Deep Learning of Micro-Doppler Features for Aided and Unaided Gait Recognition

Mehmet Saygın Seyfioğlu[1], Sevgi Zübeyde Gürbüz[2], Ahmet Murat Özbayoğlu[3], Melda Yüksel[1]
[1]Dept. of Electrical-Electronics Engineering, TOBB University of Economics and Technology, Ankara, Turkey
[2]Dept. of Electrical and Computer Engineering, Utah State University, Logan, UT, USA
[3]Dept. of Computer Engineering, TOBB University of Economics and Technology, Ankara, Turkey
msseyfioglu@etu.edu.tr, sevgi.gurbuz@usu.edu, mozbayoglu@etu.edu.tr, yuksel@etu.edu.tr

***Abstract*—Remote health monitoring is a topic that has gained increased interest as a way to improve the quality and reduce costs of health care, especially for the elderly. Falling is one of the leading causes for injury and death among the elderly, and gait recognition can be used to detect and monitor neuromuscular diseases as well as emergency events such as heart attack and seizures. In this work, the potential for radar to discriminate a large number of classes of human aided and unaided motion is demonstrated. Deep learning of micro-Doppler features is used with a 3-layer auto-encoder structure to achieve 89% correct classification, a 17% improvement in performance over the benchmark support vector machine classifier supplied with 127 pre-defined features.**

Keywords—micro-Doppler classification; radar; deep learning

I. Introduction

Remote indoor health monitoring is a topic that has been gaining increased interest in recent years as a way to improve the healthcare of outpatients, quality of life of elderly, and treatment of patients with chronic diseases. Each year, over one-third of elderly people fall in their own homes or in nursing homes. More than half of elderly who remain untreated for more than an hour after a fall ultimately die within 6 months of the fall [1-2]. Rapid response is critical to minimizing complications and thus costs of health related emergencies, including not just falling, but also heart attack, brain stroke, and other life threatening illnesses. Moreover, as gait analysis is often used in the treatment and assessment of neuromuscular and other disorders, continual in-home gait and life signs monitoring can also provide doctors with a baseline of patient health and vital additional data, outside of that obtained through in-hospital visits, which can be used to improve treatment. Thus, remote health monitoring could potentially revolutionize the way healthcare is provided today, saving both lives and money.

Until recently, primarily wearable sensors, video cameras, infrared, seismic or thermal sensors have been researched for remote heath monitoring. With the advent of small, wireless software defined radio platforms in the mid-2000's, however, radar has begun to be explored for use within wireless sensor systems. In particular, radar-based fall detectors have a subject of intense research. Radar offers unique advantages in comparison to other sensors; namely, it is capable of operating with little or no light and can only sense motion – it cannot record conversation or any personal visual information. This makes its utilization in environments requiring privacy, such as residences or hospitals, quite beneficial.

Through exploitation of the micro-Doppler effect, radar can detect much more than just falling, but in fact can recognize a wide range of indoor activities. The micro-Doppler effect refers to frequency modulations centered about the main Doppler frequency that are caused by the vibration or rotation of parts of a moving object. The periodic motion of arms and limbs that occurs during an activity generates unique patterns in the time-frequency distribution of the radar return, which can then be used to classify measurements.

The classification process typically begins with the extraction of a pre-defined set of features from either the raw data or time-frequency transform (spectrogram) of the data. Examples of such pre-defined features include physical features [3], discrete cosine transform coefficients [4], linear predictive coding coefficients [5], and cepstral coefficients [6], among numerous others [7]. A subset of the extracted features may then be chosen using feature selection [8] or dimension reduction [9] before being supplied to a classifier of choice, such as support vector machines (SVM), which is a quite popular choice.

The use of pre-defined features, however, has the drawback that the ability of the feature to discriminate between classes is highly dependent upon the activities being considered and number of classes. An alternate approach is offered by deep learning, which learns features directly from the data itself. In one of the first works on deep learning for micro-Doppler, Kim and Moon [10] used a convolution neural network (CNN) to classify 6 visually distinguishable activities (running, walking, walking while holding a stick, crawling, boxing while moving forward, boxing while standing in place, and sitting still) at a correct classification rate of 90.9%, a rate quite similar to that previously achieved with SVM [11]. Later, Jokanovic [12] used a 2-layer autoencoder structure with a total of 120 data samples to classify 4 classes of activities: falling, sitting, bending, and walking. Each spectrogram had a size of

76x70, resulting in an input vector dimensionality of 5320 and a correct classification rate of 87%.

