

Article

MDSCNet: A Lightweight Radar Image-Based Model for Multi-Action Classification in Elderly Healthcare

Xiangbo Kong ^{1,*}, Kenshi Saho ² and Akari Takebayashi ³¹ Department of Intelligent Robotics, Faculty of Information Engineering, Toyama Prefectural University, Imizu, Toyama 939-0398, Japan² Department of Electronic and Computer Engineering, College of Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan; saho@fc.ritsumei.ac.jp³ Graduate School of Engineering, Toyama Prefectural University, Imizu, Toyama 939-0398, Japan; u554018@st.pu-toyama.ac.jp

* Correspondence: kong@pu-toyama.ac.jp

Abstract

This study presents MDSCNet, a compact radar image-based deep learning model for multi-action classification in elderly healthcare scenarios. Motivated by the need for real-time deployment on resource-constrained devices, MDSCNet employs a streamlined architecture with a small number of lightweight expansion–depthwise–projection blocks, removing complex attention and squeeze-and-excitation modules to minimize computational overhead. The model is evaluated on a millimeter-wave radar dataset covering five healthcare-related actions: lying, sitting, standing, bed-exit, and falling, performed by 15 participants on an actual electric nursing bed. The experimental results demonstrate that MDSCNet achieves accuracy comparable to state-of-the-art CNN-based methods while maintaining an extremely compact model size of only 0.29 MB, showing its suitability for practical elderly care applications where both accuracy and efficiency are critical.

Keywords: radar image; elderly healthcare; multi-action classification

Academic Editor: Chiara Iacovelli

Received: 17 August 2025

Revised: 20 October 2025

Accepted: 30 October 2025

Published: 31 October 2025

Citation: Kong, X.; Saho, K.; Takebayashi, A. MDSCNet: A Lightweight Radar Image-Based Model for Multi-Action Classification in Elderly Healthcare. *Inventions* **2025**, *10*, 98. <https://doi.org/10.3390/inventions10060098>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The global elderly population is increasing rapidly. According to the U.S. Census Bureau's 2017 National Population Projections, the number of elderly people in the United States was 49.2 million in 2016, accounting for 15% of the total population, and it is projected to reach 94.7 million by 2060, representing 23% of the total [1]. Data from the State Council of the People's Republic of China indicate that the population aged 60 years and above in China reached 300 million by 2025, accounting for one-fifth of the total population, will exceed 400 million by 2033, accounting for approximately one-quarter, and will reach 487 million by around 2050, representing roughly one-third of the total [2]. According to Eurostat, in 2024, the European Union's population was estimated at 449.3 million, of which more than one-fifth were aged 65 years or older, corresponding to slightly more than three persons of working age for every person aged 65 years and over [3]. Data from the Cabinet Office of Japan show that the population aged 65 and over in Japan is 36.24 million, representing 29.0% of the total population, and it is expected to rise to 39.53 million by 2043 [4]. Furthermore, the World Health Organization projects that the proportion of people aged 60 years and above worldwide will nearly double, from 12% in 2015 to 22% by 2050 [5].

With the increase in the elderly population, the number of elderly individuals requiring care has also risen. Taking Japan as an example, the number of certified care recipients was 5.15 million in 2011 and increased to 6.76 million in 2021, representing a rise of 1.61 million [4]. This steady growth in the elderly population has led to a continual increase in the number of individuals certified as requiring long-term care. Meanwhile, due to the progression of declining birth rates, the number of newly available nurses and caregivers is expected to decrease compared with previous years. As a result, the workload borne by each nurse and caregiver has become progressively heavier. Given this situation, continuous 24-h monitoring in elderly welfare facilities and similar institutions is necessary to confirm whether care recipients are lying in bed and to prevent incidents such as falls from bed or wandering. This requirement becomes a major contributing factor to the heavy burden placed on nurses and caregivers.

Since fall accidents cause serious harm to the elderly [6], fall events require special attention in elderly action recognition. Therefore, many related studies focus specifically on fall detection. These methods can generally be categorized into five groups: RGB image-based approaches [7,8], depth image-based approaches [9,10], wearable device-based approaches [11,12], pressure sensor-based approaches [13,14], and radar-based approaches [15,16]. These methods have been widely studied and have achieved significant results, with some already being commercialized [17]. However, the main limitation of these works is that they focus solely on the binary problem of fall detection. In real healthcare scenarios, caregivers also need to be aware of other states of the elderly, such as long-sitting, standing up, or exiting. Therefore, from a practical perspective, multiclass classification problems also need to be investigated.

Since falls can cause severe harm to elderly individuals, fall detection has been widely studied in the field of elderly healthcare action classification, and these methods can generally be categorized into five groups: RGB image-based approaches, depth image-based approaches, wearable device-based approaches, pressure sensor-based approaches, and radar-based approaches.

