Contactless Fall Detection Using Time-Frequency Analysis and Convolutional Neural Networks

Hamidreza Sadreazami, *Member, IEEE*, Miodrag Bolic, *Senior Member, IEEE*, and Sreeraman Rajan, *Senior Member, IEEE*

***Abstract*—Automatic detection of a falling person based on noncontact sensing is a challenging problem with applications in smart homes for elderly care. In this article, we propose a radar-based fall detection technique based on time-frequency analysis and convolutional neural networks. The time-frequency analysis is performed by applying the short-time Fourier transform to each radar return signal. The resulting spectrograms are converted into binary images, which are fed into the convolutional neural network. The network is trained using labeled examples of fall and nonfall activities. Our method employs high-level feature learning, which distinguishes it from previously studied methods that use heuristic feature extraction. The performance of the proposed method is evaluated by conducting several experiments on a set of radar return signals. We show that our method distinguishes falls from nonfalls with 98.37% precision and 97.82% specificity, while maintaining a low false-alarm rate, which is superior to existing methods. We also show that our proposed method is robust in that it successfully distinguishes falls from nonfalls when trained on subjects in one room, but tested on different subjects in a different room. In the proposed convolutional neural network, the hierarchical features extracted from the radar return signals are the key to understand the fundamental composition of human activities and determine whether or not a fall has occurred during human daily activities. Our method may be extended to other radar-based applications such as apnea detection and gesture detection.**

***Index Terms*—Convolutional neural network, fall detection, time-frequency analysis, ultrawideband (UWB) radar.**

I. Introduction

HUMAN activity recognition has become increasingly important in recent years as it enables automatic assessment of subjects' health and well-being [1]–[3]. Among the elderly, most injury-related hospitalizations are due to falls [4]. Therefore, accidental fall detection is an important subset of human activity recognition [5]–[7]. If no preventive measures are undertaken, the number of injuries due to falls can double by 2030 due to aging population [8]. Unfortunately, continuous 24-h human monitoring of elderly to prevent falls is impossible, and therefore, an automated way of monitoring them is needed in order to detect and prevent falls.

Activity-monitoring methods currently use inexpensive wearable sensors such as accelerometers [9]. With the emergence of smart homes, noncontact indoor monitoring using radar-based technology is gaining interest as radar signals can penetrate obstacles such as wall to locate targets [10]. In addition, radar-based methods do not violate the privacy of the monitored subjects [11] and preclude the need for wearing a sensor [2]. More recently, ultrawideband (UWB) radars have gained popularity as active monitoring sensors. UWB radars are more suited for activity and fall recognition because of robustness to multipath fading, higher penetration ability, finer time resolution [12], and higher spatial resolution when compared to continuous wave radars [13]–[17].

Radar-based monitoring methods use either a threshold [18] or learning-based approach [19] to detect falls. Threshold-based fall detection requires effective features or descriptors, and when such descriptors exceed preset thresholds during a fall, an alarm would be triggered. Learning-based approaches require models trained using features extracted from either the time or the frequency domain.

In [19], a human activity recognition method based on radar micro-Doppler data was designed by extracting features from time-velocity and cadence-velocity domains while in [20], a similar design was used to classify people. In [21], a human gait classification method was proposed using the motion signature from arm and leg movements. In [22], using micro-Doppler signature, different human activities were classified over an extended time duration, through a wall, and at multiple angles to the radar. A wavelet-based approach was devised in [23] for fall detection using Doppler radars. In [24], fall detection was achieved exploiting time-frequency characteristics of the radar

Manuscript received August 14, 2020; revised November 5, 2020 and December 3, 2020; accepted January 1, 2021. Date of publication January 5, 2021; date of current version June 30, 2021. Paper no. TII-20-3918. *(Corresponding author: Hamidreza Sadreazami.)*

Hamidreza Sadreazami and Miodrag Bolic are with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: hamidreza.sadreazami@mail.mcgill.ca; mbolic@eecs.uottawa.ca).

Sreeraman Rajan is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: sreeramanr@sce.carleton.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3049342.

Digital Object Identifier 10.1109/TII.2021.3049342

Doppler signatures and the events were classified by a sparse Bayesian classifier using the joint statistics of three different features. In all these methods, feature engineering had to be done by a human and the classification accuracy depended on the engineered features.

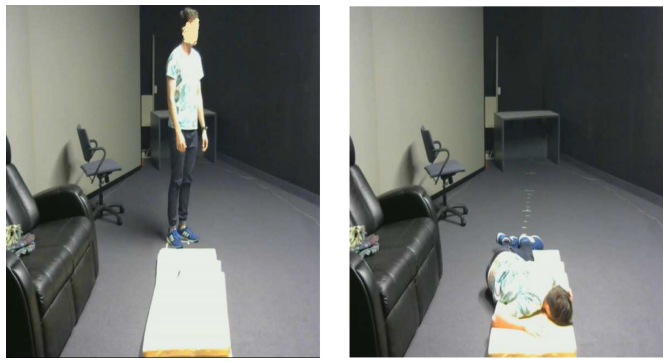
Instead of using engineered features, automatic feature extraction for fall detection can be achieved through deep learning methods [25]. A deep neural network approach for fall detection was presented in [26]. Gray-scale images were obtained from spectrograms, and stack auto-encoders were used to extract the features from the gray-scale images. Softmax regression was used to classify different kinds of activities including falls. In [27], transfer learning was investigated for classifying activities using data collected from wearable sensors. The pretrained AlexNet was used for feature extraction from spectrogram images followed by a linear or nonlinear support vector machine (SVM) for classification. In [28], human micro-Doppler signatures of different simulated activities were classified using a convolutional (convolutional) neural network using both color and gray-scale time-frequency representations as input to the network. All the previous research cited above used object-specific feature extraction. None of them addressed radar-based fall detection using automatic features that were based on shape of the distribution of the energy in the joint time-frequency domain.

This article proposes a fall detection method by projecting the energy content of the fall and nonfall activities, contained in their corresponding spectrograms, into binary image representations. The binary image is enhanced using a morphological operation and then is used as input into a deep convolutional neural network that performs automatic feature extraction. The proposed method will be compared with other machine learning approaches such as decision trees (DTs), k-nearest neighbors (KNN), and SVMs. These approaches are presented with the same input that was applied to the deep convolutional network used in the proposed approach.

The rest of this article is organized as follows. In Section II, the UWB radar used in the data acquisition and experiment setup in the proposed method is described. In Section III, the proposed fall detection method comprising of target range selection, time-frequency analysis, binary image generation, and data augmentation along with automatic feature extraction are described. In Section IV, experimental results, current challenges, and future directions on radar-based fall detection are presented. Finally, Section V concludes this article.

