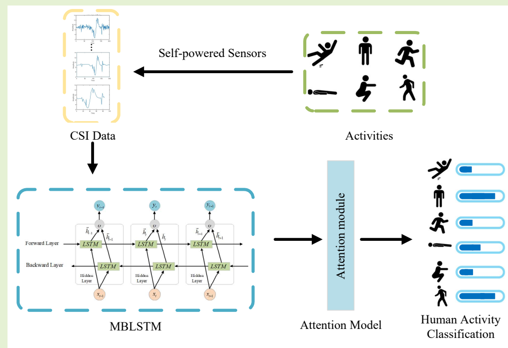
Human Activity Recognition Using Self-powered Sensors Based on Multilayer Bi-directional Long Short-Term Memory Networks

Jian Su, *Member, IEEE,* Zhenlong Liao, Zhengguo Sheng, *Senior Member, IEEE,* Alex X. Liu, *Fellow, IEEE,* Dilbag Singh and Heung-No Lee

***Abstract*—Sensor-based Human Activity Recognition (HAR) requires the acquisition of Channel State Information(CSI) data with time series based on sensors to predict human behavior. Many existing approaches are based on wearable sensors and cameras, which increase the burden and privacy issues for patients. Self-powered sensors are capable of non-contact collection of time series data generated by human activity while ensuring their own stable operation. In this paper, we propose a deep learning-based framework for contactless real-time activity detection of humans using self-powered sensors, which is called Multilayer Bi-directional Long Short-Term Memory (MBLSTM). The collected WIFI CSI data are fed into our proposed network model, which is then used to learn representative features of both sides from the original continuous CSI measurements. And the attention model is used to assign different weights to the learned features, and finally activity recognition is performed. Experimental results show that our proposed method achieves an accuracy of more than 96% for the recognition of six activities in multiple rounds of testing, outperforming other benchmark methods used for comparison.**



***Index Terms*—Self-Powered Sensors, WIFI, Channel State Information, Human Activity Recognition, BLSTM, Deep Learning**

I. Introduction

In recent years, thanks to the rapid development of IoT technology, we can get a lot of useful information from different types of sensors in IoT. This information can help IoT technology to be applied in smart cities, smart farms, medical and health services, etc. The application of IoT sensors in the livestock industry can help practitioners reduce costs and increase efficiency [1]. 3D sound source localization using fiber optic sensors [2]. IoT technologies are also gradually entering our daily life and can identify human daily activities. Human activity recognition is also receiving increasing attention for research in the field of health detection. For example, we need to understand the health status of elderly people and need to monitor their daily activities [3] over time for fall detection [4] and identification of some diseases that the elderly are prone to, such as Parkinson's [5].

To identify various human activities, many methods have been used in previous work. Cameras [6]–[9], wearable sensors [10], [11], and RFID [12]–[14]have been used for activity recognition. Camera-based systems have the advantage of being able to detect minor human movements. However, these systems face severe problems such as object blocking and privacy issues. Because of the great recognition accuracy, wearable sensors are also useful in human activity recognition [15]. Wearable sensor-based ones, on the other hand, need the use of additional devices for action recognition, which is both uncomfortable and ineffective. The mobile phone is another popular sensor for recognizing human activity. Smartphones may be considered electricity sensing platforms for human activity recognition since different sensors, such as accelerometers, gyroscopes, and barometers, are incorporated in phones. If the user forgets to carry their smartphone,

This work was supported in part by the Natural Science Foundation of China under Grant 61802196, 61872082, and 61472184, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180791, in part by the Engineering Research Center of Digital Forensics, Ministry of Education, and in part by the National Research Foundation of Korea (NRF) funded by the Government of Korea (MSIP) under Grant NRF-2021R1A2B5B03002118.

J. Su and Z. Liao are with the School of Computer and Software, Nanjing University of Information Science and Technology, Jiangsu 210044, China (e-mail: sj890718@gmail.com; liaozhenlong1105@163.com)

Z. Sheng is with the Department of Engineering and Design, University of Sussex, Brighton BN1 9RH, U. K. (e-mail: z.sheng@sussex.ac.uk).

A. X. Liu is with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: alexliu@cse.msu.edu)

D. Singh and H. N. Lee are with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, 61005, South Korea (e-mails: dilbagsingh@gist.ac.kr and heungno@gist.ac.kr).

activity recognition will be turned off. Simultaneously, the operation of the sensors in the phone will be affected by its battery capacity. The usage of WiFi devices for human activity recognition has also been successful currently [16]–[19]. WiFi provides new research directions for universal, non-visual human activity recognition due to its universality, low cost, and contactless operation. Use self-powered sensors to obtain stable and continuous WIFI signal information.

The basic point of using WiFi to recognize activities is that human motion influences the nearby WiFi signal, and that WiFi signals reflected by different activities exhibit different characteristics. Received Signal Strength (RSS), which is most widely practiced in the field of indoor positioning research [20], [21], is the most extensively utilized signal for WiFi. Although it can be used to recognize the human activity, it has disadvantages because of noise and unsteady RSS data. Distinguishing different human actions is mainly a matter of analyzing the pattern of the signal, CSI. The most advanced work showed pretty decent recognition accuracy while using a clean WiFi channel in the experiment. However, in the real world, WiFi channels are less than clean. Nowadays, wireless signals abound in indoor places such as homes, offices, and supermarkets, and there are numerous private Access Points (APs). Because most systems now utilize stationary WiFi channels for action recognition and CSI acquisition, their performance is extremely vulnerable to co-channel interference, which can significantly decrease the quality of the receiver and distort the extracted recognition features. When classifying activities, traditional classification models utilized in present systems are highly influenced by such distortions. Recently, with the rapid development of deep learning techniques, the method of automatically learning activity features in CSI using deep learning has provided a completely new way of thinking for human activity recognition. [22]. There is also experience in combining machine learning and sensors in previous work. For example, fiber optic tactile sensors combined with machine learning algorithms for surface roughness recognition [23].