Due to the low cost of monocular cameras and the recent advances in image processing techniques, many studies investigate fall detection based on RGB images. The study in [7] proposes a fall detection approach using multivariate exponentially weighted moving average (MEWMA) and support vector machine (SVM) approaches, validated on the University of Rzeszow fall detection dataset (URFD) and the fall detection dataset (FDD), showing improved discrimination of real falls from fall-like actions. The paper in [8] proposes a toddler fall detection framework combining GELAN-integrated YOLOv8 and Generalized Hough Transform, achieving 96.33% accuracy and outperforming related works in precision, recall, mAP, and F1. RGB image-based fall detection methods can achieve relatively high accuracy, and the use of color images facilitates visual confirmation. However, approaches based on RGB images are sensitive to lighting conditions and often fail to capture correctly in nighttime environments, even though falls may also occur under such dark conditions. In addition, methods relying on RGB images may raise privacy concerns.

Depth image-based methods are less affected by lighting conditions and achieve higher accuracy because they include distance information. The study in [9] presents a fall detection approach using a single depth camera and an enhanced randomized decision tree (RDT) for joint extraction, combined with an SVM classifier, achieving 11.8% higher accuracy than state-of-the-art methods and robustness under dark environments. The approach in [10] proposes a depth camera-based fall detection system using convolutional neural networks (CNNs). However, most of the depth image-based methods focus only on binary fall detection and do not address multiclass classification, which is required in elderly healthcare. Moreover, they often do not use data captured from real nursing beds.

Wearable-based fall detection methods are widely studied and commercially available [17] due to their convenience. The study in [11] uses a low-complexity finite-state machine algorithm, distinguishing daily activities from four types of falls and achieving 97.9% sensitivity, 99.9% specificity, and 99.7% overall accuracy on 6750 samples, demonstrating strong applicability for long-term outdoor use. The paper in [12] presents a fall detection system using consumer smartwatches and smartphones with edge computing. However, these devices require the elderly to wear them continuously, and they fail to operate if the elderly forget to charge or wear them. This limitation makes such systems less user-friendly for elderly individuals, who are prone to forgetfulness.

Fall detection methods based on pressure sensors are studied because they do not require the elderly to wear any device and are not subject to occlusion, as in camera-based approaches. The study in [13] proposes PIFall, a pressure insole-based fall detection system using a ResNet(2+1)D model. The system achieves 91% overall accuracy in fall detection and 94% in classifying specific fall actions, demonstrating both feasibility and user acceptance among elderly participants. The approach in [14] involves a smart carpet-based fall detection system using differential piezoresistive pressure sensors. Prototype evaluation achieves 88.8% sensitivity and 94.9% specificity, showing potential for deployment in home care and integration into beds for fall prevention during sleep. While these methods are suitable for home use, their applicability in clinical or caregiving environments is limited as they struggle to accurately detect essential activities of the elderly, such as lying down or long-sitting.

In the healthcare domain, some foundational radar-based studies investigate binary classification between falls and non-falls. The study in [15] proposes four fall detection systems using FMCW mmWave radar with x-y scatter and Doppler range images. By integrating an attention mechanism into Long Short-Term Memory (LSTM) architectures, the HOG-LSTM-Attention model achieves the best performance, reaching 95.3% accuracy and 95.5% F1-score, effectively distinguishing falls from daily activities. The work in [16] presents a non-contact fall detection system combining 4D imaging radar sensors with AI. Using point cloud data and a CNN for posture classification, the system achieves 98.66% accuracy in posture recognition and 95% in fall detection, offering a privacy-preserving alternative to wearable and camera-based solutions. Although these methods achieve good results, they only focus on binary classification of falls versus non-falls. In actual healthcare settings, additional activities, such as long-sitting and exiting, also need to be detected and reported to caregivers in order to reduce their workload. Some advanced investigations further explore radar-based multi-action classification. The work in [18] proposes a human action recognition method using FMCW radar based on micro-Doppler spectrograms and an enhanced ResNet18 architecture. By integrating asymmetric convolution, the Mish activation function, and an improved convolutional block attention module (ICBAM), the method achieves 98.28% accuracy, demonstrating superior performance over conventional deep learning models and strong robustness in noisy environments. The study in [19] explores radar-based human activity recognition by combining three radar preprocessing techniques with four CNN architectures (VGG-16, VGG-19, ResNet-50, and MobileNetV2). Among the twelve evaluated configurations, MobileNetV2 with STFT achieves the best trade-off, reaching 96.3% accuracy with low inference latency, highlighting the potential of radar spectrograms as visual features for real-time edge-deployable HAR systems. Previous work [20] demonstrates a 60 GHz FMCW radar-based bed monitoring system that classifies behaviors such as leaving, lying, and sitting on the bed using CNNs on time-range images. With MobileNet, bed-leaving is detected with 95.9% accuracy, and the overall accuracy across six representative behaviors is 83.1%, highlighting the potential of mmWave radar as a non-invasive tool for nursing care, although further improvements are needed for

