Fall Detection Using Standoff Radar-Based Sensing and Deep Convolutional Neural Network

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

***Abstract*—Automatic fall detection using radar aids in better assisted living and smarter health care. In this brief, a novel time series-based method for detecting fall incidents in human daily activities is proposed. A time series in the slow-time is obtained by summing all the range bins corresponding to fast-time of the ultra wideband radar return signals. This time series is used as input to the proposed deep convolutional neural network for automatic feature extraction. In contrast to other existing methods, the proposed fall detection method relies on multi-level feature learning directly from the radar time series signals. In particular, the proposed method utilizes a deep convolutional neural network for automating feature extraction as well as global maximum pooling technique for enhancing model discriminability. The performance of the proposed method is compared with that of the state-of-the-art, such as recurrent neural network, multi-layer perceptron, and dynamic time warping techniques. The results demonstrate that the proposed fall detection method outperforms the other methods in terms of higher accuracy, precision, sensitivity, and specificity values.**

***Index Terms*—Biomedical signal processing, smart homes, fall detection, convolutional neural network, ultra-wideband radar.**

I. Introduction

FALLING down is considered one of the leading causes of accidental deaths and one of the major causes of injury for seniors [1]. In view of this, developing technologies for fall detection and prevention systems is of the utmost importance in elder care systems. So far, the current fall detection methods are mostly based on wearable devices, video cameras and smart-phone sensors [2]. On the other hand, non-contact indoor monitoring using radar has become more popular in recent years [4], since it avoids privacy issue of the video-based techniques and preclude the need for wearing a device [3]. For instance, human activity and posture classification were studied in [5] using a radar sensor. In [6], a human activity recognition method was developed based on radar micro-Doppler data by extracting features from time-velocity and cadence-velocity domains. In [7], in order to detect gait abnormalities and transitional motions, the key Doppler features associated with gait motions were extracted. A wavelet-based approach was devised in [8] for fall detection purpose using Doppler radar. In [9], a fall detection scheme was presented by exploiting time-frequency characteristics of radar Doppler signatures, where events were classified by a sparse Bayesian classifier using statistics of different features.

Most of the existing radar-based fall detection methods are based on extracting a set of features from the radar data and developing a model to differentiate between the fall and non-fall daily activities. These features were extracted in time domain such as low and/or high order moments, in frequency domain such as cepstrum coefficients [10], peak frequency [11], and in time-frequency domain such as spectrogram [12]. However, these extracted features are limited in type, requiring expert knowledge to manually engineer. To overcome this problem, few attempts have so far been made to design an automated feature extraction method from radar data. In [13], a radar signal recognition was proposed using a deep restricted Boltzmann machine. In [14], a deep neural network approach was presented to reduce dimensionality of the extracted features from radar signals based on stack auto-encoder. A gait-based human identification was presented in [15] using an acoustic sensor and deep neural network. In [16], a generative model based on PixelCNN was presented to synthesize speech time series signals. Dilated causal convolution was employed to deal with the long-range temporal dependencies required. In [17], a sequence to sequence modeling method was proposed based only on convolutional neural network. This model was equipped with gated linear units and residual connections. It was shown that hierarchical convolutional structure may provide a more efficient way to capture long-range dependencies of the time series compared to the chain structure modeled by recurrent networks. However, automatic feature extraction directly from the radar time series used for various classification problems including fall detection has not been proposed in the literature.

In view of this, to extract features automatically from radar data and avoid manual feature extraction and engineering, in this brief, a fall detection method is proposed by incorporating time series derived from ultra wideband radar (UWB) returns and a deep neural network. The proposed method is realized by adopting deep convolutional neural network for extracting multi-level features from radar time series data. In particular, signals reflected from the target are summed up over all the range bins in fast-time and the resulting time series in

Manuscript received February 4, 2019; accepted March 7, 2019. Date of publication March 11, 2019; date of current version December 26, 2019. This brief was recommended by Associate Editor C.-T. Cheng. *(Corresponding author: Hamidreza Sadreazami.)*

