A Feature Engineering Focused System for Acoustic UAV Payload Detection

Yaqin Wang[1], Facundo Esquivel Fagiani[2], Kar Ee Ho[1] and Eric T. Matson[1]
[1]*Computer and Information Technology, Purdue University, West Lafayette, IN, U.S.A.*
[2]*Renard Analytics, San Miguel de Tucumán, Argentina*

Keywords: Audio Classification, UAV Classification, Machine Learning, Drone Security, Payload Detection, Acoustic Classification, Neural Network, Feature Extraction.

Abstract: The technology evolution of Unmanned Aerial Vehicles (UAVs) or drones, has made these devices suitable for a wide new range of applications, but it has also raised safety concerns as drones can be used for carrying explosives or weapons with malicious intentions. In this paper, Machine Learning (ML) algorithms are used to identify drones carrying payloads based on the sound signals they emit. We evaluate and propose a feature-based classification. Five individual features, and one combinations of features are used to train four different standard machine learning models: SupportVector Machine (SVM), Gaussian Naive Bayes (GNB), K-Nearest Neighbor (KNN) and a Neural Network (NN) model. The training and testing dataset is composed of sound samples of loaded drones and unloaded drones collected by the team. The results show that the combination of features outperforms the individual ones, with much higher accuracy scores.

1 INTRODUCTION

Unmanned Aerial Vehicles (UAVs), also called drones, have seen an exponential increase in popularity in recent years (Intelligence, 2021). As the technology evolves, drones have become cheaper and smaller, which allows a wide range of new applications, from filming sporting events to homeland security, the possibilities are endless. On the negative side, the increase accessibility to these devices also poses a threat. Smaller and more potent UAVs allow them to invade restricted zones without being detected, and to carry potentially harmful payloads, such as weapons and explosives. Alarming examples of UAV threats can be cited, such as the case of a drone landing inside the perimeter of the U.S. White House in 2015 (Schmidt and Shear, 2015), the attacks to German chancellor Angela Merkel in 2013 (Lee, 2013), and Venezuelan president Nicolás Maduro in 2018 (Koettl and Marcolin, 2018). Under this context, it is of special interest to be able to identify the presence of loaded drones, since they usually represent a higher risk than an unloaded drone.

The topic to be addressed in this study is the classification of loaded and unloaded drones, and acoustic detection is the chosen approach. It is a cost-effective solution, and despite the limitation in noise in the real scenarios, such as bird singing or wind, (Case et al., 2008). The acoustic method has provided promising results on drone sound classification (Bernardini et al., 2017; Seo et al., 2018). Regarding loaded and unloaded drone classification, when a drone carries a payload, its rotors have to unfailingly increase the rotational speed in order to keep its height, this produces a different sound profile, which can potentially be identified by an acoustic recognition model(Li et al., 2018).

This study was developed from (Fagiani, 2021; Wang et al., 2021), by using acoustic signals to detect UAVs and to map their exact location. However, this study focuses on using features (and combinations of features) as inputs for the ML models, and compare their performances. The selected feature extraction methods include mfcc, chroma, mel, contrast,and tonnetz (librosa Development Team, 2021). We also used a combination of the five features to compare with their individual performance. Features from these different methods are used to feed into four different standard machine learning models: Support Vector Machine (SVM), Gaussian Naive Bayes (GNB), K-Nearest Neighbor (KNN) and a Neural Network (NN) model. The dataset we use for training and testing was collected by the team. The two drones that are used to collect audio recordings are DJI Phantom 4 and an

Wang, Y., Fagiani, F., Ho, K. and Matson, E.
A Feature Engineering Focused System for Acoustic UAV Payload Detection.
DOI: 10.5220/0010843800003116
In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022) - Volume 3*, pages 470-475
ISBN: 978-989-758-547-0; ISSN: 2184-433X

EVO 2 Pro.

