2022 Iraqi International Conference on Communication & Information Technologies (IICCIT-2022) , Basrah University , Basrah , Iraq

Human Activity and Gesture Recognition based on WiFi

Sokienah K. Jawad
Department of Computer Engineering
University of Basrah,
Basrah, Iraq
engpg.sokienah.jawad@uobasrah.edu.iq

Musaab Alaziz
Department of Computer Engineering
University of Basrah,
Basrah, Iraq
mosab.adil@uobasrah.edu.iq

2022 Iraqi International Conference on Communication and Information Technologies (IICCIT) | 978-1-6654-7220-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/IICCIT55816.2022.10010433

***Abstract*— Due to the wide availability of WiFi devices in almost every modern building, activity and gesture recognition systems using WiFi have recently been in the spotlight. Given the low cost and touch-free working of WiFi, this sensing approach has found success in various fields, including elder care, child safety, and smart homes that enable householders to monitor everything. This paper aims to investigate the human activity and gesture recognition scheme that uses the Channel State Information (CSI) value provided by WiFi devices. To achieve high accuracy in the measurement, firstly, a median filter is used to filter out the noise from Channel State Information extracted from the WiFi signal. Secondly, massive features are extracted to present intrinsic characteristics for each gesture and activity. Finally, Random Forest (RF) and Support Vector Machine (SVM) algorithms with cross-validation techniques are used for data classification. The system is implemented on a public dataset (WiAR) involving ten volunteers, each of whom performs 16 activities in each experiment. Experimental results show an average recognition accuracy of 92% and 91% for Random Forest and Support Vector Machine, respectively, which outperforms the existing system by 3.14 percent for RF and 1.14 percent for SVM.**

Keywords*— *Human gesture recognition, Channel State Information (CSI), Random Forest, Support Vector Machine, Classification

I. Introduction

Human gesture and activity recognition systems have earned much interest as science and technology have progressed. This field was widely employed in various fields, including tracking and detection [1],[2],[3], and health care monitoring [4][5]. A new type of human activity and gesture recognition system that exploit WiFi signals to identify human gestures has recently been developed [6]–[15]. The fundamental principle that enables these systems to exploit WiFi signals is that when a person moves in a wireless environment, wireless channel measurements vary, such as Channel State Information (CSI) and Received Signal Strength Indicator (RSSI). The CSI depicts how a signal travels via a wireless channel, including time delay, energy attenuation, and phase shift [16]. Various movements enable these wireless channel metrics to vary in different ways. WiFi-based gesture recognition systems first understand these variations for each preset gesture using machine learning techniques and then discover these motions when the user executes the activity. CSI represents fine-grained information of WiFi signals with thirty subcarriers; each of which includes both the amplitude and phase as shown in "(1)" [16]:

$$H(k) = \|H(k)\|\, e^{\,j\angle H(k)} \qquad (1)$$

where $H(k)$ denotes a CSI of the kth subcarrier. CSI amplitude and CSI phase are represented by $//H(k)//$ and $\angle H(k)$, respectively. When compared to the RSSI, CSI is more resistant to environmental changes and is measured at the orthogonal frequency division multiplexing (OFDM) subcarrier level from the received packet. In complicated situations, the RSSI value drops with increasing distance and is vulnerable to multipath fading [17]. Because of these promising properties of CSI, it was employed in this paper to build a more stable and robust system

In this research, a system that can detect both gesture and activity for 16 specified human movements is presented. The following are the work's main contributions:

- Human activity and gesture recognition system were proposed based on WiFi, which addresses the performance degradation caused by environmental dependence.
- For denoising, a median filter was used, which would improve system efficiency.
- To increase the accuracy of recognition, massive features are extracted in a time domain and frequency domain (Amplitude, Mean, Maximum value, Standard Deviation, Variance, Root Sum square (RSS), Percentile, Skewness, Kurtosis, Median, Continuous Wavelet Transform (CWT)). Then these features are applied to Random Forest and Support Vector Machine classifiers with the 10th cross-validation technique.

II. RELATED WORK

Several related existing works were mentioned in this section.