H. Sadreazami and M. Bolic are with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: h_sadrea@encs.concordia.ca).

S. Rajan is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSII.2019.2904498

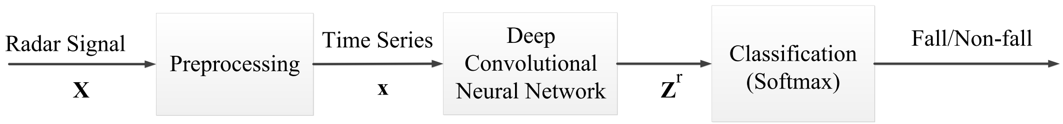


Fig. 1. Block diagram of the proposed radar-based fall detection method. A radar time series is obtained from the radar scattering matrix. A deep convolutional neural network is used for feature extraction and classification.

slow-time is fed into the deep convolutional neural network. The proposed network consists of convolutional layers and a global maximum pooling to enhance model discriminability. The following are the distinguishing contributions of this brief:

a) Unlike other published research works, this brief proposes the use of a time series derived from the radar scattering matrix as the input to the deep neural network for fall detection.

b) Unlike the other existing works that rely on heavy preprocessing and hand engineered feature extraction from time, frequency and time-frequency domains, the proposed method automatically derives the features from the time series input using deep convolutional neural networks.

II. Experimental Setup and Measurement

In our experiments, Xethru X4M03 development kit is used for data collection. UWB radar operates in $5.9 - 10.3$ GHz, providing high spatial resolution, unobtrusiveness and privacy-preserving monitoring. In particular, UWB radar works on the basic principle of sending short pulses at high frequencies, resulting in a high range resolution. The radar is placed 1.5 m above the floor level. The room is cluttered which is mostly static and is filtered out by removing mean from the radar scattering matrix before further processing. Each scan is repeated for 15 seconds and digitized at a rate of 200 samples/second. The data in each scan is summed up and it forms a time series over scans. The range of the radar used in this brief is set to 10 m. Thus, with 5.35 cm range resolution of the radar, each scan is divided into 189 bins.

The dataset includes different types of fall and non-fall activities, performed by five healthy male subjects aging from 20 to 25, namely, walking toward radar and falling down, standing in front of radar and falling down, standing and falling down perpendicular to the radar line of sight, lying down and standing up in front of the radar, and lying down and standing up perpendicular to the radar line of sight. The number of different fall and non-fall experiments performed are 121 and 85, respectively.

III. Proposed Fall Detection Method

In this section, the proposed fall detection method based on automatic feature extraction and classification using deep convolutional neural network is presented. Fig. 1 depicts block diagram of the proposed fall detection method.

A. Preprocessing

The radar return signals are recorded into a matrix $\mathbf{X} = x_{i,j} \in \mathbb{R}^{m \times n}$, where n columns represent the spatial samples from different ranges (fast-time, a scan), while m rows correspond to observations recorded at different time intervals (slow-time, across scans). The range of the radar used in this brief is set to 10 m. In order to prepare the input time series for the proposed deep convolutional neural network, the radar fast-time data are integrated to obtain a normalized time series with slow time samples as

$$\mathbf{x}_i = \sum_j \frac{x_{i,j}}{\underbrace{\max}_{j}(|x_{i,j}|)}. \tag{1}$$

TABLE I
Configuration of the Proposed Deep Neural Networks

	Layer type	Filter shape	Output shape
0	Input	-	[N_b, 3000, 1]
1	Conv.	300×(9, 1)	[N_b, 3000, 300]
2	Conv.	200×(7, 1)	[N_b, 3000, 200]
3	Conv.	100×(5, 1)	[N_b, 3000, 100]
4	Global max pooling	-	[N_b, 3000, 1]
5	Softmax	-	[2]

B. Deep Convolutional Neural Network

In most of the existing works on fall detection [8], [9], [11], a set of features from the radar data either in time or frequency domain are extracted and used as inputs for classification to detect fall incidents. The results in these works have shown that any improvement in the accuracy of the fall detection system depends on the features extracted. With the advent of deep neural networks, manual engineering of features can be avoided. In the proposed fall detection method, convolutional neural network is applied to the radar time series data, obtained in Section III-A, for automatic feature learning and classification tasks. The convolutional network is selected since it can capture dependency between a time instance and its neighboring instances in receptive field of the convolutional filters. In addition, their local connectivity and shared weights properties result in reducing the total number of trainable parameters and efficient training. The schematic of the proposed deep neural network is shown in Fig. 2. Table I gives details of configuration of the network; filter sizes and output shapes for the convolutional layers and global max pooling operation. N_b is the number of instances in each batch, which is set to 8 in our experiments.

