Semi-Supervised Convolutional Autoencoder With Attention Mechanism for Activity Recognition

Zahra Sadeghi-Adl and Fauzia Ahmad
Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA
E-mail:{zahra.sadeghi.adl, fauzia.ahmad}@temple.edu

***Abstract*—In this paper, we consider human activity recognition with a semi-supervised convolutional autoencoder (CAE), augmented by an attention mechanism, using radar micro-Doppler signatures. The attention-augmented CAE (AA-CAE) learns both global information and spatially localized features, thus enabling the classifier to overcome the limited receptive field of a CAE. Considering training data comprising both labeled and unlabeled samples, a semi-supervised training regime is implemented through a joint loss function, with training of the encoder part performed in an unsupervised fashion using all training samples and the classifier and attention mechanism trained at the same time using only the labeled samples. Using real-data measurements of six different human activities, we demonstrate that the jointly trained AA-CAE classifier yields higher classification accuracy with fewer labeled data than the semi-supervised AA-CAE trained via a conventional two-step process.**

***Index Terms*—Human activity recognition, micro-Doppler, machine learning, latent-variable models, semi-supervised learning.**

I. Introduction

Intelligent healthcare and remote monitoring technologies for human wellness have witnessed remarkable progress, owing to advances in smart sensors and evolution of high-speed data processors [1]–[6]. Over the past decade, the demand for these technologies has grown with the objectives of providing improved chronic disease management [7], and addressing both shortage of healthcare providers and high cost of healthcare services [8]. Radar modality provides a viable solution for remote patient monitoring and eldercare due to its non-contact and privacy-aware nature, accuracy, and robustness against various operating conditions [9].

Gross motor activity recognition and vital signs measurements lie at the core of radar-based remote monitoring systems. A majority of the activity classification methods employ discriminant features extracted primarily from micro-Doppler signatures of distinct human activities [10]–[17]. Recently, a plethora of deep learning methods have been proposed which leverage the power of artificial neural networks to automatically learn features from the micro-Doppler signatures and accurately classify various activities [18]–[24]. Deep convolutional neural network (CNN) based architectures were proposed for fall detection in [18] and to classify major sleep stages in [19]. In [20], fall detection was achieved using a channel attention network for fusion of features extracted from micro-Doppler signatures with pre-trained Alexnet, VGG-16-Net, and VGG-19-Net models. In [21], a deep convolutional generative adversarial network was used to overcome the limited number of training samples for fall detection. A convolutional autoencoder (CAE) was augmented with an attention mechanism in [22] to capture global and local features from micro-Doppler signatures for human activity recognition. Assuming the entire training dataset to comprise labeled samples, the encoder part was trained in an unsupervised manner first to reconstruct the input data, i.e., representation learning. Next, the decoder was removed and fully-connected layers of the classifier were attached, followed by supervised training of the attention and classification layers using all training samples. Considering semi-supervised training data comprising labeled and unlabeled sets, a two-step training process was implemented for the attention-augmented(AA)-CAE-based classifier in [25]. Therein, the entire training dataset was first used to train the encoder part in an unsupervised fashion, followed by training of the attention mechanism and the classifier using only the labeled samples. In [23], an attention mechanism was presented to improve the classification performance of Alexnet without significantly increasing the computational load. A semi-supervised transfer learning algorithm that combined unsupervised domain adaptation with supervised semantic transfer was proposed in [24] for human activity recognition. A detailed review of deep learning based human activity recognition is provided in [26].

Similar to the work in [25], we consider in this paper semi-supervised training of an AA-CAE model for human activity recognition using a mix of unlabeled and labeled micro-Doppler signatures. However, unlike [25] where training of the encoder and that of the attention mechanism plus classifier proceeded sequentially, we introduce a joint training method, inspired by [27], which enables simultaneous determination of the classifier weights and global feature maps using labeled samples only and the local feature maps from all samples. In so doing, representation learning can be influenced by the overarching classification task, thereby benefiting the classifier training. Based on real data measurements of six human activities, we show that the proposed semi-supervised AA-CAE yields higher classification accuracy compared to conventionally-trained AA-CAE using fewer labeled samples.

The remainder of this paper is organized as follows. In Section II, we present the human micro-Doppler signatures. In Section III, we review AA-CAE model and describe the proposed joint semi-supervised training method. Performance of the proposed model in terms of activity classification

accuracy is evaluated in Section IV. Concluding remarks are provided in Section V.

