Wi-SensiNet: Through-Wall Human Activity Recognition Based on WiFi Sensing

Fu Xiaoyi
Southwest University
Wang Chenlu
Southwest University
Li Shenglin
lishenglin@swu.edu.cn

Southwest University

Research Article

Keywords: Human activity recognition, Deep learning, WiFi sensing, Through wall,Channel state information

Posted Date: March 19th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4106293/v1

Additional Declarations: No competing interests reported.

Wi-SensiNet: Through-Wall Human Activity Recognition Based on WiFi Sensing

Fu Xiaoyi[1], Wang Chenlu[1], Li Shenglin[1*]

[1*]College of Artificial Intelligence, Southwest University, Chongqing, 400715, China.

*Corresponding author(s). E-mail(s): lishenglin@swu.edu.cn;
Contributing authors: swu_fuxiaoyi@163.com;

Abstract

With the advancement of Wi-Fi sensing technology, its significant benefits in convenient operation and privacy protection have become apparent, particularly in fields like smart homes, medical monitoring, and security surveillance, where the application prospects of Human Activity Recognition (HAR) technology are increasingly broad. This study focuses on a novel approach to HAR using Wi-Fi Channel State Information (CSI), especially under complex conditions such as Non-Line of Sight (NLoS) paths and through-wall transmissions. Traditionally, most research has concentrated on Line of Sight (LoS) path HAR, sensitive to environmental changes, while the NLoS path signals, especially through-wall signals, present unpredictability due to weak reflections caused by walls. Addressing this issue, we propose Wi-SensiNet, an innovative deep learning-based method that combines the spatial feature extraction capabilities of Convolutional Neural Networks (CNN) with the temporal sequence processing power of Bidirectional Long Short-Term Memory networks (BiLSTM). This method also incorporates an attention mechanism to enhance the accuracy of human activity recognition in complex environments. Wi-SensiNet is specially designed for through-wall settings, effectively handling the complexity of CSI data, and achieving accurate through-wall human activity detection. In our experiments, we collected a through-wall CSI dataset comprising seven common activities, including running, sitting, standing, squatting, falling, punching, and walking, and verified Wi-SensiNet's average accuracy exceeded 99% on the original test set. These results not only demonstrate the model's robustness and high accuracy in handling HAR tasks in complex environments but also highlight the potential of CNN and BiLSTM working in tandem to enhance performance.

Keywords: Human activity recognition, Deep learning, WiFi sensing, Through wall,Channel state information

1 Introduction

In recent years, Human Activity Recognition (HAR) has increasingly found applications in fields such as smart homes, medical monitoring, and security surveillance.Particularly with the proliferation of smart devices and the advancement of Internet of Things (IoT) technologies, Wi-Fi sensing technology has been increasingly applied, harnessing a detectable feature within Wi-Fi signals known as Channel State Information (CSI).This type of data can intricately illustrate how wireless signals are transmitted between sending and receiving devices, where movement of a person within the Wi-Fi signal propagation path induces changes in the signal's phase and amplitude.These alterations are reflected in the CSI, a characteristic of Wi-Fi signals, offering detailed insights into the signal propagation path, transmission time, and attenuation.Consequently, by analyzing the variations in CSI data, one can infer changes in human movement and positioning.Compared to traditional methods reliant on visual or wearable sensors, this approach offers a more covert and non-line-of-sight dependent solution.Wi-Fi-based Human Activity Recognition (HAR) is achieved through the reception of both Signal Strength Information (RSSI) and Channel State Information (CSI).In comparison to RSS, CSI can provide detailed channel frequency response information[1] on multiple channels at the physical layer, offering finer granularity than RSSI and the ability to distinguish multipath components, rendering it more effective in recognizing complex human movements.Owing to its high-resolution characteristics, CSI has emerged as a pivotal technology in the HAR domain, particularly excelling in applications requiring precise capture and analysis of human dynamics.

The majority of solutions for Human Activity Recognition (HAR) based on Channel State Information (CSI) in specific environments employ deep learning techniques, including gesture recognition[2][3], motion detection[4][5], and fall detection[6][7].However, researchers often conduct their studies in ideal environments with relatively simple Wi-Fi signal propagation. Due to the poor interference resistance of Wi-Fi signals, the accuracy of these solutions may significantly diminish in more complex scenarios such as through-wall or Non-Line-of-Sight (NLOS) conditions.However, these complex scenarios often represent the primary application contexts for human activity recognition.The challenges encountered in human perception through walls using Wi-Fi include technical limitations such as signal attenuation, multipath interference, signal noise, environmental factors like wall composition, dynamic environments, as well as privacy and ethical considerations.Consequently, achieving high-accuracy human activity recognition in complex environments has become a critical issue.Human Activity Recognition based on CSI fundamentally involves analyzing the impact of human activity on wireless signals, suggesting that a potential solution to the aforementioned issues is to extract and utilize the most representative features within the wireless signals.Addressing these challenges is crucial for the advancement of this field.Robust algorithms, models, and frameworks are required to mitigate the effects of signal issues and adapt to dynamic environments.Furthermore, the latest advancements in Multiple-Input Multiple-Output (MIMO) communications and the utilization of Wi-Fi signals offer a promising avenue to realize this goal.

In this study, we conducted an in-depth analysis of the original Channel State Information (CSI), with a particular focus on the impact of signal attenuation and minor movements on CSI in through-wall scenarios.To fully extract features, we treated the data from each antenna as an independent input channel, incorporating both time-domain and frequency-domain information.Additionally, we employed median filtering to optimize data quality and reduce noise, coupled with the introduction of Gaussian noise enhancement to simulate uncertainties in real-world environments, effectively mitigating noise interference caused by minor and inadvertent movements.These strategies collectively enhanced the reliability of the data, improving the model's performance and adaptability in complex environments.