multi-behavior recognition. However, the existing studies still have limitations. The work in [18] is not designed for nursing-care scenarios. Ref. [19] does not employ data captured on real nursing beds, and the adopted model is relatively large, which hinders deployment on edge devices in practical applications. Radar data can take various forms, including Range–Doppler (RD) maps, Range–Angle (RA)/Range–Azimuth heatmaps, and point clouds. Although RA heatmaps and point clouds offer their own advantages, this study, similar to related works, adopts RD maps as the primary representation considering the trade-offs in data volume and the constraints of edge computing.

In radar image-based action classification methods, CNNs are often required. Since the introduction of AlexNet [21], CNN-based classification methods have been widely adopted. GoogLeNet [22] explores parallel structures to improve accuracy, but the network becomes excessively large. While deeper networks generally achieve higher accuracy, their development is constrained by issues such as vanishing and exploding gradients. ResNet [23] addresses this problem through the introduction of residual connections, which enable substantially deeper architectures and further performance improvements. Compared with ResNet, ResNeXt [24] improves accuracy by introducing a multi-branch architecture with a unified topology, while maintaining a comparable computational cost. However, despite their accuracy gains, these models remain relatively large, making them unsuitable for deployment on edge devices. To mitigate this, MobileNet [25–27] employs depthwise-separable convolutions to significantly reduce model size, achieving wide adoption on mobile and edge platforms. ShuffleNet [28,29] further enhances model compactness through group convolution and channel shuffle techniques, although at the cost of reduced accuracy. More recently, with the incorporation of attention mechanisms, Vision Transformer (ViT) [30] achieves remarkable success in terms of accuracy. Nevertheless, ViT requires extremely large-scale training data and has a very large model size, which limits its applicability for deployment on edge devices.

In this study, a multiclass database is constructed using millimeter-wave radar recordings from 15 participants, which includes five actions: lying down, long-sitting, standing, exiting, and falling. Although a few related studies investigate radar-based multiclass action recognition for elderly healthcare, they have two limitations. First, they do not use data recorded with actual nursing-care beds. Second, the employed models are too large in size, which limits their deployment on edge devices in real applications. In this study, all data are recorded using a real electric nursing-care bed, and we develop a lightweight model, MDSCNet, with a size of only 0.29 MB. Despite its compactness, MDSCNet achieves higher accuracy than ShuffleNetV2 (5.45 MB) and only slightly lower accuracy than MobileNetV3 Small (5.93 MB) and MobileNetV3 Large (16.2 MB). The contributions of this study are as follows:

- Previous studies do not use data recorded with real nursing-care beds. In this study, we record radar data from 15 subjects using an actual electric nursing-care bed.
- Previous studies often focus only on binary classification problems such as fall detection, whereas real applications require the recognition of more postures. Based on practical nursing scenarios, this study performs a multiclass classification of five actions: lying down, long-sitting, standing, exiting, and falling.
- In real applications, lightweight models are required for deployment on edge devices. Based on extensive ablation experiments, we develop a lightweight network, MDSCNet, with a size of only 0.29 MB, which is significantly smaller than other models while achieving comparable accuracy. Furthermore, to ensure the reliability of the results, this work performs extensive cross-validation experiments. All the baseline comparisons and ablation experiments are evaluated using cross-validation of leave-one-subject-out (LOSO).

The rest of this paper is organized as follows. Section 2 presents the proposed method. Section 3 describes the experiments and experimental results. Section 4 concludes the paper.

2. Methods

Based on the discussion in the preceding chapter, this work proposes MDSCNet, a lightweight network designed for radar image-based classification tasks in nursing-care environments, as shown in Figure 1. The study in [31] shows that, for small-scale datasets, parallel-branch architectures often result in larger models with suboptimal accuracy. To reduce model size, MDSCNet adopts a sequential architecture similar to ResNet. Moreover, since max-pooling layers typically lead to information loss, the Stem stage of our network employs strided convolutions with a stride of 2 for downsampling.