II. Human Micro-Doppler Signatures

Consider a frequency-modulated continuous-wave (FMCW) radar with transmitted signal

$$s_T(t) = A_T(t)\cos[2\pi(f_c t + \frac{1}{2}\alpha t^2)], \quad (1)$$

where $A_T(t)$ is the amplitude, f_c is the carrier frequency, and α is the chirp rate. The radar return from a moving point target can be expressed as

$$s_R(t) = A_R(t)\cos[2\pi(f_c(t-\tau) + \alpha(\frac{1}{2}t^2 - \tau t) + f_D t)], \quad (2)$$

where $A_R(t)$ is the received signal amplitude, τ is the two-way travel time, and f_D is the Doppler shift. The received signal is demodulated by the I/Q demodulator, providing the in-phase and quadrature-phase components of the complex baseband signal as

$$s(t) = I(t) + jQ(t) = A(t)e^{j2\pi((f_D - \alpha\tau)t - f_c\tau)}, \quad (3)$$

where $A(t)$ is the amplitude of $s(t)$. For the underlying activity recognition problem, the human can be viewed as a collection of point scatterers. As such, the corresponding return is the superposition of individual point-target returns of the form of (3),

$$s(t) = \sum_i A_i(t)e^{j2\pi((f_{D_i} - \alpha\tau_i)t - f_c\tau_i)}. \quad (4)$$

After sampling, the complex baseband signal in (4) can be arranged as a two-dimensional matrix, $s(n,m)$, where n and m represent fast-time and slow-time, respectively. The range map, $R(p,m)$, is computed by taking the discrete Fourier transform (DFT) along the matrix columns as

$$R(p,m) = \frac{1}{N}\sum_{n=0}^{N-1} s(n,m)e^{-j2\pi\frac{pn}{N}}, \quad (5)$$

where N is the number of samples (range bins) in one pulse repetition interval, $p = 0,\ldots,N-1$, and $m = 0,\ldots,M-1$, with M representing the total number of considered intervals. To obtain the micro-Doppler signature, we first sum the data over the range bins of interest as

$$v(m) = \sum_{p=p_1}^{p_2} R(p,m), \quad (6)$$

where p_1 and p_2 denote the minimum and maximum range bins considered. Then, we apply the Short-Time Fourier Transform (STFT) to $v(m)$, followed by computation of the squared-magnitude of the STFT, i.e., spectrogram, to yield the corresponding micro-Doppler signature, $D(n,k)$,

$$D(n,k) = |\sum_{m=0}^{L-1} V(n-m)w(m)e^{-j2\pi\frac{mk}{L}}|^2, \quad (7)$$

where $w(m)$ denotes a window function that determines the trade-off between time and frequency resolutions [28].

III. Proposed Methodology

A CAE is a type of neural network that represents the input (micro-Doppler signatures for the application at hand) in a latent-variable space using "local" feature maps produced by convolutional layers in the encoder block. To reduce the computational complexity, each convolutional layer in the encoder is typically followed by a pooling layer. The decoder block of the CAE employs a convolution transpose layer that recreates the input from the encoder output, with an unpooling layer following each convolution transpose layer in the decoder. The AA-CAE architecture adds an attention layer which uses the input data in its entirety to construct a query, thereby obtaining global information that is appended to the encoder output [22].

A. Convolutional Autoencoder

Consider a single-layer CAE. Given an $h \times w$ input image x (micro-Doppler signature including the bias term) and P filters, the pth convolutional map, o^p, can be expressed as,

$$o^p = \sigma(x * \omega_e^p), \quad p = 0,1,\ldots,P-1 \quad (8)$$

where '$*$' represents two-dimensional (2-D) convolution, σ is the activation function, and ω_e^p is the pth convolutional filter. The output of the convolutional layer can, thus, be represented as

$$O = \text{concat}[o^0,...,o^{P-1}]. \quad (9)$$

To reconstruct x, a convolution transpose layer is utilized in the decoder. That is, the reconstructed image x' is obtained as

$$x' = \sigma(O * W_d), \quad (10)$$

where W_d is the 2-D filter of the convolution transpose layer.

Unsupervised training can be applied to the CAE which amounts to minimizing the reconstruction error, defined as

$$\sum_{x_i \in X} \|x_i - x_i'\|_F^2, \quad (11)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, X is the set of all training samples, x_i is the ith training sample, and x_i' is the reconstructed version of x_i.

