). *Arthropod-Plant Interactions*, 10(1) :1–8, 2016.

[154] Alef Iury Siqueira Ferreira, Nádia Felix Felipe Da Silva, Fernanda Neiva Mesquita, Thierson Couto Rosa, Victor Hugo Monzón, and José Neiva Mesquita-Neto. Automatic acoustic recognition of pollinating bee species can be highly improved by deep learning models accompanied by pre-training and strong data augmentation. *Frontiers in Plant Science*, 14 :1081050, 2023.

[155] Martin Bencsik, Joseph Bencsik, Michael Baxter, Andrei Lucian, Julien Romieu, and Mathias Millet. Identification of the honey bee swarming process by analysing the time course of hive vibrations. *Computers and electronics in agriculture*, 76(1) :44–50, 2011.

[156] Steven Davis and Paul Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 28(4) :357–366, 1980.

[157] B Yegnanarayana, SR Mahadeva Prasanna, Jinu Mariam Zachariah, and Cheedella S Gupta. Combining evidence from source, suprasegmental and spectral features for a fixed-text speaker verification system. *IEEE Transactions on speech and audio processing*, 13(4) :575–582, 2005.

[158] Edward Farrington Woods. Electronic prediction of swarming in bees. *Nature*, 184(4690) :842–844, 1959.

[159] Norden E Huang. Introduction to the hilbert–huang transform and its related mathematical problems. In *Hilbert–Huang transform and its applications*, pages 1–26. World Scientific, 2014.

[160] Bishnu S Atal and Suzanne L Hanauer. Speech analysis and synthesis by linear prediction of the speech wave. *The journal of the acoustical society of America*, 50(2B) :637–655, 1971.

[161] Beáta T Szabó, Susan L Denham, and István Winkler. Computational models of auditory scene analysis : a review. *Frontiers in Neuroscience*, 10 :524, 2016.

[162] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3) :273–297, 1995.

[163] Leo Breiman. Bagging predictors. *Machine learning*, 24(2) :123–140, 1996.

[164] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.

[165] Tianqi Chen and Carlos Guestrin. Xgboost : A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[166] Telma Garção, Joana Sousa, Luis André, and João Ferreira. Bee-yond buzz : Exploring deep learning techniques for beehive audio classification. In *2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–5. IEEE, 2024.

[167] Mingxing Tan and Quoc Le. Efficientnetv2 : Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.

[168] Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, and Mark D Plumbley. Panns : Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28 :2880–2894, 2020.

[169] Open source beehives project, available online :. `https://www.osbeehives.com/pages/about-us`. Accessed : 2021-10-13.

[170] Mahsa Abdollahi, Yi Zhu, Heitor R Guimarães, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H Falk. Urban : Urban beehive acoustics and phenotyping dataset. *Scientific Data*, 12(1) :536, 2025.

[171] Vladimir Kulyukin. Audio, image, video, and weather datasets for continuous electronic beehive monitoring. *Applied Sciences*, 11(10) :4632, 2021.

[172] Stefania Cecchi, Alessandro Terenzi, Simone Orcioni, Paola Riolo, Sara Ruschioni, and Nunzio Isidoro. A preliminary study of sounds emitted by honey bees in a beehive. In *Audio Engineering Society Convention 144*. Audio Engineering Society, 2018.

[173] Yi Zhu, Mahsa Abdollahi, Ségolène Maucourt, Nico Coallier, Heitor R Guimarães, Pierre Giovenazzo, and Tiago H Falk. Mspb : a longitudinal multi-sensor dataset with phenotypic trait measurements from honey bees. *Scientific Data*, 11(1) :860, 2024.

[174] Anna Yang. Smart bee colony monitor : Clips of beehive sounds. `https://www.kaggle.com/datasets/annajyang/beehive-sounds/data`. Accessed : 2024-05-23.

[175] Stan Chabert, Fabrice Requier, Joël Chadoeuf, Laurent Guilbaud, Nicolas Morison, and Bernard E Vaissiere. Rapid measurement of the adult worker population size in honey bees. *Ecological Indicators*, 122 :107313, 2021.