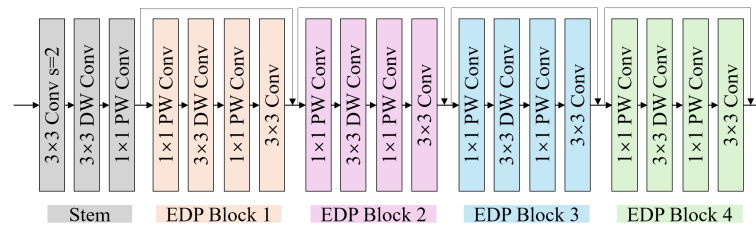


Figure 1. Network architecture of MDSCNet, where PW Conv denotes pointwise convolution and DW Conv denotes depthwise convolution. The figure illustrates the case where the network employs four expansion–depthwise–projection (EDP) blocks with residual connections. In the ablation studies, we evaluated different configurations by varying the number of EDP blocks from 2 to 7 and by comparing the performance with and without residual connections, validating both classification accuracy and model size.

To balance accuracy and efficiency, MDSCNet adopts depthwise-separable convolutions. A standard convolution with kernel size $K \times K$ has a computational cost of

$$\text{FLOPs}_{\text{standard}} = K^2 \cdot M \cdot N \cdot H \cdot W, \quad (1)$$

where M and N are the input and output channels, and $H \times W$ is the spatial dimension of the output feature map.

In contrast, a depthwise-separable convolution decomposes the operation into a depthwise and a pointwise convolution:

$$\text{FLOPs}_{\text{DW+PW}} = K^2 \cdot M \cdot H \cdot W + M \cdot N \cdot H \cdot W, \quad (2)$$

significantly reducing the computational burden when K is small relative to N .

While depthwise-separable convolutions from MobileNet provide an efficient alternative to standard convolutions, this work specifically adopts the EDP block to balance efficiency and representational capacity. The expansion layer increases the channel dimensionality, allowing depthwise convolutions to capture richer spatial patterns in a higher-dimensional feature space. The subsequent projection layer compresses the channels back to a compact representation, thereby reducing redundancy and preserving efficiency. This structure, validated in prior lightweight models such as MobileNetV2/V3, is particularly suitable for our radar-based elderly healthcare dataset, where the limited data size requires compact models with sufficient expressive power. Our ablation results further confirm that EDP blocks, especially when combined with and without residual connections, enable deeper architectures without incurring significant model size or degradation issues.

Implementation details. In our implementation, the Stem stage consists of a 3×3 convolution with stride 2 producing 16 channels, followed by a depthwise 3×3 convolution and a pointwise 1×1 convolution to produce a 32-channel feature map. The

main body then applies K EDP blocks (we vary K from 2 to 7 in ablations). The first EDP block uses stride 2 and an expansion ratio of 4. The remaining $K-1$ blocks use stride 1 and an expansion ratio of 3, all operating at 32 channels. Residual connections are enabled only when the input and output have identical shape and no projection shortcut is used. In the Tail, a 1×1 pointwise convolution expands channels from 32 to 64, followed by global average pooling and a two-layer MLP head ($64 \rightarrow 128 \rightarrow C$) for classification.

In the expansion phase, the input feature map of M channels is first expanded to αM channels through a 1×1 pointwise convolution:

$$X_{\text{exp}} = \sigma(X * W_{1 \times 1}^{\text{exp}}), \quad (3)$$

where α is the expansion ratio, $\sigma(\cdot)$ denotes a nonlinear activation, and $W_{1 \times 1}^{\text{exp}}$ is the weight matrix.

The expanded feature map is then processed by a depthwise convolution, which applies a separate spatial filter to each channel independently:

$$y_{i,j}^{(m)} = \sum_{u=1}^K \sum_{v=1}^K x_{i+u,j+v}^{(m)} \cdot W_{u,v}^{(m)}, \quad (4)$$

where $K \times K$ is the kernel size. This step enables efficient extraction of spatial features while keeping the number of parameters low.

Finally, the output of the depthwise convolution is projected back to N channels by another 1×1 convolution:

$$X_{\text{out}} = X_{\text{dw}} * W_{1 \times 1}^{\text{proj}}. \quad (5)$$

This stage reduces the dimensionality while preserving useful features.

To alleviate the vanishing gradient problem in deep networks and to enhance information flow, the EDP block incorporates a residual connection. When the input and output feature maps have the same number of channels and the stride is 1, the input X is directly added to the projected output X_{out} :

$$Y = X_{\text{out}} + X. \quad (6)$$

This design not only preserves the low-dimensional information from the input but also enables direct information propagation through identity mapping, thereby improving training stability and model performance.

It should be noted that Figure 1 illustrates the case with four EDP blocks for clarity of presentation. In our experiments, however, the ablation studies systematically vary the number of blocks from 2 to 7 to evaluate the impact of network depth on accuracy and model size.

