Available online at www.jpit.az

13 (2)
2022

NEURAL NETWORK MODELS FOR DETECTION OF UNMANNED AERIAL VEHICLES BASED ON SPECTROGRAM ANALYSIS

Fargana J. Abdullayeva [a], Rashad P. Ibrahimov [b]

[a,b]Institute of Information Technology, Azerbaijan National Academy of Sciences, B. Vahabzade str., 9A, AZ1141 Baku, Azerbaijan

[a] a_farqana@mail.ru; [b] ibrahimovrashad99@mail.ru

[a] 0000-0003-2288-6255

ARTICLE INFO

http://doi.org/10.25045/jpit.v13.i2.02

Article history:
Received 25 February 2022
Received in revised form 14 April 2022
Accepted 1 June 2022

Keywords:
Unmanned Aerial Vehicles (UAVs)
Drone
Spectrogram
Convolutional Neural Network
Radar detection system

ABSTRACT

The widespread use of unmanned aerial vehicles (UAVs) in both the national and military fields has made them the focus of industrial organizations. However, the use of UAVs has seriously affected the violation of the confidentiality of personal data, posed a threat to states, national institutions, nuclear power plants, historical places. One way to reduce this threat is to detect harmful UAVs. The article develops machine learning and deep learning methods based on sound signal analysis to detect harmful UAVs. Features were extracted from the sound signals and their ensemble was created. The created new data was transmitted to the input of neural network models in the form of vectors and drones were detected. The effectiveness of the proposed approach has been tested on a database open to scientific research.

1. Introduction

Unmanned Aerial Vehicle (UAV), also known as drone, is a pilotless and remotely controlled plane (Abdullayeva & Valikhanli, 2022). UAVs are widely used in our everyday lives. UAVs can be used to deliver medications and humanitarian supplies to inaccessible territories. They can be used during inspection process of power lines and pipelines. Ministries of Emergency Situations operating around the world use unmanned aerial vehicles to monitor and forecast dangerous situations and monitor dangerous objects. UAVs also monitor road traffic jams and river traffic jams. UAVs play a significant role in many social missions such as border control, wildlife research, weather monitoring and military trainings. There are also drawbacks of technical progress in UAV. Recently, UAVs have also been used for terrorist and reconnaissance purposes. In recent years, UAVs have become one of the most actively developed areas of military aviation in the world (Abdullayeva, 2021). The absence of a pilot on the aircraft allows to increase the range and duration of UAV flights. Emergence of UAVs is perceived as a trend of robotization of the military forces of different countries. UAVs are used as a tool to combat against international terrorism. This has significantly changed the traditional warfare methods.

In every country, there are territories where unauthorized flight of UAVs can cause serious security and privacy problems and even lead to accidents. These territories may include airports, military facilities, hospitals, prisons, border lines, nuclear power plants and oil refineries. UAVs are most commonly used for remote video surveillance in civilian areas where confidentiality is required. In order to prevent such cases, it is necessary to detect the unauthorized flight of UAVs over vulnerable territories in time.

In (Taha & Shoufan, 2019), machine learning methods were analysed in order to detect drones. In (Al-Emadi et al., 2019), an autonomous system based on acoustic characteristics was developed in order to detect drones. In (Peacock & Johnstone, 2013) an approach was proposed in order to detect UAVs using the pathway on 802.11 protocol.

The disadvantage of the existing work is that these studies used one or two features to detect UAVs. This greatly reduces the accuracy of drone detection. In the presented work, the detection of UAVs is based on an ensemble of acoustic signals created by the integration of spectrograms. "Mel-frequency cepstral coefficients (MFCC)", "Spectral Centroid", "Zero Crossing Rate", "Chroma Frequencies", "Spectral Roll-off" are among features extracted from the sound signals. Here, creation of an ensemble of signals extracted from sound signals allowed to detect drones with high accuracy.

