**SUPPLEMENTARY ONLINE MATERIAL:**

**Improving the automatic target recognition algorithm’s accuracy through an examination of the different time-frequency representations and data augmentation**

**S-I. Introduction**

This online supplementary material contains some additional information for a better understanding of experiments and results in the research study. Therefore, images and tables that we could not include in the manuscript are shown here.

The problem of the automatic target recognition (ATR) algorithm of ground-moving targets is a very significant subject in many scientific research papers. In the publicly available research reports, different approaches can be found and several datasets containing data from all types of reconnaissance radars such as pulsed-Doppler radar, range-Doppler radar, or synthetic aperture radar (SAR). Additionally, it is interesting to note that state-of-the-art ATR algorithms usually involve neural networks to solve such problems.

However, the main focus of this research paper was to investigate and improve the ATR algorithms that can be applied to pulsed-Doppler radar. The proposed ATR algorithm includes the short-time Fourier transform (STFT) for the time-frequency representation (TFR) of a radio signal, with the signal segmentation and data augmentation processes, and with the AlexNet model for detection and identification purposes. The results are obtained with the RadEch dataset introduced in [1] from various ground-moving targets. In addition, the proposed algorithm was tested on a separate dataset namely the SPL RadEch dataset described in [2]. The improvement of the proposed algorithm was verified by a series of experiments:

* the ablative experiment to study the effect of TFRs,
* the data augmentation and the decision-making time experiment to study the accuracy,
* the experiment on the discriminative power of the proposed algorithm and
* the influence of the deep learning (DL) algorithm to find the best model.

The important results of experiments have been shown in the comparative analysis in the manuscript, while in this supplementary online material all obtained results are present.

**S-II. The RadEch dataset description**

The RadEch dataset was obtained using records of a short-range ground surveillance radar that operates at the frequency of 16.8 GHz (Ku-band). Because of that, the Doppler frequencies are within the audio band (0 to 2 kHz) and can be reproduced to the operator via headphones or speaker. Therefore, the RF signals were recorded on a computer in \*.wav format with a sampling frequency of 4 kHz and after that saved in the MatLab in \*.mat format.

The ground surveillance radar is a coherent and pulsed-Doppler with an average power of 5 mW, a pulse width of 15 μs, a range resolution of 150 m, and an azimuth beamwidth of 5 degrees. The RadEch dataset contains radar echoes from various targets, which are recorded and published by the Military Academy, the University of Defence in Belgrade, Serbia [1].

The detailed specification of the RadEch dataset is presented in Table S1. It is important to note that the representation and labeling techniques used for this purpose are based on Binary Unique Identification (BUId).

Table S1. The RadEch dataset specification.

| **Class name (Target)** | **Subclass name** | **BUId** | **Number of class records** | **[%]** | **Number of subclass records** | **[%]** |
| --- | --- | --- | --- | --- | --- | --- |
| - | Clutter | 0000 | 18 | 3.97 | 18 | 3.97 |
| One-Person (OP) | OP crawling | 0100 | 188 | 41.50 | 18 | 3.97 |
| OP running | 0101 | 71 | 15.67 |
| OP walking | 0110 | 99 | 21.85 |
| Group of persons (GoP) | GoP running | 1000 | 174 | 38.41 | 50 | 11.04 |
| GoP walking | 1001 | 124 | 27.37 |
| Vehicle | Truck | 1100 | 73 | 16.11 | 47 | 10.38 |
| Wheeled | 1101 | 26 | 5.74 |

First of all, it must be pointed out that the first column of Table S1 shows three classes (ground-moving targets). Furthermore, the second column shows eight subclasses characterizing different states of corresponding classes and clutter. Next, the BUIds are specified in the third column. Finally, the remaining columns, contain information about the number of recordings and the percentage of classes/subclasses in the complete dataset.

The following scenarios were recorded within the RadEch dataset:

* Person and Group of Persons:
  + the number of persons: one, three, or more (pedestrian, soldier, group of persons, group of soldiers);
  + type of target’s motion: crawling, normal walking, and running (synchronous/asynchronous motion of persons in a group);
  + the direction of motion: go away from the radar (otherwise toward the radar);
  + angles of motion toward the radar: 0˚;
* Wheeled and Tracked vehicle:
  + type of vehicle: all-terrain wheeled vehicle (jeep) or truck;
  + speed of motion: normal (20 to 30 km/h), and fast (30 to 60 km/h);
  + angles of motion toward the radar: 0˚;
* Vegetation clutter (trees, bushes, and high grass).

It is worth mentioning that, continuous target echo records were acquired from the moment radar detects the target and during automatically tracking of the target. In order to gain a good signal-to-noise ratio (SNR), the range between the radar and the target was set to 100 up to 1000m. Only one target, which motion was fully controlled, was engaged per recording (one target per scenario). Each scenario was recorded several times (80 different scenarios of minimum 20 seconds were recorded), thus a total of 453 records of 4 seconds can be found in the RadEch dataset. Although the moving targets were within the line-of-sight, all recordings were performed in the presence of ground clutter with vegetation, and without any interference.

