Classification Strategies for Radar-Based Continuous Human Activity Recognition With Multiple Inputs and Multilabel Output

Ullmann, Ingrid; Guendel, Ronny G.; Christian Kruse, Nicolas; Fioranelli, Francesco; Yarovoy, Alexander

DOI
10.1109/JSEN.2024.3429549

Publication date
2024

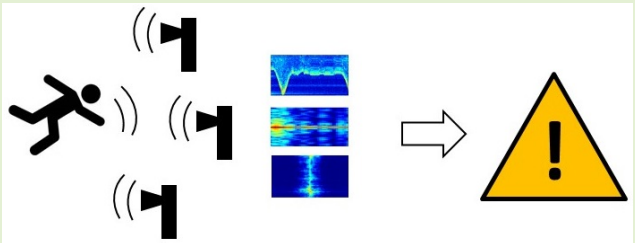
Document Version
Final published version

Published in
IEEE Sensors Journal

Classification Strategies for Radar-Based Continuous Human Activity Recognition With Multiple Inputs and Multilabel Output

Ingrid Ullmann, *Member, IEEE*, Ronny G. Guendel, Nicolas Christian Kruse, *Graduate Student Member, IEEE*, Francesco Fioranelli, *Senior Member, IEEE*, and Alexander Yarovoy, *Fellow, IEEE*

***Abstract*—Fall detection systems can play an important role in assuring safe independent living for vulnerable people. These sensors not only have to detect falls but also have to recognize uncritical, normal activities of daily living in order to differentiate them from falls. Radar sensors are very attractive for human activity recognition thanks to their contactless capabilities and lack of plain videos recorded. In this article, a novel approach to recognize single activities in a continuous stream of radar data is proposed, whereby the stream is divided into windows of fixed length and, then, multilabel classification is used to recognize all activities taking place in these time segments. While the initial feasibility of this approach was presented in an earlier contribution presented at the 2023 IEEE SENSORS conference, in this extended work, additional in-depth studies on critical parameters are performed. Specifically, multiple combinations of different radar data domains/representations (e.g., range-time maps, range-Doppler maps, and spectrograms) and different radar nodes in a network of five cooperating sensors are considered as inputs to two considered multilabel classification networks. In addition, a parametric study on the probability thresholds of the networks to assign labels to specific classes is also performed.**



***Index Terms*— Activities of daily living, deep learning, human activity recognition, multilabel classification, radar.**

I. Introduction

FALLING is one of the major risks for older and vulnerable people. Nevertheless, people want to live in their own homes for as long as possible. With an aging society, providing independent living for older and vulnerable people will be one of the great societal challenges. To allow for safe, independent living, automated fall detection systems are desirable. Such systems can detect falls and react to them quickly. To recognize dangerous situations reliably, fall detection systems must be able to distinguish falls from other types of human motion related to normal activities of daily living such as walking and sitting down. This kind of motion understanding is often termed human activity recognition.

Various sensor principles can be applied to fall detection systems [1]. Wearable sensors are well established, but they have the disadvantage that they have to be worn permanently, which may be difficult for people suffering from mental diseases such as dementia. For this reason, contactless sensors are desirable. Optical systems, such as cameras, can serve this task but have some main disadvantages. First, they are strongly dependent on the light conditions; second, they are critical with respect to privacy, and third, the individual must be observable by the line-of-sight view [2]. For these reasons, radar has emerged as an interesting alternative. Radar is a contactless sensing principle as well, but it does not have the disadvantages of cameras. Radar sensing is independent of light and sight and it raises less concerns about privacy since radar data are not readily interpretable for humans.

Manuscript received 10 June 2024; accepted 9 July 2024. Date of publication 23 July 2024; date of current version 13 December 2024. This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG), German Research Foundation-SFB 1483, under Project 442419336, EmpkinS; and in part by the Dutch Research Council [Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)] through the *RAD-ART* Project. This article is an extended paper of "Radar-Based Continuous Human Activity Recognition with Multi-Label Classification," by the authors, presented at the 2023 IEEE SENSORS. The associate editor coordinating the review of this article and approving it for publication was Dr. Juanjuan Shi. *(Corresponding author: Ingrid Ullmann.)*

Ingrid Ullmann is with the Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany (e-mail: ingrid.ullmann@fau.de).

Ronny G. Guendel, Nicolas Christian Kruse, Francesco Fioranelli, and Alexander Yarovoy are with the Microwave Sensing, Signals and Systems Group, Delft University of Technology, 2628 CD Delft, The Netherlands.

Digital Object Identifier 10.1109/JSEN.2024.3429549

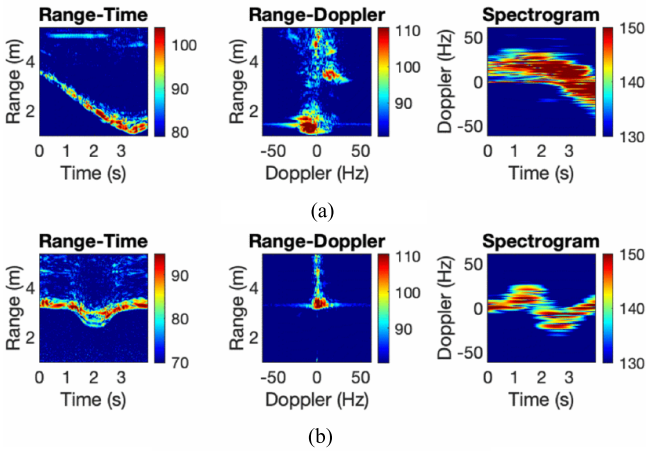


Fig. 1. Different radar data representations (copied from [6]). From left to right: range-time, range-Doppler, and spectrogram for two activities of daily living (walking and bending from standing). (a) Label 1: walking. (b) Label 6: bending from standing.

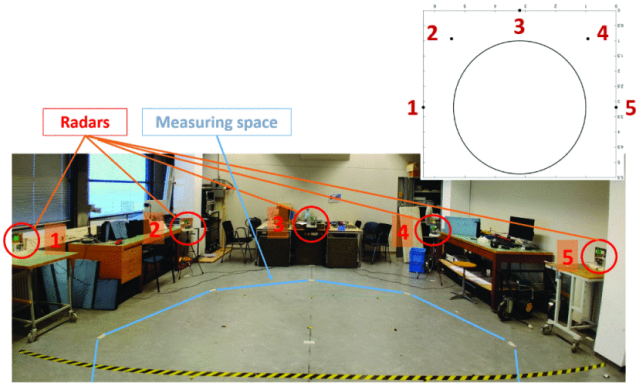


Fig. 2. Measurement setup that was used to collect the dataset we employed in this article (copied from [13]). Five radars are arranged in a semicircle as shown in the photograph and illustrated in the sketch on the upper right.