The contribution of this paper is to provide an alternative approach in building an audio-based payload classification for drones, using feature extraction. We also built a small size drone audio database that will be available to the public. We used the same ML model structures for different feature extraction settings, and the results showed that the combination of features have a better performance than individual ones. The rest of this paper is organized in the sections as follow. Section 2 reviews the current sound detection methods for drone classification, and the payload classification. Section 3 shows the methodology proposed for the feature extraction methods and three different ML models. Section 4 describes our experiments and results. Lastly, section 5 presents the conclusion and future works.

2 LITERATURE REVIEW

2.1 Sound Recognition Solution

A variety of methods have been provided to detect drones using sound detection (Mezei et al., 2015; Jeon et al., 2017; Fagiani, 2021; Kim et al., 2017). The rotation of the drone's rotor blades create an audible signature that can be sensed and recorded, even within the range of human hearing, but the question is if and how these signatures can be distinguished from other sounds. In one particular study, two different methods to achieve drone sound detection included mathematical correlation and audio fingerprinting (Mezei et al., 2015).

In the first case, the researchers employed a method similar to global positioning systems (GPS) work. To apply this methodology to the sound of drones, the researchers created a library of sounds by taking audio recordings of a lawnmower, hair dryer, music, a model airplane, and two drones. The sounds were dismantled to isolate their individual components. The samples were then compared through the process of correlation using two techniques: Pearson's correlation coefficient and normalized maximum correlation. In summary, the researchers demonstrated that the correlation techniques worked with the system correctly identifying the drones sounds versus other sounds at a level of 65.6% and 77.9% accuracy. For reference, the drone sounds were recorded at a distance of approximately 3 meters or less and in a relatively sound proof room.

In the second case, the researchers employed a technique called audio fingerprinting, which is basically the algorithm that operates the popular mobile device application called Shazam. Shazam operates by allowing the user to record a short audio sample from the ambient sound of a song playing nearby using the mobile device's built-in microphone. To simulate this capability, the researchers used an open source tool called MusicG from GitHub. Then they recorded samples of drone sounds, but this time within 1 meter distance and again in a sound controlled room. Overall, the researchers found both methods to achieve acceptable results. Future work intends to overcome limitations with regards to equipment quality and distance from subject to microphone.

Another promising study to detect drone sounds was conducted by (Jeon et al., 2017) using MFCC with GMM and two types of deep neural networks (DNN), convolutional neural network (CNN) and recurrent neural network (RNN). The unique aspect of this research was the emphasis on using polyphonic sound data from real-life environments. In other words, the focus was on identifying and classifying drone sounds from a diverse background of competing noises in the environment. One significant challenge that the research team faced was the paucity of publicly available drone sound data. To remedy this problem, the team implemented a novel technique by synthesizing tracks of drone sounds with tracks of background noise to create a coherent audio clip. In this case, the sample drone sounds were generated from DJI Phantom 3 and Phantom 4, with the background noise of people talking, car traffic, and airplane noise. The drone sounds were recorded at distances of 30m, 70m, and 150m while both hovering and approaching. Overall, RNN achieved the best performance with F-score being (RNN $>$ CNN $>$ GMM: 0.8809 $>$ 0.6451 $>$ 0.5232) coming from 240 ms of audio input. Precision and recall were also highest with RNN at (0.7953, 0.8066).

In (Kim et al., 2017), the researchers sought to develop a real-time drone detection and analysis system using sound data from DJI Phantom 1 and 2 drones and environmental noise from a European football stadium. Two different machine learning algorithms were employed. The plotted image machine learning (PIL) technique resulted in 83% accuracy and K-nearest neighbor (KNN) achieved 61% accuracy. These self-learning techniques also resulted in improvements of detection efficiency as well. The downsides of using PIL is that it requires large data sets and has a tendency to reveal bias in the result. For KNN, the limitation includes a difficulty to distinguish between similar but different drone targets, despite it being a fast and simple approach. The study's intent to produce a general UAV detection system were also limited by not being able to test both algorithms with the same drone types.

