A comparison of machine learning and human performance in the real-time acoustic detection of drones

Vishwa Alaparthy
Duke University
304 Research Dr
Durham, NC 27708
vishwa.alaparthy@duke.edu

Sayan Mandal
Duke University
304 Research Dr
Durham, NC 27708
sayan.mandal@duke.edu

Mary Cummings
Duke University
304 Research Dr
Durham, NC 27708
m.cummings@duke.edu

***Abstract*— Usage of drones has increased substantially in both recreation and commercial applications and is projected to proliferate in the near future. As this demand rises, the threat they pose to both privacy and safety also increases. Delivering contraband and unauthorized surveillance are new risks that accompany the growth in this technology. Prisons and other commercial settings where venue managers are concerned about public safety need cost-effective detection solutions in light of their increasingly strained budgets. Hence, there arises a need to design a drone detection system that is low cost, easy to maintain, and without the need for expensive real-time human monitoring and supervision. To this end, this paper presents a low-cost drone detection system, which employs a Convolutional Neural Network (CNN) algorithm, making use of acoustic features. The Mel Frequency Cepstral Co-efficients (MFCC) derived from audio signatures are fed as features to the CNN, which then predicts the presence of a drone. We compare field test results with an earlier Support Vector Machine (SVM) detection algorithm. Using the CNN yielded a decrease in the false positives and an increase in the correct detection rate. Previous tests showed that the SVM was particularly susceptible to false alarms for lawn equipment and helicopters, which were significantly improved when using the CNN. Also, in order to determine how well such a system compared to human performance and also explore including the end-user in the detection loop, a human performance experiment was conducted. With a sample of 35 participants, the human classification accuracy was 92.47%. These preliminary results clearly indicate that humans are very good at identifying drone's acoustic signatures from other sounds and can augment the CNN's performance.**

TABLE OF CONTENTS

1. INTRODUCTION

With recent technological and societal advancements, there has been a substantial increase in number of drones operated in public spaces [1]. Outdoor spaces such as prisons and recreational venues are more susceptible to a malicious activity through an interloping drone [2]. Affordability of drones coupled with the dearth of efficient counter measures, make it difficult for site administrators to effectively counter the threat drones pose. Hence, there arises a need to design a drone detection system that is not only reliable, but also is cost effective. Keeping these constraints in mind, this paper proposes a low-cost drone detection system, which is embedded on a raspberry pi and uses acoustic footprints to detect the drone's presence. We use a deep learning algorithm that leverages Mel Frequency Ceptral Coefficients (MFCC) as its features.

Previous research [3] on this problem lists RADAR [4], acoustics [5], optics [6] and radio frequencies [7] as the four prominent techniques commonly used in a designing a drone detection system. However, a drone detection system built using RADAR is expensive and ineffective in detecting small drones [8]. Visibility is a hindrance while adopting optical sensing techniques [9]. Most of the previous works [10] were confined to training and testing a drone detector in a laboratory setup and not tested in an outdoor setting. In order to build a cheap and reliable system, we made use of the drone's acoustics, realizing that detection range is traded off for cost effectiveness.

Previously, we adopted a Support Vector Machine (SVM)-based approach for this problem [11]. The current work will underline the merits of using a deep learning model by comparing it to the SVM-based model's outcomes. To determine how humans performed in relation to the algorithm, 35 participants and the algorithm classified sounds containing drone and similar-sounding files.

The remainder of this paper is organized as follows:

978-1-7281-7436-5/21/$31.00 ©2021 IEEE

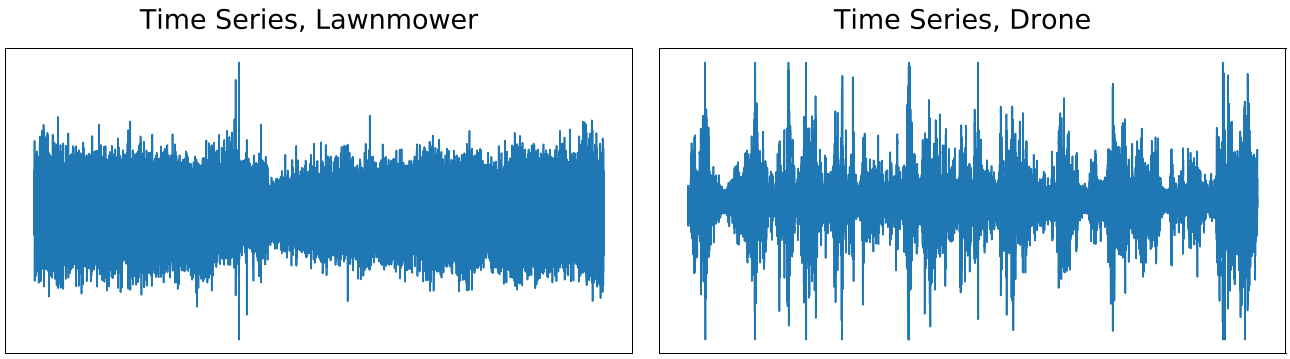

Figure 1. Time series representation of a lawnmower and a drone

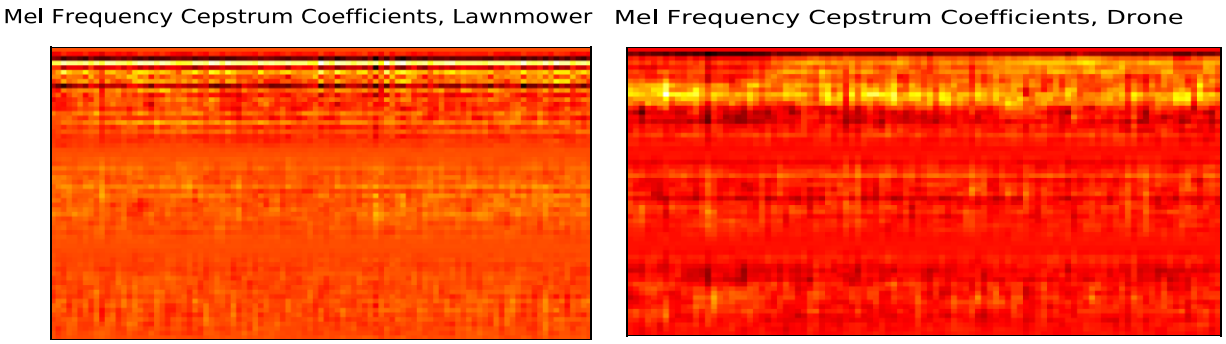

Figure 2. MFCC spectrogram of a lawnmower and a drone

Section 2 gives an overview of the algorithm design. Section 3 briefly describes the test setup. Section 4 briefly describes the details of a survey carried out to compare human and machine performance in classifying sounds and Section 5, the results and analysis. Section 6 provides a conclusion.

2. Building a Deep Learning Classifier:

Generating a Dataset

Quality and size of the dataset plays a significant role to help develop a well-trained machine learning model [12], To this end, during the first phase of data collection, we used a DJI Inspire 2 and a 3DR Iris+ to collect drone sounds at different altitudes and distances from an off-the-shelf microphone. Additional data was collected using a DJI Phantom 4 to collect audio files at different radial distances.

The dataset was further expanded by recording drones at a drone racing competition [13]. Drone audio files were also gathered at an outdoor amphitheater to expand the dataset to include a relevant application environment. Given this myriad of data sources, which are available at "https://sites.duke.edu/prisdatabase/", the SVM algorithm was trained on 73% of the data, while the CNN algorithm used 100%.

The rest of the dataset included the ESC-50 dataset [14], which is used for environmental sound classification. The negative dataset also included white noise, periods of silence and other audio clips from noisy environments, with no drone flying. In addition, the system was trained with sounds of lawnmowers, leaf blowers, helicopters and aircraft to help reduce the number of false positives occurring from the previous approach.

Contrary to the previous SVM algorithm which had five categorical classifications, this CNN-based system only used a binary classification, 'drone detected' or 'no drone'. To generate this classification, a recorded audio clip of ten seconds is sliced into five equal parts and the algorithm is run on each of the five audio clips. A 'drone detected' alert is generated when at least three of the five audio clips generate a positive outcome of the drone's presence from the algorithm. The recorded clip is then superseded with a new ten second clip.