In summary, this research makes the following key contributions:

- In addressing the challenges of datasets in through-wall scenarios, we devised and applied a data processing strategy combining median filtering with Gaussian noise, which not only effectively reduced data noise but also significantly enhanced the quality and generalizability of the data.Training with pre-processed data led to a substantial increase in accuracy on the original dataset.
- We introduced the Wi-SensiNet approach, which integrates Convolutional Neural Networks (CNN) with Bidirectional Long Short-Term Memory (BiLSTM) networks, incorporating an attention mechanism to establish an advanced framework for time-series analysis. The specially tailored BiLSTM modules process deep features extracted by the CNN layers, and the incorporated attention layer further optimizes the model's recognition of critical temporal steps, enhancing the overall accuracy and robustness of the model in understanding and predicting complex time-series data.
- We collected and constructed a through-wall dataset encompassing nine common human activities, including running, sitting, standing, squatting, falling, punching, and standing.By applying the Wi-SensiNet method, we achieved an average accuracy of 99% on the original dataset, validating the efficacy and precision of our approach in practical applications. application.

The remainder of this paper is structured as follows: Chapter 2 provides a comprehensive review of existing studies in the related field,Chapter 3 details the dataset employed in this research,Chapter 4 describes the experimental procedure and results, and includes a comparative analysis with existing studies.Chapter 5 concludes the paper with key findings and insights.

2 Related works

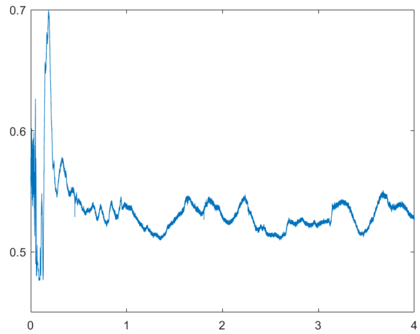
In this chapter, we conduct a thorough review of the major works related to our study, including the application of Channel State Information (CSI) data, Human Activity Recognition (HAR) technologies in Non-Line-of-Sight (NLOS) environments, and the application of deep learning in this field.In 2000, Bahl et al. [8] introduced Radar, an innovative system for indoor positioning based on Wi-Fi Received Signal Strength (RSS), marking a pioneering use of Wi-Fi signals for sensing purposes.Subsequently, CSI data became accessible in a multitude of commercial devices, with prominent tools such as the Intel 5300 NIC [9], Atheros CSI Tool[10], and Nexmon CSI Tool[11] emerging.These tools facilitated the creation of numerous platforms, significantly simplifying the process of CSI data collection.CSI data, being a signal rich in environmental information, has been extensively utilized for indoor activity monitoring.

Wang et al. developed the CARM system[12], a human activity recognition and monitoring system based on Channel State Information (CSI), grounded in two theoretical models: the CSI-Speed Model and the CSI-Activity Model. This system correlates CSI dynamics with human movement speeds and associates these speeds with specific activities.This model-based approach, however, exhibits limitations in environmental adaptability, diversity in activity recognition, real-time processing efficiency, generalization capabilities, and in handling complex data.With the advancement of deep learning, researchers are increasingly leveraging this technology for motion recognition.Ma et al. introduced the SignFi method[13], a system architecture primarily based on the Channel State Information (CSI) of Wi-Fi signals, employing Convolutional Neural Networks (CNN) for the classification and recognition of sign language gestures.The system analyzes Wi-Fi signal variations caused by gesture movements to extract features of sign language.Shi et al. described a method for human activity recognition in their study[14].This method initially employs an algorithm to preprocess CSI signals, enhancing activity-related signals and reducing noise.Subsequently, it extracts features by computing correlations in both time and frequency domains.Utilizing these processed signals, the method automatically extracts deeper features through a deep learning model.

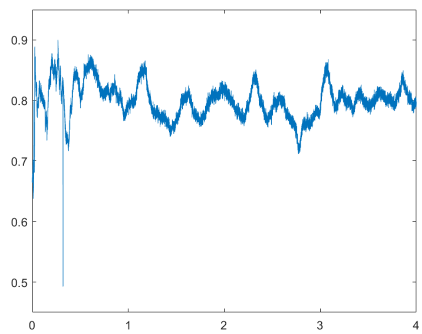
While exploring Wi-Fi-based motion recognition technologies, researchers have also expanded their focus to more challenging environments, leveraging the through-wall capabilities of Wi-Fi signals for human activity recognition in such scenarios.Wang et al.[15] introduced a method utilizing Device-Free Sensing (DFS) technology in complex scenarios, with a particular emphasis on its application in through-wall and Non-Line-of-Sight (NLOS) environments.The paper proposed a novel strategy to enhance DFS system performance by utilizing spatial structural information, integrating multi-dimensional features across time, frequency, and spatial domains for more accurate identification of target positions and activities.In conducting research on through-wall activity recognition using commercial Wi-Fi devices, we encounter several challenges.In this paper, we propose a deep learning-based approach that effectively mitigates the impact of noise and environmental factors.This approach demonstrates strong generalization capabilities, introducing a novel method to the field of through-wall activity recognition.

3 Dataset

To ensure the quality of the collected data and the generalization ability of the model, we initially conducted a series of preprocessing steps on the original dataset.This included the application of various noise reduction techniques to minimize noise interference in the data, as well as the implementation of data augmentation strategies to optimize the dataset and increase sample diversity.Noise reduction processing aims to enhance the clarity of data signals, thereby enabling the model to more accurately capture subtle variations in human movements.Data augmentation, by simulating potential variations and disturbances, enhances the model's adaptability to new scenarios and conditions.In the following sections, we will provide a detailed description of the noise reduction and data augmentation methods employed.



