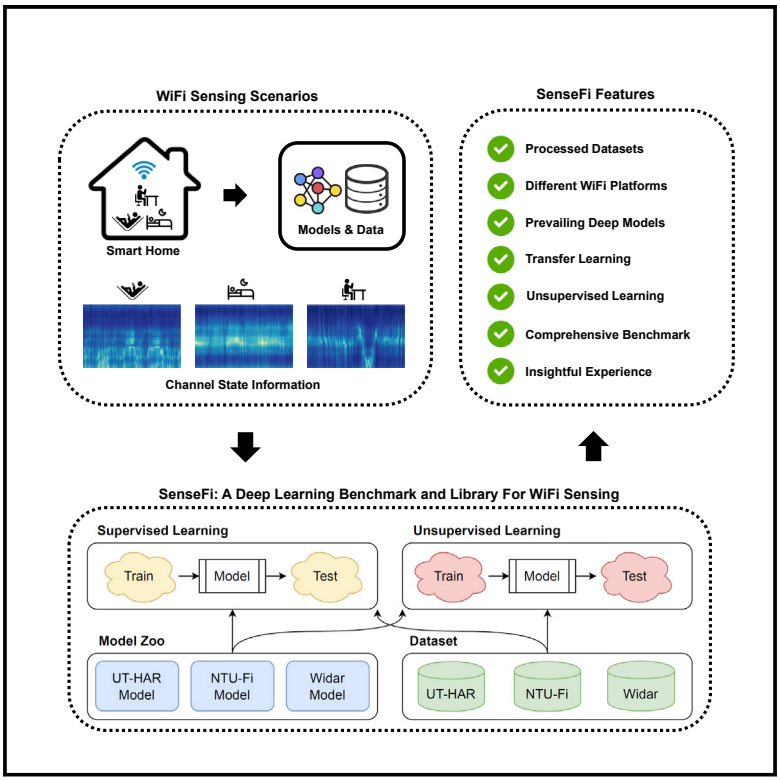
SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing

Article

Patterns

SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing

Graphical abstract



Highlights

- SenseFi offers a model zoo and a comprehensive benchmark for WiFi sensing
- Pre-training models on large datasets can be generalized to downstream WiFi-sensing tasks
- Shallow models outperform very-deep models across different environments
- Processed datasets for different WiFi-sensing platforms are available to use

Authors

Jianfei Yang, Xinyan Chen, Han Zou, Chris Xiaoxuan Lu, Dazhuo Wang, Sumei Sun, Lihua Xie

Correspondence

yang0478@ntu.edu.sg (J.Y.), elhxie@ntu.edu.sg (L.X.)

In brief

WiFi sensing is a method to detect the presence or motion of humans or objects within a wireless network by analyzing signal propagation changes utilizing learning algorithms. SenseFi, presented in this article, is a comprehensive framework for WiFi-sensing researchers. It integrates current learning algorithms, hardware platforms, and datasets for different WiFi-sensing tasks, which could support future research on designing WiFi-sensing models.

Yang et al., 2023, Patterns *4*, 100703
March 10, 2023
https://doi.org/10.1016/j.patter.2023.100703

CellPress

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

Patterns

CellPress
OPEN ACCESS

Article

SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing

Jianfei Yang,[1,4,*] Xinyan Chen,[1] Han Zou,[1] Chris Xiaoxuan Lu,[2] Dazhuo Wang,[3] Sumei Sun,[3] and Lihua Xie[3,*]
[1]School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798, Singapore
[2]School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK
[3]Institute for Infocomm Research (I²R), Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore
[4]Lead contact
*Correspondence: yang0478@ntu.edu.sg (J.Y.), elhxie@ntu.edu.sg (L.X.)
https://doi.org/10.1016/j.patter.2023.100703

THE BIGGER PICTURE WiFi is extensively used in wireless communication to connect devices within a network. Along with the development of machine-learning algorithms and the widespread use of Internet of Things (IoT) products, applications of WiFi have recently expanded from communication to sensing. The presence or motion of humans or any objects within the wireless environment can be interpreted from signal propagation patterns. Compared with traditional video forms of sensing, WiFi sensing has the benefits of privacy protection, non-line-of-sight (NLOS) detection, and broad coverage. A comprehensive WiFi-sensing framework called SenseFi is proposed in this article that integrates hardware platforms, learning algorithms, and datasets applied for different WiFi-sensing tasks. We hope that SenseFi will contribute to future algorithm design and evaluation for real-world applications.

Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

Over the recent years, WiFi sensing has been rapidly developed for privacy-preserving, ubiquitous human-sensing applications, enabled by signal processing and deep-learning methods. However, a comprehensive public benchmark for deep learning in WiFi sensing, similar to that available for visual recognition, does not yet exist. In this article, we review recent progress in topics ranging from WiFi hardware platforms to sensing algorithms and propose a new library with a comprehensive benchmark, SenseFi. On this basis, we evaluate various deep-learning models in terms of distinct sensing tasks, WiFi platforms, recognition accuracy, model size, computational complexity, and feature transferability. Extensive experiments are performed whose results provide valuable insights into model design, learning strategy, and training techniques for real-world applications. In summary, SenseFi is a comprehensive benchmark with an open-source library for deep learning in WiFi sensing research that offers researchers a convenient tool to validate learning-based WiFi-sensing methods on multiple datasets and platforms.

INTRODUCTION

With the proliferation of mobile Internet usage, wireless networks, such as WiFi access points (APs), have become a ubiquitous component of infrastructure in smart environments, ranging from commercial buildings to domestic surroundings. By analyzing the patterns of its wireless signals, current-day APs have evolved beyond being pure WiFi routers and also serve as a type of ''sensor device'' to enable new services for human sensing. In particular, recent studies have pointed out that WiFi signals in the form of channel state information (CSI)[1,2] are extremely promising for a variety of device-free human-sensing tasks, such as occupancy detection,[3] activity recognition,[4–7] fall detection,[8] gesture recognition,[9,10] human identification,[11–13] people counting,[14,15] and pose estimation.[16] Unlike coarse-grained received signal strengths, WiFi CSI records more fine-grained information in terms of the propagation of a signal between WiFi devices and their reflection with respect to the environment of human beings. However, as WiFi signals (2.4 or 5 GHz) lie in the non-visible band of the electromagnetic spectrum, WiFi CSI-based human sensing is intrinsically more privacy friendly than camera-based surveillance. Thus, it has

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

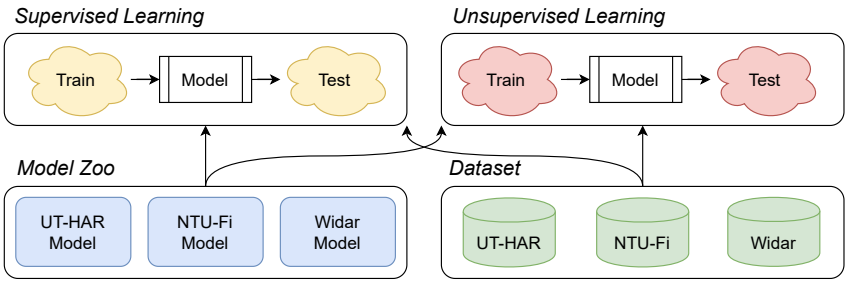


Figure 1. The diagram of the SenseFi library and benchmark

drawn significant attention from both academic and industrial agents. In response to the increasing interest, a new WiFi standard, 802.11bf,[17] was designed by the IEEE 802.11bf Task Group (TGbf), which will be used to amend the current WiFi standard both at the medium access control (MAC) and physical layer (PHY) to include WiFi sensing as part of regular WiFi service officially by late 2024.

Existing WiFi-sensing techniques can be categorized into model- and learning-based methods. Model-based methods, such as Fresnel Zone,[18] rely on physical models that describe the propagation of WiFi signals. They explicate the underlying mechanism of WiFi sensing and design sensing methods for periodic or single motions, such as respiration[19–22] and falling.[8,23,24] Nevertheless, model-based methods are incapable of accurately sensing complicated human activities consisting of a series of different motions. For example, human gait comprises the synergistic movement of arms, legs, and bodies, differences that are difficult to depict using physical models. In contrast, by feeding significant amounts of data into machine-learning[25] or deep-learning models,[5,9] learning-based methods exhibit remarkable performance in complicated sensing tasks. Various deep neural networks have been designed to enable applications such as human activity recognition[26] and gesture recognition.[9] Although deep-learning models have performed admirably in function approximation, they require very high amounts of labeled data, which are expensive to collect and are adversely affected by distribution shifts induced by environmental dynamics.[27]

Most state-of-the-art deep-learning models have been developed for computer vision tasks,[28] such as human activity recognition,[29,30] and natural language processing,[31] such as sentiment classification.[32] These models have demonstrated their ability to process high-dimensional and multimodal data. These approaches have inspired the use of deep learning in WiFi sensing for data pre-processing, network design, and learning objectives. As a result, an increasing number of deep-learning models have been developed for WiFi sensing.[33,34] These have successfully addressed the shortcomings of traditional statistical learning methods. However, this article mainly focuses on achieving high accuracy on specific sensing tasks by customizing deep neural networks and does not explore the relationship between various deep-learning models and different WiFi-sensing data collected using different devices and CSI tools. It is unclear whether the impressive results reported in research papers on WiFi sensing should be primarily attributed to the deep model design or the WiFi platform. Therefore, there are still significant gaps in our understanding of the relationship between deep learning and WiFi sensing. For instance, the following questions may be considered: (1) how can a deep neural network be customized for a WiFi-sensing task by integrating existing network modules (e.g., fully connected layer, convolutional layer, recurrent neural unit, and transformer block) into a synergistic framework? (2) How do existing models perform when compared fairly on multiple WiFi-sensing platforms and data modalities? (3) How can a trade-off be achieved between recognition accuracy and efficiency?

To address these questions, we propose SenseFi, a benchmark and model zoo library for WiFi CSI sensing, using deep learning. To this end, we first introduce the prevalent deep-learning models, including multilayer perceptron (MLP), convolutional neural network (CNN), recurrent neural network (RNN), variants of RNN, CSI transformers, and CNN-RNN, and summarize their effectiveness in CSI feature-learning and WiFi-sensing tasks. Then, we investigate and benchmark these models on three WiFi human activity recognition datasets comprising both raw CSI data and processed data collected using the Intel 5300 CSI tool[1] and the Atheros CSI tool.[25,35] The accuracies and efficiencies of these models are analyzed and compared to demonstrate their potential for real-world applications. We also explore the benefits to different WiFi-sensing tasks afforded by transfer learning and the utilization of unsupervised learning to extract features without labels, reducing the need for annotation. As depicted in Figure 1, all deep models in our model zoo, along with the dataset loading and learning schemes, are incorporated into a single library, allowing researchers to develop and evaluate their models easily in a variety of environments.

The contributions of this study are summarized as follows.

- We systematically introduce WiFi-sensing technology, analyze the benefits afforded to WiFi sensing by cutting-edge deep-learning models, and review recent progress on deep-learning-empowered WiFi sensing.
- We propose SenseFi, a comprehensive WiFi-sensing framework that enables systematic evaluation and comparison of various deep-learning models in an open-source manner. SenseFi benchmarks advanced deep models and learning schemes for WiFi sensing, providing evidence and tools for future research.
- In addition to existing datasets (UT-human activity recognition [HAR][36] and Widar[37]), we construct and release two new datasets (NTU-Fi HAR and human identification [Human-ID]) using a CSI platform that provides higher-resolution CSI data than that currently available, enabling us to benchmark deep-learning methods and evaluate their feasibility for WiFi sensing.
- Unlike existing works that focus on supervised learning, SenseFi also investigates and benchmarks transfer learning and unsupervised learning schemes, which enable knowledge transfer across different sensing tasks and data-efficient WiFi sensing, respectively.

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

- We summarize the observations and experiences that may benefit real-world applications of WiFi-sensing research in terms of model design, model training, and learning strategies. We also discuss current challenges and outline future directions of research.

Concept of WiFi sensing

WiFi-sensing data: CSI

In WiFi communication, CSI captures information on the propagation of wireless signals in a physical environment after diffraction, reflection, and scattering by describing the channel properties of the communication link. Modern wireless communication networks following the IEEE 802.11 standard, such as multiple-input multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM), at the PHY, aim to increase data capacity and improve orthogonality in transmission channels affected by multipath propagation. As a result, modern WiFi APs typically involve multiple antennas with many subcarriers for OFDM. Corresponding to a pair of transmitter and receiver antennas, CSI describes the phase shift and amplitude attenuation of multiple paths on each subcarrier. Compared with received signal strength, CSI data are of higher resolution and can be considered to be "WiFi images" of the environment of propagation. Specifically, the channel impulse response (CIR), $h(\tau)$, of WiFi signals is defined as follows in the frequency domain:

$$h(\tau) = \sum_{l=1}^{L} \alpha_l e^{j\varphi_l} \delta(\tau - \tau_l), \quad \text{(Equation 1)}$$

where α_l and φ_l denote the amplitude and phase of the l-th multipath component, respectively; τ_l denotes the time delay; L denotes the number of multipaths; and $\delta(\tau)$ denotes the Dirac delta function. To estimate CIR, the OFDM receiver samples the signal spectrum at the subcarrier level in a realistic implementation, which represents the amplitude attenuation and phase shift using a complex number. In WiFi sensing, CSI recording functions are realized using specific tools.[1,35] The estimation can be represented by

$$H_i = \|H_i\| e^{j\angle H_i}, \quad \text{(Equation 2)}$$

where $\|H_i\|$ and $\angle H_i$ denote the amplitude and phase of the i-th subcarrier, respectively.

WiFi-sensing tools and platforms

The number of subcarriers is determined by the bandwidth and tool used. The number of subcarriers is directly proportional to the resolution of the CSI data. Existing CSI tools include Intel 5300 NIC,[1] Atheros CSI Tool,[35] and Nexmon CSI Tool.[38] Several realistic sensing platforms have been constructed using these tools. The Intel 5300 NIC is the most commonly used tool, and it was the first CSI tool to be released. It can record 30 subcarriers for each pair of antennas during operation using a 20-MHz bandwidth. The Atheros CSI Tool increases the resolution of CSI data by improving the recording of CSI to 56 subcarriers at 20 MHz and 114 subcarriers at 40 MHz. These tools have been widely used in many applications.[5,6,9,25,39] The Nexmon CSI Tool was the first tool that enabled CSI recording on smartphones and Raspberry Pi—it can capture 256 subcarriers at 80 MHz. However, previous studies[40,41] have demonstrated that CSI data collected using the Nexmon CSI Tool are quite noisy. In this study, we only investigate the effectiveness of deep-learning models trained on representative CSI data obtained using the widely used Intel 5300 NIC and Atheros CSI tools.

CSI data transformation and cleansing

In general, CSI data consist of a vector of complex numbers, including the amplitude and phase. The processing of these data for deep-learning models is the primary aim in WiFi sensing. Based on existing works, we summarize the following approaches.

1. Only the amplitude data are used as input. Raw phases obtained from a single antenna are randomly distributed due to random phase offsets,[42] stabilizing the amplitude of CSI and enhancing its suitable for WiFi sensing. A simple denoising scheme, such as wavelet denoising,[25] can be used to filter the high-frequency noise in CSI amplitudes. This is the most common practice in most WiFi-sensing applications.
2. The CSI difference between antennas is used for model-based methods. Although raw phases are noisy, the phase difference between two antennas is quite stable[9] and can reflect subtle gestures more accurately than amplitudes. The CSI ratio[43] was proposed to mitigate noise using the division operation and by increasing the sensing range. These techniques are mostly designed for model-based solutions as they require clean data for setting thresholds.
3. The processed Doppler representation of CSI is used. The dependence of CSI data on the environment is eliminated by simulating the Doppler feature, which only reflects human motion, using the body-coordinate velocity profile (BVP).

In our benchmark, as we focus on learning-based methods, we use the most common data modality (i.e., amplitude only) and the BVP modality designed to be domain invariant.