The proposed network is constructed by stacking convolutional layers and global maximum pooling [18]. The output feature map $\mathbf{Z}$ of each layer c_i is obtained as

$$\mathbf{Z}^{c_i} = \mathrm{ReLU}\big(\mathbf{Z}^{c_{i-1}} * \mathbf{K}^{c_i} + \mathbf{b}^{c_i}\big), \tag{2}$$

where $*$ is the convolutional operator, c denotes the layer index, $\mathbf{K}$ and $\mathbf{b}$ denote the trainable filters and biases, respectively, and the rectified linear unit activation function (ReLU) is defined as $f(x) = \max(0, x)$. In particular, in the first

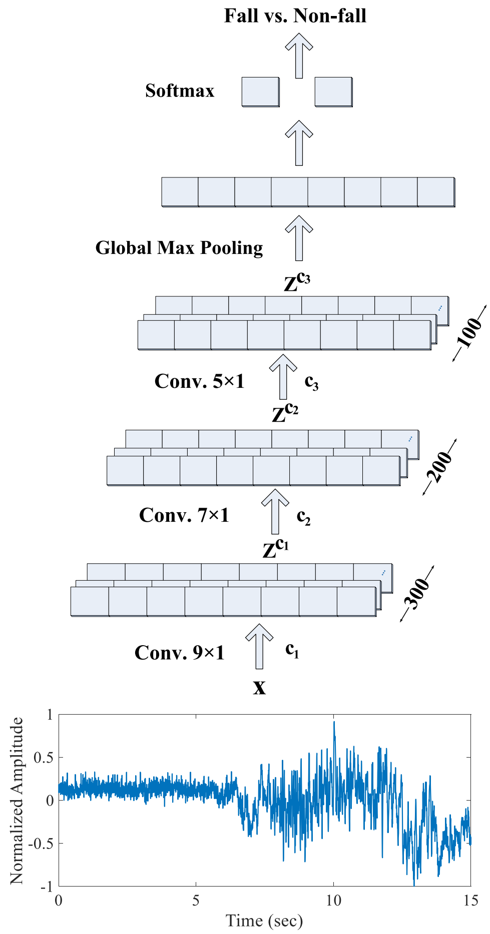


Fig. 2. Architecture of the proposed deep neural network for fall detection.

convolutional layer c_1, 300 filters $\{k_j^{c_1}\}_{j=1}^{300}$ of size 9×1, are convolved with the input time series $\mathbf{x}$ with stride 1. A bias value is added and the activation function is applied to the output resulting in a feature map $\mathbf{Z}^{c_1}$ with a depth of 300. In the second convolution layer c_2, feature map from the previous layer $\mathbf{Z}^{c_1}$ is convolved with 200 filters $\{k_j^{c_2}\}_{j=1}^{200}$ of size 7×1 with stride 1. A bias value is added and ReLU activation function is applied to the output resulting in a feature map $\mathbf{Z}^{c_2}$ with a depth of 200. Finally, in the third convolution layer c_3, feature map from the previous layer $\mathbf{Z}^{c_2}$ is convolved with 100 filters $\{k_j^{c_3}\}_{j=1}^{100}$ of size 5×1 with stride 1. After adding the bias value and applying the ReLU activation function, the resulting feature map $\mathbf{Z}^{c_3}$ has a depth of 100. It should be noted that in order to capture the local temporal information, all the trainable filters used have small sizes, i.e., 9×1, 7×1 and 5×1. The convolutional layers are followed by one pooling layer and one output layer for class prediction. The output of the third convolutional layer $\mathbf{Z}^{c_3}$, is down-sampled by the pooling technique. To this end and in order to lower the spatial dimensionality of the convolved extracted features, a global maximum pooling layer is employed. The pooling mechanism is performed by selecting the maximum value of the convolved features in the last convolutional layer. In the output layer, the softmax activation function is used as

$$h^r = \frac{\exp(\mathbf{Z}^r)}{\sum_{v=1}^{C} \exp(\mathbf{Z}^v)}, \quad \text{for } r = 1, 2 \tag{3}$$

where $C = 2$ is the number of classes, $\mathbf{Z}^r$ 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 class. Adam optimizer is used in the learning process [19] and categorical cross-entropy cost function is used to measure the performance of the model based on the true labels and probabilistic outputs of the softmax function. The entire network is trained in batch mode, i.e., the number of instances evaluated before a weight update in the network, using the back-propagation algorithm to iteratively update the weights and minimize the cost function. It is noted that the network architecture is optimized for a higher classification performance using grid searching and cross validation on the training set, which constructs and evaluates the model for each combination of parameters. To this end, the training set is divided into 3 folds, two folds for training and the other one for validation.