**S-III. The SPL RadEch dataset description**

The SPL RadEch was developed in the Signal Processing Laboratory (SPL) at Israel’s Ben Gurion University. The SPL RadEch dataset contains military (person, group of person, vehicle) and non-military (animal and clutter) targets under different conditions of appearance (different angles of target’s motion toward the radar, speed of target, target’s locomotion without hand motions, straight/zigzag target’s locomotion, etc.). The recordings were acquired with a ground surveillance pulse-Doppler radar operating in the X-band (9 GHz) frequency, with a receiver bandwidth of 3 MHz, a pulse width of 12 µs, a range resolution of 125 m, and an azimuth resolution of 4˚. The proposed algorithm was also tested on the SPL RadEch dataset to demonstrate a possible application to other scenarios. The detailed SPL RadEch dataset specification is presented in Table S2.

Table S2. The SPL RadEch dataset specification.

| **Class name (Target)** | **Subclass name** | **BUId** | **Number of class records** | **[%]** | **Number of subclass records** | **[%]** |
| --- | --- | --- | --- | --- | --- | --- |
| Non-military | Clutter[[1]](#footnote-1) | 0000 | - | - | - | - |
| Animal | 0001 | 37 | 7.6 | 37 | 7.6 |
| One-Person  (OP) | Slow walking | 0100 | 243 | 50 | 49 | 10.1 |
| Normal walking | 0101 | 140 | 28.8 |
| Fast walking | 0110 | 22 | 4.5 |
| Running | 0111 | 32 | 6.6 |
| Group of persons  (GoP)[[2]](#footnote-2) | Slow walking | 1000 | 138 | 28.4 | 16 | 3.3 |
| Normal walking | 1001 | 84 | 17.2 |
| Fast walking | 1010 | 27 | 5.6 |
| Running | 1011 | 11 | 2.3 |
| Vehicle - Wheeled | Slow | 10000 | 59 | 12.1 | 36 | 7.4 |
| Normal | 10001 | 8 | 1.6 |
| Fast | 10010 | 15 | 3.1 |
| Vehicle - Tracked | Slow | 10100 | 9 | 1.9 | 9 | 1.9 |

The following scenarios were recorded within the SPL RadEch dataset:

* Non-military Targets:
  + Animal (dog, cow, horse, sheep, pig);
  + Vegetation clutter (trees, bush, wheat);
  + Rain;
  + Fixed location rotating bodies (motors, water sprinklers);
* Military Targets:
  + Person and Group of Persons:
    - speed of motion: slow (2 to 3 km/h), normal (3 to 5 km/h), fast (5 to 8 km/h) walking, and running (8 to 9 km/h);
    - continuous and piece-wise locomotion;
    - locomotion without hand motions;
    - straight/zigzag locomotion;
    - angles of motion toward the radar: 0˚, 15˚, 30˚, 45˚, 60˚;
    - synchronous/asynchronous motion of persons in a group;
  + Wheeled and Tracked vehicle:
    - speed of motion: slow (10 to20 km/h), normal (20 to 30 km/h) and fast (30 to 90 km/h);
    - angles of motion toward the radar: 0˚, 15˚, 30˚, 45˚, 60˚.

Each scenario was recorded at least three times (30 frames of 4 s each). A total of 31 tests were conducted at 21 different locations. In these tests, 14 different people were recorded.

The proposed algorithm has been tested for identification purposes only: 3-class identification and 6-class identification. In the first case, the identification was performed with three classes: “one person”, “wheeled vehicle”, and “tracked vehicle” without modification of the second DL algorithm (output layer consist of three classes). In the second case, the identification was performed with six classes (“one person”, “GoP - two persons”, “GoP - three persons”, “wheeled vehicle”, “tracked vehicle”, and “animal”) with modification of the last layers in the third DL algorithm (output layer was modified to handle six classes instead of eight).

The algorithm proposed in the manuscript successfully solved these classification problems. This is very important as there is very little work to introduce new ATR algorithms and test them on different datasets.

**S-IV. The signal segmentation and data augmentation processes**

The signal segmentation process is performed on every recording from both datasets to divide the whole acquired RF signal into snapshots of data. This process speeded up the TFRs calculation and created more data inputs for the DL algorithm. Moreover, signal segmentation is also important because it influences the decision-making time in the proposed algorithm (4 seconds, 1 second, and 0.25 seconds). By dividing the whole RF signal into smaller segments, a different decision-making time can be simulated, so the influence of this phenomenon can be studied accordingly.

Likewise, the data augmentation process is very important for DL algorithms due to the overfitting problem. For this reason, two different approaches were studied, to observe the influence of the data augmentation process on the accuracy of the proposed algorithm.

The first data augmentation process was performed in the time domain on the original signal or signal segment. The second data augmentation process was executed in the spectral domain, i.e. on the image representing the TFR using operations on images. The first type of data augmentation process is intentionally performed on raw data in the time domain, being careful not to disturb the essential features. In this case, every recording was mirrored or modified by a set of circular shifts in the time domain (2, 3, 4, 5, 6, 10, 20, 30, 40, 50, 100, 150, 200, 250, and 300 samples). The second type of data augmentation process is based on image manipulation and is performed on an obtained TFR.

