Journal of Engineering Science and Technology
Vol. 20, No. 5 (2025) 1403 - 1413

WI-FI BASED HUMAN ACTIVITY RECOGNITION WITH TEMPLATE MATCHING

CHIA YEE SAW[1], YAN CHIEW WONG[1,*],
RANJIT SINGH SARBAN SINGH[2]

[1]Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer, Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal Melaka, Malaysia
[2]School of Engineering and Technology, Sunway University, Selangor, Malaysia
*Corresponding Author: ycwong@utem.edu.my

Abstract

Channel State Information (CSI) based human activity recognition (HAR) has gained popularity because of its privacy-protecting, light-insensitive, and devicefree benefits. The raw WiFi CSI data obtained from commercial WiFi devices is rich in human activity information and extremely sensitive to changes in the environment, human behavior, and the subject's weight and height. The data additionally consists of random noise from various sources and inter-personal variations, which may lead to false detection. This research proposes CSI-based HAR with pattern matching that is robust to noise and inter-personal adaptation. Human movements at consistent intervals over a set period of time are converted to pseudocolor images without preprocessing, allowing visual observation of whole pattern changes. The CSI features depicted in the image are then retrieved using a convolution neural network (CNN), and the resulting feature arrays are fed into a stochastic spiking neural network (SSNN) model-based simple cycle reservoir (SCR) for template matching. The proposed method demonstrates the ability to perform HAR based on partially captured signals, whereby the signal pattern of each segment can be observed in a single line of sight and used for person-to-person template recognition. This enables HAR to address interpatient variability with minimal computational complexity. The experimental results demonstrate that the proposed system achieves impressive performance in recognizing human activities with an overall accuracy of 94.81%. This approach simplifies the original complex CSI data with only 64 features and is robust in recognizing incomplete HAR signals and is capable of forecasting and predicting human activity based on historical data in future.

Keywords: Channel state information, Convolution neural network, Human activity recognition, Template matching, Time series prediction.

1. Introduction

HAR based on Wi-Fi signals have recently emerged as a promising approach, leveraging the reflective properties of WiFi signals. The core principle behind these systems is that human bodies are largely constituted of water, which may reflect WiFi signals [1]. The CSI is impacted by human movements via multipath effects, carries motion features and is more dependable than Received Signal Strength Indicator (RSSI) as it provides more information and fluctuates often based on the time and location. This wireless channel information includes information such as multipath fading, scattering, reflection, and other information is obtained during signal transmission between the transmitter and the receiver as shown in Fig. 1 without the use of any additional devices.

CSI-based human activity recognition (HAR) is a non-intrusive and privacy-preserving approach that utilizes Channel State Information (CSI) to detect human activities. Wireless sensing technologies, including millimeter-wave radar and Wi-Fi devices, provide distinct benefits, with Wi-Fi systems leveraging existing infrastructure to enable real-time environmental sensing through CSI analysis [2]. This approach proves advantageous over conventional HAR [3-5] in scenarios like elder care and intrusion detection. Additionally, it outperforms camera-based systems in coverage as WiFi signals can penetrate obstacles, overcoming challenges related to lighting conditions, blind spots, and high energy consumption, while also reducing the risk of privacy leakage during the detection process [6].

However, Wi-Fi-based Human Activity Recognition (HAR) encounters challenges in environments with significant signal noise [7]. The raw Channel State Information (CSI) data acquired from commercial Wi-Fi network interface cards (NICs) is often contaminated by random noise, particularly surrounding electrical devices. This noise introduces false edges, thereby compromising the accuracy and robustness of HAR systems.

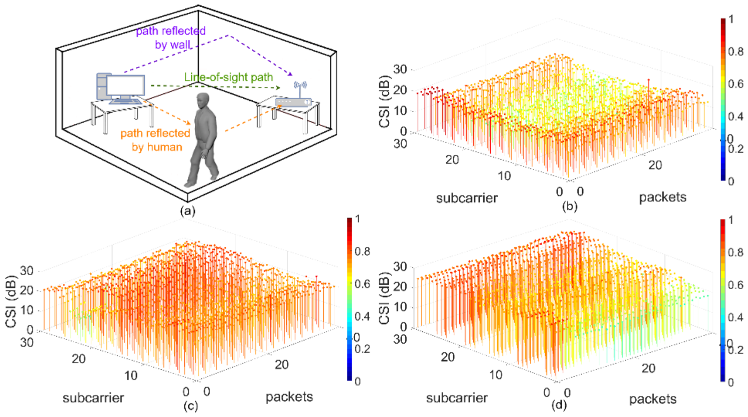

Fig. 1. Three-dimensional distribution of CSI amplitude in (a)Multipath for (b)Antenna 1 (c)Antenna 2 (d)Antenna.