[176] Mahsa Abdollahi, Evan Henry, Pierre Giovenazzo, and Tiago H Falk. The importance of context awareness in acoustics-based automated beehive monitoring. *Applied Sciences*, 13(1) :195, 2022.

[177] Vincent Dietemann, Francesco Nazzi, Stephen J Martin, Denis L Anderson, Barbara Locke, Keith S Delaplane, Quentin Wauquiez, Cindy Tannahill, Eva Frey, Bettina Ziegelmann, et al. Standard methods for varroa research. *Journal of apicultural research*, 52(1) :1–54, 2013.

[178] Patrick L Kohl, Benjamin Rutschmann, Luis G Sikora, Norbert Wimmer, Volker Zahner, Paul D'Alvise, Martin Hasselmann, and Ingolf Steffan-Dewenter. Parasites, depredators, and limited resources as potential drivers of winter mortality of feral honeybee colonies in german forests. *Oecologia*, 202(3) :465–480, 2023.

[179] Melanie Kempers, Paul Kozak, Chris Maund, Cameron Menzies, Derek Micholson, Nuria Morfin, Samantha Muirhead, Medhat Nasr, Steve Pernal, Jason Sproule, et al. Statement on honey bee wintering losses in canada for 2023.

[180] Mahsa Abdollahi, Yi Zhu, Heitor R. Guimarães, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H. Falk. Urban : Urban beehive acoustics and phenotyping dataset, 2024. `https://doi.org/10.20383/103.0972`.

[181] Mahsa Abdollahi, Nico Coallier, Pierre Giovenazzo, and Tiago H Falk. Performance comparison of voice activity detectors for acoustic beehive monitoring. In *2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 320–323. IEEE, 2023.

[182] Stephane G Mallat. Multiresolution approximations and wavelet orthonormal bases of $l^2(r)$. *Transactions of the American mathematical society*, 315(1) :69–87, 1989.

[183] Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

[184] Urszula Libal and Pawel Biernacki. Mfcc selection by lasso for honey bee classification. *Applied Sciences*, 14(2) :913, 2024.

[185] Kirill Sakhnov, Ekaterina Verteletskaya, and Boris Simak. Dynamical energy-based speech/silence detector for speech enhancement applications. In *Proceedings of the world congress on engineering*, volume 1, page 2, 2009.

[186] Xiao-Lei Zhang and Ji Wu. Deep belief networks based voice activity detection. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(4) :697–710, 2012.

[187] Google. Webrtc. `https://www.webrtc.org/`. Accessed : 2023-04-28.

[188] Lawrence R Rabiner. *Digital processing of speech signals*. Pearson Education India, 1978.

[189] John Wiseman. Python interface to the webrtc voice activity detector. `https://www.github.com/wiseman/py-webrtcvad/`, 2016. Accessed : 2023-04-28.

[190] Mirco Ravanelli, Titouan Parcollet, Peter Plantinga, Aku Rouhe, Samuele Cornell, Loren Lugosch, Cem Subakan, Nauman Dawalatabad, Abdelwahab Heba, Jianyuan Zhong, et al. Speechbrain : A general-purpose speech toolkit. *arXiv preprint arXiv :2106.04624*, 2021.

[191] Sibo Tong, Hao Gu, and Kai Yu. A comparative study of robustness of deep learning approaches for vad. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5695–5699. IEEE, 2016.

[192] Mahsa Abdollahi, Nico Coallier, Pierre Giovenazzo, and Tiago H. Falk. Performance comparison of voice activity detectors for acoustic beehive monitoring. In *2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 320–323, 2023.

[193] Heitor R. Guimarães, Mahsa Abdollahi, Yi Zhu, Ségolène Maucourt, Nico Coallier, Pierre Giovenazzo, and Tiago H. Falk. Adapting self-supervised features for background speech detection in beehive audio recordings. In *2023 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, pages 663–667, 2023.

[194] Steven Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on acoustics, speech, and signal processing*, 27(2) :113–120, 1979.

