UAV Velocity Prediction Using Audio data

Eunyoung Bang*, Yeongmin Seo†, Jeongyoun Seo‡,
Raymond Zeng§, Aminata Niang¶, Yaqin Wang§, and Eric T Matson§
*Kangwon National University, Chuncheon, South Korea
yeong35@kangwon.ac.kr
†DaeguCatholic University, Daegu, South Korea
dudals1003@cu.ac.kr
‡Sangmyung University, Seoul, South Korea
201810773@sangmyung.kr
§Purdue University, West Lafayette, IN
{zeng172, wang4070, ematson}@purdue.edu
¶Institute Polytechnique de Paris, Paris, France
aminata.niang@telecom-sudparis.eu

***Abstract*—The Federal Aviation Administration (FAA) set the Unmanned Aerial Vehicles' (UAV) speed limit to 100 mph. This research focused on detecting when the UAV exceeds a speed limit for experiment, and using the sound dataset to predict the velocity of a UAV. It is hard to detect a malicious UAV, but we can assume that a UAV over 100 mph is most likely malicious. An indoor environment will be used as a controlled environment and the dataset is divided into two classes: slow (0-9mph) and fast (over 10mph). Support Vector Machine (SVM), Random Forest, and Light Gradient Boosting algorithm (LGBM) were the Machine Learning model used for this research, and Convolutional Neural Network (CNN) was the Deep Learning model used for this research. The result shows that the CNN model has the highest performance (F-1 score: 1.0, Accuracy: 1.0, Recall: 1.0, Precision: 1.0) for classifying the sound of the UAV speed.**

***Index Terms*—UAV, Velocity Prediction, Audio data, Machine Learning, Deep Learning**

I. Introduction

The use of Unmanned Aerial Vehicle (UAV) has increased around the world in many industries such as the police [1], medical, and agricultural fields. EMERGEN Research shows that the UAV industry in the US reached 23.6 billion (in USD) in 2021, and CONSORTIQ and Goldman Sachs forecast that the UAV industry in the US will grow from 90 to 100 billion in 2030 [2], [3]. India's medical industries utilize these UAVs in certain remote regions with the intent of delivering many kinds of packages such as blood, medical supplies, or food [4]. Yamaha RMAX, the first UAV approved by the Federal Aviation Administration (FAA), carried more than 55 pounds of fertilizers and pesticides to spray crops [5]. Even though these UAVs have good intentions, they can be abused by bad actors for malicious purposes. For instance, a UAV crashed into one of the electrical grids in Pennsylvania in July 2020 as stated by the FBI (Federal Bureau of Investigation) which is an example of a UAV kamikaze attack [6]. And, two UAVs crashed into the Novoshakhtinsk oil refinery, in Rostov, setting off a fiery explosion at an inside of Russia's borders in July 2022 as stated by the governor of Russia's Rostov region, which is an example of drone strike attack [7]. To minimize the damage from above instances, the FAA limits UAV velocity to under 100 mph as a law [8]. If some UAVs fly over 100 mph, it will be considered an illegal or malicious UAV.

Many pieces of research have been published to detect malicious UAVs using various datasets. For example, Yang et al. have experimented to detect UAVs with sound signals [9]. Another example is Knoedler et al. experimented with detecting and tracking a small UAV using passive Radar [10]. However, there is little research to predict the UAV velocity. The test was conducted by utilizing microphones and recording these UAVs at different speeds. The reason is that there is little research about using a microphone to detect a UAV compared to other research using RADAR, LiDAR, and cameras. This paper aims to detect if a UAV exceeds velocity boundary using UAV driving sound. The MFCC (Mel Frequency Cepstral Coefficient) has been used to feature extraction from the sound dataset. Machine learning and Deep learning models are compared to classify accuracy. To conduct the experiments, This research used Support Vector Machine (SVM), Random Forest, and Light Gradient Boosting Machine (LGBM) as a Machine Learning model for this research, and Convolutional Neural Network (CNN) was Deep Learning model used for this research as well.