Due to its benefits, considerable research efforts have been directed toward investigating human activity recognition and fall detection by means of radar [3], [4], [5]. Typically, the data captured by the radar are subsequently processed by machine learning, and in this way, single activities can be recognized. Since radar data typically are stored as a multidimensional data tensor, various representations of the data can be used for the classification task. These include range-time data, range-Doppler maps, and spectrograms (Doppler over time). The three representations are illustrated in Fig. 1 for two activities of daily living.

Most of the work published so far has simply measured a set of separate activities and classified those [7], [8]. However, in realistic scenarios, activities will not take place separately but in a continuous stream with unknown start and stop times. Recognizing single activities in a continuous signal is what we call continuous human activity recognition. Various papers have addressed the challenge of continuous activity recognition [9], [10], [11], [12], [13]. Two main approaches to tackle the issue of continuity have emerged. The first is to use recurrent neural networks (RNNs) for classification, such as the long-short term memory (LSTM) [9], [10]. This works well, but RNNs need temporal sequences as input. Therefore, not all information can be used, e.g., range-Doppler maps are no sequence. Thus, potentially, some information is lost, which could augment classification performance. It has been shown that using multiple data inputs is beneficial for the classification [14].

The second approach to continuous human activity recognition is to recognize transitions between single activities in the first step and then classify these single activities individually [11], [12], [15]. However, this approach depends critically on the segmentation step and it will fail completely when multiple activities take place concurrently but start at different times.

To overcome the aforementioned problems, our previous work [16], presented at the 2023 IEEE SENSORS conference in Vienna, introduced a novel approach. Our approach is a two-step procedure of first segmenting the continuous time stream into windows of a fixed length and subsequently using multilabel classification to recognize which activities are performed in each window. The initial results presented in that work introduced the general concept and proved its feasibility. In this extended paper, additional in-depth studies on critical parameters are performed. Specifically, multiple combinations of different radar data domains/representations and different radar nodes in a network of five cooperating sensors are considered, highlighting the importance of finding the most suitable combinations to boost classification performances.

The rest of this article is outlined as follows. In Section II, the general concept introduced in the conference paper is briefly recapitulated for clarity. Section III provides an overview of the parameter study that we conducted, as well as details on the network implementations and evaluation metrics. Section IV presents the obtained results from the parameter study. Finally, a conclusion and an outlook on future work are provided in Section V.

II. Activity Recognition With Multilabel Classification

In this section, our concept of using multilabel classification for human activity recognition is briefly recapitulated.

For our work, we used a publicly available dataset [17], which was captured in previous work by some of the authors of this article. The dataset consists of radar measurement data of nine human activities. For capturing the data, five impulse radars of type PulsON P410 were used. They operate at 4.3 GHz with a bandwidth of 2.2 GHz. The five radars were arranged in a semicircle, as shown in Fig. 2. Using a distributed radar network instead of one radar requires more efforts such as synchronization of the stations. However, the advantage of a distributed setup is that more information can be captured. For example, a single, monostatic radar is not capable of capturing Doppler information from movement tangential to the aperture. With the setup in Fig. 2, all human movement can be captured independently of its direction.

For the dataset, 15 subjects performed nine activities of daily living. The activities are given as follows:

- walking;
- sitting down;

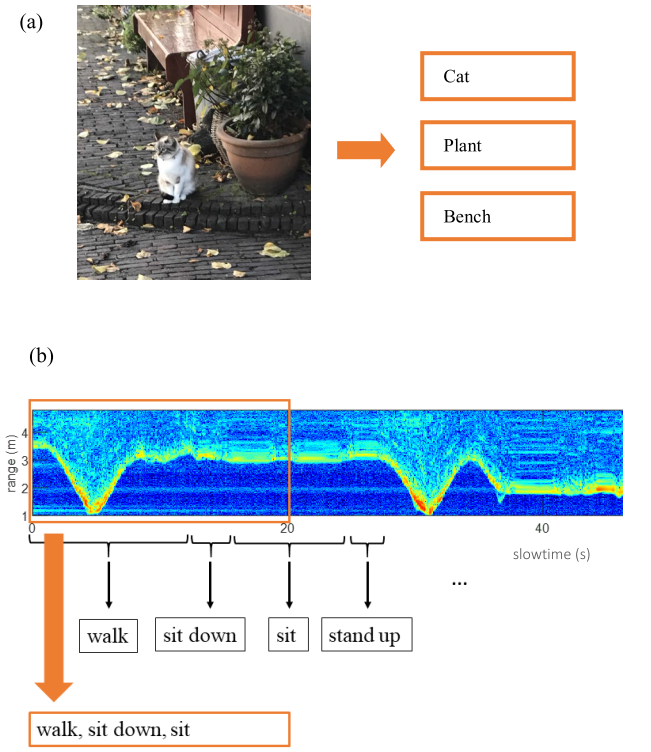


Fig. 3. Illustration of multilabel classification in (a) computer vision and (b) for human activity recognition from a time series. It can detect multiple labels, e.g., that there is a cat, a plant, and a bench on the photograph and, analogously, that the time window in orange contains radar signatures of the activities walking, sitting down, and sitting. (b) is copied from [16].

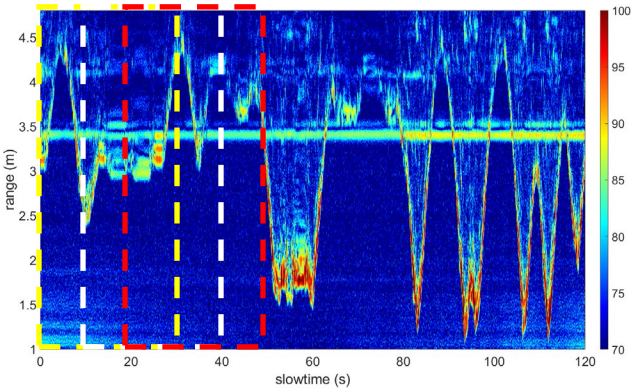


Fig. 4. Illustration of the proposed segmentation process The data stream is split into windows of 30 s, starting every 10 s. In this example, the first frame is marked by the yellow dashed line, lasting from time $t = 0$ to $t = 30$ s. The second frame (white dashed line) lasts from $t = 10$ to $t = 40$ s; the third frame (red line) lasts from $t = 20$ to $t = 50$ s and so on.