3. Results

3.1. Experimental Environment

Although prior studies investigate action classification in the context of elderly health-care, they often rely on datasets that are not collected in authentic care environments. In contrast, as illustrated in Figure 2, the dataset in this study is acquired using a real electric nursing-care bed. The experimental setup employs an SC1220AT2-B-113 60 GHz millimeter-wave radar sensor [32]. The experimental environment for the ablation studies is configured as follows. The CPU is an Intel Core i9-13900KF, and the GPU is an NVIDIA RTX 6000 Ada with 48 GB of GDDR6 memory. The system is equipped with 128 GB of RAM. The software

environment uses Python version 3.10.12 and PyTorch version 2.2.1+cu121. This work uses the Adam optimizer with an initial learning rate of 0.001, and a StepLR scheduler with a step size of 10 and decay factor 0.1. The batch size is fixed at 32 for both training and testing. The number of epochs is set to 100 for all the models to ensure convergence. For fairness, all the compared models, including ResNet18, ResNext50, MobileNetV3-Small/Large, ShuffleNetV2, ViT, and the proposed MDSCNet, are trained under the same hyperparameter settings.



Figure 2. Nursing-care bed used for data acquisition in this study. The radar sensor was fixed at the head of the bed, positioned 60 cm above the mattress surface.

3.2. Dataset and Implementation Details

Although prior studies have generated datasets for fall detection or action classification in elderly healthcare, these works often either restrict the task to a binary fall/non-fall setting, omitting other actions required in caregiving practice, or collect data without using a real nursing-care bed. To address these gaps, this work records a dataset in a realistic environment using a real electric nursing-care bed. Range (distance) information is measured by a Frequency-Modulated Continuous Wave (FM-CW) radar from 15 subjects, with each action maintained for 70 s in a single trial [31]. The five target actions are as follows: lying in bed; long-sitting (sitting upright with the legs extended forward); standing near the bed; bed-exit (the subject located away from the bed); and falling. For the fall class, this work further annotates three representative configurations: lying on the floor outside the bed parallel to the bed, long-sitting on the floor outside the bed, and a diagonally collapsed posture on the floor outside the bed. These three specific postures are uniformly distributed in both the training and testing sets. In consideration of research ethics and safety, the study recruited university students aged 18–25 as participants rather than elderly persons.

Representative radar images corresponding to each action class are presented in Figure 3, where panels (a) through (e) depict lying in bed, long-sitting, standing, falling, and exiting, respectively.

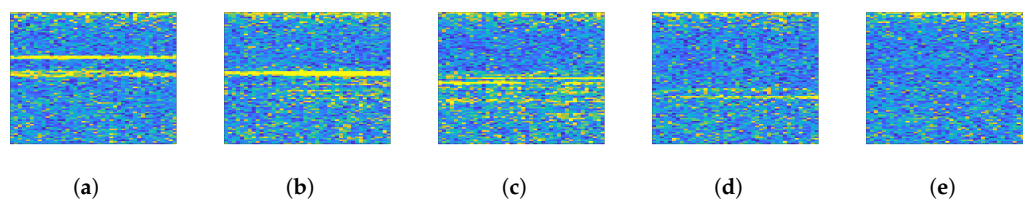


Figure 3. Radar images of different actions: (a) lying on the bed, (b) sitting, (c) standing, (d) falling, and (e) exiting.

To ensure reliability, the LOSO cross-validation strategy is employed. In Fold01, the data from Subjects 1–14 are used for training, and the data from Subject 15 are used for testing. In Fold02, the data from Subjects 1–13 and Subject 15 are used for training, and the data from Subject 14 are used for testing. This process is repeated until each subject has been used once as the test set. The final experimental result is obtained by averaging the results from all 15 folds, as shown in Figure 4.

	Test	Train													
Fold01	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold02	Subject 2	Subject 1	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold03	Subject 3	Subject 1	Subject 2	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold04	Subject 4	Subject 1	Subject 2	Subject 3	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold05	Subject 5	Subject 1	Subject 2	Subject 3	Subject 4	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold06	Subject 6	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold07	Subject 7	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold08	Subject 8	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold09	Subject 9	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold10	Subject 10	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Fold11	Subject 11	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 12	Subject 13	Subject 14	Subject 15
Fold12	Subject 12	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 13	Subject 14	Subject 15
Fold13	Subject 13	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 14	Subject 15
Fold14	Subject 14	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 15
Fold15	Subject 15	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14

Figure 4. LOSO cross-validation used in the experiments. In each fold, one subject is designated as the test set, while the remaining subjects are used for training. Since the dataset includes 15 subjects, the cross-validation is repeated 15 times, and the mean accuracy across all folds is reported as the evaluation metric.

