Contents lists available at ScienceDirect

High-Confidence Computing

journal homepage: www.sciencedirect.com/journal/high-confidence-computing

Research article

An investigation of the private-attribute leakage in WiFi sensing

Yiding Shi [a,1], Xueying Zhang [a,1], Lei Fu [b], Huanle Zhang [a,*]

[a] School of Computer Science and Technology, Shandong University, China
[b] School of Modern Finance, Jiaxing Nanhu University, China

ARTICLE INFO

Article history:
Received 13 June 2023
Revised 22 September 2023
Accepted 11 December 2023
Available online xxxx

Keywords:
WiFi sensing
Private attribute
Deep learning
Privacy protection

ABSTRACT

WiFi sensing is critical to many applications, such as localization, human activity recognition, and contact-less health monitoring. With metaverse and ubiquitous sensing advances, WiFi sensing becomes increasingly imperative. However, as shown in this paper, WiFi sensing data leaks users' private attributes (e.g., height, weight, and gender), violating increasingly stricter privacy protection laws and regulations. To demonstrate the leakage of private attributes in WiFi sensing, we investigate two public WiFi sensing datasets and apply a deep learning model to recognize users' private attributes. Our experimental results clearly show that our model can identify users' private attributes in WiFi sensing data collected by general WiFi applications, with almost 100% accuracy for gender inference, less than 4 cm error for height inference, and about 4 kg error for weight inference, respectively. Our finding calls for research efforts to preserve data privacy while enabling WiFi sensing-based applications.

1. Introduction

WiFi sensing technology uses WiFi signals to detect and track human movements and activities. WiFi systems have increasingly adopted Channel State Information (CSI) for various sensing applications. By analyzing the changes in the WiFi signals, we can extract valuable information about the user's location, movements, and interactions from the environment. WiFi sensing provides several advantages over other sensing techniques, such as being cost-effective, non-intrusive, and easy to install [1]. Furthermore, it is not influenced by lighting conditions and can be adapted to a broader range of settings. As new WiFi technologies are developed and deployed, more WiFi sensing opportunities will emerge, with applications expanding beyond humans to encompass environments, animals, and objects [2].

However, WiFi sensing causes private-attribute leakage intentionally or unintentionally. Thus, the advantages of WiFi sensing, such as non-intrusiveness and easy deployment, instead turn to severe privacy concerns. For example, an adversary may leverage WiFi sensing signals to steal users' personal attributes. As WiFi signals can penetrate walls, common countermeasures, e.g., visual occlusions, fail to work. Moreover, numerous WiFi sensing applications have already been deployed to life-log various activities, such as daily routines, hand gestures, and keystrokes [2].

This paper demonstrates the leakage of private attributes in WiFi sensing. We conduct experiments in practical scenarios where smart applications use WiFi sensing systems. Specifically, we investigate two public WiFi sensing datasets, i.e., a gesture recognition dataset named Widar [3] and an activity recognition named Wiar [4]. Both datasets include CSI features of WiFi signals when users conduct various actions. We train a Deep Learning (DL) model and infer users' private attributes from each piece of WiFi sensing data. Our model achieves significantly better accuracy in predicting private attributes than the baseline (i.e., statistical guess) and thus proves that WiFi sensing indeed has privacy issues.

In summary, we make the following contributions:

1. To the best of our knowledge, this is the first work that reveals the private attribute leakage in WiFi sensing. Please note that we do not design a WiFi system to infer private attributes but unveil that existing WiFi sensing systems have byproducts of user privacy leakage.
2. We conduct thorough experiments to show that accurate private information (height, weight, and gender) can be inferred. Specifically, we apply a DL model to two public WiFi sensing datasets [3,4]. We achieved almost 100% accuracy in gender prediction, an average of less than 4 cm error in height prediction, and about 4 kg error in weight prediction. We also conducted an ablation study to show that better performance can be reached when a single action rather than the aggregate action is considered. Overall, our DL model performs significantly and consistently better than the baseline and thus proves by experiments that WiFi sensing indeed leaks private information.

