DCT-Based Feature Extraction For Human Activity Recognition using WiFi Channel State Information Data

Ishtiaque Ahmed Showmik
Department of EEE, BUET
Dhaka, Bangladesh
ishtiaque.eee.buet@gmail.com

Tahsina Farah Sanam
Institute of Appropriate Technology, BUET
Dhaka, Bangladesh
tahsina@iat.buet.ac.bd

Hafiz Imtiaz
Department of EEE, BUET
Dhaka, Bangladesh
hafizimtiaz@eee.buet.ac.bd

***Abstract*—Human Activity Recognition (HAR) exploits WiFi signals to offer behavioral sensing for various practical applications, such as patient monitoring in hospitals, and children/elderly monitoring in smart homes. Particularly for medical applications, HAR is of considerable importance. HAR determines the type of activity precisely and effectively using the correlation between the Channel State Information (CSI) data and physical movements. However, various constraints often make the process challenging for usage in practical systems. Large feature dimensionality and variable activity signal duration are the two major difficulties for developing an efficient HAR framework. Additionally, the frequency components of the CSI activity signal are time-varying, which makes CSI data inherently non-stationary. In this work, we present a feature extraction approach using Discrete Cosine Transform (DCT) to address the aforementioned issues – we integrate Principal Component Analysis (PCA) based dimensionality reduction and a solution to the non-stationarity problem. We assess the performance of our proposed approach on real data using some off-the-shelf reference models, such as support vector machine (SVM), random forest (RF), k-nearest neighbor (KNN), and convolutional neural network (CNN). Empirical results emphasize the quality of our proposed feature extraction approach by showing that even off-the-shelf classification models perform very well in challenging scenarios.**

***Index Terms*—Human activity recognition (HAR), channel state information (CSI), principal component analysis, and discrete cosine transform.**

I. Introduction

Human Activity Recognition (HAR) has many modern applications, including smart home monitoring and surveillance, intruder detection, patient monitoring in hospitals, children or elderly monitoring in smart homes, and military and security applications. HAR offers insightful information on a person's behavior and physical health. For instance, following a certain fitness routine is often a requirement of treatment for diabetes, obesity, or heart disease. Evidently, it is helpful to identify activities like walking, running, and cycling to inform the caregiver about the patient's behavior. Similarly, helping individuals with cognitive impairments (e.g., patients with dementia, Parkinson's disease, or Alzheimer's disease) by monitoring for abnormal activities and thereby preventing undesirable consequences, is another potential use [1], [2].

There is a multitude of techniques (camera-based and wearable sensor-based) that are used to recognize human activity, and new research is continuously improving existing models [3], [4]. In recent years, using Channel State Information (CSI) from WiFi signals to identify human activity has expanded significantly by utilizing the signal penetration properties [5]. Modern WiFi systems employ Multiple Input Multiple Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM). The subcarriers of the OFDM system provide responses to human activity at various frequency levels. It has been demonstrated that any movement in an indoor environment has an impact on the CSI. As a result, differences in the CSI signals may be used to identify human activities. Additionally, CSI is useful for activity detection because of its unique ability to reduce multi-path effects. However, only a few devices, such as the Intel WiFi 5300 NICs, offer access to the underlying CSI data [6]. The development of CSI tools has made it possible to collect CSI data and investigate the relationship between signals and human activity [7].

The use of routine physical activity monitoring and recognition in the healthcare industry has been considered in several recent works. Such monitoring systems can assist in controlling and minimizing the risk of a number of health disorders; e.g., obesity and cardiovascular disease. Additional applications for activity monitoring systems include elderly rehabilitation, fall detection, and smoking cessation. The majority of activity detection techniques have relied on video, accelerometer-based data, and the integration of several types of sensors. However, both the camera-based method (which has the limitations of requiring line of sight and optimal illumination) and the wearable sensor method (which is more accurate and simple but inconvenient and expensive) have the issue of potentially breaching individual privacy. Wi-Fi-based HAR systems overcome the aforementioned difficulties by utilizing the correlations between CSI signal variations and body motions. In this work, we present an effective HAR system that i) can be employed in real-time applications, such as smart hospitals; ii) can outperform existing methods in terms of robustness and efficiency.