Related work

Previous researchers have done extensive work on CSI data preprocessing for denoising, removing outliers, and merging large matrices [8-11]. Many studies averaged all 30 subcarriers into one signal, which was subsequently utilized to perform classification [8-12]. Principal component analysis (PCA) is then introduced and used in [13-15] to eliminate the influence of the background environment and perform dimensionality reduction over a small number of features. However, under such complicated multipath situations, even after applying a PCA based denoising procedure, commercial WiFi devices are unable to exhibit significant fluctuations in CSI signals caused by human activity. Support Vector Machines (SVM) and their advanced variants, such as CNN-SVM classifiers, have been utilized for identifying Channel State Information (CSI) signals associated with various human activities [16-19].

However, these methods exhibit notable limitations, including inefficiencies in handling high-dimensional data and challenges in addressing inter-patient variability. In comparison, Neural Networks demonstrate superior adaptability particularly in managing extensive datasets and non-linear relationship, which well-suited for diverse applications such as image recognition, natural language processing, and time-series analysis. LSTM-CNN models, despite their ability to capture temporal dependencies, suffer from high computational costs, prolonged training times, and significant memory requirements due to their intricate architecture of gates and memory cells [20]. These limitations hinder their scalability and real-time applicability. A comprehensive comparison of the methods, evaluating features, accuracy, computational complexity, and noise handling, is presented in Table 1.

Table 1. Comparison of methods based on features, accuracy, computational complexity, and noise handling.

	Features	**Accuracy**	**Computational Complexity**	**Noise Handling**
PCA	Dimensionality reduction, extracts principal components	Low to Moderate	Low	Sensitive to noise; requires preprocessing
SVM	Linear/ Non-linear classification	Moderate to High	Moderate (scales with data)	Struggles with high noise and outliers
CNN-SVM	CNN for feature extraction, SVM for classification	High	High	Dependent on CNN architecture
LSTM-CNN	CNN for spatial features; LSTM for temporal dependencies	Very High	Very High	Handles noisy time-series data
Proposed method	CNN for feature extraction, SSNN for classification	Very High	Moderate to High	Robust to noise and variability

To address these challenges, a robust CSI-based HAR with hybrid framework, Convolutional Neural Network-Template Matching (CNN-TM), is urged to be developed due to the lower computational complexity, easier in dealing with inter-patient variability and supports either fully or partial signals recognition. The 2D-CNN extracts difficult-to-detect features from pseudocolor images of CSI data, enabling visual observation of pattern changes without preprocessing, which avoids information loss [21].

A simple low pass filter can be added if the data were collected in high frequency noise environment. The extracted features are then fed into stochastic spiking neural network (SSNN) model for the implementation of template matching, which significantly reduces inter-patient variability and enables accurate recognition even with partially captured signals [22]. By leveraging CNN, template matching and stochastic computing, the proposed method achieves lower computational complexity, improved hardware efficiency, and robust noise handling, making it highly scalable, adaptable to dynamic and unpredictable real-world conditions and suitable for real-world deployments.

2. Method

This study proposes a recognition system that enables commercial WiFi devices to detect personal movement based on collected CSI readings. Human activities cause fluctuations in the signals captured by the Linux 802.11n CSI Tool [23], employed by the preceding researcher [13] in this study. The receiver is outfitted with a commercial Intel 5300 NIC with the sampling rate of 1 kHz. The dataset used is based on the assumption that a person who is not wearing any specific sensor or device undertakes random activity in a target region during data collection. A total of 59 samples of CSI sequence data, each comprising 30 subcarriers for every human activity category are collected using three antennas in an office space where the transmitter and receiver are 3 m apart in line-of-sight conditions.

