Device-Free Single-User Activity Recognition Using Diversified Deep Ensemble Learning

Wei Cui[1], $BingLi^2$, $LeZhang^1$, $ZhenghuaChen^1$
[1] I2R, Singapore
[2] The University of New South Wales, Australia

Abstract

WiFi-based human activity recognition (HAR) aims to recognize human activities in an off-the-shelf manner that only relies on the commercial Wi-Fi devices already installed in environments. The recent trend in HAR research is to train classifiers on top of statistical or deep neural features extracted from channel state information (CSI) data. Unfortunately, existing methods only take into account the temporal-correlation within each CSI subcarrier, while ignoring the spatial-correlation between different subcarriers. This issue has not been fully exploited yet, resulting a limited performance. To address this issue, we propose WiAReS, a WiFi-based device-free activity recognition system that takes both temporal-correlation and spatial-correlation into account. WiAReS embarks on diversified deep ensemble methods for single-user activity recognition where one user performs a single activity at a given time. More specifically, it adopts convolutional neural network (CNN) to automatically extract features from CSI measurements with the preservation of the locality of both spatial patterns and temporal patterns. To further improve recognition accuracy upon CNN-extracted features, we propose a novel ensemble architecture that fuses a multiple layer perception (MLP), a random forest (RF) and a support vector machine (SVM). Our system obtains the CSI data in PHY layer of off-the-shelf WiFi devices by installing Atheros-CSI-Tool on AR9590 based WiFi network interface cards (NICs). Comprehensive experiments have been conducted in three real environments with environmental variation to evaluate the performance of the proposed WiAReS. The experimental results demonstrate that the proposed WiARes system significantly outperforms existing methods.

Preprint submitted to . *October 13, 2021*

Keywords: WiFi, channel state information, device-free, activity recognition, random forest

1. Introduction

Human activity recognition (HAR) is a key component for a number of applications such as smart homes, security monitoring, fall detection for elderly people, search-and-rescue systems, and many other Internet of Things (IoT)-based applications [1].

The importance of HAR has led to many research efforts in recent years. Traditional HAR approaches rely on *dedicated devices* such as cameras [2], wearable sensors [3] and radars [4]. However, they all suffer from high device cost and low efficacy, since their abilities are limited by the prerequisites of deployment. Specifically, the camera-based methods requires line-of-sight (LOS) condition, wearable sensors need to be carried all the time, and radar-based methods have limited sensing range due to the short wavelength. The requirements for both devices and their deployments limit their sensing ability and hinder their further applications in daily-life scenes.

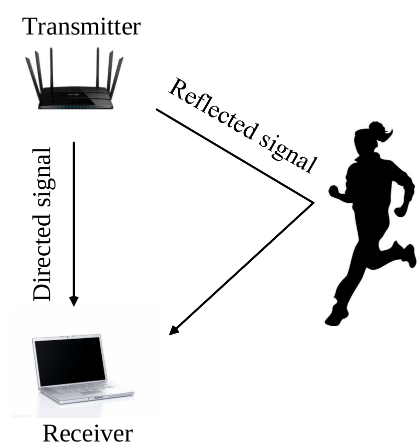


Figure 1: Multi-path environment with human activity.

Observing the disadvantages of above approaches, recently, WiFi, as an ubiquitously available devices, has been introduced to perform *device-free* human activity recognition [5]. The assumption on which WiFi-based solution relies is that a certain human activity could incur unique variations in

WiFi signals, as shown in Fig. 1. WiFi-based solutions use commercial off-the-shelf (COTS) Wi-Fi devices that have already installed in environments to recognize activitiess [6, 7, 8, 9, 10]. These methods regard human activity recognition as a classification problem that trains classifiers (e.g., SVM and MLP) on top of statistical features from WiFi signals. There are many advantages of WiFi-based solutions. They does not require LOS path, has better coverage, without carrying any sensors or devices, and could better protect users' privacy. Due to their ubiquitous availability and easy deployment, WiFi-based solutions become a popular choice for human activity recognition applications.

Commonly, there are two types of WiFi signal information that are employed in WiFi-based HAR systems: received signal strength (RSS) and channel state information (CSI). The RSS is a scalar value indicating the received signal strength, whereas the CSI is the multiple-channels information directly derived from the PHY layer. Compared with the RSS, the CSI contains more environment information, which could implicitly benefits the recognition of activities. Thus, the challenging issue is how to effectively exploit and refine the useful information from the plentiful CSI data to facilitate performance improvement in HAR system. Recently, pioneer research efforts have sought for automatically generating meaningful features from raw data with the advance of deep learning (DL) techniques. For example, Yousefi et al. [11] and Chen et al. [12] adopted long short-term memory (LSTM), an RNN-family neural network, to recognize activities, and achieved state-of-the-art results on many datasets.

Despite their usefulness, current DL-based methods share a common drawback. They only take into account the temporal-correlation within each CSI subcarrier, while ignoring the spatial-correlation between different subcarriers. In facts, due to the physical structure of CSI channels, correlation exists between adjacent subcarriers, indicating that a human activity would incur coupling effects among adjacent subcarriers. Ignoring the spatial-correlation will make HAR system not be able to fully exploit the implicit spatial patterns implied in multi-subcarriers CSI data, and as a consequence, weaken their recognition ability.

In this paper, we propose a novel method that takes both temporal-correlation and spatial-correlation into account. Our model adopt convolutional neural network (CNN) to automatically extract CSI features. The proposed methods take advantages of both highly discriminative features and efficient yet diversified classifiers. CNN has been proven successful in extracting features from data that presents locality of spatial patterns. Thus it is a perfect choice for representing spatial and temporal patterns, both computationally and semantically. On top of CNN-extracted features, we use a novel ensemble architecture that fuses a MLP, a random forest (RF) and a support vector machine (SVM) to achieve higher recognition accuracy. Moreover, we have built a prototype WiFi-based human activity recognition system (WiAReS) that applies our proposed HAR method on multiple subcarriers with Athero NIC. With this prototype system, we demonstrate that our method is practically better in recognizing human activities.