B. Attention Mechanism

Given a vectorized version $\bar{x}$ of the input image x, we follow the multi-headed dot-product attention mechanism proposed in [29]. The attention output for a single head is formulated as

$$a = \text{softmax}\left(\frac{(\bar{x}W_q)(\bar{x}W_k)^T}{\sqrt{d_k}}\right)(\bar{x}W_v), \quad (12)$$

where W_q, W_k, and W_v are the respective linear transformations that map $\bar{x}$ to queries $\bar{x}W_q$, keys $\bar{x}W_k$, and values $\bar{x}W_v$, and d_k is the depth of the query/key in each head. The output of the multi-headed attention is obtained by a linear transformation W_o of the concatenated outputs of N_h heads as

$$H = \text{concat}[a^1,...,a^{N_h}]W_o. \quad (13)$$

Reshaping H such that its spatial dimensions match the dimensions of x yields the output of the multi-headed attention layer.

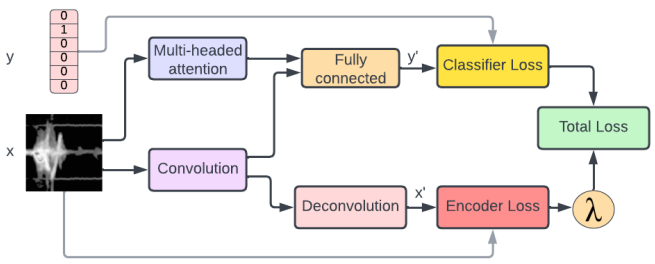

Fig. 1: Framework for joint semi-supervised training of a single-layer AA-CAE based classifier.

C. Classifier

For an AA-CAE based classifier, we consider a single fully-connected layer whose input is the concatenation of the convolutional and attention maps extracted from the input image. The output of the classifier is calculated as

$$y' = \text{concat}[H, O]W_c, \tag{14}$$

where y' is the prediction of the one-hot encoded label y and W_c denotes the weights of the fully connected layer.

An AA-CAE classifier is typically trained via a two-step approach, with the assumption that the labels of the entire training data are known. Specifically, the CAE undergoes unsupervised pre-training using all data. After the pre-training, the decoder is replaced with a fully-connected layer and the final configuration of the AA-CAE classifier is learned in a fully supervised fashion using the entire data with labels [22]. In practice, however, labels may be available for some but not necessarily all samples in the training data. In such cases, semi-supervised training can follow the same two-step process with the exception that the training of attention mechanism-plus-classifier is achieved using labeled samples only [25].

D. Proposed Joint Semi-supervised Training

The aforementioned method trains the encoder for representation learning without taking into account the classification task. Inspired by [27], we propose a joint semi-supervised training method for learning the encoder and attention weights and the classifier of an AA-CAE model at the same time. This enables the local feature map extractions to be guided by the ultimate classification objective. The main constituent blocks of the proposed approach are shown in Fig. 1.

For the considered AA-CAE based classifier, the joint training can proceed by minimizing the combined loss function, defined as

$$\sum_{y_j \in X_l} \|y_j - y'_j\|_F^2 + \lambda \sum_{x_i \in X} \|x_i - x'_i\|_F^2, \tag{15}$$

where λ is a scalar, $X = [X_u|X_l]$ with X_u and X_l denoting the sets of unlabeled and labeled training samples, respectively, and y_j is the one-hot encoded label of the jth input image in X_l.. However, backpropagation techniques cannot be used to solve this problem since there is no unambiguous way to backpropagate these two errors. To resolve this issue, we use the Split Bregman optimization technique, [27], [30], as detailed below.

We define the proxy variable Z, which captures the latent space representations of all training samples. As such, its component z_i corresponds to the ith image such that $z_i = O_i(W_e)$, with W_e being the weights of the P filters of the encoder. Substituting x'_i and y'_j from (9) and (13) into (15), the corresponding relaxed optimization problem is given by

$$\begin{aligned} \min_{W_e, W_d, W_a, W_c, Z} & \sum_{j \in |X_l|} \|y_j - \text{concat}[H_j(W_a), z_j]W_c\|_F^2 \\ & + \lambda \sum_{i \in |X|} \|x_i - \sigma(z_i * W_d)\|_F^2 \\ & + \mu_e \sum_{i \in |X|} \|z_i - o_i(W_e) - \beta\|_F^2, \end{aligned} \tag{16}$$

where W_a are the composite weights for the attention layer (i.e., W_q, W_k, W_v, and W_o), β is the Bregman relaxation variable, μ_e is a constant, and $|\cdot|$ denotes the cardinality of its argument. The proxy variable Z equals $[Z_u|Z_l]$, with Z_u and Z_l denoting its unlabeled and labeled components.