A. Related works

In recent years, researchers have proposed several techniques for CSI-based activity recognition using machine learning (ML) algorithms. For instance, an antenna selection method is proposed in [8]. a viable channel selective activity

recognition (CSAR) system was described in [9]. A simple deep neural network for HAR is proposed in [10]. To demonstrate the utility of the Internet-of-Things platform, an occupancy detection system based on the CSI curve of human presence is proposed in [11]. A deep recurrent neural network (RNN)-based HAR system (HARNN) is proposed in [12]. Another system, MCBAR, uses CSI measurements to monitor human activities and the generative adversarial network (GAN) [13]. WiKey and WiGest are two noteworthy keystroke and gesture recognition approaches, respectively, that use variations in WiFi signal strength [14], [15]. Finally, an end-to-end deep subdomain adaptive network for more fine-grained feature extraction is proposed in [16].

Apart from these, the work presented in [17] transformed the CSI data into images and used those as the inputs for a 2D Convolutional Neural Network (CNN) classifier. DeepSeg, a deep learning-based framework for activity segmentation that includes a CNN-based segmentation algorithm and a feedback mechanism is introduced in [18]. An effective HAR system that is suitable for practical applications is proposed in [19] called the Principal Component-based Wavelet Convolutional Neural Network (or PCWCNN). The segmentation algorithm is improved based on the feedback calculated from the outcomes of activity recognition. A Gaussian mixture model-hidden Markov model (GMM-HMM) is employed in [20] to describe the CSI feature data of each activity, and the CSI phase difference expansion matrix is produced as a more evident activity detection feature.

Despite these intriguing developments in CSI-based activity recognition algorithms, HAR system development still has room for improvement and optimization in terms of resource usage and computational complexity. In this work, we propose an implementation-oriented feature extraction technique that significantly improves and streamlines the performance of existing recognition systems. Our novel feature extraction approach for an efficient HAR framework employs two efficient and widely explored algorithms: Principal Component Analysis (PCA)-based filtering of activity information from the subcarriers, and Discrete Cosine Transform (DCT)-based feature extraction, which further reduces dimensionality while addressing non-stationarity constraints. We demonstrate that by virtue of a clever feature extraction approach, even off-the-shelf classification algorithms can perform superbly for CSI-based activity recognition.

II. Proposed System Architecture

The activity recognition system employing the proposed DCT-based feature extraction approach can be divided into four major steps, as shown in Figure 1. We discuss the steps in detail in the following.

A. Data Acquisition

The input data at the receiver is collected as a $T \times R \times S$-dimensional complex-valued tensor, where T is the number of transmitting antennae, R is the number of receiving antennae, and S is the number of OFDM subcarriers. Let $\mathbf{X_i} \in \mathbb{R}^{D\times 1}$ denote the i-th CSI stream (i.e., the amplitude data of the i-th subcarrier of the CSI), where D is the length of the CSI stream. We can arrange all these CSI streams to form the matrix $\mathbf{X} = [\mathbf{X_1},\ \mathbf{X_2},\ \cdots \mathbf{X_N}] \in \mathbb{R}^{D\times N}$ where N is the total number of subcarrier streams. For our case, $T = 1$, $R = 3$, and $S = 30$, and therefore, we have $N = 90$ CSI streams.