The advantage of long short-term memory (LSTM) networks to automatically learn meaningful features and encode data is widely used in deep learning. The traditional LSTM only handles the forward continuous CSI data, which means that the backward CSI data are not used effectively in training. Future CSI data, we believe, will be important for recognizing human activities. Furthermore, typical LSTM sequence properties may contribute differently to the human activity recognition challenge. The learnt characteristics, on the other hand, make an equivalent contribution to the final identification of human actions in the classic LSTM technique. We provide a multilayer bidirectional LSTM based on WiFi CSI data for human activity recognition in this research paper. Stacking LSTM hidden layers gives more depth to the model and more accurate descriptions obtained as a deep learning technique, while increasing the depth of the network, improving the efficiency of training, and obtaining higher accuracy. An MBLSTM network consisting of multiple forward and backward LSTM layers can handle both forward and backward continuous CSI measurements. Furthermore, the attention mechanism can give more weight to more essential characteristics and time steps, resulting in higher generalization for human activity detection. The effectiveness of the proposed personnel activity detection algorithm based on wireless channel state information measurement is verified by real experiments. The results are compared to several published benchmark approaches.

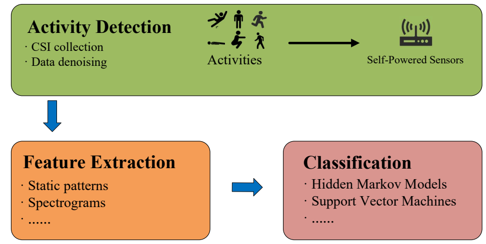


Fig. 1. Basic framework of WiFi-based activity recognition system.

In this paper, the major contribution of our work is that we establish a framework called MBLSTM to recognize human activities. The following is a detailed description:

1) We designed an MBLSTM network to collect WiFi CSI data for autonomous feature extraction and selection using self-powered sensors. Use self-powered sensors to continuously and steadily collect WIFI time series information under different activities, and match this different information with different activities.
2) Continuous CSI data in both forward and reverse directions are processed by layering several Bi-directional Long Short-Term Memory (BLSTM) networks. The MBLSTM can simultaneously consider the information of different past and future actions in CSI data, thus bringing richer information reference for feature learning, and using it can also speed up the convergence process of the training dataset.
3) The MBLSTM network uses an attention model to learn the relevance between activity features and time series. For final personnel activity recognition, more main features and time series are assigned greater weights, resulting in improved recognition performance.

The rest of the paper is organized as follows. Section 2 reviews some state-of-the-art work on using WiFi signals to identify human activities, and Section 3 describes the channel sensing model and the MBLSTM network, as well as the proposed approach. Section 4 describes the experimental setup and data. Then, this section shows and analyzes the experimental results. Finally, Section 5 summarizes this work.

II. RELATED WORK

As illustrated in Fig. 1, a conventional WiFi-based activity recognition system is composed of three parts.

1) *Filtering and monitoring channel status.* Human activity is detected using a self-powered sensor. The human body activity affects the WiFi signal, and this pair of signals can be observed. Therefore, the first step of the activity recognition system is to collect the original signal and denoise it to reveal the changes caused by human activity.

2) *Extraction of Features.* The CSI data from this denoising step is still not directly usable, and the next task is to discover and extract features from the existing behavioral data that are initially compatible with the technical requirements. At present, the signal feature extraction method contains the following three kinds: the time domain analysis method, the frequency domain analysis method ,and the combined method of time-frequency analysis.
3) *Training and Recognition.* After getting the feature dataset, the first operation is to distinguish the dataset into training set and test set, and choosing the proper division ratio is a key part to ensure the effect of behavior recognition.The next step is to select the appropriate classification algorithm to train and test the data.

Due to the common presence of WiFi in everyday life, many research teams have developed several activity recognition systems using WiFi signals. Sigg *et al.* [24] proposed a wireless human activity recognition system which analyzes the RSS information of the interfered WiFi signal for activity recognition. They extracted several important features from RSS data and used a k-Nearest Neighbor (KNN) classifier to recognize four daily activities. Abudulaziz *et al.* [25] designed an RSS-based gesture recognition system on cell phones. The system uses deep learning networks for gesture recognition and achieves high recognition accuracy. Due to multipath and fading effects, the collection of raw RSS data containing actions can be unstable and noisy, so the performance of using RSS to recognize activities with actions is very limited, even for simple actions. WiFi's more steady and informative CSI has received a lot of attention recently. Zhang *et al.* investigated the sensitivity of WiFi signals theoretically and proposed a Fresnel zone method to recognize human activity using WiFi CSI data [26].