2.1. Feature extraction using 2D-CNN

The human movements at consistent intervals over a set period are initially converted into pseudocolor images using the pcolor function. The color variations within these images, which correspond to the reflected raw CSI signals generated by human activities, are extracted utilizing a Convolutional Neural Network (CNN) incorporating a global average pooling layer. This process facilitates the direct transformation of human movements into pseudocolor images without preprocessing, while allowing observation of the signal pattern for each segment during human activity in the target area.

The CSI signal received is pseudocolor displayed along the number of packets in the order of the amplitude of the first, second, and third antennas as shown in Fig. 2(b). The pseudocolor plot of different activities differed in terms of motion duration and overall shape of the pattern. The active region of the CSI signal can be observed by the change in energy from high amplitudes in the yellow region to low amplitudes in the blue region.

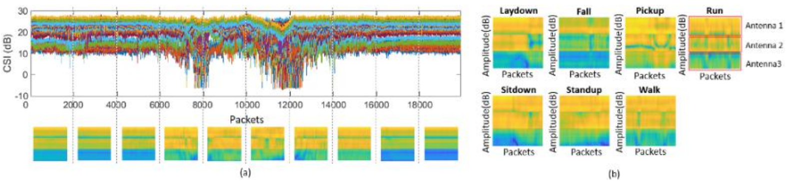

Fig. 2. Pseudocolor plot of (a) each segment during walking activity (b) different human activities.

2.2. Template matching based on simple cycle reservoir (SCR)

The retrieved feature in preceding section is then divided into 70% as training set and 30% as testing set and are used as input signals for the SSNN, down to each of the 20 stochastic neurons arranged based on simple cycle reservoir architecture, as depicted in Fig. 3. The input signals to a stochastic neuron are converted into pulsed signals, which are then multiplied and summed with the stochastic bitstream from the preceding neuron [22], facilitating neural information processing for enhanced hardware efficiency without compromising accuracy.

The resultant bitstream produced by the weighted sum of the inputs is converted into a binary integer and also further processing by another neuron. The evaluation duration has been set to Teval = (28 −1)·Tclk and the clock frequency is 1 kHz (Tclk = 1 ms). Figure 4 illustrates the output of several reservoir nodes in reaction to the input signal. Each neuron state oscillates in response to input stimuli, and a notable alignment becomes evident with an increasing number of trained neurons. While the 20th stochastic neuron closely resembles the initial input and achieves satisfactory accuracy, adding more stochastic neuron will increase the difficulty of computational time and also hardware implementation.

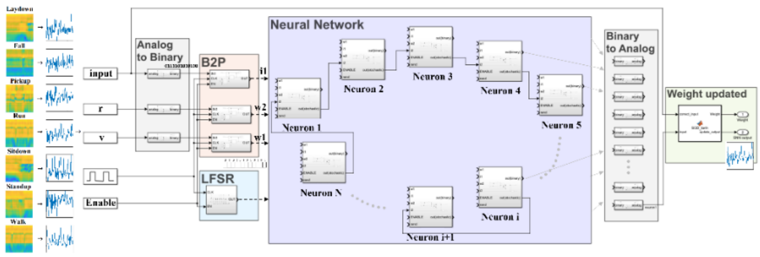

Fig. 3. SSNN that composed of 20 stochastic neurons with cyclic topology[22].

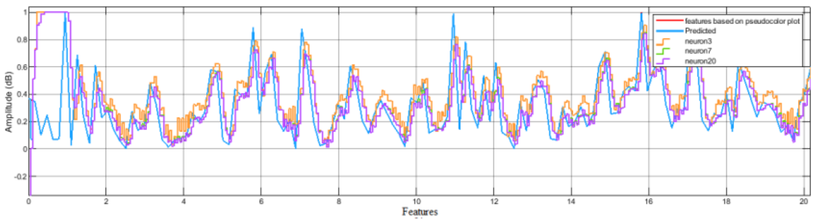

Fig. 4. The output of three selected neurons from a reservoir in response to extracted feature.

At the output layer, the training set for each activity group will be used as golden templates for pattern matching on the test dataset. This involved by calculating correlations with multiple duplicates of each template and subsequently classifying based on the highest correlation value. The cross-correlation function in Fig. 5 reveals a significant peak with a value of 0.91 at lag = 0, indicating a flawless match between the extracted feature and template. Template matching addresses inter-patient variability by comparing incoming CSI signals directly with patient-specific templates, instead of relying on a generalized model, ensuring accurate recognition. This approach successfully addresses concerns related to inter-patient variability while maintaining high accuracy in time series HAR. Additionally, it allows for easy updates to templates for specific patients, making the system adaptable and robust in real-world scenarios.