Now, the optimization problem can be segregated into a number of sub-problems:

$$\text{P1:} \min_{W_a, W_c} \sum_{j \in |X_l|} \|y_j - \text{concat}[H_j(W_a), z_j]W_c\|_F^2$$

$$\text{P2:} \min_{W_d} \sum_{i \in |X|} \|x_i - \sigma(z_i * W_d)\|_F^2$$

$$\text{P3:} \min_{W_e} \sum_{i \in |X|} \|z_i - O_i(W_e) - \beta\|_F^2$$

$$\begin{aligned} \text{P4:} \min_{Z_u} & \sum_{\iota \in |X_u|} [\lambda \|x_\iota - \sigma(z_\iota * W_d)\|_F^2 \\ & + \mu_e \|z_\iota - O_\iota(W_e) - \beta\|_F^2] \end{aligned}$$

$$\begin{aligned} \text{P5:} \min_{Z_l} & \sum_{j \in |X_l|} \|y_j - \text{concat}[H_j(W_a), z_j]W_c\|_F^2 \\ & + [\lambda \|x_j - \sigma(z_j * W_d)\|_F^2 \\ & + \mu_e \|z_j - O_j(W_e) - \beta\|_F^2]. \end{aligned}$$

Using these problems, the training can proceed by calculating the error for each layer and updating via backpropagation. After each epoch, β is updated as

$$\beta \Leftarrow \sum_{i \in |X|} z_i - O_i(W_e) - \beta. \tag{17}$$

IV. Experimental Results

To evaluate the performance of the proposed jointly trained semi-supervised AA-CAE classifier, we use micro-Doppler signatures corresponding to data collections of human activities at the University of Glasgow, UK [31]. The dataset comprises measurements using an FMCW radar operating at 5.8 GHz, with a 400 MHz bandwidth and a 1 ms chirp

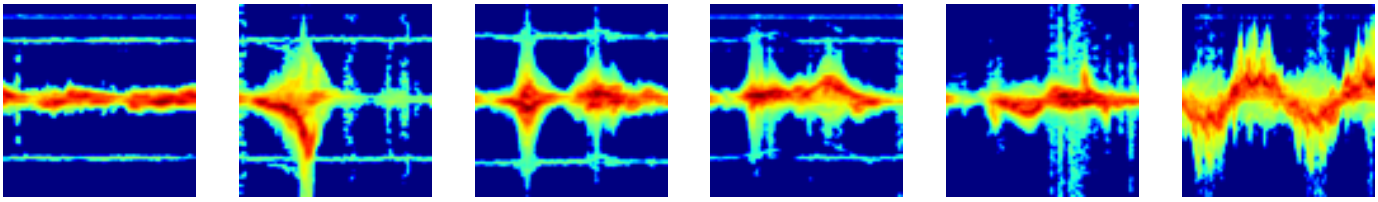

Fig. 2: Micro-Doppler signatures of six human activities: (From left to right) drinking water, falling, bending to pick up, sitting down, standing up, and walking.

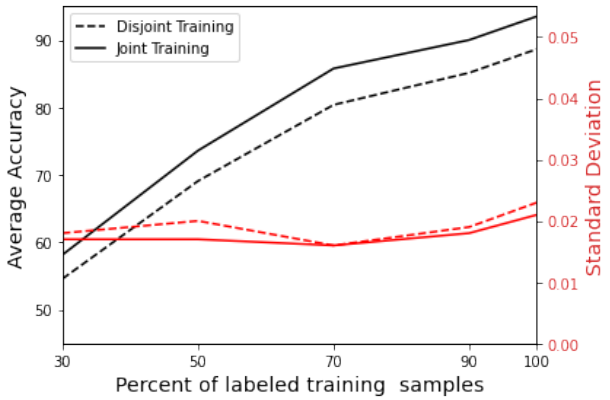

Fig. 3: Average and standard deviation of classification accuracy of AA-CAE models trained with joint and disjoint (two-step) methods.

TABLE I: Classification accuracy with 70% of training samples considered to be labeled.