Some special features may need to be carefully designed using domain knowledge in order to recognize certain actions using WiFi CSI measurements. When used to recognize other activities, these features may not perform effectively. For example, the traditional KNN method, which has a simple idea, is applicable to multi-classification problems. However, when the sample distribution is unbalanced, the new sample will be classified as the dominant sample, so it cannot better approximate the actual classification result. Furthermore, hand-crafted characteristics will gradually lose several of the implicit qualities that are important for recognizing human activity. Deep learning is an useful tool for automatically learning the differentiating features that are used to recognize human activity.

Deep learning is a type of machine learning method that uses a deep neural network to classify data. In most cases, accurate features need to be identified for input to the training model, and the model classifies and outputs results based on these features. As a result, well-designed features are essential for accurate behavior recognition and have a significant affect on classification accuracy. Some feature extraction, on the other hand, may depend on empirical experience, lowering classification accuracy. Deep learning, unlike machine learning, generally does not require feature extraction stages since a deep neural network is able to automatically identify and extract features from training data. Deep learning allows us a new method to classify data and can deal with enormous amounts of data. In other words, the most significant advantage of deep learning is that it does not require pre-processing of data in order to obtain data features. Meanwhile, deep learning can automatically compute large-scale unknown parameters in neural networks through the training process. Usually, the process of neural network training consumes a lot of practice, but the results achieved are satisfactory. Deep learning algorithms are widely applied in various fields, including picture target identification, natural language processing, video classification, visual arts and so on [27].

Damodaran *et al.* [28] used a device-free approach (CSI) to identify human activities. Wavelet analysis was used for preprocessing and feature extraction. As a result, they were able to recognize walking, sitting, standing, and running activities. High-bandwidth noise was removed using principal component analysis by Moshiri *et al.* [29]. The signal was transformed to the frequency domain using Short Time Fourier Transform (STFT) and new data was generated using Generative Adversarial Networks (GAN). The LSTM algorithm was used for classification. The accuracy was 87.2% using 50% of the ”real” data plus 50% of the synthetic data, and 92.8% using a set of all ”real” data.

Convolutional Neural Network (CNN) is a very popular deep learning method that automates feature extraction and can easily handle high-latitude data. However, when the network level is too deep, modifying the parameters using BP propagation will cause the parameters near the input layer to change more slowly, and the pooling layer will lose a lot of valuable information and ignore the local-to-whole correlation.

Since for different activities, CSI measurements are continuous measurements with temporal information, BLSTM capable of encoding temporal information is good candidate for automatic feature learning. BLSTM includes both forward and backward processes of feature learning. As a result, when evaluating the current hidden state of the LSTM, BLSTM can take into account both past and future information, resulting in richer information features. We propose stacked multilayer BLSTM networks for human action recognition. Each layer of the BLSTM neural network automatically learns the input action features and passes the learned features to the next layer. At the same time, the feature sequences learned in one temporal instance may contribute differently to the final human activity recognition. Furthermore, the significance of channel state information collected at different time stages may differ. Therefore, in order to assign different weights to different action features in the training for the purpose of reducing the training time and improving the accuracy of the model, we add an attention mechanism to the proposed network model.

III. PROPOSED METHOD

A. System Overview

The proposed MBLSTM framework is shown in Fig. 2. First, we use a router and a self-powered sensor to collect CSI signals from WiFi of human actions. Secondly, we input

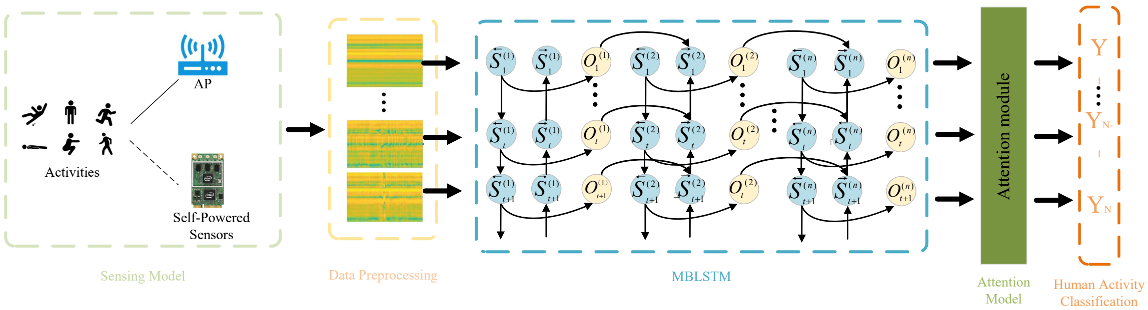


Fig. 2. The proposed MBLSTM framework for CSI based human activity recognition.

the processed CSI signals into the MBLSTM framework to automatically learn the forward and backward features. There are 200 hidden nodes in the bidirectional LSTM used for feature learning in this experiment. Since the attention model has no available prior information, it can only use the features learned from BLSTM as input to derive an attention matrix representing the importance of features and time steps. Then, we use element multiplication to merge the learned features with the attention matrix to obtain the modified feature matrix with attention. After that, the feature matrix is flattened into feature vectors for final classification using the flattened layer. Finally, the softmax classification layer is used to identify different activities with the final feature vectors.

B. Channel Sensing Model

WiFi signals are known to fluctuate significantly when objects move within the region of interest. The Fresnel zone model is introduced as a result of this to explore how the WiFi signals on these receiving antennas change as a result of different activities. Furthermore, we infer potential behavioral information from such activity-induced signal fluctuations. So we use the Intel 5300 NIC, a self-powered sensor, to collect the reflected WiFi information.