(a) Non-through-wall action



(b) Through-wall actionn

Fig. 1: Comparison of amplitude values after normalization and variance processing. The horizontal axis represents the variation in time, while the vertical axis indicates the changes in amplitude data post-normalization and variance processing.

3.1 Original Dataset

The original dataset was collected using the Intel 5300 NIC, which defines the Channel State Information (CSI)[15] as a function, given by the formula

$$H_t^{c,m} = A_t^{c,m} e^{j\phi_t^{c,m}} \tag{1}$$

Here, $H_t^{c,m}$,$A_t^{c,m}$and$\phi_t^{c,m}$ represent the complex channel response, amplitude response, and phase response, respectively, for channel c on antenna m at time t.In the original dataset collected for this study, the dimensions of each data sample are defined by the equation $X_t = N_T \times N_R \times N_{sub}$, where N_T, N_R, N_{sub} denote the number of transmitting antennas, receiving antennas, and subcarriers on each antenna, respectively.For the through-wall Channel State Information (CSI) dataset, due to the complexity of through-wall environments, the original CSI data is frequently subject to various interferences, such as signal attenuation, noise disturbances, and environmental changes.

As illustrated in Fig. 1, we utilize Normalized Variance (NV) to evaluate both non-through-wall and through-wall data.To accentuate the differences between the two scenarios, we apply normalized variance analysis to through-wall and non-through-wall data within a specific action time window.Fig.1a demonstrates the amplitude variations caused by a regular standing action, whereas Fig.1b exhibits the amplitude

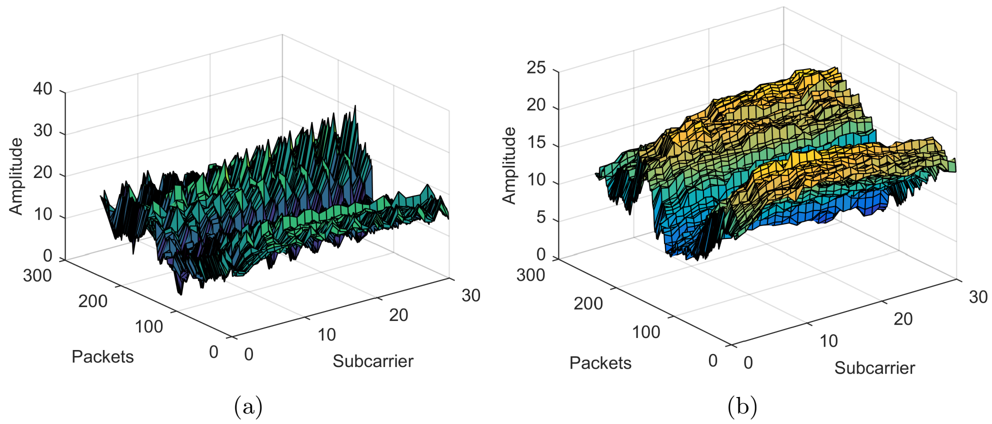


Fig. 2: Comparison of data before and after preprocessing. In the visual representation, (a) denotes the original data, while (b) represents the data post-preprocessing. The x-axis signifies the carrier index, the y-axis denotes the packet index, and the z-axis corresponds to the amplitude.

variations of the same action under through-wall conditions.The distinction between the two is evident; the through-wall data exhibits a greater normalized variance, indicating an increase in the volatility of the channel state. The calculation formula for Normalized Variance (NV) is as follows:

$$NV = \frac{\sigma^2}{\mu^2} \tag{2}$$

here,σ^2represents the variance, calculated as:

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{3}$$

μ denotes the sample mean, computed as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

Here, x_i represents an individual sample value, and N denotes the number of samples.Utilizing this formula allows for the quantification of channel quality variations under through-wall and non-through-wall conditions.The observation of a higher Normalized Variance in through-wall scenarios indicates increased signal volatility, potentially caused by multipath effects from walls or other environmental factors.This increase in volatility suggests the need for appropriate data preprocessing strategies to adapt to such changes. Preprocessing may include filtering, denoising, or employing more sophisticated signal processing techniques to stabilize the signal, thereby enhancing the accuracy and reliability of subsequent analyses.

3.2 Data Preprocessing

In this study, we employed a median filter for noise reduction in Channel State Information (CSI) data.A non-linear digital filtering technique, often used in signal processing to reduce noise. This method preserves edges while removing noise, by replacing each entry with the median of neighboring entries.Median filters are highly advantageous in image and signal processing, particularly effective in eliminating salt-and-pepper noise, while efficiently preserving the edge characteristics of images.As a nonlinear filtering technique, it is less sensitive to outliers, making it particularly effective in processing signals with anomalies or noise.The fundamental equation for median filtering is

$$y(i) = \mathrm{Med}\{x(i-k), ..., x(i+k)\} \tag{5}$$

where $x(i)$ represents the original signal, $y(i)$ is the signal post-filtering, and k is half the window size. In this context, we have set the window size of the median filter to 3x3, to effectively eliminate noise while preserving the edge features of the signal.

Additionally, considering the potential attenuation of features due to through-wall signals, we incorporated Gaussian noise augmentation to enhance the feature representation of the dataset.Adding Gaussian Noise is a process in signal processing where Gaussian noise (a statistical noise having a probability density function equal to that of the normal distribution) is added to a signal. This technique is commonly used in algorithms to test robustness against noise or to improve generalization.The addition of Gaussian noise follows the principle of $z(i) = x(i) + N(\mu, \sigma^2)$ In this context,,$N(\mu, \sigma^2)$ represents a Gaussian distribution with a mean of $\mu$and a variance denoted by $\varepsilon = \sigma^2$.In our experiments, we set the mean of the Gaussian noise to 0 and the standard deviation to 0.1, ensuring that the noise level was moderate and did not obscure the characteristics of the original signal.This approach not only simulates the uncertainties of real-world environments but also enhances the data's representational capacity while preserving the characteristics of the original signal, making it particularly suitable for through-wall signal processing scenarios.As demonstrated in Fig. 2, we present a comparison of the data before and after preprocessing.