Insight: The effect of human activities on CSI

As depicted in Figure 2, CSI data for human sensing consist of two dimensions: subcarrier and packet number (i.e., time duration). At each packet or timestamp, t, we have $X_t = N_T \times N_R \times N_{sub}$, where N_T, N_R, and N_{sub} denote the numbers of transmitter antennas, receiver antennas, and subcarriers per antenna, respectively. This can be considered to comprise a "CSI image" of the surrounding environment at the time, t. The CSI images from subsequent timestamps form a "CSI video" that can describe human activity patterns. To connect CSI data with deep-learning models, we summarize the data properties that offer a better understanding of deep model design.

1. Subcarrier dimension → spatial features. The values of multiple subcarriers can capture the propagation of signals after diffraction, reflection, and scattering, thereby describing the spatial environment. These subcarriers can be considered to be analogous to image pixels, from which convolutional layers extract spatial features.[44]
2. Time dimension → temporal features. For each subcarrier, its temporal dynamics represent changes in the environment. In deep learning, temporal dynamics are usually modeled using RNNs.[45]
3. Antenna dimension → resolution and channel features. As each antenna captures a different propagation path of the

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

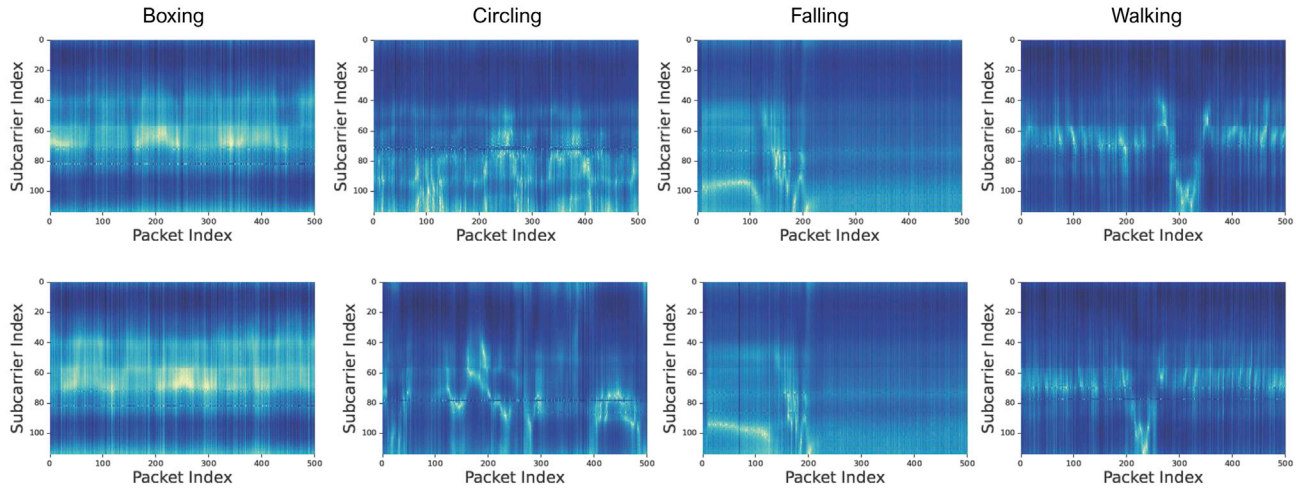


Figure 2. The CSI samples of three human activities in NTU-Fi, collected by Atheros CSI Tool

signal, it can be considered to be a channel in deep learning, similar to the RGB channels of an image. CSI data gathered from a single pair of antennas are similar to a grayscale image with only one channel. Thus, the number of pairs of antennas is directly proportional to the resolution of the CSI data. Antenna features should be processed separately in convolutional layers or recurrent neurons.

RESULTS

Models: Deep neural networks for WiFi sensing

Deep learning is a branch of machine learning that involves the use of models composed of numerous processing layers to learn data representations.[44] Unlike classical statistical learning, which relies primarily on handcrafted features designed manually using prior knowledge,[46] deep learning aims to extract features automatically from large amounts of labeled data and optimize model performance using backpropagation.[44] Although the deep-learning theories were developed in the 1980s, the high computational requirements made their implementation impractical until the development of graphical processing units (GPUs). Deep learning has since been widely used in fields such as computer vision,[28] natural language processing,[31] and interdisciplinary research.[47]

Deep-learning models for WiFi sensing typically consist of a feature extractor and a classifier. The classifier usually consists of several fully connected layers that can learn a good decision boundary, and the design of the feature extractor is critical to success. A large number of deep architectures have been proposed for feature extractors,[48] with each enjoying specific advantages for certain types of data. Deep-learning models for WiFi sensing are constructed using these prevailing architectures to extract patterns of human motion.[5] We summarize the latest works on deep models for WiFi sensing in Table 1 and observe that the network architectures of these works include MLP, CNN, RNN and its variants, combinations of these networks, and transformers.

Here, we briefly introduce these network architectures for benchmarking purposes. MLP relies on dense connections between neuron layers[71] and is commonly used as a classifier. When used as a feature extractor, it mixes the spatial and temporal dimensions of CSI data, damaging the intrinsic structure of the data. CNN preserves spatial information and uses shared weights for convolution kernels, which helps to reduce the number of parameters compared with MLP. 1D and 2D convolutions are widely used in WiFi sensing and have achieved good results.[37,54] However, CNNs can suffer from a limited receptive field and the equal importance of all convolutions, which can be partially addressed using an attention mechanism. Deep CNN, such as a series of residual neural networks (ResNets),[72] exhibits good performance in classification tasks on high-dimensional data. An RNN is designed to capture temporal patterns[73]; however, its training complexity is high owing to its high depth and number of parameters. Several variants of RNN have been developed to learn long-term patterns more efficiently, such as long short-term memory (LSTM) and GRU. Combining an RNN and a CNN enables the learning of spatiotemporal representations with fewer parameters. The transformer extracts features by exploiting attention among spatial or temporal dimensions and exhibits good performance. However, it relies significantly on a large number of parameters for training.[74] Figure 3 illustrates the input of WiFi CSI data into various network types.

Scenarios: Learning schemes for deep WiFi-sensing models

Traditional training of deep models relies on supervised learning based on high volumes of labeled data, but data collection and annotation are bottlenecks in realistic WiFi-sensing applications. For example, the recognition of human gestures requires users to perform gestures hundreds of times, which is impractical in real-world scenarios. Figure 4 illustrates the learning methods and their contribution to WiFi sensing in the real world.

Supervised learning is a method of training deep models using input data that has been labeled for a specific output. This is the most commonly used learning strategy in current WiFi-sensing

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

Table 1. A survey of existing deep-learning approaches for WiFi sensing

Method	Year	Task	Model	Platform	Scenario
Yousefi et al.[36]	2017	human activity recognition	RNN, LSTM	Intel 5300 NIC	supervised learning
WiCount[49]	2017	people counting	MLP	Intel 5300 NIC	supervised learning
EI[50]	2018	human activity recognition	CNN	Intel 5300 NIC	transfer learning
CrossSense[34]	2018	human identification, gesture recognition	MLP	Intel 5300 NIC	transfer ensemble learning
Chen et al.[51]	2018	human activity recognition	LSTM	Intel 5300 NIC	supervised learning
DeepSense[5]	2018	human activity recognition	CNN-LSTM	Atheros CSI Tool	supervised learning
WiADG[27]	2018	gesture recognition	CNN	Atheros CSI Tool	transfer learning
WiSDAR[52]	2018	human activity recognition	CNN-LSTM	Intel 5300 NIC	supervised learning
WiVi[7]	2019	human activity recognition	CNN	Atheros CSI Tool	supervised learning
SiaNet[9]	2019	gesture recognition	CNN-LSTM	Atheros CSI Tool	few-shot learning
CSIGAN[53]	2019	gesture recognition	CNN, GAN	Atheros CSI Tool	semi-supervised learning
DeepMV[54]	2020	human activity recognition	CNN (attention)	Intel 5300 NIC	supervised learning
WIHF[55]	2020	gesture recognition	CNN-GRU	Intel 5300 NIC	supervised learning
DeepSeg[56]	2020	human activity recognition	CNN	Intel 5300 NIC	supervised learning
Sheng et al.[57]	2020	human activity recognition	CNN-LSTM	Intel 5300 NIC	supervised learning
Schäfer et al.[41]	2021	human activity recognition	LSTM	Nexmon CSI Tool	supervised learning
Moshiri et al.[58]	2021	human activity recognition	CNN	Nexmon CSI Tool	supervised learning
Ding et al.[59]	2021	human activity recognition	CNN	Intel 5300 NIC	few-shot learning
Widar[37]	2021	human identification, gesture recognition	CNN-GRU	Intel 5300 NIC	supervised learning
WiONE[60]	2021	human identification	CNN	Intel 5300 NIC	few-shot learning
Ma et al.[61]	2021	human activity recognition	CNN, RNN, LSTM	Intel 5300 NIC	supervised learning
THAT[62]	2021	human activity recognition	Transformers	Intel 5300 NIC	supervised learning
WiGr[63]	2021	gesture recognition	CNN-LSTM	Intel 5300 NIC	supervised learning
MCBAR[64]	2021	human activity recognition	CNN, GAN	Atheros CSI Tool	semi-supervised learning
CAUTION[12]	2022	human identification	CNN	Atheros CSI Tool	few-shot learning
CTS-AM[65]	2022	human activity recognition	CNN (attention)	Intel 5300 NIC	supervised learning
WiGRUNT[66]	2022	gesture recognition	CNN (attention)	Intel 5300 NIC	supervised learning
Zhuravchak et al.[67]	2022	human activity recognition	LSTM	Nexmon CSI Tool	supervised learning
EfficientFi[39]	2022	human activity recognition, human identification	CNN	Atheros CSI Tool	multitask supervised learning
SecureSense[68]	2022	human activity recognition, human identification	CNN	Atheros CSI Tool	supervised learning
AirFi[69]	2022	gesture recognition	CNN-MLP	Atheros CSI Tool	transfer learning
AutoFi[70]	2022	human activity recognition, human identification	CNN-MLP	Atheros CSI Tool	unsupervised learning

research.[5,7,36,52] These methods usually use cross-entropy loss between the ground-truth label and prediction to optimize the model. Although supervised learning is easy to implement and exhibits high performance in many tasks, it requires a large amount of labeled data, which limits its widespread use in realistic applications.

Transfer learning is a method of transferring knowledge from one domain to another.[75] When the two domains are similar, a model can be pre-trained on one domain and fine-tuned in a new environment, which can lead to significant performance improvement. In contrast, when the two domains are distinct, such as the different environments in which CSI data are collected, the distribution shift can hinder performance, and domain adaptation should be used. Domain adaptation is a type of semi-supervised learning that addresses domain shifts in transfer learning. It has been successfully implemented in object recognition[76–78] and text classification.[79] Multidisciplinary scenarios are common in WiFi sensing because CSI data are significantly dependent on the training environment. Several methods have been developed to address this problem.[27,50,64,80,81]

Unsupervised learning is a method of learning data representations without any labels. The resulting feature extractor can be used to facilitate downstream tasks by training a specific classifier. Unsupervised learning has been demonstrated to improve the generalizability of models in visual recognition tasks[82] by avoiding reliance on any specific task. Current unsupervised learning models are based on self-supervised learning.[83] Despite its effectiveness, unsupervised learning has not been widely used in WiFi sensing, with only AutoFi being developed to enable model initialization for automatic user configuration in WiFi-sensing applications.[70]

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

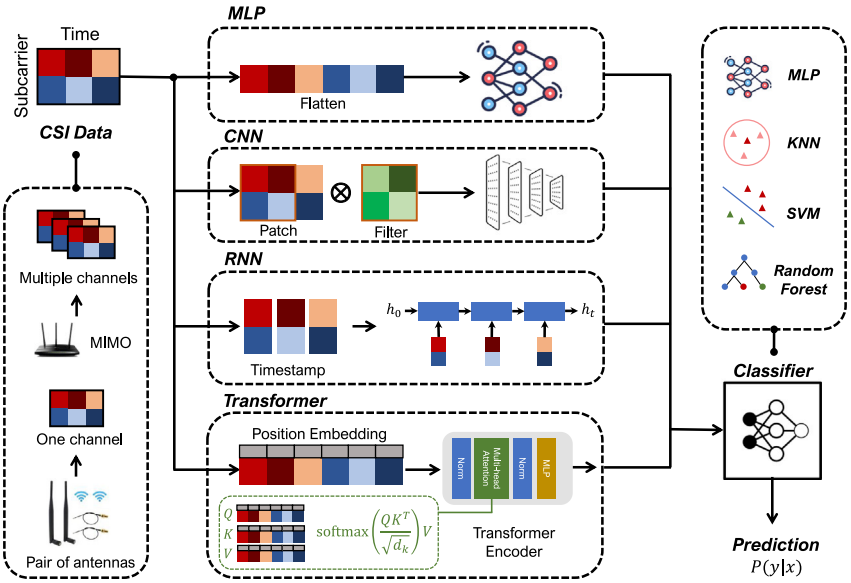


Figure 3. The illustration of how CSI data are processed by MLP, CNN, RNN, and Transformer

Few-shot learning is a data-efficient learning strategy that uses only a few samples from each category during training. This is typically achieved using contrastive or prototypical learning. SiaNet[9] is a pioneering method for exploring few-shot learning in WiFi sensing using a Siamese network. Subsequent studies[12,60] have extended prototypical networks from visual recognition to WiFi sensing and achieved good recognition results. In particular, the use of a single sample for each class during training is referred to as one-shot learning. Few-shot learning has potential applications in practical WiFi-based gesture recognition and Human-ID because of its reliance on only a small number of samples.

Ensemble learning uses multiple models to improve predictive performance.[84] The ensemble process can operate at the feature or prediction level. The feature-level ensemble concatenates features obtained from multiple models and subsequently trains a final classifier. Prediction-level ensemble is more common, usually referring to voting or probability addition. Ensemble learning can enhance performance, but it incurs heavy computation overhead. CrossSense[50] developed a mixture-of-experts approach, selecting one appropriate expert corresponding to each input, thereby reducing computation cost.

In this study, we empirically explore the effectiveness of supervised learning, transfer learning, and unsupervised learning for WiFi CSI data as they are the most commonly used learning strategies in WiFi-sensing applications.

Benchmarking deep learning in WiFi sensing

In this subsection, an empirical study is reported on WiFi CSI data in terms of the aforementioned deep models, including MLP, CNN, series of ResNet, RNN, GRU, LSTM, bidirectional LSTM (BiLSTM), CNN + GRU, and vision transformer (ViT; i.e., transformer). The number following the model indicates the number of layers, and all specific designs are described in the experimental procedures. The four datasets are detailed in Table 2. Briefly, UT-HAR and Widar were collected using Intel 5300 NIC,[1] and the NTU-Fi series of data were collected using the Atheros CSI tool.[35] 11 types of deep models are evaluated on these datasets using the three learning strategies. Eventually, detailed analytics are obtained on the convergence of the optimization, network depth, and network selection.

Evaluations of various deep architectures

Overall comparison. We summarize the performances of all baseline models in Table 3. On UT-HAR, ResNet-18 exhibits the best accuracy of 98.11%, followed by CNN-5. The shallow CNN-5 achieves good results on all datasets, but the deep networks, i.e., the ResNet series, do not exhibit significant improvement in comparison. In fact, ResNet-101 yields degenerating results on NTU-Fi. BiLSTM yields the best performance on the two NTU-Fi benchmarks. To compare these results, we visualize them in Figure 5, based on which we can derive the following conclusions.