In recent years, the Fresnel zone model has been applied to the research of human action recognition based on wireless sensing. It refers to the wireless electromagnetic wave in the transmission process, the formation of the transceiver at both ends of the transceiver device, the location of the transceiver device as the focus of the ellipse-shaped area, the area is the wireless electromagnetic wave intensity concentration area.One of the most important zones is the first Fresnel zone, where most of the energy of the wireless signal is located. If there is an obstacle in this region, it will affect the wireless signal. The wireless signal will form multiple propagation paths from the receiver (Rx) to the transmitter (Tx), and the direct propagation path that passes through both the transmitter and receiver is called the line-of-sight path (LoS). When the wireless signal propagation encounters obstacles due to reflection, scattering, and diffraction generated by the propagation path is called non-Line-of-Sight (NLoS).

Through the analysis and study of the received signals, the researchers found the characteristics of the changes brought by the human body movements on the signal propagation. And established the relationship between these features and the mapping of different activities, which built the foundation for WiFi-based human activity recognition.

The phenomenon that different actions have different effects on WiFi signals is a major discovery that the Fresnel zone model can be applied to the field of action recognition. Specifically, different activities cause significant differences in the speed of signal dynamic paths. Furthermore, CSI's amplitude attenuation and phase change can capture these specific pattern. It demonstrates the feasibility and application of using unique CSI variations to effectively and precisely identify and recognize different human activities.

C. MBLSTM Neural Network

In the case of multilayer stacking, each layer of the BLSTM neural network is composed of a forward recurrent network and a backward recurrent network. The combination of the output results of the forward LSTM and the backward LSTM of the previous layer is sent to the next layer of the network. Fig. 3 illustrates the MBLSTM network framework structure.

$$\begin{aligned}
S_t'^{(1)} &= f\left(U'^{(1)} x_t + W^{(1)} S_{t-1}'\right)\\
S_t^{(1)} &= f\left(U^{(1)} x_t + W^{(1)} S_{t-1}\right)\\
&\bullet\bullet\quad\bullet\\
S_t'^{(i)} &= f\left(U'^{(i)} S_t^{(i-1)} + W'^{(i)} S_{t+1}'\right)\\
S_t^{(i)} &= f\left(U^{(i)} S_t^{(i-1)} + W^{(i)} S_{t-1}\right)\\
\mathrm{O}_t &= g\left(V^{(i)} S_t^{(i)} + V'^{(i)} S_t'^{(i)}\right)
\end{aligned} \tag{1}$$

The output is determined by the sum of each layer's positive and negative computations. Where $S_{t-1}^{(i)}$ and $S_t^{(i)}$ are the values of the i-th hidden layer at time $t-1$ and t, respectively.Forward and backward computations do not share weights, $V^{(i)}$, $U^{(i)}$ and $W^{(i)}$ are the weight matrices of the i-th hidden layer to the output layer, the input layer to the hidden layer , and the hidden layer . $V'^{(i)}$,$U'^{(i)}$and $W'^{(i)}$ are the backward weight matrices used for the computations. And i is the number of BLSTM layers, and $i = 0, 1, 2 \ldots \infty$ is the output layer's value.

D. Attention Model

The attentional model is developed to be used for image recognition [30]. The concept was inspired by the human visual system, which says that during picture recognition,

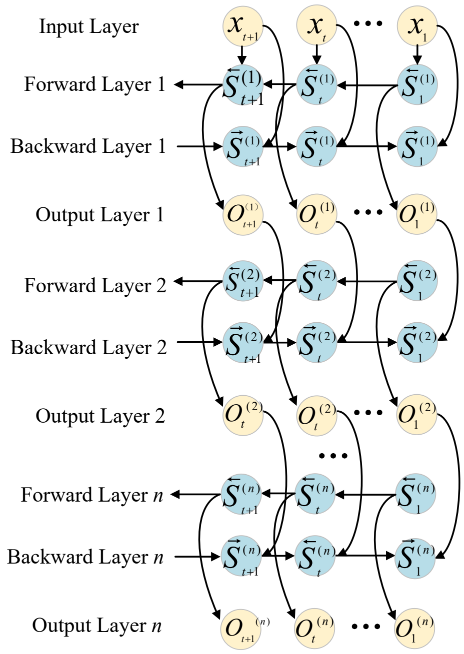


Fig. 3. Structure of MBLSTM neural network.

humans always focus on a certain portion of the image and adjust their attention over time. During the recognition work, the attention model allows the computer to attend to the area of interest while blurring other areas. Recently, the attention models have been used in language processing, proving that it is clearly effective [31]. For example, in the popular encoder-decoder method for natural language processing, the input sentence is encoded as a fixed vector that is translated throughout the translation process, meaning that at every time step, all words in the input sentence contribute equally to the translation.This task of processing sentence translation is inefficient. When the encoder model is utilized with the attention model, translations will focus more on the words that are more relevant to the current translation process at different time steps. Since the MBLSTM network learns high-dimensional sequence features, individual features and time - series may contribute differently to the final recognition results. We try to use an attention model to intelligently learn the effects of different actions of features and assign weights according to their importance.

