Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

DOI: 10.37943/19XNOV6347

Dana Utebayeva
PhD, Researcher, Department of Electronics, Telecommunications and ST
d.utebayeva@satbayev.university, orcid.org/0000-0002-5535-9200
Satbayev University, Kazakhstan

Lyazzat Ilipbayeva
Candidate of Technical Sciences, Acting associate professor,
Department of Radio-engineering, Electronics, Telecommunications
l.ilipbayeva@iitu.edu.kz , orcid.org/0000-0002-4380-7344
International Information Technology University, Kazakhstan

INVESTIGATION OF DEEP LEARNING MODELS BASED ON SINGLE-LAYER SimpleRNN, LSTM AND GRU NETWORKS FOR RECOGNIZING SOUNDS OF UAV DISTANCES

Abstract: In recent years, the potential risks posed by easily moving objects have highlighted the need for intelligent surveillance systems in protected areas, primarily to ensure the safety of human lives. Among the most common of these objects are unmanned aerial vehicles (UAVs). Recent advances in deep learning techniques for recognizing audio signals have made these techniques effective in identifying moving or aerial objects, especially those powered by engines. And the growing deployment of UAVs has made their rapid recognition in various suspicious or unauthorized circumstances critical. Detecting suspicious drone flights, especially in restricted areas, remains a significant research challenge. It is vital to perform the task of determining their distance in order to quickly detect drones approaching people in such protected areas. Therefore, this paper aims to study the research question of recognizing UAV audio data from different distances. That is, recognizing drone audio at different distances was experimentally studied using Simple RNN, LSTM and GRU based deep learning models. The main objective of this study is based on finding one of the capable types of recurrent network for the task of recognizing UAV audio data at different distances. During the experimental study, the recognition abilities of Single-layer Simple RNN, LSTM and GRU recurrent network types were studied from two basic directions: with recognition accuracy curves and classification reports. As a result, LSTM and GRU based models showed high recognition ability for these types of audio signals. It was noted that UAVs can reliably predict distances greater than 10 meters based on the proposed deep learning architecture.

Keywords: UAVs; UAV states; UAV sound recognition; UAV sound distance recognition; suspicious drone; SimpleRNN network; LSTM network; and GRU network.

Introduction

Nowadays, UAVs with various functions are being launched more frequently [1], [2]. Shows featuring them also became popular during holidays and large-scale events [3]. As well, their use is also noted in border areas [4]. Such drone incidents are discussed comprehensively in [5]. Drone incidents can occur both during recreational use and intentionally for malicious purposes, as [5] reveals. In both cases, the appearance of suspicious UAVs requires timely determination of their distance in protected areas. This is due to the fact that incidents with UAVs can harm human life or protected objects [6], [7]. Drone distance control systems [8] are especially

Received: 23.08.2024 Accepted: 24.09.2024 Published: 30.09.2024

DOI: 10.37943/19XNOV6347

necessary in border areas of any country. This is due to the fact that border areas are among the most protected areas. On the other hand, areas in need of protection include educational institutions, especially kindergartens and schools, where children and adolescents who have not yet reached adulthood and have not yet mastered the ability to fully protect themselves. That is why detecting the distance of drones is becoming a significant problem [8].

It is possible to separate out the following significant factors as to why estimating or detecting UAV distances is relevant:

1) To promptly conduct investigations of UAVs approaching a person in crowded places during various events;

2) Insurance measures in accordance with the legislation on maintaining minimum distances to other flying objects and people in the air when flying unmanned aerial vehicles that are in a suspicious condition;

3) Additional control system in case of UAV failure;

4) Measures for the timely protection of objects and people from incidents that may fail in the case of a UAV with a load.

All the above reasons indicate the relevance of the problem of UAV distance recognition. These compelling reasons require the development of a system for estimating the flight range of drones. So, one of the new directions is the study of the system for recognizing the distance of UAVs by their sounds. Therefore, this research work intended to develop systems for recognizing sound signals of various UAV distances. To achieve this goal, our previous research work [8] investigated the system for recognizing sound signals of various UAV distances using a single-layer GRU network. In that study, other recurrent networks were not investigated for comparison and finding an effective one. Therefore, in this research work, the goal is to study the deep learning model using the Single-layer SimpleRNN network, LSTM network and GRU network for recognizing sound signals of various UAV distances. The main objective was to study experimentally these recurrent networks to find out their recognition capabilities for UAV distance sound recognition. That is, to test comprehensively the recognition capabilities of deep learning neural models based on SimpleRNN, LSTM and GRU networks in a single RNN based layer architecture only. The main reason for this is to test and study the capabilities of these three recurrent networks on our chosen audio signal. That is, to determine the most capable of them. The next section will review the literature on UAV audio detection systems.

Literary review

With the development of deep learning frameworks, new possibilities for sound recognition are opening up. Creating deep neural layers based on RNNs is especially effective for time-varying signals. This is due to the fact that some works such as Classification of Environmental Sounds [9], Dangerous Sound Detection in Urban Area [10], Recognition of Heart Sound Segments [11], [12], Urban Sound Classification using Deep Learning [13], Polyphonic Sound Event Detection [14], and Bowel Sound Detection [15] studied from the point of view of experience can serve as a reliable basis for processing sound signals. The following sound recognition tasks are mainly considered using RNN types: speech recognition [16], [17], [18], recognition of environmental sounds [10], [19], animal sounds [20], sounds of cars and motorized objects [21], medical sounds [12], [22], safety signals [23] and sounds of musical instruments [24]. One of the areas that requires further research in this direction is the sounds of cars and motorized objects. This is due to the fact that when sounds of this category overlap, frequency data may present some difficulties in recognition. Therefore, research in this direction is still needed. Therefore, this paper considers the sounds of unmanned aerial vehicles, which are one of such motorized objects.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

In general, UAV sounds are studied using machine learning and deep learning methods for binary and multi-class classification tasks. Binary classification implies the presence of a UAV in protected areas or not. And multi-class classifications of this direction consider the problems of recognizing the presence of a certain load on a drone or its absence, or recognizing other states of the UAV and their distances. For instance, according to the research work [25], a machine learning model and an empirically optimized node configuration for deployment were used to propose a multi-acoustic node UAV detection system. For training, short-time Fourier transform (STFT) features and Mel frequency cepstral coefficients (MFCC) were employed. Training of Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) was investigated. Their test set's optimal configuration for maximizing the detection range without blind spots was chosen from the four sensor node arrangements that were put in place. The authors claimed that that their STFT-SVM model performed the best, and that the ideal configuration was a semicircle formation with a distance of 75 meters between the node and the protected location.

In order to protect protected areas, a research paper [26] investigated how well Machine Learning methods worked in solving the problem of detecting unauthorized UAV flights. That is, the goal of their efforts was to determine whether it would be feasible to create and use an intelligent, adaptive, and robust acoustic system designed to recognize, monitor, and report the location of unmanned aerial vehicles (UAVs). Their proposed alert system used Wigner-Ville time-frequency analysis to recognize specific audio signals to obtain an audio fingerprint of the UAV. To improve the recognition accuracy, CoNN, MFCC and MIF (mean instantaneous frequency) for each drone type were also included in their process.