- standing up from sitting;
- bending from sitting;
- bending from standing;
- no activity;
- falling from walking;
- falling from standing;
- standing up.

In addition, there is a null class for data corresponding to none of the previously mentioned activities or corrupted data. Thirty continuous time streams are available for each subject, with each stream containing 120 s of data. The data are single activity sequences as well as mixed activity sequences. For more details on the dataset, we refer to [18].

Consider a time stream of human activities captured by a radar system, as a range-over-time signal [as shown in, e.g., Fig. 3(b)]. Our approach is to first segment this time stream into windows of fixed length (e.g., 30 s). This approach avoids the physical segmentation as proposed in [11], [12], and [15]. In addition, in contrast to using RNNs, the evaluation of range-Doppler map information is possible with this approach because the window can be transformed to a range-Doppler map via a Fourier transform along the slow-time axis.

Therefore, some of the possible limitations of the state of the art are circumvented. A problem that arises from segmenting the time stream into fixed-length windows is that the windows are not unique regarding activities that take place within that window. For this reason, we introduced multilabel classification into human activity recognition. Multilabel classification is well established in computer vision (see, e.g., [19]). It can detect several entities in an image, as illustrated in Fig. 3(a). Consequently, multilabel classification is able to detect several activities within a time window. This is illustrated in Fig. 3(b).

Another problem that can arise with fixed-length windowing is that an activity can take place at the very beginning or end of a window, which could cause problems in recognizing it. To avoid this, we chose to have the windows overlapped, as illustrated in Fig. 4.

With this strategy, even if one activity is not recognized in one window, it may be well recognized in the following, time-shifted window. For our implementation, we chose a window length of 30 s and an overlap starting every 10 s. We chose a 30 s window length as it appeared a good compromise between capturing enough information, being able to react fast and at the same time not having to compute a classification too often. In a practical setting, all these issues will be relevant. With a classification twice in a minute, we thought that a good compromise might be reached.

From the 30-s windows of range-time data, we computed the range-Doppler maps by means of a Fourier transform along time and the spectrograms by means of a short-time Fourier transform after summing up along the range dimension. The three representations (range-time, range-Doppler, and spectrogram) of the radar data all contribute to the overall information for the classification. In the implementation presented in the conference paper [16], individual classification networks were trained for all three representations and the results were fused in a postprocessing step, at the decision level. However, fusing at an earlier stage is possible as well. Investigating such early fusion strategies is a part of this extended paper.

The output of the multilabel classification network is a vector containing a probability for each activity class. Each probability lies in the interval between 0 and 1. To determine the actual presence or absence of an activity, a threshold has

to be set. Then, each activity having a probability above this threshold is said to be present within the segment and each activity having a probability below this threshold is said to be absent. This binary output can be compared to the ground truth. The ground truth for each segment is a binary vector as well, where the presence/absence of an activity is indicated by 1/0. It is therefore very important to investigate the effect of different values of this threshold on the overall classification performance. This is another objective of this extended paper.

In the conference implementation, we chose the threshold value to be 0.5. This seemed an intuitive value because it implies that if the network computes a probability above 50%, then we say that the activity is rather there than not there and set the binary decision to "present." In contrast, using a higher threshold implies that the network has to be more strongly "convinced" that an activity is actually present and might avoid misclassifications (reducing the number of false positives). Instead, using a lower threshold value implies a higher acceptance of any activity above a certain noise level, which reduces the number of false negatives.

In the conference work, a ResNet50 (residual network with 50 layers) was used for each classification. It performed well, but ResNet50 is a very deep network and therefore is computationally costly. Other networks that are shallower perform faster and therefore are investigated in this extended paper.

Another degree of freedom in the network design is the question of how to treat the data from the five radar nodes. So far, all data were used separately, and however, other fusion strategies are possible as well. In [20], such investigations were performed and it was reported that fusion at the signal level, i.e., summation of the five radars' data performed best. Therefore, we will compare this approach to the no-fusion approach in this article.

III. Parameter Studies

A. Parameter Study Overview

As indicated in the previous section, there are various possibilities to treat the multiple inputs when using a multisensor network and multiple radar data representations. Equally, the output decision based on the multilabel classification result can be computed in different ways. The selection of the classification network is a further design choice.

Based on the aforementioned degrees of freedom, we performed parameter studies with respect to the following in this article:

1) no fusion versus signal-level fusion of the data from the five radars;
2) signal-level fusion and decision-level fusion for the three input data representations range-time, range-Doppler, and spectrogram;
3) comparison of two neural network architectures, namely, a simple three-stage convolutional neural network (CNN) and a ResNet18, in order to use more efficient networks compared to the ResNet50 as used in [16];
4) a range of 0.01–1 for the output decision threshold of the multilabel classification.

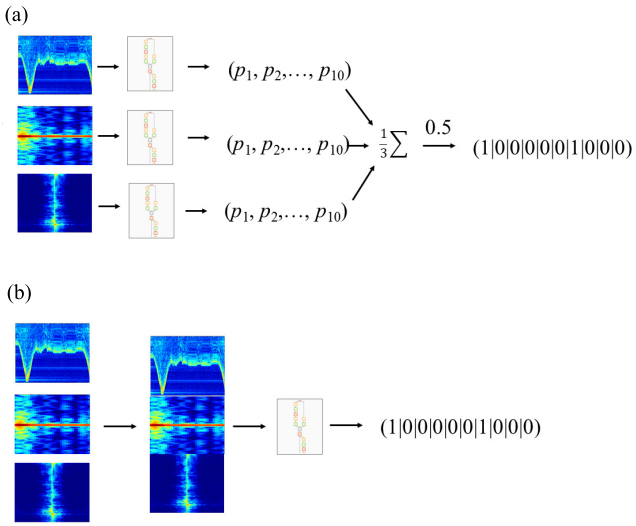


Fig. 5. (a) Decision-level fusion as implemented in the conference paper and (b) signal-level fusion.

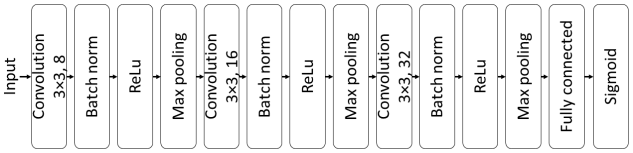


Fig. 6. Architecture of the employed three-stage CNN.