In the recognition system, there is no usable a priori information for training. As a result, the attention model, also known as self-attention, will utilize the sequential features learned by MBLSTM as input. This attention model is shown in a simple example here. Given n feature vectors $\mathbf{h}_i, i = 1, 2, \ldots, n$ that can be obtained from the feature learning network. We build a score function $\Phi(\bullet)$ to evaluate the significance of each feature vector by computing the score s_i as follows:

$$s_i = \Phi\left(\mathbf{W}^T \mathbf{h}_i + b\right) \tag{2}$$

Where $\mathbf{W}^T$ and b are the weight vector and bias,respectively. Any activation function in a neural network, such as tanh, relu, or linear, can be used to build the score function. We can normalize each feature vector's score utilising softmax function, that is written as:

$$a_i = softmax(s_i) = \frac{esp(s_i)}{\sum_i(s_i)} \tag{3}$$

The final output feature $\mathbf{O}$ of the attention model is the product of the vector and its normalization score, as follows:

$$\mathbf{O} = \sum_{i=1}^{n} a_i * \mathbf{h}_i \tag{4}$$

E. Training Proposed Method

To identify the model parameters, the proposed MBLSTM framework is trained using CSI data with real labels. At first, all parameters are are randomly given. The CSI data is then sent into MBLSTM, which uses it to predict the labels. The category cross-entropy errors are measured and back-propagated using a gradient-based optimization approach to adjust the model parameters utilizing the given true labels. We utilize ADAM [32] to calculate the adaptive learning rate for each parameter in the optimization process efficiently.

In learning-based systems, overfitting is a typical problem. To avoid overfitting, we utilize the ADAM optimizer. It provides adaptive learning rates for various parameters. Furthermore, the suggested attention method will only choose a few significant features and time series, decreasing the possibility of overfitting.

IV. EXPERIMENTS AND ANALYSIS

In this section, we first introduce our experimental settings in detail and then present the extensive experimental results that validate the effectiveness of our model.

A. Experiments Settings

We compared the proposed method to several benchmark CSI-based human activity identification algorithms to evaluate how effective it is. According to [33], the Random Forests (RF) model outperforms Support Vector Machines (SVM), Logistic Regression (LR), and Decision Tree (DT) in WiFi-based human activity recognition. In [34], Hidden Markov Models (HMM) have also been found to be useful for recognizing human activity. As a result, we compared our method to these two handmade methods. Manual feature extraction is described in detail in [33]. We also compare it to other deep learning-based approaches that can learn features automatically, such as Sparse Autoencoder (SAE) [34], [35] and traditional LSTM [33]. The SAE algorithm is an unsupervised algorithm that automatically learns features from unlabeled data and can give a better feature description than the original data. Validation sets from the training examples were used to fine-tune the parameters of all methods. For evaluation, 10-fold cross-validation has been used. We divided all of the data into 10-folds at random. Then, we select one fold of data for testing and the rest for training and finally get 10 times. The average of all 10 runs determines the final recognition accuracy. The dataset used for comparison was taken by the authors in [33] from an office. A router was used as a transmitter and a

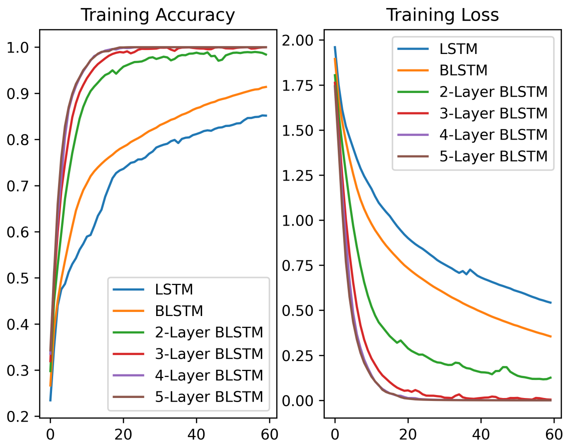


Fig. 4. Trend of accuracy rate of different network training.

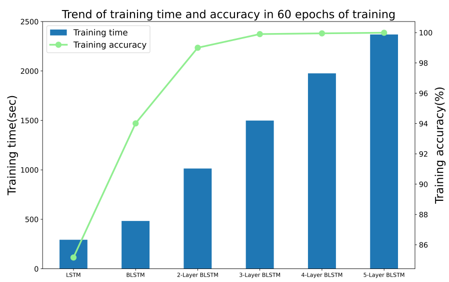


Fig. 5. Trend of training time and accuracy in 60 epochs of training.

laptop with an Intel 5300 NIC was used as a receiver. The sampling frequency was 1 kHz, with three antennas and 30 subcarriers, and the size of the original CSI data was 90. The window size used for data segmentation was a sliding window of 2s. Transmitters and receivers were separated by three meters under line-of-sight (LOS) conditions. Each person performed each activity for 20 seconds during data collection. Note that the person remains stationary at the beginning and end of the activity. Six persons were involved in the data process of collecting, which included six normal daily activities: Lie down, Fall, Run, Sit down, Stand up and Walk. Every volunteer performed 20 rounds of each activity, the resulting dataset was approximately 17 GB in size. All experiments were performed on a workstation in our lab, using python to run the code. The workstation is equipped with an 8-core, 16-thread Intel i9-9900 CPU and an NVIDIA GeForce RTX 2080 GPU.

We compare the trend of accuracy and loss of BLSTM networks with the different number of layers in the training dataset. Fig. 4 shows that the LSTM and BLSTM networks converge more slowly, with accuracy barely reaching 90% at the 60th round of training. The multilayer BLSTM network, on the other hand, converges quickly, with accuracy exceeding 90% at about 10 rounds of training, approaching 100% at close to 20 rounds, and preserving stability in accuracy during subsequent training.