In both cases, when the original radar signals (without data augmentation) are counted, the data augmentation methods increase the amount of input data by 17-fold and improve the flow of training without the occurrence of overfitting. The RadEch dataset specification after signal segmentation and data augmentation is shown in Table S3.

Table S3. The RadEch dataset specification after signal segmentation and data augmentation.

| **Signal segmentation** | **Data augmentation** | **Overall augmentation** | **Overall dataset dimensions** |
| --- | --- | --- | --- |
| 1x (without) | 17x | 17x | 453\*17= 7,701 records |
| 4x | 17x | 68x | 453\*4\*17=30,804 records |
| 16x | 17x | 272x | 453\*16\*17=123,216 records |

The first type of data augmentation (time-domain augmentation with 15 different circular shifts) can be time demanding operation, therefore we examined a simplified time domain data augmentation based on only a subset of circular shifts (2, 10, 50, 100, and 300 samples that are shifted). In this case, overall augmentation without signal segmentation increases the amount of input data by 7-fold, so the dataset contains records. Furthermore, overall augmentation with 4x signal segmentation increases the amount of input data by 28-fold, so the dataset contains records. Finally, overall augmentation with 16x signal segmentation increases the amount of input data by 112-fold, so the dataset contains records. The data amount reduction in the case of the simplified time domain data augmentation is approximately 2.43 times.

**S-V. The time-frequency representation of signals**

Three TFRs were used in this research: short-time Fourier transform (STFT), continuous wavelet transform (CWT), and the Wigner-Ville distribution (WVD). The examples of used TFRs of all classes have been presented in Figures S1 to S3.

Radar target echo spectrograms obtained with STFT are presented in Figure S1 for the entire record (the length of the radar signal is four seconds, i.e., signal segmentation is not engaged).

|  |  |
| --- | --- |
|  |  |
| a) | b) |
|  |  |
| c) | d) |
| Figure S1. Spectrograms of clutter and 3 ground-moving targets from the RadEch dataset obtained by STFT: a) clutter, b) one person, c) group of persons, and d) vehicle. | |

Radar target echo scalograms obtained with CWT are presented in Figure S2 for the entire record (the length of the radar signal is four seconds, i.e., signal segmentation is not engaged).

|  |  |
| --- | --- |
|  |  |
| a) | b) |
|  |  |
| c) | d) |
| Figure S2. Scalograms of clutter and 3 ground-moving targets from the RadEch dataset obtained by CWT: a) clutter, b) one person, c) group of persons, and d) vehicle. | |

It is important to note that the default Morse analytic wavelet was used to compute the CWT. Two basic parameters can be adjusted in the Morse analytic wavelet: the symmetry parameter () and time-bandwidth product (). The Morse symmetry parameter (), controls the symmetry of the wavelet in time through the skewness, while the square root of the time-bandwidth product () is proportional to the wavelet duration in time. Scalograms in Figure S2 were obtained with the MatLab function CWT which gives the Morse wavelets in the function of the time-bandwidth product () and the symmetry parameter (). It is possible to obtain wavelets with different characteristics by adjusting the time-bandwidth product and symmetry parameters of a Morse wavelet. However, the default values were used for the time-bandwidth product () and gamma () since the authors in [3] presented that the Morse wavelet has a minimum Heisenberg area and the skewness equal to zero.

Radar target echo WVDs are presented in Figure S3 for the entire record (the length of the radar signal is four seconds, i.e., signal segmentation is not engaged).

|  |  |
| --- | --- |
|  |  |
| a) | b) |
|  |  |
| c) | d) |
| Figure S3. WVD of clutter and 3 ground-moving targets from the RadEch dataset: a) clutter, b) one person, c) group of persons, and d) vehicle. | |

It is worth saying that the spectrogram and scalogram belong to the same group of TFRs which can be represented with joint energy densities. In the case of the spectrogram, the STFT is employed as a signal transform, together with the calculation of the squared magnitude. In the case of the scalogram, first, the computation of wavelet transform is employed followed by the calculation of the squared magnitude. The result is a non-negative energy density which is one of the key requirements that a joint energy density should satisfy.

The WVD takes a different approach by calculating a local auto-covariance function followed by the Fourier transform so there are no special parameters for calculation like in the previous cases (STFT and CWT). The good characteristic of WVD is that the concept of windowing operation (the window type and length) is not to be applied, because it is already built into the calculation of the auto-covariance function. Another good behavior of WVD is the perfect localization for chirp and modulated chirp signals, which can be observed in Figure S3 (d). The WVD does not have major smearing of the signal energy, in contrast to the CWT, presented in Figures S2 (b) and (d). However, the WVD satisfied the key requirements of a joint energy density but is not guaranteed to be non-negative for all signals in the entire time-frequency plane.

It can be observed that each TFR obtained with STFT, CWT, and WVD has unique properties (the Doppler frequency shift and visible oscillations around the central Doppler frequency).