2.2 Loaded and Unloaded Drones Recognition

In recent years, drones are becoming more and more popular in both recreational and commercial purposes. A micro-drone is relatively cheap and not too difficult to use. It can be used for recreational activities, such as filming, as well as in farming, package deliveries, and more (Ritchie et al., 2017; Pallotta et al., 2020). However, the popularity of using drones has led to potential criminal and dangerous activities, including privacy invasion, illegal flying in restricted areas such as airports, interference in public events, and terrorist attacks with armed drones. Hence, to detect and classify drones with different payloads is crucial in terms of security and safety.

There are only limited number of research in the topic of loaded drones recognition, and most of them are focusing on using micro-Doppler radar to detect and classify loaded and unloaded drones (Ritchie et al., 2017; Pallotta et al., 2020). Palotta et al. proposed a new micro-Doppler feature extraction procedure based on spectral kurtosis to classify UAVs with different payloads. Both of the narrowband and wideband spectrograms from the radar are used in calculating spectral kurtosis, which is used as input to a classifier after a dimensionality reduction stage using principal component analysis (PCA). They have reached an average accuracy of 92.61% for different payloads on the proposed feature extraction procedure.

3 METHODOLOGY

In order to evaluate using feature-based methods for UAV's payload classification, we compared features and ML methods. Using UAV audio recording files as input, we extracted five features (mfcc, chroma, mel, contrast, and tonnetz). In addition one combinations of these features is used. These 5 individual and 1 combinations of features are used to train four ML models. We compared four different ML models, which are SVM, GNB, KNN and NN.

We used a DJI Phantom 4 and an EVO 2 Pro to collect audio recordings, with and without payload. UAV samples were collected at McAllister Park, Lafayette, IN, 47904. We collected a total of 1232 number of samples for loaded and unloaded data, for a total of 204.5 minutes long, as shown in Table 1. The payload we used for both UAVs is a bottle of water with 500ml capacity, which is about 16.9 oz. All the data processing and ML models training are done on a Macbook Air, with 1.1 GHz Quad-Core Intel Core i5 and 8 GB memory.

Table 1: UAV Audio Recording Data.

UAV Type	Quantity	Total Time
Loaded DJI Phantom 4	343	57.16 min
Unloaded DJI Phantom 4	302	50.33 min
Loaded EVO 2 Pro	297	49.50 min
Unloaded EVO 2 Pro	290	48.33 min
Total	1232	204.5 min

Table 2: Feature Extraction Methods.

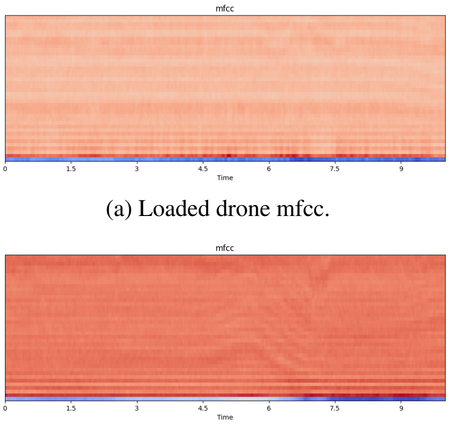
Feature	Shape
chroma_stft	12
chroma_cqt	12
chroma_cqt	12
mel	128
mfcc	40
rms	1
centroid	1
bandwidth	1
contrast	7
flatness	1
bandwidth	1
rolloff	1
poly shape	2
tonnetz	6
zero_crossing	1

3.1 Features Extraction

When a human is asked to identify the sound of an object, they might try to recognize melodic or rhythmic patterns and use them to differentiate them, e.g. traffic sounds, bird singing, and music. Using features for classification may provide "explanations" to understand on how the ML classification was produced. In analyzing and preparing audio files for machine learning training, the process of learning the patterns is feature extraction. In this project, we used the python library, Librosa, for audio feature processing (librosa Development Team, 2021).