The focus of this work is the discrimination of highly similar classes of aided and unaided walking, as might be encountered especially in assisted living environments for the elderly. Thus, twelve classes are considered: 1) walking, 2) limping, 3) walking using a cane, 4) walking using a walker, 5) walking using crutches with one leg bent at the knee, 6) jogging, 7) creeping, 8) crawling, 9) falling from an upright position, 10) falling from a chair, 11) quickly sitting down, and 12) using a wheelchair. Note that the micro-Doppler signatures of many of these activities are not visually distinguishable from each other. A 3-layer auto-encoder structure is used in this work to classify these twelve classes with a significant performance improvement in comparison to a multi-class majority voting SVM classifier supplied with 127 pre-defined features. For both classification tasks, 10 fold cross validation is employed in order to have robust classification accuracies.

In Section II, the experimental test setup and micro-Doppler recordings collected are described. In Section III, the proposed 3-layer auto-encoder design is detailed, followed by experimental results for both the auto-encoder and SVM methods in Section IV. Section V concludes with discussion of the results and plans for future directions.

II. Radar Micro-Doppler Measurements

Radar measurements for this study was conducted by programing a NI-USRP 2922 model software defined radio platform to transmit a 4 GHz continuous wave signal. The USRP-2922 has a 20 MHz bandwidth and is capable of broadcasting between frequencies of 400 MHz and 4.4 GHz. Two SAS-571 model horn antennas having 48° azimuthal beam width were used for transmit and receive. Both antennas and USRP were mounted upon a vertical panel raised 1 meter above the ground (Figure 1) and pointing directly in-line with the direction of motion for each data collect.

A total of 10 test subjects were utilized to collect a database of 864 measurements, each measurement 4 seconds in duration. Walking was enacted with both arms swinging in a mild fashion, while limping was done by role-playing that the left leg was injured and dragging behind the right. Two types of canes were used – a single poled cane and tripod – from which were randomly selected in the final database. Usage of a cane primarily constrains the motion of one arm; however, due to the subconscious coupling of arm motion, movement of the left arm while using a cane with the right hand is of lower amplitude than that experience in normal walking. Movement with the aid of a walker, on the other hand, physically constrains both arms. Use of crutches also constrains both arms, but signatures may experience more Doppler bandwidth due to the motion of the crutches. Crawling was performed both hands and knee on the ground and moving slowly. Creeping was enacted with the whole body on the ground and advancing lizard-style by pushing with both arms and legs. Falling was enacted by tripping on a brick and falling forward onto a

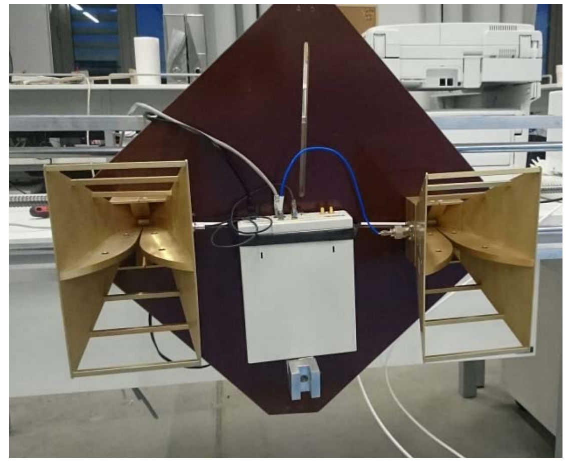

Figure 1. Panel-mounted CW radar.

mat placed on the ground. Falling from chair was performed by falling sideways on a mat. Rapid sitting was performed by the actor throwing himself backwards towards the chair without restraint. A manual wheelchair was employed in the experiments, with test subjects turning the wheel rims with their hands to move forward.

A short-time Fourier transform (STFT), or spectrogram, is used to represent the time-frequency distribution of the return signal. Example spectrograms for several activities are shown in Figure 2. Spectrograms were generated using a window length of 2048 samples, 4096 FFT points, and an overlap of 128 samples. Afterwards, the spectrograms were resized to 121 x 91 and converted to gray scale, resulting in an initial input vector to the auto-encoder with a dimension of 11,011.