Pu et al. [12] employed USRP (Universal Software Radio Peripheral) to extract the Doppler Shifts from the arrived WiFi signals as features and then used them to recognize nine actions performed by a human. Kellogg et al. [11] utilized a specific circuit to separate the amplitudes of signals received and then matched the signal's properties to the gestures. Wang et al. [18] collected WiFi signals and then extracted CSI data using a laptop with a 5300 NIC (Network Interface Card). Principal Component Analysis (PCA) was performed for feature extraction. The CSI-speed and CSI-activity models were designed based on the CSI value and employed to define three essential human activities. Wang et al. [10] proposed identification for human activities at home by using WiFi signals. Abdelnasser et al. [6] built an RSSI-based gesture recognition system that was less sensitive and less responsive compared to CSI. In the initial step, the primitive gesture sketches were introduced, which included waving the hand in various directions and at varying speeds. Virmani et al. [19] used the transfer learning principle to minimize training effort. The system decomposed the gestures into two parts (linear and non-linear), and then these models were converted as attributes. Li et al. [20] built a system that recognized 9 digits of finger-grained gestures in American sign language. Ali et al. [21] tracked the keystrokes when the user executed a movement with their fingers in the air. Al-Qaness and Li [22] presented a gesture recognition system throughout the home with 7 hand movements as interactive gestures, 3 hand movements for security schemes, and the others for device selection gestures. Wavelet analysis and short-term energy were used for the segmentation method, and a dynamic time wrapping algorithm was implemented for classification. Jiang et al. [23] employed a Generative Adversal Network for features creation and extraction. Then, the SVM classifier was used in the classification stage. This system's implementation was handy for keeping time as well as coping with small samples. Zhang et al. [24] used USRP with WiFi for gesture recognition. The timestamps included in CSI were used for partitioning continuously received packets. PCA is used for denoising and Random Forest with two stages is used for classification.

III. METHODOLOGY

The proposed method contains three stages: data pre-processing, feature extraction, and classification plus accuracy calculation. The workflow of the proposed method is depicted in Fig. 1, and the following sections go over each stage in detail.

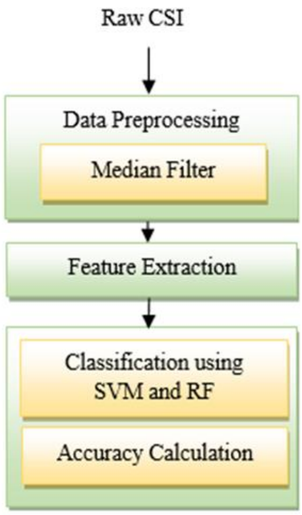

Fig. 1. Workflow of the proposed method.

A. *Data Preprocessing*

The first issue in data pre-processing is to minimize the impact of noise in the CSI data. In this paper, a median filter was applied for this purpose. The median filter is a non-linear digital noise-reduction approach for images and signals. This filter works by allowing a window to traverse across the points in a sequence and changes the value at the window midpoint with the median of the original values within the window. This frequently results in a smoother output sequence than the original. A linear low-pass filter is the traditional smoothing method, and, in many circumstances, it is the most appropriate method. However, in some cases, median filtering can be highly effective at smoothing spiky noise. This form of noise reduction is a typical pre-processing technique used to improve the results of later processing [25]. To improve accuracy, the window was empirically set to 5. Fig. 2 shows the signal before and after the median filter.

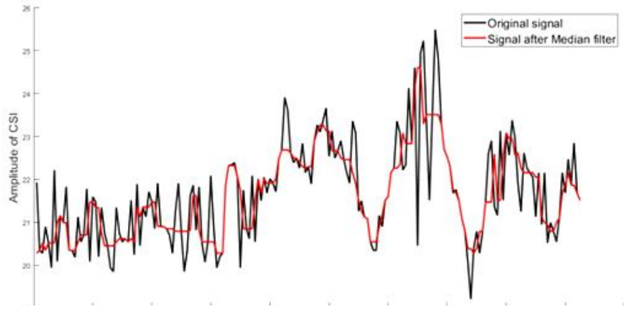

Fig. 2. The original signal before and after the median filter.

