

Human Activity and Gesture Recognition Based on WiFi Using Deep Convolutional Neural Networks

Sokienah K. Jawad*, Musaab Alaziz

Department of Computer Engineering, University of Basrah, Basrah, Iraq

Correspondence

* Sokienah K. Jawad

Department of Computer Engineering,
University of Basrah, Basrah, Iraq

Email: engpg.sokienah.jawad@uobasrah.edu.iq
mosab.adil@uobasrah.edu.iq

Abstract

WiFi-based human activity and gesture recognition explore the interaction between the human hand or body movements and the reflected WiFi signals to identify various activities. This type of recognition has received much attention in recent years since it does not require wearing special sensors or installing cameras. This paper aims to investigate human activity and gesture recognition schemes that use Channel State Information (CSI) provided by WiFi devices. To achieve high accuracy in the measurement, deep learning models such as AlexNet, VGG 19, and SqueezeNet were used for classification and extracting features automatically. Firstly, outliers are removed from the amplitude of each CSI stream during the preprocessing stage by using the Hampel identifier algorithm. Next, the RGB images are created for each activity to feed as input to Deep Convolutional Neural Networks. After that, data augmentation is implemented to reduce the overfitting problems in deep learning models. Finally, the proposed method is evaluated on a publicly available dataset called WiAR, which contains 10 volunteers, each of whom executes 16 activities. The experiment results demonstrate that AlexNet, VGG19, and SqueezeNet all have high recognition accuracy of 99.17 %, 96.25%, and 100 %, respectively.

KEYWORDS: Human gesture recognition, Channel State Information, deep convolutional neural networks, data augmentation.

I. INTRODUCTION

Human activity and gesture recognition have become a thriving study field in human society, ubiquitous computing, and security monitoring [1]–[3]. Daily actions [4][5] were viewed as vital ways of communicating in our daily lives, and we can interact with our bodies rather than words. As a result, human activity and gesture recognition systems have been proposed in terms of application demand, technological support, and auxiliary devices.

Previous similar activity recognition research works were generally divided into three groups: vision-based, sensor-based [6][7], and WiFi-based. Camera-based systems are effective in tracking human activity but they do not provide individuals with privacy. Sensor-based systems can also be used to track and monitor daily activities, such as elderly care, smart sensing, sports applications, and tracking[8]–[10]. Although the sensor-based technique has no privacy concerns, it can be difficult at times due to the sensors that users must wear. In recent years, WiFi-based human activity recognition systems have been developed to solve the foregoing restrictions, including activity-based [11]–[13], gesture-based [14], and keystroke-based [15].

Conventional power attributes, such as Received Signal Strength Indicator (RSSI), are unable to provide adequate differentiation and robustness in complicated indoor situations since RSSI is the combination of multipath signals having fast-changing phases. The latest developments in wireless technologies offer new answers to the issues mentioned above. Channel information can be received in part through Orthogonal Frequency Division Multiplexing (OFDM) receivers in the form of Channel State Information (CSI), which displays a set of channel measurements defining the amplitudes and phases of subcarriers in the 802.11 a/g/n standards [16].

CSI-based studies have shown that by employing simply WiFi signals, researchers can accurately estimate position and detect activity [17]–[19]. Activity recognition systems based on WiFi, on the other side, offer numerous advantages. They overcome problems of security in camera-based systems as well as battery life in wearable-based devices. They are also contactless, unnoticeable, and can be utilized through walls.

The Channel State Information, which defines how a signal propagates from transmitter to receiver and contains both the amplitude and phase for each subcarrier, is used in



this research to provide a human activity and gesture recognition system.

In latest years, deep learning has been successfully utilized in vision, audio, and other domains with remarkable success in extracting features and classification methods. Deep learning offers an excellent classification capability and a solution for manually extracted features. The goal of feature extraction is to take information about a specified task from a signal. The majority of recent research use low-level features, including the statistical features used in [20][21] and the frequency domain features utilized in [22]. Therefore, this paper intends to utilize certain effective techniques to automatically extract high-level semantic features that are well met by the present task in order to achieve a better recognition result. As a result, deep convolutional neural networks, such as Alexnet, VGG 19, and SqueezeNet were used in this research. In a WiFi-based field that have never been done previously, these networks were also used for the classification task.