Signal-level fusion and decision-level fusion are illustrated in Fig. 5. In decision-level fusion, as was explained in detail in Section II, the three representations (range-time, spectrogram, and range-Doppler) are each fed to an individual neural network. Each network outputs a probability vector for the ten activity classes. These vectors are averaged and then compared to a threshold, e.g., 0.5. If the average probability is larger than the threshold, then the final output is 1 (activity present), and otherwise, it is 0 (absent). In contrast, for signal-level fusion, the data from the three representations are concatenated and the concatenated matrix is input to a classification network, which computes a probability for each of the ten classes, as shown in Fig. 5(b). It is evident that this has the advantage that only one network is required. However, the input to this network is thrice the size compared to decision-level fusion.

The employed three-stage CNN consists of the following:

1) an image input layer;
2) three stages consisting of convolution layer, batch normalization layer, ReLu layer, and maximum pooling layer;
3) a fully connected layer;
4) a sigmoid layer;
5) a binary cross-entropy loss output layer.

The network structure is illustrated in Fig. 6. It is smaller than ResNet18. Details on ResNet18's architecture can be found in, e.g., [21] and [22].

Investigating different window lengths and their influence on classification performance might be interesting as well. However, it is to be assumed that if the network can recognize a certain activity in a time of 30 s as used in this work, it will

also be able to recognize the activity in a longer or slightly shorter window of, e.g., 1 min or 20 s, respectively. Because of this and to keep this article reasonably concise, we chose to focus on the fusion strategies and network architectures here as it holds more scientific potential.

Another possible aspect in the parameter study might be to vary the number of radar nodes. In the previous work [20], a number of different configurations were tested, including usage of a single radar or two orthogonal radars from the semicircle (cf. Fig. 2). The results in [20] showed that using all five radar nodes gave the best classification performance. Based on these findings, we chose all five radars in this study as well.

B. Network Training, Testing, and Validation

To allow for automated classification, the neural network has to be trained with a set of training data. After training, it has to be tested to evaluate its performance. For our previous work as well as for the results presented in this article, we used leave-person-out testing. This means that the classifier is trained with measurement data from a number of test persons and tested with measurements from persons whose data have not been included in the training. In this way, the classifier is expected to be more robust and able to generalize to unseen data. We divided the dataset into nine persons' data for training, two persons' data for validation, and four persons' data for testing.

The validation, which is an intermediate testing during training, is used to monitor the training process and avoid overfitting. When training and validation loss diverge, this can be an indication for overfitting to the training data [23]. Therefore, in our implementation, we set the training options so that training is abandoned if the validation loss no longer decreases for more than ten consecutive validation iterations. Then, the network with the lowest validation loss is stored for the following classification.

The stochastic gradient descent with momentum was used as a solver and the learning rate was set to 0.0005. We used MATLAB R2022a for the implementation.

C. Evaluation Metrics

To evaluate the performance of a classification strategy, various metrics can be used. In this article, we use the accuracy, which is defined as the number of correct predictions divided by the total number of predictions [24]. Note that for the multilabel classification, the number of correct predictions includes all correctly identified present activities as well as all correctly identified nonpresent activities.

Particular attention is paid to the detection performance of falls since this is a critical activity. To investigate the fall detection performance, we examined the results for the classes "falling from walking" and "falling from standing" together. It is of particular interest to investigate the percentage of recognized falls, which, in turn, determines the percentage of missed falls. Another relevant issue is the false alarm rate, i.e., a fall that is recognized when, in reality, there is no fall.

IV. Results

In this section, the results are presented. The following subsections are divided according to the input fusion variations. For each of them, results with the two investigated network architectures for various thresholds are shown.

A. No Fusion of the Radars' Data—Decision-Level Fusion of the Data Representations

This approach was demonstrated in the conference paper, which is why we show it first. For decision-level fusion, three neural networks were trained separately for the range-time data, the range-Doppler maps, and the spectrograms. All radar data were input individually, without any fusion.

1) CNN: Fig. 7(a) shows the overall classification accuracies for all activity classes and the two combined fall classes, respectively, as well as the fall detection performance and false alarm rate, all as a function of varying the network threshold to declare an activity as detected. The thresholds were varied in steps of 0.01, ranging from 0.01 to 1.

In this configuration, a maximum classification accuracy of 89.37% was achieved for the overall activity classification (all activities). The maximum was reached for a threshold of 0.45. However, the fall detection rate at this threshold value was no more than 36.18%. Fall detection reached its maximum value (99.76%) for the lowest threshold value (0.01). This would indicate that it is beneficial to use a low threshold for fall detection; however, at this point, the false alarm rate was 93.97%.

2) ResNet18: ResNet18 is a deeper network than the aforementioned CNN, and therefore, better results could be obtained for the classification, as can be seen in Fig. 7(b). When using the ResNet18, a maximum activity classification accuracy of 95.25% was possible when selecting a threshold of 0.43. This result is comparable to the ResNet50 used in the previous work [16].

At the threshold of maximum accuracy, the fall detection rate amounted to 69.43%. When selecting low threshold values, again, fall detection could be improved. A maximum percentage for the fall detection of 99.70% was observable, again for a threshold value of 0.01. However, at this point, the false alarm rate was still 49.38%.

B. No Fusion of the Radars' Data—Signal-Level Fusion of the Data Representations

For signal-level fusion, the three 2-D matrices containing range-time, range-Doppler, and Doppler-time data are concatenated along the rows dimension and fed to one classification network. While the range-time and range-Doppler data have the same size, the Doppler-time representation does not necessarily. This is because it is obtained via a short-time Fourier transform (STFT) and therefore depends on the STFT's window length. To match dimensions, we zero-padded the Doppler-time data and subsequently concatenated the three data representations. In the STFT implementation presented here, we chose Hann windowing with a window length of 150 samples and a window overlap of ten samples. With a sample duration of 8.2 ms within the radar data, this corresponds to 1.23-s window size and 82-ms overlap. The