Feature Extraction

We chose to use MFCC's as the feature set, which are widely used in speech recognition [15]. MFCC's provide an image of how the acoustic signal looks with respect to time. The acoustic signal in the time domain is converted to a

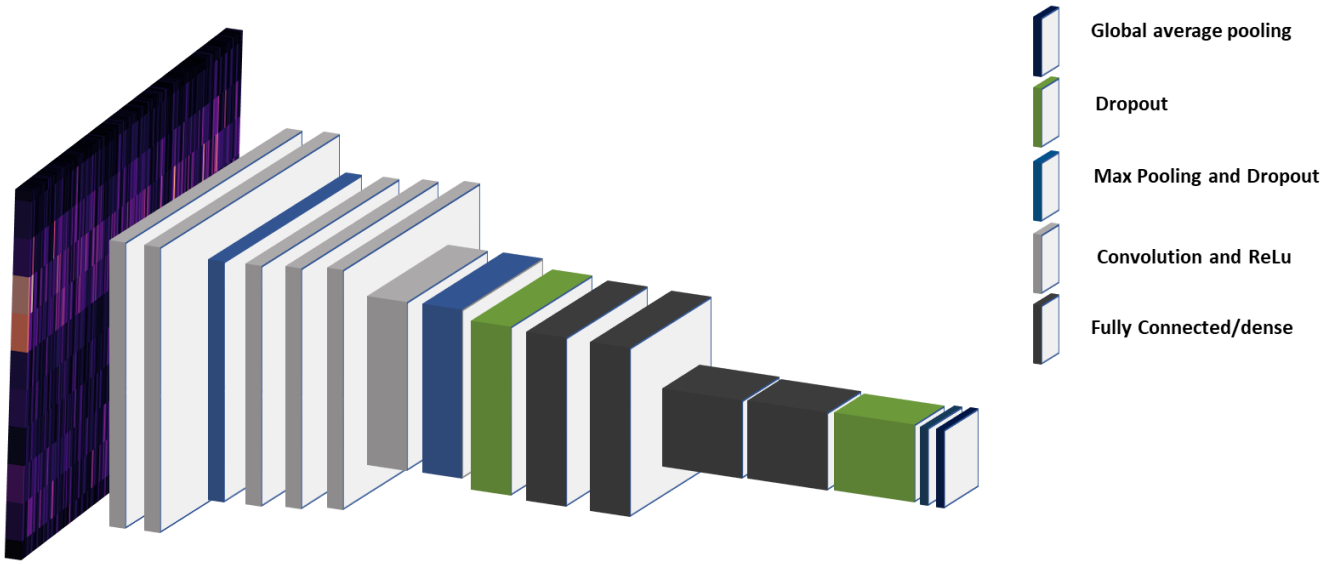

Figure 3. CNN Architecture

periodogram using Fast Fourier Transforms (FFT). The periodogram, which is a Power Spectral Density (PSD) estimate of the signal, is then converted to a spectrogram by stacking them linearly together. Using the mel scale [16], a mel frequency spectrogram is computed (1). MFCC's are derived by taking a Discrete Cosine Transform (DCT) on it log powers. While Figure 1 shows the time domain representation of a lawnmower and a drone, Figure 2 provide a depiction of the MFCC's for a lawnmower and a drone.

$$M(f) = 1125\ln(1 + \frac{f}{700}) \quad (1)$$

Other low priority features used in the previous approach such as tonal centroids and chromagram were discarded to lower the overhead of the deep learning algorithm.

Using a Convolutional Neural Network (CNN)

CNN is a popular deep learning model widely used in natural language processing [17] and image processing [18]. A CNN is made up of an input layer and an output layer, with several hidden layers between them. The number of hidden layers is usually determined based on the size of the dataset. A mathematical operation, namely convolution, is used between the layers. The CNN model for the acoustic drone detection application used a six-convolution layer approach, with each stacked next to one other. The number of filters were progressively increased from 16 till 512. The architecture is shown in Figure 3.