III. Deep Auto-Encoder Neural Network

The auto-encoder is an unsupervised neural network, which aims to learn a representation of its inputs. For a given training vector $\boldsymbol{x}$, an auto-encoder network aims to adjust its weights and biases to achieve $\boldsymbol{h_w}(\boldsymbol{x}) \approx \boldsymbol{x}$. However, if one would add constraints to the network, the network would be forced to learn a compressed representation of the input vector. For example, if the number of neurons in the first layer is selected smaller than length of the input vector $\boldsymbol{x}$, the network naturally compresses the input or one would add sparsity parameters to the cost function to have a sparse representation of the input. Therefore, the auto-encoder network can be used for the purposes of automatic feature extraction or dimension reduction.

A new training strategy for deep auto-encoders, proposed in [13], uses semi-supervised learning. In semi-supervised learning there are two main stages: first, the unsupervised layer-by-layer pre-training takes place, where the network is trying to reconstruct its inputs at the output subject to the given constraints. This stage ensures that the network weights are initialized not randomly, but using the

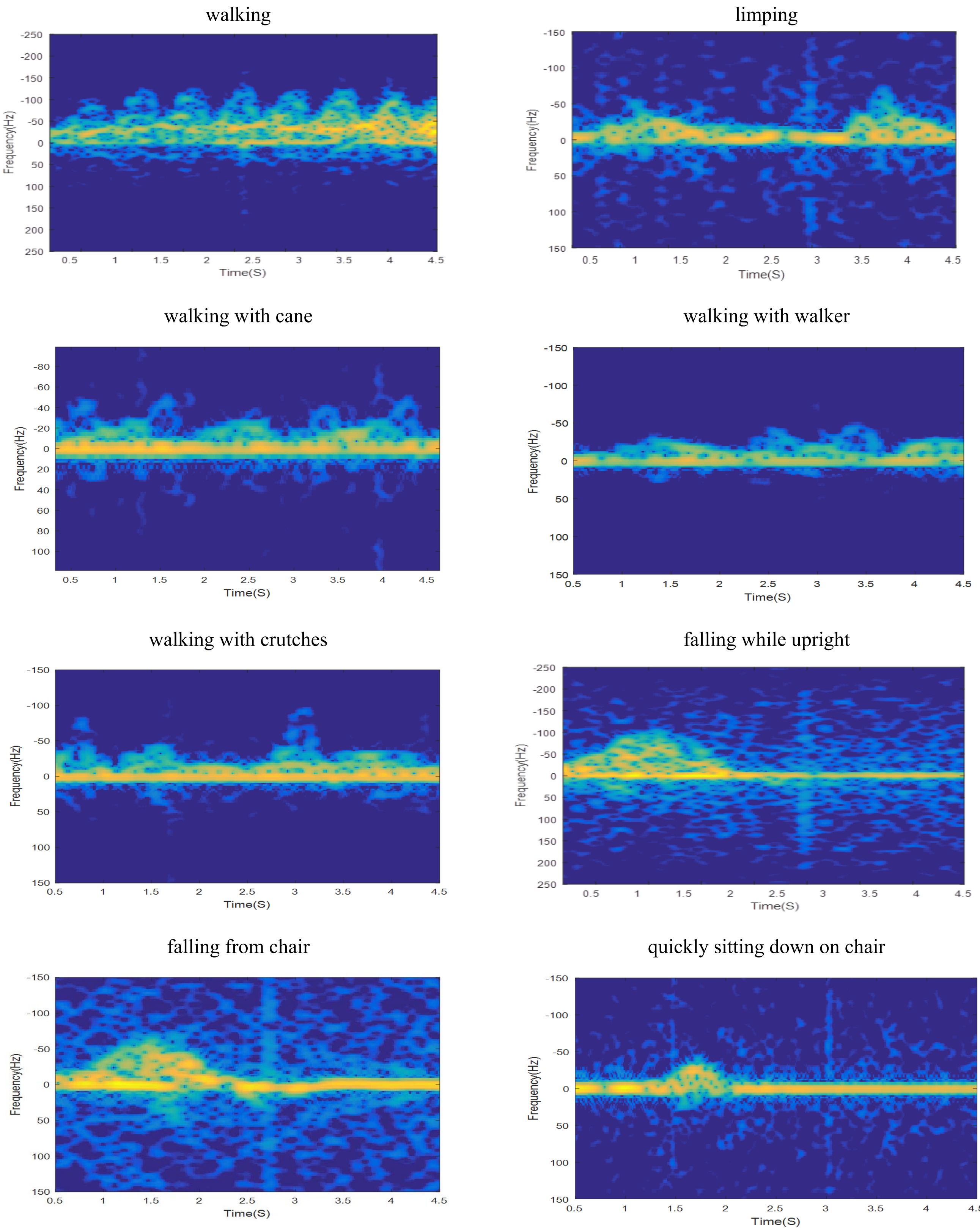

Figure 2. Example spectrograms of measured radar data for similar aided and unaided human activities.

information from the input vector. Therefore, unsupervised pre-training helps to avoid bad local minima in the non-convex optimization space. The unsupervised pre-training stage is applied in layer-by-layer fashion in which the previous layer's output is the next layer's input.

Since auto-encoder networks are very flexible, they tend to suffer from over-fitting. Thus, a regularization parameter is needed to prevent over-fitting. Also, to ensure a sparse representation of the given input vector $\boldsymbol{x}$, one can add a sparsity parameter to the cost function. The cost function is defined as follows:

$$E = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}(x_{kn} - \bar{x}_{kn})^2 + \dot{\rho}_{weight} + \beta * \dot{\rho}_{sparsity} \qquad (1)$$

where $\sum_{n=1}^{N}\sum_{k=1}^{K}(x_{kn} - \bar{x}_{kn})^2$ denotes the mean square error, $\dot{\rho}_{weight}$ represents L_2 regularization parameter and $\dot{\rho}_{sparsity}$ denotes the sparsity parameter.

The L_2 regularizer prevents the network weights from having large values. The sparsity parameter ensures that some of the neurons stay inactive during the pre-training process. Thus, a sparse representation of the inputs is provided. β is the scaling parameter of the sparsity regularizer. All of these regularizer terms are hyper-parameters and their optimal values are problem variant.