**S-VI. Convolutional neural network models**

Three convolutional neural network models were used in this research: the AlexNet model, the simplified CNN model, and the CRNN model. These three CNN models were intentionally used in this research in order to test their application to the introduced classification problem with short-range ground surveillance pulsed-Doppler radar. It should be noted that simplified CNN and CRNN models are from our laboratory “Anti Drone Deep Learning (ADRO-DL)”. These CNN models were introduced in [4], and [5] and tested for the classification problem of drones in the radio frequency domain.

**S-VII. Confusion matrix explanation**

To better understand the performance of the proposed ATR algorithm, an example of a confusion matrix for two classes is shown in Figure S4 with an explanation of the corresponding rows, columns, and cells. It is important to emphasize that true predictions (green cells in the confusion matrix) represent the sum of TP and TN, while false predictions (red cells in the confusion matrix) represent the sum of FP and FN results.

|  |
| --- |
|  |
| Figure S4. Resultant rows, columns, and cells for confusion matrix with an explanation. |

The corresponding performance metrics used in this study within confusion matrices are defined with equations (1) to (8) [6]:

* Accuracy (ACC) measures the proportion of correctly predicted samples. A larger value indicates better prediction accuracy. ACC is calculated as the number of all correct predictions (TP and TN) divided by the total number of samples in the dataset.

|  |  |
| --- | --- |
|  | (1) |

* Error rate (ERR) is calculated as the number of all incorrectly predicted samples (FN and FP) divided by the total number of samples in the dataset.

|  |  |
| --- | --- |
|  | (2) |

or

|  |  |
| --- | --- |
|  | (3) |

* Precision (PREC) measures the proportion of samples that are positive among samples predicted to be positive. A larger value indicates better prediction accuracy. It is also called the positive predictive value.

|  |  |
| --- | --- |
|  | (4) |

* False Discovery Rate (FDR) measures the expected proportion of Type I error. FDR controls the expected proportion of discovered samples that are false.

|  |  |
| --- | --- |
|  | (5) |

* Sensitivity (SN) measures the relation of actual positive samples that are predicted as positive. SN is calculated as the number of correct positive predictions divided by the total number of positives. It is also called the true positive rate.

|  |  |
| --- | --- |
|  | (6) |

* False Negative Rate (FNR) measures the rate of incorrectly identified tests, false negative and true positive samples.

|  |  |
| --- | --- |
|  | (7) |

* F1 score (F1SCR) is a measure of a model’s accuracy on a dataset. It is calculated like the harmonic mean of the model’s PREC and SN.

|  |  |
| --- | --- |
|  | (8) |

**S-VIII. The RadEch dataset detection and identification results**

Figures S5, S6, and S7 present the results of the performance assessment of the proposed ATR algorithm (the STFT-TFR method with the AlexNet model). This is convenient because it is easy to compare the results of detection and identification of ground-moving target for all scenarios. Figure S5 shows the performance of the first AlexNet model for detecting the presence of a ground-moving target. The corresponding labels used for this model are “1” for the target absence class (clutter) and “2” for the target presence class.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| a) | b) | c) |
|  |  |  |
| d) | e) | f) |
| Figure S5. Confusion matrices for the detection of the presence of a target: a) without data augmentation and signal segmentation, b) without data augmentation and with signal segmentation (length of segment = 4,000 samples), c) without data augmentation and with signal segmentation (length of segment = 1,000 samples), d) with data augmentation and without signal segmentation, e) with data augmentation and with signal segmentation (length of segment = 4,000 samples), and f) with data augmentation and with signal segmentation (length of segment = 1,000 samples). | | |

To properly interpret the results presented through the confusion matrices, it is important to emphasize the following facts: 20% of the total number of class samples was used for testing purposes, and the RadEch dataset is divided into K = 5 stratified folds, taking into account that each overlap represents the corresponding representative of the source data. Accordingly, in Figure S5 (a) a total of 91 samples were used to test the first AlexNet model when there was no augmentation. In that way, there is a total of samples in the confusion matrix (the RadEch dataset contains 453 real echo samples). This principle is followed to obtain all confusion matrices for all AlexNet models.

Figure S6 shows the performance of the second AlexNet model for the identification of ground-moving target classes. The corresponding labels used for this model are “1” - OP, “2” - GoP, and “3” - Vehicle.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| a) | b) | c) |
|  |  |  |
| d) | e) | f) |
| Figure S6. Confusion matrices for the identification of the class of a target: a) without data augmentation and signal segmentation, b) without data augmentation and with signal segmentation (length of segment = 4,000 samples), c) without data augmentation and with signal segmentation (length of segment = 1,000 samples), d) with data augmentation and without signal segmentation, e) with data augmentation and with signal segmentation (length of segment = 4,000 samples), and f) with data augmentation and with signal segmentation (length of segment = 1,000 samples). | | |

Figure S7 shows the performance of the third AlexNet model for the identification of ground-moving target subclasses. The corresponding labels used for this model are “1” - Clutter, “2” - OP crawling, “3” - OP running, “4” - OP walking, “5” - GoP running, “6” - GoP walking, “7” - Truck and “8” – Wheeled vehicle.