The paper [27] also used Machine Learning (ML) algorithms to recognize UAVs based on their audio data. The authors aimed to estimate the audio characteristics of UAVs and provide a classification system. Their methods trained four ML classification approaches, namely Neural Network (NN), Support Vector Machine (SVM), naive Gaussian Bayes (GNB), and K-nearest neighbours (KNN), using five individual features and one combination of features.

In the research paper [28], a method for identifying small aircraft using a sound array of five microphones was proposed. That is, they proposed a system with four microphones, symmetrical to the fifth in the horizontal plane along the orthogonal direction. It is proposed to use correlation or phase detection methods depending on the properties of the sound radiation. The assessment of the angular coordinates of the object included two stages. The angle of arrival of the acoustic wave was determined after the arrival sector was established. The method they proposed simplifies the algorithm for processing the received audio data and reduces the amount of equipment required.

The study [29] attempted to investigate whether Deep Neural Networks could possibly be employed for assessing audio data from commercial hobby drones in order to recognize them in actual-life situations. Their effort was intended to improve the detection system for drones used for malicious activities, like terrorism. They specifically demonstrated a technique based on audio event detection that can recognize the existence of commercial hobby drones as a binary classification. In their work, the sounds made by a number of well-known commercial hobby drones were recorded and UAV presensense class was conducted using these recordings. And then added additional environmental sound data to make up for the absence of drone audio data.

Deep learning (DL) methods were also used in the study [30] to identify and categorize payload-carrying UAVs according to the sound they produce. To examine the final dataset, the authors considered hybrid convolutional recurrent neural networks (CRNNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs). They demonstrated results of

DOI: 10.37943/19XNOV6347

their ability to accurately classify the noise class and classes of unloaded, single and dual payloads of unmanned aerial vehicles using only sound. Therefore, the accuracy scores of 0.9493, 0.8133, and 0.9174 were explained by the better performance of MFCC in CNN, RNN, and CRNN. The cost-effective approach to collecting the study data, which used laptop microphones, was the authors' contribution. The limitation of this paper is due to the fact that the data was collected using only two UAV models and one type of payload.

The study [31] presented various methods for UAV identification and detection. They used audio data from previous studies to compare UAV identification and detection in their study. They tried to study Transformer Encoders (TE), Long Short-Term Memory (LSTM), Convolutional Long Short-Term Memory (CLSTM), Deep Neural Networks (DNN), and Convolutional Neural Networks (CNN) as deep learning methods. In their proposed work, the UAV identification task showed good performance with LSTM network. So, the research works [5], [6], [8] provide a thorough analysis of other research on drone sound detection.

Methods and Materials

Since UAV sound recognition is performed for different distances, first the sounds of the UAV distance was recorded, which is described in detail in the following section "Data Preparation". And, in the next step, deep learning model architectures based on SimpleRNN, LSTM and GRU recurrent networks were built. Experimental work on UAV sounds was carried out using these deep learning models based on SimpleRNN, LSTM and GRU.

Data preparation

The sounds in this database were recorded using the sounds of «DJI Mini 2» and «Qazdrone» drones, as we mentioned in the previous study [8]. We focused on the speed of «DJI Mini 2» drone. Because «DJI Mini 2» is a fast-flying drone compared to «Qazdrone». Therefore, the distances of the drones were obtained relative to the flight speed of DJI Mini 2 promised by the manufacturer of this drone. In addition, the flight speed promised by the manufacturer of this drone is shown in Table 1 below in three modes **«S», «N» and «C»,** [32]. The reason is that the speeds of these modes work in windy weather and other environmental conditions.

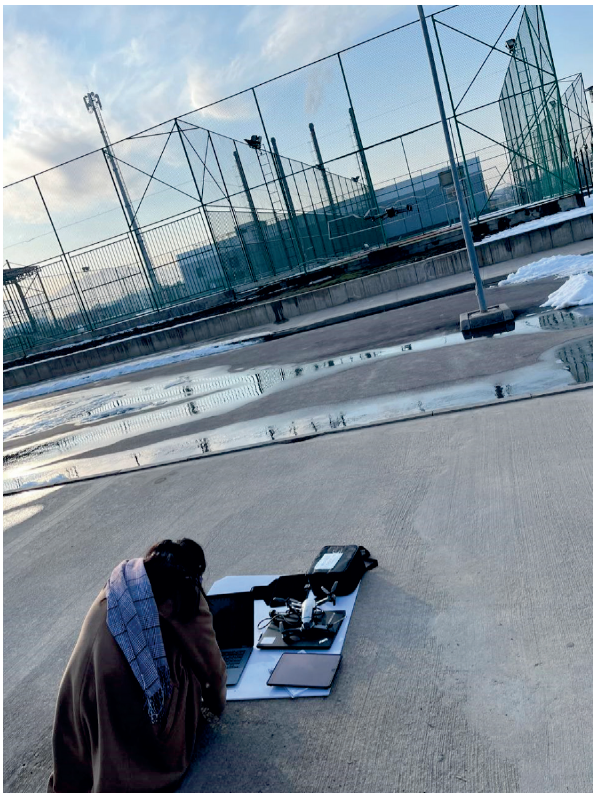
Table 1. Technical characteristics of the UAV "DJI Phantom mini 2"

Total flying duration: 31 min	**Mode «S»**	**Mode «N»**	**Mode «C»**
Maximum ascent speed:	5 m/s	3 m/s	2 m/s
Maximum descent speed:	3.5 m/s	3 m/s	1.5 m/s

As it can be seen from Table 1 above, the drones in our database can fly a maximum of 5 meters in 1 second in "S" mode. That is why the difference in distance in the database is 5 meters. We set the maximum speed at which the drone can fly. In reality, these drones fly at an average speed of 3 m/s per second. Thus, the sounds of the UAVs launched from the ground were recorded every 5 meters. In this experiment, a laptop microphone was chosen, Figure 1. This is because laptop microphones are not as good as other high-quality microphones, and our research goal was to study a system that can function well even with poor signal quality. Therefore, the sounds were recorded using a microphone of a regular laptop and saved in a special folder.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

a)



b)

Figure 1. Moments of recording the sounds of UAV distances

Moreover, suppression of other background noises was not considered. This is due to the need to distinguish sounds reliably in cases where there were other overlapping sounds in parallel. The microphone of the laptop was placed on the terrain, as shown in Figure 1b. It was designed as a station projecting an acoustic recognition sensor onto the ground. At first, the sounds of UAVs flying at an altitude of 5 meters with a radius of 2-3 meters were recorded. Then this practice of recording at higher altitudes was continued. Sounds at this distance were taken as one class. In the next step, another height of 5 meters was added, so that up to a height of 50 meters, sounds were recorded and saved as classes. In general, an experiment was conducted to study the possibility of processing such a system so that it can be carried out in several places with a diameter of 50 meters, Table 2.