- MLP, CNN-5, GRU, LSTM, and transformer all achieve good results on all benchmarks, indicating that these networks are suitable as feature extractors for WiFi CSI data.
- MLP, GRU, and CNN exhibit stable and superior performance compared with the other networks. CNN and GRU involve fewer parameters and lower computational complexity.
- Very-deep networks (such as the series of ResNets) do not consistently outperform CNN-5 and sometimes even perform slightly worse on the NTU-Fi benchmark. This suggests that shallow networks, such as CNN-5, are already sufficient as feature extractors for WiFi sensing.
- Vanilla RNN performs worse than LSTM and GRU because of its difficulty in capturing long-term patterns and susceptibility to gradient vanishing.
- The transformer architecture, i.e., ViT, does not operate satisfactorily when the size of training dataset is not adequate and the task is difficult, e.g., in NTU-Fi Human-ID. In this case, simple MLP outperforms ViT.
- As depicted in Figure 5, no model performs well on all datasets consistently, especially on the more difficult Widar dataset.

Computational complexity. The computational complexity of the models, as measured in terms of the floating-point operation (Flop) values presented in Table 3, is an important consideration during the selection of a model for WiFi sensing. The vanilla RNN exhibits low complexity but does not perform well. GRU and CNN-5 are the second-best performing models, achieving good results with relatively low computational complexity. It should be noted that the transformer (ViT) exhibits very high computational complexity owing to the attention computation,

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

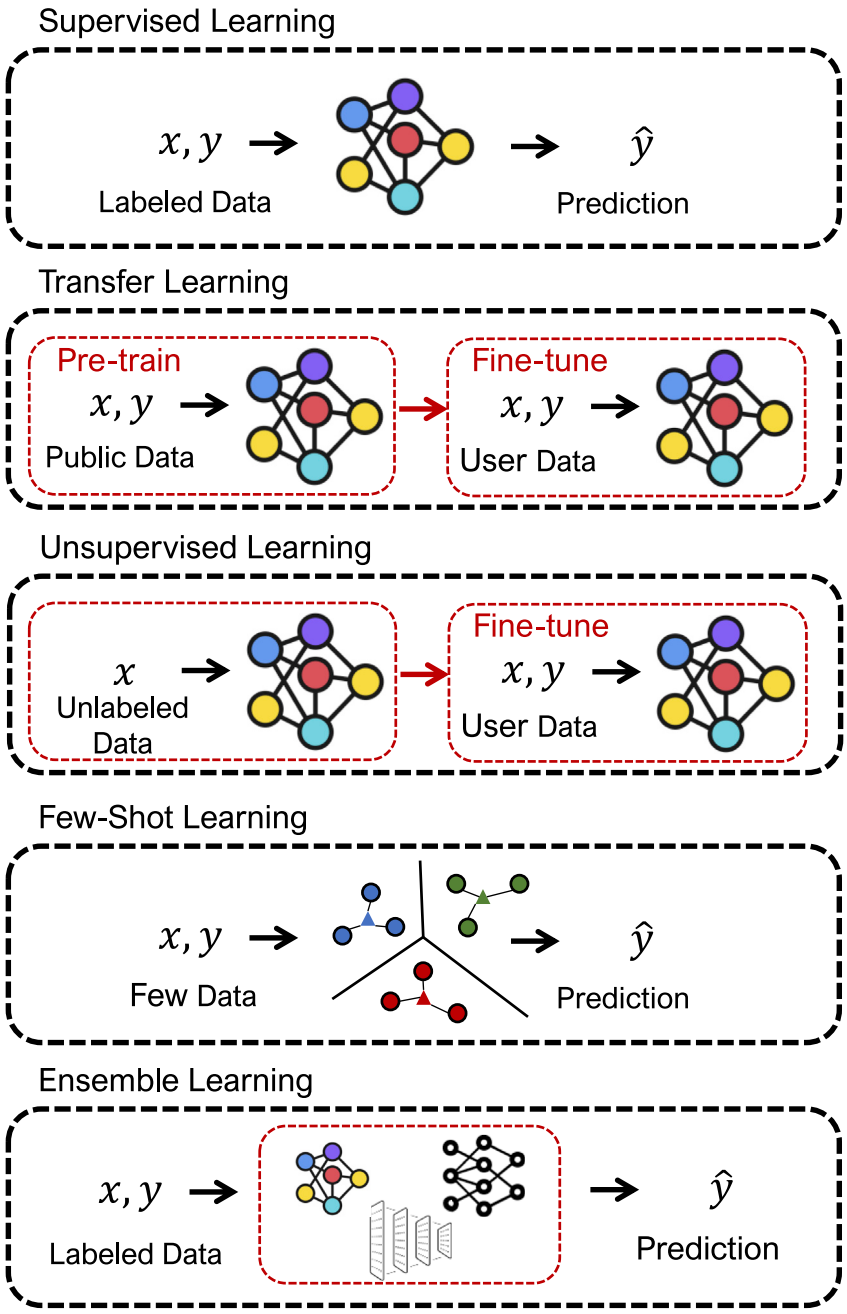


Figure 4. The illustration of the learning strategies

and its performance is similar to that of CNN, MLP, and GRU. Thus, its utilization for supervised learning tasks in WiFi sensing is inappropriate.

Model parameters. The number of model parameters also affects the amount of GPU memory required for the inference. As presented in Table 3, vanilla RNN involves the smallest number of parameters, followed by CNN-5 and CNN + GRU. The numbers of parameters of CNN-5, RNN, GRU, LSTM, BiLSTM, and CNN + GRU are all small and acceptable for model inference on the edge. When considering both the number of parameters and the accuracy, CNN-5, GRU, BiLSTM, and CNN + GRU are good choices for WiFi sensing. Although the number of model parameters can be reduced via techniques, such as model pruning,[85] quantisation,[85] or by fine-tuning hyperparameters, we only evaluate pure models with the minimum number of parameters required to converge in the training split in this study.

Evaluation of learning schemes

Apart from supervised learning, other learning schemes are also useful to realize robust and data-efficient WiFi-sensing applications. In this subsection, we evaluate two prevailing learning strategies in these models.

Evaluation using transfer learning. Transfer-learning experiments are conducted using NTU-Fi. The model is transferred from the HAR dataset to the Human-ID dataset by pre-training it on the entire HAR dataset and then fine-tuning a new classifier on the training split of Human-ID. This simulates a situation in which the model can be trained using a large amount of labeled data collected in a laboratory, and then a small amount of data can be used to customize the model for specific tasks for users. Human activities in the HAR dataset and human gaits in the Human-ID dataset are composed of human motions; therefore, the feature extractor should be able to generalize over these two tasks. We evaluate this setting for all baseline models, and the results are listed in Table 4. It is observed that the CNN-5 feature extractor exhibits the best transferability, achieving a score of 96.35% on the Human-ID task. The series of ResNets also exhibits superior performance; however, their model complexities are much higher. ResNet-18 outperforms ResNet-50 and ResNet-101, which suggests that greater depth may not always improve transferability in WiFi-sensing tasks. Similar to CNN-5, MLP and BiLSTM also exhibit good transferability. However, RNN, CNN + GRU, and ViT are observed to achieve scores of only 57.84%, 51.73%, and 66.20%, respectively, which demonstrates their weaker capacity for transfer learning. This could be attributed to overfitting, such as a simple RNN that only memorizes specific patterns on the HAR dataset but cannot recognize new patterns. This could also be attributed to the feature-learning mechanism. For example, the transformer (ViT) learns the connections between local patches via self-attention, but these connections may be different on the HAR and Human-ID datasets. Recognizing different activities often relies on detecting differences in a series of motions. However, most human gaits are very

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

Table 2. Statistics of four CSI datasets for our SenseFi benchmarks

Datasets	UT-HAR[36]	Widar[37]	NTU-Fi HAR[39]	NTU-Fi Human-ID[64]
Platform	Intel 5300 NIC	Intel 5300 NIC	Atheros CSI Tool	Atheros CSI Tool
Category number	7	22	6	14
Category names	lie down, fall, walk, pick up, run, sit down, stand up	push&pull, sweep, clap, slide, 18 types of draws	box, circle, clean, fall, run, walk	gaits of 14 subjects
Data size	(330,250) (antenna, subcarrier, packet)	(22,20,20) (time, x_velocity, y_velocity)	(3,114,500) (antenna, subcarrier, packet)	(3,114,500) (antenna, subcarrier, packet)
Training samples	3,977	34,926	936	546
Testing samples	996	8,726	264	294
Training epochs	200	100	30	30

similar, so only subtle patterns can be used as indicators for gait identification.

Evaluation using unsupervised learning. We further investigate the effectiveness of unsupervised learning for CSI feature learning. We adopt the self-supervised approach AutoFi[70] to construct two parallel networks and use Kullback-Leibler (KL) divergence, mutual information, and kernel density estimation loss to train the two networks using only unlabeled CSI data. After unsupervised learning, an independent classifier is trained based on the fixed parameters of the two networks. All backbone networks are tested using the same strategy—unsupervised training on NTU-Fi HAR and supervised learning (i.e., fine-tuning) on NTU-Fi Human-ID. The evaluation is performed using Human-ID, and the results are presented in Table 5. CNN-5 exhibits the best accuracy of 97.62%, followed by MLP and ResNet-18. Recurrent networks such as RNN, GRU, and LSTM do not perform well after unsupervised learning. These results demonstrate that unsupervised learning can help deep models learn discriminative features for CSI data—thus, CNNs are the most effective. Compared with the results of transfer learning, unsupervised learning yields better cross-task evaluation results, which suggests that unsupervised learning on a large dataset may aid model training on a smaller labeled dataset. CNNs and MLP-based networks are more suitable for unsupervised learning of WiFi CSI data.

Model analysis

Convergence of deep models. Although all the models converge eventually, their training difficulties vary and affect their practical usage. To compare their relative ease of convergence, we depict the training losses of MLP, CNN-5, ViT, and RNN in terms of epochs in Figure 6. It should be noted that the CNN converges very quickly within 25 epochs on all four datasets, and MLP also converges quickly. The transformer requires a higher number of training epochs owing to a greater number of model parameters. In comparison, RNN hardly converges on UT-HAR and Widar and converges more slowly on NTU-Fi. We further explore the convergence of RNN-based models, including GRU, LSTM, BiLSTM, and CNN + GRU in Figure 7. Although strong fluctuations are observed during the training phases of GRU, LSTM, and BiLSTM, these three models exhibit much lower training losses. In particular, GRU achieves the lowest loss among all RNN-based methods. The training phase of CNN + GRU is more stable, but its convergence loss is larger than those of the others.

The role of transfer learning. We further plot the training losses of all models on NTU-Fi Human-ID with pre-trained parameters obtained from NTU-Fi HAR in Figure 8. Compared with the training procedures of the randomly initialized models depicted in Figures 6C and 7C, convergence is achieved and becomes much more stable. Two conclusions are derived based on these results: (1) the feature extractors of these models are transferable across two similar tasks, and (2) the fluctuations of training losses are caused by the feature extractor because only the classifier is trained during transfer learning.

Comparison with traditional machine learning (TML). To demonstrate the superiority of deep-learning models, we evaluate the performance of TML methods on UT-HAR. UT-HAR contains low-dimensional CSI data; therefore, TML is expected to handle it more easily. Support vector machine (SVM), K-nearest neighbor (KNN), decision tree, random forest, and several configurations of naive Bayes are selected for comparison, and the results are listed in Table 6. The best model is observed to be random forest, achieving 87.75% accuracy on UT-HAR. It outperforms vanilla RNN and LSTM slightly. In particular, most deep-learning methods are observed to outperform TML methods by large margins. These results demonstrate that deep models are better suited to WiFi sensing than TML methods.

Poor performance of very-deep CNNs. As is evident from the data presented in Table 3, deeper ResNets do not perform better than the basic ResNet-18 in supervised learning, transfer learning, and unsupervised learning. In visual recognition, deeper networks perform better on large-scale datasets,[72] such as ImageNet. This raises the question of whether the poor performance of these deep models should be attributed to underfitting or overfitting in WiFi sensing. We attempt to answer this question by conducting an experiment on Widar and plotting its training and testing accuracy, as depicted in Figure 9. Figure 9A indicates that although the training accuracies of the three ResNets reach almost 100%, their testing accuracies remain low. This indicates that the poor performance of ResNets may be attributed to overfitting induced by domain shifting due to different environments in Widar.[37] As CSI captures both moving objects and the environment, a distribution shift is induced when the training and testing environments are different. The generalizability of deep models can be improved further via domain adaptation.[86] In addition, overfitting leads to unstable testing performance of deep models.

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

Table 3. Evaluation of deep neural networks (using supervised learning) on four datasets

Dataset	UT-HAR			Widar			NTU-Fi HAR			NTU-Fi Human-ID		
Method	Acc (%)	Flops (M)	Params (M)	Acc (%)	Flops (M)	Params (M)	Acc (%)	Flops (M)	Params (M)	Acc (%)	Flops (M)	Params (M)
MLP	92.00	23.17	23.170	67.24	9.15	9.150	99.69[a]	175.24	175.240	93.91	175.24	175.240
CNN-5	97.61[b]	31.68	0.296[b]	70.19[b]	3.38	0.299	98.70	28.24[b]	0.477	97.14	28.24[b]	0.478
ResNet18	98.11[a]	49.93	11.180	71.70[a]	38.39	11.250	95.31	54.19	11.180	96.42	54.19	11.190
ResNet50	97.21	86.40	23.550	68.56	69.70	23.640	99.38[b]	90.66	23.550	92.91	90.67	23.570
ResNet101	94.99	162.58	42.570	68.71	145.87	42.660	95.31	166.83	42.570	88.40	166.85	42.590
RNN	83.53	2.51[a]	0.010[a]	47.05	0.66[a]	0.031[a]	84.64	13.09[a]	0.027[a]	89.30	13.09[a]	0.027[a]
GRU	94.18	7.60[b]	0.030	62.50	1.98[b]	0.091[b]	97.66	39.39	0.079	98.96[b]	39.39	0.079
LSTM	87.18	10.14	0.040	63.35	2.64	0.121	97.14	52.54	0.105	97.19	52.54	0.105
BiLSTM	90.19	20.29	0.080	63.43	5.28	0.240	99.69[a]	105.09	0.209	99.38[a]	105.09	0.210
CNN + GRU	96.72	39.99	1.430	63.19	3.34	0.092	93.75	48.38	0.058[b]	87.48	48.39	0.058[b]
ViT	96.53	273.10	10.580	67.72	9.28	0.106	93.75	501.64	1.052	76.84	501.64	1.054

[a]Best.
[b]Second best.

The testing accuracy of ResNet-101 is observed to undergo greater fluctuation than that of ResNet-18, whereas other networks (i.e., MLP, CNN, GRU) converge more easily with more stable testing accuracies. This indicates that very-deep networks are prone to overfitting in cross-domain tasks in WiFi sensing. Thus, they may not be a good choice for current WiFi-sensing applications because of their performance and computational overhead. The varied performances of very-deep models also make the selection of an appropriate model for practical deployment more challenging.

Evaluation of more deep models. In addition to the aforementioned networks mentioned in our benchmark, some more commonly used models that are used in deep-learning applications are also evaluated on NTU-Fi Human-ID and Widar. These models include the first deep CNN (AlexNet),[87] deep models with intermediary depth (VGG-16 and VGG-19),[88] an inception model (GoogleNet),[89] and a small network tailored for low-computational devices (EfficientNet-b0).[90] The results are listed in Table 7. These models achieve accuracies exceeding 93% on NTU-Fi Human-ID and exceeding 63% on Widar. It is noteworthy that EfficientNet-b0 performs satisfactorily on both datasets with small numbers of flops and parameters. This conclusion reinforces a previous one—very large models may not be necessary for WiFi-sensing applications.

Choices of optimizer. Although the adoption of the Adam optimizer hastens the convergence of models during training, it also leads to considerable training instability, especially for very-deep neural networks. As depicted in Figure 10A, ResNet-18 converges stably, but ResNet-50 and ResNet-101 exhibit fluctuating losses every 20–30 epochs. This may be attributed to the rapidly changing values of WiFi data and Adam's adaptive learning rate.[91] As an alternative to Adam, the more stable optimizer, stochastic gradient descent (SGD), is considered. As is evident from Figure 10B, this makes the training procedure more stable. This implies that SGD is a better choice when very-deep models are implemented in WiFi sensing. On the other hand, the Adam optimizer improves and hastens the convergence of simple models that are sufficient for sensing tasks.