|  |  |
| --- | --- |
|  |  |
| a) | d) |
|  |  |
| b) | e) |
|  |  |
| c) | f) |
| Figure S7. Confusion matrices for the identification of the subclass of a target: a) without data augmentation and signal segmentation, b) without data augmentation and with signal segmentation (length of segment = 4,000 samples), c) without data augmentation and with signal segmentation (length of segment = 1,000 samples), d) with data augmentation and without signal segmentation, e) with data augmentation and with signal segmentation (length of segment = 4,000 samples), and f) with data augmentation and with signal segmentation (length of segment = 1,000 samples). | |

First of all, Figures S5 (a), (b), and (c) show the performance of the detection of the presence of a ground-moving target when data augmentation is not engaged, while Figures S5 (d), (e), and (f) show the classification performance when data augmentation is engaged. The best result for ground-moving target detection with an average accuracy of 100% is achieved for the decision-making time of 4 seconds (without signal segmentation) and presented in Figures S5 (a) and (d). Furthermore, it is important to emphasize that the detection accuracy is relatively stable with data augmentation (decreased by 0.2%) when the decision-making time is reduced. Based on this, Figure S5 (e) shows an average accuracy of 99.9%, and an average F1 score of 99.4% for signal segments of 4,000 samples, while Figure S5 (f) shows an average accuracy of 99.8%, and average F1 score of 98.5% for signal segments of 1,000 samples.

Secondly, Figures S6 (a), (b), and (c) show the classification performance of the identification of the type of ground-moving target when data augmentation is not engaged, while Figures S6 (d), (e), and (f) show the classification performance when data augmentation is engaged. The best result is shown in Figure S6 (d) for ground-moving target class identification with an average accuracy of 99.9%, and an average F1 score of 99.9% (with the data augmentation and without signal segmentation). Moreover, the classification performance is slightly decreased by 1.3% on augmented shorter segments. This can be observed in Figure S6 (e) where an average accuracy is 99.8%, and an average F1 score is 99.9% for signal segments of 4,000 samples, while Figure S6 (f) shows an average accuracy of 98.6% and average F1 score of 98.9% for signal segments of 1,000 samples.

Finally, Figures S7 (a), (b), and (c) show the classification performance of the identification of the eight types of ground-moving target subclasses, while Figures S7 (d), (e), and (f) show the classification performance when data augmentation is engaged. The best result is shown in Figure S7 (d) for ground-moving target subclass identification with an average accuracy of 100.0% (with the data augmentation and without signal segmentation). Moreover, the classification performance is also slightly decreasing by 1.3% when the data segmentation process is engaged. This can be observed in Figure S7 (e) where an average accuracy is 99.9%, and an average F1 score is 99.8% for signal segments of 4,000 samples, while Figure S7 (f) shows an average accuracy of 98.7% and average F1 score of 98.8% for signal segments of 1,000 samples.

**S-IX. The comparative analysis**

**TFR’s dependencies.** The first comparison was done by averaged accuracy considering the used TFRs (STFT, CWT, and WVD). The overall results are presented in Tables S4 to S6.

Table S4. The comparative analysis of the proposed ATR algorithm accuracy from the ablative experiment for 4 seconds of decision-making time without data augmentation.

| **4 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_CNN** | 99.80 | 85.64 | 86.67 |
| **STFT\_AlexNet\*** | ***100.00*** | 96.80 | 95.80 |
| **STFT\_CRNN** | ***100.00*** | ***97.98*** | ***96.95*** |
|  | | | |
| **CWT\_CNN** | 98.02 | 85.38 | 85.43 |
| **CWT\_AlexNet** | 98.20 | *94.70* | 94.50 |
| **CWT\_CRNN** | *99.17* | 93.66 | *95.29* |
|  | | | |
| **WVD\_CNN** | 97.04 | 79.74 | 76.54 |
| **WVD\_AlexNet** | 98.20 | *88.30* | *82.20* |
| **WVD\_CRNN** | *98.70* | 78.93 | 75.82 |

Table S5. The comparative analysis of the proposed ATR algorithm accuracy from the ablative experiment for 1 second decision-making time without data augmentation.

| **1 sec** | **target detection** | **3-class identification** | **8-subclass identification** | |
| --- | --- | --- | --- | --- |
| **STFT\_CNN** | 95.32 | 90.61 | 88.98 | |
| **STFT\_AlexNet\*** | 99.10 | ***96.40*** | 96.00 | |
| **STFT\_CRNN** | ***99.72*** | 95.69 | ***98.62*** | |
|  | | | |
| **CWT\_CNN** | 96.38 | 94.09 | 87.46 | |
| **CWT\_AlexNet** | 99.40 | *95.40* | 94.80 | |
| **CWT\_CRNN** | *99.57* | 94.83 | *95.86* | |
|  | | | |
| **WVD\_CNN** | 96.38 | 79.48 | 75.59 | |
| **WVD\_AlexNet** | 97.50 | 88.50 | 81.80 | |
| **WVD\_CRNN** | *98.81* | *90.87* | *86.33* | |

Table S6. The comparative analysis of the proposed ATR algorithm accuracy from the ablative experiment for 0.25 seconds of decision-making time without data augmentation.