Table 2. Classes and durations of UAV sounds

2a

Classes	Duration, in seconds (s)		
	Total, (s)	*Train (s)*	*Validation/Test,s*
5 meters	534	434	100
10 meters	519	419	100
15 meters	510	410	100
20 meters	501	401	100
25 meters	491	391	100
30 meters	534	434	100
35 meters	784	684	100
40 meters	642	542	100
45 meters	587	487	100
50 meters	737	637	100
Ambient Noise	798	698	100

2b

Classes	Duration, in seconds (s)		
	Total, (s)	*Train (s)*	*Validation/Test,s*
5 meters	534	434	100
15 meters	510	410	100
25 meters	491	391	100
35 meters	784	684	100
45 meters	587	487	100
Ambient Noise	798	698	100

After the database was collected, the sounds were filtered at 16 kHz. This is because information with a frequency higher than 16 kHz was not reflected for our data. This filter was used according to the concept of previous studies in this direction [8]. The recorded databases

DOI: 10.37943/19XNOV6347

were reorganized into two different compositions and received two different database names as given in Table 2a, 2b. The first database was the original database with all the recordings. There were 11 classes: from 5 meters to 50 meters, sounds at a distance of 5 meters and other background noises were recorded, "Table 2a". The second database was reassembled to study reliable recognition. In this case, a distance of 10 meters was observed. Some distances were not taken into account in the second database, "Table 2b".

Deep Learning Model Based on SimpleRNN network.

RNN neural networks belong to a group of networks widely used in time-varying event recognition and processing [13], [14]. The widely used RNN neural network types are "SimpleRNN", "LSTM" and "GRU". SimpleRNN is a basic form of recurrent networks. These networks are suitable for sequential data processing. Therefore, this paper considered the RNN networks types for UAV sounds. First of all, a deep learning model was created based on the SimpleRNN network as in Figure 2.

UAV sounds are type of time-varying signal. And applying this type of network to the UAV sound distance recognition task from an experimental point of view was one of the first objectives of this study. Deep learning frameworks are models that consist of a set of multiple deep neural layers, and recently consider one of the CNN or RNN network types as the core layer. Since RNN network types are used in sound processing, this research work considered the experiments with recurrent network types. And in our experimental studies, the core structure of the deep learning model consisted of 17 layers, as shown in Figure 2-4. In all three models, the 9th layer is a recurrent type network.

Model: "SimpleRNN"

Layer (type)	Output Shape	Param #	Connected to
stft_9_input (InputLayer)	[(None, 16000, 1)]	0	
stft_9 (STFT)	(None, 100, 257, 1)	0	stft_9_input[0][0]
magnitude_9 (Magnitude)	(None, 100, 257, 1)	0	stft_9[0][0]
apply_filterbank_9 (ApplyFilter	(None, 100, 128, 1)	0	magnitude_9[0][0]
magnitude_to_decibel_9 (Magnitu	(None, 100, 128, 1)	0	apply_filterbank_9[0][0]
batch_norm (LayerNormalization)	(None, 100, 128, 1)	256	magnitude_to_decibel_9[0][0]
reshape (TimeDistributed)	(None, 100, 128)	0	batch_norm[0][0]
td_dense_tanh (TimeDistributed)	(None, 100, 64)	8256	reshape[0][0]
SimpleRNN (SimpleRNN)	(None, 100, 64)	8256	td_dense_tanh[0][0]
skip_connection (Concatenate)	(None, 100, 128)	0	td_dense_tanh[0][0] SimpleRNN[0][0]
dense_1_relu (Dense)	(None, 100, 64)	8256	skip_connection[0][0]
max_pool_1d (MaxPooling1D)	(None, 50, 64)	0	dense_1_relu[0][0]
dense_2_relu (Dense)	(None, 50, 32)	2080	max_pool_1d[0][0]
flatten (Flatten)	(None, 1600)	0	dense_2_relu[0][0]
dropout (Dropout)	(None, 1600)	0	flatten[0][0]
dense_3_relu (Dense)	(None, 32)	51232	dropout[0][0]
softmax (Dense)	(None, 11)	363	dense_3_relu[0][0]

Figure 2. Architecture of Deep Learning Model based on "SimpleRNN"

Deep Learning Model Based on LSTM network. In the second step, the hyper-parameters of the LSTM-based structure were set while preserving the original structure, Figure 3. Here, the LSTM network is represented as the main recurrent network layer in layer 9. Its cells were "64".

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

Model: "LSTM"

Layer (type)	Output Shape	Param #	Connected to
stft_2_input (InputLayer)	[(None, 16000, 1)]	0	
stft_2 (STFT)	(None, 100, 257, 1)	0	stft_2_input[0][0]
magnitude_2 (Magnitude)	(None, 100, 257, 1)	0	stft_2[0][0]
apply_filterbank_2 (ApplyFilter	(None, 100, 128, 1)	0	magnitude_2[0][0]
magnitude_to_decibel_2 (Magnitu	(None, 100, 128, 1)	0	apply_filterbank_2[0][0]
batch_norm (LayerNormalization)	(None, 100, 128, 1)	256	magnitude_to_decibel_2[0][0]
reshape (TimeDistributed)	(None, 100, 128)	0	batch_norm[0][0]
td_dense_tanh (TimeDistributed)	(None, 100, 64)	8256	reshape[0][0]
LSTM (LSTM)	(None, 100, 64)	33024	td_dense_tanh[0][0]
skip_connection (Concatenate)	(None, 100, 128)	0	td_dense_tanh[0][0] LSTM[0][0]
dense_1_relu (Dense)	(None, 100, 64)	8256	skip_connection[0][0]
max_pool_1d (MaxPooling1D)	(None, 50, 64)	0	dense_1_relu[0][0]
dense_2_relu (Dense)	(None, 50, 32)	2080	max_pool_1d[0][0]
flatten (Flatten)	(None, 1600)	0	dense_2_relu[0][0]
dropout (Dropout)	(None, 1600)	0	flatten[0][0]
dense_3_relu (Dense)	(None, 32)	51232	dropout[0][0]
softmax (Dense)	(None, 11)	363	dense_3_relu[0][0]

Figure 3. Architecture of Deep Learning model based on “LSTM”.

In [31], LSTM networks gave better recognition performance for UAV sound recognition. Our work’s method considered to test the recognition ability of the LSTM network when it is connected as a single layer and considered individually, based on the same layers as other architectures. Other types of RNN networks were also tested with a single-layer architecture.

Deep Learning Model Based on GRU network. In the third step, keeping the first two structures, the structure and hyper-parameters of the GRU network with cells “64” were built, Figure 4.