In this paper, a human activity and gesture recognition system is presented based on the channel state information of the WiFi devices. The Deep convolutional neural networks were used to distinguish 16 classes of movement. The main contributions of our work are summarized as follows:

- Proposing a new concept for improving the performance of WiFi-based human activity and gesture recognition by implementing data augmentation to minimize the overfitting which appears in deep CNN models.
- After conversion from CSI data to RGB images, a set of powerful algorithms in deep learning were proposed to achieve feature extracting and classification tasks.
- To increase the recognition accuracy, the proposed models, such as Alexnet, VGG19, and SqueezeNet, were modified (fine-tuned) to meet our task.
- The efficiency of the proposed method is verified via empirical results.

The remainder of this paper is structured as follows: Section II presents the description of the dataset used. The related works are introduced in Section III. The proposed methodology is fully described in Section IV. The conclusion of the paper is discussed in Section V.

II. DATASET DESCRIPTION

In this study, WiAR [23] is employed, which is a publicly available dataset that includes 16 gestures and activities (horizontal arm wave, high arm wave, two hands wave, high throw, draw x, draw tick, toss paper, forward kick, side kick, bend, hand clap, walk, phone call, drink water, sit down, squat) performed by ten people. Each volunteer does 30 trials of the movement. A unique shape for each signal is produced at the receiver when a person performs a specific action. Barriers and impediments distort, reflect, and disperse a signal as it travels from the transmitter to the receiver. Whenever the signal touches obstacles and objects, this leads to multipath overlay signals at the receiver. This technique can be described using fine-grained CSI [24]. The major benefit of CSI over RSSI is that it can identify changes at a specific frequency rather than averaging out

changes across all WiFi bandwidth, which is why it is utilized in this experiment. Because the Intel 5300 NIC is utilized to extract it, CSI represents fine-grained information of WiFi signals with thirty subcarriers. Since the number of subcarriers is determined by the tool used to extract it. The main components of the Channel State Information are the amplitude and the phase. Each activity has a unique CSI signal pattern. CSI can be expressed as follows [23]:

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

Where $H(k)$ denotes the CSI of the k th subcarrier. The amplitude and the phase of CSI represent by $\| H(k) \|$ and $\angle H(k)$, respectively. Only the amplitude is used in this paper.

III. RELATED WORK

In this part, various similar existing works were reviewed. Wang *et al.* [12] exploited a laptop with the use of an Intel 5300 Network Interface Card to gather WiFi signals and subsequently extract CSI data. Pu *et al.* [14] exploited Universal Software Radio Peripheral to capture Doppler Shifts as features from the arriving WiFi signals and then utilized them to recognize nine motion patterns. For feature extraction, Principal Component Analysis (PCA) was used. Based on the CSI value, the CSI-speed model and CSI-activity model were created to describe three main human activities. Chien and Xiu [16] proposed a human activity detection based on WiFi. Deep neural networks were used for activity classification. To reduce training effort, Virmani *et al.* [22] applied the transfer learning principle. The gestures were decomposed into two parts (linear and non-linear) by the system, which was then transformed into attributes. Parisa *et al.* [24] created a CSI-based dataset for human activity recognition using Raspberry pi with a WiFi device. The CSI data is transformed into images and fed into a variety of Convolutional Neural Network classifiers, including (Long Short Term Memory (LSTM), 1D-CNN, 2D-CNN, and Bidirectional LSTM). Kellogg *et al.* [25] employed a specialized circuit to segregate the amplitudes of received signals before matching the signal's attributes to the motions. Jiang *et al.* [26] introduced a Generative Adversal Network to create and extract features. In the classification stage, the SVM classifier was utilized. The implementation of this approach has proven helpful for keeping track of time and dealing with small samples. Cong *et al.* [27] presented user authentication using deep learning in daily activities based on WiFi. Jiang *et al.* [28] presented the Fine-grained Indoor Fingerprinting System design. This system investigates a physical layer of Channel State Information that provides the channel status over all subcarriers. Using Channel State Information (CSI), Xuyu *et al.* [29] developed a revolutionary deep learning-based indoor fingerprinting method. The system architecture contains an off-line training phase and an online localization phase based on three CSI assumptions. Deep learning is used in the off-line training phase to train all of the weights as fingerprints.