| **0.25 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_CNN** | *98.71* | 89.98 | 88.32 |
| **STFT\_AlexNet\*** | 98.70 | 92.10 | 90.50 |
| **STFT\_CRNN** | 97.01 | *95.70* | *94.25* |
|  | | | |
| **CWT\_CNN** | 97.65 | 87.69 | 87.67 |
| **CWT\_AlexNet** | 98.60 | 91.30 | 87.90 |
| **CWT\_CRNN** | ***99.03*** | ***97.67*** | ***97.05*** |
|  | | | |
| **WVD\_CNN** | 97.15 | 87.34 | 80.50 |
| **WVD\_AlexNet** | *97.60* | 87.30 | 81.40 |
| **WVD\_CRNN** | 95.50 | *95.00* | *92.75* |

Notation (\*) in Tables S4 to S6 represents the results that were obtained from the proposed ATR algorithm (the STFT-TFR method with the AlexNet model). Moreover, it can be noted that CRNN also provides good results with the STFT and CWT TFR methods along with the proposed algorithm. The CRNN model improved the accuracy compared to the CNN model because it used a sequence of two images as an input.

**The decision-making time and the data augmentation process.** The second comparison was done by averaged accuracy considering the used data augmentation process. The overall results are shown in Tables S7 to S9 for different types of data augmentation.

Table S7. The comparative analysis of the proposed ATR algorithm accuracy with the data augmentation for 4 seconds of decision-making time.

| **4 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_AlexNet** **w/o augmentation** | ***100.0*** | 96.8 | 95.80 |
| **STFT \_AlexNet\_A1 (time domain)\*** | ***100.0*** | ***99.9*** | ***100.0*** |
| **STFT \_AlexNet\_A2 (spectral domain)** | 99.6 | 99.2 | 99.3 |
| **STFT \_AlexNet\_A3 (simplified time domain)** | 99.3 | ***99.9*** | 99.4 |

Table S8. The comparative analysis of the proposed ATR algorithm accuracy with the data augmentation for 1 second decision-making time.

| **1 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_AlexNet w/o augmentation** | 99.1 | 96.4 | 96.0 |
| **STFT \_AlexNet\_A1 (time domain)\*** | ***99.9*** | ***99.8*** | ***99.9*** |
| **STFT \_AlexNet\_A2 (spectral domain)** | 99.4 | 95.1 | 96,8 |
| **STFT \_AlexNet\_A3 (simplified time domain)** | 99.5 | 98.6 | 99.5 |

Table S9. The comparative analysis of the proposed ATR algorithm accuracy with the data augmentation for 0.25 seconds of decision-making time.

| **0.25 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_AlexNet w/o augmentation** | 98.7 | 92.1 | 90.5 |
| **STFT \_AlexNet\_A1 (time domain)\*** | ***99.8*** | ***98.6*** | ***98.7*** |
| **STFT \_AlexNet\_A2 (spectral domain)** | 98.4 | 80.6 | 85.8 |
| **STFT \_AlexNet\_A3 (simplified time domain)** | 99.4 | 97.9 | 97.2 |

Notation (\*) in Tables S7 to S9 represents the results that were obtained from the proposed ATR algorithm (the STFT-TFR method with the AlexNet model) with the data augmentation in the time domain with a 17-fold data increase. In addition, notations (A1, A2, and A3) represent the results obtained with the time domain data augmentation, spectral domain data augmentation, and simplified time domain data augmentation, respectively. It is interesting to note that data augmentation in the spectral domain only shows good results for a decision-making time of 4 seconds, while in other circumstances declines.

Moreover, it can be noted that simplified time domain data augmentation provides better results compared to the results obtained with data augmentation in the spectral domain. This is interesting because the data amount in this situation is approximately 2.5 times smaller. Because of this, the simplified data augmentation in the time domain is our second choice.

An analysis of the effect of decision-making intervals on accuracy was investigated. These results are presented in Figure S8. The comparison was done by averaged accuracy considering decision-making time and data augmentation.

|  |
| --- |
|  |
| a) |
|  |
| b) |
| Figure S8. Comparative analysis of the proposed ATR algorithm accuracy: a) without data augmentation, and b) with data augmentation. |

The comparative analysis points to an interesting phenomenon when the data augmentation process in the time domain is employed. The accuracy for the 3 classes identification is slightly worse compared to the 8-subclasses identification presented in Figure S8 (b), although the rational assumption is usually in favor of the smaller number of classes. Moreover, this can be noted in Figure S6 by observing false negative and false positive predictions for “OP” and “GoP” classes. We believe this degradation is due to the similarities in Doppler frequency shift and the surrounding oscillations between the named classes, which is consistent with the spectrograms (Figure S1 (b) and Figure S1 (c)). The effect of this ambiguity can also be found in Figure S7 observing false negative and false positive predictions for “OP walking” and “GoP walking”, and “OP running” and “GoP running” classes. Both results are not unexpected and must be interpreted with caution as the RadEch dataset consists of very similar and slow-moving targets in the presence of the clatter.

**DL algorithms.** The third comparison was done by averaged accuracy considering the used DL algorithms. The results are contained in Tables S4 to S6. Moreover, the data augmentation process was tested on different DL algorithms and the results are presented in Tables S10 to S12.

