Data Segmentation and Fusion for Classification of Armed Personnel Using Micro-Doppler Signatures

Focante, Edoardo; Martin, Lilian De; Coutino, Mario

DOI
10.1109/RADAR54928.2023.10371198

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the 2023 IEEE International Radar Conference, RADAR 2023

Citation (APA)
Focante, E., Martin, L. D., & Coutino, M. (2023). Data Segmentation and Fusion for Classification of Armed Personnel Using Micro-Doppler Signatures. In *Proceedings of the 2023 IEEE International Radar Conference, RADAR 2023* (Proceedings of the IEEE Radar Conference). IEEE. https://doi.org/10.1109/RADAR54928.2023.10371198

Data Segmentation and Fusion for Classification of Armed Personnel Using Micro-Doppler Signatures

Edoardo Focante†, Lilian de Martín‡, Mario Coutino‡
†*Delft Center for Systems and Control, TU Delft, Delft, The Netherlands*
‡ *Radar Technology Department, TNO, The Hague, The Netherlands*
e.focante@tudelft.nl,{lilian.demartin,mario.coutino}@tno.nl

2023 IEEE International Radar Conference (RADAR) | 978-1-6654-8278-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/RADAR54928.2023.10371198

***Abstract*—In recent years, convolutional neural networks (CNNs) have been increasingly used for classifying radar micro-Doppler signatures of various targets. However, obtaining large amounts of data for efficient CNN training in defence and surveillance scenarios can be challenging. Therefore, designing techniques that maximize the use of available samples is critical. In this paper, we propose an approach built on the hypothesis that certain classes of radar spectrograms, such as those used for discerning armed from unarmed walking personnel, do not have information about the class encoded in the trajectory. Therefore, our method entails segmenting each input spectrogram into individual frames that correspond to a distinct step of human locomotion. Subsequently, we classify each segment independently and combine the resulting classification scores to obtain the final score for the entire spectrogram. As a result of this segmentation, the size of the training set is increased, whereas the dimensions of each sample—and therefore the number of parameters in the classifier—is decreased, reducing the risk of overfitting. Our experimental results demonstrate the effectiveness of our approach and its potential to enhance CNN-based classification of micro-Doppler signatures.**

***Index Terms*—micro-Doppler, human detection, target classification, neural networks**

I. Introduction

Security surveillance in public spaces, such as airports or parks, is typically done using video systems. Optical cameras, in general, are relatively cheap and provide high-resolution imagery that is easy to interpret. However, the quality of the images can be heavily degraded in environments with poor lighting conditions, dust or smoke. In addition, simple cameras cannot provide range information and their use is usually restricted due to privacy concerns. Some of these drawbacks could be mitigated by complementing camera systems with radars, since radars are insensitive to light conditions, can provide range information and incur less privacy problems [1], [2].

This paper focuses on the classification of armed and unarmed individuals in surveillance scenarios. This task is usually performed with convolutional neural networks (CNNs) taking spectrograms as input [3]–[5]. Unfortunately, CNNs require a substantial amount of labeled data to be trained effectively, which may be inadequate in defense and security applications due to the high costs related to obtaining real-world measurements. In such scenarios, CNNs may overfit the limited available data, leading to the the memorization of the training set by the model and rendering it incapable of generalizing its outcomes to unseen data. This issue could be mitigated by leveraging appropriate data properties specific to the application.

Components with periodic motion, such as the rotors in a drone, the wings of a bird or the limbs of a walking person will appear in the micro-Doppler signature (spectogram) of the object of interest as quasi-periodic structures that repeat over time, which we will refer to here as ”cycles”. From the point of view of classification, the optimal number of cycles in a spectrogram will depend on the characteristics of the component (rotor, wing, limb) and the correlation between the cycles, which is related to the trajectory and maneuverability of the main object (drone, bird, person). If the number of cycles in the spectrogram is too large, the input to the CNN will have high dimensionality, resulting in an unnecessarily large model that will overfit the data. If the number of cycles is too low, information about the trajectory or maneuverability, which might be useful from classification purposes, will be lost.