The core contributions of this paper are summarized as follows.

1. We propose a novel ensemble model for human activity recognition that takes both spatial and temporal patterns into consideration. This model uses CNNs to extract comprehensive features from raw CSI data in terms of both temporal-correlation and spatial-correlation, and combines triple classifiers to improve the accuracy in activity recognition.
2. We develop a prototype HAR system WiARes. Our WiAReS was built upon Atheros devices that have 56 subcarriers of CSI measurements for each TX-RX antenna pair. This push the accuracy even further to achieve new state-of-the-art results on many benchmark datasets.
3. We conduct experiments with hardware testbed in two clutter indoor laboratories, and evaluate the proposed schemes extensively. The results demonstrate that our system significantly outperform existing systems.

The rest of the paper is structured as follows. Section 2 reviews the related work on device-free wireless activity recognition. Section 3 introduces the CSI, followed by the motivation behind the proposed schemes. Section 4 presents the proposed WiAReS using CSI features. The experimental evaluations are demonstrated in Section 5. Section 5 gives the conclusion remarks.

2. Related work

Due to ubiquitous availability in indoor areas, WiFi based human activity recognition has gained tremendous attention in recent years. Sigg et al. [13, 14] explored methods to realize activity recognition using the RSS of wireless signals, which relies on the fluctuations of the received signal strength affected by human behavior. WiGest [6] leveraged the effect of hand motion on received wireless signal strength to recognize hand gesture with off-the-shelf WiFi devices. The proposed system starts with a primitives extraction step using the Discrete Wavelet Transform (DWT) to reduce noise, followed by a gesture identification step and a action mapping step. However, as the RSS is highly unstable even when there is no dynamic changes in indoor areas, the RSS based systems have limited performance for activity recognition.

Recently, in some WiFi network interface cards (NICs), the fine-grained PHY layer CSI can be collected using tools provided in [15, 16]. Since the CSI is more robust to environment changes, researchers have developed a number of CSI-based human activity recognition applications. [7, 17] and [18] built theoretical models to quantify the relationships between CSI changes and human mobility without prior training. However, a large number of access points (APs) are required to accurately track the user's motions, making them impractical in some applications.

In contrast, [19, 20] adopt the amplitude and phase information of CSI measurements on multiple channels to detect human activities by employing learning techniques. [10] presented an automatic fall detection system WiFall for elders' healthcare by employing the CSI as the indicator of human activities. WiFall detects a fall activity using commercial NICs with the following two steps. The first step is anomaly detection, which uses local outlier factor (LOF) based method to find abnormal human motions. The second step is activity classification in which WiFall employs an one-class Support Vector Machine to distinguish human fall based on the features extracted from anomaly patterns. [21] proposed a CSI based human activity recognition system (CARM). In the CARM, principal component analysis (PCA) is first applied to filter out the noise from CSI data, and then a Hidden Markov Model (HMM) is adopted for classification with the features extracted by a 12-level DWT. [22] used fine-grained radio reflections to detect mouth movements via off-the-shelf WiFi infrastructure. The system also achieves lip reading and speech recognition in through-wall scenarios. [23] adopted data fusion for activity recognition system. The basic idea of the aforementioned methods is to match CSI patterns against activity profiles. [24] designed an unobtrusive user behaviors analysis system leveraging CSI signals. The general process used in human activity recognition systems includes the following steps: preprocessing the CSI signals, feature extraction, and classifying the activities using machine learning techniques. Data de-Noising (e.g., DWT or wavelet) and any specific handcraft feature extraction techniques are required during the processing. In the recognition processing, the extraction of certain features is considered as the most important factor for increasing the performance.

Instead of employing the traditional machine learning techniques, deep learning techniques are employed to recognize activity without manual feature extraction. Gao *et al.* [25] transformed CSI measurements into radio images to preserve channel correlation to realize human localization and activity. A sparse auto-encoder (SAE) network based deep learning processing framework was employed to characterize the influence of human behaviors on surrounding WiFi signals. Researchers proposed some systems utilizing CNNs to automatically detect human behaviors [26, 27, 28]. For example, Ma *et al.* [26] proposed a 9-layer CNN to recognize sign language gestures using CSI, Zhou *et al.* [27] utilize a 7-layer CNN to detect finger gestures. The experimental results confirm that the deep learning networks can optimize the features effectively. [11] proposed to employ deep learning techniques such as long short-term memory (LSTM) to recognize activity without manual feature extraction. An improved version of LSTM was considered by [12] to learn features from raw CSI measurements for the improvement of activity recognition performance. The experimental evaluation validates that the LSTM based approaches outperforms the conventional machine learning approaches with hand-crafted features since LSTM networks are well-suited to classify based on time series data. Unfortunately, LSTM networks are unable to learn the useful information involved in the correlation among subcarriers. The summary of main related studies for CSI-based activity recognition is shown in Table.1.

Different with above schemes, we apply CNNs to extract features from raw CSI measurements on multiple subcarriers simultaneously and adopt the ensemble strategy to further boost the system's performance. It is widely accepted that the strength and diversity are the two keys factors for ensemble learning [29, 30]. Motivated by this, we adopt strong yet diversified classifiers in the system. More specifically, we use a novel ensemble architecture that fuses a multiple layer perception (MLP), a random forest (RF) and a support vector machine (SVM). Real-world experiments demonstrate that the proposed system significantly outperforms existing methods.

3. Preliminary information

In this section, we first reviews the CSI. Next, the motivation of the proposed ensemble scheme for activity recognition is discussed.

Table 1: A summary of main related studies on CSI-based activity recognition.