IV. METHODOLOGY

To obtain good results, the suggested method involves various stages, including Preprocessing, Image Construction of CSI, and finally, Activity and Gesture recognition system using Deep Convolutional Neural Networks (DCNN). The architecture of the human activity and gesture recognition system is shown in Fig. 1.

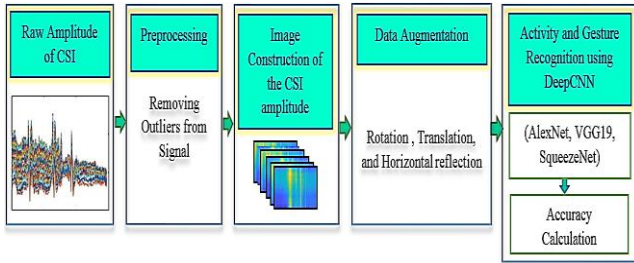


Fig. 1: The system architecture of the proposed approach for recognizing human gestures and activities.

A. Preprocessing

The noise and outliers must be removed from the amplitude extracted from the channel state information since they have a direct impact on classification outcomes. Furniture, interference from other neighboring devices, and transmitter transmitting power adaption are all potential causes. As a result, when the volunteer performs a gesture or movements between the transmitter and receiver, there is more noise. This results in modifications to the receiver's channel state information, which represents three main channels (antennas) with 30 subcarriers each, and hence incorrect CSI values. The abnormal points were removed using the Hampel identifier algorithm [30]–[32]. Outliers are points that are outside the range of Formula (2).

$$[\mu - \gamma \times \sigma, \mu + \gamma \times \sigma] \quad (2)$$

Where μ is the signal's mean value, σ is the standard deviation, and γ is the removal factor. In different contexts, the value of γ can be different. It was set to 3 because it is the most common in similar studies. Fig. 2 shows the original signal before and after outliers removal.

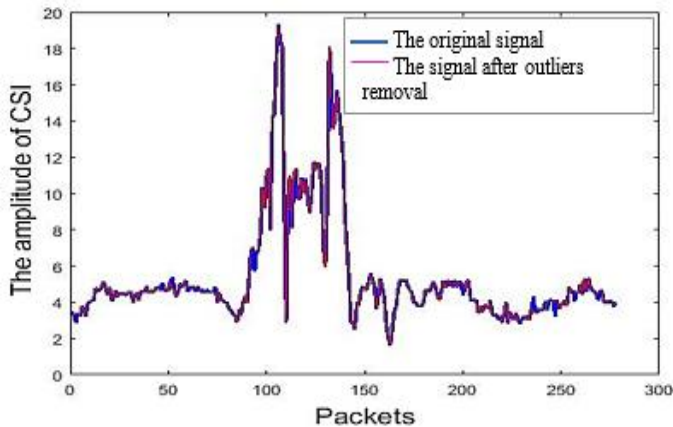
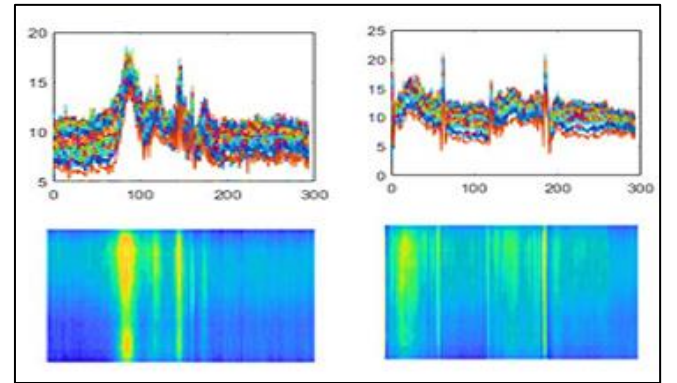


Fig. 2: The original signal before and after outliers removal.