B. *Feature Extraction*

The proposed method extracts a total of 20 features from Channel State Information (Amplitude extraction using Weighted Moving Average (WMA), Mean, Maximum value, Standard Deviation, Variance, Root Sum square (RSS), Percentile, Skewness, Kurtosis, Median, Continuous Wavelet Transform (CWT) with 10 features). The meaning of each feature is as follows :

1) Amplitude Feature Extraction: Weighted Moving Average method was used to extract and smooth CSI Data [9][25]. The amplitude sequence of the first subcarrier at time t is denoted by $\{CSI_{1,1},..., CSI_{t,1}\}$. According to the WMA algorithm, the most recent CSI is assigned the highest weight m. The formula of CSI after using WMA is shown as follows:

$$CSI_{t,1} = \frac{1}{m+(m-1)+\cdots+1}\times(m\times CSI_{t,1} + (m-1)\times CSI_{t-1,1}+\cdots+1\times CSI_{t-m-1,1}) \quad (2)$$

In this paper, the weight was set to 10.

2) Mean: The mean can be computed as follows [26] :

$$\mu = \frac{1}{|x|}\sum_{i=1}^{|x|} x_i \quad (3)$$

since the amplitude of CSI is represented by a time series sequence $x=(x_1,x_2,..., x_n)$.

3) Standard Deviation: It determines the degree of variability, or dispersion, between individual data values and the mean as described in “(4)” :

$$\delta = \sqrt{\frac{\sum_{i=1}^{|x|}(x_i-\mu)^2}{|x|}} \quad (4)$$

4) Maximum values: The maximum value of x.
5) Variance: Variance is a measure of the spread between numbers in a data collection, and it is represented by δ^2.
6) Root Sum square (RSS):

$$RSS = \sqrt{\sum_{i=1}^{|x|} x^2} \quad (5)$$

7) Median: The middle value of x.
8) Skewness: This method is used to measure symmetry or asymmetry of data distribution as defined in the following expression [27] :

$$\mathrm{Skew}(x) = \frac{\sum_{i=1}^{N} (x-\mu)^3}{(N-1).\delta^3} \quad (6)$$

where N represents the number of data points, μ is the mean, and δ is the standard deviation.

9) Kurtosis: It is the degree of presence of outliers in the distribution. It can be expressed as in "(7)" [27]:

$$kurtosis(x) = \frac{\sum_{i=1}^{N} (x-\mu)^4}{(N-1).\delta^4} \quad (7)$$

where N represents the number of data points, μ is the mean, and δ is the standard deviation.
10) K-th Percentile: It refers to the value below which a percentage k of data falls. In this experiment, k =50 was used.
11) Continuous Wavelet Transform: Signal processing methods are frequently used to increase a model's prediction accuracy. A continuous wavelet transform is a powerful approach for this application. It depicts time frequency or, more precisely, time scale. Given the input signal x(t), the CWT is defined as follows [28]:

$$CWT_x(a,\tau) = \int_{-\infty}^{\infty} x(t)\psi_{a,\tau}^{*}(t)dt \quad (8)$$

Where $a \in \Re^+$ indicates the scale parameter, $\tau \in \Re$ means the translation diameter of time-shifting and the basis function $\psi_{a,\tau}^{*}$ is generated by scaling the mother wavelet $\psi(t)$ at time τ and scale a. The asterisk denotes that the complex conjugate of the wavelet function is employed in the transform. The mathematical formula of a wavelet family, which is made up of members or daughter wavelets $\psi_{a,\tau}$ is determined by the scaling and time-shifting of the mother wavelet $\psi(t)$ described as follows [29]:

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-\tau}{a}) \quad (9)$$

When a it grows large enough, the basis function becomes a stretched version of the prototype, emphasizing the low-frequency components. The basis function is contracted when a it is small, which emphasizes the high-frequency components. The shape of the basic function, on the other hand, will never change.
This stage extracts 10 features, giving a total of 20 features to feed into the classifiers.

C. Classification and Accuracy Calculation:

At this stage, a WiFi-based activity recognition dataset called WiAR [16] was employed with two classifiers (Random Forest and Support Vector Machine). This dataset contains 16 classes of different gestures and activities (horizontal arm wave, high arm wave, two hands wave, high throw, draw x, draw tick, toss paper, forward kick, side kick, bend, hand clap, walk, phone call, drink water, sit down, squat) performed by ten volunteers. The 802.11n CSI tool was used to acquire data for this dataset. Following the classification, the proposed system is evaluated using a variety of performance indicators.