The remaining portions of the paper are organized as follows: Section 2 presents reviews of Machine Learning, Deep Learning research using audio data to detect the UAV, and Velocity prediction research. Section 3 explains the methodology of the feature extract method, and four different algorithms including three Machine Learning algorithms, and one Deep Learning algorithm. Section 4 describes the experiment and shows the result. The last section summarizes the research and provides future works.

II. Literature Review

A. UAV Research Using Audio Data

UAV detection research has become active due to the potential malicious threats of these UAVs. To solve these threats of UAVs, research has also been conducted to detect UAVs using computer vision, RADAR, and audio data [11], [12]. Research using audio data shows promising results.

Wang et al. used multiple feature extraction methods (mfcc, mel, contrast, chroma, and tonnetz) to try and find the best feature extraction method to use when extracting UAV characteristics from audio data. SVM, Gaussian Naive Bayes (GNN), K-Nearest Neighbor (KNN), and Neural Network (NN) were used to find the best feature extraction. 300 audio data was collected from a DJI Phantom 4 and an Evo 2 Pro, and 600 more dataset will be collected from an outdoor environment. The mfcc shows over 99% accuracy compared to other models (SVM: 99.6%, GNB: 99.5%, KNN: 99.0%, NN: 99.7%). A combination of machine learning and feature extraction shows an accuracy over 100% [9].

Al-Emadi et al. focused on the malicious activities of drones and conducted a study to identify UAVs using audio data. According to the paper, the use of SVM is effective for drone detection, but requires optimization of hand-created functions. To remediate this, Deep Learning was used as a way to have feature extraction and optimization. Experiments were conducted using three Deep Learning models: CNN, Recurrent Neural Network (RNN), and Convolutional Recurrent Neural Network (CRNN). RNN shows the shortest time required (389.02 seconds) that is required to process this data, but it also showed the lowest accuracy (57.16%) among the three models. CRNN has both RNN and CNN features. It has a time of 605.67 seconds and an accuracy above 90%. CNN has the longest time to train (807.10 second), however it shows an accuracy above 90%. CNN has better accuracy than CRNN by 0.72 percent-point [11], [13], [14].

B. Vehicle Velocity Prediction Using Audio Data

There have been some trial researchers to predict the speed of the objects by using audio data not only to detect.

One research has been conducted to predict the car's speed and the gear's position using the audio data of the engine using Gradient Boosting. The author confirmed the relationship between audio and the car's speed. The author uses three microphones to get the dataset in controlled condition. Two microphones attached on the inside and outside the windshiel to induce wind noise. The other microphone is placed next t the engine. After normalizing the dataset, the dataset wer under feature extraction using MFCC, zero crossing rate and spectral centroid. The audio contains features of spee and gear state, and this audio went under feature extractio where the result will be used as input to Gradient Boosting Gradient Boosting accuracy increases when the speed interva is large and time frame is minimal. When only using Gradier Boosting accuracy is under 75%. The author uses correlatio matrix to optimize Gradient Boosting. As a result, Gradier Boosting shows over 90% accuracy in gear position and spee prediction [15].

Kubera et al. used Machine Learning to find out the changes in the velocity of passing vehicles. The dataset was created by installing a microphone about five feet from the road and obtaining audio data from vehicles passing by the road. Acquired audio data were parameterized through different methods based on audio features and spectrogram data before putting the audio data through Machine Learning. There are five types of Machine Learning used: Random Forest, SVM with linear kernel (SVML), SVM with quadratic kernel (SVMQ), SVM with Radial Basis Function kernel (SVMR), and Multi-Layer Perceptron (MLP). The result shows Random Forest, SVML, SVMQ, SVMR, and MLP accuracy results as 90.5%, 85.4%, 87.1%, 90.9%, and 88.6% respectively. Models that have the highest accuracy were put into a classifier ensemble that consists of the best performing classifiers. The ensemble resulted in an accuracy of 94.7% [16].