3.3. Evaluation Metrics

Following the definition in [33], this study employs per-class accuracy as the evaluation metric for classification performance. The confusion matrix is defined as follows: true positive (TP) denotes the number of samples correctly identified as the target class, true negative (TN) represents the number of samples correctly identified as not belonging to the target class, false positive (FP) refers to the number of samples incorrectly predicted as the target class, and false negative (FN) corresponds to the number of samples incorrectly predicted as not belonging to the target class. The per-class accuracy for class i is computed using Equation (7):

$$\text{Accuracy}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \quad (7)$$

where the numerator indicates the number of samples correctly classified as class i , and the denominator corresponds to the total number of ground-truth samples in class i . It is noteworthy that, in a multiclass classification setting, this definition of per-class accuracy is mathematically equivalent to recall since the denominator reflects the total population of samples belonging to class i .

3.4. Ablation Study

The ablation study is conducted to investigate the optimal network architecture components in MDSCNet. Specifically, it aims to determine whether the performance improvements arise from the introduction of the EDP block, the presence of residual connections, or the combined effect of both. Furthermore, by analyzing the classification accuracy and model size under different configurations, the ablation study provides insight into the optimal number of EDP blocks required to balance accuracy and computational efficiency. In the ablation experiments, the number of EDP blocks is varied from two to seven, and the effect of residual connections is examined by comparing models with and without skip connections. All the other hyperparameters, including learning rate, batch size, and optimizer settings, are kept constant to ensure fairness. The evaluation is based on per-class

accuracy under leave-one-subject-out cross-validation, while the model size is also reported to assess efficiency. Table 1 summarizes the results of the ablation study. From the experimental results, it is observed that classification accuracy consistently improves as network depth increases, reaching its peak when the number of blocks is set to six. At this depth, where the network exceeds 20 layers, the incorporation of residual connections effectively mitigates degradation issues in deep architectures, thereby further enhancing performance. Therefore, in the subsequent experimental results section, we select the model with residual connections and six EDP blocks as our proposed architecture and compare it against other baseline algorithms, including ResNet18 used in [18] and MobileNet employed in [19].

Table 1. Per-class accuracy (%) for ablation study under different block numbers, with and without residual connections. The best result for each class is in bold. The symbol “✓” denotes the use of residual connections, whereas “×” denotes that residual connections is not used.

Blocks	Residual	Lying	Sitting	Standing	Exiting	Falling	Average
2	✓	83.21	73.33	76.99	81.73	67.21	76.49
3	✓	82.22	74.61	78.62	79.56	70.52	77.10
4	✓	85.97	74.85	80.44	77.04	71.95	78.05
5	✓	87.21	77.52	80.25	75.11	72.10	78.44
6	✓	87.01	78.35	80.79	77.14	72.20	79.10
7	✓	86.42	76.53	78.81	76.05	72.15	77.99
2	×	86.66	71.06	75.80	82.62	66.37	76.50
3	×	85.93	72.13	77.23	79.90	69.68	76.97
4	×	85.88	74.16	80.49	77.28	70.62	77.69
5	×	86.12	73.31	79.46	76.00	72.15	77.41
6	×	85.68	79.00	79.36	76.35	72.59	78.59
7	×	87.45	75.05	80.54	76.10	71.21	78.07

3.5. Experimental Results and Discussion

From the results in Table 2, lightweight CNN-based models such as MobileNetV3-Small (79.95%, 0.13 GFLOPs) and ShuffleNetV2 (78.80%, 0.29 GFLOPs) achieve competitive performance [31]. The proposed MDSCNet attains a comparable average accuracy of 79.10% with 0.40 GFLOPs, remaining in the same computational order as these lightweight baselines. Although MobileNetV3-Large achieves the highest average accuracy of 80.49%, its computational cost (0.44 GFLOPs) and model size (16.20 MB) are considerably higher, with the latter being more than 55 times larger than MDSCNet (0.29 MB). Furthermore, MDSCNet allows flexible adjustment of the number of blocks, enabling trade-offs between accuracy and efficiency depending on the available edge computing resources.

It is observed that the ViT approach achieves significantly lower accuracy compared with CNN-based models. We attribute this to two possible reasons. First, ViT relies heavily on large-scale training data due to the lack of convolutional inductive biases. In contrast, our dataset is relatively small, consisting of only 15 participants and five action classes, which is insufficient to fully exploit the capacity of ViT. Second, radar images differ substantially from natural images, with lower resolution, higher noise, and distinctive spatial patterns. While CNNs are effective in capturing local spatial features that are critical in radar-based classification, ViT focuses on modeling global dependencies, which may not align well with radar image characteristics. For future work, we plan to investigate data augmentation and synthetic radar data generation to alleviate the data scarcity issue, as well as explore hybrid architectures (e.g., CNN–Transformer models) and localized attention mechanisms better suited to radar signals.