1) Classification using Support Vector Machine (SVM):

SVM is a type of supervised machine learning approach that can be used to solve classification and regression problems [30]. Its goal is to find the best potential border between the various outputs. It does not support multiclass classification natively; however, it does enable binary classification and splitting data points into two groups. The same strategy is used for multiclass classification after breaking down the multiclassification problem into many binary classification problems. The goal is to map data points to a high-dimensional space to achieve mutual linear separation between classes. SVM needs the features extracted from the data as input to be able to classify. This is done by dividing the data into two parts: training and testing set. Each instance in the training sets includes numerous attributes and a single target value called a label. The fundamental goal of this classifier is to create a model for predicting the label of test data based on its attributes [31]. A kernel function is used to compute the separation of data points. Based on the results, Radial Basis Function (RBF) kernel has provided the best performance. The hyper-parameters (gamma γ and cost parameters) were chosen using 10-fold cross-validation.

2) Classification using Random Forest(RF):

The random forest method has been proven to be a state-of-the-art learning model, which not only has a good classification, regression, performance, and fast and efficient operations but also can effectively handle multiple classification problems. The random forest also has an obvious advantage in dealing with noise. This approach, which is not limited by memory and has a fast processing speed and good parallel scalability, is an efficient classification tool for handling large amounts of data and is similar to a decision tree classification algorithm [32]. The term "RF classifier" refers to a collection of tree-structured classifiers. It is a more sophisticated type of bagging that includes randomization. RF splits each node using the best combination among a group of predictors randomly picked at that node, rather than the best split among all variables. The original data set is used to create a new training data set with a replacement. The tree is then grown using random feature selection. The user must specify two parameters to start the RF algorithm. N and m are the two parameters in concern; the number of trees that must be established and the number of variables that must be considered to split each node, respectively. This method is also quick, resistant to overfitting, and allows the user to create as many trees as they wish. The number of trees was chosen at 100 to ensure excellent classification accuracy and avoid significant overfitting.

3) Accuracy Calculation :

Several evaluation metrics were employed to calculate the performance of the proposed system, as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (11)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (12)$$

$$\text{F1-score} = \frac{2\times \text{Recall}\times \text{Precision}}{\text{Recall}+\text{Precision}} \quad (13)$$

Where:

TP: represents if the sample is positive and correctly classified as positive, it is considered a true positive.

TN: represents if the sample is negative and correctly classified as negative, it is considered a true negative.

FN: denotes a false negative and occurs when a sample is positive but classified as negative.

FP: denotes a false positive and occurs when a sample is negative and classified as positive [33].

A Recall is the ratio of correct predictions in a class divided by the total number of elements in this class. Precision is the ratio of the number of correct predictions of a class divided by the total number of predictions for this class.

The dataset, which consisted of 16 classes, was trained using two models RF and SVM. The average accuracy was 92% for Random Forest and 91% for SVM . In addition, the confusion matrix for these classifiers is computed as shown in Fig. 3 and Fig. 4, respectively.

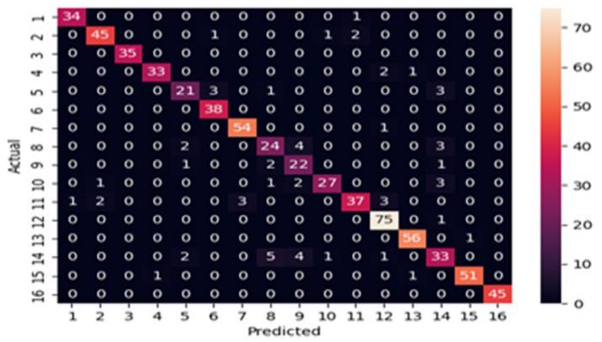

Fig. 3. Confusion Matrix for Random Forest classifier.

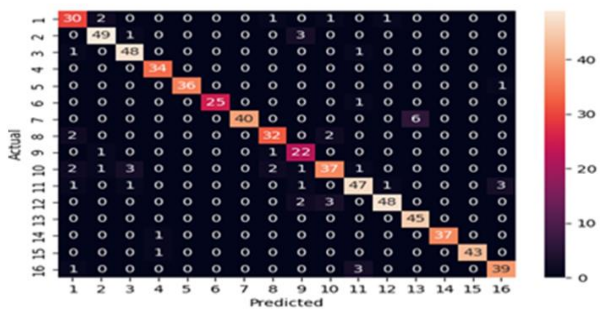

Fig. 4. Confusion Matrix for SVM classifier.