To minimize the cost function, Scaled Conjugate Gradient (SCG) method is employed during backpropagation. In SCG, the gradient of the error follows the conjugate of the previous steps. At the last layer, the classification results are obtained by using the Softmax Classifier [14] which yields the probability of a given sample to belong each class label. After layer-by-layer unsupervised pre-training, a global fine tuning process is applied to the network. In this step, the network is trained in supervised manner, in which the weights are 'fine-tuned' by using the class label information. In this work, three auto-encoders are stacked to form a deep network, where layers have 400-150-50 neurons, respectively, from the first layer to third. Lastly, at the end of the network, a Softmax classifier with 12 neurons is employed to estimate the label probability of each data sample.

IV. EXPERIMENTAL RESULTS

The performance of the auto-encoder structure described in Section III is compared with that obtained from an SVM classifier supplied with 127 different predefined features; namely, 13 physical features, 3 cepstral coefficients, 10 DCT coefficients and 101 LPC coefficients. A performance comparison of the deep auto-encoder neural network and baseline SVM classifier is shown in confusion matrices provided in Tables 1 and 2. Also an overall comparison between two methods is shown in Table 3.

The SVM classifier has an overall correct classification rate of 72%, while the deep auto-encoder neural network has significantly higher performance at a rate of 89%. The SVM classifier exhibits the most confusion between activities such as walking-creeping-crawling, wheelchair-cane-walker-limping, creeping-crawling-jogging, and falling while upright-falling from a chair. Some of these confusions are not surprising, such as confusion between creeping and crawling, because it is difficult for many subjects to move military-style low to the ground. Any upward motions or attempt to use the knees to advance would render the enactment potentially similar to crawling. Also, activities such as use of cane and walker or limping are quite similar in micro-Doppler signature (see Figure 2). However, it would have been hoped that short-duration activities like falling would not be confused with walking, and prior work [15] has shown that generally wheelchair usage should be easily differentiated.