Model: "GRU"

Layer (type)	Output Shape	Param #	Connected to
stft_8_input (InputLayer)	[(None, 16000, 1)]	0	
stft_8 (STFT)	(None, 100, 257, 1)	0	stft_8_input[0][0]
magnitude_8 (Magnitude)	(None, 100, 257, 1)	0	stft_8[0][0]
apply_filterbank_8 (ApplyFilter	(None, 100, 128, 1)	0	magnitude_8[0][0]
magnitude_to_decibel_8 (Magnitu	(None, 100, 128, 1)	0	apply_filterbank_8[0][0]
batch_norm (LayerNormalization)	(None, 100, 128, 1)	256	magnitude_to_decibel_8[0][0]
reshape (TimeDistributed)	(None, 100, 128)	0	batch_norm[0][0]
td_dense_tanh (TimeDistributed)	(None, 100, 64)	8256	reshape[0][0]
GRU (GRU)	(None, 100, 64)	24960	td_dense_tanh[0][0]
skip_connection (Concatenate)	(None, 100, 128)	0	td_dense_tanh[0][0] GRU[0][0]
dense_1_relu (Dense)	(None, 100, 64)	8256	skip_connection[0][0]
max_pool_1d (MaxPooling1D)	(None, 50, 64)	0	dense_1_relu[0][0]
dense_2_relu (Dense)	(None, 50, 32)	2080	max_pool_1d[0][0]
flatten (Flatten)	(None, 1600)	0	dense_2_relu[0][0]
dropout (Dropout)	(None, 1600)	0	flatten[0][0]
dense_3_relu (Dense)	(None, 32)	51232	dropout[0][0]
softmax (Dense)	(None, 11)	363	dense_3_relu[0][0]

Figure 4. Architecture of Deep Learning model based on “GRU”

Therefore, these three different deep learning models were created. And for each model, experimental works were conducted, and the recognition accuracy curves were first given for the database that are at a distance of every 5 meters. Then, the recognition accuracy curves were obtained for the database that are at a distance of every 10 meters. The experiment at a

DOI: 10.37943/19XNOV6347

distance of 10 meters was considered because the first experiment failed with the recognition accuracy on certain classes. In both experiments, the recognition curves were analyzed first, and then the classification reports were answered. All the results are analyzed in the next section.

Results

The deep learning models based on SimpleRNN, LSTM and GRU were first experimented with the database in Table 1a, which was collected from sounds every 5 meters. Their recognition curves are shown in Figure 5.

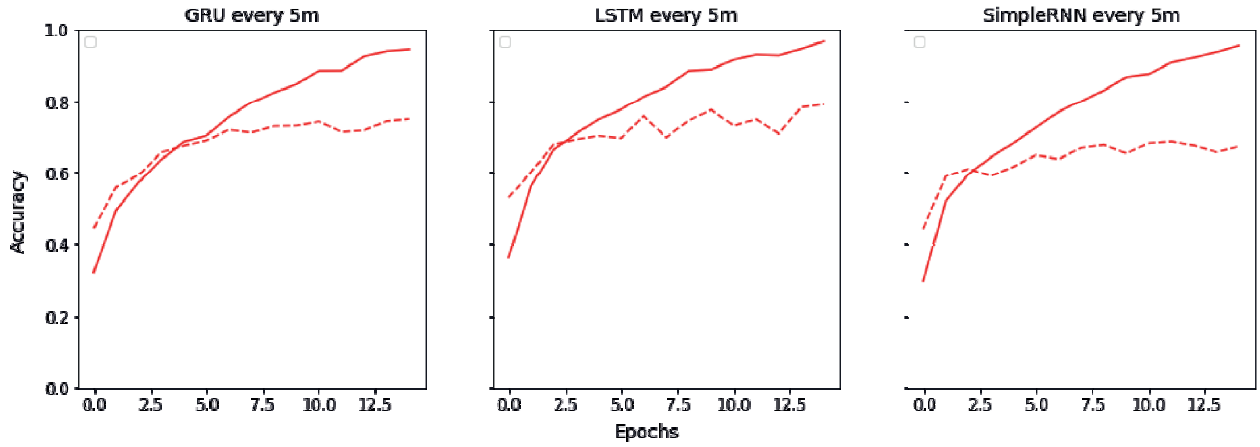


Figure 5. Recognition curves for exploring the architectures of Deep Learning models based on "Simple RNN", "LSTM" and "GRU" for drone sounds at a distance of each 5 meters

If we conceptualize Figure 5, our generated models and our database at a distance of 5 meters will have a "non-representative" appearance. This points to the limitation of Experimental Work 1 due to the small amount of data for each class in the database consisting of distances every 5 meters. However, among these three networks, it is clear that the GRU network is more flexible to the model. Since these recognition curves cannot provide complete information about the recognition, there is now a need to openly discuss the complete classification reports for each class. For this reason, the classification responses have been presented in an expanded form in Figure 6. As can be seen from Figure 6, areas near the microphone gave reliable recognition, and the further away, the lower the recognition accuracy.

	precision	recall	f1-score	support
10 meters	0.88	0.99	0.93	100
15 meters	0.88	0.67	0.76	100
20 meters	0.57	0.63	0.60	100
25 meters	0.49	0.49	0.49	100
30 meters	0.58	0.67	0.62	100
35 meters	0.48	0.82	0.61	100
40 meters	0.51	0.37	0.43	100
45 meters	0.48	0.29	0.36	100
5 meters	0.99	0.93	0.96	100
50 meters	0.56	0.49	0.52	100
Ambient Noise	0.95	0.96	0.96	100
accuracy			0.66	1100
macro avg	0.67	0.66	0.66	1100
weighted avg	0.67	0.66	0.66	1100

Figure 6. Results of UAV sound classification recognition at a distance of 5 meters when studying the architecture of the Deep Learning model based on "SimpleRNN"

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

A confusion matrix was obtained to see how close the intervals are confused with each other, Figure 7. Nearby regions such as "20 m" and "25 m" or "40 m" and "45 m" showed severe confusion in the "SimpleRNN" based Deep Learning model from 15m to 50m intervals.

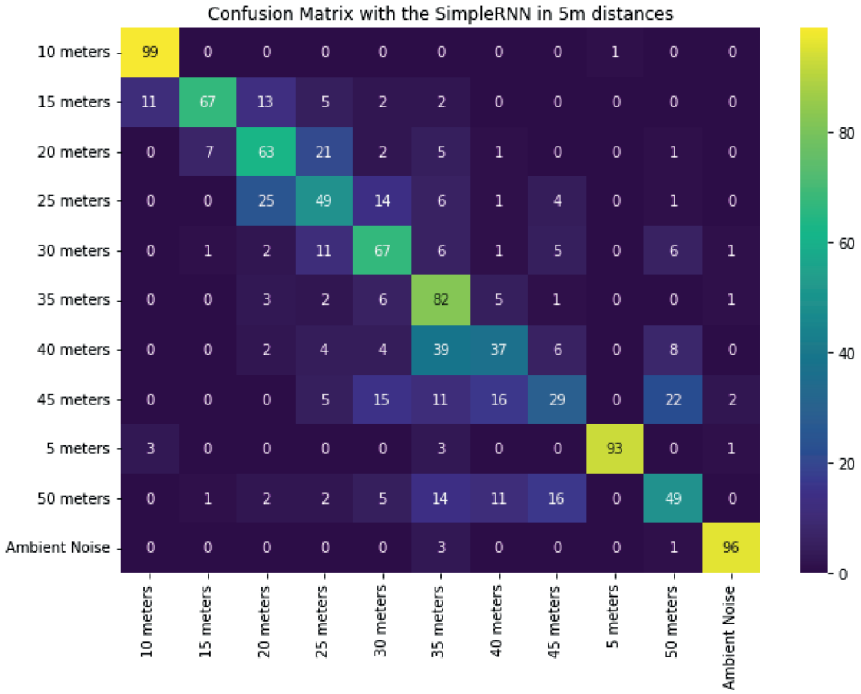


Figure 7. Confusion matrix when studying the architecture of a Deep Learning model based on "Simple RNN" for drone sounds every 5 meters