II. Experimental Setup and Measurement

The radar used in this experiment was the Xethru X4M03 development kit, manufactured by Novelda (Oslo, Norway) [29]. This radar uses a UWB transceiver operating in the range of 5.9–10.3 GHz and a patch antenna with a 65° aperture in both azimuth and elevation axes. This particular radar was chosen because of its low cost, small size, and high spatial resolution. The experiments were conducted in two realistic room environments at the University of Ottawa. Both rooms were cluttered and had the following dimensions: 12.6×4.1 and 5.7×2.2 m^2. The radar was mounted 1.5 m above the floor level. The sampling rate of the radar was 200 Hz, which is high enough to capture high-frequency components of the radar signal that appear during a fall (around 60 Hz). The dataset used in our experiments included different types of fall and nonfall activities performed by ten different healthy subjects aging from 20–35, namely, walking toward radar and falling down at different distances to the radar, i.e., 3 or 4 m, standing in front of radar and falling down at different distances to the radar, standing and falling down perpendicular to the radar line of sight, lying down with or without side rolling or other movements, and standing up in front of the radar, and lying down and standing up perpendicular to the radar line of sight. Each activity lasted 15 s, within which one of these activities occurred. The signals were then digitized at a rate of 200 samples/s. The range of the radar used in this study was set to 10 m. Since it has a 5.35 cm range resolution, there are 189 range bins. The type and number of different activities performed are given in Table I. Fig. 1 shows the posture of one of the subjects before and after a fall incident. The data were manually labeled as fall or nonfall. Ethics approval for conducting the experiments was obtained from the Research Ethics Board at the University of Ottawa.



(a) (b)

Fig. 1. Postures in one of the room environments. (a) Before. (b) After a fall.

III. Proposed Fall Detection Method

In this section, the proposed fall detection method based on time-frequency representation of the radar return signals and automatic feature extraction using a convolutional neural network is presented. Fig. 2 shows the block diagram of the proposed radar-based fall detection method.

A. Target Range Selection

The radar return signals were recorded into a matrix, where each column represents the spatial samples from different ranges (fast-time), while the data in each row corresponds to observations recorded at different time intervals (slow-time). The first 20 range bins, corresponding to all ranges less than 1 m from the radar, were noisy and were removed before further processing. The observations from the range bin having the highest variance was chosen to be the target range bin and used

TABLE I
TYPES AND NUMBER OF ACTIVITIES PERFORMED BY TEN DIFFERENT SUBJECTS IN TWO DIFFERENT ROOMS

Class	Description	# of Exp.	# of Exp. after augmentation ($\times 10$)
Fall	Stand in front of the radar and fall down	61	610
Fall	Walk toward the radar and fall down	59	590
Fall	Stand and fall down perpendicularly to the radar line of sight	67	670
Non-fall	Lie down and stand up	85	850
Non-fall	Lie down and stand up perpendicularly to the radar line of sight	64	640

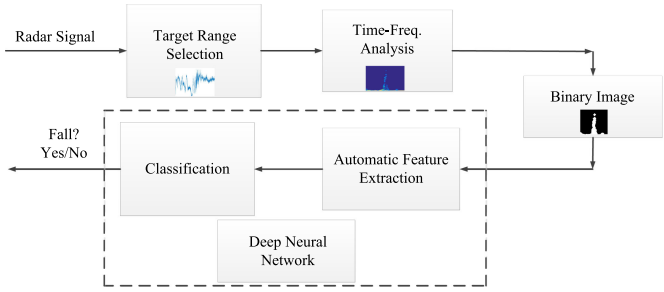


Fig. 2. Block diagram of the proposed radar-based fall detection method which is built upon time-frequency analysis and convolutional neural network.

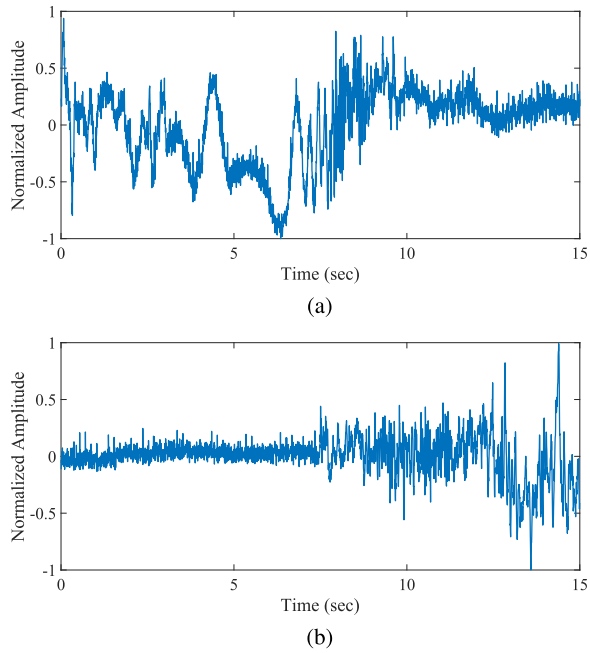


Fig. 3. Radar return signal in the target range bin. (a) Falling down. (b) Standing up.

for further processing. Fig. 3(a) and (b) shows the radar return signals normalized by the maximum amplitude for the target range bin after clutter removal, corresponding to falling down and standing up activities, respectively. More specifically, in Fig. 3(a), the subject is standing in front of the radar and then falls, while in Fig. 3(b), the subject is lying down and then stands up while facing the radar.

B. Time-Frequency Analysis

It is known that the nature of the radar return signal from moving human subjects is nonstationary with time-varying

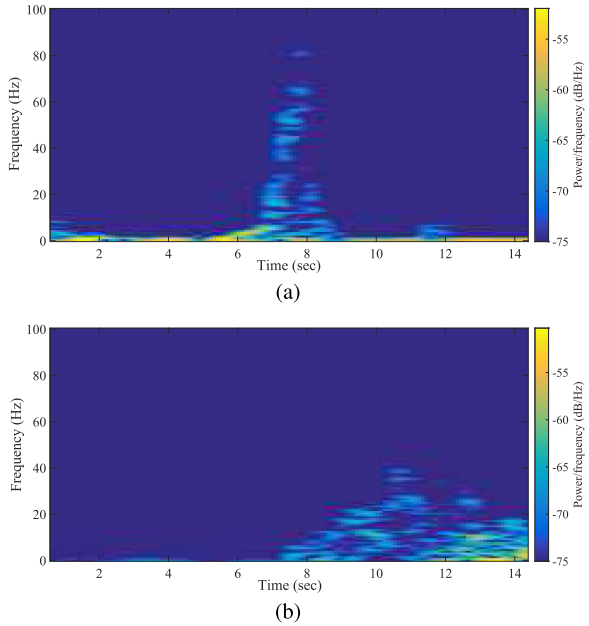


Fig. 4. Time-frequency signature. (a) Falling down. (b) Standing up.