B. Image Construction of CSI Amplitude

From the CSI amplitude, 3-channel RGB images are created [24]. To start, all 30 subcarriers of the amplitude of channel state information from the most sensitive antenna are used to create images without losing any sensitive data. Using the (*imagesc*) Matlab function, the signal from one activity was transformed into its color image. Each element in the amplitude of a specific movement corresponds to a rectangular area in the image. The values of the elements of original data are indices into the current colormap that determine the color of each patch. The images are scaled to the desired dimensions (227 -by-227 for AlexNet and SqueezeNet and 224-by-224 for VGG19). Fig. 3 shows the combinations of 30 subcarriers of CSI amplitude for some



activities before and after conversion to RGB images.

(a)

(b)

Fig. 3: The CSI amplitude of 30 subcarriers before and after conversion to the RGB- images for (a) horizontal-arm wave and (b) high-arm wave.

C. Data Augmentation

Data augmentation describes the methods to increase the amount of data by introducing additional, slightly altered copies from the current data or synthesizing new data from the existing data. The dataset, including images, has now been augmented. Because the models now have adequate samples for learning, augmentation is a very crucial feature of all deep learning classification models and it improves the accuracy quotient by lowering the bias factor [33]–[35]. Rotation, translation, and horizontal reflection are three methods of data augmentation applied with the following values. The rotation value is $[-30, +30]$, and the translation is equal to 10 degrees.

D. Activity and Gesture Recognition using Deep Convolutional Neural Networks (DCNN)

Deep CNN algorithms have been utilized to solve a variety of computer vision problems with great success. To achieve the desired result, CNN takes an input image and sends it through various layers, such as convolutional, nonlinear, fully connected, and pooling. Transfer learning (TL) is commonly expressed in computer visualization by using pre-trained models. The use of pre-trained models is

possible due to the high computing cost of training such models from scratch [36].

The majority of recent research makes use of statistical features employed in [20] and frequency domain features [22] that are extracted manually from signals. This research introduces some techniques that are capable to automatically extract high-level features and perform classifications. Models (such as AlexNet, VGG19, and SqueezeNet) trained on naturalistic image classification have been applied to Wi-Fi activity and gesture recognition systems. These models were chosen for their outstanding robustness and efficiency in a variety of applications. Numerous studies have shown that a deep network's intermediary layers can collect features that provide a suitable balance between depiction and target independence [37]. The CSI amplitude image is utilized to fine-tune the CNN because the features generated from CNN trained on ImageNet are not good for Wi-Fi gesture and activity recognition. This technique is known as Fine-Tuning CNN. The models that were implemented are briefly described below.

AlexNet is a convolutional neural network variation. Alex Krizhevsky proposed this model in 2012. There are eight layers in this model, including 5 convolutional layers and 3 fully connected layers. The maximum pooling layers are placed after some convolutional layers of the model [38].

The VGG-19 is a CNN with 19 layers deep [39]. The Visual Geometry Group at Oxford was responsible for its creation, hence the VGG name. This network uses deep Convolutional neural layers to improve accuracy, based on some of its predecessors' ideas. This network contains 47 layers. The 19 layers with learnable weights include 16 convolutional layers and 3 fully connected layers.

Forrest *et al.* [42] released SqueezeNet, an 18-layer deep neural network for computer vision, in 2016. The authors' purpose in developing SqueezeNet was to produce a smaller neural network with fewer parameters that could fit into computer memory and be communicated over a computer network more easily. It is a more compact version of AlexNet. It has nearly half as many parameters as AlexNet but performs three times faster.

The AlexNet, VGG19, and SqueezeNet models were fine-tuned to meet our classification requirement and produce good results. Because the first layer in AlexNet, the image input layer, requires images with a size of $227 \times 227 \times 3$, where 3 is the number of color channels for the image, the size of the color images is set to 227×227 . The last three layers were replaced with a fully connected layer, a softmax layer, and a classification output layer in the new classification task. The fully connected layer was set to the number of classes used in our dataset (16 classes). For these newly created fully connected layers, the rate of learning is set very high. Because the rest of the model is unaltered except for these three layers, the augmented datasets are used to train it.

The size of the color images in VGG19 was set to 224×224 because the first layer, the image input layer, needs images with a size of $224 \times 224 \times 3$. The network's last three layers were eliminated and replaced with new ones. These

layers in the new classification task include a fully connected layer, a softmax layer, and a classification output layer.