2. Related work

Many studies have been conducted in the field of drone detection. Juhyun et al. (2020) proposes a drone detection method based on sound signals analysis. It uses FFT (Fast Fourier Transform) method to extract features, and Plotted Image Machine Learning (PIL) and K Nearest Neighbors (KNN) algorithms to implement classification. Dumitrescu et al. (2020) explores the possibility of creating and using an intelligent, flexible and reliable acoustic system designed to determine, find and transmit the location of unmanned aerial vehicles (UAVs). Software functional components of proposed detection and positioning algorithm were developed using acoustic signal analysis and convolutional neural networks. In this study Cohen class fragmentation of a sound sample collected from a spiral microphone array, log-Mel spectrograms, harmonic-pulse source separation, and spectrograms in raw audio waveforms were used as input data to the Convolutional Neural Networks. Al-Emadi et al. (2019) proposed drone detection and identification methods using deep learning methods. Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Convolutional Recurrent Neural Network (CRNN) algorithms are used in order to conduct classification. Boban et al. (2022) establish a radio frequency drone database, and develops detection and identification methodologies to detect single or multiple drones and to identify a single detected drone type. The article focuses on the use of deep learning algorithms, especially fully connected deep neural networks as an anti-drone solution in two different radio frequency ranges.

3. UAV detection methods

Mass proliferation of drones has led security structures of many countries to develop new technologies capable to confront drones. The methods and means of neutralizing UAVs include followings (Yaacoub et al., 2020):

- Video-based detection. Video-based detection uses both graphical and electrical camera sensors to recognize moving objects in a monitored environment. Overall, advertising cameras can reach a working range of about 350 feet, which leads to a quiet, desirable neighborhood for monitoring. This method uses features such as color, contour lines, shapes, and edges to classify a typical drone object. Remote monitoring algorithms can also be used to evaluate elements over sequential frames. This can help to identify different objects of the same shape through ordinary gestures, such as drones and birds (for example, to distinguish between artificial drone movements and natural bird movements) (Ganti & Kim, 2016). Cameras installed in these systems are also very sensitive to lighting conditions and require the target to be in their line of sight in order to detect flying objects. In addition, numerous studies have contributed to the development of a system for detecting and identifying drones from surveillance videos.
- Sound-based detection. Recent studies are directed at using sound detection method in order to detect and identify drones using tools such as correlation analysis. Methods used in voice recognition and differentiation of drones from other objects include Basic Neural Network and Convoluted Neural network.
- Radar-based detection. Radar-based detection uses the electromagnetic principle of reverse dissipation theory to identify drones. The traditional radar method is based on the observation that aircraft or flying objects usually show a wide radar cross section (RCS). It shows that most modern drones are mechanical quadcopters with low RCS. The main

disadvantage of this method depends on the building materials, some of which have dielectric properties close to air and, as a result, are less reflected in the conductor. Therefore, new research is using updated forms of radar detectors that use the energy dissipated back from propellers and rotors.

- Radio frequency-based detection. Radiofrequency-based detection systems are based on use of radio frequency signals in order to contact ground stations of the drones. Drone network protocols are usually performed on 2.4 and 5 GHz bands that are also used for Wi-Fi communication. Additionally, camera-equipped drones also usually transmit the video streams to the control system using the same wireless channel.
- Laser-based detection. The U.S. Navy is testing a low-power laser system that can detect, track, and destroy air targets moving on the battlefield. The system was developed by Boeing and can destroy drones, artillery shells and very low-flying aircraft approaching the ship. This system is called the Laser Weapon System (LWS). It is the most compact of the laser defense type systems currently available. The availability of laser here brings great mobility to the system. The laser can detect a target at a distance of up to 3 km, and its destruction zone is 1.6 km (US Marines Test Boeing Laser To Knock Down Drones, Enemy Artillery, 2022). Solid matter laser forms the basis of the system; this laser operates in infrared band. It can operate in low power mode to disable target sensors or in high power mode to destroy the target. Its power is up to 30 kW. Target destruction time is 2 seconds.