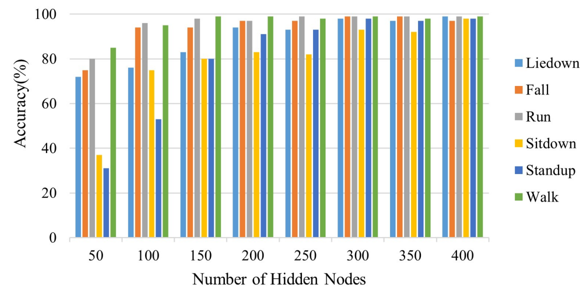


Fig. 6. Recognition accuracy of each activity with different number of hidden nodes.

TABLE I
TRAINING AND VALIDATION TIME FOR DIFFERENT NUMBER OF HIDDEN NODES

Nodes / Time	50	100	150	200	250	300	350	400
Training(s)	245.85	349.14	545.11	785.69	880.83	1043.76	1238.80	1444.24
Validation(s)	7.63	7.46	7.04	7.64	7.21	7.24	7.86	8.35

TABLE II
THE TRAINING AND TESTING TIME FOR DIFFERENT METHODS

Time	RF	HMM	SAE	LSTM	MBLSTM
Training(s)	5.31	0.024	158.16	493.23	1551.14
Testing(s)	0.008	0.17	0.19	3.54	8.72

Although the training converges faster as the number of BLSTM layers increases, it is not better to have more layers. As the number of layers increases, the network structure becomes increasingly large, which means that more and more computational resources will be used, and more time will be consumed in training. As shown in Fig. 5, we run experiments using 200 hidden nodes. The results show that the more complex the network structure is, the time for training increases significantly. It is obvious that, with the same number of training rounds, the overall training accuracy does not improve much after increasing the BLSTM network to 3 layers, which are close to 100%, indicating that the limit has been approached. However, the training time spent by each network differs greatly. Considering all factors, we choose the 3-layer BLSTM network as the network model for this experiment in order to minimize the computer resources consumed while ensuring high accuracy.

Impact of the Number of Hidden Nodes: We find that the number of LSTM hidden nodes has a large impact on the experimental results. As a result, we performed a second experiment to see how this parameter affected the accuracy of activity recognition. The results of the experiment are shown in Fig. 6. When using 50 hidden nodes, the recognition accuracy was low for actions, especially for the two activities "Sit down" and "Stand up", which we guess are too similar. When the number of hidden nodes is raised, the recognition performance of each activity is improved, and after the number reaches 300, the accuracy tends to be stable. As shown in Table. I, we use a 3-layer BLSTM network, and in the same 30 rounds of training, the more hidden nodes, the longer the training time, and we choose to use 200 hidden nodes in the MBLSTM.

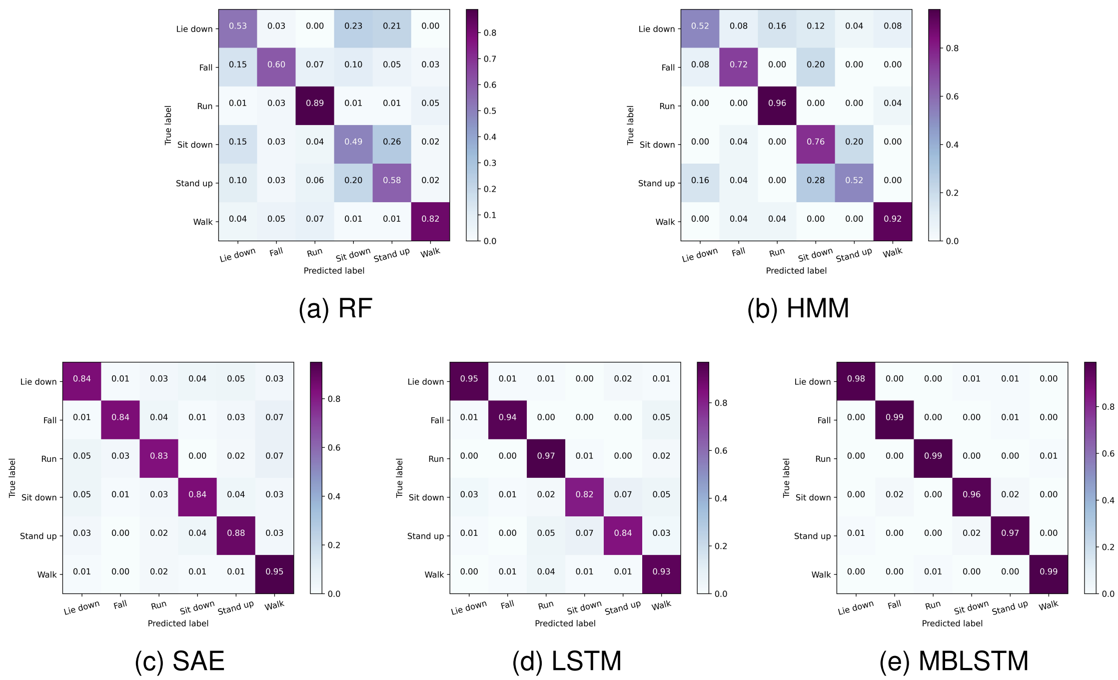


Fig. 7. The confusion matrix of all benchmarks and proposed MBLSTM methods on the dataset.