For our application, we hypothesize that the only difference between unarmed or armed personnel is that the latter carries an object, for instance, a rifle. This implies that the main difference in their spectrograms will be in the presence or absence of the signature corresponding to the swinging of the arms [6]. Overall motion characteristics such as trajectory and variations in target speed are not informative for inferring the target's class. This is, the key features are contained within a single movement cycle, in this case a single step of the human motion, rather than across the entire observation time. As a result, the relationship between subsequent steps could be overlooked with minimal loss of information for this classification task. Note that this assumption may not be true for other types of targets, such as unmanned aerial vehicles.

Based on this hypothesis, we propose a segmentation approach for the micro-Doppler spectrogram data, in which individual steps are classified and their information is fused in a subsequent stage. Such approach has the potential to yield a smaller model size and a larger number of training samples, mitigating the risk of overfitting the classifier to the training data.

This paper is structured as follows: in Section II, we present the dataset employed for the experimentation. In Section III, we detail our proposed approach and discuss its contribution to the state of the art. Section IV showcases the experimental

978-1-6654-8278-3/23/$31.00 ©2023 IEEE

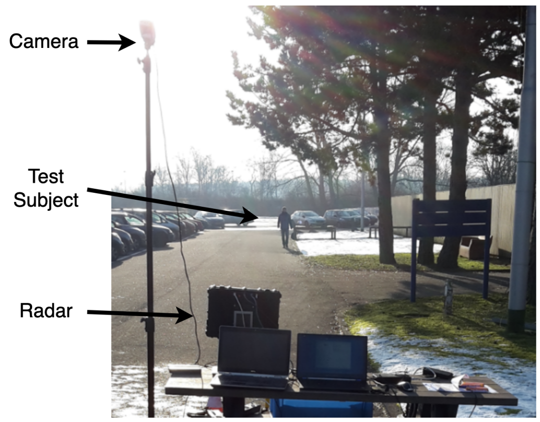


Fig. 1: The setup used in the measurement campaign. The test subject is walking towards the radar which is set on a table.

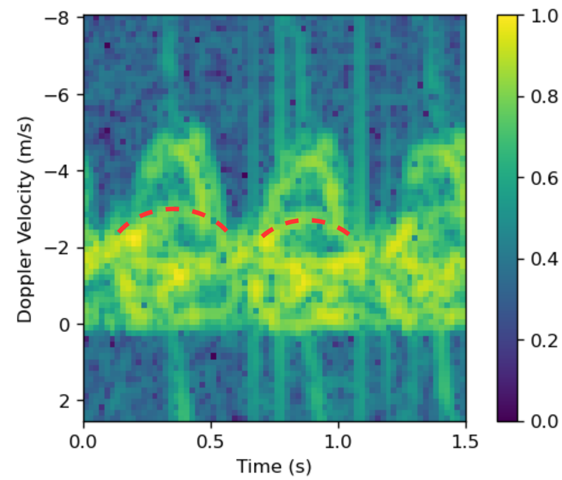
findings and evaluation of our proposed methodology. Finally, in Section V, we provide a summary of the main conclusions and suggest potential directions for future research.

II. Dataset

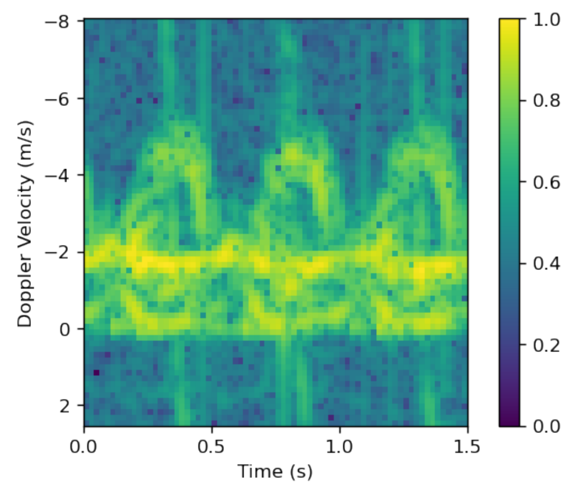
Our classification approach will be evaluated on the dataset described in [2]. Fig. 1 presents an overview of the experimental setup that was used. The measurements were acquired using a frequency-modulated, continuous-wave radar operating in X-band. During the acquisition campaign, 35 subjects were instructed to walk following a straight trajectory towards the radar from a distance of 40 m at a consistent pace of approximately 1.5 to 2 m/s. The subjects were left free to walk without following a specific form, and exhibited different walking behaviours (e.g. walking with their hands in their pockets). The subjects were asked to walk without any objects in their hands and then to walk while holding a metal object in both hands to simulate the action of carrying a weapon. Each subject was recorded performing each activity twice, resulting in a total of 140 measurements.