3.1. Proposed approach

An architecture of the acoustic feature-based UAV detection system is depicted in Figure 1.

The microphones installed here record sounds from the environment. The sounds recorded by the microphones are then transmitted to the Ground Control Stations, where the data is collected.

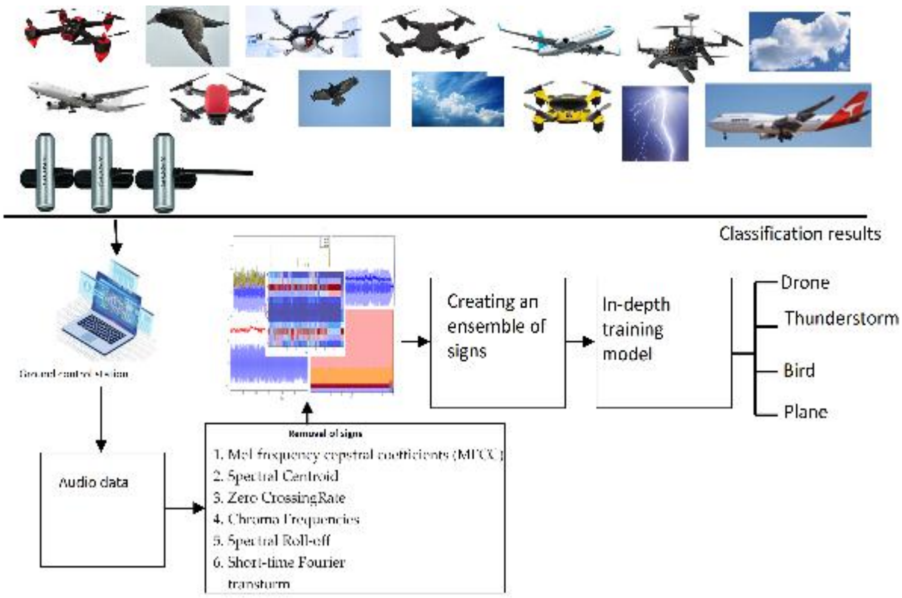


Figure 1. UAV detection system based on ensemble of audio features

Features are extracted and ensembled by applying signal extraction methods to the data collected in the form of sound signals. Data is classified by applying classification algorithms to data formed as vectors. Here, features are extracted using "Mel-frequency cepstral coefficients (MFCC)", "Spectral Centroid", "Zero Crossing Rate", "Chroma Frequencies", "Spectral Roll-off" methods. Data classification is based on a classifier that combines the capabilities of a neural network and a deep convolutional neural network.

4. Features

UAV detection is performed based on unification of spectrograms of following features extracted from audio signals: "Mel-frequency cepstral coefficients (MFCC)", "Spectral Centroid", "Spectral bandwidth", "Chroma Frequencies", "Spectral Roll-off".

4.1. Spectrogram

A spectrogram is a visual representation of the frequency spectrum of a signal as it varies with time. Here, axis x represents time and axis y represents frequency. Main idea of the spectrogram is to measure signal frequency and amplitude. Signal amplitude and frequency are the two main components of any spectrogram. Frequency is the number of waves passing through a given point in a single instant of time. Spectrogram can be in different colors. Here, density of colors reflects the strength of the signal. Frequency characteristics of the signal can be obtained by performing a Fourier transform on the signal. Visual representation of the spectrogram is depicted in Figure 2.