frequency contents [7]. In view of this, in order to analyze the radar return signals, a joint time-frequency representation is obtained by applying the short-time Fourier transform (STFT) [30]. For a radar return signal in the target range bin $x[\cdot]$, STFT is defined as follows:

$$X[n,k] = \sum_{r=-\infty}^{\infty} x[r]W[r-n]\exp(-j2\pi rk/N) \qquad (1)$$

where $W[\cdot]$ is a finite length sliding window function (e.g., a Hamming window), n is the time index, $k = 0, 1, \ldots, N-1$ is the frequency index, and N is the number of frequency points. The squared magnitude of STFT yields the spectrogram, i.e., $S(n,k) = |X[n,k]|^2$. In this experiment, a Hamming window of size 256 samples was applied to the radar returns. The main lobe of the Hamming window is inversely proportional to the window length and the magnitude of its sidelobes is well attenuated. We use an 80% overlap between adjacent windows in order to minimize the effect of spectral leakage [30] and to better localize the fall events (to make sure the event is not missed). We found that lower amounts of overlap do not work as well. For example, with 50% overlap, falls were not always completely localized. With larger overlaps, the whole event is captured in at least one window [31]; this cannot be assured with smaller overlaps. Fig. 4(a) and (b) shows the time-frequency plot of the falling

Fig. 5. Binary time-frequency signature. (a) Falling down. (b) Standing up.

down and standing up activities, corresponding to the radar return signals of Fig. 3, respectively, where the horizontal axis is time and the vertical axis is frequency. The image intensities indicate the energy corresponding to the micro motion signature at each time instant [32]. It is seen from Fig. 4 that the energy content of each activity is distinguishable in the time-frequency domain. More specifically, a fall incident results in an instantaneous high-frequency content with a specific distribution of energy over time, whereas a nonfall activity exhibits a lower frequency peak with different energy distribution.

C. Binary Image Generation

Radar signals of fall and nonfall activities were processed and converted into binary image representations. Each binary image has a size of $m_1 \times m_2$ pixels, corresponding to m_1 frequencies and m_2 time instants at which the spectrogram is computed. It is known that the raw spectrogram images exhibit a high level of noise, which may obscure the true signature of the activity under study [26]. This may result in a lower classification performance, especially when using a neural network. To address this issue, a binary time-frequency signature of each activity was first obtained by applying the *k*-means clustering algorithm [33] to the time-frequency representations, i.e., pixel clustering. The clusters were then passed through a median filter to remove the outliers. In addition, a morphological opening operation was performed to remove disconnected regions and fully preserve the shape of the energy content of the activity. We do this by creating a structuring element [34], which identifies the pixel in the image that is being processed together with the adjacent neighboring pixels. The result of this postprocessing is the signature of different activities. Fig. 5 shows binary time-frequency signatures corresponding to the fall and nonfall spectrograms in Fig. 4, respectively.

D. Deep Neural Network

Automatic feature extraction precludes the domain specific feature extraction crafted manually by the experts. In order to automatically extract features from the radar binary images, a neural network comprising of convolutional and fully connected (FC) layers was constructed. Fig. 6 shows the architecture of the proposed convolutional neural network, and Table II gives kernel sizes and output shapes for each layer. As discussed in Section III-C, the time-frequency representation containing the energy of an activity is converted into a binary image, which will be denoted as $\mathbf{Z}$. The image is then fed into the convolutional neural network. The key attribute of this network is that it contains different processing units such as convolution, pooling, activation, and normalization.

TABLE II
CONFIGURATION OF THE PROPOSED NETWORK

Layer	Layer type	Kernel shape	Output shape
0	Input	-	$[N_{tr}, 129, 139, 1]$
1	Conv.	64×[3, 3]	$[N_{tr}, 129, 139, 64]$
2	MP	[2, 2]	$[N_{tr}, 64, 69, 64]$
3	Conv.	128×[3, 3]	$[N_{tr}, 64, 69, 128]$
4	MP	[2, 2]	$[N_{tr}, 32, 35, 128]$
5	FC	-	[500, 200, 100, 50]
6	Softmax	-	[2]

Inspired by [35] and [36], we construct a neural network to solve our binary classification problem by stacking two convolutional layers and four FC layers. In the first convolutional layer c_1, 64 kernels $\{k_j^{c_1}\}_{j=1}^{64}$ of size 3×3 were convolved with the input images with a stride of 1. A bias value was then added and an activation function was applied to the output resulting in a feature map $\mathbf{Z}^{c_1}$ with a depth of 64. The rectified linear unit activation function (ReLU), defined as $f(x) = \max(0, x)$, was employed. Each unit in a convolutional layer was connected to a local patch in the feature maps of the previous layer via a set of convolutional kernels.

In order to obtain distortion-invariant features and reduce the spatial resolution of the feature map, a nonoverlapping 2×2 max-pooling (MP) layer p_1 was used. The pooling was performed by selecting the maximum values of the convolved features among the adjacent neurons that are located in the preceding convolutional layer.

The number of filters increases from the first to second convolutional layer since the spatial size of the feature maps decreases after pooling operation, and thus, the depth needs to be increased by adding more filters. In view of this, in the next convolution layer c_2, the feature map from the previous pooling layer $\mathbf{Z}^{c_1p_1}$ was convolved with 128 kernels $\{k_j^{c_2}\}_{j=1}^{128}$ of size 3×3 with a stride of 1. A bias value was added and the ReLU activation function was applied to the output, resulting in a feature map $\mathbf{Z}^{c_2}$ with a depth of 128. The second MP layer p_2 of size 2×2 was then applied. The convolutional layers capture low-level features, while higher layers extract higher-level features by combining low-level ones. In addition, in the convolutional layers, a weight sharing mechanism exists among the neurons that are located in the same feature map. In the proposed network, the convolutional layers were followed by four FC layers and one output layer for class prediction. The four FC layers have 500, 200, 100, and 50 neurons, respectively. The output of the second convolutional layer $\mathbf{Z}^{c_2p_2}$ was flattened into a 1-D vector, resulting in parameter concatenation and used as the input to the first FC layer. Neurons in each FC layer have full connections to all neurons in the previous layer. Their activations can be computed by applying a matrix multiplication and a bias offset followed by the ReLU activation function for the first four layers and the softmax function for the output layer.