The activation function used is a Rectified Linear Unit (ReLU). Two max pooling layers were included in the design, as well as a dropout layer, which helps to prevent overfitting. In the end, the data is pooled down by using a series of fully connected dense layers. The final dense layer was activated by using a softmax function, categorical crossentropy was used as the loss function and Adam was used as the optimizer.

The stray noises from the audio clip were removed by using an envelope function, which outlined the extremes of a signal. This helped in cleaning the dataset and decreased the number of false positives. The dataset was split into training and validation sets in a 4:1 ratio and trained over 30 epochs with a batch size of 32. We obtained a validation accuracy of 96.1% for the CNN algorithm, compared to the 96.86% accuracy of the previously-developed SVM algorithm. It should be noted that the SVM algorithm was not trained on a comparable negative dataset.

3. FIELD TESTS

While laboratory tests are important intermediate steps for assessing CNN performance, field tests are critical in determining actual performance in realistic settings. To this, similar to the field tests for the SVM-enabled system [11], the system was tested against a drone at difference distances, altitudes and with possible sources of false alarms.

Hardware Setup

When any sound is detected in the environment by an Apex 220 microphone, 10s audio clips are sent to a raspberry pi with a 4gb RAM, which is embedded with the CNN algorithm. A mobile hotspot is used to relay the drone detection alerts from the raspberry pi to an android app [19]. An alarm is triggered at the user end via an android device and the algorithm provides a "drone detected" notification. The user also receives an audio clip of 10 seconds along with the alert.

Test Conditions

Local flight tests were conducted on a day where the temperature was 44 ◦F with a wind of 3mph and

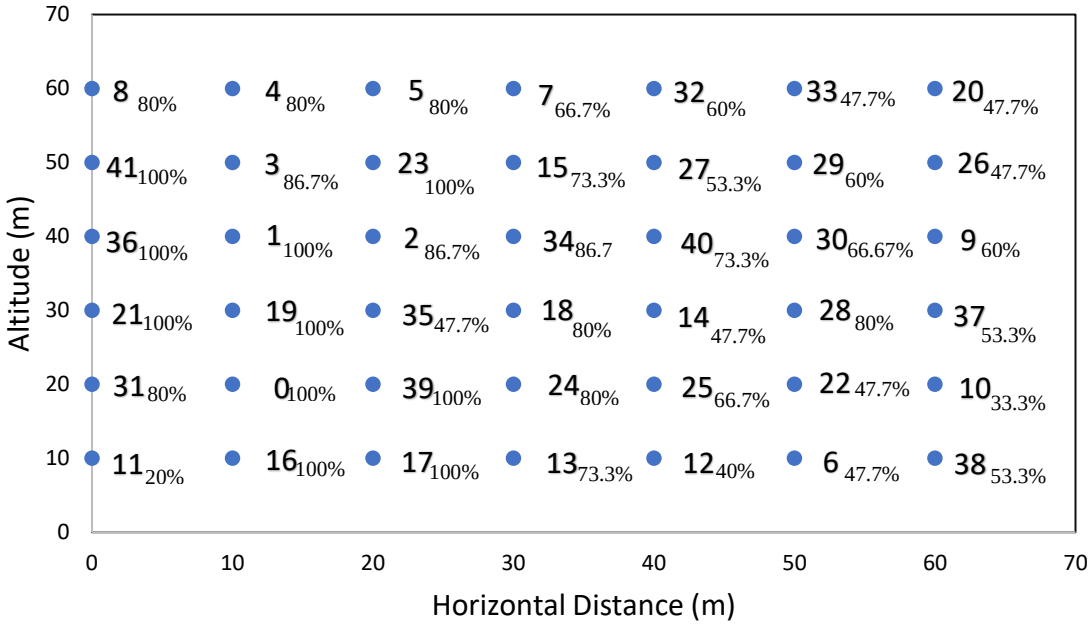

Figure 4. Randomized drone hovering points with confidence levels

visibility of 10 miles. Two commercial aircraft flew over during the test procedure. In addition, two racing drones and a remote-controlled aircraft were also operating in the vicinity. However, it is assumed that there is no real effect on the classification because the remote-controlled aircraft and the race drones were at least 100m away from the detector.