The image input size for the SqueezeNet is $227 \times 227 \times 3$. However, the image dataset contains images of various sizes. To automatically resize the training images, the augmented image datastore was employed. The network's convolutional layers capture features of the image that are used by the final learnable layer and final classification layer for classifying the entered image. The last two layers include instructions for combining the characteristics extracted by the network into probabilities for the class, a value for loss, and predicted labels. These layers were replaced with new layers appropriate to the dataset used for retraining a CNN model to classify new images.

A fully connected layer is the final layer in most networks with learnable weights. The final learnable layer in some nets, such as SqueezeNet, is a 1×1 convolutional layer. In this situation, the convolutional layer was replaced with a new one with the same number of filters as classes.

The augmented dataset is divided into two parts: training (50%) and validation (50%). After each iteration, the model training performance is checked with a validation check. The hyperparameters used in AlexNet, VGG19, and SqueezeNet are listed in TABLE I. where "sgdm" stands for stochastic gradient descent with momentum and is used as a weights adjustment optimization approach.

TABLE I
THE HYPERPARAMETERS USED IN DEEP CNN MODELS

| Model | AlexNet | VGG-19 | SqueezeNet |
|------------------|---------------------------|---------------------------|---------------------------|
| Parameters | | | |
| Input Image Size | $227 \times 227 \times 3$ | $224 \times 224 \times 3$ | $227 \times 227 \times 3$ |
| Learning Rate | 0.0001 | 0.0002 | 0.0002 |
| Batch Size | 10 | 10 | 10 |
| Epochs | 20 | 20 | 30 |
| Optimizer | sgdm | sgdm | sgdm |

E. Results and Accuracy Calculation

The following measure is selected as a performance metric:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Where TP is a correctly classified sample (true positive), FP is an incorrectly classified sample (false positive), TN represents if the sample is negative and correctly classified as negative, it is considered a true negative, FN denotes a false negative, and occurs when a sample is positive but classified as negative. AlexNet, VGG-19, and SqueezeNet exhibited the best classification accuracy of 99.17 %, 96.25 %, and 100 %, respectively, as shown in Fig. 4, Fig. 5, and Fig. 6.

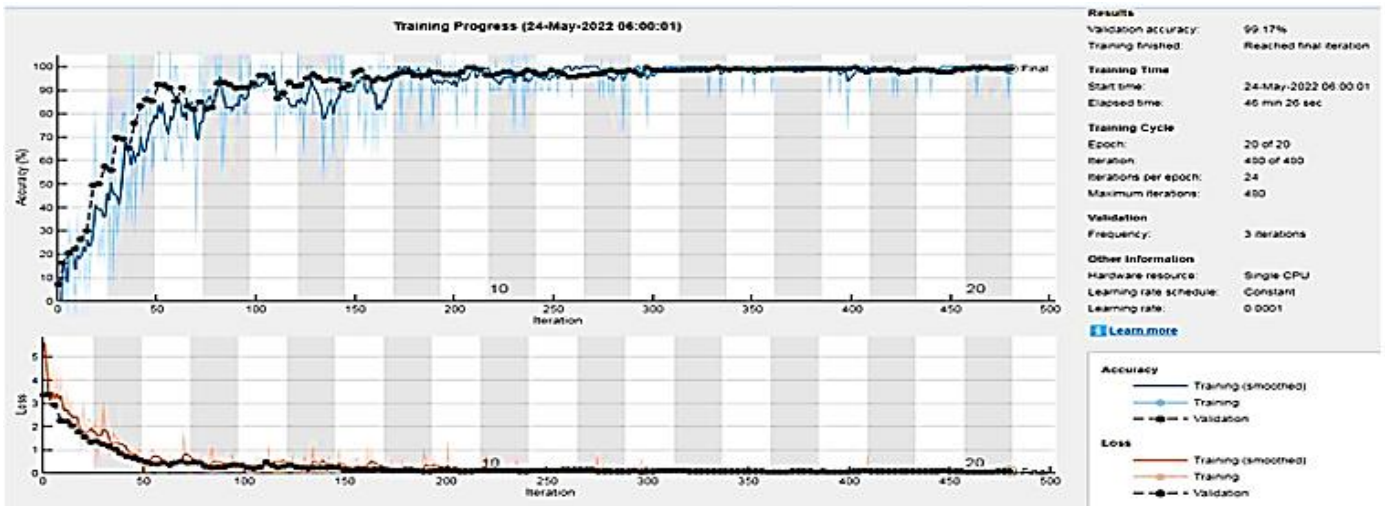


Fig. 4: The fine-tuned model AlexNet's, loss, training, and validation accuracy curve.