4 System Architecture

In this study, a deep learning-based algorithmic framework is developed, aimed at processing and classifying Channel State Information (CSI) data collected through walls. As depicted in Fig. 4, the entire processing pipeline is segmented into several pivotal stages: Initially, the preprocessed through-wall CSI data is fed into a Convolutional Neural Network (CNN) module. This module leverages its multiple convolutional layers to autonomously extract spatial features from the data. The features outputted by the CNN module are then reshaped to conform to the input specifications of subsequent modules. Subsequently, the reshaped time-series data is input into a Bidirectional Long Short-Term Memory (BiLSTM) module. The architecture of the BiLSTM module is meticulously designed to process time-series data, with a specific focus on capturing extended dependencies in the temporal dimension. Following this, the output from the BiLSTM module is conveyed to a fully connected layer, which is tasked with mapping the temporal features onto specific categories of actions.

To augment the model's proficiency in interpreting time-series data, an attention mechanism has been integrated. This mechanism dynamically modulates the model's focus across various time steps in the sequence. By assigning attention weights to the hidden states of the Bidirectional Long Short-Term Memory (BiLSTM) module, the model intensively processes dynamic features that are essential for the classification task. This enhancement notably escalates the model's accuracy in recognizing human activities through walls, especially in environments with complex signal patterns. Consequently, the proposed CNN-BiLSTM framework, enriched with the attention mechanism, amalgamates the strengths of spatial feature extraction and temporal dependency capture. Furthermore, it attains an in-depth comprehension of time-series data via the incorporation of the attention model. This culminates in a novel and efficacious methodology for human activity recognition in through-wall scenarios.

4.1 Convolutional Neural Network

In our research, for the three-dimensional through-wall Channel State Information (CSI) dataset, we employed a two-dimensional Convolutional Neural Network (CNN2D) module to perform feature extraction tasks.This dataset possesses dimensions of 3x500x30, corresponding respectively to the number of antennas, time steps, and amplitude information for each antenna.The design of the CNN2D module specifically accounts for the unique structure of CSI data, aiming to efficiently extract spatial features embedded within the time-series data.

As depicted in Fig. 3, the CNN2D module comprises two convolutional layers, each followed by a Rectified Linear Unit (ReLU) and a max pooling layer.The first convolutional layer utilizes 3x3 kernels (with a stride of 1 and padding of 1) to process signals from different antennas, generating feature maps containing primary spatial features.Subsequently, the ReLU activation function is applied to introduce nonlinearity, aiding in capturing more complex data patterns. Subsequently, a max pooling layer with 2x2 kernels and a stride of 2 is utilized to reduce the spatial dimensions of

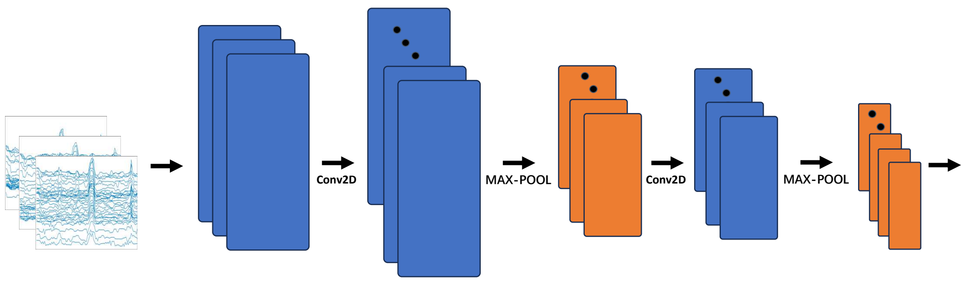


Fig. 3: Schematic Diagram of the CNN Module.

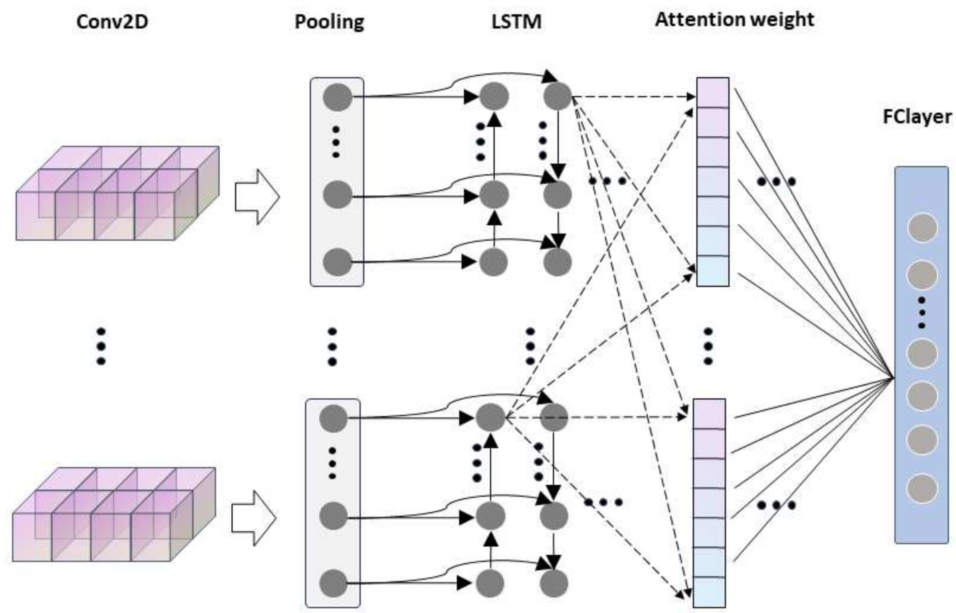


Fig. 4: Algorithm process framework.

the feature maps while retaining essential features.This downsampling step is aimed at reducing computational load and preventing overfitting.The second convolutional layer further processes the features, and another ReLU activation function is employed to maintain nonlinearity.Then, max pooling is applied again to further reduce the size of the feature maps, facilitating higher-level feature abstraction.The design of this CNN module takes into account the uniqueness of through-wall CSI data. With carefully selected parameters and layer structure, it ensures that the network effectively extracts crucial spatiotemporal features from the data, significantly enhancing the accuracy and efficiency of subsequent classification tasks.