The objective of the tests was to understand how well the algorithm performed when the drone hovered at a certain altitude and at a certain horizontal distance from the detector. There were 42 hovering points spread 10m apart from each other across a grid of 60m*60m (Figure 4). The DJI phantom pro V 2.0 drone was moved from one data point to another in a random fashion without any specific sequential ordering to ensure that there is no effect of the test sequence on the potential results. The drone held its position for 30 seconds at each of the 42 points. The algorithm was considered successful when a notification of 'Drone detected' appeared within this time period. At each of the hovering points, sounds were recorded and stored as a wav file to feed the CNN.

Additionally, noise resilience tests at a local field were conducted using an electric leaf blower to observe how well the CNN performed when exposed to sounds similar to a drone's harmonics. Also, the device was placed at the local amphitheater for seventeen days, running uninterrupted to see what might trigger false alarms.

4. Human Classification Test

In order to benchmark the effectiveness of the CNN algorithm as compared to humans, we designed a survey to identify how well humans could distinguish the sound of a drone from similar sounds. In an IRB-approved online survey, 35 participants were provided training with four sample audio files. Two contained the sound of a drone and the other two contained sounds that are similar to a drone's sound like a leaf blower and a helicopter. They were instructed to listen to the sound files in a quiet environment under normal volumes.

After the training, participants were sent a randomized order of 30 precompiled audio files containing various sounds. Half of these were a variety of quadrotor drone sounds, and the other half were sounds similar to a drone in terms of frequency and pattern, such as helicopters, lawn mowers, leaf blowers, bees or wasps, wind and planes or fixed wing UAVs. Participants listened to the sound clips and chose if each one was a drone or not a drone. If they choose 'not drone', then they were asked to label the sound.

5. Results and Analysis

Field Test Results

The system's accuracy is 91.67% for data points in the 30m*30m region. The accuracy is 86.7% when the area of observation is expanded to 40*40. The accuracy comes down to 76.19%, when the coverage increases to 60*60. The ratio of the number of true positives to the total number of observation points is used to determine the confidence level at each data point. All the data points, with their respective confidence levels is given in Figure 4.

Figure 4 indicates that both the detection rate and the confidence level improve as the drone gets closer to the detector. However, an anomaly from this observation is seen at data point 11, located at a horizontal distance of 0m and altitude of 10m, where the confidence level is 20%. It is presumed that this occurs due to the similarity of the drone's

harmonics to that of a lawn mower sound that it was trained on.

Figure 5 compares the current algorithm's performance to the previous SVM algorithm. The current algorithm has a better detection rate when compared to the SVM-based algorithm for coverage areas of 30m*30m and 60m*60m. The SVM-based algorithm has a detection rate of 86.7% and 62% at 30m*30m and 60m*60m respectively and is outperformed by the CNN algorithm by at least 5.9%.

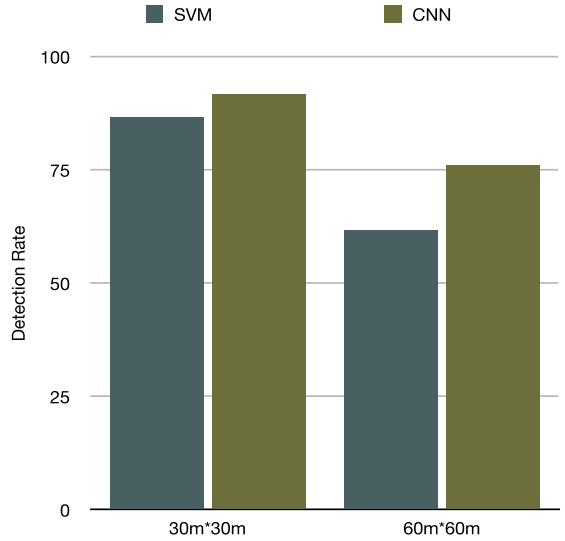

Figure 5. Comparison of CNN and SVM for UAV classification

As mentioned earlier, the previous SVM model experienced a high number of false positives from lawn equipment such as lawn mowers and leaf blowers. This was lowered by the current algorithm, which used such sounds in the training data. When exposed to an electric leaf blower with a radial distance of 100m and below, there was a 'drone present' notification 12.5% of the time. The CNN was trained on gas-powered leaf blower data, so this demonstrates a source of brittleness for the CNN. It was, however, very effective in distinguishing background noises, with no additional filters used for background noises. This test demonstrates that with a well-trained CNN and a quality dataset, the effectiveness of a drone detection systems increases to a considerable extent.