Table 2 shows 12 different feature extraction tools in Librosa (librosa Development Team, 2021). The right column is the number of features on each method calculated. All of the feature extraction methods in this table are spectral features except the last two, which are rhythmic features. Spectral features represent sound based on the amount of vibration at each individual frequency.

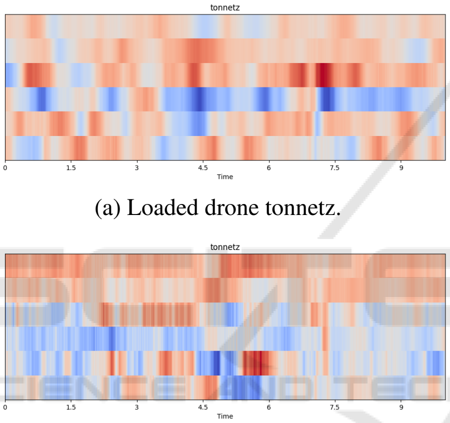
Among the 12 different extraction methods, we removed the ones with the shape of 1 or 2 because individually they would not provide enough information for classification purposes. The selected methods include: mfcc, mel, contrast, chroma, and tonnetz. We explored the individual features and the combination



(a) Loaded drone mfcc.

(b) Unloaded drone mfcc.

Figure 1: MFCC Feature Plot.



(a) Loaded drone tonnetz.

(b) Unloade drone tonnetz.

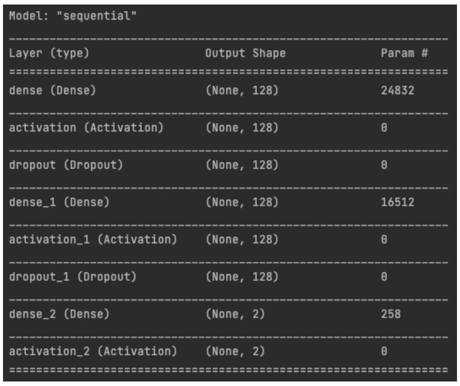
Figure 2: Tonnetz Feature Plot.

of them to see how they perform with different machine learning models. By using the Librosa feature extraction methods, the features from the audio files are saved into numpy array format.

It is difficult for a human to see and understand how these features (a vector of numbers) represent the audio features. Spectral display functions enable visualization of the features. Figure 1 and Figure 2 compare the visualized feature between two audio recordings from loaded and unloaded drones. It is visually obvious that the two audio files are different in sound.

3.2 ML Models

The four models used for training are: SVM, GNB, KNN, and neural network (NN). In these three linear models, we used all default settings for all parameters. The neural network has 3 dense layers, 2 activation layers, and 2 dropout layers, as shown in Figure 3.



```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 128)               24832
_________________________________________________________________
activation (Activation)      (None, 128)               0
_________________________________________________________________
dropout (Dropout)            (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               16512
_________________________________________________________________
activation_1 (Activation)    (None, 128)               0
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 258
_________________________________________________________________
activation_2 (Activation)    (None, 2)                 0
=================================================================
```

Figure 3: Neural Network Model Structure.

In the SVM model, the parameter C is set to 10, and the kernel we chose is linear. For GNB model, we used all default parameters for model training. And in KNN, the parameter n_neighbors is set to 6, with all other ones as default.

4 EXPERIMENTS AND RESULTS

4.1 Experiments

The loaded and unloaded drones audio recordings dataset contains 1232 samples, and each one of them has a sample length of 10 seconds. Experiments on the dataset we collected will provide valid information about which feature extraction method or combination works the best, as well as which machine learning model has the best performance.