[195] L Lin, WH Holmes, and E Ambikairajah. Adaptive noise estimation algorithm for speech enhancement. *ELECTRONICS LETTERS-IEE*, 39(9) :754–754, 2003.

[196] Mahsa Abdollahi, Yi Zhu, Heitor R. Guimarães, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H. Falk. Audio modulation spectral features for improved honeybee colony population prediction. *IEEE Sensors Journal*, pages 1–1, 2025.

[197] Mahsa Abdollahi, Yi Zhu, Heitor R. Guimarâes, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H. Falk. On the prediction of varroa mite infestations in honeybee colonies via acoustic monitoring. *IEEE Sensors Journal*, pages 1–1, 2026.

[198] Abhishek Tiwari, Raymundo Cassani, Shruti Kshirsagar, Diana P Tobon, Yi Zhu, and Tiago H Falk. Modulation spectral signal representation for quality measurement and enhancement of wearable device data : A technical note. *Sensors*, 22(12) :4579, 2022.

[199] Raymundo Cassani and Tiago H Falk. Spectrotemporal modeling of biomedical signals : Theoretical foundation and applications. *Reference module in biomedical sciences*, 2018.

[200] Yi Zhu and Tiago H Falk. Spectral–temporal saliency masks and modulation tensorgrams for generalizable covid-19 detection. *Computer Speech & Language*, 86 :101620, 2024.

[201] Benjamin Cauchi, Kai Siedenburg, Joao F Santos, Tiago H Falk, Simon Doclo, and Stefan Goetze. Non-intrusive speech quality prediction using modulation energies and lstm-network. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(7) :1151–1163, 2019.

[202] Anderson R Avila, Zahid Akhtar, Joao F Santos, Douglas O'Shaughnessy, and Tiago H Falk. Feature pooling of modulation spectrum features for improved speech emotion recognition in the wild. *IEEE Transactions on Affective Computing*, 12(1) :177–188, 2018.

[203] Mohammed Senoussaoui, João F Santos, and Tiago H Falk. Speech temporal dynamics fusion approaches for noise-robust reverberation time estimation. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5545–5549. IEEE, 2017.

[204] Garima Sharma, Kartikeyan Umapathy, and Sridhar Krishnan. Trends in audio signal feature extraction methods. *Applied Acoustics*, 158 :107020, 2020.

[205] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8) :1226–1238, 2005.

[206] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[207] Milton Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200) :675–701, 1937.

[208] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan) :1–30, 2006.

[209] Peter Bjorn Nemenyi. *Distribution-free multiple comparisons.* Princeton University, 1963.

[210] George H John and Pat Langley. Estimating continuous distributions in bayesian classifiers. *arXiv preprint arXiv :1302.4964*, 2013.

[211] Leo Breiman. Random forests. *Machine learning*, 45 :5–32, 2001.

[212] Jerome H Friedman. Greedy function approximation : a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[213] Evelyn Fix. *Discriminatory analysis : nonparametric discrimination, consistency properties*, volume 1. USAF school of Aviation Medicine, 1985.

[214] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088) :533–536, 1986.

[215] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn : Machine learning in python. *the Journal of machine Learning research*, 12 :2825–2830, 2011.

[216] Geoffroy Peeters et al. A large set of audio features for sound description (similarity and classification) in the cuidado project. *CUIDADO Ist Project Report*, 54(0) :1–25, 2004.

[217] Honey Bee Health Coalition. Tools for varroa management : a guide to effective varroa sampling & control, 2018.

[218] Virginia Tech Entomology. Varroa mite sampling methods. `https://www.ento.vt.edu/the-bee-group-at-vt/beekeeping/mites2.html`, 2021.

[219] Ontario Beekeepers' Association. Treatment recommendations, 2017. Accessed : 2025-08-18.

[220] Mahsa Abdollahi, Yi Zhu, Heitor R. Guimarães, Nico Coallier, Ségolène Maucourt, Pierre Giovenazzo, and Tiago H. Falk. Improved monitoring of honeybee colony strength via audio iot sensors, modulation tensorgrams and recurrent neural networks. *Submitted to IEEE Sensors Journal*, 2026.

[221] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam : Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.