III. METHODOLOGY

The FAA limits UAV velocity to under 100 mph. Therefore, if the UAV flies over 100 mph, it will be an illegal act. In our experiment, the speed boundary has been set as 10 mph, and if the UAV exceeds the speed limit, we assume that is an illegal act (or a threatening act). This paper classifies two types of speed: slow and fast. UAVs flying under 10 mph will be considered slow, and over 10 mph will be considered fast. The dataset was collected in an indoor environment to control the noise. In this paper, one type of UAV model, X8SW, was used to collect the dataset. Fig. 1 shows the environment of the dataset collection using a microphone. X8SW is used for the experiment. There are various ways when a malicious actor tries to attack someone or someplace. Therefore, we collect the dataset in two-way; flying back and forth, and flying in a circle over the microphone. The audio files were split into 3-second snippets. The slow velocity sound was labeled 0, the fast velocity was labeled 1. Table 1 shows the size of the dataset. After collecting the dataset, MFCC is used to feature extraction. MFCC shows promising result variable research [9], [17]. The dataset is used to train Machine Learning and Deep Learning models. The performance is measured by F-1 Score, Accuracy, Recall, and Precision.

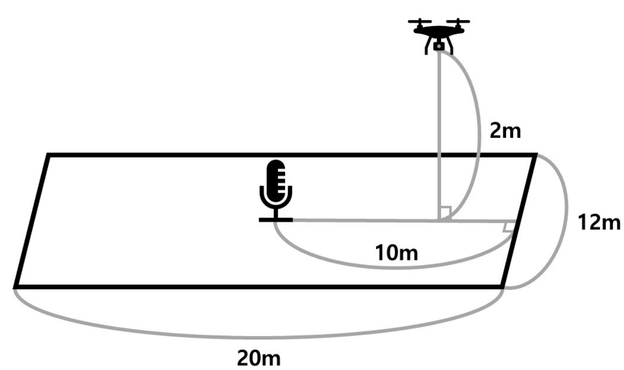


Fig. 1: Dataset collection environments.

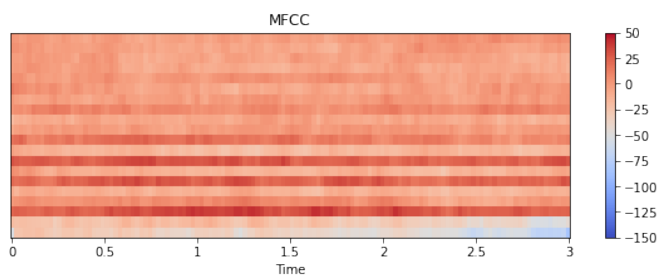
TABLE I: The Size of Dataset.

Velocity	Number of Samples	Total Time (Sec)
Slow	1059	3177
Fast	1296	3888
Total	2355	7065

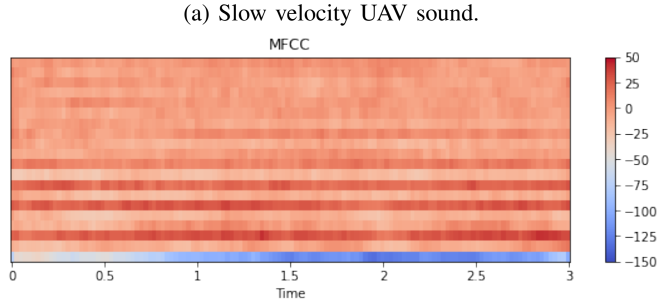
A. Feature Extraction

MFCC is a feature extraction that uses Fourier transform to extract features from audio signals. Wang et al. used other feature extensions in their studies as well as MFCC and compared the performance of each feature extraction. As a result, MFCC showed the highest average accuracy of 99.45% and it is 12-26 percent-point higher than other features. Therefore, MFCC is the most suitable feature for accurately measuring the velocity of the UAV using acoustic data [9].

To predict UAV velocity, MFCC is used to extract features from audio data. Fig 2 shows the MFCC images in our dataset. The bottom of the image shows a different appearance.



(a) Slow velocity UAV sound.



(b) Fast velocity UAV sound.