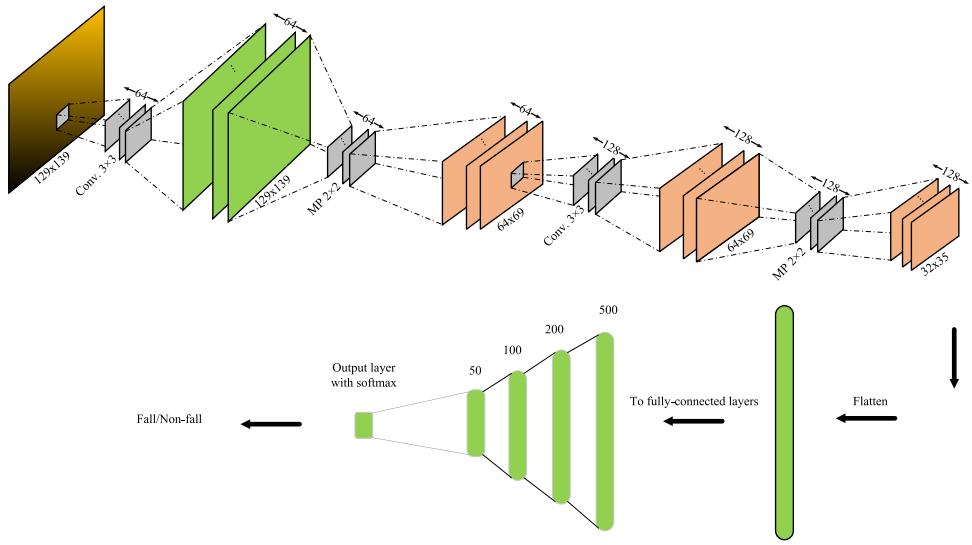


Fig. 6. Architecture of the proposed convolutional neural network with Convolutional, MP, and FC layers.

The softmax function is given by the following:

$$h^r = \frac{\exp(\mathbf{Z}^r)}{\sum_{v=1}^{2}\exp(\mathbf{Z}^v)}, \quad \text{for } r = 1, 2 \tag{2}$$

where $\mathbf{Z}^r$ is the rth score of the output layer and h^r denotes the output of the softmax function, i.e., the probability of the predicted classes. If N_{tr} is the number of trials, the actual labels for each fall or nonfall are $\{l_i\}_{i=1}^{N_{\text{tr}}}$, and the probabilistic outcomes are h_i, then the softmax cross-entropy cost function L is defined as follows:

$$L = -\frac{1}{N_{\text{tr}}}\sum_{i=1}^{N_{\text{tr}}}\sum_{v=1}^{2} I(l_i = v)\log(h_i^v) \tag{3}$$

where $I(.)$ is the indicator function. The entire network was trained by means of the back-propagation algorithm [37] to iteratively update the weights and minimize the cost function. The Adam optimizer was used in the learning process, which has been shown to provide fast convergence rates [38]. It is known that the Adam optimizer combines the properties of the AdaGrad and RMSProp algorithms and enables an optimization algorithm to handle sparse gradients [38]. Through the optimization model, the feature learning and classification are mutually enhanced and the learned features gain more discriminative power for the ultimate classification task.

In the proposed convolutional neural network, hyperparameter selection was done by tuning the model with different values of hyperparameters and selecting the one based on the validation accuracy, i.e., a grid search optimization technique. The hyperparameters of the proposed network are the learning rate, which determines the magnitude of weight changes when a classification error occurs during the training phase, and the drop-out regularization factor, which restricts the network adaptation to the training data and, thus, avoids overfitting and high-dimensionality issues.

E. Data Augmentation

Because falling is not a common human activity, fall data are scarce. In order to ensure that the fall detection model has good generalization, we employ data augmentation. Data augmentation artificially increases the number of exemplars or patterns in the training set, so that the network never sees twice the exact same image. This prevents the network from overfitting. There are several ways of augmenting the data described in the literature. They can be broadly divided into two groups, namely, oversampling and data warping [39]. Oversampling methods such as random over sampling or synthetic minority oversampling technique artificially increase the exemplars. The exemplars are nothing but copies of the existing ones in the dataset. Data warping uses image transformation processes such as geometric transformations, flipping, cropping, and translation to produce exemplars, which are different from the existing exemplars in the dataset. Fall detection using radars has several dependencies such as different return power, presence or absence of clutters, elevation and angle of the radars. As all these variabilities cannot be captured in a dataset, data augmentation is needed. Oversampling augmentation is unsuitable because it will not produce exemplars that are representative of these variabilities. Therefore, image transformations were used to realize such feature variability and account for physically meaningful and intuitive interpretation of the activities. The image transformations used in this work were rotation, width shifting, height shifting, horizontal flipping, shearing, and zooming. For instance, patterns resulting from horizontal flipping and width shifting could represent fall/nonfall activities at different distances from the radar sensor, while those obtained using zooming and height shifting may account for falling down at different angles relative to the radar. Note that the proposed convolutional neural network aims to extract robust features to correctly classify a variety of instances. Such robust features can

TABLE III
CONFUSION MATRIX DEFINITION FOR THE FALL DETECTION PROBLEM

Class	Fall	Non-fall
Fall	TP	FP
Non-fall	FN	TN

be learned only if a sufficiently large number of different types of instances are presented to the network.

IV. RESULTS, CHALLENGES, AND FUTURE DIRECTIONS

In this section, the experimental results, obtained using the proposed method, are presented. The performance of the convolutional neural network is compared with an auto-encoder (AE) [26], a SVM [40], DTs, and KNN [4]. To evaluate the performance of the proposed fall detection method, a set of radar data, collected as described in Section II, was used. The radar return signal was processed to obtain the spectrogram. The spectrogram was treated as an image containing the energy content of a particular activity. The image was then binarized and further enhanced using the morphological opening operation. In the experiments, the image size, discussed in Section III-C, was fixed by setting the values of m_1 and m_2 to 129 and 139, respectively. The resulting binary image was augmented to obtain enough data to train the proposed convolutional neural network. The trained network is then used to test whether or not a test image represents a fall incident.

To evaluate the classification performance of the proposed method, a fivefold cross validation was performed. For the ease of defining metrics, the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), corresponding to the fall and nonfall activities, are defined in Table III. The main diagonal entries are the correct classification rates for fall and nonfall activities. The off-diagonal entries indicate the misclassification rates.

The following metrics are used to evaluate the performance of the proposed fall detection method:

1) Precision (PR), $\text{PR} = \frac{\text{TP}}{\text{TP+FP}}$
2) Recall or sensitivity (SE), $\text{SE} = \frac{\text{TP}}{\text{TP+FN}}$
3) Specificity (SP), $\text{SP} = \frac{\text{TN}}{\text{TN+FP}}$
4) False Positive Rate (FPR), $\text{FPR} = \frac{\text{FP}}{\text{FP+TN}}$
5) False Negative Rate (FNR), $\text{FNR} = \frac{\text{FN}}{\text{FN+TP}}$
6) F-score, $F = \frac{2\text{TP}}{2\text{TP+FP+FN}}$.

Table IV gives various classification metrics obtained using the proposed method and those of the other methods, namely, KNN, DT, AE, and SVM. It should be noted that for all these methods, the resulting binary images from Section III-C were vectorized and used as the input to the classifier.