DISCUSSION

Summary of benchmarks and recommendations

Based on the analysis of the empirical results and characteristics of deep-learning models in the context of WiFi sensing, the following experiences and observations are summarized that are expected to facilitate future research on model design, model training, and real-world applications.

- Model choices. We recommend CNN, GRU, and BiLSTM due to their high performance, low computational cost, and small number of parameters. These shallow models are observed to achieve remarkable results in activity recognition, gesture recognition, and Human-ID, in contrast to very-deep models, which are affected by overfitting, especially in cross-domain scenarios.
- Optimization. We recommend using the Adam or SGD optimizer. The Adam optimizer can help the model converge quickly, but it causes instability during training in some cases. In such cases, SGD is a more secure option, but the hyperparameters of SGD (i.e., the learning rate and momentum) require manual specification and tuning.
- Recommended transfer-learning applications. We recommend utilizing transfer learning when the task is similar to existing applications and when a single CSI sensing platform is used. The pre-trained parameters provide good initialization and better generalizability. CNN, MLP, and BiLSTM exhibit superior transferability.
- Recommendations regarding unsupervised learning. We recommend applying unsupervised learning to initialize the model for similar tasks since it extracts more generalizable features than transfer learning. In general, CNN, MLP, and ViT are more suitable in the unsupervised learning framework.

Grand challenges and future directions

Deep learning continues to thrive in many research fields and is constantly empowering more challenging applications and

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

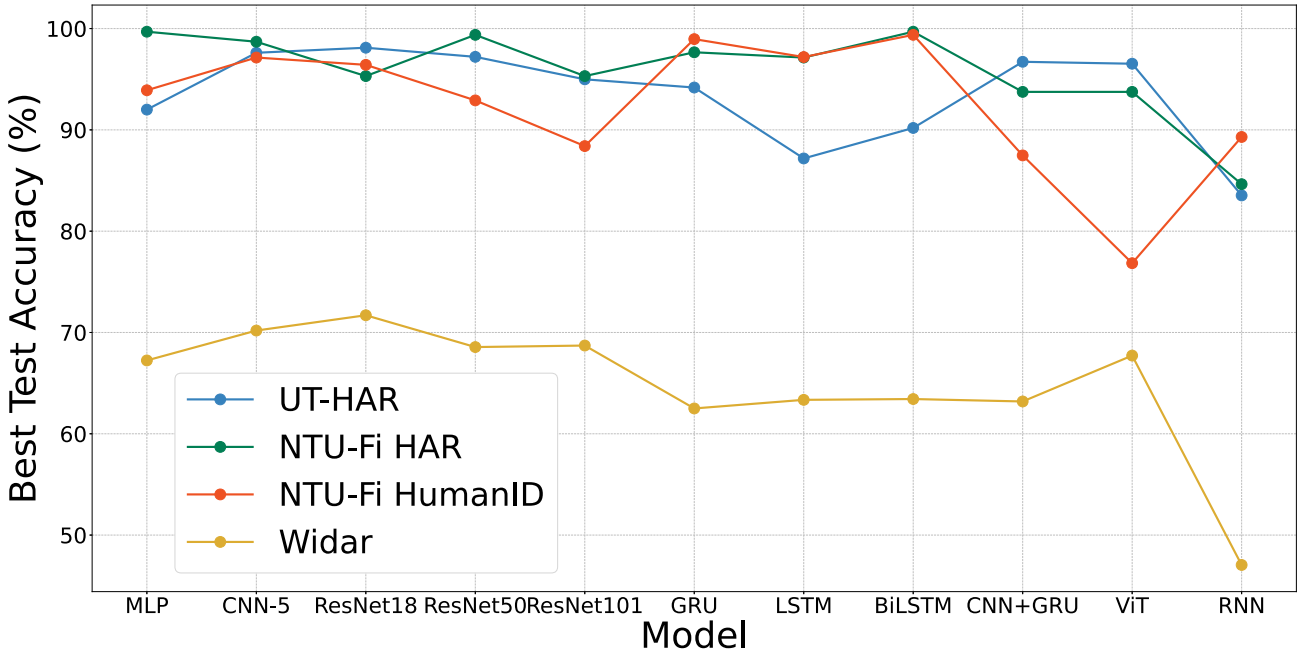


Figure 5. The performance comparison across four datasets

scenarios in WiFi sensing. Based on the new progress, we look into the future directions of deep-learning-empowered WiFi sensing and summarize them as follows.

Data-efficient learning

As CSI data are expensive to collect, data-efficient learning methods should be further explored. Existing works have utilized few-shot learning, transfer learning, and domain adaptation, which yield satisfactory results in a new environment with limited training samples. However, since the testing scenarios are simple, the transferability of these models cannot be well evaluated. In the future, meta-learning and zero-shot learning could further help learn robust features across environments and tasks.

Model compression or lightweight model design

In the future, it is necessary to achieve real-time computation for certain applications of WiFi sensing, such as vital sign monitoring.[92] Therefore, model compression techniques will play a crucial role, such as model pruning,[85] quantization[?], and distillation,[93] which decrease the model size via an extra learning step. The design of lightweight model, such as the EfficientNet[90] for visual recognition, is also favorable. It aims to construct a model from scratch by balancing network depth, width, and resolution.

Multimodal learning

WiFi sensing is ubiquitous, cost effective, and privacy preserving and can work without the effect of illumination and part of occlusion, which is complementary to existing visual sensing techniques. To achieve robust sensing 24/7, multiple modalities of sensing data should be fused using multimodal learning. WiVi[7] pioneers HAR by integrating WiFi sensing and visual recognition. Multimodal learning can learn joint features from

Table 4. Evaluations on transfer learning

Method	Accuracy (%)	Flops (M)	Params (M)
MLP	84.46	175.24	175.240
CNN-5	96.35[a]	28.24[b]	0.478
ResNet18	85.94[b]	54.19	11.190
ResNet50	79.21	90.67	23.570
ResNet101	68.88	166.85	42.590
RNN	57.84	13.09[a]	0.027[a]
GRU	75.89	39.39	0.079
LSTM	71.98	52.54	0.105
BiLSTM	80.20	105.09	0.210
CNN + GRU	51.73	48.39	0.059[b]
ViT	66.20	501.64	1.054

[a]Best.
[b]Second best.

Table 5. Evaluations on unsupervised learning

	Accuracy (%)			
Method	Classifier1	Classifier2	Flops (M)	Params (M)
MLP	90.48[a]	89.12[a]	175.24	175.240
CNN-5	96.26[b]	97.62[b]	28.24[a]	0.478
ResNet18	85.03[b]	82.99	54.19	11.190
ResNet50	47.28	45.58	90.67	23.570
ResNet101	36.05	35.37	166.85	42.590
RNN	53.74	51.36	13.09[b]	0.027[b]
GRU	65.99	64.63	39.39	0.079
LSTM	53.06	55.10	52.54	0.105
BiLSTM	51.36	55.78	105.09	0.210
CNN + GRU	50.34	53.40	48.39	0.059[a]
ViT	78.91	84.35	501.64	1.054

[a]Second best.
[b]Best.

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

CellPress
OPEN ACCESS

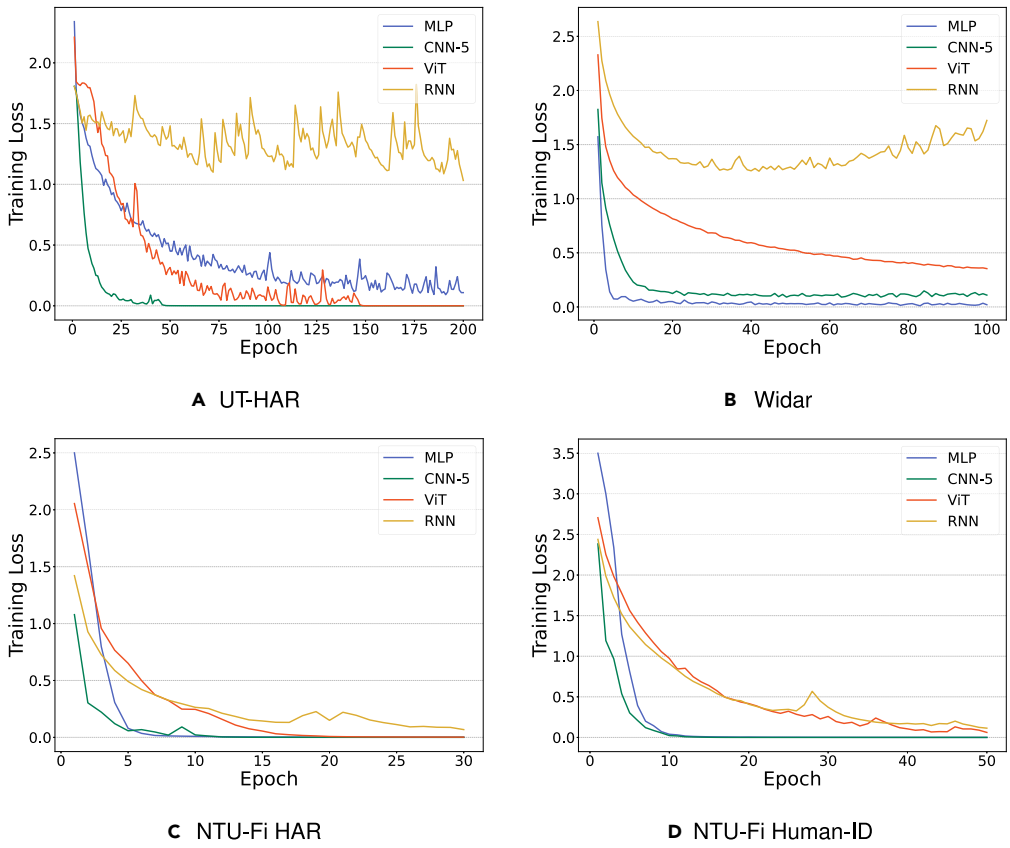


Figure 6. The training losses of MLP, CNN, transformer, and RNN for the four datasets
(A) UT-HAR.
(B) Widar.
(C) NTU-Fi HAR.
(D) NTU-FI Human-ID.

multiple modalities and make decisions by choosing reliable modalities.

Cross-modal learning

Cross-modal learning aims to supervise or reconstruct one modality from another modality, which helps WiFi truly "see" the environment and visualize it in other modalities. Wi2Vi[94] manages to generate video frames from CSI data by transferring knowledge from video to WiFi. The human pose is then estimated by supervising the model with the results of pose estimation, such as OpenPose.[95] In the future, cross-modal learning may enable the WiFi model to learn from other supervision sources such as radar and Lidar.

Model robustness and security for trustworthy sensing

When deploying WiFi-sensing models in the real world, the model should be secure to use. Adversarial attacks have received attention in video-based human sensing.[96] However, existing WiFi-sensing works study the accuracy of models, but few pay attention to the security issue. First, during communication, the sensing data may leak the privacy of users. Second, if any adversarial attack is made on the CSI data, the model can perform incorrectly and trigger the wrong actions of smart appliances. SecureSense[68] seeks to overcome adversarial attacks through augmentation and adversarial training. EfficientFi[39] proposes a variational autoencoder to quantize the CSI for efficient and robust communication. WiFi-ADG[97] protects user privacy by enforcing that the data are not recognizable by general classifiers. More work should be focused on the safety of WiFi sensing and trustworthy models for large-scale sensing, such as federated learning.

Complicated human activities and behaviors analytics

While current methods have shown good recognition accuracy for single activities or gestures, human behavior is depicted by more complicated activities. For example, to indicate if a patient may have a risk of Alzheimer's disease, the model should record the routine and analyze the anomaly activity, which is still difficult for existing approaches. Precise user behavior analysis can contribute to daily healthcare monitoring and behavioral economics.

Model interpretability for a physical explanation

Model-based and learning-based methods are developing quickly, but in different ways. In WiFi sensing, there could be a connection between the data-driven model and the physical model based on model interpretability research, which may inspire us to develop new theories of physical models for WiFi sensing. In contrast, the existing model (e.g., Fresnel Zone) may enable us to propose new learning methods based on

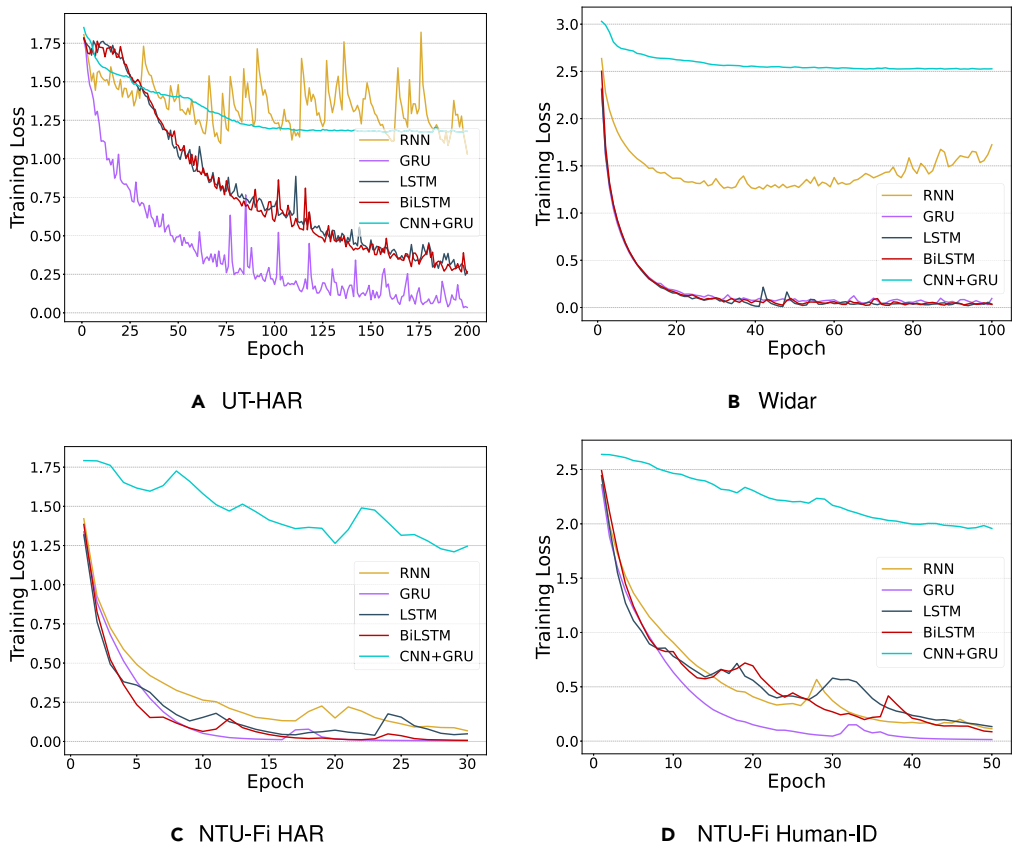


Figure 7. The training losses of RNN-based models for the four datasets
(A) UT-HAR.
(B) Widar.
(C) NTU-Fi HAR.
(D) NTU-Fi Human-ID.

physical models. It is hoped that these two directions of methods can be unified theoretically and practically.

Concluding remarks

Deep-learning methods have been shown to be effective for challenging applications in WiFi sensing, but these models exhibit different characteristics on WiFi-sensing tasks, and a comprehensive benchmark is highly needed. To this end, this work illustrates the recent progress in deep learning for WiFi human sensing and benchmarks current deep neural networks and deep-learning strategies on WiFi CSI data across different platforms. We summarize the conclusions drawn from the experimental observations, which provide valuable experiences for model design in practical WiFi-sensing applications. Finally, we propose grand challenges and future directions to address the research issues emerging from future large-scale WiFi-sensing scenarios.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and materials should be directed to and will be fulfilled by the lead contact, Dr. Jianfei Yang (yang0478@ntu.edu.sg).

Materials availability

All the well-trained model weights in this benchmark have been deposited in an online drive.