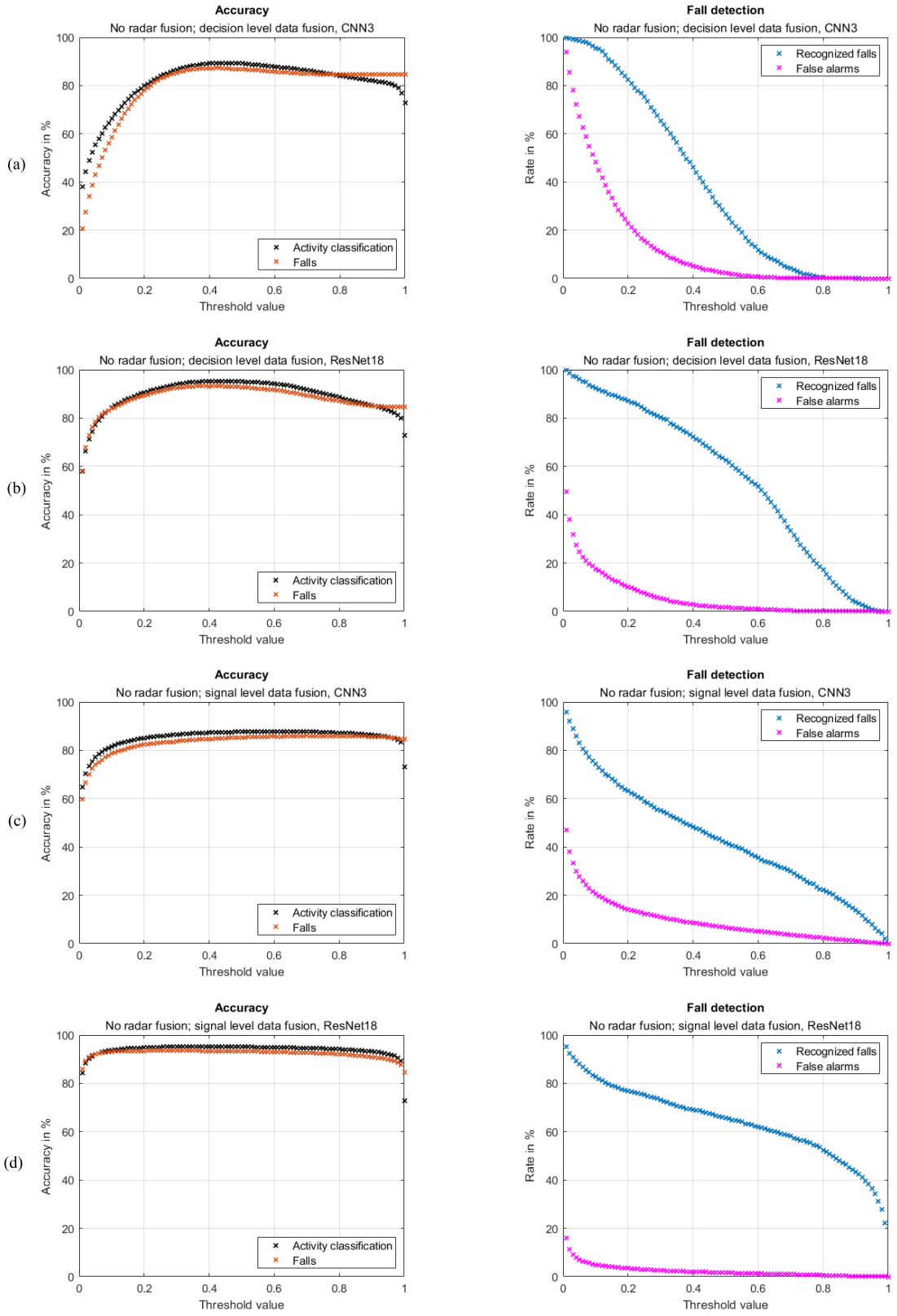


Fig. 7. Accuracies for activity recognition and falls (left column) and fall detection rate and false alarms (right column), all as a function of the threshold value. This figure sums up the results when no fusion of the radars' data was performed. The configurations of the subfigures are (see respective subtitle) (a) decision-level representation fusion + CNN3, (b) decision-level representation fusion + ResNet18, (c) signal-level representation fusion + CNN3, and (d) signal-level representation fusion + ResNet18.

two values were found to give good results in [20], which is why we chose them.

The concatenated data are three times as long as single-representation data, which results in larger images to classify,