The 3-layer auto-encoder topology proposed in this work handily identifies the most distinct activities at a perfect 100% rate or quite close to that for wheelchair usage, jogging, falling while upright, falling while sitting, and quickly sitting. Identification of walking, using a walker or using crutches is also correctly done for over 90% of the data. The primary sources of confusion are between creeping and crawling, as well as use of a cane versus a walker. Nevertheless, a tremendous classification improvement of 17% is achieved using the learned features of the auto-encoder as opposed to pre-defined features with SVM.

V. DISCUSSION AND CONCLUSIONS

The groupings of activities that are confused in by the auto-encoder are indeed quite similar in nature – to distinguish them a substantial increase in the amount of training data is needed. Indeed, a rule of thumb for the ideal training data size for deep networks is 10 times the degrees of freedom, which in this work is the number of neurons used in each layer, 400x150x50. This gives a required training data size of over $3\text{x}10^7$ samples!

Herein lays the principle challenge of implementing deep learning on radar data: how to collect a sufficiently large data set, containing a sufficient number of people of varying height and weight, so as to obtain a data set reflecting the full spectrum of statistical variation across different realizations of the same activity?

Collecting thirty million samples of data just for the training set is not a feasible or cost-effective process. In extensions of this work, we plan to consider alternative methods for building up a large training database, including using micro-Doppler simulations for training, which has shown some success in previous work [16], as well as exploiting multiple sensors in the training process.

Despite issues that need to be addressed in establishing the requisite training set, this work has shown that deep learning of micro-Doppler features can be a powerful technique for differentiating between and identifying human signatures that are so similar so as to be indistinguishable by the human eye.

Table 1. Confusion matrix for SVM classifier given pre-defined features.

%	Walking	Wheel-chair	Limping	Cane	Walker	Falling	Using Crutches	Creeping	Crawling	Jogging	Fast sitting	Falling from chair
Walking	98.0	0	0	0	0	0	0	0.5	0.2	1.3	0	0
Wheel chair	0	78.3	5.3	8.9	5.4	0	1.4	0.3	0.5	0	0	0
Limping	0	9.0	65.0	20.1	3.5	0	2.4	0	0	0	0	0
Cane	0	13.1	12.1	61.5	13.3	0	0	0	0	0	0	0
Walker	0	5.5	5.2	9.8	79.5	0	0	0	0	0	0	0
Falling	20.6	0.6	0.6	0	0	52.5	0	0	1.9	0	1.3	22.5
Using Crutches	0.4	2.8	3.0	0	0	0	91.3	0	2.4	0	0	0
Creeping	29.6	0	0	0	0	0	1.2	40.8	22.7	5.8	0	0
Crawling	30.4	4.6	6.4	0	0.7	0	0.7	16.1	30.1	11	0	0
Jogging	2.4	0	0	0	0	0	0	0	2.1	95.5	0	0
Fast sitting	7.5	0	0	0	0	1.5	0	0	0	0	86	5.0
Falling from chair	3.3	0	0	0	0	2.9	0	0	0.4	1.3	5.0	87.1

Table 2. Confusion matrix for deep auto-encoder neural network.

%	Walking	Wheel chair	Limping	Cane	Walker	Falling	Using Crutches	Creeping	Crawling	Jogging	Fast sitting	Falling from chair
Walking	90.3	0	0	0	0	0	0	6.5	0	3.2	0	0
Wheel chair	0	98.2	0	0	1.8	0	0	0	0	0	0	0
Limping	0	0	88.5	7.7	3.8	0	0	0	0	0	0	0
Cane	0	0	11.9	71.4	16.7	0	0	0	0	0	0	0
Walker	0	0	0	7.7	92.3	0	0	0	0	0	0	0
Falling	0	0	0	0	0	100	0	0	0	0	0	0
Using Crutches	7.7	0	0	0	0	0	92.3	0	0	0	0	0
Creeping	0	0	0	0	0	0	0	64.3	35.7	0	0	0
Crawling	0	0	0	0	0	0	0	20	70	10	0	0
Jogging	0	0	0	0	0	0	0	0	0	100	0	0
Fast sitting	0	0	0	0	0	0	0	0	0	0	100	0
Falling from chair	0	0	0	0	0	0	0	0	0	0	0	100

Table 3. Precision, recall and accuracy values for both baseline SVM and proposed Autoencoder methods