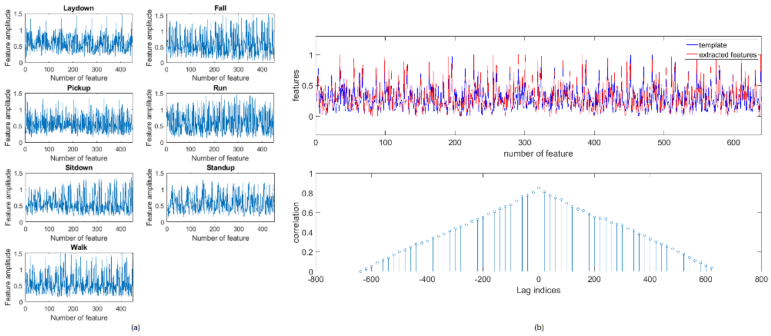

Fig. 5. (a) Cross-correlation between each template and (b) matches and results with lag index.

3. Results and Discussion

3.1. Performance of the proposed CNN-TM

Human activities have different CSI values, resulting in different recognition accuracy [24]. The performance of proposed method for each activity is summarized using a confusion matrix as shown in Fig. 6(b) and benchmarked against the conventional CNN as shown in Fig. 6(a). From the confusion matrices, activities involving more significant body movement, such as walking and running, have a better recognition accuracy due to their greater impact on CSI features. The pickup activity which has a quite similar body movement compared to the stand activity, giving that the ending position is identical and only the starting positions are different, causes to a less comparable effect on CSI values for both activities and lower classification accuracy in conventional CNN.

By employing pattern matching in time series, the proposed method reduces confusion between the two activities, resulting in significantly improved accuracy for both pickup and standing up activities. To further address this, Discrete Wavelet Transform (DWT) can be employed to decompose the CSI signals into multiple frequency components, capturing subtle variations in movement patterns. The system can better differentiate between activities with similar body postures, significantly improving recognition accuracy while maintaining computational efficiency.

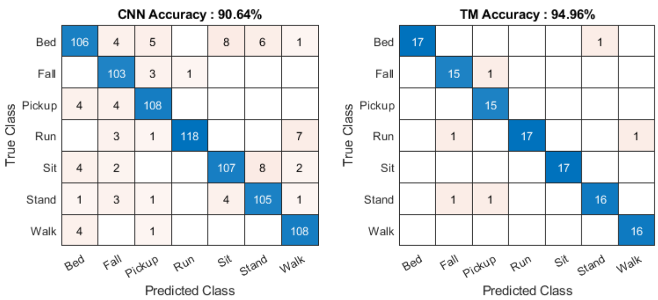

Fig. 6. Confusion matrix of (a) CNN and (b) TM.

3.2. CNN-TM with partial input signals

Time series recognition pattern matching based on partially captured signals, through cross-correlation with sliding windows between extracted features and templates, the accuracy distribution is shown in Fig. 7. The extracted features of every sample are denoted as $F_1, F_2, F_3, \cdots, F_n$ where F_n represents the number of features. A first reading of the sliding window size W_z is included in the first test signal. The $W z$ windows then slide for another s steps, and the following sample comprises of readings from $s + 1$ to $s + W_z + 1$. As a result, every three extracted feature segments are used as test signals and cross-correlated with the corresponding templates. The accuracy distribution based on 40 iterations, 30 epoch, 7:3 training ratio demonstrates that the third, fourth, and fifth steps achieves relatively higher accuracy.