Literature	Equipment	Behavior	Preprocessing and Classification	Performance
Wu *et al.* [19]	Intel 5300 NIC	Stationary human detection	Bandpass filter Sinusoidal model	94.82%
Wang, Wu *et al.* [10]	Intel 5300 NIC	Fall	Anomaly detection SVM	94%
Li *et al.* [20]	Intel 5300 NIC	Five daily activities	DWT SVM	96.6%
Wang, Liu *et al.* [21]	Intel 5300 NIC	Eight daily activities	PCA DWT HMM	96.5%
Wang, Zou *et al.* [22]	USRP	Mouth motion	Butterworth filter DWT, MCFS, SVM	91%
Wang, Chen *et al.* [23]	Intel 5300 NIC	Nine daily activities	Low-pass filter DWT	96%
Gu *et al.* [24]	Intel 5300 NIC	Micro-gestures	Butterworth Filter HMM	93.3%
Gao *et al.* [25]	Intel 5300 NIC	Activities and locations	SAE	93.8%
Ma *et al.* [26]	Intel 5300 NIC	Sign gestures	CNN	86.66%
Zhou *et al.* [27]	Intel 5300 NIC	Ten finger gestures	CNN	98%
Yousefi *et al.* [11]	Intel 5300 NIC	Six daily activities	LSTM	90.5%
Chen *et al.* [12]	Intel 5300 NIC	Six daily activities	ABLSTM	97.3%

3.1. Channel State Information

WiFi propagation that supports IEEE 801.11n/ac with multiple transmit and multiple receive antennas can be modeled as MIMO (multiple inputs multiple outputs) leveraging Orthogonal Frequency Division Modulation (OFDM). Current WiFi devices, such as Intel 5300 wireless network interface card, could provide physical layer channel frequency response information of subcarriers to upper layers in the format of CSI as the following structure

$$H_t = [H_t^1, ..., H_t^k, ..., H_t^K]^T \tag{1}$$

where H_t denotes the CSI measurement vector obtained at time t, K represents the total number of multipath components and are measured at $K = 30$ for Intel 5300 wireless NICs, T indicates the transposition operation, and H_t^k denotes the CSI for subcarriers k at time t which is defined by

$$H_t^k = A_t^k e^{j\phi_t^k} \tag{2}$$

where A_t^k and $e^{j\phi_t^k}$ indicate the amplitude and phase of k-th path at time t respectively. Note that the received phase information is highly distorted due to several sources of errors such as Carrier Frequency Offset (CFO), Symbol Timing Offset (STO) and Sampling Frequency Offset (SFO) in a WiFi system [11]. In contrast to phase, the amplitude information of CSI is generally reliable and widely used for feature extraction and classification in human activity recognition systems. Therefore, the amplitude of CSI is adopted for activity recognition in this work. Let $\mathbf{x}_k \in \mathbb{R}^{N_T}$ and $\mathbf{y}_k \in \mathbb{R}^{N_R}$ denote transmitted and received signals for k-th subcarrier in frequency domain, where N_T and N_R represent the number of transmit antennas at the device and the number of receive antennas at the access point (AP), respectively. A $N_T \times N_R$ communication system at any time instant can be modeled by the following equation.

$$\mathbf{y}_k = H^k \mathbf{x}_k + \eta_k \tag{3}$$

where η_k represents an N_R dimensional noise vector.

Most existing WiFi CSI-based human behavior recognition systems employed the Intel 5300 NIC to collect CSI measurements. However, the Intel 5300 tool only provides CSI with 30 subcarriers for each TX-RX antenna pair. In order to obtain more fine-grained information than conventional CSI tools, we equip Atheros-CSI-Tool on AR9590 based WiFi network interface cards (NIC) and develop our platform with a laptop served as a RX and a wireless router served as a TX. Our system reports CSI measurements on 56 subcarriers for each TX-RX antenna pair. At each time instant, $N_T \times N_R \times 56$ CSI streams are available to analyze the influence of WiFi signals caused by the movements of the person.

3.2. Rationale

Compared with conventional WiFi MAC layer RSS, the CSI is more sensitive and accurate indicator for activity recognition. The fundamental and crucial problem is to characterize the influence of human activities on WiFi CSI signals. Pioneer works have done valuable exploration on how to extract features from CSI measurements on each subcarrier independently to achieve activity recognition. However, multiple OFDM subcarriers are adjacent to each other, and thus the correlation information between different subcarriers may loss. Motivated by this, we explore algorithms to extract features from multiple CSI subcarriers jointly. Meanwhile, we realize that convolutional neural networks can be utilized to extract elaborate features from image based signals with two-dimensional perspective. This inspires us to employ CNNs to learn both the time-varying correlation information in each CSI subcarrier and the subcarrier correlation information between different subcarriers.

Based on this idea, we notice that notwithstanding CNN architectures have shown excellent performance for nonlinear features extraction. However, they are not always optimal for classification, because the performance may depend on the MLP layer which contains a lot of trainable parameters (e.g., the number of neurons and the depth). The study of Huang and LeCun [31] presented a hybrid technique that integrating the CNN and SVM achieves a superior performance compared to the use of CNN only and the use of SVM only. In the hybrid method, the CNN was employed for automatically feature extraction, and SVM was used for object categorization. Typically, the ensemble model that aggregates multiple weighted classifiers is able to outperform every single classifier in it [32]. We realize that ensemble techniques can successfully boost the performance of CNNs in many applications [33, 34].

Meanwhile, RF and SVM have been shown to be effective for CSI based human activity recognition with well-defined features [12]. Specifically, SVM [35] is one of the most popular machine learning algorithms due to its ability to deal with nonlinear classification problems, such as face recognition [36], image processing [37] and others. The reason why random forest has great activity recognition performance is that it builds multiple decision trees and merges them together to get a more accurate and stable prediction[38, 39]. Consider these aspects, we proposed an ensemble architecture that combines the respective outputs of RF, SVM and MLP of CNN with the features extracted from the feature extraction layer of CNN (i.e., the RF, SVM and MLP share the same features) to achieve even higher classification accuracy for CSI-based activity recognition.