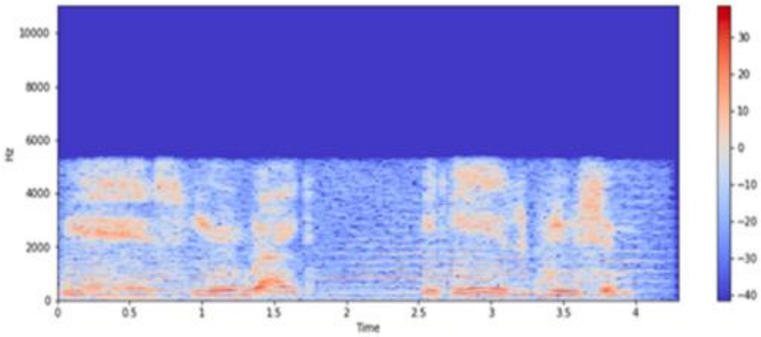


Figure 2. Spectrogram

In this study, each audio file was converted into its relevant spectrogram and different features were extracted from these spectrograms. These features include: “Mel-frequency cepstral coefficients (MFCC)”, “Spectral Centroid”, “Zero Crossing Rate”, “Chroma Frequencies”, “Spectral Roll-off” (Figure 3).

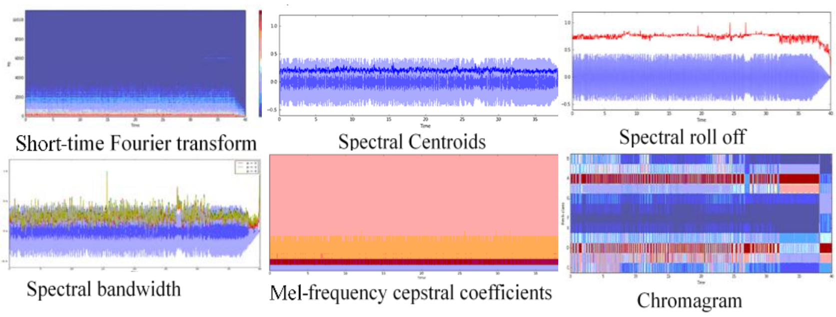


Figure 3. Audio file spectrograms

4.2. Mel-Frequency Cepstral Coefficients (MFCC)

Mel spectrum models capacity of the signal derived from Fourier transform to resolve non-linear frequencies based on Filterbank human ear audio spectrum and is reduced at higher frequencies (Young et al., 2002). Filterbank is an array of band conducting filters that separate input signals into many components (Figure 4).

$$Mel(f) = 2595 \log 10 \ (1 + \frac{f}{700}) \qquad (1)$$

where, f is frequency and M is mel frequency.

Used filters are triangular and they are located at equal mel-distances defined by following:

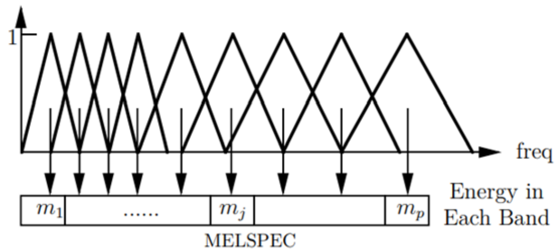


Figure 4. Mel-Scale Filterbank

Mel-Frequency Cepstral Coefficients (MFCCs) calculate log filterbank amplitudes {mj} using the Discrete Cosine Transform.

$$C_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{n} m_j \ \cos(\frac{\pi i}{N}(j - 0.5)) \qquad (2)$$

Where N is the number of filterbank channels and C_i are the cepstral coefficients.

4.3. Spectral Centroid

Spectral centroid is a measure used in digital signal processing in order to characterize a specter. This demonstrates the location of spectral mass center. It is calculated as the average weight of frequencies found in the signal, defined by Fourier transform and their magnitude is given as weights:

$$Centroid = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \quad (3)$$

where $x(n)$ is the weighted frequency value or the magnitude of the box number n, $f(n)$ is the central frequency of that box.

4.4. Zero Crossing Rate

The zero crossing rate (ZCR) is the rate at which a signal changes from positive to zero to negative or from negative to zero to positive. Its value is widely used in speech recognition and audio data search, and is the key feature for classification of percussive sounds. Zero crossing rate is defined by following formula:

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{R<0}(s_t s_{t-1}) \quad (4)$$

where S is a signal and $1_{R<0}$ is an indicator function.