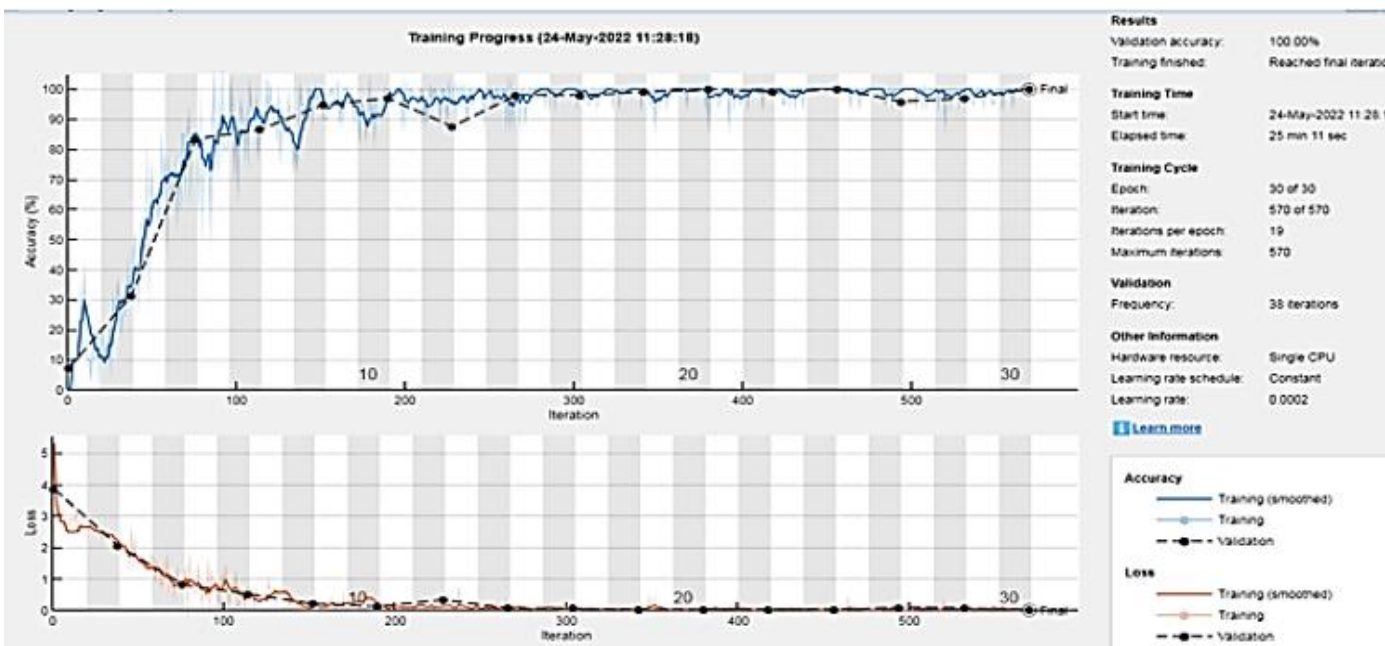


Fig. 5: The squeezeNet fine-tuned model's, loss, training, and validation accuracy curve.