4. The proposed WiAReS with WiFi CSI

In this section, we first present the design of our WiAReS using WiFi CSI measurements. Then, the proposed hybrid structure of CNN, RF and SVM for activity recognition is introduced in details.

4.1. Proposed System Architecture

Fig.2 presents the architecture of our proposed human activity recognition system with CSI measurements. The system mainly consists of the following modules: CSI data acquisition module which collects CSI amplitude measurements. Deep feature extraction module that extracts the features automatically from raw CSI data. Activity classification module that utilizes several classification algorithms to recognize the activity of the user. Activity decision module that employs an ensemble approach to combine the results of different classifiers. The parameters of the deep learning network and the classification module should be learned offline using the training data sets. In the online phase, the system performs the activity estimation directly using the currently CSI measurements.

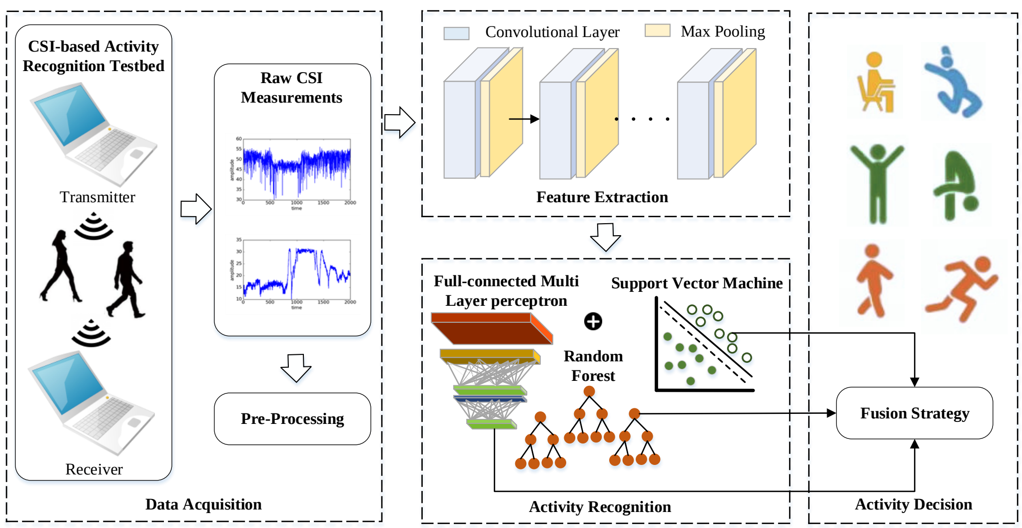


Figure 2: Full schematic diagram of our system architecture.

4.2. Activity recognition based on proposed ensemble model

In contrast to existing hand-crafted feature extraction and deep learning techniques, we use CNN to extract the features from multiple CSI subcarriers simultaneously. Fig.3 shows our proposed model for activity recognition using CSI measurements. The correlative parameters, such as the size of each filter, and the stride of each sliding window, are provided in the figure. After the input layer, three convolutional layers interleave with three maxpooling layers. The convolution operation generates multiple feature maps with kernels of size 5×5 in all convolutional layers. The ReLU function is used as activation function within the whole network. The three maxpooling layers use kernels of size 2×2.

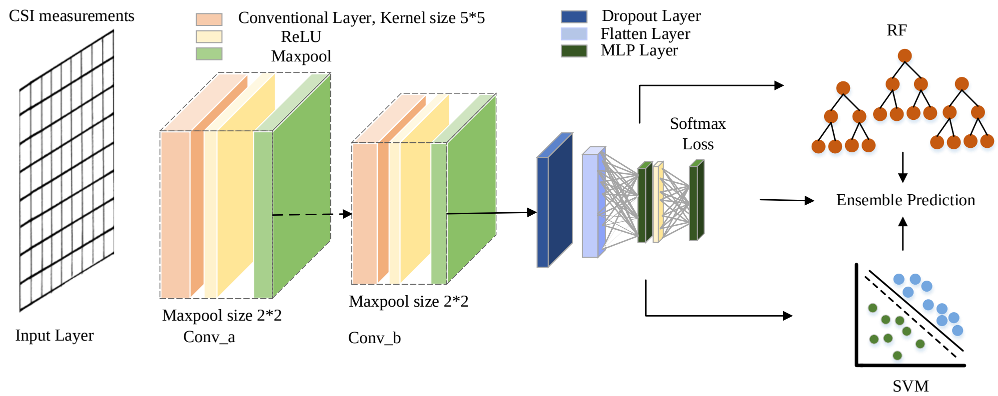


Figure 3: Architecture of the proposed ensemble model for human activity recognition

In the proposed WiAReS, a laptop equipped with Atheros network interface card is used to record the CSI measurements on multiple subcarriers while a person performs certain activities. Suppose we have N_S subcarriers, after collecting the CSI measurements for consecutive N_P packages, we can get a two dimensional $(N_P \times N_S)$ measurement matrix H.

In our proposed model, the time-series CSI signals H are fed into the convolutional layer as input. Then, convolutional operations are performed between the previous layer and a series of learnable kernel filters to detect local features from the input feature map. Next, an additive bias is added to the outputs of the convolutions. The input and kernel filters of the previous layer can be expressed as:

$$z_j^l = \sum_{l=1}^{M^{l-1}} (H_i^l \otimes k_{i,j}^{l-1}) + b_j^l \tag{4}$$

where H_i^{l-1} and z_j^l represent the input and current output of the lth layer, respectively. $k_{i,j}^{l-1}$ is defined as the weight kernel filter from the ith neuron in layer $l-1$ to the jth neuron in layer l. $\otimes$ expresses the convolutional operation and b_j^l denotes the bias of the jth neuron in the lth layer. M^{l-1} is the number of kernel filters in the $l-1$th layer. Then, we apply the activation function of $ReLU = max(0, x)$ over the outputs of convolutional layers to resolve vanishing or exploding gradients problem. The ReLu activation function is performed as follows:

$$H_j^l = ReLU(z_j^l) \tag{5}$$

where H_i^l is the intermediate output of layer l. We can obtain the output of layer l after the max pooling operation:

$$y_j^l = max\_pool(H_j^l) \tag{6}$$

The purpose of the max pooling is to detect a more useful representation while downsampling the input feature maps. We reconstruct the output of last pooling layer as the input of the next layer, i.e., $H_j^{l+1} = y_j^l$. The procedures of the other convolutional layers and max pooling layers are the same as those of the layers mentioned above, except for different parameters.