Classification Accuracy	Training/testing Split	Training method	
		Disjoint	Joint
Average	20/80	71.13	76.43
	50/50	77.43	81.78
	80/20	80.39	85.78
Standard Deviation	20/80	0.016	0.016
	50/50	0.017	0.015
	80/20	0.016	0.016

duration. The number of samples per recorded beat-note signal was set to be 128. There are six activity classes in the data: walking, sitting down, standing up, picking up an object from the floor, drinking water, and falling. The considered dataset corresponds to a total of 33 participants, 31 of them are male and two are female, ranging in height from 149 cm to 188 cm and aged between 22 to 36 years. Each participant repeated each activity two to three times. We utilized a Hanning window length of 256 with 2048 frequency points and 254 points overlap to compute the spectrograms. After cropping and downsampling, each micro-Doppler signature has dimensions of 76×76, with the pixel values ranging from 0 to 255 in grayscale. There are 570 micro-Doppler signatures in total, with 95 signatures per class. Fig. 2 shows representative signatures, one for each of the six considered activities.

The single-layer AA-CAE architecture used for performance evaluation comprised a multi-headed attention block with $N_h = 2$ and $d_k = 4$; a convolutional layer with eight filters and a kernel size of 7×7, followed by a max-pooling layer with a stride of 3; a convolution transpose layer; and a fully-connected layer with a one-hot encoded output. A 10% dropout was used after the convolutional layer. All layers used the ReLU activation function, except for the classifier which used the softmax function. To optimize the model, we utilized stochastic gradient descent with a batch size of 10. We employed adaptive learning rate during training, for a maximum of 80 epochs. Moreover, we set λ to 0.2, μ_e to 0.2, and the initial value of β to 0.05.

We use 80% of the micro-Doppler signatures for training the model and hold out the remaining 20% for testing. From the training data, we varied the percentage of labeled training samples from 30% to 90%, with an increment of 20%. The remaining data samples in each case are assumed to be unlabeled. We also train the classifier with 100% labeled training data as a benchmark case. We compare the performance of two separate models, one trained using the proposed joint semi-supervised method and the other via disjoint (two-step) semi-supervised training. For the latter, unsupervised pre-training was conducted for 30 epochs with all training samples, and fine-tuning with labeled samples for 60 epochs. Using 30 experiments, we computed the mean and the standard deviation of the classification accuracy using 30 repetitions of the experiment for each percentage of labeled samples. The results corresponding to both models are depicted in Fig. 3. We see that the learned model with the proposed training method outperformed its counterpart trained via the two-step process in terms of average accuracy for all labeled data percentages. Moreover, the jointly trained model achieved the same accuracy as the model trained in a disjoint manner, but using a significantly lower percentage of labeled samples for training. Specifically, the proposed model achieved the same average accuracy as the fully supervised benchmark case of the disjointly trained model, with roughly 20% less labeled data. Additionally, the standard deviation corresponding to the jointly trained model were mostly lower than those of the disjointly trained model, thereby indicating greater stability in the proposed training process.

Next, we compare the performance of the two models under different training and testing samples splits, namely, 20/80,

50/50, and 80/20, with the first value denoting the percentage of training samples and the second value representing the testing data percentage. For all three splits, we maintained the percentage of labeled training samples at 70%, with the remaining 30% as unlabeled samples. The mean and standard deviation of the accuracy of 30 trained models under each training category are provided in Table I. We can see that joint training improves the accuracy and reduces the standard deviation across the board, as compared to its counterpart trained via the two-step method.

V. Conclusion

In this paper, we presented a jointly semi-supervised attention-augmented CAE-based classifier for human activity recognition to support radar-based remote monitoring for human wellness. We trained the encoder, attention, and classifier parts of the network simultaneously with both labeled and unlabeled data samples, using a joint loss function comprising both reconstruction and prediction errors. We demonstrated the superiority of the proposed approach in terms of classification accuracy over the two-step disjoint training method, utilizing real radar measurements of six different human activities. Further, we showed that the joint training improved the standard deviation of the trained model's classification accuracy for the considered activities. Overall, joint semi-supervised AA-CAE classifier provided performance enhancements for remote activity monitoring applications of radar.

References

[1] M. J. Rodrigues, O. Postolache, and F. Cercas, "Physiological and behavior monitoring systems for smart healthcare environments: A review." *Sensors*, vol. 20, 2020, Article ID: 2186.

[2] J. Bajo, J. F. de Paz, Y. de Paz, and J. M. Corchado, "Integrating case-based planning and RPTW neural networks to construct an intelligent environment for health care," *Expert Systems with Applications*, vol. 36, pp. 5844–5858, 2009.

[3] Q. Hao, F. Hu, and Y. Xiao, "Multiple human tracking and identification with wireless distributed pyroelectric sensor systems," *IEEE Systems Journal*, vol. 3, no. 4, pp. 428–439, 2009.