In the next step, these experimental sequences were performed with the "LSTM" network. Its classification reports can be seen in Figure 8. Here, it can be seen that the recognition indicators are higher than the indicators of the "Simple RNN" network.

	precision	recall	f1-score	support
10 meters	0.97	0.95	0.96	100
15 meters	0.89	0.84	0.87	100
20 meters	0.76	0.68	0.72	100
25 meters	0.63	0.62	0.62	100
30 meters	0.68	0.73	0.71	100
35 meters	0.70	0.88	0.78	100
40 meters	0.72	0.66	0.69	100
45 meters	0.66	0.60	0.63	100
5 meters	0.96	0.96	0.96	100
50 meters	0.66	0.63	0.65	100
Ambient Noise	0.89	0.97	0.93	100
accuracy			0.77	1100
macro avg	0.77	0.77	0.77	1100
weighted avg	0.77	0.77	0.77	1100

Figure 8. Recognition results of classification report when studying the architecture of Deep Learning model based on "LSTM" for UAV sounds at a distance of 5 meters

It can be seen that the confusion matrix shows that the confusion intervals are slightly improved compared to the deep learning model based on "SimpleRNN", Figure 9.

DOI: 10.37943/19XNOV6347

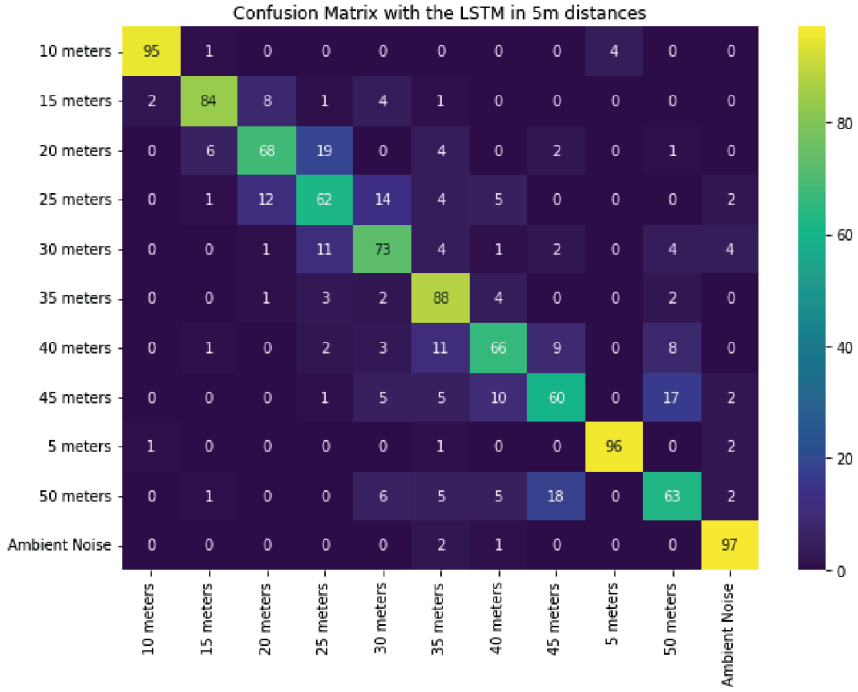


Figure 9. Confusion matrix when studying the architecture of a Deep Learning model based on "LSTM" for UAV sounds every 5 meters

The next step was carried out with the "GRU" network. The recognition accuracies of this network were found to be comparable in classification reports, Figure 10 and Confusion matrix Reports, Figure 11 .

	precision	recall	f1-score	support
10 meters	0.93	0.99	0.96	100
15 meters	0.81	0.86	0.83	100
20 meters	0.67	0.56	0.61	100
25 meters	0.62	0.51	0.56	100
30 meters	0.66	0.78	0.71	100
35 meters	0.70	0.78	0.74	100
40 meters	0.62	0.56	0.59	100
45 meters	0.65	0.60	0.62	100
5 meters	1.00	0.96	0.98	100
50 meters	0.58	0.70	0.63	100
Ambient Noise	0.98	0.90	0.94	100
accuracy			0.75	1100
macro avg	0.75	0.75	0.74	1100
weighted avg	0.75	0.75	0.74	1100

Figure 10. Classification report for recognition results of a UAV sounds at a distance of 5 meters when studying the architecture of the Deep Learning model based on "GRU"

Confusion is reduced for some classes. And for some classes it has increased. This may also be due to the informational data of the sounds.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

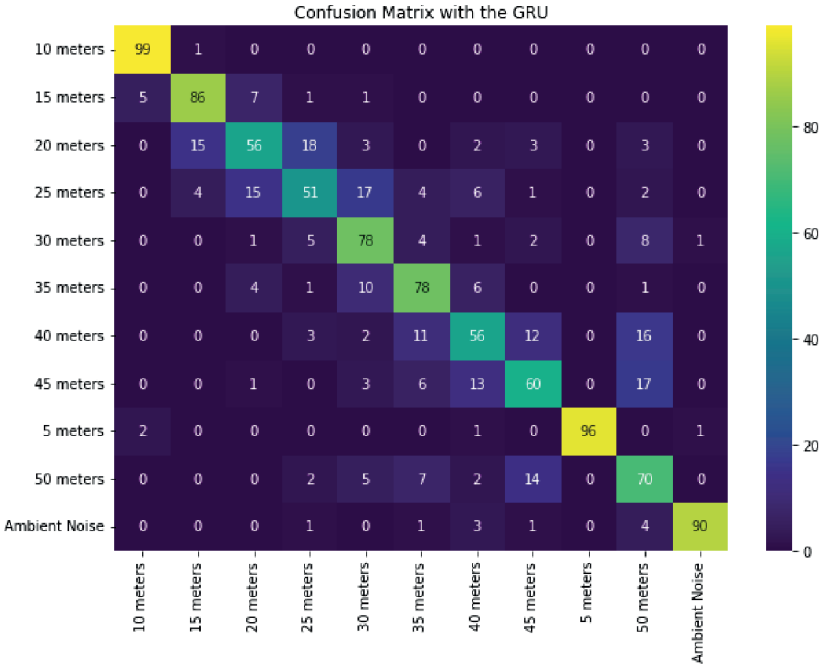


Figure 11. Confusion matrix when exploring the architecture of a Deep Learning model based on "GRU" for UAV sounds at a distance of every 5 meters

Overall, LSTM and GRU networks showed similar recognition capabilities in these three experiments. However, the "GRU" network has shown that it works flexibly with this 1st database with a recognition curve. And the next series of experiments was carried out with the new database in Table 2b. The reason for that was the insensitivity of the proposed system for distances of 5 meters. Therefore, further experiments were carried out to test the sensitivity to distances of 10 meters, Fig. 12. This time, "GRU" was able to show that the framework of deep learning has the ability to learn flexibly with the database.