It should be noted that the detection accuracy for the “falling” class is lower than that for the other classes, with an accuracy of only 72.20%, of which 23.06% are misclassified as

“exiting.” Similar tendencies are also observed in other baseline algorithms, indicating that this is not unique to MDSCNet. We hypothesize that the main reason lies in the physical properties of radar reflection: compared with upright postures such as standing, sitting, or exiting, a fallen body lies close to the ground with a substantially reduced overall height. As radar systems rely on the reflection of electromagnetic waves, low-lying targets generally exhibit a smaller radar cross-section (RCS) than upright bodies. Moreover, when the body lies parallel to the ground after a fall, its reflection pattern tends to be weak or ambiguous and may resemble background clutter, which explains the frequent misclassification as “exiting.” To address this issue, future work will investigate multimodal approaches, such as radar–RGB fusion, and the integration of multiple radars placed at different viewpoints. These strategies may help to capture complementary spatial information and enhance the robustness of fall detection in real-world healthcare environments.

In terms of model size, MDSCNet is by far the most compact architecture, requiring only 0.29 MB while still maintaining accuracy on par with state-of-the-art methods. For example, compared to MobileNetV3-Small (5.93 MB), the proposed model achieves nearly identical performance while reducing the storage requirements by more than 95%. Similarly, compared with ShuffleNetV2 (5.45 MB), MDSCNet reduces the model size by about 95% yet provides slightly higher accuracy.

This remarkable balance between efficiency and accuracy demonstrates the effectiveness of the proposed architecture. The compact size makes MDSCNet highly suitable for deployment in resource-constrained environments, such as wearable healthcare devices or embedded systems, where memory and computational resources are limited.

Table 2. Per-class accuracy (%), model size, and GFLOPs for different models. The best result for each class is in bold.

Model	Lying	Sitting	Standing	Exiting	Falling	Average	Model Size (MB)	GFLOPs
ResNet18	84.63	76.98	80.25	77.33	75.70	78.98	42.70	1.80
ResNeXt50	83.50	75.59	78.72	75.56	76.10	77.89	88.00	8.50
MobileNetV3-Small	86.81	77.87	84.64	72.99	77.43	79.95	5.93	0.13
MobileNetV3-Large	87.25	79.05	84.20	74.52	77.43	80.49	16.20	0.44
ShuffleNetV2	87.70	75.00	81.04	75.01	75.26	78.80	5.45	0.29
ViT	64.97	58.70	63.46	78.96	52.05	63.63	327.00	35.13
MDSCNet (proposed)	87.01	78.35	80.79	77.14	72.20	79.10	0.29	0.40

4. Conclusions

In this study, we propose a lightweight MDSCNet model and validate its effectiveness for radar image-based action classification. The ablation study demonstrates that incorporating residual connections significantly enhances performance when network depth increases. The experimental results show that the proposed model achieves competitive classification accuracy compared to mainstream deep learning models while maintaining an extremely compact model size of only 0.29 MB. This balance between accuracy and efficiency highlights the practical value of our approach. Although MDSCNet relies on well-established techniques, such as depthwise-separable convolutions and residual connections, due to the strict computational constraints of industrial edge devices, future work will explore integrating more recent advances, including efficient attention mechanisms, dynamic channel pruning, and radar-specific modules, to further enhance performance while maintaining deployability.

Author Contributions: Conceptualization, X.K. and K.S.; methodology, X.K.; software, X.K.; validation, X.K.; formal analysis, X.K.; investigation, X.K. and A.T.; data curation, X.K. and A.T.;

writing—original draft preparation, X.K.; writing—review and editing, X.K. and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable to this article due to the involvement of commercially available products and user privacy-sensitive data.

Acknowledgments: The authors acknowledge the use of ChatGPT (OpenAI, version GPT-5) to support the preparation of this manuscript, specifically for language refinement and technical review, code review, and debugging support. All generated content was critically reviewed and revised by the authors, who accept full responsibility for the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDSCNet	Multi-Depthwise-Separable Convolution Network
EDP Block	Expansion-Depthwise-Projection Block
CNN	Convolutional Neural Network
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