Data and code availability

- The codes generated during this study have been deposited in GitHub and Zenodo, and a tutorial is provided to guide the users. Any additional information required to reproduce this work is available on Github, https://github.com/xyanchen/WiFi-CSI-Sensing-Benchmark. The Zenodo link is https://doi.org/10.5281/zenodo.7501869.
- The article analyzes existing public data and releases new data. All the processed datasets and well-trained model parameters are available in https://data.mendeley.com/datasets/dzvgyxkx2f/draft?a=9ff61de8-9565-4543-8278-8072329a0a16.

Method details of deep models in benchmarks

In the following, we introduce these key architectures and how they are applied to WiFi-sensing tasks. To better instantiate these networks, we first formulate the normal WiFi CSI sensing task. The CSI data are defined as $x \in \mathbb{R}^{N_a \times N_s \times T}$, where N_a denotes the number of antennas, N_s denotes the number of subcarriers, and T denotes the duration. The CSI data of each pair of antennas are relatively independent and are thus regarded as one channel of CSI, similar to the RGB channels of image data. The deep-learning model $f(\cdot)$ aims to map the data to the corresponding label: $y = f(x)$, according to different tasks. We denote $\Phi_i(\cdot)$ and z_i as the i-th layer of the deep model and the feature of the i-th layer, respectively. After feature extraction, the classifier is trained to seek a decision boundary in the feature space. In deep learning, the feature extractor is the key module that reduces the feature dimension while preserving the

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

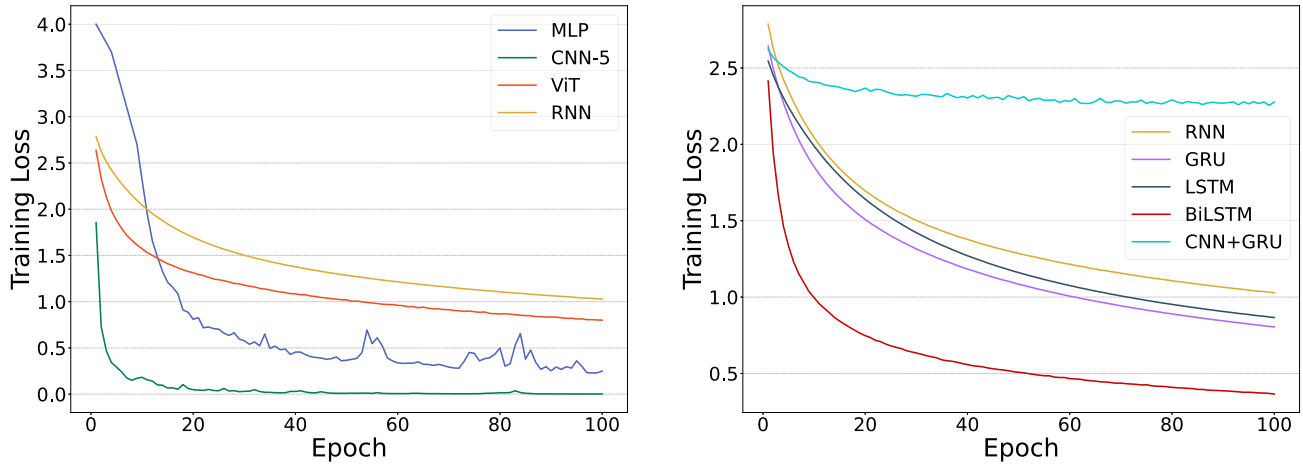


Figure 8. The training losses of all baseline models on NTU-Fi Human-ID with pre-trained parameters of NTU-Fi HAR

manifold.[44] With a discriminative feature space, the choices of classifiers are flexible and can be either deep classifiers (e.g., MLP) or traditional classifiers (e.g., KNNs, SVM, and random forest). In addition to this illustration, we visualize the intuition of how to feed these CSI data into various networks in Figure 3.

MLP

MLP[71] is one of the most classic architectures and has played the role of classifier in most deep classification networks. It typically consists of multiple fully connected layers followed by activation functions. The first layer is called the input layer, which transforms the input data into the hidden latent space, and after several hidden layers, the last layer maps the latent feature into the categorical space. Each layer is calculated as

$$\Phi_i(z_{i-1}) = \sigma(W_i z_{i-1}), \quad \text{(Equation 3)}$$

where W_i is the parameter of Φ_i and $\sigma(\cdot)$ is the activation function that aims to increase the non-linearity of the MLP. The input CSI has to be flattened to a vector and then fed into the MLP such that $x \in \mathbb{R}^{N_s T}$. This process mixes the spatial and temporal dimensions and damages the intrinsic structure of the CSI data. Despite this, the MLP can still work with a large amount of labeled data because the MLP has a fully connected structure with a large number of parameters. However, this leads to slow convergence and high computational costs. Therefore, although the MLP can achieve satisfactory performance, it is not common to stack many layers in an MLP for feature learning, which makes the MLP typically serve as a classifier. In WiFi sensing, MLP is commonly used as a classifier.[5,7,34,37,50,52]

CNN

CNN was first proposed for image recognition tasks by LeCun.[98] It addresses the drawbacks of MLP through weight sharing and spatial pooling.

Table 6. Evaluations of traditional methods on UT-HAR

Method	Accuracy (%)
Support vector machines	86.95[a]
K-nearest neighbors	84.04
Decision tree	64.56
Random forest	87.75[b]
Multinomial naive Bayes	33.94
Gaussian naive Bayes	46.39
Complement naive Bayes	30.62
Bernoulli naive Bayes	29.42

[a]Second best.
[b]Best.

CNN models have achieved remarkable performances in classification problems of 2D data in computer vision[99,100] and sequential data in speech recognition[101] and natural language processing.[102] CNN learns features by stacking convolutional kernels and spatial pooling operations. The convolution operation refers to the dot product between a filter $\mathbf{k} \in \mathbb{R}^d$ and an input vector $\mathbf{v} \in \mathbb{R}^d$, defined as follows:

$$\mathbf{k} \otimes \mathbf{v} = \sigma(\mathbf{k}^T \mathbf{v}). \quad \text{(Equation 4)}$$

The pooling operation is a down-sampling strategy that calculates the maximum (max pooling) or mean (average pooling) inside a kernel. CNNs typically consist of several convolutional layers, max-pooling layers, and an MLP classifier. In general, increasing the depth of CNNs can lead to better model capacity. However, when the depth of CNN is too large (e.g., greater than 20 layers), the gradient vanishing problem leads to degrading performance. This degradation is addressed by ResNet,[72] which uses residual connections to reduce the difficulty of optimization.

In WiFi-sensing tasks, the convolution kernel can operate on a 2D patch of CSI data (i.e., Conv1D) that includes a spatial-temporal feature or on a 1D patch of each subcarrier of CSI data (i.e., Conv2D). For Conv2D, a 2D convolution kernel $\mathbf{k}_{2D} \in \mathbb{R}^{h \times w}$ operates on all patches of the CSI data via the sliding window strategy to obtain the output of the feature map, while Conv1D only extracts the spatial feature along the subcarrier dimension. Conv2D can be applied independently as it considers both spatial and temporal features, while Conv1D is usually used with other temporal feature-learning methods. To enhance the capacity of CNN, multiple convolution kernels with random initialization are used. The advantages of CNNs for WiFi sensing include fewer training parameters and the preservation of the subcarrier and time dimension in CSI data. However, the disadvantage is that CNN has a small receptive field due to the limited kernel size and thus fails to capture dependencies that exceed the kernel size. Another drawback is that CNN stacks all the feature maps of kernels equally, which has been revamped by an attention mechanism that assigns different weights at the kernel or spatial level while stacking features. For CSI data, due to the varying locations of human motions, the patterns of different subcarriers should have different importance, which can be depicted by spatial attention.[65] More attention techniques have been successfully developed to extract temporal-, antenna-, and subcarrier-level features for WiFi sensing.[54,103,104]

RNN

RNN is one of the deepest network architectures that can memorize arbitrary-length sequences of input patterns. The unique advantage of RNN is that it enables multiple inputs and multiple outputs, making it very effective for time sequence data, such as video[105] and CSI.[5,106,107] Its principle is to create an internal memory to store historical patterns, which are trained via backpropagation through time.[73]

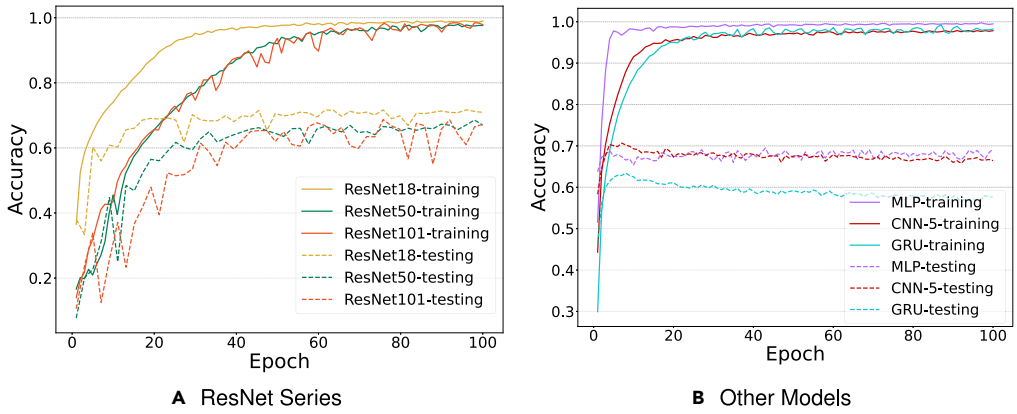


Figure 9. The training procedures of deep learning models on Widar data in terms of training and testing accuracy

(A) ResNet series.

(B) Other models.

For a CSI sample P, we denote a CSI frame at the t as $x_t \in \mathbb{R}^{N_s}$. The vanilla RNN uses two sharing matrices W_x, W_h to generate the hidden state h_t:

$$h_t = \sigma(W_x x_t + W_h h_{t-1}), \quad \text{(Equation 5)}$$

where the activation function $\sigma(\cdot)$ is usually tanh or sigmoid functions. RNN is designed to capture temporal dynamics, but it suffers from the vanishing gradient problem during backpropagation and thus cannot capture long-term dependencies of CSI data.

Variants of RNN (LSTM)

To tackle the problem of long-term dependencies of RNN, LSTM[108] is proposed by designing several gates with varying purposes and mitigating the gradient instability during training. The standard LSTM sequentially updates a hidden sequence by a memory cell that contains four states: a memory state c_t, an output gate o_t that controls the effect of output, an input gate i_t, and a forget gate f_t that decides what to preserve and forget in the memory. The LSTM is parameterized by weight matrices $W_i, W_f, W_c, W_o, U_i, U_f, U_c, U_o$ and biases b^i, b^f, b^c, b^o, and the whole update is performed at each $t \in \{1, \ldots, T\}$:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b^i), \quad \text{(Equation 6)}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b^f), \quad \text{(Equation 7)}$$

$$\tilde{c}_t = tanh(W_c x_t + U_c h_{t-1} + b^c), \quad \text{(Equation 8)}$$

$$c_t = i_t \odot \tilde{c}_t + f_t \odot c_{t-1}, \quad \text{(Equation 9)}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b^o), \quad \text{and (Equation 10)}$$

$$h_t = o_t \odot tanh(c_t), \quad \text{(Equation 11)}$$

where σ is a sigmoid function.

In addition to the LSTM cell,[109–111] the use of a multilayer and bidirectional structure can further increase the model's capacity. The BiLSTM model processes the sequence in two directions and concatenates the features of the forward input x and backward input $\dot{x}$. It has been demonstrated that BiLSTM performs better than LSTM in Chen et al.[51] and Kadir et al.[112]

Recurrent CNN

While LSTM addresses long-term dependencies, it can lead to a large computation overhead. To address this issue, the GRU was introduced. GRU combines the forget and input gates into one gate and does not use the memory state present in LSTM, simplifying the model while still being able to capture long-term dependencies. GRU is considered a simple yet effective version of LSTM. By using a simple recurrent network, we can integrate Conv1D and GRU to extract spatial and temporal features, respectively. Studies in Dua et al.[113] and Chen et al.[114] have shown that a CNN + GRU model is effective for HAR. In WiFi sensing, DeepSense[5] was the first to propose using Conv2D with LSTM for HAR. CNN + GRU has also been used for CSI-based human gesture recognition in Widar.[37] SiaNet[9] further proposed using Conv1D with BiLSTM for few-shot gesture recognition. As these models perform similarly, we use the CNN + GRU model with fewer parameters as a benchmark in this article.

Transformers

Transformer[74] was initially proposed for natural language processing (NLP) applications to extract sequence embeddings using attention between words and was later extended to the computer vision field where each patch is treated as a word and an image is composed of multiple patches.[115] The vanilla transformer consists of an encoder and a decoder used for machine translation, but we only need the encoder. The transformer block consists of a multihead attention layer, a feedforward neural network (MLP), and layer

Table 7. Evaluations of more deep models using supervised learning

Dataset	NTU-Fi Human-ID			Widar		
Method	Acc (%)	Flops (M)	Params (M)	Acc (%)	Flops (M)	Params (M)
MLP	93.91	175.24	175.240	67.24	9.15	9.150
CNN-5	97.14	28.24[a]	0.478	70.19	3.38	0.299
AlexNet	95.92	736.25	57.060	68.74	64.70	23.350
VGG16	95.61	15.52k	134.320	66.10	317.09	14.740
VGG19	93.32	19.69k	139.630	63.67	402.13	20.050
EfficientNet	93.95	422.44	4.030	71.14[a]	34.08	3.610
GoogLeNet	96.13	1.54k	9.970	64.31	1.54k	6.180
ResNet18	96.42	54.19	11.190	73.49[b]	38.39	11.250
ResNet50	93.21	90.67	23.570	69.40	69.70	23.640
ResNet101	92.89	166.85	42.590	67.59	145.87	42.660
RNN	89.77	13.09[b]	0.027[b]	46.77	0.66[b]	0.031[b]
GRU	98.96[a]	39.39	0.079	62.50	1.98[a]	0.091[a]
LSTM	97.17	52.54	0.105	63.35	2.64	0.121
BiLSTM	99.38[b]	105.09	0.210	63.43	5.28	0.240
CNN + GRU	90.82	48.39	0.058[a]	61.21	3.34	0.092
ViT	78.27	501.64	1.054	64.85	9.28	0.106

[a]Second best.

[b]Best.

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

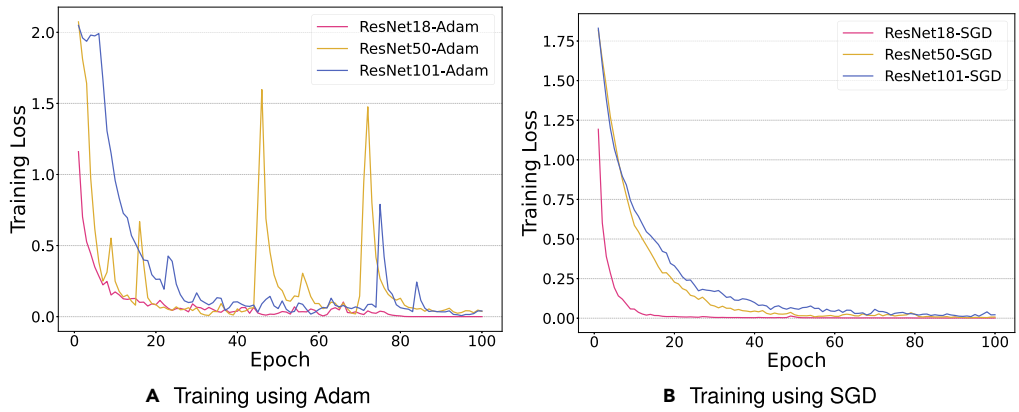


Figure 10. The training procedures of ResNet-18/50/101 using Adam and SGD optimizers on UT-HAR

(A) Training using Adam.