Afterward, the outputs of the last pooling layer y_j^l are fed into both RF, SVM and MLP for classification. For the fully connected MLP, we use 50% dropout to avoid overfitting and calculate the predicted error ξ with y_j^l according to the cost function using cross entropy with a softmax function, defined as:

$$\xi = -\sum_n \sum_i (T_{ni} \log t_{ni}) \tag{7}$$

where T_{ni} and t_{ni} denote the target labels and the predicted outputs of i-th class in n-th training set. We use stochastic gradient descent as the optimization algorithm to minimize the predicted error ξ. When training SVM, Radial Basis Function (RBF) is used as the kernel function with Penalty Parameter $C = 1$ and the gamma parameter $gamma = 0.0001$. For RF, the number of trees in the forest is set as 20 and entropy function is applied to measure the quality of a split. Additional details of SVM, RF and MLP are available in [40], [11] and [41].

We aggregate the outputs of RF, SVM and MLP of the CNN according to the ensemble principle to obtain an enhanced results that achieve a superior performance than the single classifier. A new fusion method is employed, which multiplies the probability of each classifier with their corresponding weighting factors. RF and softmax function predict the outputs with corresponding probabilities. Platt's trick is applied for extracting probabilities from SVM outputs[42]. The probability of the proposed ensemble method can be described as follows:

$$P(T_c) = \omega_1 \times P(t_{c,1}) + \omega_2 \times P(t_{c,2}) + \omega_3 \times P(t_{c,3}) \tag{8}$$

where $P(T_c)$ denotes the combined probability for activity c. We use $P(t_{c,1})$, $P(t_{c,2})$ and $P(t_{c,3})$ to denote the probability of RF, SVM and MLP for activity c, respectively, and we set $P(t_{c,1}) + P(t_{c,2}) + P(t_{c,3}) = 1$. The weights ω_1,

ω_2 and ω_3 are the weighting factors, which are fine-tuned using 5-fold cross validation on training data. According to the results, the weights of MLP, SVM and RF are set to be 0.2, 0.4 and 0.4, respectively.

5. Experimental Evaluation

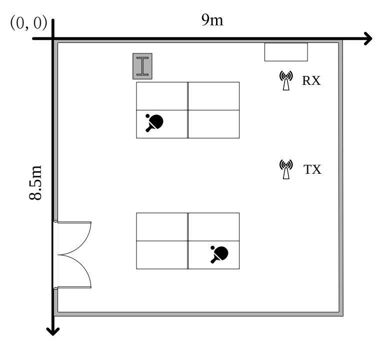
In this section, we discuss the implementation of WiARes and present the activity evaluation of users under various environments to show its accuracy and robustness.

5.1. Data Collection

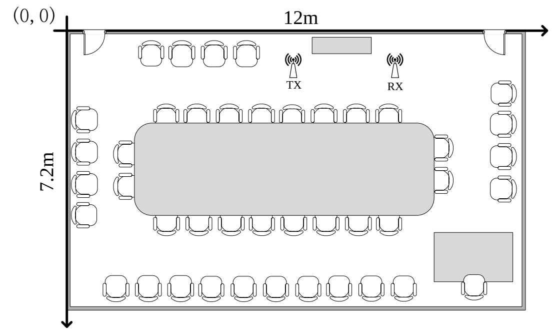
We implement the proposed system using a commercial TP-LINK wireless router and a laptop equipped with Atheros AR9590 NIC and Ubuntu operating system. The laptop works as the a receiver, while the router acts as a transmitter. The NIC is equipped with three receiving antennas and the router has one transmitting antenna. To acquire and record CSI measurements from three antennas, the laptop is installed with the 802.11n CSI Tool [16].

We analyze the performance of the proposed system in two typical indoor scenarios: (a) A typical activity room covering a $8.5 \times 9m^2$ area and consisting of two table tennis tables. (b) The second test scenario is a computer laboratory in a university. There are many tables and desktops crowded in the $16 \times 32m^2$ room. The layouts of the activity room and the laboratory are shown in Fig. 4. TX and RX represent the locations of the transmitter router and receiver laptop, respectively. While collecting the CSI data, TX and RX are located $3m$ apart under Line-of-Sight (LOS) condition according to the theoretical foundation for WiFi radio propagation in indoor environment [9]. The coordinators of TX and RX in activity room are (7.2,4), (7.2,1), and in laboratory are (6,1), (9,1). In each environment, the transmitter and receiver are placed $0.8m$ above the floor.

Totally, seven volunteers are involved for data collection with several feasible activities, i.e., Jump, Pick up, Run, Sit down, Wave hand, and Walk. Since most of the common daily activities happen in several seconds, a sampling frequency of 500 Hz is adopted to capture the CSI affected by those short time activities. A sliding window with a window size of 4s is used to construct CSI vector. It should be noted that all activities are performed by one user at separate time points. All volunteers are asked to perform each activity 101 times in each testing scenarios. The entire data collection



(a) activity room



(b) laboratory

Figure 4: The two scenarios layouts

process is recorded by the camera to mark all data. The volunteers include 5 male and 2 female graduate/undergraduate students with ages in the range of 18-27, and heights in the range of 1.6m to 1.8m. The experiments are conducted in different days in a month, with 4242 samples for each environment and 8484 samples in total. We randomly divide all the data into ten folds, one fold is selected for testing and the remaining for training.