[4] F. Ahmad, A. E. Cetin, K. C. Ho, and J. Nelson, "Signal processing for assisted living: Developments and open problems," pp. 25–26, mar 2016.

[5] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 859–869, 2017.

[6] S. Iyer et al., "mm-wave radar-based vital signs monitoring and arrhythmia detection using machine learning," *Sensors*, vol. 22, no. 9, pp. 1–20, 2022, Art. ID 3106.

[7] J. L. Nielsen, D. Bakula, and M. Scheibye-Knudsen, "Clinical trials targeting aging," *Frontiers in Aging*, vol. 3, 2022.

[8] A. M. Iancu, M. T. Kemp, and H. B. Alam, "Unmuting medical students' education: Utilizing telemedicine during the covid-19 pandemic and beyond," *Journal of Medical Internet Research*, vol. 22, no. 7, 2020, Article ID: e19667.

[9] M. Amin (Ed.), *Radar for Indoor Monitoring: Detection, Classification, and Assessment*. Boca Raton, FL: CRC Press, 2018.

[10] B. Erol, M. G. Amin, and S. Z. Gurbuz, "Automatic data-driven frequency-warped cepstral feature design for micro-Doppler classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1724–1738, August 2018.

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

[12] P. P. Markopoulos and F. Ahmad, "Indoor human motion classification by L1-norm subspaces of micro-Doppler signatures," in *Proc. IEEE Radar Conference*, June 2017, pp. 1807–1810.

[13] J. Zabalza et al., "Robust PCA micro-doppler classification using SVM on embedded systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 2304–2312, July 2014.

[14] B. Jokanovic, M. Amin, F. Ahmad, and B. Boashash, "Radar fall detection using principal component analysis," in *Radar Sensor Technology XX*, K. I. Ranney and A. Doerry, Eds., vol. 9829. SPIE, May 2016, p. 982919.

[15] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, may 2009.

[16] P. P. Markopoulos, S. Zlotnikov, and F. Ahmad, "Adaptive radar-based human activity recognition with L1-norm linear discriminant analysis," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 3, no. 2, pp. 120–126, June 2019.

[17] T. D. Bufler and R. M. Narayanan, "Radar classification of indoor targets using support vector machines," *IET Radar, Sonar and Navigation*, vol. 10, no. 8, pp. 1468–1476, 2016.

[18] H. Sadreazami, M. Bolic, and S. Rajan, "Contactless fall detection using time-frequency analysis and convolutional neural networks," *IEEE Trans. Ind. Informatics*, vol. 17, no. 10, pp. 6842–6851, 2021. [Online]. Available: https://doi.org/10.1109/TII.2021.3049342

[19] M. Zhao et al, "Learning sleep stages from radio signals: A conditional adversarial architecture," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, 2017, pp. 4100–4109.

[20] F. J. Abdu, Y. Zhang, and Z. Deng, "Activity classification based on feature fusion of FMCW radar human motion micro-Doppler signatures," *IEEE Sensors Journal*, vol. 22, no. 9, pp. 8648–8662, 2022.

[21] R. C. Tewari, P. Palo, J. Maiti, and A. Routray, "GAN-based radar micro-Doppler augmentation for high accuracy fall detection system," in *Proceedings of the 48th Annual Conference of the IEEE Industrial Electronics Society*, 2022, pp. 1–6.

[22] C. Campbell and F. Ahmad, "Attention-augmented convolutional autoencoder for radar-based human activity recognition," in *Proceedings of the IEEE International Radar Conference*, 2020, pp. 990–995.

[23] S. Vishwakarma *et al.*, "Attention-enhanced Alexnet for improved radar micro-Doppler signature classification," *IET Radar, Sonar & Navigation*, 2022.

[24] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[25] C. Campbell and F. Ahmad, "Semi-supervised attention-augmented convolutional autoencoder for radar-based human activity recognition," in *Proc. SPIE Radar Sensor Technology XXVI Conference*, vol. 12108, 2022, p. 121080B.

[26] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

[27] A. Gogna, A. Majumdar, and R. Ward, "Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2196–2205, 2017.

[28] W. Chen and N. Griswold, "An efficient recursive time-varying Fourier transform by using a half-sine wave window," in *Proceedings of IEEE-SP International Symposium on Time- Frequency and Time-Scale Analysis*, 1994, pp. 284–286.

[29] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[30] R. Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5810–5819, 2012.

[31] F. Fioranelli *et al.*, "Radar sensing for healthcare," *Electronics Letters*, vol. 55, no. 19, pp. 1022–1024, 2019.