4.2 BiLSTM

In the architecture we propose, a Bidirectional Long Short-Term Memory network (BiLSTM) is employed following the Convolutional Neural Network (CNN) module to effectively process sequential data. As illustrated in Fig. 5,the BiLSTM enhances the learning of sequence features by capturing the contextual information of the time series from both forward and backward directions simultaneously. In its specific implementation, the BiLSTM consists of a single LSTM layer, with the input feature dimension specified as the reshaped output of the preceding layer, while the dimension of the hidden layer is set to 64. Furthermore, we ensure that the batch size of the input and output tensors is positioned in the first dimension. The bidirectional processing capability of the LSTM is also activated. This forms a comprehensive framework for analyzing time-series data. The LSTM module is renowned for its efficiency in capturing long-term dependencies in sequential data and is specifically tailored to handle the feature-rich outputs extracted by the preceding CNN layer. At each time step t, the forward LSTM segment computes the hidden state $\overrightarrow{h_t}$ using the current input x_t and the hidden state $\overrightarrow{h_{t-1}}$ from the previous time step, while the backward LSTM calculates $\overleftarrow{h_t}$ based on the same input and the hidden state $\overleftarrow{h_{t+1}}$ from the subsequent time step. The output h_t of the bidirectional LSTM at each time step is a concatenation of the forward and backward hidden states, formulated as $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$. This structure allows the model to integrate the information of the entire input sequence at each time step, thus more comprehensively capturing the long-term dependencies in the sequence.

4.3 Attention mechanism

Additionally, our model incorporates an attention mechanism, aimed at further enhancing the identification of key sequential information. This attention mechanism assigns a weight to each time step output of the bidirectional LSTM, thereby emphasizing more significant features while suppressing less relevant information. Specifically, we initially transform the BiLSTM output at each time step into a scalar using a linear layer, a step that can be viewed as scoring the importance of each time step. Subsequently, we apply a softmax function to normalize these scores, ensuring that the sum

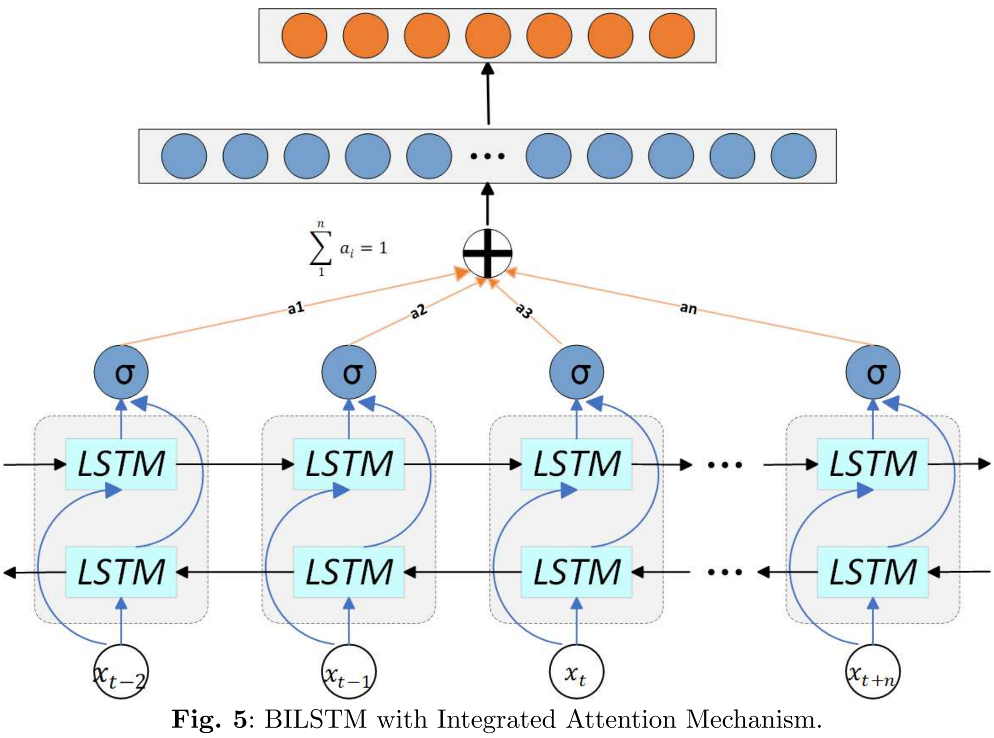


Fig. 5: BILSTM with Integrated Attention Mechanism.