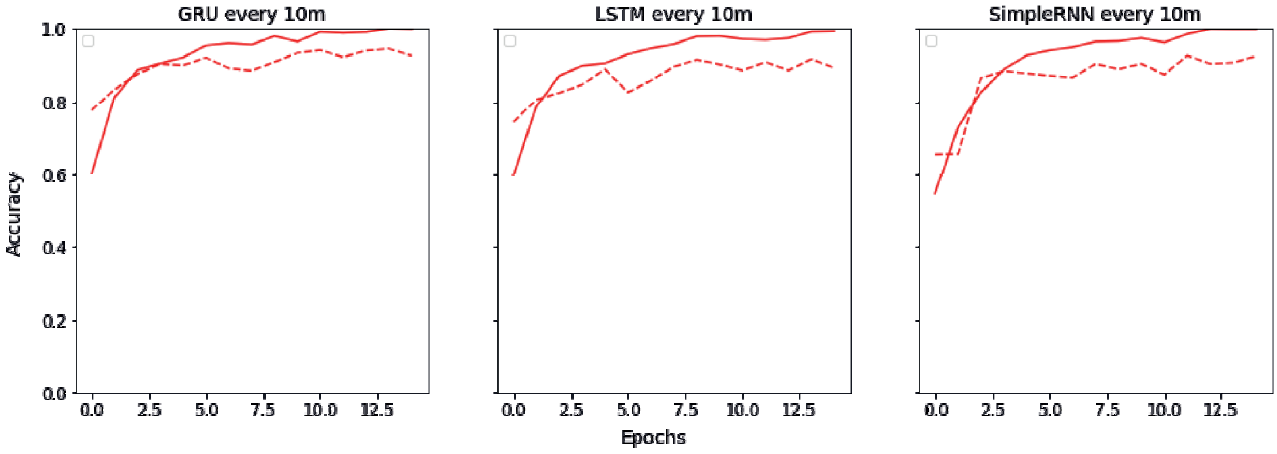


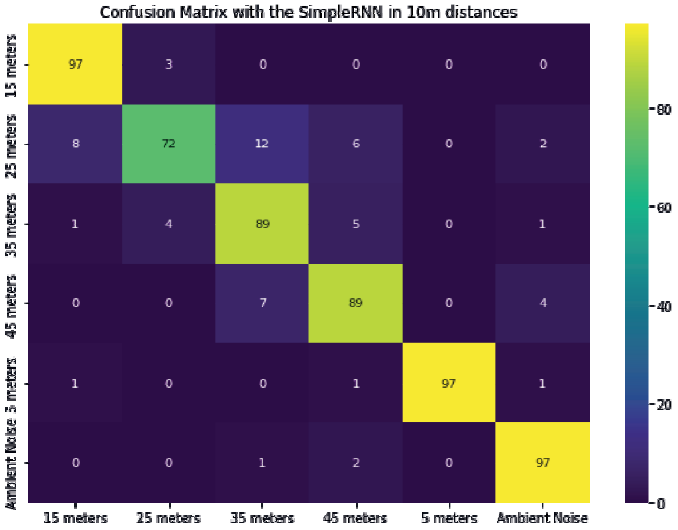
Figure 12. Recognition curves for studying the architectures of Deep Learning models based on "Simple RNN", "LSTM" and "GRU" for UAV sounds at a distance of 10 meters

As above, the "SimpleRNN" model was first studied. In this research step, both the average recognition accuracy and the ability to recognize other individual classes increased. That is, it can be observed that the probability of predicting sounds is high with this system.

DOI: 10.37943/19XNOV6347

	precision	recall	f1-score	support
15 meters	0.91	0.97	0.94	100
25 meters	0.91	0.72	0.80	100
35 meters	0.82	0.89	0.85	100
45 meters	0.86	0.89	0.88	100
5 meters	1.00	0.97	0.98	100
Ambient Noise	0.92	0.97	0.95	100
accuracy			0.90	600
macro avg	0.90	0.90	0.90	600
weighted avg	0.90	0.90	0.90	600

a)



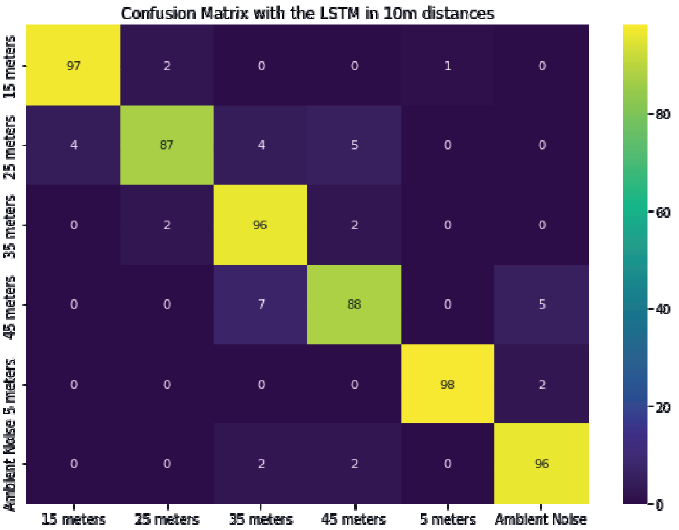
b)

Figure 13. Classification reports and confusion matrix when exploring the architecture of the Deep Learning model based on "SimpleRNN" for drone sounds at a distance of 10 meters

The results of classification of models based on "LSTM" and "GRU" are shown in Figures 14-15. Based on the results, distance from the sensor was tried to estimate the average recognition ability.

	precision	recall	f1-score	support
15 meters	0.96	0.97	0.97	100
25 meters	0.96	0.87	0.91	100
35 meters	0.88	0.96	0.92	100
45 meters	0.91	0.88	0.89	100
5 meters	0.99	0.98	0.98	100
Ambient Noise	0.93	0.96	0.95	100
accuracy			0.94	600
macro avg	0.94	0.94	0.94	600
weighted avg	0.94	0.94	0.94	600

a)



b)

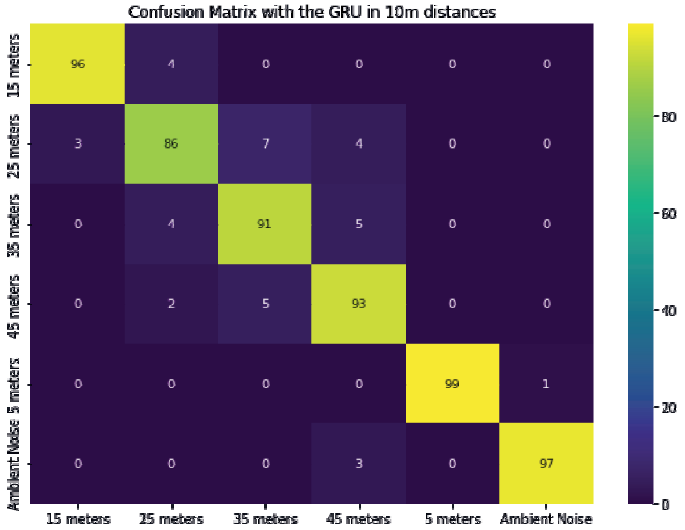
Figure 14. Classification reports and confusion matrix when studying the architecture of Deep Learning model based on "LSTM" for UAV sounds at a distance of 10 meters