To obtain the spectrograms, we computed the short-time Fourier transform (STFT) of each measured radar signal. The STFTs were computed from overlapping sequences of the radar signals, with an integration length of 0.1 s and an overlap of 80%. We then took the squared magnitude of the resulting STFTs to obtain the spectrograms. Each captured spectrogram was then divided into frames in time of length 1.5 s, corresponding to approximately one and a half human steps. Finally, we further cropped the spectrograms in frequency to a size of 72 frequency bins by 72 time bins, to focus only on the relevant frequency content. The resulting dataset consists of around 1000 spectrograms, which are equally distributed between the two classes of interest: unarmed and armed.

Fig. 2 depicts two spectrograms representing the micro-Doppler signature of individuals from the armed and unarmed classes, respectively. It can be observed that the return due to the swinging motion of the arms is highlighted, which distinguishes it from the sample belonging to the armed class.



(a) Unarmed



(b) Armed

Fig. 2: Framed spectrograms from the measured datasets belonging to (a) a subject walking hands-free and (b) a different subject carrying a rifle. The micro-Doppler response from the arm is highlighted in (a). The velocity is assumed positive when walking away from the radar. The magnitude in dB scale of the spectrograms is scaled to be within $[0, 1]$.

III. Approach

In this study, we make the assumption that the primary distinguishing factor between the spectrograms of unarmed and armed personnel is the presence or absence of the micro-Doppler signature produced by the arms, as illustrated in Fig. 2. In view of this, we hypothesize that the correlation between cycles does not provide significant information for classification purposes. It is important to note that the cycles analyzed in this study refer specifically to the individual steps taken during human locomotion, rather than a complete gait cycle which comprises a left step followed by a right step.

Under the aforementioned assumptions, it is feasible to train a classifier that can predict whether a target is armed or unarmed by analyzing a single step of the target's gait. Traditional CNN-based approaches that utilise longer cyclic spectrograms as input run into the risk of learning less relevant dependencies between consecutive cycles, a symptom of overfitting in low-data regimes. By training a classifier on individual cycles, the model is compelled to concentrate on the distinctive characteristics of each step rather than on the properties of the entire spectrogram of the walking motion, which has the potential to result in models that better capture the underlying features of the data. Moreover, by training a classifier on individual cycles, the dimensionality of the input is reduced, resulting in a reduced number of parameters required to model the input-output relationship between spectrograms and class labels. This segmenting of the input would also allow for the use of more data points in the training process, which in turn would mitigate the risk of overfitting to the training dataset when the number of training spectrograms is limited.

Our pipeline is therefore defined as follows:

1) Segment the original spectrogram S into individual cycles S_i with $i = 1, 2, 3$
2) Obtain the classification score $\mathbf{d}_i$ for the respective individual cycle S_i
3) Fuse the classification scores $\mathbf{d}_i$ originated from the same initial spectrogram S to obtain a final classification score $\hat{\mathbf{d}}$ for S. The final decision is taken as the class that corresponds to the maximum entry of $\hat{\mathbf{d}}$.

Fig. 3 shows a diagram of the classification pipeline. In the first step, we segment the original spectrograms (Fig. 2) into single cycles. The segmentation is performed by fitting a sinusoid to the upper envelope of the spectrogram to maintain phase consistency across cycles. To properly determine the envelope, we first apply a hard threshold to remove the background noise. The threshold value is set to 65% of the maximum amplitude among the spectrogram bins. Subsequently, we apply a rectangular window of 0.5 s, centered around the peaks of the envelope of the micro-Doppler signature, to perform the segmentation. For each spectrogram, we obtain three framed cycles. The newly obtained dataset contains approximately 3000 segments, with 3 segments per original spectrogram. Each segment has a size of 72 frequency bins by 24 time bins. As a result, the dataset has been expanded threefold, providing a substantial increase in data quantity, while simultaneously allowing for a reduction in the number of parameters required for the classifier.