Time Complexity: Deep learning-based approaches' time complexity is a common issue. We compared the training time and testing time of some methods using the same dataset. The Table. II shows the training time and testing time for all methods. It can be clearly seen that algorithms using deep learning methods have much longer training times than a typical machine learning algorithm. The proposed MBLSTM consumes the longest training time of all methods using deep learning. All of the approaches have short testing times, according to the Table. II. The proposed MBLSTM, for example, has a testing time of 8.72 seconds for 420 test samples. This signifies that each sample will be tested for 0.0208 seconds. The window size for data segmentation is 4 seconds for each case. We believe that our proposed MBLSTM approach, which is based on WiFi CSI, may be utilized for real-time personnel activity recognition.

B. Experimental Results

Fig. 7 shows the confusion matrix of all benchmarks and proposed MBLSTM methods on the dataset. Activity recognition algorithms that need manual feature extraction, such as RF and HMM, perform the poorest. The HMM algorithm performs significantly better than the RF algorithm. Unlike RF and HMM manual feature extraction, SAE algorithms using deep learning methods have better performance. This demonstrates the effectiveness of using the SAE method for automatic feature learning. The LSTM network outperforms the SAE method because it incorporates the temporal factors in the CSI sequences into feature learning. Due to the inclusion of the attention model and the structure of the multilayer bidirectional LSTM in our proposed method, our MBLSTM method achieves excellent recognition results in recognizing six daily activities. Accuracy of 96% and above for all six daily activities recognition, which is sufficient for most recognition situations.

The accuracy of recognition varies greatly depending on the activity. Higher physical activities, such as "Fall", "Walk" and "Run" show greater recognition performance. This is due to the fact that these activities have a large impact on the features of the collected CSI data. It is also evident that most methods have relatively low accuracy for recognizing the activity of "Sit down". This might be because this activity has the same effect on CSI features as the "Lie down" and "Stand up" activities. It's worth noting that the RF method's recognition accuracy with hand-made features is much lower than 50%. The "Fall" activity is the most important of these six, especially for the elderly [36]. The proposed MBLSTM approach can recognize "Fall" activities with 99% accuracy, which will be useful in a wide variety of medical applications. The extended training period of the deep learning-based approach is one of its drawbacks. However, this time-consuming procedure only has to be completed once. It's worth noting that deep learning-based methods can be tested online quickly enough for most real-time applications.

V. CONCLUSION

In this paper, we use self-powered sensors to collect WiFi time series information and propose a multilayer BLSTM network for extracting WiFi signal feature information used for human activity recognition by improving the traditional LSTM model. In both directions, the BLSTM network can learn important sequential features from original WiFi CSI data. The multilayer BLSTM network can enhance the accuracy by accelerating convergence during training. We evaluated the method in real environments and compared it to a variety of benchmark methods, such as , Random Forest , Hidden

Markov Models, sparse autoencoders and traditional LSTM. The proposed MBLSTM for WiFi CSI-based personnel activity recognition has demonstrated higher performance in experiments. Although our method has a high recognition rate for single-person activities, there is still a big room for improvement in multi-person activities. For future work, we hope to improve the accuracy of multi-person activity recognition and the compatibility of the system with different environments. In [37], the authors used an inertial measurement device to calculate acceleration. This inspired our proposed method helps to recognize the type of body movement. In case of a car accident, it can help to determine the posture of the injured person.

Abbreviation List

B-LSTM	Bi-directional Long Short-Term Memory
MBLSTM	Multilayer Bi-directional Long Short-Term Memory
CSI	Channel State Information
HAR	Human Activity Recognition
RSS	Received Signal Strengt
APs	Access Points
KNN	k-Nearest Neighbor
STFT	Short Time Fourier Transform
GAN	Generative Adversarial Networks
CNN	Convolutional Neural Network
RF	Random Forest
SVM	Support Vector Machine
LR	Logistic Regression
DT	Decision Tree
HMM	Hidden Markov Models
SAE	Sparse Autoencode

References

[1] S. Neethirajan, "The role of sensors, big data and machine learning in modern animal farming," *Sensing and Bio-Sensing Research*, vol. 29, p. 100367, 07 2020.

[2] S. E. Hayber and S. Keser, "3d sound source localization with fiber optic sensor array based on genetic algorithm," *Optical Fiber Technology*, vol. 57, p. 102229, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1068520020302194

[3] T. Huynh-The, C.-H. Hua, N. A. Tu, and D.-S. Kim, "Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1533–1543, 2021.

[4] Y. Wang, S. Yang, F. Li, Y. Wu, and Y. Wang, "Fallviewer: A fine-grained indoor fall detection system with ubiquitous wi-fi devices," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12 455–12 466, 2021.

[5] X. Cao, Y. Xue, J. Chen, X. Chen, Y. Ma, C. Hu, H. Ma, and H. Ma, "Video based shuffling step detection for parkinsonian patients using 3d convolution," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 641–649, 2021.

[6] J. Aggarwal and M. Ryoo, "Human activity analysis: A review," vol. 43, no. 3 https://doi.org/10.1145/1922649.1922653

[7] S. Jiang, Y. Qi, H. Zhang, Z. Bai, X. Lu, and P. Wang, "D3d: Dual 3-d convolutional network for real-time action recognition," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4584–4593, 2021.

[8] D. C. Luvizon, D. Picard, and H. Tabia, "Multi-task deep learning for real-time 3d human pose estimation and action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2752–2764, 2021.