(B) Training using SGD.

normalization. Since MLP was explained in the previous section, we will mainly introduce the attention mechanism in this section. For a CSI sample P, we divide it into P patches $x_p \in \mathbb{R}^{h \times w}$, of which each patch has contained spatial-temporal features. Then, these patches are concatenated and added by positional embeddings that infer the spatial position of patches, which makes the input matrix $\in \mathbb{R}^{d_k}$, where $d_k = P \times hw$. This matrix is transformed into three different matrices via linear embedding: the query Q, the key K, and the value V. The self-attention process is calculated by

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V. \quad \text{(Equation 12)}$$

Intuitively, this process calculates the attention between any two patches using the dot product (i.e., cosine similarity) and then normalizes the weights to improve gradient stability during training. Multihead attention simply repeats the self-attention process several times to increase the diversity of attention. The transformer architecture can connect with every patch of CSI, making it strong when given sufficient training data, such as in Li et al.[62] However, the transformer has a large number of parameters that can make training expensive, and it is difficult to collect large amounts of labeled CSI data, making it less attractive for supervised learning.

Generative models

Unlike the aforementioned discriminative models that primarily conduct classification, generative models aim to capture the distribution of CSI data. Generative adversarial network (GAN)[116] is a classic generative model that learns to generate data that appears real through an adversarial game between a generative network and a discriminator network. In WiFi sensing, GAN can help deal with environmental dependency by generating labeled samples in the new environment from based on a well-trained model in a different environment.[53,64] GAN has also inspired domain-adversarial training, which enables deep models to learn domain-invariant representations for both the training and real-world testing environments.[27,117–119] Variational network[120] is another common generative model that maps the input variable to a multivariate latent distribution. Variational autoencoder (VAE) learns the data distribution by a stochastic variational inference and learning algorithm[120] and has been used in CSI-based localization[121,122] and CSI compression.[39] For instance, EfficientFi[39] leverages a quantized variational model to compress the CSI transmission data for large-scale WiFi sensing in the future.

Datasets

We use two public CSI datasets (UT-HAR[36] and Widar[37]) collected using Intel 5300 NICs and two new datasets (NTU-Fi HAR and NTU-Fi Human-ID) collected using Atheros CSI Tool[35] and our embedded Internet of Things (IoT) system[25] to validate the effectiveness of deep-learning models on CSI data from different platforms. The statistics of these datasets are summarized in Table 2.

UT-HAR[36] is a public CSI dataset for HAR that consists of seven categories. It was collected using an Intel 5300 NIC with three pairs of antennas that record 30 subcarriers per pair. All the data were collected in the same environment, but they were collected continuously and do not have golden labels for activity segmentation. Following existing works,[62] the data were segmented using a sliding window, which inevitably leads to many repeated data among samples. Although the total number of samples reaches around 5,000, it is a small dataset with intrinsic limitations.

Widar[37] is the largest WiFi-sensing dataset for gesture recognition, consisting of 22 categories and 43,000 samples. It was collected using an Intel 5300 NIC with a 3×3 array of antenna pairs in many distinct environments. To eliminate environmental dependencies, the data were processed into the BVP.

NTU-Fi is our proposed dataset for this benchmark, including both HAR and Human-ID tasks. Different from UT-HAR and Widar, our dataset was collected using the Atheros CSI Tool and has a higher resolution of subcarriers (114 per pair of antennas). Each CSI sample is segmented by a threshold and a specific time duration to ensure that the entire activity is included. For the HAR dataset,

Table 8. Optimizer and hyperparameters of the experiments

	Training stage	Batch size	Optimizer	lr	weight_decay
Supervised learning	–	64	SGD	1e−03	0
	–	64	Adam	1e−03	0
Unsupervised learning	self-supervised training	64	AdamW	1e−03	1.5e−06
	supervised training	64	Adam	1e−03	1e−05
Transfer learning	encoder training	64	Adam	1e−03	0
	classification header training	64	Adam	1e−03	0

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

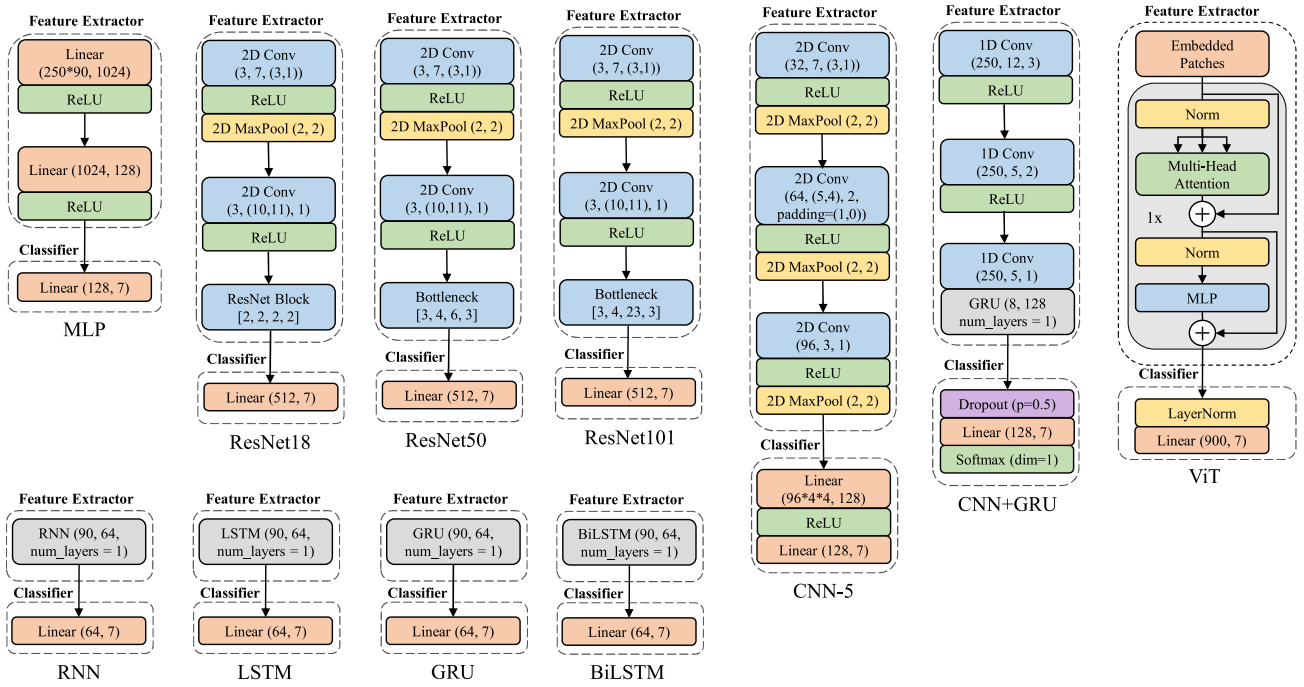


Figure 11. The network architectures used in UT-HAR experiments

we collected data in three different layouts. For the Human-ID dataset, we collected human walking gaits in three situations: wearing a t-shirt, a coat, or a backpack, which presents many challenges. The NTU-Fi data were collected simultaneously in Wang et al.[12] and Yang et al.[39] which provide detailed descriptions of the data collection layouts.

Implementation details

We normalize the data for each dataset using min-max normalization and implement all the aforementioned methods using the PyTorch framework.[123] To ensure convergence, we train the models on UT-HAR, Widar, and NTU-Fi for 200, 100, and 30 epochs, respectively. As the vanilla RNN is hard to

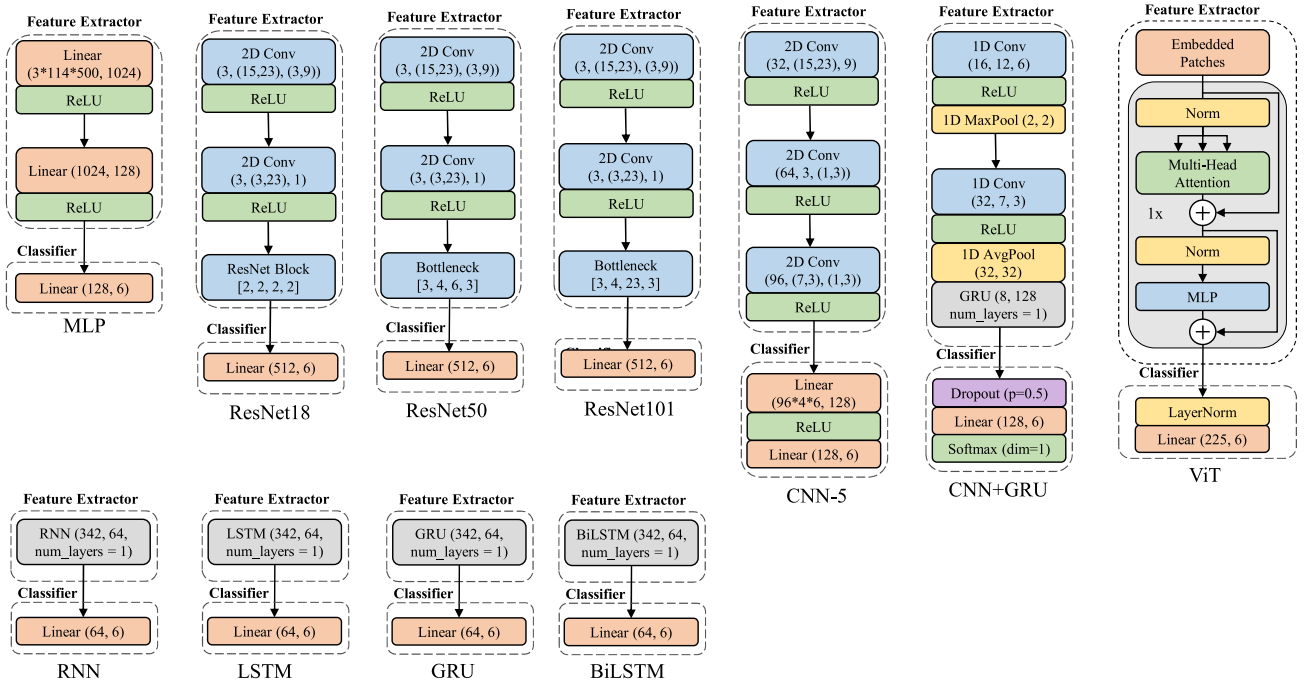


Figure 12. The network architectures used in NTU-Fi HAR experiments

Please cite this article in press as: Yang et al., SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100703

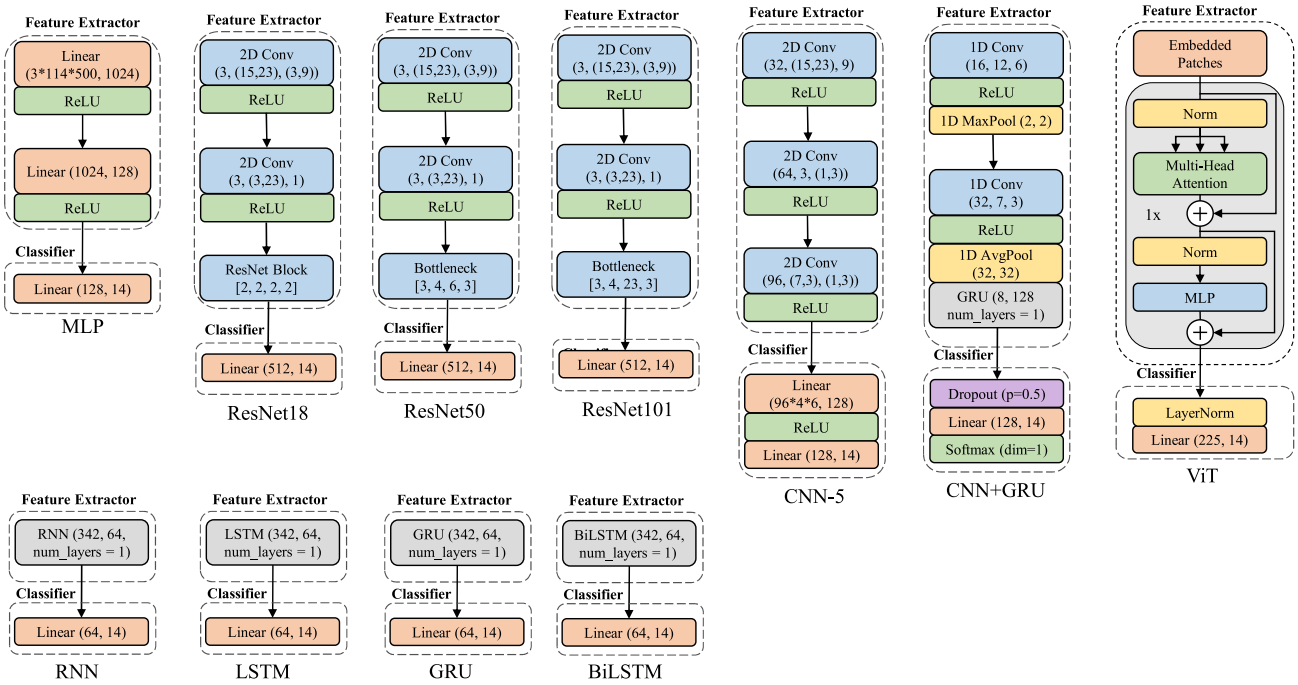


Figure 13. The network architectures used in NTU-Fi Human-ID experiments

converge due to the gradient vanishing, we train it for more epochs. We use the Adam optimizer with a learning rate of 0.001 and betas of 0.9 and 0.999, as recommended in the original Adam paper.[91] The ratio of training and testing splits is 8:2 for all datasets, using stratified sampling. We present all the hyperparameters in Table 8. The results reported in the benchmark are all obtained using the benchmark codes, and the accuracy is the mean value of three independent runs with random seeds.

Baselines and criterion

As baselines, we design the MLP, CNN, RNN, GRU, LSTM, BiLSTM, CNN + GRU, and transformer networks based on experiences from existing works (in Table 1). The CNN-5 is modified from LeNet-5,[98] and we also introduce a series of ResNets[72] with deeper layers. The transformer network is based on the ViT[115] and allows each patch to contain spatial and temporal dimensions. It is found that, given sufficient parameters and reasonable depth of layers,

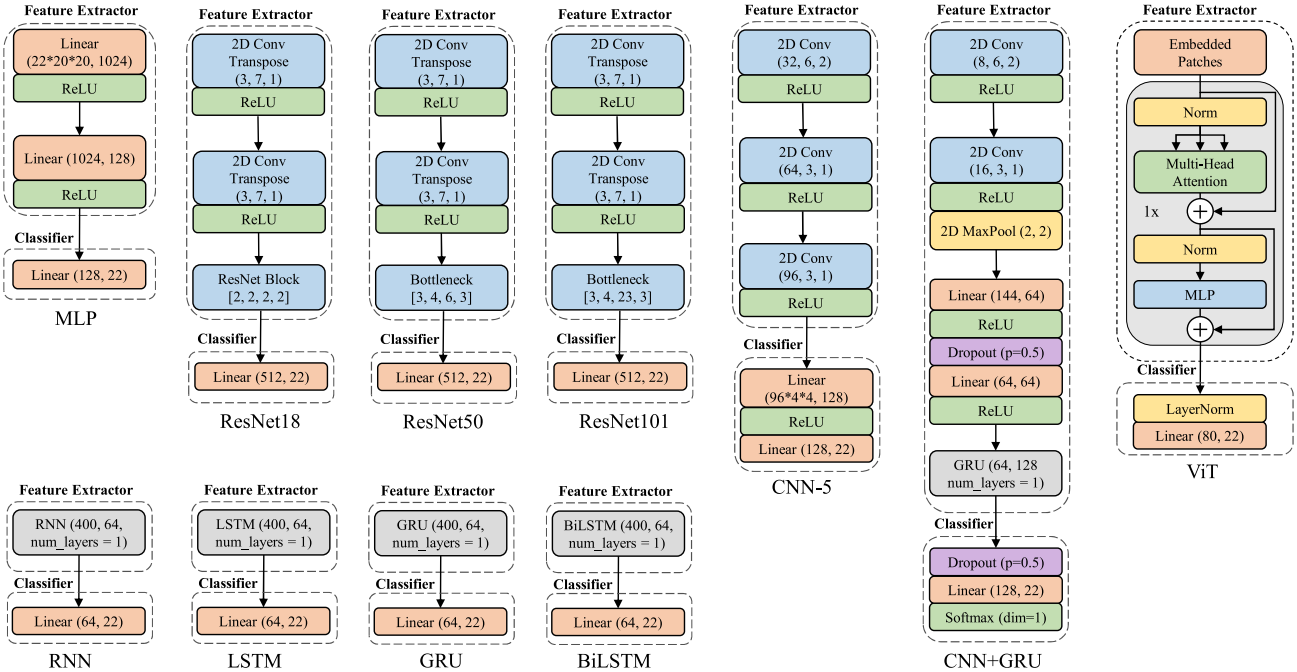


Figure 14. The network architectures used in Widar experiments

these networks can converge to more than 98% accuracy on the training split. Since the data sizes of UT-HAR, Widar, and NTU-Fi are different, we use a convolutional layer to map them to a unified size, enabling us to use the same network architecture for all datasets. The specific network architectures for all models are illustrated in Figures 11, 12, 13, and 14. The hyperparameters of the networks have been tuned to ensure satisfactory convergence, i.e., over 98% training accuracy. To compare the baseline models, we use three criteria: accuracy (Acc), which evaluates the prediction ability; Flops, which evaluates the computational complexity; and the number of parameters (Params), which measures the requirement for GPU memory. In WiFi sensing, which is usually performed on edge devices, Flops and Params also matter due to limited resources. Achieving good accuracy with fewer Flops and Params represents a good trade-off between accuracy and efficiency.