4.5. Chroma Frequencies

librosa.sign.chroma stft() function from python library was used to convert an audio file into a chromagram.

4.6. Spectral Roll-off

It can be defined as the movement of a special type of a filter intended to remove the frequencies from a defined bandwidth. It's called "roll-off" as it is a step-by-step procedure. There are two types of filters: high-pass and low-pass and both can erase the frequency that deviates from its bandwidth. Figure 5 depicts high-pass roll-off on the blue line and white low-pass slide.

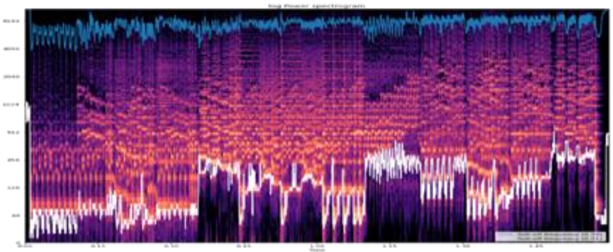


Figure 5. Spectral roll-off

4.7. Data scaling methods

Data scaling methods include standardization and normalization methods. During normalization, data are scaled so that their value varies between 0 and 1. This method is also known as min-max scale. Data is normalized using following formula:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5)$$

where X_{max} and X_{min} are the maximum and minimum values of the feature.

Standardization is a method of conversion of feature values in the database. When performing normalization based on standardization, an average value is subtracted from each feature sample and the result is divided into standard deviation of that feature. Feature of each standard element in the database is calculated as following:

$$x_i = \frac{x_i - \mu}{\sigma} \quad (6)$$

where x_i - is the standardized element of the feature, x_i - initial element, μ - arithmetic mean, i.e., standard deviation.

4.8. Evaluation metrics

Confusion matrix. Confusion matrix is a matrix in an $x \times x$ dimention showing how accurately the model works. Here, n is the number of classes. Columns of the matrix demonstrate the true classes of target object and in our case, there are four classes: drone, thunderstorm, bird and plane. Rows of the matrix show the classes predicted by the proposed method.

Precision indicates what percentage of positive predicted input data is from a true positive class sample. Precision metrics are calculated as following.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Precision metrics values vary between 0 and 1. Precision is separately calculated for each class. For example, the precision value of a drone class indicates what percentage of all input data predicted as a drone class is actually a drone.

Accuracy is the overall efficacy indicator of the model. Accuracy shows that the proposed method accurately determines what percentage of the data is positive and what percentage is negative. Accuracy is calculated using following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

Recall shows what percentage of total data in a positive class is positively predicted. Recall is separately calculated for each class. For example, the recall value of a drone class indicates what percentage of all input data in the drone class is correctly classified as a drone. Recall is calculated as following:

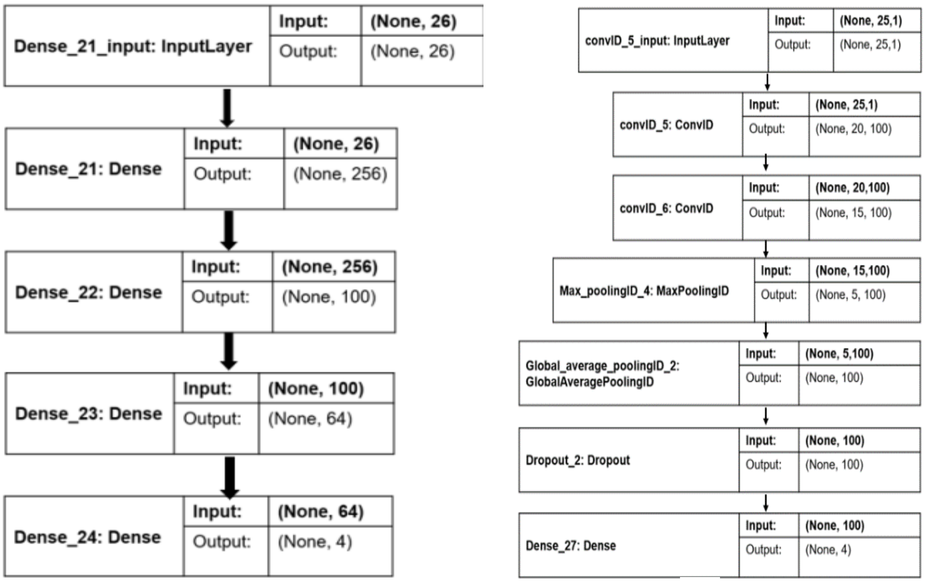
$$Recall = \frac{TP}{TP+FN} \quad (9)$$

F1-score−is the harmonic mean of recall and precision, and separately calculated for each class.

Loss is a method of estimating how well a method models data. The loss function gets a high value when the prediction model makes a lot of mistakes, and a low value when it makes no mistakes. There are many types of loss functions in machine learning. Mean squared error (MSE) and cross-entropy loss are the most widely used loss functions.

5. Experiments

Simple Neural Network and Convolutional Neural Network (CNN) architectures were constructed to detect UAVs. The structure of constructed models is depicted in Figure 6.



(a) (b)

Figure 6. Simple Neural Network (a) Convolutional Neural Network (b)

256 neurons are used in the first layer of the neural network, 100 - in the second layer, and 64 - in the third layer. In the model, Relu is the activation function, and cross-entropy is the loss function. Parameter optimization is based on the Adam function.

An experimental study of the approach was conducted based on the "Malicious UAVs Detection" database (Malicious UAVs Detection, n.d.). The Sony ECMCS3 Stereo microphone is used to collect audio files.

Database consists of four classes of audio data titled "Drones (0)", "birds (1)", "Thunderstorms (2)" and "Planes (3)". Overall, the database stores 1053 samples. Recording of each audio file sample for each class of the database is depicted in Figure 7.

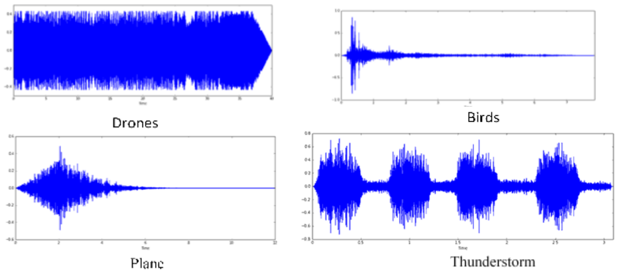


Figure 7. Audio files in "Malicious UAVs Detection" database

In order to test the neural network architectures built on the database, a value is calculated for each audio file by applying feature extraction methods to them. These values are integrated to create a feature ensemble. Created feature ensemble is provided in Table 1.

Table 1. Creating a feature ensemble

№	File name	Class	chroma_stft	spec_cent	spec_bw	rolloff	zcr	mfcc_0	...	mfcc_19
1	93.wav	Plane	0.609105	853.243	969.246	1597.22	0.048543	-363.688		-1.60867
2	204.wav	Plane	0.519653	1302.58	1174.79	2291.07	0.096203	-292.252		6.39921
3	216.wav	Plane	0.519653	1302.58	1174.79	2291.07	0.096203	-292.252		6.39921
4	11.wav	Plane	0.597159	2520.02	2014.77	4604.78	0.174823	-266.186		-7.77581
5	178.wav	Plane	0.597159	2520.02	2014.77	4604.78	0.174823	-266.186		-7.77581

Since the value of each feature is a number in a different format, a standardization operation is performed on the obtained values to bring them into a single form. Drones are detected by applying classification algorithms to a new database created on the basis of performed operations. The effectiveness of the proposed model wisas assessed for accuracy, precision, recall, f1-score, and loss metrics. Results of the experiments are provided in Table 2.