[9] M. Moniruzzaman, Z. Yin, Z. He, R. Qin, and M. C. Leu, "Human action recognition by discriminative feature pooling and video segment attention model," *IEEE Transactions on Multimedia*, vol. 24, pp. 689–701, 2022.

[10] L. Tong, H. Ma, Q. Lin, J. He, and L. Peng, "A novel deep learning bi-gru-i model for real-time human activity recognition using inertial sensors," *IEEE Sensors Journal*, pp. 1–1, 2022.

[11] D. Chen, S. Yongchareon, E. M.-K. Lai, J. Yu, and Q. Z. Sheng, "Hybrid fuzzy c-means cpd-based segmentation for improving sensor-based multiresident activity recognition," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 193–11 207, 2021.

[12] X. Fan, W. Gong, and J. Liu, "Tagfree activity identification with rfids," vol. 2, no. 1 Article 7, 2018. [Online]. Available: https://doi.org/10.1145/3191739

[13] G. A. Oguntala, Y. F. Hu, A. A. S. Alabdullah, R. A. Abd-Alhameed, M. Ali, and D. K. Luong, "Passive rfid module with lstm recurrent neural network activity classification algorithm for ambient-assisted living," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10 953–10 962, 2021.

[14] F. Wang, J. Liu, and W. Gong, "Multi-adversarial in-car activity recognition using rfids," *IEEE Transactions on Mobile Computing*, vol. 20, no. 6, pp. 2224–2237, 2021.

[15] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys &Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[16] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-lstm for human activity recognition using wifi signal," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4628–4641, 2021.

[17] C. Xiao, Y. Lei, Y. Ma, F. Zhou, and Z. Qin, "Deepseg: Deep-learning-based activity segmentation framework for activity recognition using wifi," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5669–5681, 2021.

[18] R. H. Venkatnarayan, S. Mahmood, and M. Shahzad, "Wifi based multi-user gesture recognition," *IEEE Transactions on Mobile Computing*, vol. 20, no. 3, pp. 1242–1256, 2021.

[19] J. Zuo, X. Zhu, Y. Peng, Z. Zhao, X. Wei, and X. Wang, "A new method of posture recognition based on wifi signal," *IEEE Communications Letters*, vol. 25, no. 8, pp. 2564–2568, 2021.

[20] C.-H. Lin, L.-H. Chen, H.-K. Wu, M.-H. Jin, G.-H. Chen, J. L. Garcia Gomez, and C.-F. Chou, "An indoor positioning algorithm based on fingerprint and mobility prediction in rss fluctuation-prone wlans," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 5, pp. 2926–2936, 2021.

[21] D. Yu and C. Li, "An accurate wifi indoor positioning algorithm for complex pedestrian environments," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 440–24 452, 2021.

[22] Z. Wang, K. Jiang, Y. Hou, W. Dou, C. Zhang, Z. Huang, and Y. Guo, "A survey on human behavior recognition using channel state information," *IEEE Access*, vol. 7, pp. 155 986–156 024, 2019.

[23] S. Keser and Ş. E. Hayber, "Fiber optic tactile sensor for surface roughness recognition by machine learning algorithms," *Sensors and Actuators A: Physical*, vol. 332, p. 113071, 2021.

[24] S. Sigg, S. Shi, F. Buesching, Y. Ji, and L. Wolf, "Leveraging rf-channel fluctuation for activity recognition: Active and passive systems, continuous and rssi-based signal features," in *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*, ser. MoMM '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 43–52. [Online]. Available: https://doi.org/10.1145/2536853.2536873

[25] M. A. A. Haseeb and R. Parasuraman, "Wisture: Rnn-based learning of wireless signals for gesture recognition in unmodified smartphones," 2017. [Online]. Available: https://arxiv.org/abs/1707.08569

[26] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with wi-fi signals," *Computer*, vol. 50, no. 1, pp. 48–57, 2017.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[28] N. Damodaran, E. Haruni, M. Kokhkharova, and J. Schäfer, "Device free human activity and fall recognition using wifi channel state information (csi)," *CCF Transactions on Pervasive Computing and Interaction*, vol. 2, no. 1, pp. 1–17, 2020. [Online]. Available: https://doi.org/10.1007/s42486-020-00027-1

[29] P. Fard Moshiri, H. Navidan, R. Shahbazian, S. A. Ghorashi, and D. Windridge, "Using gan to enhance the accuracy of indoor human activity recognition," 04 2020.

[30] J. Ji, C. Xu, X. Zhang, B. Wang, and X. Song, "Spatio-temporal memory attention for image captioning," *IEEE Transactions on Image Processing*, vol. 29, pp. 7615–7628, 2020.

[31] A. Galassi, M. Lippi, and P. Torroni, "Attention in natural language processing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 10, pp. 4291–4308, 2021.

[32] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 2014.

[33] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[34] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial wifi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[35] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "Csi-based device-free wireless localization and activity recognition using radio image features," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 346–10 356, 2017.

[36] C. Wang, W. Lu, M. R. Narayanan, D. C. W. Chang, S. R. Lord, S. J. Redmond, and N. H. Lovell, "Low-power fall detector using triaxial accelerometry and barometric pressure sensing," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2302–2311, 2016.

[37] S. Glowinski, S. Majdanik, A. Glowinska, and E. Majdanik, "Trauma in a shaken infant? a case study," *Aggression and Violent Behavior*, vol. 56, p. 101515, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1359178920302196