In the case of AE, we followed the approach in [26] to reduce the dimensionality of the feature vector in such a way that the binary images were compressed into 300 and 150 features using two stacked encoders with l_2 regularization. In addition, the softmax regression classifier was used in the output layer. In the case of KNN, different values for k were examined and the best results, i.e., $k = 1$ were reported. In the case of SVM, linear SVM (LSVM) and Gaussian SVM (GSVM) were employed. The LSVM uses the l_2 regularization and the squared hinge loss function [41]. The parameters of the GSVM are the regularization parameter $C = 1$, which aims at classifying the training dataset by providing the model with freedom to select more samples as support vectors and the cutoff parameter $\gamma = 0.001$, which identifies the extent of the influence of a single training sample (exemplar) [41]. The abovementioned model parameters were tuned using a grid search on a validation set for higher classification accuracy.

Table IV describes the classification metrics obtained when each classifier was trained on 336 data samples, as well as the same metrics when the classifiers were trained on an augmented dataset (augmented by a factor of 10). The classifiers were trained using the radar data for nine human subjects and tested using the data from the tenth subject in a leave-one-subject-out cross-validation sense (tenfold). It is seen from this table that the proposed method outperforms the other methods when tested with unseen set of data. When the number of data samples increases, the performance improvement using the proposed method is more significant than that using the other methods, as evidenced by the specificity values. It should be noted that the hierarchical features extracted from radar binary images are the key to understand the fundamental composition of human activities and determine whether or not a fall occurs during human daily activities.

Tables V and VI give classification metrics obtained using a fivefold cross validation for the proposed fall detection method and those obtained using other methods, with original data and with the data augmented by a factor of 10, respectively. In the case of the DT approach, two attribute selection measures, namely, the Gini index (DTG) and information gain or entropy (DTE) were considered [42]. It is seen from these tables that the proposed method is capable of detecting falls with higher accuracy, precision, and specificity values, indicating the capability to better detect a fall incident when it occurs and avoiding false alarms. Without data augmentation, DTG, DTE, LSVM, GSVM, KNN, and AE achieve 81.30%, 82.47%, 89.01%, 85.71%, 92.84%, and 92.85% classification accuracies, respectively, while the proposed method outperforms the other methods by providing 95.83% classification accuracy and 97.82% specificity. It is also seen from Table VI that when the data are augmented, the proposed method detects the fall incidents with a lower false alarm rate as evidenced by the higher value of specificity. The overall result indicates that the proposed method is effective in classifying fall from nonfall activities. The superior performance of the proposed convolutional neural network-based fall detection method can be attributed to the fact that the proposed model finds the saliency of the signal in feature representation more accurately than the other methods. This is due to the fact that the convolutional neural networks make use of hierarchical distributed representations.

The statistical significance of the proposed method was also investigated. The p-value for a balanced one-way analysis of variance (ANOVA) between different methods, when comparing the classification accuracies, is 0.0299; when comparing precision values, it is 0.0333. This indicates that the differences

TABLE IV
PR, SP, SE, FPR, AND FNR (%) FOR VARIOUS METHODS, IN A LEAVE-ONE-SUBJECT-OUT CROSS VALIDATION

	No augmentation					With augmentation ($\times 10$)				
Method	PR	SP	SE	FPR	FNR	PR	SP	SE	FPR	FNR
LSVM	77.32	89.52	76.89	23.11	10.48	79.56	77.12	73.57	26.43	22.88
GSVM	78.78	88.34	78.12	21.88	11.66	83.66	81.34	79.12	20.88	18.66
KNN	85.89	90.42	86.10	13.9	9.58	87.92	91.51	86.22	13.78	8.49
AE	89.03	90.42	88.15	11.85	9.58	87.64	93.33	88.02	11.98	6.67
Proposed	**92.97**	**93.18**	**92.79**	**7.21**	**6.82**	**91.27**	**95.19**	**91.68**	**8.32**	**4.81**

The bold entities show the result of the proposed method.

TABLE V
PERFORMANCE METRICS FOR VARIOUS CLASSIFIERS

Method	Accuracy	Precision	Sensitivity	Specificity	F-score
DTG	81.30	85.39	83.21	80.47	83.69
DTE	82.47	85.46	84.23	81.32	84.41
LSVM	89.01	90.69	89.95	88.32	90.17
GSVM	85.71	90.75	86.43	87.88	87.77
KNN	92.84	92.94	94.52	91.78	93.41
AE	92.85	93.25	93.40	92.34	93.30
Proposed	**95.83**	**98.37**	**94.37**	**97.82**	**96.28**

The bold entities show the result of the proposed method.

TABLE VI
PERFORMANCE METRICS FOR VARIOUS CLASSIFIERS AFTER DATA AUGMENTATION

Method	Accuracy	Precision	Sensitivity	Specificity	F-score
DTG	85.54	86.43	88.25	83.35	86.93
DTE	86.28	88.66	87.59	85.53	87.78
LSVM	78.24	80.44	81.30	75.83	81.71
GSVM	83.52	87.37	84.54	83.43	86.45
KNN	90.33	90.73	92.47	88.59	91.28
AE	90.89	90.90	92.97	88.85	91.72
Proposed	**93.54**	**94.04**	**94.22**	**92.54**	**92.65**

The bold entities show the result of the proposed method.

between the methods are significant. Fig. 7 shows box plots for classification metrics of various methods. It is seen from this figure that the proposed method using the convolutional neural network provides higher accuracy and precision on a fivefold cross validation. This figure also shows that the proposed method provides lower standard deviation in both the metrics, indicating the high level of agreement between the results of each fold.

Receiver operating characteristic (ROC) curves for various methods were also obtained. Fig. 8 shows the ROC curves for KNN-, AE-, and CNN-based methods obtained through fivefold cross validation. The area under the ROC curve (AUC) for each method was obtained from the corresponding average ROC curve. It is seen from these figures that the proposed method based on the convolutional neural network provided the highest AUC, indicating its superiority to the other methods in discriminating fall from nonfall activities.

We now compare the performance of the proposed fall detection method to those presented in [26] and [43]. To this end, as discussed in [26], three features, namely, extreme frequency magnitude, extreme frequency ratio, and time-span of event, were extracted from the spectrograms. These features were used to classify the fall and nonfall activities using a SVM classifier. In addition, as discussed in [43], several features including higher-order statistics such as moments, cumulants, and the energy of the normalized signal were extracted. We also extracted 27

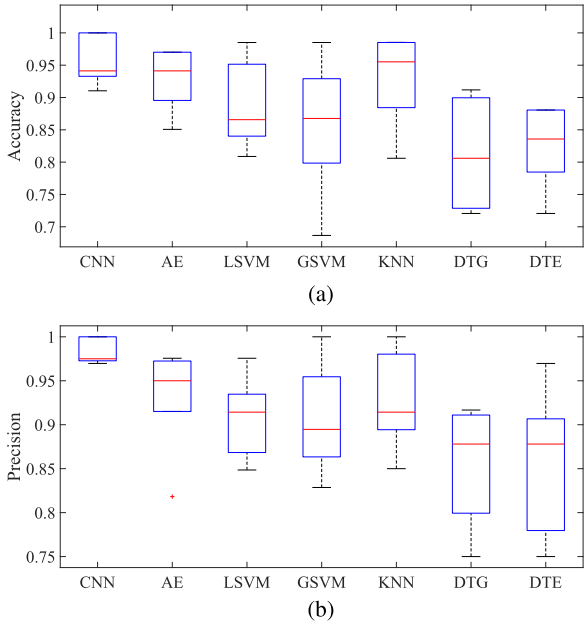


Fig. 7. Box plots for classification metrics of various methods. (a) Accuracy. (b) Precision.