Table 2. Effectiveness indicators of SimpleNN və CNN

Method	Classes	Accuracy	Precision	Recall	F1-score
SimpleNN	Drone	1.00	1.00	1.00	1.00
	Birds	1.00	0.94	1.00	0.97
	Thunderstorm	0.89	1.00	0.89	0.94
	Plane	1.00	1.00	1.00	1.00
CNN	Drone	0.74	1.00	0.74	0.85
	Birds	0.67	0.60	0.67	0.63
	Thunderstorm	0.31	1.00	0.30	0.47
	Plane	1.00	0.50	1.00	0.67

As seen in Table 2, SimpleNN performs the highest result in drone detection. CNN performs low values for all indicators. The algorithm is almost unable to recognize Thunderstorm class samples, and Accuracy, Recall, and F1-score metrics of the model account for 0.31, 0.30, and 0.47, respectively. Results from Table 2 are visually depicted in Figure 8 (SimpleNN – first row, CNN – second row).

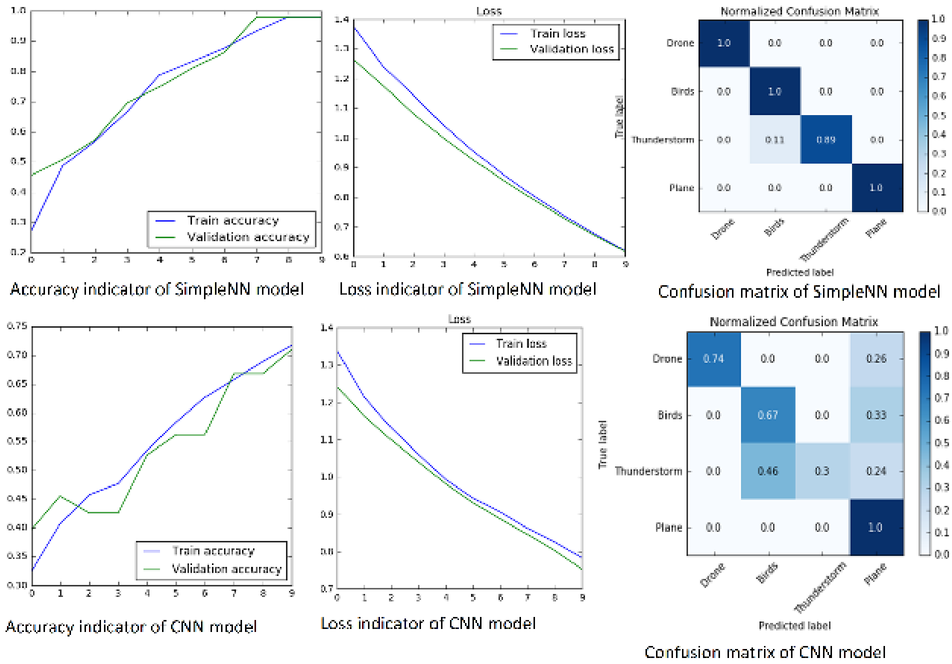


Figure 8. Visual representation of experiment results

As can be seen from the confusion matrix of the SimpleNN model in Figure 8, the algorithm is able to diagonally assemble the points across all classes. Here the model demonstrates an accuracy of 0.89, incorrectly recognizing 4 points during recognition of samples from only Thunderstorms class. CNN is only capable to accurately recognize samples from Plane class, and made mistakes while recognition of samples from other 3 classes. Model is capable to recognize samples from Drone class with 0.74 accuracy and from Bird class with 0.67 accuracy. It incorrectly includes samples from the Thunderstorm class into the Birds class with 0.46 accuracy.