In the second step, each segment obtained from the first step is classified separately using the same classifier. To keep the comparison consistent, the structure of this classifier is kept the same as the one that was found to be most effective to classify the entire spectrograms. This classifier comprises two stages. First, a series of convolutional layers with increasing amount of filters, separated by max pooling layers to reduce the dimensionality of the input after each convolution operation. Second, an stage where the previous output is flattened and classified by means of three dense layers. The first two are

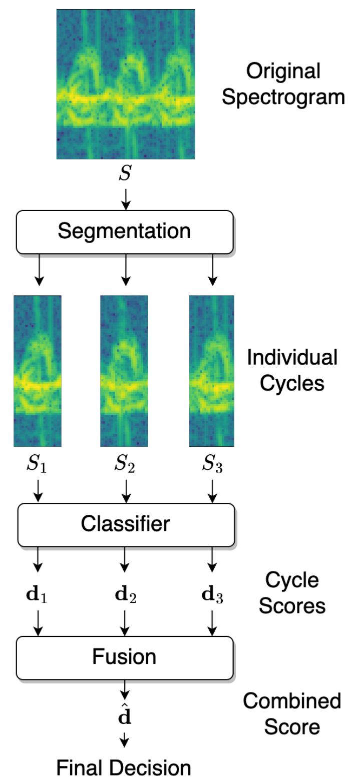


Fig. 3: Schematic of the classification pipeline. The input spectrogram is segmented into N cycles, each of which is separately input to the classifier. The resulting softmax outputs, $\{\mathbf{d}_i\}$, are then combined to obtain and unanimous decision on the class.

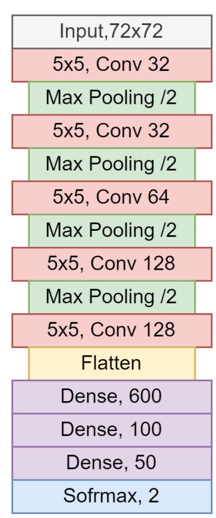


Fig. 4: Structure of the classifier used in the second step of the proposed approach.

followed by ReLU activation functions, while the last one is followed by a softmax layer used to obtain the final classification score. A diagram of the proposed architecture is shown in Fig. 4.

The classification of distinct individual cycles belonging to the same original spectrogram results in multiple classifier outputs per spectrogram. Consequently, it is crucial to investigate how the classification scores belonging to different cycles can be combined to reach a unanimous decision which may enhance the accuracy achieved when classifying the original spectrograms as a whole. In fact, although the classification is performed on a cycle basis, a unified decision can be made by merging the predictions that the classifier produces for each of the segments that compose the original micro-Doppler spectrogram.

Therefore, as a third step, we perform the fusion of the scores of the individual input cycles, by employing techniques that can be adapted from the task of ensemble learning [7]. This area of study primarily concerns methods to combine the output of multiple classifiers with the aim of enhancing the accuracy of a given classification task. Although the approach proposed in this study involves a single model, the classification of multiple input cycles poses a similar challenge in terms of determining an appropriate approach for integrating multiple classification scores.

Among the approaches adopted for ensemble learning, particularly relevant to our problem are voting techniques [8], where each combined score is considered as a vote towards the final decision on the input's class. Voting strategies are divided into the categories of hard and soft voting. Hard voting involves the combination of the hard scores, which are generated by applying a hard threshold to the each softmax classification output. Among hard voting techniques, majority voting treats each non-zero hard score as a vote for the corresponding class. The final classification decision is made based on the class with the highest number of votes. In the event of a tie, no decision is made on the final class.

On the other hand, soft voting relies on the combination of the soft scores, namely the softmax outputs of each separate classification. To reach the final decision, different combination schemes of the soft scores can be considered, such as the arithmetic mean, geometric mean, or median of the scores. The first two are defined as $\frac{1}{3}\sum_{i=1}^{3}\mathbf{d}_i$ and $\sqrt[3]{\prod_{i=1}^{3}\mathbf{d}_i}$ respectively, where $\mathbf{d}_i$ is the classification score of the i-th frame of the original spectrogram. The fusion via the median rule, instead, is achieved by considering the median across the different frames as the final score for each class. In all cases, the decision on the final class label is then determined by selecting the class that achieves the maximum score after applying the chosen combination rule.