The features of each recording under different category are calculated and extracted first, and then saved in the form of numpy arrays. Five individual features are extracted, as well as one combination. The extracted features are the input for the machine learning models. The models will learn the features from the audio files, not the raw audio data. Once the training is done, the program evaluates the model, and provides a test accuracy score. We split the dataset in 70% training and 30% testing then report and compare on the test accuracy, recall, precision and F-1 score.

4.2 Results

Table 3 shows the results for model SVM. We calculated and extracted individual features and combinations described above. Accuracy on combination of features has an obvious increase compared to individual ones. MFCC outperforms the other individual features with the highest accuracy and F-1 score.

Table 3: Test Results for SVM.

Feature	Accuracy	Recall	Precision	F-1
1. chroma	0.965	1.000	0.989	0.995
2. mel	0.914	0.984	0.864	0.920
3. mfcc	0.986	1.00	0.988	0.986
4. contrast	0.786	0.818	0.773	0.795
5. tonnetz	0.676	0.722	0.665	0.692
Combo	0.992	1.00	1.00	1.00

Table 4: Test Results for GNB.

Feature	Accuracy	Recall	Precision	F-1
1. chroma	0.924	0.850	1.00	0.919
2. mel	0.924	1.000	0.870	0.930
3. mfcc	0.997	1.000	0.995	0.997
4. contrast	0.546	0.107	0.952	0.192
5. tonnetz	0.689	0.765	0.668	0.713
Combo	0.997	1.000	0.995	0.997

Table 5: Test Results for KNN.

Feature	Accuracy	Recall	Precision	F-1
1. chroma	0.995	0.995	0.995	0.995
2. mel	0.954	0.947	0.962	0.954
3. mfcc	0.989	0.989	0.989	0.989
4. contrast	0.800	0.790	0.780	0.770
5. tonnetz	0.708	0.781	0.684	0.730
Combo	0.989	0.984	0.995	0.989

Table 6: Test Results for NN.

Feature	Accuracy	Recall	Precision	F-1
1. chroma_stft	0.924	0.850	1.00	0.919
2. mel	0.924	1.000	0.870	0.930
3. mfcc	0.997	1.000	0.995	0.997
4. contrast	0.546	0.107	0.952	0.192
5. tonnetz	0.689	0.765	0.668	0.713
Combo	0.997	1.000	0.995	0.997

Chroma feature also performs well with a recall of 1.

Table 4 shows the test results for the model GNB. Accuracy on combination of features has the best performance in accuracy and all other scores. MFCC and Mel are the two best individual features with the highest accuracy scores. Chroma also has very high accuracy, precision and F-1 score, but a relatively lower recall.

Table 5 shows the results for model KNN. Surprisingly, chroma feature outperforms the others, including the combination features. The combination feature has a slightly lower accuracy score, but still promising.

Table 6 shows the results for the neural network (NN) model. The combination features and MFCC have the same accuracy, recall, precision and F-1 score. Chroma and Mel have similar performance.

In all four machine learning models, the combination of features has the better performance compared to individual ones. MFCC is the best individual feature method for SVM, GNB and NN, while chroma performs the best in KNN.

5 CONCLUSION AND FUTURE WORKS

This paper explored five different feature extraction methods and a combination to classify whether a drone carries payload. The five selected feature extraction methods are chroma, mel, mfcc, contrast, and tonnetz, and the combinations of the five individual ones is also applied and evaluated. Features of each audio recording under each category (loaded and unloaded) are calculated and saved. The four machine learning models that we we used for trainings are SVM, GNB, KNN, and a Neural Network. Those saved features are used as input to feed into the training models. The dataset was collected and labeled by using two different brands and models of the drones. The dataset includes 1232 audio samples of loaded and unloaded drones. The results show that the combination of features have a better performance than individual ones. The combination feature reaches about 99% average accuracy in all four ML models. The best individual features are MFCC and chroma for all four ML models. Our method of feature combination outperforms the research of Palotta et al, with the average accuracy of 92.61%. And our approach requires fewer computational resources, and has higher explainability.