For the seventeen-day field test, the detector experienced a total of 1226 false positives over 46650 observations amounting to a false positive rate of 2.62%. A significant amount were caused by heavy winds and gusts from hurricane remnants. There was very limited activity on the detector when there was light rain (or no rain) or no heavy winds and gust. While the false positive rate was low when compared against all observations, 1226 false positives would not be tolerated over a 17-day period by users. Thus, more work is needed to dramatically lower the absolute numbers of false positives.

Human Classification Results

From the survey, the participants had an accuracy of 92.47%, which is superior to the 80% accuracy obtained from the CNN algorithm on the same dataset. Figure 6 shows the confusion matrix, which includes the number of correct classifications and the number of misclassifications. We also obtained a precision of 90.97%, recall of 93.79% and a F1 score of 92.36%. Table 1 summarizes and compares the outcomes of machine learning and human classification, which clearly indicates that the humans are better suited to classify the sounds in this dataset.

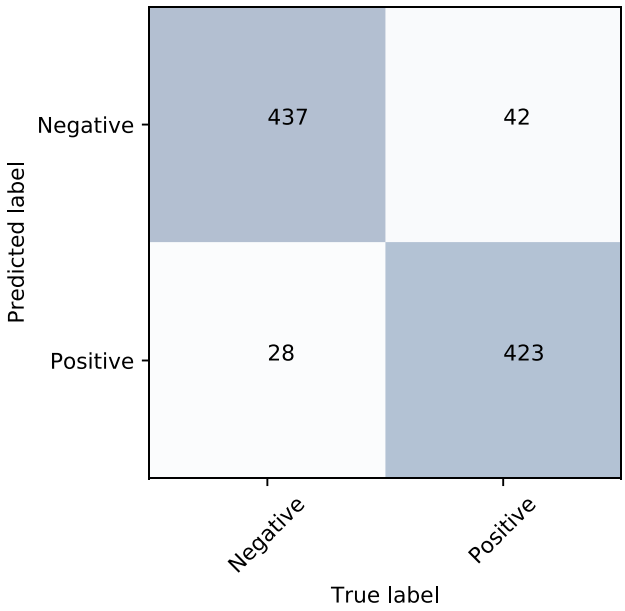

Figure 6. Confusion matrix for human classification

We also performed a non-metric multidimensional scaling (NMDS) on the human-classified data to determine the similarities across multiple dimensions. Figure 7 shows how similar the participants responses are with respect to each other. This shows there is not much deviance in the surveyed population and that most humans can successfully and consistently classify a drone sound. At the same time, different machine learning model can end up producing different results. It should also be noted that most of the human misclassifications were due to a fixed wing drone aircraft, which was classified as 'not a drone'.

Table1. Comparison of Machine learning vs Human performance

Metric	**CNN**	**Human Classification**
Accuracy	80%	92.47%
Precision	90.9%	90.97%
Recall	66.7%	93.79%
F1 score	76.9%	92.36%