In the experimental work presented in this paper, we will evaluate different voting schemes based on their achieved classification accuracy on the experimental dataset.

IV. Experimental Results

In this section, we aim to evaluate the performance of the proposed classification approaches by analyzing their ability to accurately classify samples from the experimental data. To achieve this, we will use 20% of the available data to create a test dataset, while the remaining samples will be used for training purposes. The splitting will be performed subject-wise, to increase the separation between training and testing data. Additionally, we will analyze the performance of the proposed approach in data-scarce scenarios by gradually reducing the size of the training set through the removal of subjects. We will validate the models on 100 random folds of the datasets for each size of the training set to ensure statistical significance of our results. For each training set size, 20% of the training data will be used for validation

Following the computation of the baseline accuracy obtained by utilizing the entire spectrograms (Fig. 2) as input to the classifier, we will assess the impact of classifying the individual cycles and applying five distinct combination techniques: majority voting using the hard thresholded scores, arithmetic mean, geometric mean and median of the soft scores, and no fusion. Specifically, in the latter case when no fusion is being performed, we consider the classification made by observing only one cycle at a time, by directly inferring the class from the scores of a single cycle. In all cases, the final chosen class will be the one with the highest final score.

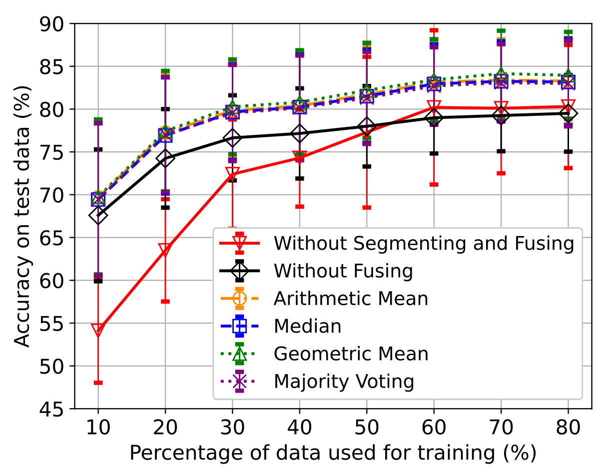


Fig. 5: Accuracy achieved using different fusion techniques over the amount of subjects used for training. Error bars represent standard deviation.

Figure 5 illustrates the accuracy achieved using the various classification and fusion approaches at different levels of data scarcity. Results indicate that approaches based on data segmentation and score fusion outperform the baseline obtained by classifying the entire spectrograms, especially when the amount of training samples is scarce. The results also indicate that a similar level of accuracy is achieved regardless of the chosen fusion technique. Despite the soft output providing

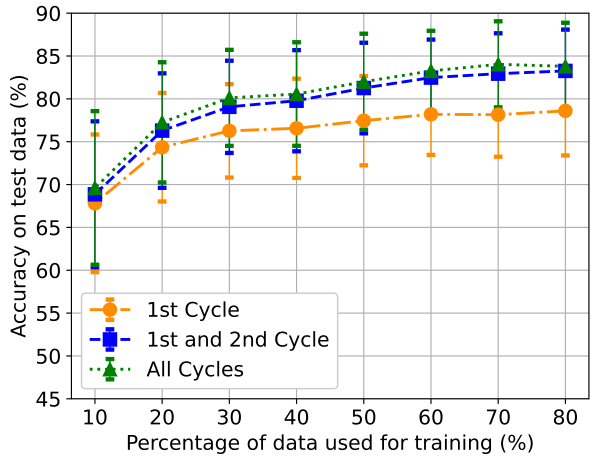


Fig. 6: Accuracy achieved when combining progressively increasing number of segments via the geometric mean of the soft classification scores. Error bars represent standard deviation.

additional information regarding the prediction's confidence level, there seems to be no considerable improvement in the fusion process in comparison to utilizing only hard thresholded labels.

Upon examining the performance curve obtained by classifying single cycles without any fusion, it can be observed that segmenting the data results in a significant accuracy improvement in the low-data regimes, due to the increased amount of training samples and the reduced number of parameters of the classifier. However, when the training set size is larger, the accuracy achieved without fusion becomes slightly worse than the baseline one. This is due to the fact that relying on a single segment for classification lacks the additional information available in the baseline case, resulting in less accuracy. Therefore, fusing scores from multiple cycles is necessary to improve performance in such cases.