TABLE VII
CLASSIFICATION METRICS (AC, PR, SE, SP) OBTAINED USING THE PROPOSED FALL DETECTION METHOD AND THOSE OBTAINED USING [26] AND [43] AND THE ONE WITH 27 FEATURES EXTRACTED FROM THE RADAR DATA ($27F$), IN A FIVEFOLD CROSS VALIDATION

Method	Accuracy	Precision	Sensitivity	Specificity
[43]	84.15	84.94	86.82	80.84
[26]	86.03	87.53	87.82	83.94
27F	89.88	91.44	90.48	91.05
Proposed	**95.83**	**98.37**	**94.37**	**97.82**

The bold entities show the result of the proposed method.

features ($27F$) including maximum, minimum, mean, standard deviation, skewness, kurtosis, interquartile range, area under the curve, area under the squared curve of the radar time series signal, its first derivative, and its discrete Fourier transformed version. The extracted features were classified using a SVM classifier. Table VII gives classification metrics in fivefold cross validation obtained using the proposed method and those of the other methods. It is seen from this table that the proposed method provides higher classification metrics as compared to the other methods based on manual feature extraction techniques.

We also investigated the performance of the proposed fall detection method when the network was trained on the dataset collected from a set of subjects in one room and tested on the dataset collected from different subjects in the another room.

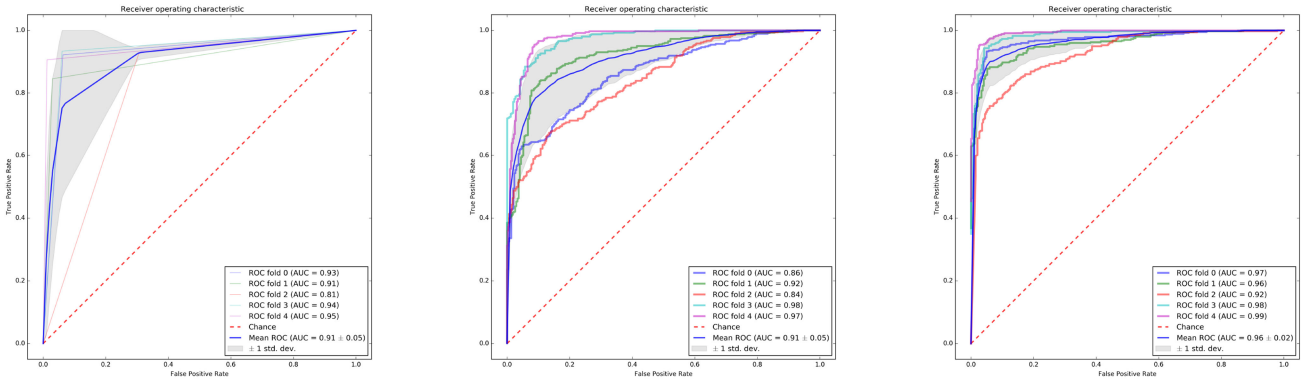


Fig. 8. ROC curves and AUCs obtained using fivefold cross validation for (a) KNN, (b) AE, and (c) proposed method.

TABLE VIII
CLASSIFICATION METRICS FOR VARIOUS METHODS, WHEN THE CLASSIFIERS ARE TRAINED USING THE DATA COLLECTED IN ONE ROOM AND TESTED AGAINST THE DATA FROM DIFFERENT SUBJECTS IN ANOTHER ROOM IN A TWOFOLD CROSS VALIDATION

Method	Accuracy	Precision	Sensitivity	Specificity
LSVM	77.94	92.37	69.43	90.52
GSVM	76.14	74.59	77.09	80.82
KNN	82.72	92.19	75.52	91.66
AE	85.03	89.04	80.06	89.81
Proposed	**89.12**	**94.32**	**83.28**	**95.04**

The bold entities show the result of the proposed method.

Table VIII gives classification metrics obtained in this scenario. It is seen from this table that the proposed method generalizes well to other unseen environments and datasets.

As discussed in Section III-D, the proposed convolutional neural network automatically extracts features from the radar returns. Visualizing the activations of the network is useful for understanding how different layers transform their input and reduce the size of the feature map. To understand the automatically extracted features at each layer, feature maps (activations) of the flatten layer, the fully connected layers, and the output layer are depicted in Fig. 9. The convolutional layers retain the shape existing in the input image. As we go deeper into the layers toward the output layer, the activations become increasingly abstract and less visually interpretable. In that case, the network begins to encode higher-level concepts. Activations in the higher layers carry increasingly less information about the visual contents of the image, and increasingly more information related to the class of the image. This is clearly visible in Fig. 9. The cross-entropy loss and accuracy values for the training and test sets were obtained using the proposed network. Fig. 10 shows cross-entropy loss and accuracy values on the training and test dataset obtained using the proposed method with and without data augmentation. The gradually decreasing trend of the loss seen from this figure demonstrates that the network is successfully trained after a fixed number of epochs. The accuracy improves with the number of epochs and finally reaches a steady state. It can also be seen from this figure that the use of data augmentation helps the model to generalize better and improves the classification accuracy.

In the proposed fall detection method, the range bin having the highest variance was chosen to be the target range bin and used for further processing. In addition, a range spread of four consecutive bins for the target and the average scattered signal over these range bins were also examined. However, only minor difference in the performance was observed.

When the subject falls down perpendicularly to the radar line of sight, the time-frequency representation and binary image are both slightly different from falling down along the radar line of sight. This results in confusion between fall and nonfall classes. This confusion can be prevented and the fall detection rate can be increased by employing multiple radar sensors [44]. Doing so would raise the precision of fall detection by covering the target from multiple directions and reducing the chance that the target may be obscured from one of the sensors. In addition, as discussed in [45], combining range features may further improve the recognition accuracy, especially when the target motion is perpendicular to the radar. Integrating range information as a parallel input to the proposed convolutional neural network can be a direction for future research.

So far, there is no automated system for human gait analysis and fall risk assessment based on the radar, which is mostly due to the lack of publicly available datasets. Previous researchers have mostly collected their own data and tried to validate their algorithms statistically. In view of this, benchmark datasets are important to realize qualitative comparisons of new analysis algorithms and pipelines. In order to help create standard benchmark datasets, the dataset used in this research work will be made publicly available.