Fig. 2: MFCCs feature plot

B. Models

Three types of Machine Learning models are used: SVM, Random Forest, and LGBM. CNN is used as a Deep Learning model to predict UAV velocity. SVM is a Machine Learning algorithm that defines a decision boundary for pattern recognition and data analysis. SVM is usually used in classification and shows high accuracy in [9], [18].

Random Forest is an ensemble model based on decision trees. If there are many features within the decision tree that often occur overfitting. However, Random Forest reduces the possibility of overfitting because it divides some features into each decision tree. Some research shows that Random Forest with MFCC shows good results constantly [19], [20].

TABLE II: The CNN model's architecture

Layer	Layer Description
Layer 1	Conv1D (20 filters with a size of 4, 2 Padding, 2 Stride) ReLU activation function
Layer 2	Conv1D (11 filters with a size of 4, 2 Padding, 2 Stride) ReLU activation function
Layer 3	Conv1D (5 filters with a size of 4, 2 Padding, 2 Stride) Sigmoid activation function

LGBM is a type of gradient boosting machine (GBM). GBM is a Machine Learning algorithm that decreases errors by repeatedly training the difference between correct answers and errors using the gradient descent method. As a kind of boosting algorithm, several models with too high a bias or variance value. LGBM does not generate trees in the level-wise method when generating trees in the learning method. So LGBM uses less memory to make trees in the Leaf-wise method. Also, LGBM shows good results in lung sound classification using MFCC [21], [22].

CNN is a Deep Learning algorithm, artificial neurons or nodes based on a human neural network. CNN shows good performance in image processing. A lot of research about UAV shows promising results using CNN [11], [13], [14].

IV. Experiments and Results

A. Experiment Design

The fast and slow velocity files consist of 2355 sound samples. Every file in the dataset is a 3-second snippet. A MFCC method is used to extract the features from the dataset.

Three kinds of Machine Learning models (SVM, Random Forest, and LGBM) and one Deep Learning model (CNN) are used to detect UAV velocity. Table 3 shows the details of CNN architecture used in the experiment. CNN architecture consists of 3 layers, a 1D convolutional layer and ReLU activation function. The last layer's activation function uses Sigmoid. The hyperparameter that we use has a batch size of 16, learning rate of 0.0001, and epoch of 100. Dataset is split into 80% for training, 10% for validation, and 10% for testing. Each model is estimated using the F-1 score, Accuracy, Recall, and Precision on binary classification.

B. Results

Table 3 shows the difference between each dataset. The slow and fast velocity sound shows a difference between 6 of 20 features. The low-frequency range shows a difference between slow and fast velocity sound over 5.

Table 4 shows the result of the experiment with Machine Learning and Deep Learning models. All of the models show an accuracy of over 0.95. The reason is table 3 shows a difference between the two datasets. It helps models to learn the difference. The best model in Machine Learning is Random Forest. Random Forest shows higher accuracy in all metrics than other Machine Learning models. CNN showed the best

TABLE III: The table shows the diffrence between the MFCC feature of the sound dataset.

Velocity	Feature 0	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	Feature 8	Feature 9
Slow	**-48.16**	**-25.84**	**19.66**	0.91	**-17.52**	21.85	-10.59	**20.11**	-15.26	9.50
Fast	**-55.64**	**-11.42**	**25.14**	-3.21	**-23.56**	22.03	-14.79	**14.18**	-18.04	7.65
Velocity	Feature 10	Feature 11	Feature 12	Feature 13	Feature 14	Feature 15	Feature 16	Feature 17	Feature 18	Feature 19
Slow	-0.63	-9.87	**7.72**	-1.24	-1.62	0.52	-4.19	-6.14	-3.47	1.66
Fast	-3.59	-6.54	**1.83**	-1.03	-5.80	-0.25	-5.22	-2.39	-3.67	0.12

TABLE IV: The result of predicting UAV velocity.

Model	F-1 Score	Accuracy	Precision	Recall
SVM	0.988	0.987	0.977	1.000
Random Forest	0.998	0.997	0.996	1.000
LGBM	0.996	0.995	0.992	1.000
CNN	1.000	1.000	1.000	1.000

performance of all models. CNN shows a perfect result of 1.000 to all of the metrics.