The four performance measures in the confusion matrix are TP, TN, FP, and FN. In each row of the matrix, the instances correspond to an actual class, whereas in each column, the instances correspond to a predicted class. The true positives for each class are represented by the diagonal matrix entries. The true positives of a single class are subtracted from the overall true positives of all classes to produce the true negatives for each class.

Table I shows the average Recall, Precision, and F1- score for each class for both Random Forest and SVM models.

TABLE I. RECALL, PRECISION, AND F1-SCORE VALUES FOR THE RANDOM FOREST(RF) AND SVM CLASSIFIERS

Class No.	Precision		Recall		F1-score	
	RF	*SVM*	*RF*	*SVM*	*RF*	*SVM*
1	0.79	0.97	0.86	0.97	0.82	0.97
2	0.89	0.94	0.89	0.92	0.89	0.93
3	0.91	1.00	0.96	1.00	0.93	1.00
4	0.94	0.97	1.00	0.92	0.97	0.94
5	1.00	0.81	0.97	0.75	0.99	0.78
6	0.96	0.90	0.96	1.00	0.96	0.95
7	1.00	0.95	0.87	0.98	0.93	0.96
8	0.91	0.73	0.89	0.73	0.90	0.73
9	0.76	0.69	0.92	0.85	0.83	0.76
10	0.86	0.93	0.81	0.79	0.84	0.86
11	0.91	0.93	0.93	0.80	0.92	0.86
12	0.98	0.91	0.89	0.99	0.93	0.95
13	0.88	0.97	1.00	0.98	0.94	0.97
14	1.00	0.75	0.97	0.72	0.99	0.73
15	1.00	0.98	0.98	0.96	0.99	0.97
16	0.93	1.00	0.88	1.00	0.90	1.00

IV. CONCLUSION

This research presents a WiFi-based human gesture recognition system that uses the advantage of the CSI of the wireless device. CSI data pre-processing, feature extraction, and classification have all been used to handle issues effectively. Random Forest and Support Vector Machine classifiers were used to classify 16 classes conducted by ten metrics, and the results demonstrated that in the WiAR dataset, the proposed approach yields high accuracy and outperforms the existing method. In future works, various pre-processing techniques, new features, and different machine learning and deep learning algorithms to increase recognition accuracy can be employed. Furthermore, the experiment only utilized signal amplitudes for the channel state information. Future studies can examine the mix of amplitude and phase information since the phase response may offer helpful information for recognition.

REFERENCES

[1] A. Jalal, M. Mahmood, and A. S. Hasan, “Multi-features descriptors for human activity tracking and recognition in Indoor-outdoor environments,” in 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 2019, pp. 371–376.

[2] J. Manjarres, P. Narvaez, K. Gasser, W. Percybrooks, and M. Pardo, “Physical workload tracking using human activity recognition with wearable devices,” Sensors, vol. 20, no. 1, p. 39, 2019.

[3] G. D’Angelo and F. Palmieri, “Enhancing COVID-19 tracking apps with human activity recognition using a deep convolutional neural network and HAR-images,” Neural Computing and Applications, pp. 1–17, 2021.

[4] L. Y. Chen et al., “Continuous wireless pressure monitoring and mapping with ultra-small passive sensors for health monitoring and critical care,” Nature communications, vol. 5, no. 1, pp. 1–10, 2014.

[5] C. Li, V. M. Lubecke, O. Boric-Lubecke, and J. Lin, “A review on recent advances in Doppler radar sensors for noncontact healthcare monitoring,” IEEE Transactions on microwave theory and techniques, vol. 61, no. 5, pp. 2046–2060, 2013.

[6] H. Abdelnasser, M. Youssef, and K. A. Harras, “Wigest: A ubiquitous wifi-based gesture recognition system,” in 2015 IEEE conference on computer communications (INFOCOM), 2015, pp. 1472–1480.

[7] B. Fang, N. D. Lane, M. Zhang, A. Boran, and F. Kawsar, “BodyScan: Enabling radio-based sensing on wearable devices for contactless activity and vital sign monitoring,” in Proceedings of the 14th annual international conference on mobile systems, applications, and services, 2016, pp. 97–110.