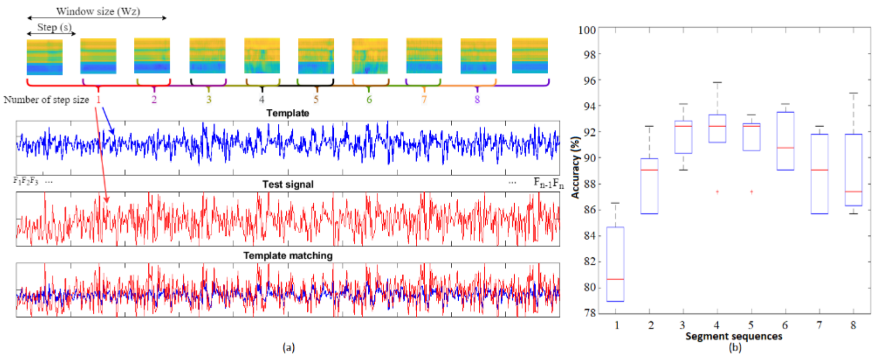

Fig. 7. (a) Every three segments cross-correlated with templates and the (b) accuracy distribution.

This is because the pseudocolor variations caused by human movements that depicted in the images only occurs during the active region; no variation is observed in the images during the inactive region. As a result, instead of requiring extracted features of the entire series for template matching, using features from a few segments that compose the active region is sufficient for categorization. This concerns is further evaluated by performing classification between various extracted feature segments and a template as shown in Fig. 8.

Recognizing activities from partial signals in human activity recognition holds significant practical importance, as it enables the system to function effectively even when data is incomplete or interrupted. This capability is particularly valuable in real-world deployments, such as elderly care monitoring, where continuous activity tracking is essential for fall detection or health assessment. By supporting partial signal recognition, the system becomes more robust, reliable, and adaptable to dynamic and unpredictable real-world conditions, enhancing its applicability across diverse scenarios.

Table 2 demonstrates that the proposed CNN-TM model surpasses previous methodologies, achieving higher accuracy in recognizing a wider range of human activities while eliminating preprocessing requirements. The proposed method allows recognition using either full or partial input signals, significantly enhancing its feasibility for time series activity recognition. Human movements at consistent intervals are converted into pseudocolor images without preprocessing and employing template matching to compare incoming CSI signals with patient-specific templates which effectively addresses inter-patient variability, ensuring adaptability and robustness in real-world scenarios. Collectively, these advancements improve noise robustness, computational efficiency, and recognition accuracy, making the system highly effective for diverse applications.

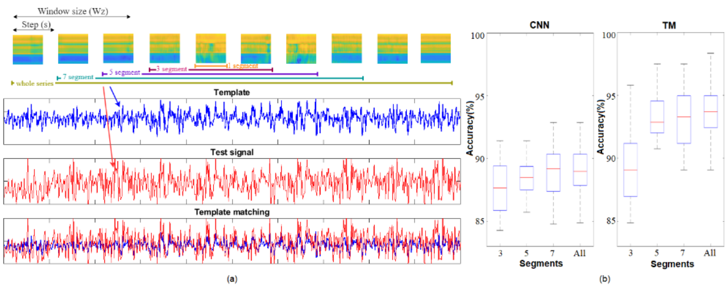

Fig. 8. (a) Different segments cross-correlated with templates and the (b) accuracy distribution.

Table 2. Comparison of proposed model and other model results.

	Number of activities	Average accuracy (%)	Average Loss	Precision (%)	Recall (%)	F1-score (%)
SVM [17]	5	91.27	-	84.01	-	85.96
LSTM [25]	5	82.72	0.5081	80.70	79.82	79.95
LSTM-CNN [25]	5	94.14	0.1269	91.03	83.81	87.10
LSTM [13]	6	90.50	-	-	-	-
CNN	7	86.58	-	86.85	86.89	86.72
CNN-TM	7	93.51	-	93.71	93.92	93.81

The proposed CSI-based HAR system is developed using Verilog hardware description and synthesized on the Xilinx Zynq-7000 FPGA using Vivado, demonstrating significant hardware efficiency. As shown in Table 3, the proposed SSNN utilizes the least number of lookup tables (LUTs) and flip-flops (FFs) for categorizing distinct human activities compared to previous studies. By leveraging stochastic computing, the system reduces the hardware required for arithmetic calculations through probabilistic techniques, while the simple cycle reservoir (SCR) architecture minimizes the number of connections, further simplifying the network design. Additionally, template matching at the output layer enables recognition using partial input signals, enhancing the system's robustness and adaptability to dynamic, real-world conditions compared to previous research. This capability, combined with hardware-efficient design, ensures reliable performance in applications such as elderly care monitoring, where data is incomplete or continuous activity tracking for fall detection.

Table 3. Hardware resource utilization comparison.