V. Conclusion

This paper proposed to detect UAV velocity using audio data generated by driving a UAV. This project used three kinds of Machine Learning models(SVM, Random Forest, and LGBM) and a Deep Learning model, CNN, to detect UAV velocity. In this paper, the dataset was collected in an indoor environment. If the UAV flies over 10 mph, it is considered a fast one. If the UAV flies under 10 mph, it is considered a slow one. The dataset consists of 3-second snippets, and there are 2355 audio files in the dataset. Each data extracted the feature using MFCC, and the features of slow and fast audio data show the differences between 6 of 20 features. The differences in features are helpful for the model to learn the difference between fast and slow. The dataset was used to train the Machine Learning and Deep Learning models. All of the models show over 0.95 f-1 score (SVM: 0.988, Random Forest: 0.998, LGBM: 0.996, CNN: 1.000).

The limitation of our research included that we only used one kind of UAV for data collection in indoor conditions. In this paper, the model only could classify binary classes, fast and slow. However, Table 3 shows the difference between the slow and fast dataset. So the model will classify the velocity of the UAV at 5 mph intervals in future experiments. Also, the amount of the dataset was sufficient for the experiment, at the next experiment, planning to collect more data for general use.

ACKNOWLEDGEMENT

This research was supported by the MSIT(Ministry of Science ICT), Korea, under the National Program for Excellence in SW (2015-0-00910) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

References

[1] N. Chen, “Drones are responding to 911 calls in this California city.” CNN.com. https://www.cnn.com/2019/04/12/us/california-drones-emergency-response/index.html (accessed May. 17, 2022)

[2] EMERGEN Research, “unmanned aerial vehicle market.” EMERGENRESEARCH.com. https://www.emergenresearch.com/industry-report/unmanned-aerial-vehicle-market (accessed May. 20, 2022).

[3] D. Daly, “Drone Industry Outlook For the United States: 2020-2030.” CONSORTIQ.com. https://consortiq.com/uas-resources/drone-industry-outlook-us-2020-2030 (accessed May. 20, 2022).

[4] A. Clarance, “What a drone picking up blood samples tells about healthcare in India.” BBC.com, https://www.bbc.com/news/world-asia-india-61267750 (accessed May. 15, 2022).

[5] A. Meola, “Precision agriculture in 2021: The future of farming is using drones and sensors for efficient mapping and spraying.” INSIDER.com. https://www.businessinsider.com/agricultural-drones-precision-mapping-spraying (accessed May. 16, 2022).

[6] B. Crumley, “FBI says PA electricity station likely ‘target’ of drone incident.” DroneDJ.com. https://dronedj.com/2021/11/04/fbi-says-pa-electricity-station-likely-target-of-drone-incident/ (accessed May. 15, 2022)

[7] L. Harding, “Drone crashes into Russian oil refinery in possible attack” The Guardian https://www.theguardian.com/world/2022/jun/22/russian-novoshakhtinsk-oil-refinery-struck-drone-possible-attack-inside-borders (accessed July. 24, 2022)

[8] Federal Aviation Administration. *Small Unmanned Aircraft Systems (UAS) Regulations (Part 107)*. [Online]. Available: FAA/newsroom/small-unmanned-aircraft-systems-uas-regulations-part-107

[9] Y. Wang, F. E. Fagian, K. E. Ho, and E. T. Matson, “A Feature Engineering Focused System for Acoustic UAV Detection,” *2021 5th IEEE Int. Conf. on Robot. Comput. (IRC)*, 2021, pp. 125-130, doi: 10.1109/IRC52146.2021.00031.

[10] B. Knoedler, C. Steffes, and W. Koch, “Detecting and Tracking a Small UAV in GSM Passive Radar Using Track-before-Detect,” *2020 IEEE Radar Conf. (RadarConf20)*, 2020, pp. 1-6, doi: 10.1109/RadarConf2043947.2020.9266673.