References

1. Vespa, J.; Medina, L.; Armstrong, D.M. Demographic Turning Points for the United States: Population Projections for 2020 to 2060. Available online: <https://www.census.gov/content/dam/Census/library/publications/2020/demo/p25-1144.pdf> (accessed on 13 August 2025).
2. The State Council of the People's Republic of China. Available online: https://www.gov.cn/xinwen/2018-07/19/content_5307839.htm (accessed on 13 August 2025).
3. Eurostat. Population Structure and Ageing. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing (accessed on 13 August 2025).
4. Cabinet Office, Government of Japan. Available online: <https://www8.cao.go.jp/kourei/whitepaper/w-2023/html/zenbun/index.html> (accessed on 13 August 2025).
5. World Health Organization. Ageing and Health. Available online: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (accessed on 13 August 2025).
6. World Health Organization. Falls. Available online: <https://www.who.int/news-room/fact-sheets/detail/falls> (accessed on 13 August 2025).
7. Harrou, F.; Zerrouki, F.; Sun, Y.; Houacine, A. Vision-based fall detection system for improving safety of elderly people. *IEEE Instrum. Meas. Mag.* **2017**, *20*, 49–55. [CrossRef]
8. Yang, Z.; Tsui, B.; Ning, J.; Wu, Z. Falling Detection of Toddlers Based on Improved YOLOv8 Models. *Sensors* **2024**, *24*, 6451. [CrossRef] [PubMed]
9. Bian, Z.-P.; Hou, J.; Chau, L.-P.; Magnenat-Thalmann, N. Fall Detection Based on Body Part Tracking Using a Depth Camera. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 430–439. [CrossRef] [PubMed]
10. Su, M.-C.; Chen, J.-H.; Hsieh, Y.Z.; Hsu, S.-C.; Liao, C.-W. Enhancing Detection of Falls and Bed-Falls Using a Depth Sensor and Convolutional Neural Network. *IEEE Sens. J.* **2024**, *24*, 23150–23162. [CrossRef]
11. Tseng, C.-K.; Huang, S.-J.; Kau, L.-J. Wearable Fall Detection System with Real-Time Localization and Notification Capabilities. *Sensors* **2025**, *25*, 3632. [CrossRef] [PubMed]
12. Leone, A.; Manni, A.; Rescio, G.; Siciliano, P.; Caroppo, A. Deep Transfer Learning Approach in Smartwatch-Based Fall Detection Systems. *Eng. Proc.* **2024**, *78*, 2.
13. Guo, W.; Liu, X.; Lu, C.; Jing, L. PIFall: A Pressure Insole-Based Fall Detection System for the Elderly Using ResNet3D. *Electronics* **2024**, *13*, 1066. [CrossRef]

14. Chaccour, K.; Darazi, R.; Hassans, A.H.; Andres, E. Smart carpet using differential piezoresistive pressure sensors for elderly fall detection. In Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Abu Dhabi, United Arab Emirates, 19–21 October 2015.
15. Yu, Y.S.; Wie, S.; Lee, H.; Lee, J.; Kim, N.H. Long Short-Term Memory-Based Fall Detection by Frequency-Modulated Continuous Wave Millimeter-Wave Radar Sensor for Seniors Living Alone. *Appl. Sci.* **2025**, *15*, 8381. [\[CrossRef\]](#)
16. Ahn, S.; Choi, M.; Lee, J.; Kim, J.; Chung, S. Non-Contact Fall Detection System Using 4D Imaging Radar for Elderly Safety Based on a CNN Model. *Sensors* **2025**, *25*, 3452. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Apple Inc. Use Fall Detection with Apple Watch. Available online: <https://support.apple.com/en-us/108896> (accessed on 13 August 2025).
18. Zhang, Y.; Tang, H.; Wu, Y.; Wang, B.; Yang, D. FMCW Radar Human Action Recognition Based on Asymmetric Convolutional Residual Blocks. *Sensors* **2024**, *24*, 4570. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Ayaz, F.; Alhumaily, B.; Hussain, S.; Imran, M.; Arshad, K.; Assaleh, K.; Zoha, A. Radar Signal Processing and Its Impact on Deep Learning-Driven Human Activity Recognition. *Sensors* **2025**, *25*, 724. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Hashimoto, S.; Kong, X.; Kamiya, K.; Saho, K. Behavior Classification for Bed Monitoring Using Short-Term 60 GHz-Band FMCW Radar Images. In Proceedings of the Advanced Technologies and Applications in the Internet of Things, Ceur Workshop Proceedings, Kusatsu, Siga, Japan, 19–21 October 2024.
21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Curran Associates Inc., Lake Tahoe, NV USA, 3–6 December 2012.
22. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
24. Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
25. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Wey, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [\[CrossRef\]](#)
26. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
27. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
28. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
29. Ma, N.; Zhang, X.; Zheng, H.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In Proceedings of the European Conference on Computer Vision, Springer, Munich, Germany, 8–14 September 2018.
30. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning, Curran Associates Inc., Virtual Conference, 3–7 May 2021.
31. Kong, X.; Urano, K.; Takebayashi, A. Radar-Based Action Classification for Elderly Healthcare in Caregiving Environment. In Proceedings of the International Conference on Advanced Mechatronic Systems, Xi'an, China, 19–22 September 2025.
32. SC1220AT2-B-113 60GHz Radar Kit. Available online: <https://www.socionext.com/jp/download/catalog/AD04-00141-1.pdf> (accessed on 13 August 2025).
33. Noury, N.; Fleury, A.; Rumeau, P.; Bourke, A.K.; Laighin, G.O.; Rialle, V. Fall Detection-Principles and Methods. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 22–26 August 2007.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.