	Lei et al. [26]	Vita et al.[27]	This Approach
FPGA	Zynq XC7Z045	Artix-7	Zynq 7000
Classifier Type	CNN	BNN	SSNN
LUT	79637	5988	692 (0.65%)
FF	48972	4299	440 (0.83%)

4. Conclusions

In this study, we have demonstrated that our proposed method enables robust sequence matching, which is much simple and capable than conventional techniques. The pseudocolor variations induced by human movements are retrieved using 2D-CNN without preprocessing, while allowing visual viewing of entire pattern alterations and minimize computational complexity. The SC approach substantially reduces the required hardware, while SCR architecture implemented in network design further minimizes the number of connections. The template matching approach significantly mitigates inter-patient variability issues and enables high-precision HAR even with only a quarter of the input signal, indicating that the system is effective and robust in time series human activities classification.

In real-world scenarios like multi-room or dynamic environments with higher interference, CSI-based human activity recognition can face challenges such as signal attenuation, noise, and overlapping activities and environmental variations can lead to reduced recognition accuracy. Addressing these limitations will be a focus of future work, with an emphasis on developing advanced algorithms to effectively handle multiple users and adapt to diverse environments. Also, the proposed method demonstrates strong performance in line-of-sight conditions with the transmitter and receiver placed 3 meters apart, future work will focus on enhancing its generalizability in non-line-of-sight (NLOS) scenarios. This includes leveraging advanced signal processing techniques and multipath propagation analysis to ensure robust activity recognition in environments with obstacles or dynamic layouts, broadening its applicability in real-world deployments.

Acknowledgements

The authors acknowledge the technical and financial support by the Ministry of Higher Education, Malaysia, under the research grant no. FRGS/1/2024/ICT02/UTEM/02/3 and Universiti Teknikal Malaysia Melaka (UTeM).

Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Yu, N.; Wang, W.; Liu, A.X.; and Kong, L. (2018). Qgesture: Quantifying gesture distance and direction with wifi signals. *Proceedings of the ACM on Interactive*, *Mobile*, *Wearable and Ubiquitous Technologies*, 2(1), 1-23.
2. Wang, Z.; Li, J.; Wang, W.; Dong, Z.; Zhang, Q.; and Guo, Y. (2024). Review of few-shot learning application in CSI human sensing. *Artificial Intelligence Review*, 57(8), 195.
3. Oleh, U.; Obermaisser, R.; and Ahammed, A.S. (2024). A review of recent techniques for human activity recognition: Multimodality, reinforcement learning, and language models. *Algorithms*, 17(10), 434.
4. Bhuiyan, M.S.H.; Patwary, N.S.; Saha, P.K.; and Hossain, M.T. (2020). Sensor-based human activity recognition: A comparative study of machine

learning techniques. *Proceedings of the* 2020 2nd *International Conference on Advanced Information and Communication Technology* (*ICAICT*), Dhaka, Bangladesh, 286-290.