B. PCA-based Noise Reduction and Dimensionality Reduction

Raw CSI data contain noise and outliers. Therefore, using CSI data in activity detection algorithms requires the application of signal processing and denoising techniques to be employed on the raw CSI streams. An approach for reducing noise and solving the issue of integrating CSI streams is CARM [4], which applies PCA on CSI streams. In the following, we review the steps in brief: after extraction of CSI as the matrix $\mathbf{X} \in \mathbb{R}^{D\times N}$, we can find the correlation estimate from the sample auto-correlation matrix $\mathbf{R} = \mathbf{X}^T\mathbf{X} \in \mathbb{R}^{N\times N}$. The symmetric positive semi-definite matrix $\mathbf{R}$ can be decomposed as $\mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ [21], where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N] \in \mathbb{R}^{N\times N}$, is the unitary matrix containing the eigenvectors of $\mathbf{R}$, and $\mathbf{\Lambda} \in \mathbb{R}^{N\times N}$ is the diagonal matrix containing the eigenvalues λ_k of $\mathbf{R}$. Typically, the maximum eigenvalue is associated with the signal component with the largest variance, whereas smaller eigenvalues correspond to the noise subspace. Now, it is possible to project the data onto the subspace spanned by the eigenvectors to identify the components in their signal subspace since the eigenvectors of $\mathbf{R}$ are orthogonal to one another. The k-th principal component can be obtained as $\mathbf{h}_k = \mathbf{R}\mathbf{v}_k$ [4]. While the first principal component $\mathbf{h}_1$ contains the majority of data variability, it also catches the burst noise that is present in all CSI streams. Additionally, the first principal component shows the most presence of noise brought on by internal state fluctuations [4]. The first principal component and the ones corresponding to the smallest eigenvalues can be ignored to effectively filter out the majority of the noise. In this way, dimensionality reduction will also be accomplished. In this paper, we only used the second and third principal components ($\mathbf{h}_2$ and $\mathbf{h}_3$), discarding the others.

C. DCT-based Feature Extraction

For a sequence of length M, the DCT coefficients can be found from the following transformation [22],

$$F(u) = \alpha(u) \sum_{x=0}^{M-1} f(x) \cos\left[\frac{\pi(2x+1)u}{2M}\right],$$

for $u \in \{0, 1, 2, \cdots, M-1\}$. Here, $\alpha(u)$ is defined as,

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{M}}, & \text{if } u = 0 \\ \sqrt{\frac{2}{M}}, & \text{if } u \neq 0. \end{cases}$$

The first coefficient corresponds to the sequence's average value, whereas the following coefficients correspond to higher frequency components. We choose DCT since it provides packing the input data into "as few coefficients as possible", and is very efficient. As a result, small-valued and high-frequency coefficients can be removed without the reconstructed signal becoming qualitatively distorted. DCT offers excellent energy compression for highly correlated signals [22]. That is, the majority of the signal's information is compressed by DCT into a relatively small number of