5.2. Experimental Setup

To evaluate the performance of the proposed CSI based activity recognition system, we compare the performance of WiAReS with several state-of-the-art approaches. In [11], the deep learning based approach, i.e. conventional LSTM achieved a superior performance than those methodologies with hand-crafted feature such as hidden markov model (HMM), decision tree (DT) and logistic regression (LR). Therefore, we compare our proposed approach with LSTM [11] and CNN [26], which are able to learn feature automatically. We also compared the performance of different features extraction techniques. We employ discrete wavelet transform to extract features from CSI, and implement a SVM classifier and a RF classifier based on certain features [21]. We abbreviate activity recognition system with the aforementioned SVM classifier with features extracted from DWT as W-SVM [10, 20], and the RF classifier with wavelet features as W-RF [11], respectively. Meanwhile, we also conduct a comparison with hybrid CNN-SVM, CNN-RF which are composed of the CNN feature extraction layer and SVM, RF classifiers to verify whether an ensemble based technique is able to reduce the noise from raw CSI measurements and improve the recognition accuracy.

Table 2: The recognition accuracy of all activities with different algorithms

Methods	Sit down	Jump	Wave	Pick up	Walk	Run	Overall
W-RF [11]	91.6%	84.9%	88.7%	84.8%	83.9%	94.7%	88.1%
W-SVM [20]	86.8%	87.9%	92.7%	94.8%	93.8%	91.7%	91.3%
LSTM [11]	92.1%	81.8%	92.2%	91.9%	94.7%	96.9%	91.6%
CNN [26]	98.4%	97.4%	97.9%	97.5%	97.6%	97.1%	97.7%
CNN-RF	98.2%	97.2%	97.6%	97.2%	97.5%	97.6%	97.6%
CNN-SVM	98.1%	96.7%	98.5%	97.6%	97.7%	96.6%	97.5%
WiAReS	98.8%	98.7%	99.7%	98.7%	98.6%	98.6%	98.9%

5.3. Performance Overview

5.3.1. Overall performance

We first report the overall activity recognition performance of WiAReS using CSI data collected by Atheros AR9590. The input feature vectors are the raw CSI amplitude data, which are 168-dimensional vectors (3 antennas and 56 sub-carriers). We integrate the CSI data collected from two environments into one dataset and randomly partition the dataset into training and testing sets with the ratio of 9:1. As shown in Table 2, the proposed WiAReS using automatic feature extraction and multiple classifiers demonstrated the best performance overall. The overall recognition accuracies With Atheros NIC, W-RF, W-SVM, LSTM, CNN, CNN-RF, CNN-SVM and WiAReS achieve average accuracies of 88.10%, 91.28%, 91.60%, 97.65%, 97.55%, 97.53% and 98.85%, respectively. From the table, we can see that LSTM, CNN, CNN-RF and CNN-SVM have higher accuracy than W-RF and W-SVM, which confirms that the automatic deep learning-based feature extraction is better than conventional handcraft feature extraction scheme. We also discover that the CNN, CNN-RF, CNN-SVM and the proposed WiAReS achieve better performance than LSTM. This is because the convolution operation in the CNN network enables extracting efficient features involved in subcarrier correlation and time-domain correlation from raw CSI data.

5.3.2. Impact of environments

The environment in which the experiments are conducted is a crucial factor for the CSI-based behavior recognition system due to the distinctive multi-path distribution. We present the recognition accuracies under the two real deployments in Table 3. We observe that all algorithms achieve better activity recognition performance in the laboratory than that in activity room. By analysing the deployment of the two experimental environments, we discover that the activity room has larger interference where people come in and move out frequently. The laboratory environment which is usually occupied by certain occupants is relatively static. Given the accuracies of W-RF, W-SVM, LSTM, CNN, CNN-RF, CNN-SVM and WiAReS in laboratory are 88.2%, 90.9%, 91.9%, 98.6%, 98.6%, 98.1% and 99.0%, respectively. There is a 6.1%, 2.6%, 2.7%, 1.4%, 1.4%, 1% and 0.9% recognition accuracy reduction for each method in activity room. The recognition accuracy of WiAReS can be as high as 99.0% in the laboratory, drops to 98.1% in activity room. The experiment results demonstrate the robustness of WiAReS under different environments.

5.4. Additional Experiments with public dataset

To verify the effectiveness of the proposed WiAReS, we conduct additional experiments using a publicly available dataset which is collected from an indoor office area in [11]. The testbed consists of a TX and a RX. The RX is equipped with a commercial Intel 5300 network interface card, and acquires CSI with a sampling frequency of 1 KHz. In the experiments, a person may perform one of six feasible activities, i.e.,lie down, fall, walk, run, sit down, and stand up. Six volunteers are requested to perform an activity within a period of $20s$, and 20 trials for each one. Note that some activities (i.e., walk, sit down, run) are the same across the public dataset and our self-collected dataset, and the others are different.

The confusion matrices for all benchmark algorithms and the proposed WiAReS is shown in Table 4. It can be found that the activities of "Walk" and "Run" could achieve decent performance for most of classifiers, because these activities are with large movement ranges and can be easily recognized. We also can see that activities with smaller movement ranges, such as "Sit down" and "Stand up", have lower recognition accuracy than the rest activities. In the table, we discover that the performance is the worst when the hand-crafted feature extraction is used with RF and SVM classifiers. The performance of LSTM is slightly better than the DWT based classifiers. Compared to conventional algorithms, CNN-based algorithms could improve the performance through feature extraction layer. The CNN is slightly more accurate than CNN-RF and CNN-SVM, due to various optimization techniques, such as dropout, weight initialization, and batch normalization. Finally, considering that the recognition accuracies of RF and SVM are distinct between different activities, our proposed WiAReS using CNN-based automatic feature extraction and triple classifiers demonstrates the best performance, which confirms that the proposed ensemble CNN-based method is effective.