In order to evaluate the impact of incorporating the output of classification for additional cycles, we assessed the classification accuracy over different training sizes when utilising the 1st cycle of the original spectrogram, the 1st and 2nd cycles and all the cycles respectively. The results, as presented in Fig. 6, demonstrate that including more cycles in the classification process leads to an increase in the final accuracy on the test data. Notably, the greatest increase in accuracy is achieved through the inclusion of the second cycle, with performance appearing to reach saturation upon the addition of the third and final frame of the original spectrogram. This observation aligns with the expectation that the individual segments representing a human step are not independent realizations, but rather the representation of actions performed consecutively by the same subject. Consequently, the inference of the class label is affected by the correlation between subsequent steps and the accuracy does not increase linearly with the number of considered frames.

V. Conclusion

This study highlights the importance of critically assessing the information content of a micro-Doppler spectrogram and its connection to the physics of the generating target before designing a classification strategy.

We have designed a classification approach based on the hypothesis that the main difference between the spectrogram of unarmed and armed personnel is in the absence or presence of the swinging arms micro-Doppler signature, respectively. The proposed approach involves segmenting the original spectrograms into individual cycles and separately classifying them, while combining their classification scores to improve accuracy and mitigate overfitting. Results reveal that this approach enhances classification performance and is particularly effective in data-scarce scenarios. Furthermore, our results demonstrate that the accuracy improvement tends to saturate with an increase in the number of cycles considered. This observation is consistent with the fact that all cycles are generated by the same dynamical processes (i.e., limb swinging), and suggests that there may be a limit to the amount of information that can be extracted by considering additional steps. Note that although all the subjects in the experiments are walking in straight line towards the radar, the hypothesis behind our methodology remains applicable even in scenarios where subjects follow different trajectories during their movement.

Further research could investigate the applicability of our approach in scenarios where the relationship between individual cycles bears relevance to the classification task. Although we assumed that a single cycle captures all the relevant features for classification in this work, our approach might also be employed in situations where this assumption does not hold. For instance, in drone classification, the information about the type of drone is encoded both within a single cycle (type of rotor, rotation speed) and within the relationship between cycles (trajectory, maneuverability). However, in a drone spectrogram, the time scale of the features related to maneuverability might be an order of magnitude larger (or more) compared to the time scales associated with the rotation of the drone's blades, depending on the drone inertia. This means that in a given observation time, the features related to blade rotation would be observed more frequently than those characterizing the drone's maneuverability. If the available data is too scarce, it might be preferable to focus the model in the well represented single cycles than in the poorly represented trajectories.

In addition, future research could explore alternative strategies for combining cycle scores other than voting. One possibility could be to apply weights to the score of each segment before the fusion. The weights could be determined by properties of the input cycles, such as an estimate of the signal-to-noise ratio. Such approaches may offer improved accuracy and robustness in the classification task.

REFERENCES

[1] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor person identification using a low-power fmcw radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.

[2] R. J. de Jong, J. J. M. de Wit, and F. Uysal, "Classification of human activity using radar and video multimodal learning," *IET Radar, Sonar & Navigation*, vol. 15, no. 8, pp. 902–914, 2021.

[3] Y. Kim and T. Moon, "Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 8–12, 2016.

[4] R. Zhang and S. Cao, "Real-time human motion behavior detection via cnn using mmwave radar," *IEEE Sensors Letters*, vol. 3, no. 2, pp. 1–4, 2019.

[5] J. S. Patel, F. Fioranelli, M. Ritchie, and H. Griffiths, "Multistatic radar classification of armed vs unarmed personnel using neural networks," *Evolving Systems*, vol. 9, no. 2, pp. 135–144, Jun. 2018.

[6] F. Fioranelli, M. Ritchie, and H. Griffiths, "Classification of unarmed/armed personnel using the netrad multistatic radar for micro-doppler and singular value decomposition features," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, pp. 1933–1937, 2015.

[7] M. Ganaie, M. Hu, A. Malik, M. Tanveer, and P. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, 2022.

[8] M. Mohandes, M. Deriche, and S. O. Aliyu, "Classifiers combination techniques: A comprehensive review," *IEEE Access*, vol. 6, pp. 19 626–19 639, 2018.