ACKNOWLEDGMENTS

This research is supported by NTU Presidential Postdoctoral Fellowship, "Adaptive Multi-modal Learning for Robust Sensing and Recognition in Smart Cities" project fund (020977-00001), at the Nanyang Technological University, Singapore.

AUTHOR CONTRIBUTIONS

Conceptualization, J.Y. and L.X.; methodology, J.Y. and X.C.; investigation, J.Y.; validation, J.Y. and X.C.; data curation, J.Y., X.C., and D.W.; writing – original draft, J.Y.; writing – review & editing, H.Z., C.X.L., S.S., and L.X.; visualization, J.Y. and X.C.; funding acquisition, J.Y.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research. We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: September 30, 2022
Revised: November 23, 2022
Accepted: February 6, 2023
Published: February 28, 2023

REFERENCES

1. Halperin, D., Hu, W., Sheth, A., and Wetherall, D. (2011). Tool release: gathering 802.11 n traces with channel state information. SIGCOMM Comput. Commun. Rev. *41*, 53. https://doi.org/10.1145/1925861.1925870.
2. Wu, C., Wang, B., Au, O.C., and Liu, K.R. (2022). Wi-fi can do more: toward ubiquitous wireless sensing. IEEE Comm. Stand. Mag. *6*, 42–49. https://doi.org/10.1109/MCOMSTD.0001.2100111.
3. Zou, H., Jiang, H., Yang, J., Xie, L., and Spanos, C. (2017). Non-intrusive occupancy sensing in commercial buildings. Energy Build. *154*, 633–643. https://doi.org/10.1016/j.enbuild.2017.08.045.
4. Wang, Y., Liu, J., Chen, Y., Gruteser, M., Yang, J., and Liu, H. (2014). 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, pp. 617–628. https://doi.org/10.1145/2639108.2639143.
5. Zou, H., Zhou, Y., Yang, J., Jiang, H., Xie, L., and Spanos, C.J. (2018). Deepsense: device-free human activity recognition via autoencoder long-term recurrent convolutional network. In 2018 IEEE International Conference on Communications (ICC) (IEEE), pp. 1–6. https://doi.org/10.1109/ICC.2018.8422895.
6. Yang, J., Zou, H., Jiang, H., and Xie, L. (2018). Carefi: sedentary behavior monitoring system via commodity wifi infrastructures. IEEE Trans. Veh. Technol. *67*, 7620–7629. https://doi.org/10.1109/TVT.2018.2833388.
7. Zou, H., Yang, J., Prasanna Das, H., Liu, H., Zhou, Y., and Spanos, C.J. (2019). Wifi and vision multimodal learning for accurate and robust device-free human activity recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. https://doi.org/10.1109/CVPRW.2019.00056.
8. Wang, H., Zhang, D., Wang, Y., Ma, J., Wang, Y., and Li, S. (2017). Rt-fall: a real-time and contactless fall detection system with commodity wifi devices. IEEE Trans. Mob. Comput. *16*, 511–526. https://doi.org/10.1109/TMC.2016.2557795.
9. Yang, J., Zou, H., Zhou, Y., and Xie, L. (2019). Learning gestures from wifi: a siamese recurrent convolutional architecture. IEEE Internet Things J. *6*, 10763–10772. https://doi.org/10.1109/JIOT.2019.2941527.
10. Zou, H., Zhou, Y., Yang, J., Jiang, H., Xie, L., and Spanos, C.J. (2018). Wifi-enabled device-free gesture recognition for smart home automation. In 2018 IEEE 14th international conference on control and automation (ICCA) (IEEE), pp. 476–481. https://doi.org/10.1109/ICCA.2018.8444331.
11. Zou, H., Zhou, Y., Yang, J., Gu, W., Xie, L., and Spanos, C.J. (2018). Wifi-based human identification via convex tensor shapelet learning. In Thirty-Second AAAI Conference on Artificial Intelligence, pp. 1711–1718. https://doi.org/10.5555/3504035.3504244.
12. Wang, D., Yang, J., Cui, W., Xie, L., and Sun, S. (2022). Caution: a robust wifi-based human authentication system via few-shot open-set gait recognition. IEEE Internet Things J. *9*, 17323–17333. https://doi.org/10.1109/JIOT.2022.3156099.
13. Deng, L., Yang, J., Yuan, S., Zou, H., Lu, C.X., and Xie, L. (2023). Gaitfi: robust device-free human identification via wifi and vision multimodal learning. IEEE Internet Things J. *10*, 625–636. https://doi.org/10.1109/JIOT.2022.3203559.
14. Zou, H., Zhou, Y., Yang, J., and Spanos, C.J. (2018). Device-free occupancy detection and crowd counting in smart buildings with wifi-enabled iot. Energy Build. *174*, 309–322. https://doi.org/10.1016/j.enbuild.2018.06.040.
15. Zou, H., Zhou, Y., Yang, J., Xie, L., and Spanos, C. (2017). Freecount: device-free crowd counting with commodity wifi. In 2017 IEEE Global Communications Conference (GLOBECOM) (IEEE), pp. 1–6. https://doi.org/10.1109/GLOCOM.2017.8255034.
16. Yang, J., Zhou, Y., Huang, H., Zou, H., and Xie, L. (2022). Metafi: device-free pose estimation via commodity wifi for metaverse avatar simulation. Preprint at arXiv. https://doi.org/10.48550/arXiv.2208.10414.
17. Restuccia, F. (2021). Ieee 802.11 bf: toward ubiquitous wi-fi sensing. Preprint at arXiv. https://doi.org/10.48550/arXiv.2103.14918.
18. Wu, D., Zhang, D., Xu, C., Wang, H., and Li, X. (2017). Device-free wifi human sensing: from pattern-based to model-based approaches. IEEE Commun. Mag. *55*, 91–97. https://doi.org/10.1109/MCOM.2017.1700143.
19. Wang, H., Zhang, D., Ma, J., Wang, Y., Wang, Y., Wu, D., Gu, T., and Xie, B. (2016). 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, pp. 25–36. https://doi.org/10.1145/2971648.2971744.
20. Wang, P., Guo, B., Xin, T., Wang, Z., and Yu, Z. (2017). Tinysense: multi-user respiration detection using wi-fi csi signals. In 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), pp. 1–6. https://doi.org/10.1109/HealthCom.2017.8210837.
21. Xu, Y.T., Chen, X., Liu, X., Meger, D., and Dudek, G. (2020). Pressense: passive respiration sensing via ambient wifi signals in noisy environments. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4032–4039. https://doi.org/10.1109/IROS45743.2020.9341474.

22. Wang, X., Yang, C., and Mao, S. (2017). Tensorbeat: tensor decomposition for monitoring multiperson breathing beats with commodity wifi. ACM Trans. Intell. Syst. Technol. *9*, 1–27. https://doi.org/10.1145/3078855.

23. Nakamura, T., Bouazizi, M., Yamamoto, K., and Ohtsuki, T. (2020). Wi-fi-csi-based fall detection by spectrogram analysis with cnn. In GLOBECOM 2020 - 2020 IEEE Global Communications Conference, pp. 1–6. https://doi.org/10.1109/GLOBECOM42002.2020.9322323.

24. Nakamura, T., Bouazizi, M., Yamamoto, K., and Ohtsuki, T. (2022). Wi-fi-based fall detection using spectrogram image of channel state information. IEEE Internet Things J. *9*, 17220–17234. https://doi.org/10.1109/JIOT.2022.3152315.

25. Yang, J., Zou, H., Jiang, H., and Xie, L. (2018). Device-free occupant activity sensing using wifi-enabled iot devices for smart homes. IEEE Internet Things J. *5*, 3991–4002. https://doi.org/10.1109/JIOT.2018.2849655.

26. Zou, H., Zhou, Y., Yang, J., Gu, W., Xie, L., and Spanos, C. (2017). Multiple kernel representation learning for wifi-based human activity recognition. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (IEEE), pp. 268–274. https://doi.org/10.1109/ICMLA.2017.0-148.

27. Zou, H., Yang, J., Zhou, Y., Xie, L., and Spanos, C.J. (2018). Robust wifi-enabled device-free gesture recognition via unsupervised adversarial domain adaptation. In 2018 27th International Conference on Computer Communication and Networks (ICCCN) (IEEE), pp. 1–8. https://doi.org/10.1109/ICCCN.2018.8487345.

28. Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: a brief review. Comput. Intell. Neurosci. *2018*. https://doi.org/10.1155/2018/7068349.

29. Shu, X., Zhang, L., Sun, Y., and Tang, J. (2021). Host–parasite: graph lstm-in-lstm for group activity recognition. IEEE Trans. Neural Netw. Learn. Syst. *32*, 663–674. https://doi.org/10.1109/TNNLS.2020.2978942.

30. Zhu, G., Zhang, L., Yang, L., Mei, L., Shah, S.A.A., Bennamoun, M., and Shen, P. (2020). Redundancy and attention in convolutional lstm for gesture recognition. IEEE Trans. Neural Netw. Learn. Syst. *31*, 1323–1335. https://doi.org/10.1109/TNNLS.2019.2919764.

31. Otter, D.W., Medina, J.R., and Kalita, J.K. (2021). A survey of the usages of deep learning for natural language processing. IEEE Trans. Neural Netw. Learn. Syst. *32*, 604–624. https://doi.org/10.1109/TNNLS.2020.2979670.

32. Wang, D., Jing, B., Lu, C., Wu, J., Liu, G., Du, C., and Zhuang, F. (2021). Coarse alignment of topic and sentiment: a unified model for cross-lingual sentiment classification. IEEE Trans. Neural Netw. Learn. Syst. *32*, 736–747. https://doi.org/10.1109/TNNLS.2020.2979225.

33. Bu, Q., Ming, X., Hu, J., Zhang, T., Feng, J., and Zhang, J. (2021). Transfersense: towards environment independent and one-shot wifi sensing. Pers. Ubiquitous Comput. *26*, 555–573. https://doi.org/10.1007/s00779-020-01480-6.

34. Zhang, J., Tang, Z., Li, M., Fang, D., Nurmi, P., and Wang, Z. (2018). Crosssense: towards cross-site and large-scale wifi sensing. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, pp. 305–320. https://doi.org/10.1145/3241539.3241570.

35. Xie, Y., Li, Z., and Li, M. (2015). Precise power delay profiling with commodity wifi. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (ACM), pp. 53–64. https://doi.org/10.1109/TMC.2018.2860991.

36. Yousefi, S., Narui, H., Dayal, S., Ermon, S., and Valaee, S. (2017). A survey on behavior recognition using wifi channel state information. IEEE Commun. Mag. *55*, 98–104. https://doi.org/10.1109/MCOM.2017.1700082.

37. Zhang, Y., Zheng, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., and Yang, Z. (2021). Widar3. 0: zero-effort cross-domain gesture recognition with wi-fi. IEEE Trans. Pattern Anal. Mach. Intell. 1. https://doi.org/10.1145/3307334.3326081.

38. Gringoli, F., Schulz, M., Link, J., and Hollick, M. (2019). Free your csi: a channel state information extraction platform for modern wi-fi chipsets. In Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (WiNTECH '19), pp. 21–28. https://doi.org/10.1145/3349623.3355477.

39. Yang, J., Chen, X., Zou, H., Wang, D., Xu, Q., and Xie, L. (2022). Efficientfi: towards large-scale lightweight wifi sensing via csi compression. IEEE Internet Things J. *9*, 13086–13095. https://doi.org/10.1109/JIOT.2021.3139958.

40. Sharma, A., Li, J., Mishra, D., Batista, G., and Seneviratne, A. (2021). Passive wifi csi sensing based machine learning framework for covid-safe occupancy monitoring. In 2021 IEEE International Conference on Communications Workshops (ICC Workshops) (IEEE), pp. 1–6. https://doi.org/10.1109/ICCWorkshops50388.2021.9473673.

41. Schäfer, J., Barrsiwal, B.R., Kokhkharova, M., Adil, H., and Liebehenschel, J. (2021). Human activity recognition using csi information with nexmon. Appl. Sci. *11*, 8860. https://doi.org/10.3390/app11198860.

42. Liu, J., Teng, G., and Hong, F. (2020). Human activity sensing with wireless signals: a survey. Sensors *20*, 1210. https://doi.org/10.3390/s20041210.

43. Zeng, Y., Wu, D., Xiong, J., Yi, E., Gao, R., and Zhang, D. (2019). Farsense: pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. *3*, 1–26. https://doi.org/10.1145/3351279.

44. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444. https://doi.org/10.1038/nature14539.

45. Schuster, M., and Paliwal, K.K. (1997). Bidirectional recurrent neural networks. IEEE Trans. Signal Process. *45*, 2673–2681. https://doi.org/10.1109/78.650093.

46. Jordan, M.I., and Mitchell, T.M. (2015). Machine learning: trends, perspectives, and prospects. Science *349*, 255–260. https://doi.org/10.1126/science.aaa8415.

47. Chen, K., Zeng, Z., and Yang, J. (2021). A deep region-based pyramid neural network for automatic detection and multi-classification of various surface defects of aluminum alloys. J. Build. Eng. *43*, 102523. https://doi.org/10.1016/j.jobe.2021.102523.

48. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F.E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing *234*, 11–26. https://doi.org/10.1016/j.neucom.2016.12.038.

49. Liu, S., Zhao, Y., and Chen, B. (2017). Wicount: a deep learning approach for crowd counting using wifi signals. In 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC) (IEEE), pp. 967–974. https://doi.org/10.1109/ISPA/IUCC.2017.00148.

50. Jiang, W., Miao, C., Ma, F., Yao, S., Wang, Y., Yuan, Y., Xue, H., Song, C., Ma, X., Koutsonikolas, D., et al. (2018). Towards environment independent device free human activity recognition. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (ACM), pp. 289–304. https://doi.org/10.1145/3241539.3241548.

51. Chen, Z., Zhang, L., Jiang, C., Cao, Z., and Cui, W. (2019). Wifi csi based passive human activity recognition using attention based blstm. IEEE Trans. Mob. Comput. *18*, 2714–2724. https://doi.org/10.1109/TMC.2018.2878233.

52. Wang, F., Gong, W., and Liu, J. (2019). On spatial diversity in wifi-based human activity recognition: a deep learning-based approach. IEEE Internet Things J. *6*, 2035–2047. https://doi.org/10.1109/JIOT.2018.2871445.