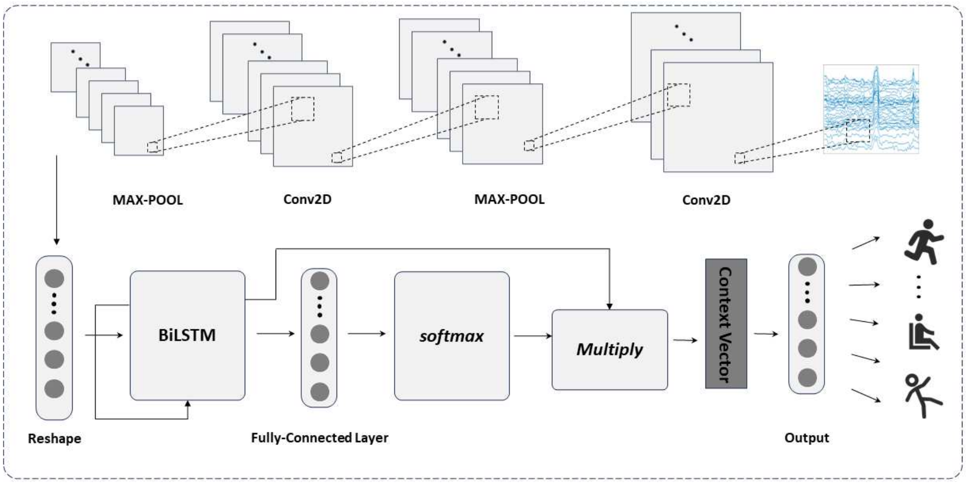


Fig. 6: System Workflow.

of attention weights across all time steps equals 1. This process can be represented as:

$$a_t = \text{softmax}(W_{attn} \cdot h_t + b_{attn}) \tag{6}$$

Here, a_t represents the attention weight for time step t, while W_{attn} and b_{attn} are the weight and bias of the linear layer, respectively, and h_t is the output of the BiLSTM at time step t. Subsequently, we compute the context vector c_t as the sum of the weighted BiLSTM outputs:

$$c = \sum_{t=1}^{T} a_t \cdot h_t \tag{7}$$

The context vector c provides a weighted representation of the sequence, where the contribution of each time step is determined based on its relative importance. This enables the model to utilize more refined and targeted sequence features in subsequent fully connected layers.

5 Experiments And Results

In this section, we detail our experimental procedures and the corresponding results.

5.1 Dataset Collection

We meticulously designed and constructed a data collection experiment specifically targeting through-wall propagation scenarios, aimed at augmenting our understanding and simulation of wireless signal behavior in real-world environments.As depicted in Fig. 7, the transmitter (TX) is strategically placed in an outdoor setting to emulate the reception of wireless signals within a home or office environment.Conversely, the receiver (RX) is positioned indoors, ensuring that the signals it receives must penetrate the walls of the building.

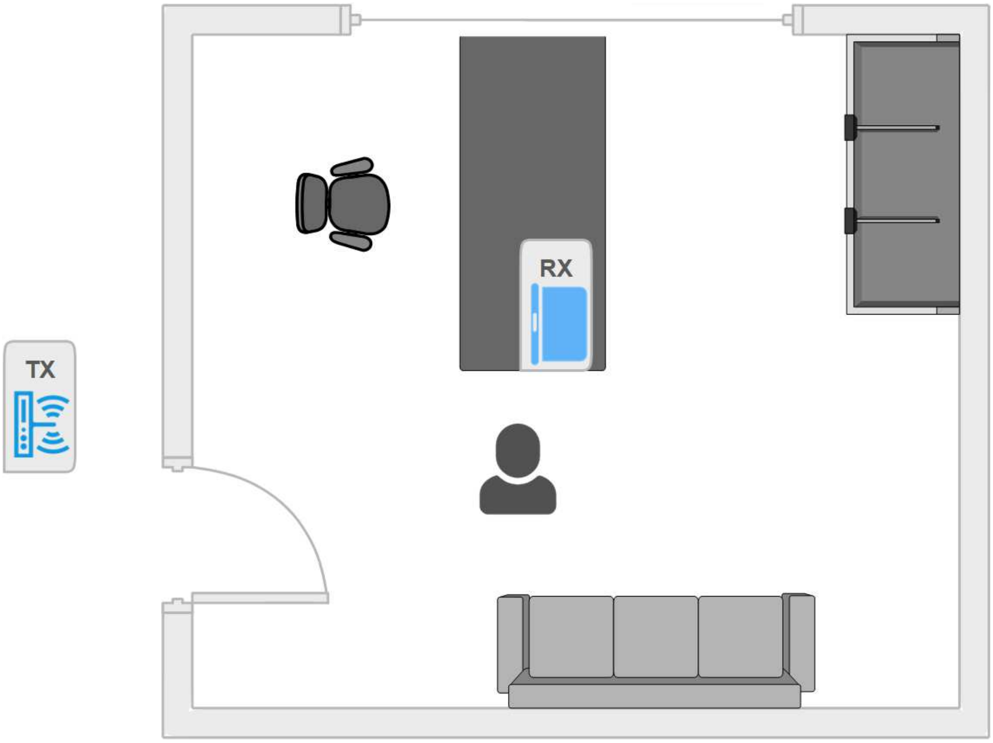
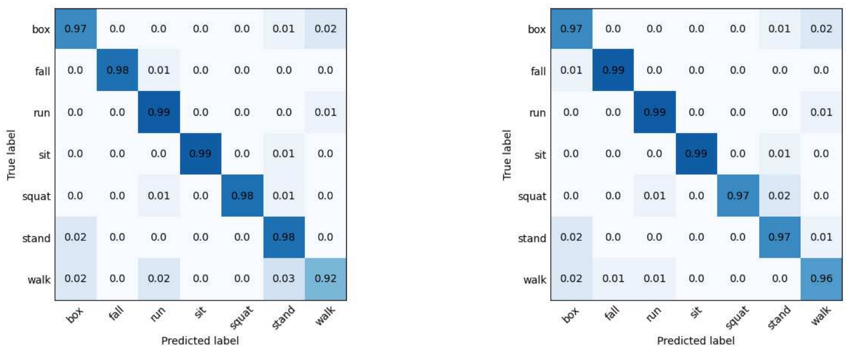


Fig. 7: Dataset collection scenario.