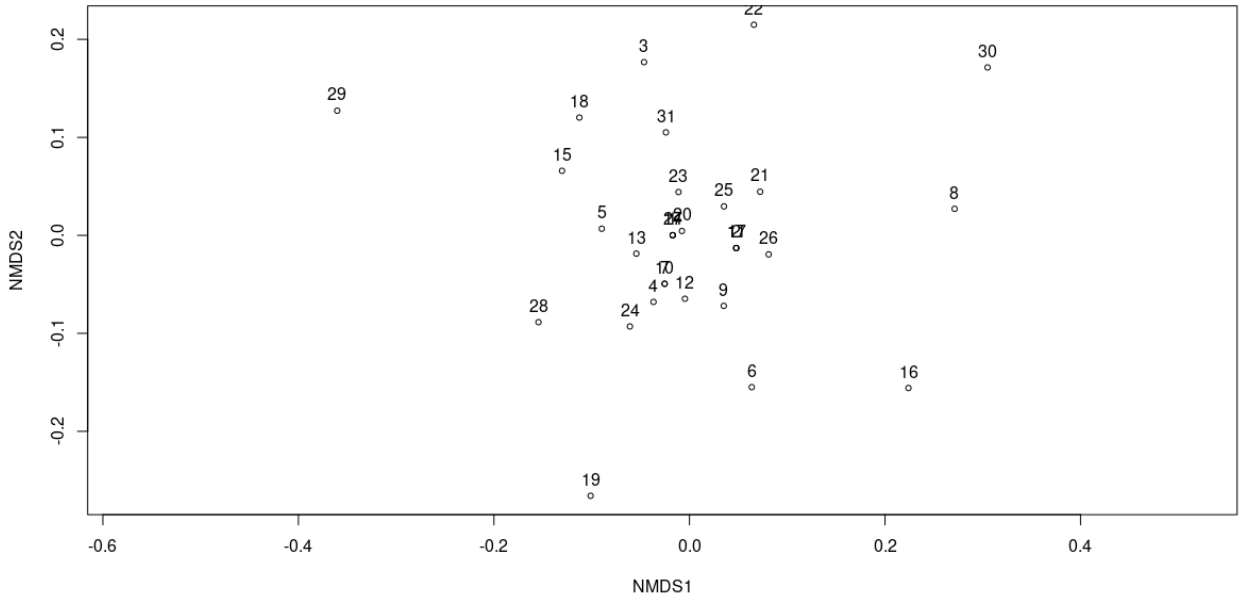

Figure 7. NMDS plot for similarity index

6. CONCLUSIONS

This paper compares the efficacy of a CNN and an SVM-based algorithm in detecting a drone's presence. Although the CNN-enabled system performed better than the previous SVM technique, it was trained on more data. It also experienced false alarms for data it had not previously seen. Because a high number of false alarms could lead to significant distrust by human users, it is important to reduce this number. For the next iteration, we are exploring a sensor fusion approach to address this problem.

Another possible intervention to reduce the number of false alarms due to never-before-seen data is to allow the human to essentially label data for CNN retraining. To this end, we are building a 'user-in-the-loop' classification mechanism, which enables the user to classify sounds in the environment. The human classification results demonstrate that humans can reliably identify sounds and we are exploring how a partnership can be developed so that humans can assist the CNN.

ACKNOWLEDGEMENTS

This work is sponsored by the National Science Foundation. Tigey Jewell-Alibhai flew the drones during flight tests, Kausthub Ramchandran Kunisi assisted in setting up the survey. We also appreciate the assistance of the Town of Cary and Anthony Campbell for helping with the field tests.

REFERENCES

[1] B. Rao, A. G. Gopi and M. Romana, "The societal impact of commercial drones," *Technology in Society*, vol. 45, pp. 83-90, 2016.

[2] M. Russon, "Drones are being used to smuggle drugs into Canadian prisons," 2013 URL https://www.businessinsider.com/drones-are-being-used-to-smuggle-drugs-into-canadian-prisons-2013-11, [Online],[Accessed: 2020-10-09].

[3] D. Doroftei and G. De Cubber, "Qualitative and quantitative validation of drone detection systems," in *International Symposium on Measurement and Control in Robotics*, Mons, 2018.

[4] Y. Liu, X. Wan, H. Tang, J. Yi, Y. Cheng and X. Zhang, "Digital television based passive bistatic radar system for drone detection," in *IEEE RADAR Conference*, 2017.

[5] A. Bernardini, F. Mangiatordi, E. Pallotti and L. Capodiferro, "Drone detection by acoustic signature identification," *Electronic Imaging*, pp. 60-64, 2017.

[6] F. Christnacher, S. Hengy, M. Laurenzis, A. Matwyschuk, P. Naz, S. Schertzer and G. Schmitt, "Optical and acoustical UAV detection," in *Electro-Optical Remote Sensing X}*, 2016.