[11] S. Al-Emadi, A. Al-Ali, A. Mohammad, and A. Al-Ali, “Audio Based Drone Detection and Identification using Deep Learning,” *2019 15th Int. Wireless Commun. & Mobile Comput Conf. (IWCMC)*, 2019, pp. 459-464, doi: 10.1109/IWCMC.2019.8766732.

[12] S.Jamil, Fawad, M.Rahman, A.Ullah, S.Badnava, M.Forsat, and S.S.Mirjavadi, “Malicious UAV Detection Using Integrated Audio and Visual Features for Public Safety Applications”, *Sensors (Basel) vol. 20,14 3923. , 2020*, doi: 10.3390/s20143923.

[13] S. Seo, S. Yeo, H. Han, Y. Ko, K. E. Ho, and E. T. Matson, “Single Node Detection on Direction of Approach,” *2020 IEEE Int. Instrum. and Meas. Technol. Conf. (I2MTC)*, 2020, pp. 1-6, doi: 10.1109/I2MTC43012.2020.9129016.

[14] S. Li, H. Kim, S. Lee, J. C. Gallagher, D. Kim, S. Park, and E. T. Matson, “Convolutional Neural Networks for Analyzing Unmanned Aerial Vehicles Sound,” *2018 18th Int. Conf. on Control, Automat. and Syst. (ICCAS)*, 2018, pp. 862-866.

[15] H. V. Koops, and F. Franchetti, “An ensemble technique for estimating vehicle speed and gear position from acoustic data,” *2015 IEEE Int. Conf. on Digit. Signal Process. (DSP)*, 2015, pp. 422-426, doi: 10.1109/ICDSP.2015.7251906.

[16] E. Kubera, A. Wieczorkowska, A Kuranc, and T. Słowik “Discovering Speed Changes of Vehicles from Audio Data.” *Sensors (Basel, Switzerland) vol. 19,14 3067.*, 2019, doi:10.3390/s19143067

[17] Y. Wang, F. Fagiani, K. Ho, and E. T. Matson, “A Feature Engineering Focused System for Acoustic UAV Payload Detection,” in Proc. 14th Int. Conf. on Agents and Artificial Intelligence (ICAART), pp. 470-475, doi: 10.5220/0010843800003116

[18] K. Racharla, V. Kumar, C. B. Jayant, A. Khairkar and P. Harish, ”Predominant Musical Instrument Classification based on Spectral Features,”

2020 7th Int. Conf. on Signal Process. and Integr. Netw. (SPIN), 2020, pp. 617-622, doi: 10.1109/SPIN48934.2020.9071125.

[19] O. K. Toffa and M. Mignotte, "Environmental Sound Classification Using Local Binary Pattern and Audio Features Collaboration," in IEEE Trans. on Multimedia, 2021, vol. 23, pp. 3978-3985.

[20] U. Garg, S. Agarwal, S. Gupta, R. Dutt and, D. Singh, "Prediction of Emotions from the Audio Speech Signals using MFCC, MEL and Chroma," in Proc. 2020 12th Int. Conf. on Comput. Intell. and Communication Netw. (CICN), 2020, pp. 87-91, doi: 10.1109/CICN49253.2020.9242635.

[21] R. K. Tripathy, S. Dash, A. Rath, G. Panda, and R. B. Pachori, "Automated Detection of Pulmonary Diseases From Lung Sound Signals Using Fixed-Boundary-Based Empirical Wavelet Transform," IEEE Sensors Letters, vol. 6, no. 5, pp. 1-4, May 2022, Art no. 7001504

[22] M. M. Nishat, T. Hasan, S. M. Nasrullah, F. Faisal, M. A. -A. -R. Asif, and M. A. Hoque, "Detection of Parkinson's Disease by Employing Boosting Algorithms," in Proc. 2021 Joint 10th Int. Conf. on Inform., Electronics & Vision (ICIEV) and 2021 5th Int. Conf. on Imaging, Vision & Pattern Recognition (icIVPR), 2021, pp. 1-7, doi: 10.1109/ICIEVicIVPR52578.2021.9564108.