IV. Experimental Results

Experiments were conducted on a set of radar data collected in a realistic room environment to evaluate the performance of the proposed fall detection method. All of the deep learning tasks were implemented using Keras that is backended by TensorFlow package.[1] For comparison purposes, several approaches are considered which are based on multi-layer perceptron (MLP), k-nearest neighbors (KNN) [1] and dynamic time warping (DTW) [20], [21]. It is known that KNN classifier when using DTW as a distance measure, i.e., KNN-DTW, is considered state-of-the-art [20], and provides a better performance than the feature-based methods [20]. In addition, long-short-term-memory recurrent neural networks (LSTM-RNN), known to be a baseline in time series classification, are used for comparison purpose. In the proposed method, the radar return data is first processed to obtain the range-integrated time series data. The resulting time series is fed into the proposed network to test whether or not an specific time series represents a fall incident.

Table II gives classification metrics obtained using the proposed method and those of the other methods, namely, three MLP networks with three hidden layers having 50, 100, neurons in each layer, KNN with different number of neighbors (e.g., $k = 5$ and 10), DTW with different warping window of sizes $W = 5$ and 10, i.e., DTW(5) and DTW(10), and LSTM-RNN. The LSTM-RNN network structure and its hyper-parameters are tuned for the best classification results on a validation set. More specifically, the optimal number of LSTM units at every time step of the network is selected over a range of 8 to 128 units and the number of LSTM layers is selected over a range of 1 to 4 layers. The number of training epochs is set to 10000 epochs and using early-stopping, training is halted as soon as the validation accuracy decreases. In a leave-one-subject-out cross validation, the classifiers are trained using the radar data for four of the subjects and tested using the data from the fifth subject. It is seen from this table that the proposed method outperforms the other methods by providing higher recognition rates in presence of an unseen set

[1]The dataset and code used in this brief work is available at http://meddev.eecs.uottawa.ca/radar.html.

TABLE II
CLASSIFICATION METRICS (%) OF THE PROPOSED FALL DETECTION METHOD AS WELL AS THOSE OBTAINED USING MLP, DTW, KNN AND LSTM-RNN, IN A LEAVE-ONE-SUBJECT-OUT CROSS VALIDATION

Method	Accuracy	Precision	Sensitivity	Specificity
5NN	56.25	100	55.32	100
10NN	60.41	96.15	58.14	80.00
MLP50	87.50	92.31	85.71	90.00
MLP100	85.42	88.46	85.18	85.71
5NN-DTW(5)	83.34	80.77	87.50	79.17
10NN-DTW(5)	87.50	80.77	95.45	80.77
5NN-DTW(10)	87.50	80.77	95.45	80.77
10NN-DTW(10)	85.42	80.77	91.30	80
LSTM-RNN	89.58	88.46	92	86.95
Proposed	**95.83**	**92.31**	**100**	**91.67**

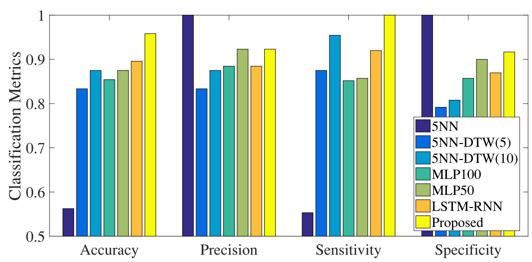
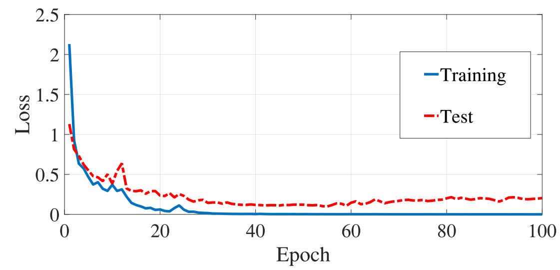