53. Xiao, C., Han, D., Ma, Y., and Qin, Z. (2019). Csigan: robust channel state information-based activity recognition with gans. IEEE Internet Things J. *6*, 10191–10204. https://doi.org/10.1109/JIOT.2019.2936580.

54. Xue, H., Jiang, W., Miao, C., Ma, F., Wang, S., Yuan, Y., Yao, S., Zhang, A., and Su, L. (2020). Deepmv: multi-view deep learning for device-free human activity recognition. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. *4*, 1–26. https://doi.org/10.1145/3380980.

55. Li, C., Liu, M., and Cao, Z. (2020). Wihf: Enable user identified gesture recognition with wifi. In IEEE INFOCOM 2020-IEEE Conference on Computer Communications (IEEE), pp. 586–595. https://doi.org/10.1109/INFOCOM41043.2020.9155539.

56. Xiao, C., Lei, Y., Ma, Y., Zhou, F., and Qin, Z. (2021). Deepseg: deep-learning-based activity segmentation framework for activity recognition using wifi. IEEE Internet Things J. *8*, 5669–5681. https://doi.org/10.1109/JIOT.2020.3033173.

57. Sheng, B., Xiao, F., Sha, L., and Sun, L. (2020). Deep spatial–temporal model based cross-scene action recognition using commodity wifi. IEEE Internet Things J. *7*, 3592–3601. https://doi.org/10.1109/JIOT.2020.2973272.

58. Fard Moshiri, P., Shahbazian, R., Nabati, M., and Ghorashi, S.A. (2021). A csi-based human activity recognition using deep learning. Sensors *21*, 7225. https://doi.org/10.3390/s21217225.

59. Ding, X., Jiang, T., Zhong, Y., Wu, S., Yang, J., and Xue, W. (2021). Improving wifi-based human activity recognition with adaptive initial state via one-shot learning. In 2021 IEEE Wireless Communications and Networking Conference (WCNC) (IEEE), pp. 1–6. https://doi.org/10.1109/WCNC49053.2021.9417590.

60. Gu, Y., Yan, H., Dong, M., Wang, M., Zhang, X., Liu, Z., and Ren, F. (2021). Wione: one-shot learning for environment-robust device-free user authentication via commodity wi-fi in man–machine system. IEEE Trans. Comput. Soc. Syst. *8*, 630–642. https://doi.org/10.1109/TCSS.2021.3056654.

61. Ma, Y., Arshad, S., Muniraju, S., Torkildson, E., Rantala, E., Doppler, K., and Zhou, G. (2021). Location-and person-independent activity recognition with wifi, deep neural networks, and reinforcement learning. ACM Trans. Internet Things *2*, 1–25. https://doi.org/10.1145/3424739.

62. Li, B., Cui, W., Wang, W., Zhang, L., Chen, Z., and Wu, M. (2021). Two-stream convolution augmented transformer for human activity recognition. Proc. AAAI Conf. Artif. Intell. *35*, 286–293. https://doi.org/10.1609/aaai.v35i1.16103.

63. Zhang, X., Tang, C., Yin, K., and Ni, Q. (2022). Wifi-based cross-domain gesture recognition via modified prototypical networks. IEEE Internet Things J. *9*, 8584–8596. https://doi.org/10.1109/JIOT.2021.3114309.

64. Wang, D., Yang, J., Cui, W., Xie, L., and Sun, S. (2021). Multimodal csi-based human activity recognition using gans. IEEE Internet Things J. *8*, 17345–17355. https://doi.org/10.1109/JIOT.2021.3080401.

65. Ding, X., Jiang, T., Zhong, Y., Wu, S., Yang, J., and Zeng, J. (2022). Wi-fi-based location-independent human activity recognition with attention mechanism enhanced method. Electronics *11*, 642. https://doi.org/10.3390/electronics11040642.

66. Gu, Y., Zhang, X., Wang, Y., Wang, M., Yan, H., Ji, Y., Liu, Z., Li, J., and Dong, M. (2022). Wigrunt: wifi-enabled gesture recognition using dual-attention network. IEEE Trans. Hum. Mach. Syst. *52*, 736–746. https://doi.org/10.1109/THMS.2022.3163189.

67. Zhuravchak, A., Kapshii, O., and Pournaras, E. (2022). Human activity recognition based on wi-fi csi data-a deep neural network approach. Procedia Comput. Sci. *198*, 59–66. https://doi.org/10.1016/j.procs.2021.12.211.

68. Yang, J., Zou, H., and Xie, L. (2022). Securesense: defending adversarial attack for secure device-free human activity recognition. IEEE Trans. Mob. Comput. 1–11. https://doi.org/10.1109/TMC.2022.3226742.

69. Wang, D., Yang, J., Cui, W., Xie, L., and Sun, S. (2022). Airfi: empowering wifi-based passive human gesture recognition to unseen environment via domain generalization. IEEE Trans. Mob. Comput. 1–12. https://doi.org/10.1109/TMC.2022.3230665.

70. Yang, J., Chen, X., Zou, H., Wang, D., and Xie, L. (2022). Autofi: towards automatic wifi human sensing via geometric self-supervised learning. IEEE Internet Things J. 1. https://doi.org/10.1109/JIOT.2022.3228820.

71. Gardner, M.W., and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. *32*, 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0.

72. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

73. Lipton, Z.C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.1506.00019.

74. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. Adv. Neural Inf. Process. Syst. *30*. https://doi.org/10.48550/arXiv.1706.03762.

75. Pan, S.J., and Yang, Q. (2010). A survey on transfer learning. Knowledge and Data Engineering. IEEE Trans. Knowl. Data Eng. *22*, 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

76. Yang, J., Yang, J., Wang, S., Cao, S., Zou, H., and Xie, L. (2023). Advancing imbalanced domain adaptation: cluster-level discrepancy minimization with a comprehensive benchmark. IEEE Trans. Cybern. *53*, 1106–1117. https://doi.org/10.1109/TCYB.2021.3093888.

77. Yang, J., Qian, H., Zou, H., and Xie, L. (2021). Learning decomposed hierarchical feature for better transferability of deep models. Inf. Sci. *580*, 385–397. https://doi.org/10.1016/j.ins.2021.08.046.

78. Yang, J., Zou, H., Zhou, Y., and Xie, L. (2020). Towards stable and comprehensive domain alignment: max-margin domain-adversarial training. Preprint at arXiv. https://doi.org/10.48550/arXiv.2003.13249.

79. Zou, H., Yang, J., and Wu, X. (2021). Unsupervised energy-based adversarial domain adaptation for cross-domain text classification. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 1208–1218. https://doi.org/10.18653/v1/2021.findings-acl.103.

80. Arshad, S., Feng, C., Yu, R., and Liu, Y. (2019). Leveraging transfer learning in multiple human activity recognition using wifi signal. In 2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), pp. 1–10. https://doi.org/10.1109/WoWMoM.2019.8793019.

81. Li, L., Wang, L., Han, B., Lu, X., Zhou, Z., and Lu, B. (2021). Subdomain adaptive learning network for cross-domain human activities recognition using wifi with csi. In 2021 IEEE 27th International Conference on Parallel and Distributed Systems (ICPADS), pp. 1–7. https://doi.org/10.1109/ICPADS53394.2021.00006.

82. Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al. (2020). Bootstrap your own latent-a new approach to self-supervised learning. Adv. Neural Inf. Process. Syst. *33*, 21271–21284. https://doi.org/10.48550/arXiv.2006.07733.

83. Wang, F., Kong, T., Zhang, R., Liu, H., and Li, H. (2021). Self-supervised learning by estimating twin class distributions. Preprint at arXiv. https://doi.org/10.48550/arXiv.2110.07402.

84. Sagi, O., and Rokach, L. (2018). Ensemble learning: a survey. WIREs Data Mining Knowl. Discov. *8*, e1249. https://doi.org/10.1002/widm.1249.

85. Chen, S., Wang, W., and Pan, S.J. (2019). Cooperative pruning in cross-domain deep neural network compression. In IJCAI, pp. 2102–2108. https://doi.org/10.24963/ijcai.2019/291.

86. Zou, H., Zhou, Y., Yang, J., Liu, H., Das, H.P., and Spanos, C.J. (2019). Consensus adversarial domain adaptation. Proc. AAAI Conf. Artif. Intell. *33*, 5997–6004. https://doi.org/10.1609/aaai.v33i01.33015997.

87. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. Commun. ACM *60*, 84–90. https://doi.org/10.1145/3065386.

88. Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Preprint at arXiv. https://doi.org/10.48550/arXiv.1409.1556.

89. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.

90. Tan, M., and Le, Q. (2019). Efficientnet: rethinking model scaling for convolutional neural networks. In International conference on machine learning (PMLR), pp. 6105–6114. https://doi.org/10.48550/arXiv.1905.11946.

91. Kingma, D., and Ba, J. (2014). Adam: a method for stochastic optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.

92. Hu, J., Yang, J., Ong, J., and Xie, L. (2022). Resfi: wifi-enabled device-free respiration detection based on deep learning. In 2022 IEEE 18th international conference on control and automation (ICCA) (IEEE), pp. 510–515. https://doi.org/10.1109/ICCA54724.2022.9831898.

93. Yang, J., Zou, H., Cao, S., Chen, Z., and Xie, L. (2020). Mobileda: toward edge-domain adaptation. IEEE Internet Things J. *7*, 6909–6918. https://doi.org/10.1109/JIOT.2020.2976762.

94. Kefayati, M.H., Pourahmadi, V., and Aghaeinia, H. (2020). Wi2vi: generating video frames from wifi csi samples. IEEE Sens. J. *20*, 11463–11473. https://doi.org/10.1109/JSEN.2020.2996078.

95. Wang, F., Zhou, S., Panev, S., Han, J., and Huang, D. (2019). Person-in-wifi: fine-grained person perception using wifi. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5452–5461. https://doi.org/10.1109/ICCV.2019.00555.

96. Liu, J., Akhtar, N., and Mian, A. (2022). Adversarial attack on skeleton-based human action recognition. IEEE Trans. Neural Netw. Learn. Syst. *33*, 1609–1622. https://doi.org/10.1109/TNNLS.2020.3043002.

97. Zhou, S., Zhang, W., Peng, D., Liu, Y., Liao, X., and Jiang, H. (2020). Adversarial wifi sensing for privacy preservation of human behaviors. IEEE Commun. Lett. *24*, 259–263. https://doi.org/10.1109/LCOMM.2019.2952844.

98. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proc. IEEE *86*, 2278–2324. https://doi.org/10.1109/5.726791.

99. Khan, A., Sohail, A., Zahoora, U., and Qureshi, A.S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. *53*, 5455–5516. https://doi.org/10.1007/s10462-020-09825-6.

100. Wang, C., Yang, J., Xie, L., and Yuan, J. (2019). Kervolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 31–40. https://doi.org/10.1109/CVPR.2019.00012.

101. Abdel-Hamid, O., Mohamed, A.R., Jiang, H., and Penn, G. (2012). Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition. In 2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP) (IEEE), pp. 4277–4280. https://doi.org/10.1109/ICASSP.2012.6288864.

102. Yin, W., Kann, K., Yu, M., and Schütze, H. (2017). Comparative study of cnn and rnn for natural language processing. Preprint at arXiv. https://doi.org/10.48550/arXiv.1702.01923.

103. Zhang, Y., Wang, X., Wang, Y., and Chen, H. (2020). Human activity recognition across scenes and categories based on csi. IEEE Trans. Mob. Comput. *21*, 2411–2420. https://doi.org/10.1109/TMC.2020.3041756.

104. Moshiri, P.F., Nabati, M., Shahbazian, R., and Ghorashi, S.A. (2021). Csi-based human activity recognition using convolutional neural networks. In 2021 11th International Conference on Computer Engineering and Knowledge (ICCKE), pp. 7–12. https://doi.org/10.1109/ICCKE54056.2021.9721516.

105. Yang, J., Wang, K., Peng, X., and Qiao, Y. (2018). Deep recurrent multi-instance learning with spatio-temporal features for engagement intensity prediction. In Proceedings of the 2018 on International Conference on Multimodal Interaction (ACM), pp. 594–598. https://doi.org/10.1145/3242969.3264981.

106. Ding, J., and Wang, Y. (2019). Wifi csi-based human activity recognition using deep recurrent neural network. IEEE Access *7*, 174257–174269. https://doi.org/10.1109/ACCESS.2019.2956952.

107. Shi, Z., Zhang, J.A., Xu, R., and Cheng, Q. (2019). Deep learning networks for human activity recognition with csi correlation feature extraction. In ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pp. 1–6. https://doi.org/10.1109/ICC.2019.8761445.

108. Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural Comput. *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

109. Kim, S.-C., and Kim, Y.-H. (2022). Efficient classification of human activity using pca and deep learning lstm with wifi csi. In 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 329–332. https://doi.org/10.1109/ICAIIC54071.2022.9722627.

110. Thariq Ahmed, H.F., Ahmad, H., Phang, S.K., Harkat, H., and Narasingamurthi, K. (2021). Wi-fi csi based human sign language recognition using lstm network. In 2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), pp. 51–57. https://doi.org/10.1109/IAICT52856.2021.9532548.

111. Tang, Z., Liu, Q., Wu, M., Chen, W., and Huang, J. (2021). Wifi csi gesture recognition based on parallel lstm-fcn deep space-time neural network. China Commun. *18*, 205–215. https://doi.org/10.23919/JCC.2021.03.016.

112. Kadir, R., Saha, R., Awal, M.A., and Kadir, M.I. (2021). Deep bidirectional lstm network learning-aided ofdma downlink and sc-fdma uplink. In 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), pp. 1–4. https://doi.org/10.1109/ICECIT54077.2021.9641123.

113. Dua, N., Singh, S.N., and Semwal, V.B. (2021). Multi-input cnn-gru based human activity recognition using wearable sensors. Computing *103*, 1461–1478. https://doi.org/10.1007/s00607-021-00928-8.

114. Chen, K., Yao, L., Zhang, D., Wang, X., Chang, X., and Nie, F. (2020). A semisupervised recurrent convolutional attention model for human activity recognition. IEEE Trans. Neural Netw. Learn. Syst. *31*, 1747–1756. https://doi.org/10.1109/TNNLS.2019.2927224.

115. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: transformers for image recognition at scale. Preprint at arXiv. https://doi.org/10.48550/arXiv.2010.11929.

116. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. Adv. Neural Inf. Process. Syst. *27*. https://doi.org/10.48550/arXiv.1406.2661.

117. Yang, J., Zou, H., Zhou, Y., and Xie, L. (2021). Robust adversarial discriminative domain adaptation for real-world cross-domain visual recognition. Neurocomputing *433*, 28–36. https://doi.org/10.1016/j.neucom.2020.12.046.

118. Yang, J., Zou, H., Zhou, Y., Zeng, Z., and Xie, L. (2020). Mind the discriminability: asymmetric adversarial domain adaptation. In European Conference on Computer Vision (Springer), pp. 589–606. https://doi.org/10.1007/978-3-030-58586-0_35.

119. Xu, Y., Yang, J., Cao, H., Chen, Z., Li, Q., and Mao, K. (2021). Partial video domain adaptation with partial adversarial temporal attentive network. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9332–9341. https://doi.org/10.48550/arXiv.2107.04941.

120. Kingma, D.P., and Welling, M. (2013). Auto-encoding variational bayes. Preprint at arXiv. https://doi.org/10.48550/arXiv.1312.6114.

121. Kim, M., Han, D., and Rhee, J.-K.K. (2021). Multiview variational deep learning with application to practical indoor localization. IEEE Internet Things J. *8*, 12375–12383. https://doi.org/10.1109/JIOT.2021.3063512.

122. Chen, X., Li, H., Zhou, C., Liu, X., Wu, D., and Dudek, G. (2020). Fido: ubiquitous fine-grained wifi-based localization for unlabelled users via domain adaptation. In Proceedings of The Web Conference 2020, pp. 23–33. https://doi.org/10.1145/3366423.3380091.

123. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. *32*. https://doi.org/10.48550/arXiv.1912.01703.