Radar-based indoor monitoring, and in particular radar-based fall detection, is in its infancy. There is huge potential for more advanced radars and algorithms to overcome the existing challenges. Radar-based systems may carry great potential as one of the leading technologies in the future of indoor monitoring.

We are currently investigating the use of the micro-Doppler radar to detect motion and falls through the fusion of the range-Doppler image and extracted spectrogram images, as it is known that the Doppler signature and rate of change in Doppler contain

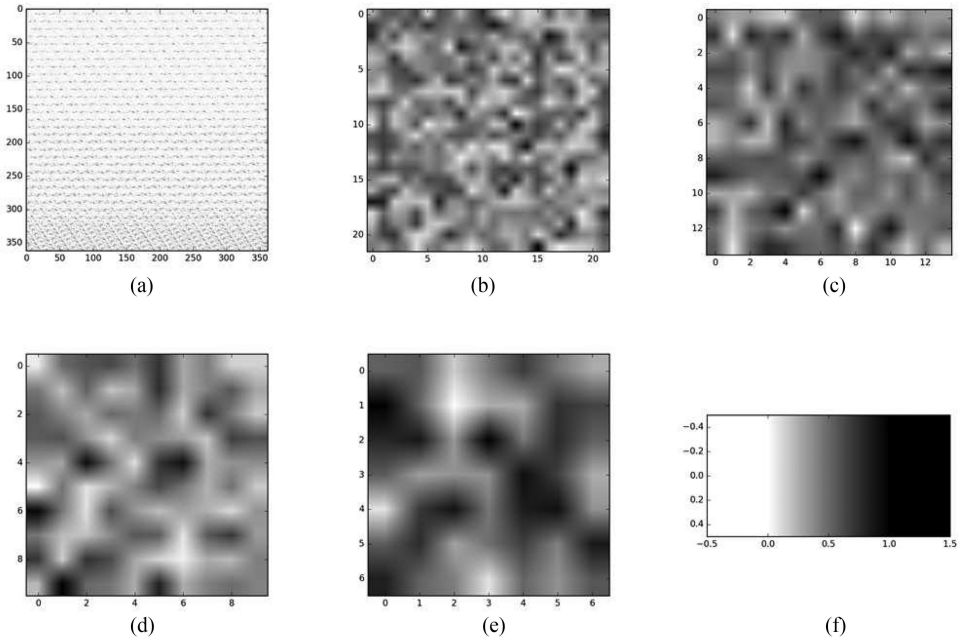


Fig. 9. Activations of (a) flatten layer, (b)–(e) FC layers, and (f) output layer, for one sample binary image.

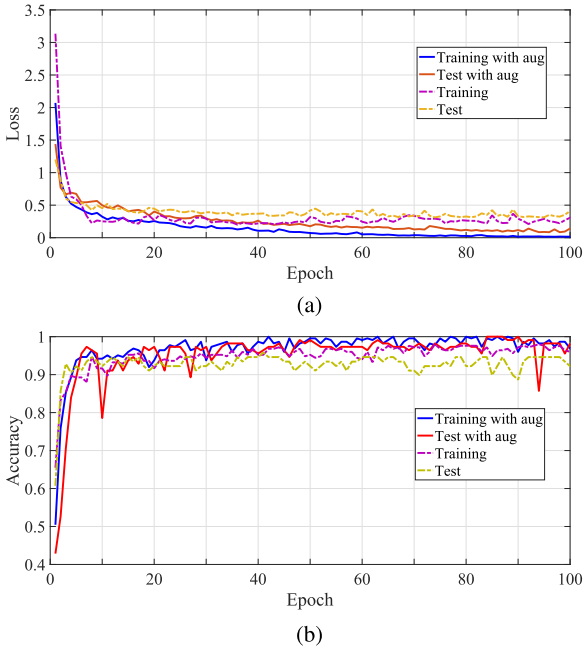


Fig. 10. (a) Cross-entropy loss, and (b) accuracy values for the training and test sets, with and without data augmentation, obtained using the proposed convolutional neural network. The results are averaged over tenfolds corresponding to each subject in a leave-one-subject-out cross validation.

rich information about posture, movement, and sudden changes. In addition, the Doppler rate may be identified by using two radars which are synchronously capturing the information.

Fall detection using radars is only in its gestation stage. Automatic real-time detection of falls using the radar is still not a reality; distinguishing other activities of daily living such as walking, running, sitting, standing from falls is also still in its infancy. This article proposed a binary classification of fall from a nonfall activity. A multiactivity classifier will be built in the future to distinguish not only falls from nonfalls, but to classify various types of activities. However, when multiple people are in the view of the radar, activity monitoring becomes very difficult. Future research will focus on automatic identification and tracking of multiple people and also on building systems that can learn activities in time, using unsupervised learning, through extended observation in time. Also, current systems require large datasets for training. Recently introduced ideas such as contrastive learning could be considered for enabling learning with fewer examples. Also, as systems often get trained offline and are required to work in a new environment, transfer learning preferably in an unsupervised manner will be considered in the future.

V. Conclusion

In this article, a novel radar-based fall detection method was proposed based on time-frequency analysis and deep learning techniques. Data were collected in room environments without any constraints, including fall and nonfall activities. The radar return data were preprocessed to determine the target range bin and removed the effect of clutters. Time-frequency analysis was performed by using the STFT and to obtain the spectrogram for different activities. The spectrograms were further processed to obtain binary images and enhanced using morphological operators. The binary images were augmented using class-preserving transformations and fed into the proposed convolutional neural network for feature extraction. The proposed network was devised by stacking convolutional and FC layers. Several experiments were conducted to evaluate the performance of the proposed fall detection method and to compare it with those of

the other methods. The results showed that the proposed method outperforms other methods in accuracy, precision, sensitivity, and specificity. The performance of the proposed fall detection method when the network was trained on a dataset collected from a set of subjects in one room and tested on a dataset collected from different subjects in another room was also investigated. The higher classification metrics indicates the robustness of the proposed method.

REFERENCES

[1] S. Kim, K. Hirota, T. Nozaki, and T. Murakami, "Human motion analysis and its application to walking stabilization with COG and ZMP," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 5178–5186, Nov. 2018.

[2] A. Ozdemir and B. Barshan, "Detecting falls with wearable sensors using machine learning techniques," *Sensors*, vol. 14, pp. 10691–10708, 2014.

[3] M. Forouzanfar, M. Mabrouk, S. Rajan, M. Bolic, H. R. Dajani, and V. Z. Groza, "Event recognition for contactless activity monitoring using phase-modulated continuous wave radar," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 479–491, Feb. 2017.

[4] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, and J. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1274–1286, Nov. 2012.