Table 3: The activity recognition accuracy under two testing environments

Environment	W-RF	W-SVM	LSTM	CNN	CNN-RF	CNN-SVM	WiAReS
Activity Room	82.1%	88.3%	89.2%	97.2%	97.2%	97.1%	98.1%
Laboratory	88.2%	90.9%	91.9%	98.6%	98.6%	98.1%	99.0%

6. Conclusion

This paper presents a WiFi CSI-based activity recognition system *WiAReS* that considers the influence of human behaviors on surrounding CSI measurements. First, this paper explores the challenges associated with subcarrier correlation information loss. We use CNNs to learn optimized features involved with subcarrier correlation and time-domain correlation from raw CSI measurements on multiple subcarriers. Second, we propose a diversified deep ensemble system using features learned from convolutional layer. Third, we develop Atheros devices based testbed for *WiAReS* so that 56 subcarriers of CSI measure-ments can be obtain for each TX-RX antenna pair. we implement *WiAReS* using commercial WiFi devices and collect a human activity database under two real environments. Our results show that the recognition performance of *WiAReS* can be enhanced compared to handcrafted feature extraction or conventional deep learning architectures.

[1] Z. Wang, K. Jiang, Y. Hou, Z. Huang, W. Dou, C. Zhang, Y. Guo, A survey on csi-based human behavior recognition in through-the-wall scenario, IEEE Access 7 (2019) 78772–78793.

[2] J. K. Aggarwal, M. S. Ryoo, Human activity analysis: A review, ACM Computing Surveys (CSUR) 43 (2011) 16.

[3] E. Ertin, N. Stohs, S. Kumar, A. Raij, M. Al'Absi, S. Shah, Autosense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field, in: Proceedings of the 9th ACM

Table 4: Confusion matrices for all the benchmark approaches and the proposed WiAReS approach.

(a) W-RF

Actual \ Predicted	Lie down	Fall	Walk	Run	Sit down	Stand up
Lie down	0.76	0.03	0.0	0.0	0.11	0.10
Fall	0.08	0.81	0.01	0.04	0.05	0.01
Walk	0.01	0.02	0.93	0.02	0.01	0.01
Run	0.01	0.05	0.09	0.83	0.01	0.01
Sit down	0.08	0.02	0.01	0.03	0.77	0.09
Stand up	0.05	0.01	0.01	0.04	0.18	0.71

(b) W-SVM

Actual \ Predicted	Lie down	Fall	Walk	Run	Sit down	Stand up
Lie down	0.89	0.01	0.01	0.05	0.03	0.01
Fall	0.01	0.90	0.00	0.00	0.04	0.00
Walk	0.00	0.05	0.95	0.05	0.00	0.00
Run	0.00	0.00	0.05	0.95	0.00	0.00
Sit down	0.00	0.02	0.00	0.00	0.80	0.18
Stand up	0.04	0.01	0.0	0.0	0.10	0.85

(c) LSTM

Actual \ Predicted	Lie down	Fall	Walk	Run	Sit down	Stand up
Lie down	0.95	0.01	0.01	0.01	0.0	0.02
Fall	0.01	0.94	0.05	0.00	0.00	0.00
Walk	0.00	0.01	0.93	0.04	0.01	0.01
Run	0.00	0.00	0.02	0.97	0.01	0.00
Sit down	0.03	0.01	0.05	0.02	0.81	0.07
Stand up	0.01	0.0	0.03	0.05	0.07	0.83

(d) CNN

Actual \ Predicted	Lie down	Fall	Walk	Run	Sit down	Stand up
Lie down	0.99	0.00	0.00	0.00	0.00	0.01
Fall	0.01	0.98	0.01	0.00	0.00	0.00
Walk	0.00	0.00	0.99	0.01	0.00	0.00
Run	0.00	0.00	0.01	0.99	0.00	0.00
Sit down	0.01	0.00	0.04	0.01	0.90	0.04
Stand up	0.00	0.00	0.02	0.01	0.03	0.94

(e) CNN-RF

Actual \ Predicted	Lie down	Fall	Walk	Run	Sit down	Stand up
Lie down	0.99	0.00	0.00	0.00	0.00	0.01
Fall	0.01	0.98	0.01	0.00	0.00	0.00
Walk	0.00	0.00	0.98	0.02	0.00	0.00
Run	0.00	0.00	0.01	0.99	0.00	0.00
Sit down	0.01	0.00	0.02	0.01	0.94	0.02
Stand up	0.00	0.00	0.02	0.01	0.05	0.92

(f) CNN-SVM

Actual \ Predicted	Lie down	Fall	Walk	Run	Sit down	Stand up
Lie down	0.99	0.00	0.00	0.00	0.00	0.01
Fall	0.00	0.99	0.01	0.00	0.00	0.00
Walk	0.00	0.00	0.99	0.02	0.00	0.00
Run	0.00	0.00	0.01	0.99	0.00	0.00
Sit down	0.01	0.00	0.02	0.01	0.94	0.02
Stand up	0.00	0.00	0.01	0.01	0.04	0.94

(g) WiAReS

Actual \ Predicted	Lie down	Fall	Walk	Run	Sit down	Stand up
Lie down	0.99	0.00	0.00	0.00	0.00	0.01
Fall	0.00	0.99	0.00	0.00	0.01	0.00
Walk	0.00	0.00	0.99	0.01	0.00	0.00
Run	0.00	0.00	0.01	0.99	0.00	0.00
Sit down	0.01	0.00	0.00	0.00	0.93	0.06
Stand up	0.03	0.00	0.02	0.00	0.00	0.95

Conference on Embedded Networked Sensor Systems, ACM, 2011, pp. 274–287.