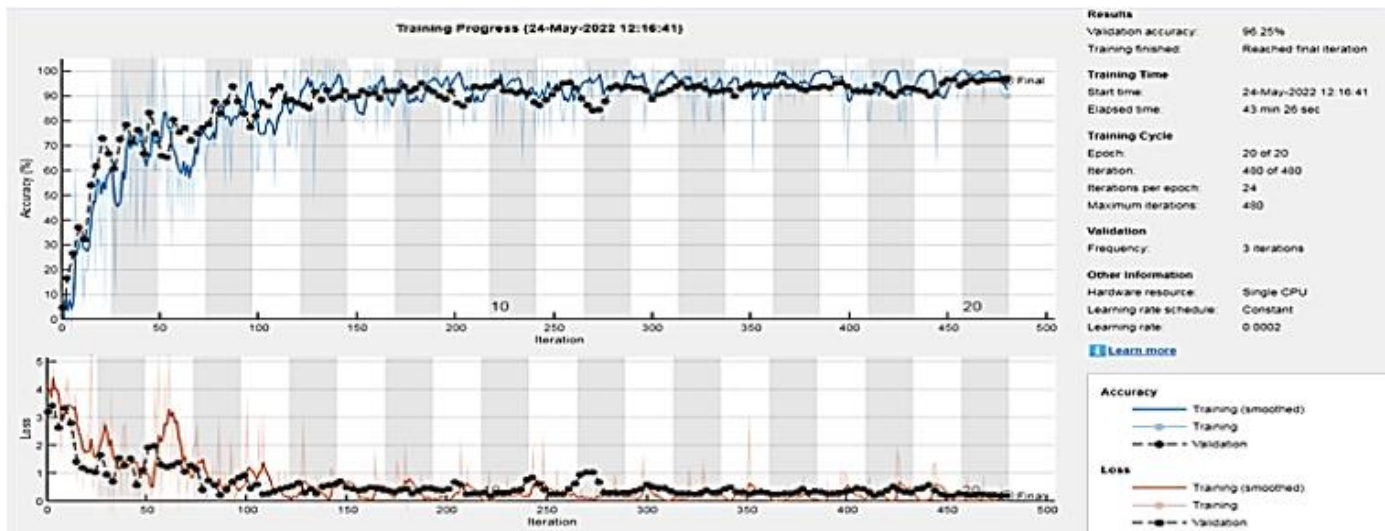


Fig. 6: The VGG19 fine-tuned model's, loss, training, and validation accuracy curve.

V. CONCLUSION

This paper demonstrates a human activity and gesture recognition system employing WiFi device channel state information. Data preprocessing, feature extraction, data augmentation, and classification techniques were all utilized to increase system efficiency. First, the Hampel identifier technique was used to eliminate outliers from the amplitude of CSI. Then, the CSI data was converted to RGB images for each movement by combining the amplitudes of thirty subcarriers of the most sensitive antenna. Moreover, New approaches in deep learning were used for automatically extracting features and classification tasks. Deep CNN models, such as AlexNet, VGG19, and SqueezeNet, were fine-tuned before being applied to a big dataset called WiAR, which contains 16 different types of movement in the form of raw signals. The results of this study show that the proposed strategy of deep transfer learning models recognizes the classes very well and with high accuracy, outperforming the existing strategies.

CONFLICT OF INTEREST

The authors have no conflict of relevant interest to this article.

REFERENCES

- [1] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.
- [2] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, 2016.
- [3] S. Gaglio, G. Lo Re, and M. Morana, "Human activity recognition process using 3-D posture data," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 586–597, 2014.
- [4] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE conference on computer communications (INFOCOM)*, 2015, pp. 1472–1480.
- [5] M. S. Aljumaily and G. A. Al-Suhail, "Towards ubiquitous human gestures recognition using wireless networks," *International Journal of Pervasive Computing and Communications*, 2017.
- [6] H. Salim, M. Alaziz, and T. Abdalla, "Human Activity Recognition Using The Human Skeleton Provided by Kinect," *Iraqi Journal for Electrical and Electronic Engineering*, vol. 17, no. 2, pp. 183–189, 2021, doi: 10.37917/ijeee.17.2.20.
- [7] M. Alaziz, Z. Jia, R. Howard, X. Lin, and Y. Zhang, "In-bed body motion detection and classification system," *ACM Transactions on Sensor Networks (TOSN)*, vol. 16, no. 2, pp. 1–26, 2020.
- [8] H. A. Salim, M. Alaziz, and T. Y. Abdalla, "People Tracking in a Smart Room Using Kinect Sensor," in *2021 International Conference on Communication & Information Technology (ICICT)*, 2021, pp. 221–226.
- [9] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, pp. 1–33, 2014.
- [10] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *23th International conference on architecture of computing systems 2010*, 2010, pp. 1–10.
- [11] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 617–628.
- [12] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015, pp. 65–76.
- [13] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 363–373.
- [14] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*, 2013, pp. 27–38.
- [15] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using wifi signals," in *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015, pp. 90–102.
- [16] C. C. Lee and X. C. Huang, "Human Activity Detection via WiFi Signals Using Deep Neural Networks," in *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*, 2018, pp. 3–4.
- [17] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1–32, 2013.
- [18] Y. Wang, X. Jiang, R. Cao, and X. Wang, "Robust indoor human activity recognition using wireless signals," *Sensors*, vol. 15, no. 7, pp. 17195–17208, 2015.
- [19] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with WiFi," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329–2342, 2015.
- [20] S. Arshad et al., "Wi-chase: A WiFi based human activity recognition system for sensorless environments," in *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2017, pp. 1–6.
- [21] M. A. A. Al-qaness, "Device-free human micro-activity recognition method using WiFi signals," *Geo-spatial Information Science*, vol. 22, no. 2, pp. 128–137, 2019.
- [22] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using wifi," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 252–264.