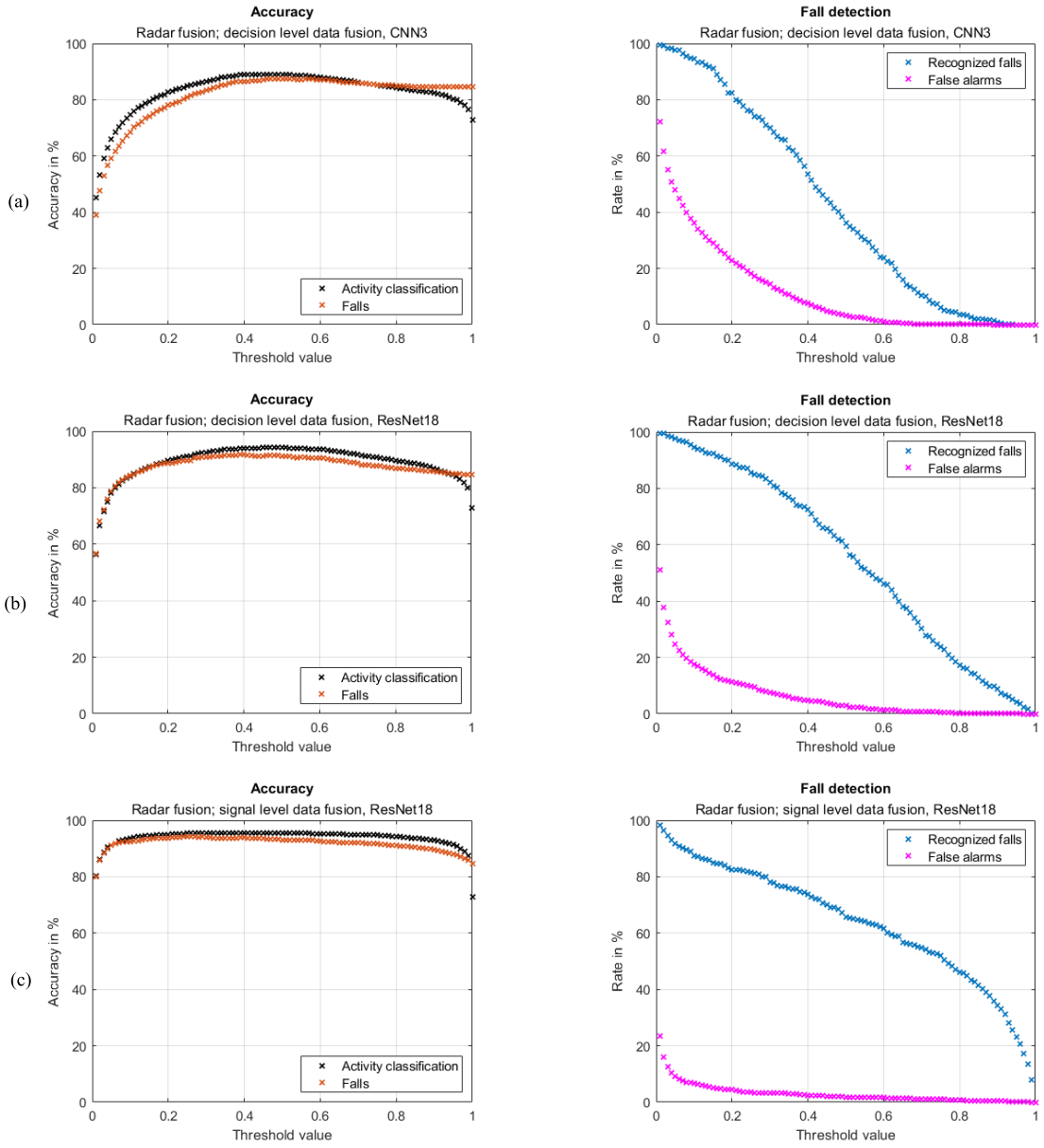


Fig. 8. Accuracies for activity recognition and falls (left column) and fall detection rate and false alarms (right column), all as a function of the threshold value. This figure sums up the results for summation fusion of the radars' data. The configurations of the subfigures are (see respective subtitle) (a) decision-level representation fusion + CNN3, (b) decision-level representation fusion + ResNet18, and (c) signal-level representation fusion + ResNet18. The configuration using signal-level representation fusion + CNN3 gave no meaningful results, which is why no results are shown for this configuration.

but the benefit is that, in contrast to decision-level fusion, only one neural network is required instead of three.

1) CNN: Results for this configuration can be seen in Fig. 7(c). A maximum classification accuracy of 87.76% was found for a threshold of 0.60. However, at this value, the fall detection rate was no more than 35.64%. Its maximum, found again at a threshold of 0.01, was 95.82%. Because of these rather low values, we deduce that this configuration is less suited for a reliable classification.

2) ResNet18: Using a ResNet18 for the same fusion concept provided better results again. The results can be seen in Fig. 7(d). Here, the classification accuracy amounted to a 95.16% maximum for a threshold of 0.42. The fall detection rate at this point is 68.66%. The maximum fall detection for low thresholds is 95.28%, which is a lower value than when using decision fusion as in Section IV-A-II.

C. Fusion of the Radars' Data—Decision-Level Fusion of the Data Representations

The five radars that form the radar network all provide data that can be fused in different fashions or used independently. In the conference publication, no fusion was performed. Fusing the radar data can take place at the signal level, feature level, and decision level. In [20], these possibilities were investigated and it turned out that a summation of the raw range-time data provided the best results. Therefore, we used this approach as well. The five range-time plots coming from the five radars were summed

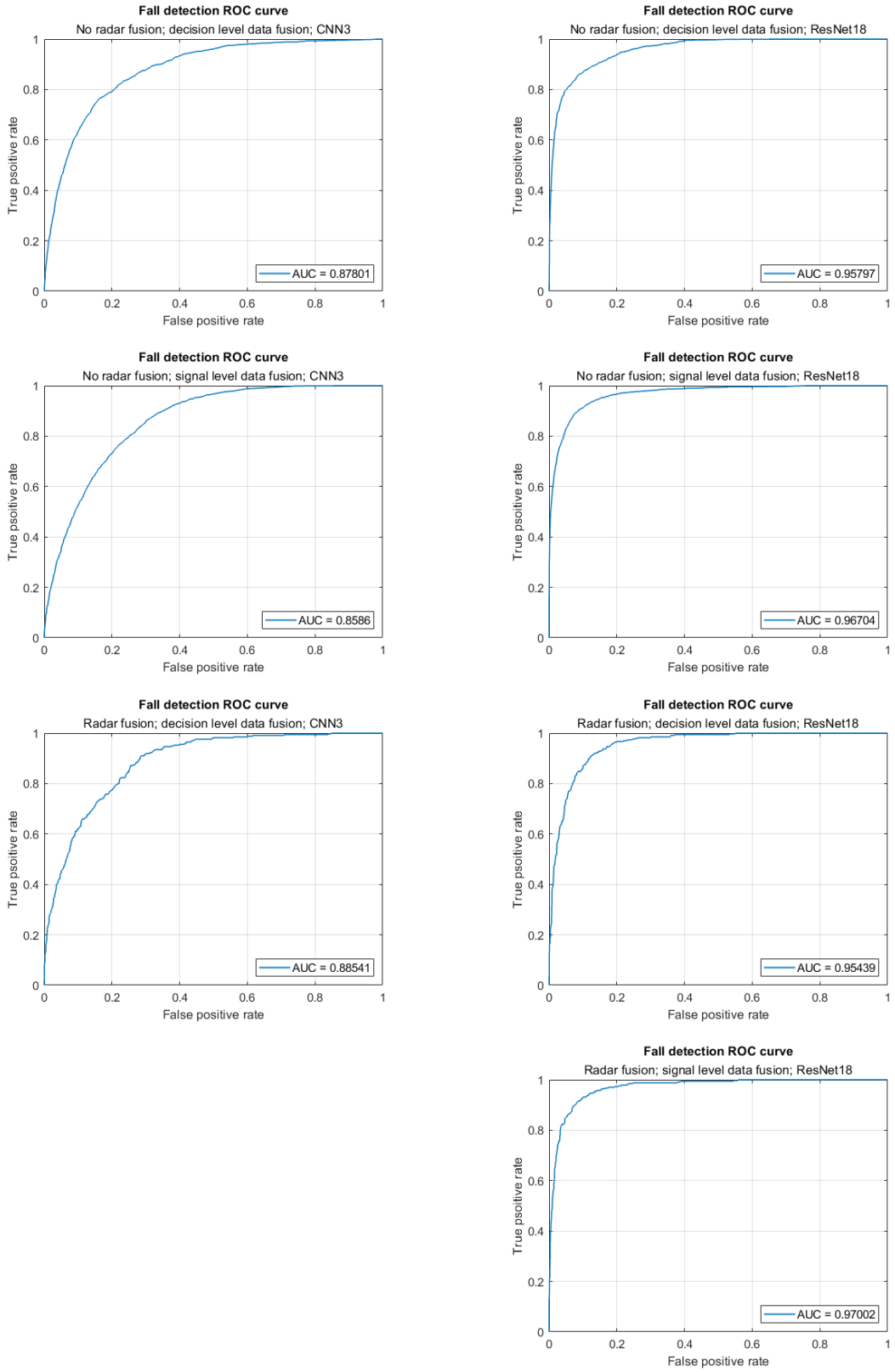


Fig. 9. Receiver operating characteristic curves and area under curve values for fall detection in the examined configurations. The configurations of the subfigures can be seen from the respective subtitles. The configuration using radar fusion, signal-level representation fusion, and CNN3 gave no meaningful results, which is why no results are shown for this configuration.

up and the resulting data were Fourier-transformed and short-time Fourier-transformed to obtain range-Doppler map and spectrogram, respectively. Results for the radar data fusion can be found in Fig. 8.