[4] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, I. Poupyrev, Soli: Ubiquitous gesture sensing with millimeter wave radar, ACM Transactions on Graphics (TOG) 35 (2016) 142.

[5] Z. Wang, K. Jiang, Y. Hou, W. Dou, C. Zhang, Z. Huang, Y. Guo, A survey on human behavior recognition using channel state information, IEEE Access 7 (2019) 155986–156024.

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

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

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

[9] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, B. Xie, Human respiration detection with commodity wifi devices: do user location and body orientation matter?, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2016, pp. 25–36.

[10] Y. Wang, K. Wu, L. M. Ni, Wifall: Device-free fall detection by wireless networks, IEEE Transactions on Mobile Computing 16 (2017) 581–594.

[11] S. Yousefi, H. Narui, S. Dayal, S. Ermon, S. Valaee, A survey on behavior recognition using wifi channel state information, IEEE Communications Magazine 55 (2017) 98–104.

[12] Z. Chen, L. Zhang, C. Jiang, Z. Cao, W. Cui, Wifi csi based passive human activity recognition using attention based blstm, IEEE Transactions on Mobile Computing 18 (2019) 2714–2724.

[13] S. Sigg, M. Scholz, S. Shi, Y. Ji, M. Beigl, Rf-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals, IEEE Transactions on Mobile Computing 13 (2014) 907–920.

[14] S. Sigg, S. Shi, F. Buesching, Y. Ji, 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, ACM, 2013, p. 43.

[15] D. Halperin, W. Hu, A. Sheth, D. Wetherall, Tool release: Gathering 802.11 n traces with channel state information, ACM SIGCOMM Computer Communication Review 41 (2011) 53–53.

[16] Y. Xie, Z. Li, M. Li, Precise power delay profiling with commodity wi-fi, IEEE Transactions on Mobile Computing (2018).

[17] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, Y. Liu, Widar2. 0: Passive human tracking with a single wi-fi link, Procs. of ACM MobiSys (2018).

[18] L. Wenyuan, W. Siyang, W. Lin, A multiperson behavior feature generation model based on noise reduction using wifi, IEEE Transactions on Industrial Electronics 67 (2020) 5179–5186.

[19] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, J. Cao, Non-invasive detection of moving and stationary human with wifi, IEEE Journal on Selected Areas in Communications 33 (2015) 2329–2342.

[20] H. Li, X. He, X. Chen, Y. Fang, Q. Fang, Wi-motion: A robust human activity recognition using wifi signals, IEEE Access 7 (2019) 153287–153299.

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

[22] G. Wang, Y. Zou, Z. Zhou, K. Wu, L. M. Ni, We can hear you with wi-fi!, IEEE Transactions on Mobile Computing 15 (2016) 2907–2920.

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

[24] Y. Gu, X. Zhang, Z. Liu, F. Ren, Besense: Leveraging wifi channel data and computational intelligence for behavior analysis, IEEE Computational Intelligence Magazine 14 (2019) 31–41.

[25] Q. Gao, J. Wang, X. Ma, X. Feng, H. Wang, Csi-based device-free wireless localization and activity recognition using radio image features, IEEE Transactions on Vehicular Technology 66 (2017) 10346–10356.

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

[27] Q. Zhou, J. Xing, W. Chen, X. Zhang, Q. Yang, From signal to image: enabling fine-grained gesture recognition with commercial wi-fi devices, Sensors 18 (2018) 3142.

[28] L. Jia, Y. Gu, K. Cheng, H. Yan, F. Ren, Beaware: Convolutional neural network (cnn) based user behavior understanding through wifi channel state information, Neurocomputing (2020).

[29] L. Zhang, P. N. Suganthan, Benchmarking ensemble classifiers with novel co-trained kernel ridge regression and random vector functional link ensembles [research frontier], IEEE Computational Intelligence Magazine 12 (2017) 61–72.

[30] R. Katuwal, P. N. Suganthan, L. Zhang, Heterogeneous oblique random forest, Pattern Recognition 99 (2020) 107078.

[31] F. Huang, Y. LeCun, Large-scale learning with svm and convolutional netw for generic object recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.

[32] X. Qiu, Y. Ren, P. N. Suganthan, G. A. Amaratunga, Empirical mode decomposition based ensemble deep learning for load demand time series forecasting, Applied Soft Computing 54 (2017) 246–255.

[33] D. Cireşan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, arXiv preprint arXiv:1202.2745 (2012).

[34] D. CireşAn, U. Meier, J. Masci, J. Schmidhuber, Multi-column deep neural network for traffic sign classification, Neural networks 32 (2012) 333–338.

[35] V. N. Vapnik, An overview of statistical learning theory, IEEE transactions on neural networks 10 (1999) 988–999.

[36] D. Gong, Z. Li, W. Huang, X. Li, D. Tao, Heterogeneous face recognition: A common encoding feature discriminant approach, IEEE Transactions on Image Processing 26 (2017) 2079–2089.

[37] W. YuKun, Z. Yong, W. JieSheng, C. Ying, Research of flotation dosing control method based on svm and image processing technology, in: Proceedings of the 31st Chinese Control Conference, IEEE, 2012, pp. 7067–7070.

[38] L. Zhang, J. Varadarajan, P. Nagaratnam Suganthan, N. Ahuja, P. Moulin, Robust visual tracking using oblique random forests, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5589–5598.

[39] L. Zhang, P. N. Suganthan, Oblique decision tree ensemble via multisurface proximal support vector machine, IEEE Transactions on Cybernetics 45 (2014) 2165–2176.

[40] Z. Chen, Q. Zhu, Y. C. Soh, L. Zhang, Robust human activity recognition using smartphone sensors via ct-pca and online svm, IEEE Transactions on Industrial Informatics 13 (2017) 3070–3080.

[41] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks 61 (2015) 85–117.

[42] J. Platt, et al., Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Advances in large margin classifiers 10 (1999) 61–74.