5. Alema Khatun, M.; and Abu Yousuf, M. (2020). Human activity recognition using smartphone sensor based on selective classifiers. *Proceedings of the* 2020 2nd *International Conference on Sustainable Technologies for Industry* 4.0 (*STI*), Dhaka, Bangladesh, 1-6.
6. Abuhoureyah, F.; Sim K.S.; and Chiew Wong, Y. (2024). Multi-user human activity recognition through adaptive location-independent wifi signal characteristics. *IEEE Access*, 12, 112008-112024.
7. Lin, C.-Y.; Lin, C.-Y.; Liu, Y.-T.; Chen, Y.-W.; and Shih, T.K. (2024). Wifi-tcn: Temporal convolution for human interaction recognition based on wifi signal. *IEEE Access*, 12, 126970-126982.
8. Yang, J.; Liu, Y.; Liu, Z.; Wu, Y.; Li, T.; and Yang, Y. (2021). A framework for human activity recognition based on wifi csi signal enhancement. *International Journal of Antennas and Propagation*, 2021, 1-18.
9. Schäfer, J.; Barrsiwal, B.R.; Kokhkharova, M.; Adil, H.; and Liebehenschel, J. (2021). Human activity recognition using csi information with nexmon. *Applied Sciences*, 11(19), 8860.
10. Ding, J.; and Wang, Y. (2019). Wifi csi-based human activity recognition using deep recurrent neural network. *IEEE Access*, 7, 174257-174269.
11. Guan, L.; Hu, F.; Al-Turjman, F.; Khan, M.B.; and Yang, X. (2019). A non-contact paraparesis detection technique based on 1D-CNN. *IEEE Access*, 7, 182280-182288.
12. Wang, H.; Zhang, D.; Wang, Y.; Ma, J.; Wang, Y.; and Li, S. (2017). Rt-fall: A real-time and contactless fall detection system with commodity wifi devices. *IEEE Transactions on Mobile Computing*, 16(2), 511-526.
13. Yousefi, S.; Narui, H.; Dayal, S.; Ermon, S.; and Valaee, S. (2017). A survey on behavior recognition using wifi channel state information. *IEEE Communications Magazine*, 55(10), 98-104.
14. Wu, X.; Chu, Z.; Yang, P.; Xiang, C.; Zheng, X.; and Huang, W. (2018). Tw-see: Human activity recognition through the wall with commodity wi-fi devices. *IEEE Transactions on Vehicular Technology*, 68(1), 306-319.
15. Al-qaness, M.A.A. (2019). Device-free human micro-activity recognition method using wifi signals. *Geo-spatial Information Science*, 22(2), 128-137.
16. Chen, Z.; Zhu, Q.; Soh, Y.C.; and Zhang, L. (2017). Robust human activity recognition using smartphone sensors via CT-PCA and online SVM. *IEEE Transactions on Industrial Informatics*, 13(6), 3070-3080.
17. Alsaify, B.A.; Almazari, M.M.; Alazrai, R.; Alouneh, S.; and Daoud, M.I. (2022). A csi-based multi-environment human activity recognition framework. *Applied Sciences*, 12(2), 930.
18. Basly, H.; Ouarda, W.; Sayadi, F.E.; Ouni, B.; and Alimi, A.M. (2020). *CNN-SVM learning approach based human activity recognition*. In El Moataz, A.; Mammass, D.; Mansouri, A.; and Nouboud, F. (Eds.), *Image and Signal Processing*. Springer International Publishing, 271-281.

19. Nawal, Y.; Oussalah, M.; Fergani, B.; and Fleury, A. (2023). New incremental svm algorithms for human activity recognition in smart homes. *Journal of Ambient Intelligence and Humanized Computing*, 14(10), 13433-13450.
20. Kandadi, Thirupathi; and Shankarlingam, G. (2025). Drawbacks of LSTM algorithm: A case study. *Available at SSRN* 5080605.
21. Fard Moshiri, P.; Shahbazian, R.; Nabati, M.; and Ghorashi, S.A. (2021). A CSI-based human activity recognition using deep learning. *Sensors*, 21(21), 7225.
22. Saw, C.Y.; and Wong, Y.C. (2022). Neuromorphic computing based on stochastic spiking reservoir for heartbeat classification. *Jordanian Journal of Computers and Information Technology* (*JJCIT*), 8(02), 184-195.
23. Halperin, D.; Hu, W.; Sheth, A.; and Wetherall, D. (2011). Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review*, 41(1), 53-53.
24. Moshiri, P.F.; Navidan, H.; Shahbazian, R.; Ghorashi, S.A.; and Windridge, D. (2020). Using GAN to enhance the accuracy of indoor human activity recognition. *arXiv preprint arXiv:*2004.11228.
25. Shang, S.; Luo, Q.; Zhao, J.; Xue, R.; Sun, W.; and Bao, N. (2021). LSTM-CNN network for human activity recognition using WiFi CSI data. *Journal of Physics*: *Conference Series*, 1883(1), 012139.
26. Lei, P.; Liang, J.; Guan, Z.; Wang, J.; and Zheng, T. (2019). Acceleration of FPGA based convolutional neural network for human activity classification using millimeter-wave radar. *IEEE Access*, 7, 88917-88926.
27. Vita, A.D.; Pau, D.; Benedetto, L.D.; Rubino, A.; Petrot, F.; and Licciardo, G.D. (2020). Low power tiny binary neural network with improved accuracy in human recognition systems. *Proceedings of the* 2020 23rd *Euromicro Conference on Digital System Design* (*DSD*), Kranj, Slovenia, 309-315.