Table S10. The comparative analysis of different DL algorithms accuracy with the time domain data augmentation for 4 second decision-making time.

| **4 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_CNN** | 99.80 | 85.64 | 86.67 |
| **STFT\_CNN\_A1** | *99.94* | *99.87* | *99.97* |
|  | | | |
| **STFT\_AlexNet** | ***100.0*** | 96.8 | 95.8 |
| **STFT\_AlexNet\_A1** | ***100.0*** | ***99.9*** | ***100.0*** |
|  | | | |
| **STFT\_CRNN** | ***100.0*** | 97.98 | 96.95 |
| **STFT\_CRNN\_A1** | ***100.0*** | *99.73* | *99.56* |

Table S11. The comparative analysis of different DL algorithms accuracy with the time domain data augmentation for 1 second decision-making time.

| **1 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_CNN** | 95.32 | 90.61 | 88.98 |
| **STFT\_CNN\_A1** | *95.80* | *96.17* | *92.96* |
|  | | | |
| **STFT\_AlexNet** | 99.10 | 96.40 | 96.00 |
| **STFT\_AlexNet\_A1** | ***99.90*** | ***99.80*** | ***99.90*** |
|  | | | |
| **STFT\_CRNN** | 99.72 | 95.69 | 98.62 |
| **STFT\_CRNN\_A1** | ***99.90*** | *99.75* | *99.44* |

Table S12. The comparative analysis of different DL algorithms accuracy with the time domain data augmentation for 0.25 seconds of decision-making time.

| **0.25 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_CNN** | 98.71 | 89.98 | 88.32 |
| **STFT\_CNN\_A1** | *99.51* | *93.98* | *94.51* |
|  | | | |
| **STFT\_AlexNet** | 98.70 | 92.10 | 90.50 |
| **STFT\_AlexNet\_A1** | ***99.80*** | ***98.60*** | ***98.70*** |
|  | | | |
| **STFT\_CRNN** | 97.01 | 95.70 | 94.25 |
| **STFT\_CRNN\_A1** | *99.25* | *98.49* | *97.50* |

Notation A1 in Tables S10 to S12 represents the results that were obtained with the data augmentation in the time domain with a 17-fold data increase.

**S-X. The overall results**

Finally, the overall results are present in Tables S13 to S15 with all measurements in the described experiments. Notation (A1) in Tables S13 to S15 represents the results that were obtained with the data augmentation in the time domain with a 17-fold data increase.

Table S13. The results (accuracy) for 4 seconds of decision-making time.

| **4 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_CNN** | 99.80 | 85.64 | 86.67 |
| **STFT \_CNN\_A1** | 99.94 | 99.87 | 99.97 |
| **STFT\_AlexNet** | ***100.00*** | 96.80 | 95.80 |
| **STFT \_AlexNet\_A1** | ***100.00*** | ***99.90*** | ***100.00*** |
| **STFT\_CRNN** | ***100.00*** | 97.98 | 96.95 |
| **STFT \_CRNN\_A1** | ***100.00*** | 99.73 | 99.56 |
|  | | | |
| **CWT\_CNN** | 98.02 | 85.38 | 85.43 |
| **CWT \_CNN\_A1** | 99.80 | 99.03 | 97.41 |
| **CWT\_AlexNet** | 98.20 | 94.70 | 94.50 |
| **CWT\_AlexNet\_A1** | 99.34 | 99.84 | 99.78 |
| **CWT\_CRNN** | 99.17 | 93.66 | 95.29 |
| **CWT\_CRNN\_A1** | ***100.00*** | ***99.90*** | ***100.00*** |
|  | | | |
| **WVD\_CNN** | 97.04 | 79.74 | 76.54 |
| **WVD\_CNN\_A1** | 99.41 | 97.70 | 96.32 |
| **WVD\_AlexNet** | 98.20 | 88.30 | 82.20 |
| **WVD\_AlexNet\_A1** | 99.78 | 98.35 | 96.34 |
| **WVD\_CRNN** | 98.70 | 78.93 | 75.82 |
| **WVD \_CRNN\_A1** | ***100.00*** | *98.65* | *98.65* |

Table S14. The results (accuracy) for 1 second decision-making time.

| **1 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_CNN** | 95.32 | 90.61 | 88.98 |
| **STFT\_CNN\_A1** | 95.80 | 96.17 | 92.96 |
| **STFT\_AlexNet** | 99.10 | 96.40 | 96.00 |
| **STFT\_AlexNet\_A1** | ***99.90*** | *99.80* | ***99.90*** |
| **STFT\_CRNN** | 99.72 | 95.69 | 98.62 |
| **STFT\_CRNN\_A1** | ***99.90*** | 99.75 | 99.44 |
|  | | | |
| **CWT\_CNN** | 96.38 | 94.09 | 87.46 |
| **CWT\_CNN\_A1** | 99.60 | 95.63 | 97.63 |
| **CWT\_AlexNet** | 99.40 | 95.40 | 94.80 |
| **CWT\_AlexNet\_A1** | 99.78 | 98.46 | 97.82 |
| **CWT\_CRNN** | 99.57 | 94.83 | 95.86 |
| **CWT\_CRNN\_A1** | *99.84* | ***99.90*** | *99.81* |
|  | | | |
| **WVD\_CNN** | 96.38 | 79.48 | 75.59 |
| **WVD\_CNN\_A1** | 98.90 | 92.89 | 92.29 |
| **WVD\_AlexNet** | 97.50 | 88.50 | 81.80 |
| **WVD\_AlexNet\_A1** | 98.63 | 95.09 | 94.57 |
| **WVD\_CRNN** | 98.81 | 90.87 | 86.33 |
| **WVD\_CRNN\_A1** | *99.03* | *98.87* | *99.06* |