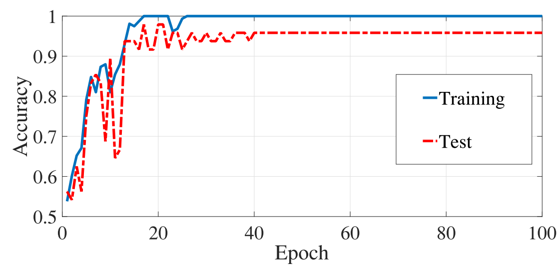
Fig. 3. Classification metrics comparison of the proposed fall detection method and those of the other methods in a subject cross-validation sense.

of data. In particular, the proposed method achieves 95.83% accuracy, 92.31% precision, 100% sensitivity and 91.67% specificity, which are higher than those yielded by the other methods. Remarkably, the proposed method performs better than the best baselines 5NN-DTW and LSTM-RNN by about 10% and 7%, respectively, in terms of accuracy. The superior performance of the proposed method using convolutional neural network is due to the fact that the structure of the convolutional model captures the nature of signals and their local relationships in feature representation more accurately than the other methods for the fall detection problem. In addition, in terms of computational complexity, the proposed method has a lower computational complexity compared to 5NN-DTW and LSTM-RNN, since its required CPU time on an Intel Core i7 2.93 GHz personal computer with 16 GB RAM is 30 sec per each training epoch and 2.36 sec at the test time, while 5NN-DTW runs in 1708 sec and the required CPU time for LSTM-RNN is 43 sec per each training epoch and 3.51 sec at the test time. In addition, the number of trainable parameters needed for the proposed convolutional neural network is 531,702, while that needed for the LSTM-RNN based method is 1,602,048. Fig. 3 illustrates classification metrics obtained using the proposed method as well those obtained using the other methods in a one-subject-out cross-validation sense. It is seen from this figure that the proposed method provides higher values for accuracy and sensitivity than the other methods.

Fig. 4 (a) shows cross-entropy loss values on the training and test sets obtained using the proposed method. It is known that the main objective in a learning model is to reduce (minimize) the loss function's value with respect to the model's parameters through optimization process. Thus, the gradually decreasing trend of loss seen from this figure demonstrates



(a)



(b)

Fig. 4. (a) Cross-entropy loss, and (b) accuracy values for the training and test sets, obtained using the proposed deep neural network.

TABLE III
CLASSIFICATION METRICS (%) OF THE PROPOSED FALL DETECTION AND THOSE USING THE OTHER METHODS IN A 5-FOLD CROSS-VALIDATION

Method	Accuracy	Precision	Sensitivity	Specificity
5NN	66.99	95.04	64.97	79.31
10NN	65.53	90.90	64.70	69.44
MLP50	69.90	78.51	72.52	65.33
MLP100	69.42	77.68	72.31	64.47
5NN-DTW(5)	86.89	89.26	88.52	84.52
10NN-DTW(5)	86.41	87.60	89.07	82.76
5NN-DTW(10)	91.26	91.73	93.28	88.50
10NN-DTW(10)	86.89	85.12	84.42	88.85
LSTM-RNN	88.35	89.26	90.76	85.06
Proposed	**92.72**	**94.21**	**93.44**	**91.67**

that the network is successfully trained after a fixed number of iterations, i.e., epochs. The results are averaged over 20 runs, and in each run, the number of epochs is set to 100. Fig. 4 (b) depicts the accuracy values on the training and test dataset obtained using the proposed method. It can be seen from this figure that the accuracy values improve as epoch value increases and reach a steady state after about 50 epochs.

In addition, to further evaluate the recognition performance of the proposed method, the dataset is partitioned into five subsets and a subset is used for testing, whereas the other four subsets for training, i.e., 5-fold cross-validation. Table III gives classification metrics obtained using 5-fold cross-validation for the proposed fall detection method and those obtained using the other methods. It is seen from this table that the proposed method is capable of detecting falls with higher accuracy, sensitivity and specificity values, indicating the capacity to better detect a fall incident when it occurs to avoid false alarms. In particular, the proposed method outperforms the other methods by providing 92.72% classification accuracy which is higher than that of the state-of-the-art.