During the experimental phase, the participants engaged in a variety of pre-defined motions within an indoor setting, encompassing fundamental activities like walking, running, and boxing, along with a range of everyday actions. This approach was designed to maximize the coverage of typical human activities within the collected data. Each activity was meticulously labeled and linked with concurrent wireless signal data collection. The primary objective was to furnish a comprehensive training and testing dataset for the development of advanced motion classification algorithms.

Within the experimental setup, particular emphasis was placed on the pathways of wireless signal propagation and potential environmental obstacles. Key parameters, including the direct line distance between the transmitter and receiver, the material properties and thickness of walls, as well as the relative indoor and outdoor positioning, were meticulously documented and analyzed. Furthermore, in order to simulate the propagation of wireless signals in varied residential and occupational settings more accurately, the positions of the subjects were varied in accordance with their movements to assess the impact of diverse factors on the behavior of wireless signals. Each activity was accurately tagged and associated with the simultaneously collected wireless signal data, with the aim of enriching the training and testing datasets for forthcoming motion classification algorithms.



(a) Performance on the test set post-preprocessing

(b) Performance on the test set with original data

Fig. 8: Confusion Matrix for Through-Wall Activity Recognition.The horizontal axis represents predicted labels, while the vertical axis denotes the true labels.

5.2 System Workflow

In our study, as illustrated in Fig. 6, we employed a human activity recognition system based on Channel State Information (CSI) data.The system's workflow encompasses several stages: initially acquiring raw CSI data through collection devices, followed by data preprocessing involving denoising and data augmentation to enhance data quality.Subsequently, the preprocessed data is fed into a deep learning model that integrates a Convolutional Neural Network (CNN) with a Bidirectional Long Short-Term Memory (BiLSTM) network equipped with an attention mechanism, responsible for feature extraction and time series analysis.Ultimately, the model outputs a probability distribution for action classification, achieving precise identification of human activities.

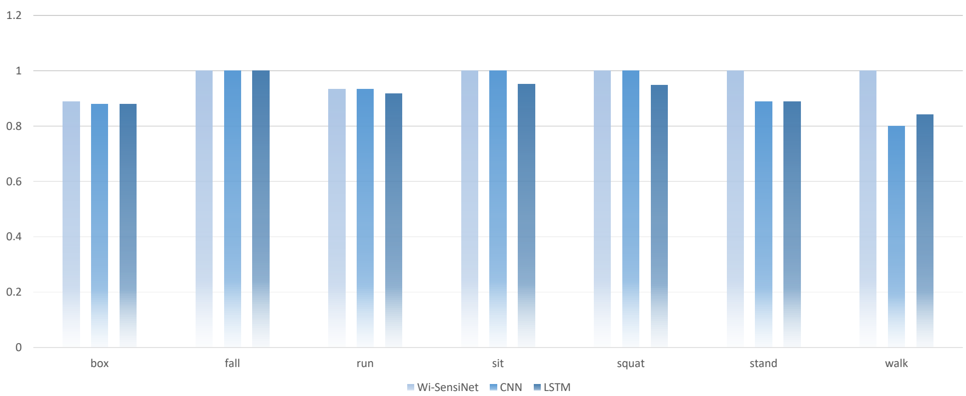


Fig. 9: Accuracy of Three Models in Recognizing Various Actions.

5.3 Experimental Results

In this section, we will present a detailed exposition of the results obtained from our experimental study.The experiment is designed to validate the efficacy of our proposed model in classifying human activities.By comparing the performance of the model under different configurations, we are able to accurately assess the impact of various factors on the classification accuracy. As illustrated in Fig. 8, we conducted a comparative analysis of the model's performance on both preprocessed and original datasets.The results of the confusion matrix clearly demonstrate that the model trained on datasets subjected to noise reduction and data augmentation exhibits superior classification accuracy on the original dataset.Notably, the diagonal elements, representing the percentage of correct classifications, are generally higher in the confusion matrix of the preprocessed dataset compared to the original dataset.This finding validates the effectiveness of our data preprocessing strategy, affirming its significant role in enhancing the model's generalization capability on unseen data. Fig. 9 presents

Table 1: Comparison of wall penetrating HAR using machine learning and recognition techniques.

Name	Method	Activity	Hardware	Accuracy
Wi-SensiNet	CNN-BiLSTM(attention)	Running, Sitting, Standing, Squatting, Falling, Punching, Walking	WiFi devices	99%
TW-See[16]	BP-network	Walking, falling, waving, boxing, standing up, sitting down, empty	WiFi devices	94.46%
WiHACS[17]	SVM(one wall)	Sitting, standing, walking, squatting, falling, lying down, standing up after lying	WiFi devices	92%
RbHAR[18]	Adaptive thresholding	walking, sitting, standing, picking up an object, drinking,falling	Radar sensors	93%

a comparative analysis of the accuracy of three distinct models in recognizing seven different activities.These activities include boxing, falling, running, sitting, squatting, standing, and walking.The three models are Wi-SensiNet, CNN, and LSTM respectively. It is observable that, in most activity recognition tasks, the accuracy of the Wi-SensiNet model slightly surpasses the other two models.Particularly in the recognition of falling activity, the Wi-SensiNet model demonstrates a significant advantage.In contrast, the performance of CNN and LSTM is relatively similar across all activities, though CNN exhibits a slight edge in recognizing squatting and standing actions.This chart significantly highlights the performance disparities among different models in activity recognition tasks and provides a quantitative basis for further discussion and analytical research.

Fig. 10: Accuracy performance of different datasets.