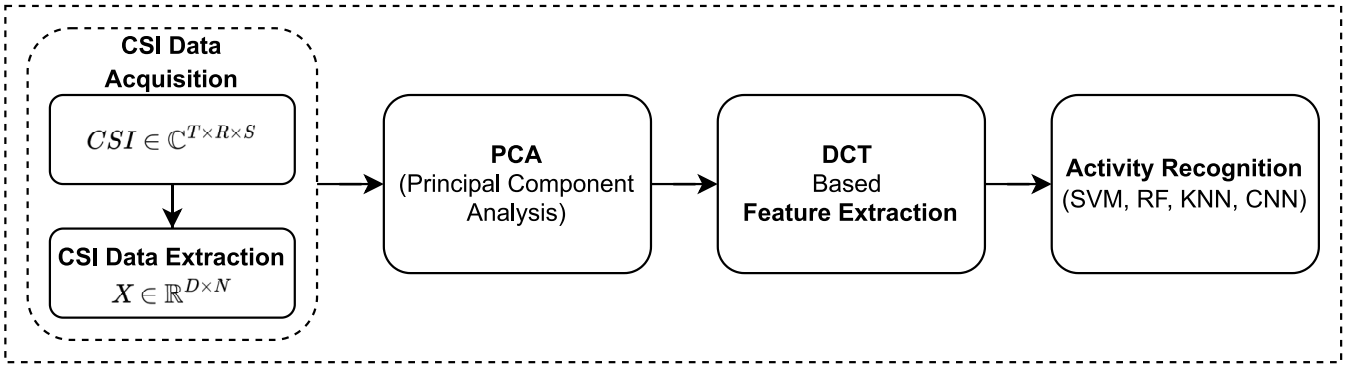


Fig. 1: Framework of the proposed DCT-based activity recognition system

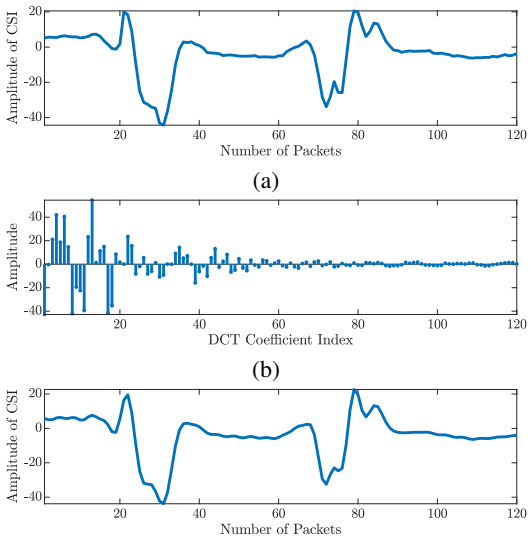


Fig. 2: (a) CSI data (b) DCT coefficients (c) Reconstructed signal

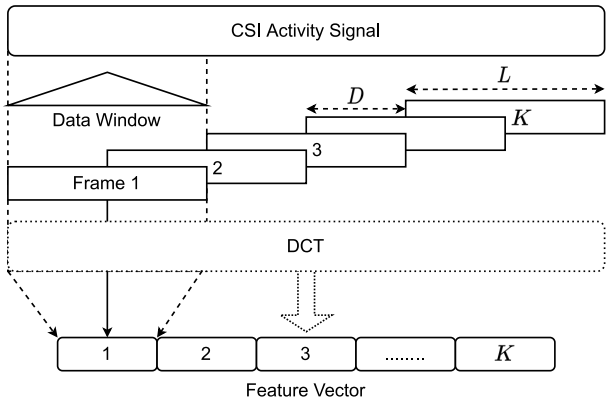


Fig. 3: DCT-based feature extraction using the sliding window method. An overlap of 50% is used. The DCT coefficients are concatenated to form the final feature vector.

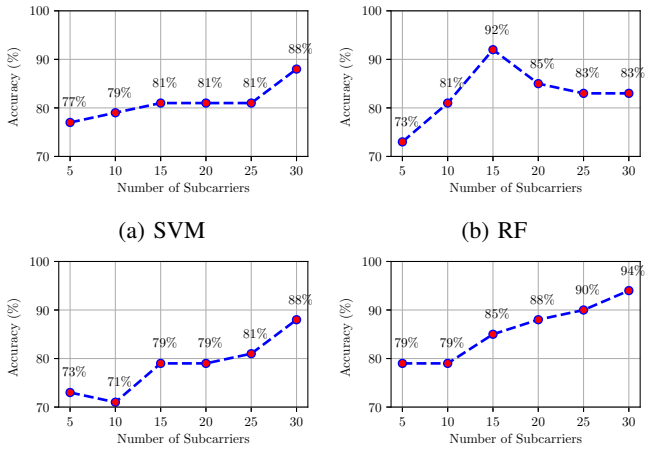
coefficients [23]. Additionally, DCT enables us to manage the fluctuating nature of the signal length of the CSI data. The removal of high-frequency components also helps the reduction of noise since activity features are most evident in the signal's low-frequency components.

The data compression and feature extraction process that we employ in this work is depicted in Figure 2. The signal with 120 samples is reconstructed leveraging only 40 DCT coefficients. The reconstructed signal resembles the original signal sufficiently well, although being reconstructed from a smaller number of coefficients. However, the frequency components of the CSI data are time-varying. Hence, we propose a modified DCT-based feature extraction approach in this work, which we summarize in the following. The activity signals $x(n)$ is divided into multiple shorter, equal-length sequences, which are assumed to be stationary. A sliding window of appropriate width is used to partition the signal into several frames of the same length. Additionally, succeeding signal frames overlap by a certain percentage in order to optimize the usage of information. To reduce spectral leakage, a data window, such as the Kaiser window, is used with the signal frames. The DCT coefficients of each frame are then computed and arranged to generate a feature vector corresponding to the CSI signal.