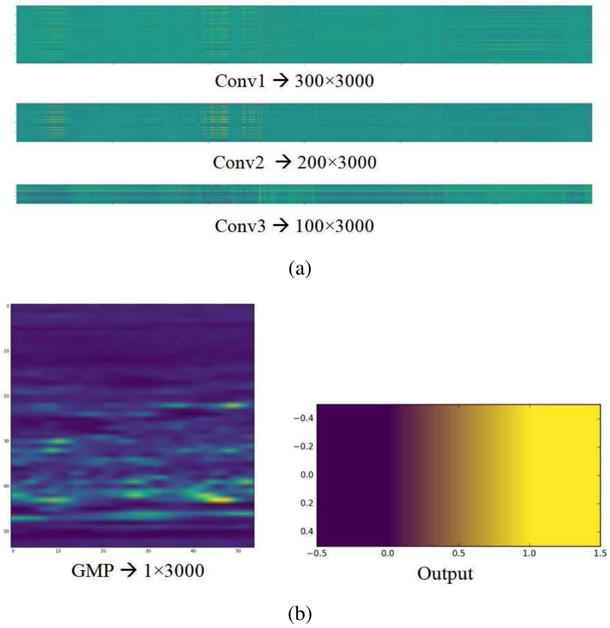


Fig. 5. (a) Feature maps (activations) of the three convolutional layers of the proposed model, and (b) activations of global max pooling and output layer, when a time series from the test set is fed into the trained network.

To understand automatically extracted features at each layer, feature maps (activations) of the convolutional layers, global max pooling and output layer are depicted. Fig. 5 shows activation of the three convolutional layers for one sample time series as well as those of the global max pooling and output layer. It is seen from this figure how a time series is decomposed into different filters learned by the network. It is noted that after global max pooling, activation becomes increasingly abstract and less visually interpretable. It is also noted that activations in the higher layers carry increasingly less information about the visual contents of the time series, and more information related to the class of the time series.

V. Conclusion

A new fall detection method has been proposed using an ultra wideband radar and a supervised learning approach based on deep neural network. Radar data has been collected in a room environment by considering the home healthcare setting. The proposed network have been devised by deriving a time series from the radar back-scattered matrix and feeding it to a deep convolutional network to automatically learn multi-level features from radar time series data. Experiments have been conducted to assess the performance of the proposed method and to compare it with that of state-of-the-art. The results have demonstrated that the proposed fall detection method outperforms LSTM-RNN, MLP and KNN-DTW algorithms in terms of providing higher classification metrics and significantly lower CPU time. It has also been shown that the proposed deep fully convolutional neural network-based method by extracting multi-level features from radar time series data can circumvent heuristic feature extraction used in time-frequency analysis of the radar return signals. The significant improved performance of the proposed method especially in leave-one-subject-out cross validation indicates its generalization capability.

References

[1] 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.

[2] 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.

[3] A. T. Özdemir and B. Barshan, "Detecting falls with wearable sensors using machine learning techniques," *Sensors*, vol. 14, no. 6, pp. 10691–10708, 2014.

[4] 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.

[5] Z. Baird, "Human activity and posture classification using single non-contact radar sensor," M.S. thesis, Syst. Comput. Eng., Carleton Univ., Ottawa, ON, Canada, 2017.

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

[7] A.-K. Seifert, M. G. Amin, and A. M. Zoubir, "New analysis of radar micro-Doppler gait signatures for rehabilitation and assisted living," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2017, pp. 4004–4008.

[8] 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.

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

[10] L. Liu *et al.*, "Automatic fall detection based on Doppler radar motion signature," in *Proc. Int. Conf. Pervasive Comput. Technol. Healthcare (PervasiveHealth)*, 2011, pp. 222–225.

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

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

[13] D. Zhou, X. Wang, Y. Tian, and R. Wang, "A novel radar signal recognition method based on a deep restricted Boltzman machine," *Eng. Rev.*, vol. 37, no. 2, pp. 165–171, 2017.

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

[15] Y. Wang *et al.*, "Gait-based human identification using acoustic sensor and deep neural network," *Future Gener. Comput. Syst.*, vol. 86, pp. 1228–1237, Sep. 2018.

[16] A. Van Den Oord *et al.*, "WaveNet: A generative model for raw audio," in *Proc. SSW*, 2016, p. 125.

[17] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 1243–1252.

[18] L. Min, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

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

[20] S. Skyler, W. Zhang, and Y. Zhou, "Multivariate time series classification using dynamic time warping template selection for human activity recognition," in *Proc. IEEE Symp. Series Comput. Intell.*, 2015, pp. 1399–1406.

[21] H. Yang, Y. Yang, W. Xu, and Y. Pang, "An automatic fall detection system based on derivative dynamic time warping, in *Proc. Int. Symp. Parallel Archit. Algorithm Program. (PAAP)*, 2017, pp. 427–438.