[7] P. Nguyen, M. Ravindranatha, A. Nguyen, R. Han and T. Vu, "Investigating cost-effective rf-based detection of drones," in *2nd workshop on micro aerial vehicle networks, systems, and applications for civilian use*,

2016.

[8] A. Laučys, S. Rudys, M. Kinka, P. Ragulis, J. Aleksandravičius, D. Jablonskas, D. Bručas, E. Daugėla and L. Mačiulis, "Investigation of detection possibility of UAVS using low cost marine radar," *Aviation,* vol. 23, pp. 48-53, 2019.

[9] I. Guvenc, F. Koohifar, S. Singh, M. L. Sichitiu and D. Matolak, ""Detection, Tracking, and Interdiction for Amateur Drones," *IEEE Communications Magazine,* vol. 56, no. 4, pp. 75-81, 2018.

[10] S. Jeon, J.-W. Shin, Y.-J. Lee, W.-H. Kim, Y. Kwon and H.-Y. Yang, "Empirical study of drone sound detection in real-life environment with deep neural networks," in *25th European Signal Processing Conference*, 2017.

[11] S. Mandal, L. Chen, V. Alaparthy and M. L. Cummings, "Acoustic Detection of Drones through Real-time Audio Attribute Prediction," in *AIAA Scitech 2020 Forum*, 2020.

[12] H. Xiao, K. Rasul and R. vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747,* 2017.

[13] "multigp website :https://www.multigp.com/about-multigp/".

[14] K. J. Piczak, "ESC: Dataset for environmental sound classification," in *23rd ACM international conference on Multimedia*, 2015.

[15] F. Zheng, G. Zhang and Z. Song, "Comparison of different implementations of MFCC," *Journal of Computer science and Technology,* vol. 16, no. 6, pp. 582-589, 2001.

[16] M. R. Hasan, M. Jamil and M. Rahman, "Speaker identification using mel frequency cepstral coefficients," *Variations,* vol. 1, no. 4, 2004.

[17] T. Young, D. P. Hazarika, S. a and E. Cambria, "Recent trends in deep learning based natural language processing," *ieee Computational intelligenCe magazine,* pp. 55-75, 2018.

[18] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[19] C. Wang and M. Cummings, "A Mobile Alerting Interface for Drone and Human Contraband Drops," in *AIAA Aviation 2019 Forum*, 2019.

BIOGRAPHY

Vishwa Alaparthy *is a Postdoctoral Associate at Duke University working in the Humans and Autonomy Lab (HAL). He received a Ph.D. degree in Electrical Engineering from the University of South Florida in 2018. His dissertation work focused on applying Immune System models towards IoT security. His current research interests include Machine learning, IoT security, Autonomous vehicles, and acoustic signal processing. At HAL, his work focuses on developing a drone detection system using Acoustic footprints and designing route planning algorithms for connected vehicles.*

Sayan Mandal *is a Ph.D. student in Electrical and Computer Engineering (ECE) at Duke University. His general research areas include multi-agent robotics, optimization, and machine learning. A KVPY scholar, he received B.Tech. in Aerospace Engineering from the Indian Institute of Technology, Kharagpur in 2019, and then joined Humans and Autonomy Lab (HAL) at Duke University. His research in HAL is to optimize processes for safety and performance on Aircraft Carrier Deck by analyzing the Spatio-temporal evolution of crew and aircraft flows. He is currently implementing statistical estimation techniques to identify risk density functions for multi-agent reinforcement learning. In the past, Sayan did a summer internship in 2018, in which he developed a machine learning algorithm for detecting drones for HAL's Prison Reconnaissance Information System (PRIS).*

Professor Mary (Missy) Cummings *received her B.S. in Mathematics from the US Naval Academy in 1988, her M.S. in Space Systems Engineering from the Naval Postgraduate School in 1994, and her Ph.D. in Systems Engineering from the University of Virginia in 2004. A naval pilot from 1988-1999, she was one of the U.S. Navy's first female fighter pilots. She is currently a Professor in the Duke University Electrical and Computer Engineering Department, and the Director of the Humans and Autonomy Laboratory. She is an AIAA Fellow, and a member of the Defense Innovation Board.*