1) CNN: Fig. 8(a) shows the obtained results for this configuration. A maximum classification accuracy of 89.09% was obtained for a threshold of 0.47. This is again a rather low value compared to the other configurations. At this threshold, the fall detection rate amounted to 41.49%, which is also quite low.

2) ResNet18: Fig. 8(b) shows the results when using a ResNet18. Here, a maximum classification accuracy of 94.30% for a threshold of 0.49 was possible. The fall detection rate was 61.19% at the threshold of 0.49. For the threshold value of 0.01, a maximum percentage for the fall detection of 99.40% was obtainable, which is quite high. However, it comes with a rather high false alarm rate of 51.07%.

D. Fusion of the Radars' Data—Signal-Level Fusion of the Data Representations

In this configuration, the data of the five radars were summed up and processed as in Section IV-C. Then, the resulting range-time, range-Doppler, and spectrogram data were concatenated and fed to the classifier.

1) CNN: This configuration did not provide satisfying results. Training and validation converged at a high loss value and the trained network was not able to classify any falls correctly. Convergence at a high loss value indicates underfitting [25], which is probably because the CNN is too shallow for the problem.

2) ResNet18: Results for this configuration are shown in Fig. 8(c). A maximum classification accuracy of 95.65% could be obtained. This is the highest for all investigated configurations. It was achieved for a threshold of 0.37. At this threshold, the fall detection rate was 75.52%. The maximum percentage for fall detection is 98.21%, which is again obtained for the lowest threshold value of 0.01. At the same time, this point exhibits a false alarm rate of 23.40%.

E. Summary and Discussion of the Results

From the results in Sections IV-A–IV-D, we can see some trends that shall be discussed in the following.

Using ResNet18 for classification outperformed the three-stage CNN in all scenarios. However, since the CNN3 architecture is shallower than the ResNet18, it requires less computational power, which might be beneficial for practical application.

The ResNet18 gave a similar performance to the ResNet50 used in [16] but is a much more efficient network in terms of computational complexity.

Fusing the five radars' data by summation and concatenating the data representations did not bring any meaningful results when using the three-stage CNN as a classifier. All other investigated fusions and networks gave $> 80\%$ classification accuracies for a large span of thresholds.

We observe that the range of about 0.3–0.6 gives the best results when fusing the data representations at the decision level. When using signal-level fusion, the accuracy is similar for a larger range of thresholds. This could indicate that the classification is more evident in this case.

When looking at the fall classes, relatively high accuracies could be obtained in all cases as well. The fall accuracy curve did not differ much from the activity recognition curve in all investigated scenarios. However, simply looking at the accuracy could be misleading here because the high number

Fig. 10. Confusion matrices for various thresholds and the configuration of using radar fusion of the five radars, signal-level fusion of the three data representations, and ResNet18 for classification. From top to bottom: thresholds 0.2, 0.5, and 0.7.

TABLE I
PARAMETER STUDY RESULTS OVERVIEW

Radar fusion	Representation fusion	Network	Max. accuracy	AUC fall detection
No fusion	Decision level	CNN3	89.37 %	0.88
No fusion	Decision level	ResNet18	95.25 %	0.96
No fusion	Signal level	CNN3	87.76 %	0.86
No fusion	Signal level	ResNet18	95.16 %	0.97
Signal fusion	Decision level	CNN3	89.09 %	0.89
Signal fusion	Decision level	ResNet18	94.30 %	0.95
Signal fusion	Signal level	CNN3	/	/
Signal fusion	Signal level	ResNet18	**95.65 %**	**0.97**

of true negatives (no fall classification when there is no fall) contributes strongly to this metric. For this reason, it is useful to look at the amount of recognized falls. We can see from the results that high fall detection rates are achieved for low thresholds. However, this comes at the price of high false alarm rates. For higher threshold values, a relatively high number of falls was missed. A good compromise for fall detection might be a threshold of about 0.2. In this case, the recognized falls amounted to ca. 80 % for most of the investigated scenarios and the false alarm rate was 20% or lower.

To elucidate this more, Fig. 9 shows the receiver operating characteristic (ROC) curves [26] for fall detection in the investigated scenarios. The ROC curve plots the true positive rate versus the false positive rate for various classification thresholds. The true positive rate is defined as the number of true positives divided by the sum of true positives and false negatives. The false positive rate is defined as the number of false positives divided by the sum of false positives and true negatives [26]. A steeper ROC curve generally indicates a better performance. Another performance measure that can be deduced from the ROC curve is the area under the curve (AUC). The AUC is the integral of the ROC curve. A higher AUC value corresponds to a better performance. An ideal classification would correspond to an AUC value of 1.

From Fig. 9, we see that for the fall detection problem, the ROC curves are steeper when using ResNet18 in all cases. The AUC values all lie above 0.9, whereas when using the CNN, they are between 0.8 and 0.9. The highest AUC value is obtained when using signal-level fusion of the five radars and signal-level fusion of the data representations together with ResNet18 as a classifier. Remarkably, this configuration also gave the highest overall classification accuracy for all activities. This might indicate that it is the most suitable fusion strategy for the problem.

To have a more detailed look at the performance of this configuration, Fig. 10 shows the confusion matrices for the single activities in this setup. Since the prediction depends on the threshold, confusion matrices for three selected thresholds are shown, namely, 0.2, 0.5, and 0.7. A high number of correct predictions can be observed for all thresholds, which confirms the results from Fig. 8(c). In addition, a detailed look at single activity classes is possible from Fig. 10. For the individual classes, the percentages of correct classifications vary with threshold. A general tendency is that the number of true and false positives decreases with threshold, whereas the number of true and false negatives increases. This is expectable because, with a higher threshold, all decisions with a not-so-high confidence will be classified as 0 (absent/negative).

In general, a rather good classification performance can be seen across all classes: The accuracy for all classes and thresholds is above 90%. Thus, the misclassification rate is below 10% for all activity classes in the selected configuration. Table I sums up the results of the parameter study.

V. CONCLUSION

This article discussed various classification approaches for radar-based continuous human activity recognition with multiple inputs and multilabel output.