Experimental results of the GRU-based deep learning model tested at distances of 10 meters are presented in Figure 15. Average recognition accuracies and correct recognition rates also yielded more reliable recognition results than 5 meters distance dataset.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

	precision	recall	f1-score	support
15 meters	0.97	0.96	0.96	100
25 meters	0.90	0.86	0.88	100
35 meters	0.88	0.91	0.90	100
45 meters	0.89	0.93	0.91	100
5 meters	1.00	0.99	0.99	100
Ambient Noise	0.99	0.97	0.98	100
accuracy			0.94	600
macro avg	0.94	0.94	0.94	600
weighted avg	0.94	0.94	0.94	600

a)



b)

Figure 15. Confusion matrix when exploring the architecture of a Deep Learning model based on "GRU" for UAV sounds at a distance of every 10 meters

In general, this work aimed to study the types of recurrent networks from the perspective of experiments on UAV distances sounds. And it can be seen that recurrent network types have been studied in the context of detailed experiments on UAV distance recognition. Not only recognition curves, but also confusion matrices for each class were tested. Along with their average recognition accuracies, false recognition and true recognition metrics were also obtained.

Discussion

So, the results of these studies were tested in two different directions: by extended classification responses and by recognition accuracy curves. When studying SimpleRNN, LSTM and GRU networks, SimpleRNN showed lower recognition performance than LSTM and GRU networks in both directions of this study. And to understand the recognition skills of all networks, a comparison tables was created based on the true-positive recognized results "Recall" only. This tables immediately showed that LSTM and GRU networks have good recognition skills in different metrics. In general, they had comparable and similar recognition abilities in the extended classification responses, Table 3, Table 4.

Table 3. Analysis of "Experiment 1" conducted on the first database

Model	Classes	«SimpleRNN»	«LSTM»	«GRU»
Recall, %	"5 meters"	93	96	96
	"10 meters"	99	95	99
	"15 meters"	67	84	86
	"20 meters"	63	68	56
	"25 meters"	49	62	51
	"30 meters"	67	73	78
	"35 meters"	82	88	78
	"40 meters"	37	66	56
	"45 meters"	29	60	60
	"50 meters"	49	63	70
	"Ambient Noise"	96	97	90

DOI: 10.37943/19XNOV6347

LSTM networks showed good recognition accuracy when there are classes at very close distances such as every 5 meters under noisy circumstances. The reason is that when recording at a distance of 20-35 meters, there were parallel noises such as vehicle driving and students singing on the guitar. That is, here we can assume that LSTM networks showed the skill of reliably recognizing under noisy circumstances. This ability of LSTM is visible through the results of the "Ambient noise" class as well.

Table 4. Analysis of "Experiment 2" conducted on the second database

Model	Classes	«SimpleRNN»	«LSTM»	«GRU»
Recall, %	"5 meters"	97	98	99
	"15 meters"	97	97	96
	"25 meters"	72	87	86
	"35 meters"	89	96	91
	"45 meters"	97	88	93
	"Ambient Noise"	97	96	97

As well, GRU networks showed very good recognition accuracy when UAVs appear at very close or large distances of the observation zone. However, only extended classification reports cannot determine the reliability of the created model. Therefore, the curves of the recognition accuracies were checked for the trained and tested data and the results were compared. The obtained recognition results and curves were able to formulate the GRU-based Deep Learning model with a more reliable performance.

To sum up LSTM and GRU networks showed effective results in general. In addition, GRU networks showed robustness at long and very close ranges and can help in timely detection of UAVs in a protected area when solving a real-time system problem. And LSTM network showed robustness in noisy environments. Therefore, there is motivation to hybridize these networks with each other or with other networks such as CNN or with various intelligent sensor tasks based on late and early fusion in future research. It should be noted here that the sound database is small for database 1 and shows the limitations of this study. To achieve this, further research on hybrid architectures, database expansion and intelligent sensor fusion will be taken into account.

Conclusion

In conclusion, it can be noted that the system of estimating the distance to the UAV using sound signals is possible with sufficient databases. Thus, the study of deep learning models based on Simple RNN, LSTM and GRU for recognizing the sound of UAVs at a distance was considered. As a result, the deep learning model based on GRU and LSTM gave more reliable results from different angles. It was also noted that UAVs can recognize and predict distances of more than ten meters with high accuracy. In this paper, an attempt was made to experimentally study deep learning models based on Single-layer Simple RNN, LSTM and GRU. And the study of Simple RNN, LSTM and GRU networks allows us to navigate the problems of recognizing the sounds of other objects.

Acknowledgement

The research was funded by the Scientific Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant IRN AP14971907, "Development of a robust frequency-based detection system for suspicious UAVs using SDR and acoustic signatures").

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

References

[1] Taha B. and Shoufan A. (2019). Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research. *IEEE Access*, vol. 7, pp. 138669-138682, doi: https://doi.org/10.1109/ACCESS.2019.2942944.

[2] First drone crash with a commercial aircraft in Canada triggers safety review and possible new rules. Available at: https://www.ediweekly.com/first-drone-crash-commercial-aircraft-canada-triggers-safety-review-possible-new-rules/

[3] Patrick H. Hundreds of drones crash after glitching during show in China. (2023). Available at: https://www.independent.co.uk/tv/lifestyle/china-drone-crash-zoo-show-b2394312.html, Wednesday 16 August.

[4] Kosenov A. Kazakhstan podtverdil proniknoveniye uzbekskogo bespilotnika na svoyu territoriyu. (2012). Available at: https://tengrinews.kz/events/kazahstan-podtverdil-proniknovenie-uzbekskogo-bespilotnika-208687/.

[5] Seidaliyeva, U.; Ilipbayeva, L.; Taissariyeva, K.; Smailov, N.; Matson, E.T. (2024). Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review. *Sensors*, 24, 125. https://doi.org/10.3390/s24010125

[6] Ilipbayeva L.B., Seydaliyeva U.O., Smaylov N.K., Matson E.T. (2024). Research of UAV detection using modified yoloalgorithm. Vestnik Almatinskogo universiteta energetiki i svyazi No 2(65) https://doi.org/10.51775/2790-0886_2024_65_2_179

[7] Zhanbirova A. (2024). UAV crashes near airport in Kyrgyzstan. Available at: https://kz.kursiv.media/en/2024-08-15/uav-crashes-near-airport-in-kyrgyzstan/ (accessed on August 15, 2024 21:41)

[8] Utebayeva D. and Yembergenova A. (2024). Study a deep learning-based audio classification for detecting the distance of UAV. *IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, Madrid, Spain, 2024, pp. 1-7, https://doi.org/10.1109/EAIS58494.2024.10569107.