The i-th frame is obtained by assuming that consecutive sequences are offset by D points and that each sequence is of length L. That is, $x_i(n) = x(n + iD)$, for $n \in \{0, 1, \cdots L-1\}$. Let F_i represent the DCT coefficients of $x_i(n)$, and N_C represent the number of components retained after transformation. The feature vector from a signal of length M will have KN_C elements if the signal is segmented into K parts. The feature vector length KN_C can be shorter than M by selecting appropriate K and N_C values. In Figure 3, we illustrate the working principle of this method. Recall that we propose to use only the second and third principal components to form the feature vectors, and use that in the classification stage.

D. Activity Classifier

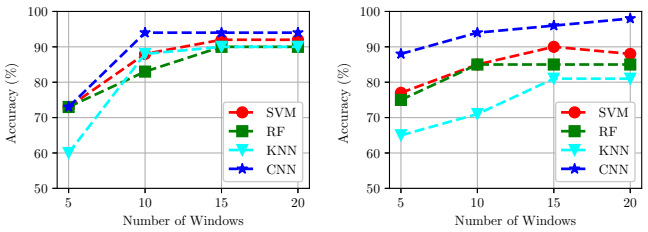
We emphasize that the primary contribution of this work is the proposed feature extraction method. To demonstrate the quality of the extracted features, we choose to use off-the-shelf classifiers for evaluating the activity classification performance. More specifically, the performance of the proposed feature extraction method was evaluated using the following state-of-the-art reference models: Support Vector Machine (SVM), Random Forest (RF), k-nearest neighbor (KNN), and 1-dimensional convolutional neural network (CNN). We used SVM with the Radial Basis Function (RBF) kernel and the parameter γ that controls the influence of individual points on the overall process. C, the regularization parameter, was set to 11, and the regularization was adjusted to be inversely



(a) SVM (b) RF

(c) KNN (d) CNN

Fig. 4: Effect of number of subcarriers on the performance of different approaches using DCT feature extraction across all the activities of an individual



(a) All subcarriers (b) PCA-based integration

Fig. 5: Effect of a number of windows on the performance of DCT feature extraction.

proportional to C. For RF, the maximum depth was set to 20 and the number of trees was set to 200. The $\mathcal{L}_2$-norm was utilized as the distance measure in the KNN method, and the value of k was set to 3. The 1D-CNN model accepts a 128-element vector as input and returns a vector with probabilities for each possible activity type. Each of these classifiers was a 16-class classifier. Before training the models, 20% of the data was reserved as the test set, and the rest was used for training.

III. Experimental Results

Dataset: In this paper, we use the publicly available WiAR data set for activity recognition [24]. WiAR consists of sixteen activities, which can be divided into three main categories depending on the body parts associated, as shown in Table I. The CSI data are acquired for three-receiver antennae and one transmitter antenna. The WiAR dataset includes data from three different locations and heights. The activities are performed by ten volunteers, and 30 samples were taken from each volunteer. We choose this dataset for demonstrating the effectiveness of our proposed feature extraction since it consists of a variety of activities performed by different volunteers at different distances and heights. We, therefore, believe this dataset provides good representative scenarios for real-world use cases and applications. Before training the models, 20% of the data was reserved as the test set, and the rest was used for training.

Effect of Number of Subcarriers: The impact of the number of subcarriers (S) on the performance of DCT feature extraction-based activity recognition is presented in Figure 4. We observe that the performance of the DCT feature extraction method improves as the number of subcarriers increases. The best overall result for DCT feature extraction is for $S = 30$. For SVM, $S = 15$ to 25 produces similar results. The best outcome for RF is when we choose $S = 15$; increasing the number of subcarriers above this value negatively affects performance for RF. For CNN, accuracy grows approximately linearly as S increases. The maximum accuracy is 94%, which is obtained using CNN.

Effect of Number of Windows: In Figure 5, we show how the number of windows (or window size) impacts the performance of DCT feature extraction. The performance of CNN is always better in comparison to that of others. The dominance of CNN for PCA-based integration is even more pronounced. When all subcarriers are being used, however, the difference in performance among the reference models is small. The difference is negligible for 20 windows. With increasing the number of windows, accuracy improves for both PCA and all subcarrier scenarios. The maximum accuracy is 98%, which is obtained for PCA-based integration with CNN. As the number of windows increases, finer details become visible, and thereby accuracy increases.