Regarding the investigated networks, ResNet18 clearly outperformed a simple CNN. All fusion methods performed similarly well when using the ResNet18. From the results in this work, it might be most beneficial to fuse the radars' data as well as the data representations at the signal level. This configuration gave the highest overall classification accuracy as well as the best performance for fall detection when using the AUC metric.

For the output thresholds, it turned out that a value of 0.2 might be a good compromise between high fall detection rate and low false alarm rate—this assuming that fall detection is the most important task in the analysis of the activities.

With the aforementioned configuration (summation fusion of the radars' data, signal-level fusion of the three data representations, ResNet18 as classifier, and threshold 0.2), the overall accuracy for activity recognition amounts to 94.98%. A fall detection rate of 82.39% could be reached and a false alarm rate of 4.4 %. For real-world applications, however, this is still a rather low value. Therefore, future research will have to find ways to improve fall detection performance still more.

Another question in a practical setting is how and where to compute the classification. Edge computing would be a desired solution, but depending on the data and network size, it might not always be applicable. As can be seen from the results, deeper networks perform better; however, they require more computational power and memory hardware for their higher number of key parameters. Therefore, finding efficient solutions to this problem is another key challenge on the way to applicability of the technique.

REFERENCES

[1] K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andrés, "From fall detection to fall prevention: A generic classification of fall-related systems," *IEEE Sensors J.*, vol. 17, no. 3, pp. 812–822, Feb. 2017, doi: 10.1109/JSEN.2016.2628099.

[2] R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Multipath exploitation for human activity recognition using a radar network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5103013, doi: 10.1109/TGRS.2024.3363631.

[3] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, Jul. 2019, doi: 10.1109/MSP.2018.2890128.

[4] J.-C. Chiao et al., "Applications of microwaves in medicine," *IEEE J. Microw.*, vol. 3, no. 1, pp. 134–169, Jan. 2023, doi: 10.1109/JMW.2022.3223301.

[5] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, "SimHumalator: An open-source end-to-end radar simulator for human activity recognition," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 3, pp. 6–22, Mar. 2022, doi: 10.1109/MAES.2021.3138948.

[6] X. Yang, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Radar-based human activities classification with complex-valued neural networks," in *Proc. IEEE Radar Conf. (RadarConf)*, Mar. 2022, pp. 1–6, doi: 10.1109/RadarConf2248738.2022.9763903.

[7] A. Dey, S. Rajan, G. Xiao, and J. Lu, "Fall event detection using vision transformer," in *Proc. IEEE Sensors*, Dallas, TX, USA, Oct. 2022, pp. 1–4, doi: 10.1109/SENSORS52175.2022.9967352.

[8] S. Waqar, M. Muaaz, and M. Pätzold, "Direction-independent human activity recognition using a distributed MIMO radar system and deep learning," *IEEE Sensors J.*, vol. 23, no. 20, pp. 24916–24929, Oct. 2023, doi: 10.1109/JSEN.2023.3310620.

[9] S. Zhu, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Continuous human activity recognition with distributed radar sensor networks and CNN–RNN architectures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5115215.

[10] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "Bi-LSTM network for multimodal continuous human activity recognition and fall detection," *IEEE Sensors J.*, vol. 20, no. 3, pp. 1191–1201, Feb. 2020.

[11] S.-W. Kang, M.-H. Jang, and S. Lee, "Identification of human motion using radar sensor in an indoor environment," *Sensors*, vol. 21, no. 7, p. 2305, Mar. 2021.

[12] E. Kurtoglu, A. C. Gurbuz, E. A. Malaia, D. Griffin, C. Crawford, and S. Z. Gurbuz, "ASL trigger recognition in mixed Activity/Signing sequences for RF sensor-based user interfaces," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 4, pp. 699–712, Aug. 2022.

[13] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "A survey on radar-based continuous human activity recognition," *IEEE J. Microw.*, vol. 3, no. 3, pp. 938–950, Jul. 2023, doi: 10.1109/JMW.2023.3264494.

[14] L. Cao, S. Liang, Z. Zhao, D. Wang, C. Fu, and K. Du, "Human activity recognition method based on FMCW radar sensor with multi-domain feature attention fusion network," *Sensors*, vol. 23, no. 11, p. 5100, May 2023, doi: 10.3390/s23115100.

[15] N. Kruse, R. Guendel, F. Fioranelli, and A. Yarovoy, "Segmentation of micro-Doppler signatures of human sequential activities using Rényi entropy," in *Proc. Int. Conf. Radar Syst. (RADAR)*, Oct. 2022, pp. 435–440, doi: 10.1049/icp.2022.2357.

[16] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "Radar-based continuous human activity recognition with multi-label classification," in *Proc. IEEE Sensors*, Nov. 2023, pp. 1–4, doi: 10.1109/SENSORS56945.2023.10324957.

[17] R. G. Guendel, M. Unterhorst, F. Fioranelli, and A. Yarovoy, "Dataset of continuous human activities performed in arbitrary directions collected with a distributed radar network of five nodes," 4TU.ResearchData, 2021, doi: 10.4121/16691500.

[18] R. G. Guendel, M. Unterhorst, E. Gambi, F. Fioranelli, and A. Yarovoy, "Continuous human activity recognition for arbitrary directions with distributed radars," in *Proc. IEEE Radar Conf.*, May 2021, pp. 1–6, doi: 10.1109/RadarConf2147009.2021.9454972.

[19] T. Ridnik et al., "Asymmetric loss for multi-label classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 82–91, doi: 10.1109/ICCV48922.2021.00015.

[20] R. G. Guendel, F. Fioranelli, and A. Yarovoy, "Distributed radar fusion and recurrent networks for classification of continuous human activities," *IET Radar Sonar Navig.*, vol. 16, no. 7, pp. 1144–1161, 2022, doi: 10.1049/rsn2.12249.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[22] F. Ramzan et al., "A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks," *J. Med. Syst.*, vol. 44, no. 2, pp. 1–16, Feb. 2020, doi: 10.1007/s10916-019-1475-2.

[23] J. Brownlee, *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Melbourne, VIC, Australia: Machine Learning Mastery.2018

[24] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

[25] S. Z. Gurbuz, *Deep Neural Netwrok Design for Radar Applications*. London, U.K.: Institution of Engineering and Technology, 2020.

[26] T. Fawcett, "ROC Graphs: Notes and Practical Considerations for Data Mining Researchers," *Mach. Learn.*, vol. 31, no. 1, pp. 1–38, 2004.