[9] Mkrtchian G. and Furletov Y. (2022). Classification of Environmental Sounds Using Neural Networks. *Systems of Signal Synchronization, Generating and Processing in Telecommunications* (SYNCHROINFO), Arkhangelsk, Russian Federation, pp. 1-4, http://dx.doi.org/10.1109/SYNCHROINFO55067.2022.9840922.

[10] Momynkulov Z., Omarov N. and Altayeva A. (2024) CNN-RNN Hybrid Model For Dangerous Sound Detection in Urban Area. *IEEE 4th International Conference on Smart Information Systems and Technologies (SIST)*, Astana, Kazakhstan, pp. 284-289, http://dx.doi.org/10.1109/SIST61555.2024.10629358.

[11] Babu K.A. and Ramkumar B. (2020). Automatic Recognition of Fundamental Heart Sound Segments From PCG Corrupted With Lung Sounds and Speech," in IEEE Access, vol. 8, pp. 179983-179994, https://doi.org/10.1109/ACCESS.2020.3023044.

[12] Naveen Sundar G., Subramanian S., Narmadha D., Malin Bruntha P., I. Thanakumar Joseph S and S.S. (2024). Improved Heart Sound Classification Using LSTM Based Deep Learning Technique. *5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Tirunelveli, India, pp. 557-561, http://dx.doi.org/10.1109/ICICV62344.2024.00094.

[13] Bubashait M. and Hewahi N. (2021). Urban Sound Classification Using DNN, CNN & LSTM a Comparative Approach. *International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Zallaq, Bahrain, 2021, pp. 46-50, https://doi.org/10.1109/3ICT53449.2021.9581339.

[14] Hayashi T., Watanabe S., Toda T., Hori T., Le Roux J. and Takeda K. (2017). Duration-Controlled LSTM for Polyphonic Sound Event Detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 11, pp. 2059-2070, Nov., https://doi.org/10.1109/TASLP.2017.2740002.

[15] Liu J. et al. (2018). Bowel Sound Detection Based on MFCC Feature and LSTM Neural Network. *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Cleveland, OH, USA, pp. 1-4, doi: https://doi.org/10.1109/BIOCAS.2018.8584723.

[16] Huang Z., Tang J., Xue S. and Dai L. (2016). Speaker adaptation OF RNN-BLSTM for speech recognition based on speaker code. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, pp. 5305-5309, https://doi.org/10.1109/ICASSP.2016.7472690.

[17] Hwang K. and Sung W. (2016). Character-level incremental speech recognition with recurrent neural networks. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 5335-5339, doi: https://doi.org/10.1109/ICASSP.2016.7472696.
[18] Lotfidereshgi R. and Gournay P. (2018). Speech Prediction Using an Adaptive Recurrent Neural Network with Application to Packet Loss Concealment. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Canada, pp. 5394-5398, https://doi.org/10.1109/ICASSP.2018.8462185.
[19] Momynkulov Z., Omarov N. and Uxikbayev Y. (2024). Detection of Dangerous Situations by Sounds in Real-Time Using Deep Learning. *IEEE 4th International Conference on Smart Information Systems and Technologies (SIST)*, Astana, Kazakhstan, pp. 278-283, http://dx.doi.org/10.1109/SIST61555.2024.10629572.
[20] Jose T. and Mayan J. A. (2023). Real-Time Sound Detection of Rose-Ringed Parakeet Using LSTM Network with MFCC and Mel Spectrogram. *Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*, Kanjirapally, India, pp. 1-6, https://doi.org/10.1109/AICERA/ICIS59538.2023.10420143.
[21] Elghamrawy S. M. and Edin Ibrahim S. (2021). Audio Signal Processing and Musical Instrument Detection using Deep Learning Techniques. *9th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*, Alexandria, Egypt, pp. 146-149, https://doi.org/10.1109/JAC-ECC54461.2021.9691427.
[22] Kamepalli S., Rao B. S. and Venkata Krishna Kishore K. (2022). Multi-Class Classification and Prediction of Heart Sounds Using Stacked LSTM to Detect Heart Sound Abnormalities. *3rd International Conference for Emerging Technology (INCET)*, Belgaum, India, pp. 1-6, https://doi.org/10.1109/INCET54531.2022.9825189.
[23] Dosbayev, Z. et al. (2021). Audio Surveillance: Detection of Audio-Based Emergency Situations. *In: Wojtkiewicz, K., Treur, J., Pimenidis, E., Maleszka, M. (eds) Advances in Computational Collective Intelligence*. ICCCI. Communications in Computer and Information Science, vol 1463. Springer, Cham. https://doi.org/10.1007/978-3-030-88113-9_33
[24] Sajad S., Dharshika S. and Meleet S. (2021). Music Generation for Novices Using Recurrent Neural Network (RNN). *International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Chennai, India, pp. 1-6, doi: 10.1109/ICSES52305.2021.9633906.
[25] Yang B., Matson E.T., Smith A.H., Dietz J.E. and Gallagher J.C. (2019). UAV Detection System with Multiple Acoustic Nodes Using Machine Learning Models. Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, pp. 493-498, doi: 10.1109/IRC.2019.00103.
[26] Dumitrescu, C.; Minea, M.; Costea, I.M.; Cosmin Chiva, I.; Semenescu, A. (2020). Development of an Acoustic System for UAV Detection. Sensors, 20, 4870. https://doi.org/10.3390/s20174870
[27] Wang Y., Fagian Y., Ho K.E. and Matson E.T. (2021). A Feature Engineering Focused System for Acoustic UAV Detection. *Fifth IEEE International Conference on Robotic Computing (IRC)*, Taichung, Taiwan, pp. 125-130, doi: 10.1109/IRC52146.2021.00031.
[28] Didkovskyi V., Kozeruk S. and Korzhik O. (2019). Simple Acoustic Array for Small UAV Detection. *IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, Ukraine, pp. 656-659, doi: 10.1109/ELNANO.2019.8783262.
[29] Jeon S., Shin J. -W., Lee Y. -J., Kim W. -H., Kwon Y. and Yang Y. (2017). Empirical study of drone sound detection in real-life environment with deep neural networks. *25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, pp. 1858-1862, doi: 10.23919/EUSIPCO.2017.8081531.
[30] Ku I., Roh S., Kim G., Taylor C., Wang C. and Matson E. T. (2022). UAV Payload Detection Using Deep Learning and Data Augmentation. *Sixth IEEE International Conference on Robotic Computing (IRC)*, Italy, pp. 18-25, doi: 10.1109/IRC55401.2022.00009.
[31] Katta S.S., Nandyala S., Viegas S. and AlMahmoud A. (2022). Benchmarking Audio-based Deep Learning Models for Detection and Identification of Unmanned Aerial Vehicles. *Workshop on Benchmarking Cyber-Physical Systems and Internet of Things (CPS-IoTBench)*, Milan, Italy, pp. 7-11, doi: 10.1109/CPS-IoTBench56135.2022.00008.
[32] Information from the Internet [mavic.kz] – Available at: https://mavic.kz/product/dron-dji-mini-2-fly-more-combo/