Effect of Distance: In Figures 6(a) and (b), we show the effect of three distances: 1 m, 3 m, and 6 m on the performance of activity recognition. For CNN, the results are the most consistent. Nevertheless, accuracy increases with distance for all subcarriers. We observe that the accuracy while using all subcarriers and the PCA-based integration is the highest for 6 m distance. This is contrary to the intuition that the accuracy of activity recognition declines with increasing distance [25]. We believe this is dataset-specific.

Effect of Height: Finally, the WiAR dataset contains a range of different height data, since different height data connect to diverse body parts. Lower body activities correspond to a height of 60 cm, the whole body to a height of 90 cm, and upper body activities to a height of 120 cm. In Figures 6(c) and (d), we illustrate the accuracy of the reference models for three distinct heights. The accuracy of CNN for all subcarriers is unexpectedly low, which highlights the excess noise at various subcarrier levels for height data. Due to the poor quality of the activity data, the KNN, SVM, and RF performances are below average. In the majority of the cases, accuracy for PCA-based integration is at its highest at 90 cm, indicating a stronger correlation between all body part movements and the WiFi signal.

TABLE I: Activity Types

Types	Activities
Upper Body Activities	Horizontal arm wave, Two hand wave, Toss paper, Draw kick, Phone, Draw X, Handclap, High arm wave, Drink water, High throw
Lower Body Activities	Forward kick, Side kick
Whole Body Activities	Squat, Sit down, Bend, Walk

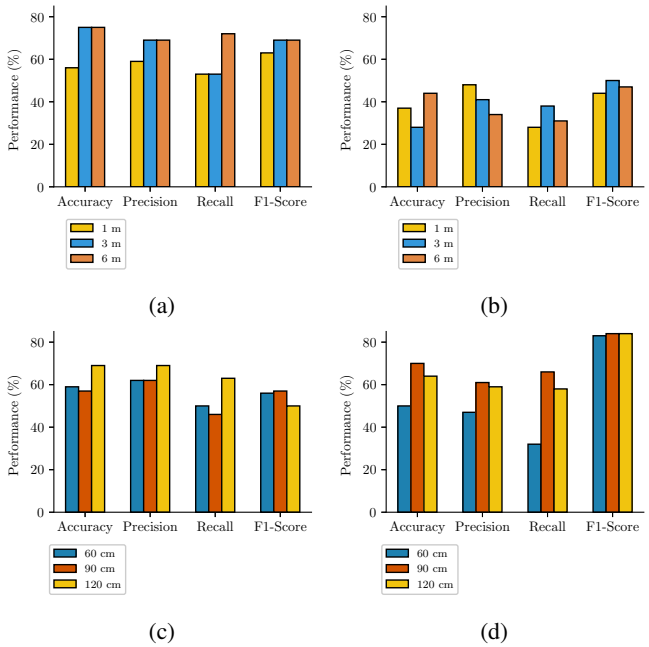


Fig. 6: Effect of distance and height on the performance of different approaches using DCT feature extraction (a) All subcarriers (Distance) (b) PCA-based integration (Distance) (c) All subcarriers (Height) (d) PCA-based integration (Height)

IV. CONCLUSION

In this work, we proposed an efficient feature extraction approach that is suitable for real-time HAR systems by utilizing commodity WiFi devices. Such systems are useful for noninvasive and device-free HAR applications, e.g. smart hospitals and elderly monitoring, and are advantageous compared to other forms of HAR systems because of the rapid growth of ubiquitous Wi-Fi technologies. We proposed a novel DCT-based feature extraction technique that handles the non-stationary nature of CSI data, and a PCA-based technique to reduce noise and dimensionality. A number of off-the-shelf classification algorithms were employed to thoroughly investigate the quality of the proposed feature extraction approach on real data. When the PCA-based integration is compared to the approach utilizing all subcarriers, it is found that the PCA-based method is superior, since it yields improved prediction accuracy and dimensionality reduction with quicker run-time. Indeed, the quality of the extracted features is highlighted by the fact that even off-the-shelf classifiers can perform very well by virtue of feature extraction.