	SVM			Autoencoder		
Activity	Precision	Recall	Accuracy	Precision	Recall	Accuracy
Walking	0.98	0.51	98	0.90	0.92	90.3
Wheel chair	0.78	0.69	78.3	0.98	1	98.2
Limping	0.65	0.66	65.0	0.89	0.88	88.5
Cane	0.61	0.61	61.5	0.71	0.82	71.4
Walker	0.79	0.78	79.5	0.92	0.81	92.3
Falling	0.52	0.92	52.5	1	1	100
Using Crutches	0.91	0.94	91.3	0.92	1	92.3
Creeping	0.41	0.70	40.8	0.64	0.71	64.3
Crawling	0.30	0.50	30.0	0.7	0.66	70
Jogging	0.96	0.83	95.5	1	0.88	100
Fast sitting	0.86	0.93	86.0	1	1	100
Falling from chair	0.87	0.76	87.1	1	1	100
Average	0.72	0.74	72.13	0.89	0.89	88.94

REFERENCES

[1] Lord, S.R., Sherrington, C., and Menz, H.B., Falls in Older People: Risk Factors and Strategies for Prevention, Cambridge University Press, 2001.

[2] Wild, D., Nayak, U.S., Isaacs, B., “How dangerous are falls in old people at home?” British Medical Journal (Clinical Research), 1981, pp. 282:266-8.

[3] Kim, Y., Ha, S., and Kwon, J., “Human detection using Doppler radar based on physical characteristics of targets,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, iss. 2, 2015, pp. 289-293.

[4] Molchanov, P., Astola, J., Egiazarian, K., Totsky, A., "Ground moving target classification by using DCT coefficients extracted from micro-Doppler radar signatures and artificial neuron network," in Proc. *Microwaves, Radar and Remote Sensing Symposium (MRRS),* 25-27 Aug. 2011, pp.173-176.

[5] Javier, R.J., and Kim, Y., “Application of linear predictive coding for human activity classification based on micro-doppler signatures,” IEEE Geoscience and Remote Sensing Letters, vol. 11, iss. 10, 2014.

[6] Yessad, D., Amrouche, A., Debyeche, M., and Djeddou, M., “Micro-doppler classification for ground surveillance radar using speech recognition tools,” in *Proc. 16th Iberoamerican Congress*, Pucon, Chile, November 15-18, 2011.

[7] Gürbüz, S.Z., Erol, B., Çağlıyan, B., Tekeli, B., “Operational assessment and adaptive selection of micro-Doppler features,” *IET Radar Sonar and Navigation*, vol. 9, iss. 9., dec. 2015, pp. 1196-1204.

[8] Tekeli, B, Gürbüz, S.Z., Yüksel, M., “Information Theoretic Feature Selection for Human Micro-Doppler Signature Classification,” *IEEE Trans. Geoscience and Remote Sensing*, Vol. 54, Iss. 5, Jan. 2016.

[9] Zabalza, J., Clemente, C., Di Caterina, G., Ren, J., Soraghan, J.J.; Marshall, S., "Robust PCA micro-doppler classification using SVM on embedded systems," in IEEE Trans. on *Aerospace and Electronic Systems,* vol.50, no.3, pp.2304-2310, July 2014.

[10] Kim, Y. and Moon, T., "Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks," in *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 8-12, Jan. 2016.

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

[12] Jokanovic, B., Amin, M., and Ahmad, F., "Radar fall motion detection using deep learning," in Proc. *IEEE Radar Conference*, Philadelphia, PA, 2016.

[13] Hinton, G., Salakhutdinov, R., “Reducing the Dimensionality of Data with Neural Networks,” Science, vol. 313, 2006.

[14] Ng, A., Ngiam, J., Foo, C.Y., Mai, Y., and Suen, C., *Unsupervised Feature Learning and Deep Learning Tutorial*, 2013.

[15] Gürbüz, S.Z, Clemente, C., Balleri, A., Soraghan, J., “Micro-Doppler Based In-Home Aided and Unaided Walking Recognition with Multiple Radar and Sonar Systems,” *IET Radar, Sonar, Navigation*, June 2016.

[16] Karabacak, C., Gürbüz, S.Z., Gürbüz, A.C., Guldogan, M.B., Hendeby, G., Gustafsson, F., “*Knowledge exploitation for human micro-doppler classification*,” IEEE Geoscience and Remote Sensing Letters, Vol. 12, No. 10, October 2015, pp. 2125-2129.