- [23] L. Guo et al., “Wiar: A public dataset for wifi-based activity recognition,” *IEEE Access*, vol. 7, pp. 154935–154945, 2019.
- [24] P. Fard Moshiri, R. Shahbazian, M. Nabati, and S. A. Ghorashi, “A CSI-Based Human Activity Recognition Using Deep Learning,” *Sensors*, vol. 21, no. 21, p. 7225, 2021.
- [25] B. Kellogg, V. Talla, and S. Gollakota, “Bringing gesture recognition to all devices,” in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014, pp. 303–316.
- [26] D. Jiang, M. Li, and C. Xu, “Wigan: A wifi based gesture recognition system with gans,” *Sensors*, vol. 20, no. 17, p. 4757, 2020.
- [27] C. Shi, J. Liu, H. Liu, and Y. Chen, “WiFi-enabled user authentication through deep learning in daily activities,” *ACM Transactions on Internet of Things*, vol. 2, no. 2, pp. 1–25, 2021.
- [28] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, “FIFS: Fine-grained indoor fingerprinting system,” in *2012 21st international conference on computer communications and networks (ICCCN)*, 2012, pp. 1–7.
- [29] X. Wang, L. Gao, S. Mao, and S. Pandey, “DeepFi: Deep learning for indoor fingerprinting using channel state information,” in *2015 IEEE wireless communications and networking conference (WCNC)*, 2015, pp. 1666–1671.
- [30] Y. Lu, S. Lv, and X. Wang, “Towards location independent gesture recognition with commodity wifi devices,” *Electronics*, vol. 8, no. 10, p. 1069, 2019.
- [31] L. Davies and U. Gather, “The identification of multiple outliers,” *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 782–792, 1993.
- [32] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, “WiFinger: Talk to your smart devices with finger-grained gesture,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 250–261.
- [33] L. Perez and J. Wang, “The effectiveness of data augmentation in image classification using deep learning,” *arXiv preprint arXiv:1712.04621*, 2017.
- [34] H. R. Roth et al., “Anatomy-specific classification of medical images using deep convolutional nets,” in *2015 IEEE 12th international symposium on biomedical imaging (ISBI)*, 2015, pp. 101–104.
- [35] D. Tiwari, M. Dixit, and K. Gupta, “Deep Multi-View Breast Cancer Detection: A Multi-View Concatenated Infrared Thermal Images Based Breast Cancer Detection System Using Deep Transfer Learning,” *Traitement du Signal*, vol. 38, no. 6, 2021.
- [36] M. Naz, J. H. Shah, M. A. Khan, M. Sharif, M. Raza, and R. Damaševičius, “From ECG signals to images: a transformation based approach for deep learning,” *PeerJ Computer Science*, vol. 7, p. e386, 2021.
- [37] Q. Bu, G. Yang, J. Feng, and X. Ming, “Wi-fi based gesture recognition using deep transfer learning,” in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation(SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)*, 2018, pp. 590–595.
- [38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [39] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size,” *arXiv Prepr. arXiv1602.07360*, 2016.