Conclusion

Wide proliferation of UAVs caused security structures in many countries to develop new technologies capable of confronting UAVs. Numerous methods and tools were developed to neutralize UAVs. Existing studies used one or two features to detect UAVs. Using a low number of features in the detection process complicates detecting UAVs with high accuracy. In the presented work, the audio data was analyzed and several spectral features were extracted from the audio files. An ensemble of extracted features was created in order to detect UAVs based on audio data. Data were classified into different classes by transmitting the formed ensemble of features to the input of Simple Neural Network and Convolution Neural Network (CNN) in form of vectors. As a result of experiments conducted on real data, Simple Neural network performed highest results with 98% detection accuracy.

Acknowledgement

This work was supported by the Science Foundation of the State Oil Company of Azerbaijan Republic (Contact No. 3 LR-AMEA).

References

Abdullayeva, F., Valikhanli, O. (2022). Development of a method for detecting GPS spoofing attacks on unmanned aerial vehicles. Problems of Information Technology, vol. 13, no. 1, pp. 3-8 http://doi.org/10.25045/jpit.v13.i1.01

Abdullayeva, F. (2021). Development of a support model to ensure the cyber security of UAV components, The second Karabakh War as a new generation warfare, Proceedings of the international scientific-practical conference dedicated to the anniversary of the victory achieved in the 44 days patriotic war, pp. 129-132. (Azerbaijani)

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell., J., Ollason, D., Povey, D., Valtchev, V., Woodland, P. (2002). The HTK Book, (for HTK Version 3.2), 355 p. https://www.danielpovey.com/files/htkbook.pdf

Taha B. & Shoufan A. (2019). Machine Learning-Based Drone Detection and Classification. State-of-the-Art in Research, IEEE Access., 69-82.

Al-Emadi S., Al-Ali A., Mohammad A., Al-Ali A. (2019). Audio Based Drone Detection and Identification using Deep Learning. Proceedings of the 15th IEEE International Wireless Communications & Mobile Computing Conference (IWCMC), Tangier, Morocco. pp. 459- 464.

Peacock M., Johnstone MN. Towards detection and control of civilian unmanned aerial vehicles, Proceedings of the 14th Australian Information Warfare and Security Conference; 2013 December 2-4; Perth, Western Australia. p. 9-15.

Yaacoub JP., Noura H., Salman O., Chehab A. (2020). Security analysis of drones systems: Attacks, limitations, and recommendations. Internet of Things, 1-39.

Ganti SR., Kim Y. (2016). Implementation of detection and tracking mechanism for small UAS. Proceedings of the IEEE International Conference on Unmanned Aircraft Systems (ICUAS), Arlington, VA, USA. p. 1254-1260.

Malicious UAVs Detection. (n.d.). https://www.kaggle.com/sonain/malicious-uavs-detection

Juhyun K., Cheonbok P., Jinwoo A., Youlim K., Junghyun P., John C.G. (2017). Real-time UAV Sound Detection and Analysis System, Proceedings of the IEEE Sensors Applications Symposium (SAS), pp. 1-5.

Dumitrescu C., Minea M., Costea I.M., Cosmin Chiva I., Semenescu A. (2020). Development of an Acoustic System for UAV Detection. Sensors, 20(17):4870. https://doi.org/10.3390/s20174870

Sara Al-Emadi, Abdulla Al-Ali, Amr Mohammad, Abdulaziz Al-Ali. (2019). Audio Based Drone Detection and Identification using Deep Learning, Proceedings of the 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp 459-464

US Marines Test Boeing Laser To Knock Down Drones, Enemy Artillery, https://www.ibtimes.com/us-marines-test-boeing-laser-knock-down-drones-enemy-artillery-2011610, Accessed 09.03.2022.

Boban SJ., Ivan P., Jovan B., Boban B., Danilo O. (2022). Single and multiple drones detection and identification using RF based deep learning algorithm. Expert Systems with Applications, Vol. 187, 11592