The limitations of our method include that we only used the same payload for all data collecting. Also the amount of data we have is sufficient for current research purpose, but we will need more data for more general UAV payload detection with different manufactures and models. We may need to increase the complexity of the ML models after collecting more data. With more effort in the future, accuracy is expected to improve with more data and optimized models.

ACKNOWLEDGEMENT

The authors appreciate the support and guidance from Dr. Eric Matson.

REFERENCES

Bernardini, A., Mangiatordi, F., Pallotti, E., and Capodiferro, L. (2017). Drone detection by acoustic signature identification. *Electronic Imaging*, 2017(10):60–64.

Case, E. E., Zelnio, A. M., and Rigling, B. D. (2008). Low-cost acoustic array for small uav detection and tracking. In *2008 IEEE National Aerospace and Electronics Conference*, pages 110–113. IEEE.

Fagiani, F. E. (2021). *UAV detection and localization system using an interconnected array of acoustic sensors and machine learning algorithms*. Master thesis, Purdue University. https://doi.org/10.25394/PGS.14502759.v1.

Intelligence, I. (2021). Drone market outlook in 2021: industry growth trends, market stats and forecast, “https://www.businessinsider.com/drone-industry-analysis-market-trends-growth-forecasts”, (accessed August 2021).

Jeon, S., Shin, J.-W., Lee, Y.-J., Kim, W.-H., Kwon, Y., and Yang, H.-Y. (2017). Empirical study of drone sound detection in real-life environment with deep neural networks. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1858–1862. IEEE.

Kim, J., Park, C., Ahn, J., Ko, Y., Park, J., and Gallagher, J. C. (2017). Real-time uav sound detection and analysis system. In *2017 IEEE Sensors Applications Symposium (SAS)*, pages 1–5. IEEE.

Koettl, C. and Marcolin, B. (August 10th, 2018). “A closer look at the drone attack on maduro in venezuela, “https://www.nytimes.com/2018/08/10/world /americas/venezuela-video-analysis.html”, (accessed August 2021).

Lee, T. B. (September 18th, 2013). Watch the pirate party fly a drone in front of germany’s chance, “https://www.washingtonpost.com/news/the-switch/ wp/2013/09/18/watch-the-pirate-party-fly-a-drone-in-front-of-germanys-chancellor/”, (accessed August 2021).

Li, S., Kim, H., Lee, S., Gallagher, J. C., Kim, D., Park, S., and Matson, E. T. (2018). Convolutional neural networks for analyzing unmanned aerial vehicles sound. In *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, pages 862–866. IEEE.

librosa Development Team (2021). Feature extraction, “https://librosa.org/doc/main/feature.html”, (accessed August 2021).

Mezei, J., Fiaska, V., and Molnár, A. (2015). Drone sound detection. In *2015 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 333–338. IEEE.

Pallotta, L., Clemente, C., Raddi, A., and Giunta, G. (2020). A feature-based approach for loaded/unloaded drones classification exploiting micro-doppler signatures. In *2020 IEEE Radar Conference (RadarConf20)*, pages 1–6. IEEE.

Ritchie, M., Fioranelli, F., Borrion, H., and Griffiths, H. (2017). Multistatic micro-doppler radar feature extraction for classification of unloaded/loaded micro-drones. *IET Radar, Sonar & Navigation*, 11(1):116–124.

Schmidt, M. S. and Shear, M. D. (January 26th, 2015). A drone, too small for radar to detect, rattles the white house, “https://www.nytimes.com/2015/01/27/us/white-house-drone.html”, (accessed August 2021).

Seo, Y., Jang, B., and Im, S. (2018). Drone detection using convolutional neural networks with acoustic stft features. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE.

Wang, Y., Fagiani, F. E., Ho, K. E., and Matson, E. T. (2021). A feature engineering focused system for acoustic uav detection. In *ccecpted for Publication International Conference on Robotic Computing (IRC2021)*. IEEE.