Table S15. The results (accuracy) for 0.25 seconds of decision-making time.

| **0.25 sec** | **target detection** | **3-class identification** | **8-subclass identification** |
| --- | --- | --- | --- |
| **STFT\_CNN** | 98.71 | 89.98 | 88.32 |
| **STFT\_CNN\_A1** | 99.51 | 93.98 | 94.51 |
| **STFT\_AlexNet** | 98.70 | 92.10 | 90.50 |
| **STFT\_AlexNet\_A1** | ***99.80*** | ***98.60*** | ***98.70*** |
| **STFT\_CRNN** | 97.01 | 95.70 | 94.25 |
| **STFT\_CRNN\_A1** | 99.25 | 98.49 | 97.50 |
|  | | | | |
| **CWT\_CNN** | 97.65 | 87.69 | 87.67 |
| **CWT\_CNN\_A1** | 98.86 | 91.91 | 92.65 |
| **CWT\_AlexNet** | 98.60 | 91.30 | 87.90 |
| **CWT\_AlexNet\_A1** | 98.65 | 97.51 | 93.61 |
| **CWT\_CRNN** | 99.03 | 97.67 | 97.05 |
| **CWT\_CRNN\_A1** | *99.37* | *98.31* | *97.50* |
|  | | | | |
| **WVD\_CNN** | 97.15 | 85.34 | 80.50 |
| **WVD\_CNN\_A1** | 97.61 | 85.81 | 82.69 |
| **WVD\_AlexNet** | 97.60 | 87.30 | 81.40 |
| **WVD\_AlexNet\_A1** | 97.58 | 95.83 | 92.11 |
| **WVD\_CRNN** | 95.50 | 95.00 | 92.75 |
| **WVD\_CRNN\_A1** | *98.56* | *97.40* | *94.68* |

It is important to note that all results from this research, except those obtained from the data augmentation experiment, are presented in Tables S13 to S15. These results present the accuracy of the employed different TFR methods (STFT, CWT, and WVD) and CNN models (AlexNet, simplified CNN, and CRNN) with data augmentation in the time domain (noted as A1 in corresponding Tables). It has been shown that the proposed algorithm (the STFT-TFR with the AlexNet model) achieves the best results for all decision-making intervals when data augmentation is enabled. The only deviance occurs for 3-class identification when the decision-making time is 1 second (Table S14). In that case, the best accuracy was achieved by using the CWT-TFR with the CRNN model (only a 0.1% better accuracy is recorded). Moreover, the same result as the proposed algorithm was achieved for 3-class identification when the decision-making time is 4 seconds for the CWT-TFR with the CRNN model (Table S13). However, the proposed algorithm outperformed the CWT-TFR with the CRNN model when the decision-making time is only 0.25 seconds (Table S15).

**References**

[1] M. Andrić, B. Bondžulić, and B. Zrnić, “The database of radar echoes from various targets with spectral analysis,” *10th Symp. Neural Netw. Appl. Electr. Eng. NEUREL-2010 - Proc.*, pp. 187–190, 2010, doi: 10.1109/NEUREL.2010.5644074.

[2] I. Bilik, J. Tabrikian, and A. Cohen, “GMM-based target classification for ground surveillance Doppler radar,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 267–278, Jan. 2006, doi: 10.1109/TAES.2006.1603422.

[3] J. M. Lilly and S. C. Olhede, “Higher-order properties of analytic wavelets,” *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 146–160, Jan. 2009, doi: 10.1109/TSP.2008.2007607.

[4] M. Mokhtari, J. Bajcetic, B. Sazdic-Jotic, and B. Pavlovic, “RF-based drone detection and classification system using convolutional neural network,” in *2021 29th Telecommunications Forum (TELFOR)*, Nov. 2021, pp. 1–4. doi: 10.1109/telfor52709.2021.9653332.

[5] M. Mokhtari, J. Bajčetić, B. Sazdić-Jotić, and B. Pavlovic, “Developing a convolutional recurrent neural network for drone classification,” in *Proceedings of the 9th Small Systems Simulation Symposium 2022*, 2022, no. March, pp. 85–89.

[6] M. F. Al-Sa’d, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, “RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database,” *Futur. Gener. Comput. Syst.*, vol. 100, pp. 86–97, Nov. 2019, doi: 10.1016/j.future.2019.05.007.

1. The “Clutter” subclass is not publicly available and was not the subject of our consideration. [↑](#footnote-ref-1)
2. The “Group of persons (GoP)” class is consist of “GoP - two persons” and “GoP - three persons” records which were examined as separate classes. [↑](#footnote-ref-2)