As depicted in Table 1, we conducted a comparative analysis with other literature that utilizes machine learning and recognition techniques for through-wall Human Activity Recognition (HAR).These approaches include systems based on WiFi devices and radar sensors, each targeting the recognition of various daily activities.By contrasting these various methods, we assessed their respective performances, and the comparison has illuminated the advancements made in the field of human activity recognition through different technological approaches, also offering valuable insights for our research.

As depicted in Fig. 10, we present a comparison of the performance of three models: Wi-SensiNet, CNN, and LSTM across different datasets and preprocessing methods. Notably, the Wi-SensiNet model consistently outperforms the other models across all datasets, as indicated by the higher accuracy metric. Additionally, the preprocessing method combining median filtering with Gaussian noise enhancement appears to significantly improve model generalization, as reflected by the increased accuracy rates across datasets when this method is employed. Furthermore, the robustness of the Wi-SensiNet model is evident from its sustained high accuracy irrespective of the dataset partitioning rules applied. This graph eloquently demonstrates the superiority of the Wi-SensiNet model in terms of accuracy and generalization capabilities, as well as its robustness to different data division strategies.

6 Conclusion

In summary, this study developed an innovative human activity recognition system based on Channel State Information (CSI), specifically targeting activity detection under Non-Line-of-Sight (NLOS) conditions. The system employs state-of-the-art signal processing techniques, significantly enhancing the predictive capability for through-wall human activity recognition using WiFi sensing technology. The primary experimental focus was on minimizing the impact of prediction errors through walls on accuracy in indoor environments. Data denoising and enhancement strategies were employed, optimizing the quality of the dataset and boosting the model's adaptability to new environments. Technically, the model amalgamates the spatial feature extraction prowess of Convolutional Neural Networks (CNN) with the temporal analysis proficiency of Bidirectional Long Short-Term Memory (BiLSTM) networks, further augmented by an attention mechanism. The experimental results underscore the high accuracy of this approach in human activity recognition, demonstrating its viability in practical applications.

References

[1] Zhou, Z., Wu, C., Yang, Z., Liu, Y.: Sensorless sensing with wifi. Tsinghua Science and Technology **20**(1), 1–6 (2015)

[2] Yang, J., Zou, H., Zhou, Y., Xie, L.: Learning gestures from wifi: A siamese recurrent convolutional architecture. IEEE Internet of Things Journal **6**(6), 10763–10772 (2019)

[3] Li, C., Liu, M., Cao, Z.: Wihf: Enable user identified gesture recognition with wifi. In: IEEE INFOCOM 2020-IEEE Conference on Computer Communications, pp. 586–595 (2020). IEEE

[4] Ding, X., Jiang, T., Zhong, Y., Wu, S., Yang, J., Xue, W.: Improving wifi-based human activity recognition with adaptive initial state via one-shot learning. In: 2021 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6 (2021). IEEE

[5] Zou, H., Zhou, Y., Yang, J., Jiang, H., Xie, L., Spanos, C.J.: Deepsense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network. In: 2018 IEEE International Conference on Communications (ICC), pp. 1–6 (2018). IEEE

[6] Palipana, S., Rojas, D., Agrawal, P., Pesch, D.: Falldefi: Ubiquitous fall detection using commodity wi-fi devices. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **1**(4), 1–25 (2018)

[7] Ding, J., Wang, Y.: A wifi-based smart home fall detection system using recurrent neural network. IEEE Transactions on Consumer Electronics **66**(4), 308–317 (2020)

[8] Bahl, P., Padmanabhan, V.N.: Radar: An in-building rf-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer

and Communications Societies (Cat. No. 00CH37064), vol. 2, pp. 775–784 (2000). Ieee

[9] Halperin, D., Hu, W., Sheth, A., Wetherall, D.: Tool release: Gathering 802.11 n traces with channel state information. ACM SIGCOMM computer communication review **41**(1), 53–53 (2011)

[10] Xie, Y., Li, Z., Li, M.: Precise power delay profiling with commodity wifi. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp. 53–64 (2015)

[11] Gringoli, F., Schulz, M., Link, J., Hollick, M.: Free your csi: A channel state information extraction platform for modern wi-fi chipsets. In: Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, pp. 21–28 (2019)

[12] Wang, W., Liu, A.X., Shahzad, M., Ling, K., Lu, S.: Device-free human activity recognition using commercial wifi devices. IEEE Journal on Selected Areas in Communications **35**(5), 1118–1131 (2017)

[13] Ma, Y., Zhou, G., Wang, S., Zhao, H., Jung, W.: Signfi: Sign language recognition using wifi. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **2**(1), 1–21 (2018)

[14] Shi, Z., Zhang, J.A., Xu, R., Cheng, Q.: Deep learning networks for human activity recognition with csi correlation feature extraction. In: ICC 2019-2019 IEEE International Conference on Communications (ICC), pp. 1–6 (2019). IEEE

[15] Wang, J., Zhang, L., Gao, Q., Pan, M., Wang, H.: Device-free wireless sensing in complex scenarios using spatial structural information. IEEE Transactions on Wireless Communications **17**(4), 2432–2442 (2018)

[16] Wu, X., Chu, Z., Yang, P., Xiang, C., Zheng, X., Huang, W.: Tw-see: Human activity recognition through the wall with commodity wi-fi devices. IEEE Transactions on Vehicular Technology **68**(1), 306–319 (2018)

[17] Chowdhury, T.Z., Leung, C., Miao, C.Y.: Wihacs: Leveraging wifi for human activity classification using ofdm subcarriers' correlation. In: 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 338–342 (2017). IEEE

[18] Li, Z., Le Kernec, J., Abbasi, Q., Fioranelli, F., Yang, S., Romain, O.: Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms. Scientific Reports **13**(1), 3473 (2023)