REFERENCES

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.

[2] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE transactions on knowledge and data engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[3] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video technology*, vol. 18, no. 11, pp. 1473–1488, 2008.

[4] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015, pp. 65–76.

[5] F. Adib and D. Katabi, "See through walls with wifi!" in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, 2013, pp. 75–86.

[6] Y. Gu, F. Ren, and J. Li, "Paws: Passive human activity recognition based on wifi ambient signals," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 796–805, 2015.

[7] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM computer communication review*, vol. 41, no. 1, pp. 53–53, 2011.

[8] J. Yang, Y. Liu, Z. Liu, Y. Wu, T. Li, and Y. Yang, "A framework for human activity recognition based on wifi csi signal enhancement," *International Journal of Antennas and Propagation*, vol. 2021, 2021.

[9] F. Wang, W. Gong, J. Liu, and K. Wu, "Channel selective activity recognition with wifi: A deep learning approach exploring wideband information," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 181–192, 2018.

[10] C.-H. Hsieh, J.-Y. Chen, C.-M. Kuo, and P. Wang, "End-to-end deep learning-based human activity recognition using channel state information," *Journal of Internet Technology*, vol. 22, no. 2, pp. 271–281, 2021.

[11] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using wifi-enabled iot devices for smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, 2018.

[12] J. Ding and Y. Wang, "Wifi csi-based human activity recognition using deep recurrent neural network," *IEEE Access*, vol. 7, pp. 174 257–174 269, 2019.

[13] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Multimodal csi-based human activity recognition using gans," *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 345–17 355, 2021.

[14] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using wifi signals," in *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015, pp. 90–102.

[15] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015, pp. 1472–1480.

[16] L. Li, L. Wang, B. Han, X. Lu, Z. Zhou, and B. Lu, "Subdomain adaptive learning network for cross-domain human activities recognition using wifi with csi," in *2021 IEEE 27th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2021, pp. 1–7.

[17] P. F. Moshiri, R. Shahbazian, M. Nabati, and S. A. Ghorashi, "A csi-based human activity recognition using deep learning," *Sensors*, vol. 21, no. 21, p. 7225, 2021.

[18] C. Xiao, Y. Lei, Y. Ma, F. Zhou, and Z. Qin, "Deepseg: Deep-learning-based activity segmentation framework for activity recognition using wifi," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5669–5681, 2020.

[19] I. A. Showmik, T. F. Sanam, and H. Imtiaz, "Human activity recognition from wi-fi csi data using principal component-based wavelet cnn," *Digital Signal Processing*, vol. 138, p. 104056, 2023.

[20] X. Cheng and B. Huang, "Csi-based human continuous activity recognition using gmm–hmm," *IEEE Sensors Journal*, vol. 22, no. 19, pp. 18 709–18 717, 2022.

[21] H. Abdi, "The eigen-decomposition: Eigenvalues and eigenvectors," *Encyclopedia of measurement and statistics*, pp. 304–308, 2007.

[22] M. B. Khan, Z. Zhang, L. Li, W. Zhao, M. A. M. A. Hababi, X. Yang, and Q. H. Abbasi, "A systematic review of non-contact sensing for developing a platform to contain covid-19," *Micromachines*, vol. 11, no. 10, p. 912, 2020.

[23] S. Dabbaghchian, M. P. Ghaemmaghami, and A. Aghagolzadeh, "Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology," *Pattern recognition*, vol. 43, no. 4, pp. 1431–1440, 2010.

[24] L. Guo, L. Wang, C. Lin, J. Liu, B. Lu, J. Fang, Z. Liu, Z. Shan, J. Yang, and S. Guo, "Wiar: A public dataset for wifi-based activity recognition," *IEEE Access*, vol. 7, pp. 154 935–154 945, 2019.

[25] L. Guo, H. Zhang, C. Wang, W. Guo, G. Diao, B. Lu, C. Lin, and L. Wang, "Towards csi-based diversity activity recognition via lstm-cnn encoder-decoder neural network," *Neurocomputing*, vol. 444, pp. 260–273, 2021.