[8] B. Fang, N. D. Lane, M. Zhang, and F. Kawsar, “Headscan: A wearable system for radio-based sensing of head and mouth-related activities,” in 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), 2016, pp. 1–12.

[9] Y. Wang, K. Wu, and L. M. Ni, “Wifall: Device-free fall detection by wireless networks,” IEEE Transactions on microwave theory and techniques, vol. 16, no. 2, pp. 581–594, 2016.

[10] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, “E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures,” in Proceedings of the 20th annual international conference on Mobile computing and networking, 2014, pp. 617–628.

[11] B. Kellogg, V. Talla, and S. Gollakota, “Bringing gesture recognition to all devices,” in 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14), 2014, pp. 303–316.

[12] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, “Whole-home gesture recognition using wireless signals,” in Proceedings of the 19th annual international conference on Mobile computing & networking, 2013, pp. 27–38.

[13] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, “Widraw: Enabling hands-free drawing in the air on commodity wifi devices,” in Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, 2015, pp. 77–89.

[14] S. Tan and J. Yang, “WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition,” in Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing, 2016, pp. 201–210.

[15] M. S. Aljumaily and G. A. Al-Suhail, “Towards ubiquitous human gestures recognition using wireless networks,” International Journal of Pervasive Computing and Communications, 2017.

[16] L. Guo et al., “Wiar: A public dataset for wifi-based activity recognition,” IEEE Access, vol. 7, pp. 154935–154945, 2019.

[17] Z. Yang, Z. Zhou, and Y. Liu, “From RSSI to CSI: Indoor localization via channel response,” ACM Computing Surveys (CSUR), vol. 46, no. 2, pp. 1–32, 2013.

[18] 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.

[19] A. Virmani and M. Shahzad, “Position and orientation agnostic gesture recognition using wifi,” in Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, 2017, pp. 252–264.

[20] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, “WiFinger: Talk to your smart devices with finger-grained gesture,” in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2016, pp. 250–261.

[21] 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.

[22] M. A. A. Al-qaness and F. Li, “WiGeR: WiFi-based gesture recognition system,” ISPRS International Journal of Geo-Information, vol. 5, no. 6, p. 92, 2016.

[23] D. Jiang, M. Li, and C. Xu, “Wigan: A wifi based gesture recognition system with gans,” Sensors, vol. 20, no. 17, p. 4757, 2020.

[24] T. Zhang, T. Song, D. Chen, T. Zhang, and J. Zhuang, “WiGrus: A WiFi-based gesture recognition system using software-defined radio,” IEEE Access, vol. 7, pp. 131102–131113, 2019.

[25] L. Guo, L. Wang, J. Liu, W. Zhou, and B. Lu, "HuAc: Human activity recognition using crowdsourced WiFi signals and skeleton data," Wireless Communications and Mobile Computing, vol. 2018, 2018.

[26] F. Zhang et al., "Towards a diffraction-based sensing approach on human activity recognition," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies vol. 3, no. 1, pp. 1–25, 2019.

[27] R. Sudha, "Measures of Skewness and Kurtosis," Research and Biostatistics for Nurses. pp. 412–412, 2017. doi: 10.5005/jp/books/13016_16.

[28] J. Kilby, G. Mawston, and H. G. Hosseini, "Analysis of surface electromyography signals using continuous wavelet transform for feature extraction," 2006.

[29] R. Merletti and P. J. Parker, Electromyography: physiology, engineering, and non-invasive applications, vol. 11. John Wiley & Sons, 2004.

[30] H. A. Salim, M. Alaziz, and T. Y. Abdalla, "People Tracking in a Smart Room Using Kinect Sensor," in 2021 International Conference on Communication & Information Technology (ICICT), 2021, pp. 221–226.

[31] W. He, K. Wu, Y. Zou, and Z. Ming, "WiG: WiFi-based gesture recognition system," in 2015 24th International Conference on Computer Communication and Networks (ICCCN), 2015, pp. 1–7.

[32] I. A. Ibrahim, T. Khatib, A. Mohamed, and W. Elmenreich, "Modeling of the output current of a photovoltaic grid-connected system using random forests technique," Energy Exploration & Exploitation, vol. 36, no. 1, pp. 132–148, 2018.

[33] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 4, pp. 1–25, 2018.