[5] A. Shahzad and K. Kim, "FallDroid: An automated smart phone based fall detection system using multiple kernel learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 35–44, Jan. 2019.

[6] C. Wang *et al.*, "Low-power fall detector using triaxial accelerometry and barometric pressure sensing," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2302–2311, Dec. 2016.

[7] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 71–80, Mar. 2016.

[8] WHO Global Report on Falls Prevention in Older Age, World Health Organization, Geneva, Switzerland, 2007.

[9] P. Corbishley and E. Rodriguez-Villegas, "Breathing detection: Towards a miniaturized, wearable, battery operated monitoring system," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 196–204, Jan. 2008.

[10] V. C. Chen, D. Tahmoush, and W. J. Miceli, *Radar Micro-Doppler Signatures: Processing and Applications, Radar, Sonar, Navigation and Avionics*. London, U.K.: Institution of Engineering and Technology, 2014.

[11] L. Hazelhoff, J. Han, and P. H. N. de-With, "Video-based fall detection in the home using principal component analysis," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2008, pp. 298–309.

[12] J. D. Bryan, J. Kwon, N. Lee, and Y. Kim, "Application of ultra-wideband radar for classification of human activities," *IET Radar, Sonar Navigation*, vol. 6, no. 3, pp. 172–179, 2012.

[13] C. Ding *et al.*, "Non-contact human motion recognition based on UWB radar," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 8, no. 2, pp. 306–315, Jun. 2018.

[14] I. Nejadgholi, H. Sadreazami, S. Rajan, and M. Bolic, "Classification of Doppler radar reflections as preprocessing for breathing rate monitoring," *IET Signal Process.*, vol. 13, no. 1, pp. 21–28, 2019.

[15] H. Sadreazami, M. Bolic, and S. Rajan, "On the use of ultra-wideband radar and stacked LSTM-RNN for at home fall detection," in *Proc. Life Sci. Conf.*, 2018, pp. 255–258.

[16] H. Sadreazami, M. Bolic, and S. Rajan, "Fall detection using standoff radar-based sensing and deep convolutional neural network," *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 67, no. 1, pp. 197–201, Jan. 2020.

[17] H. Sadreazami, M. Bolic, and S. Rajan, "Residual network-based supervised learning of remotely sensed fall incidents using ultra-wideband radar," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2019, pp. 1–4.

[18] S. Abdelhedi, R. Bourguiba, J. Mouine, and M. Baklouti, "Development of a two-threshold-based fall detection algorithm for elderly health monitoring," in *Proc. IEEE Int. Conf. Res. Challenges Inf. Sci.*, 2016, pp. 1–5.

[19] S. Bjorklund, H. Petersson, and G. Hendeby, "Features for micro-Doppler based activity classification," *IET Radar, Sonar Navigation*, vol. 9, no. 9, pp. 1181–1187, 2015.

[20] M. Otero, "Application of a continuous wave radar for human gait recognition," *Signal Proc., Sensor Fusion, Target Recognit.*, vol. 5809, pp. 538–548, 2005.

[21] I. Orovic, S. Stankovic, and M. Amin, "A new approach for classification of human gait based on time-frequency feature representations," *Signal Process.*, vol. 91, no. 6, pp. 1448–1456, 2011.

[22] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.

[23] B. Y. Su, K. Ho, M. J. Rantz, and M. Skubic, "Doppler radar fall activity detection using the wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 865–875, Mar. 2015.

[24] Q. Wu, Y. D. Zhang, W. Tao, and M. Amin, "Radar-based fall detection based on Doppler time-frequency signatures for assisted living," *IET Radar, Sonar Navigation*, vol. 9, no. 2, pp. 164–172, 2015.

[25] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: A survey," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 84–100, Jan. 2018.

[26] B. Jokanovic, M. Amin, and F. Ahmad, "Radar fall motion detection using deep learning," in *Proc. IEEE Radar Conf.*, 2016, pp. 1–6.

[27] D. Wagner, K. Kalischewski, J. Velten, and A. Kummert, "Activity recognition using inertial sensors and a 2D convolutional neural network," in *Proc. Int. Workshop Multidimensional Syst.*, 2017, pp. 1–6.

[28] Y. Lang, C. Hou, Y. Yang, D. Huang, and Y. He, "Convolutional neural network for human micro-Doppler classification," in *Proc. Eur. Microw. Conf.*, 2017, pp. 1–4.

[29] Novelda's XeThruX4M03, 2018. [Online]. Available: https://www.xethru.com/x4-radar-chip.html/

[30] L. Stankovic, M. Dakovic, and T. Thayaparan, *Time-Frequency Signal Analysis With Applications*. Norwood, MA, USA: Artech House, 2013.

[31] B. Erol, M. Francisco, A. Ravisankar, and M. Amin, "Realization of radar-based fall detection using spectrograms," *Compressive Sensing VII: From Diverse Modalities Big Data Anal.*, vol. 10658, 2018, Art. no. 106580B.

[32] A. Gadde, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Fall detection and classifications based on timescale radar signal characteristics, in *Proc. Radar Sensor Tech. XVIII*, vol. 9077, 2014, Art. no. 907712.

[33] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm," *Procedia Comput. Sci.*, vol. 54, pp. 764–771, 2015.

[34] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2008.

[35] O. C.-Reyes, P. Scully, and K. B. Ozanyan, "Deep neural networks for learning spatio-temporal features from tomography sensors," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 645–653, Jan. 2018.

[36] L. Li, K. Ota, and M. Dong, "Deep learning for smart industry: Efficient manufacture inspection system with fog computing," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4665–4673, Oct. 2018.

[37] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[38] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–13.

[39] C. Shorten and T. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 60, pp. 1–48, 2019.

[40] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on Doppler radar motion signature," in *Proc. Int. Conf. Pervasive Comput. Tech. Healthcare*, 2011, pp. 222–225.

[41] C. Cortes and V. Vapnik, "Support vector machine," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[42] L. Raileanu and K. Stoffel, "Theoretical comparison between the Gini index and information gain criteria," *Annu. Math. AI*, vol. 41, no. 1, pp. 77–93, 2004.

[43] M. A. Kiasari, S. Y. Na, and J. Y. Kim, "Classification of human postures using ultra-wideband radar based on neural networks," in *Proc. Int. Conf. IT Convergence Secur.*, 2014, pp. 1–4.

[44] S. Gurbuz, C. Clemente, A. Balleri, and J. Soraghan, "Micro-Doppler based in-home aided and unaided walking recognition with multiple radar and sonar systems," *IET Radar, Sonar Navigation*, vol. 11, no. 1, pp. 107–115, 2017.

[45] B. Erol and M. Amin, "Fall motion detection using combined range and Doppler features," in *Proc. IEEE Eur. Signal Conf.*, 2016, pp. 2075–2080.