[*] Corresponding author.
E-mail address: dtczhang@sdu.edu.cn (H. Zhang).
[1] Contributed equally.

https://doi.org/10.1016/j.hcc.2024.100209

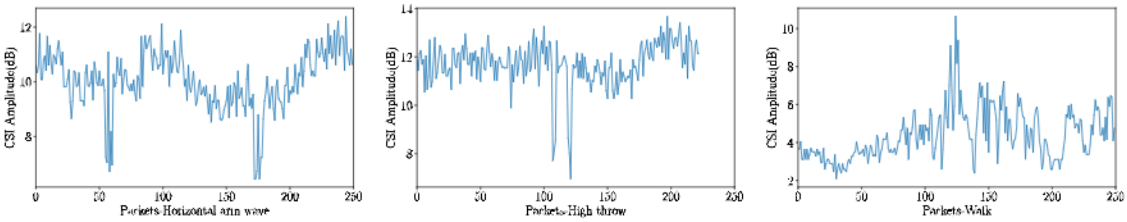

Fig. 1. The amplitude of CSI subcarrier signal when a volunteer performs three different actions. (x-axis represents time).

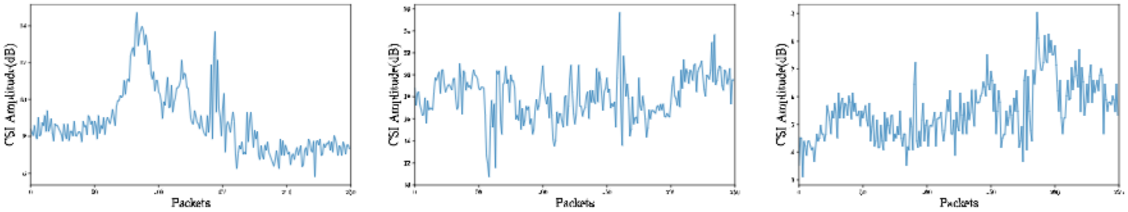

Fig. 2. The amplitude of CSI subcarrier signal when three volunteers performing the same action (arm wave).

Table 1
Related work of WiFi sensing applications.

Applications	Publication
Human activity recognition	[3,4,6,7]
Healthcare	[8–12]
The Internet of Things (IoT) and smart home	[1,13–16]

3. We release our code at https://github.com/SnoopD201/Private-Attribute-Leakage-Investigation to facilitate the research of privacy leakage in WiFi sensing.

2. Related work

In addition to being used in communications, WiFi signals are also widely used in wireless sensing, such as human activity recognition, fall detection, and other sensing tasks. Table 1 lists representative WiFi sensing-based applications. Overall, WiFi sensing is a versatile technology that can be applied to various domains, providing valuable insights and enabling new applications and services [5].

Meanwhile, WiFi sensing is gradually exhibiting its importance in metaverse. As an emerging field, the metaverse is still in its early stages. Nevertheless, many research projects have been explored in this field. For example, metaverse can be combined with AI, IoT [17], blockchain [18], etc., to build the next generation applications [19–22]. Because of the ubiquitous deployment of WiFi systems, WiFi sensing has become an indispensable component of the metaverse. We expect more WiFi-based metaverse systems to be designed in the coming years.

The observation that sensing signals carry private attributes has been introduced previously. For example, ObscureNet [23] explores the private attributes of inertial sensors and designs attribute-invariant latent representation to protect the private attributes. However, this paper is the first work demonstrating that WiFi sensing signals also leak private attributes, which is more concerning since WiFi sensing is ubiquitous and contactless.

3. Preliminaries

This section presents the principles of WiFi sensing and inferring private attributes from WiFi sensing data.

3.1. Principle of WiFi sensing

The state-of-the-art WiFi systems exploit Channel State Information (CSI) of WiFi signals to sense users. CSI characterizes signals propagating through the wireless channel, including the impact of time delay, energy attenuation, and phase shift [24]. In a typical indoor environment, the transmission signal reaches the receiver through various paths, i.e., the multipath effect. The combination of several alias versions of the transmitted signal forms the received signal. Usually, we use the following equation to represent CSI,

$$H = \sum_{k=1}^{N} \|H_k\| \, e^{-j\theta_i} \tag{1}$$

where $\|H_k\|$ is the CSI amplitude of the kth subcarrier, N is the total number of subcarriers and θ_i is the phase of the kth subcarrier. For a multi-antenna WiFi system, the CSI data is represented by a four-dimensional matrix, that is,

$$H \in C^{N \times M \times K \times T} \tag{2}$$

while $H(n, m, k, t)$ represents the sampled value of the Channel Frequency Response (CFR) between the ith antenna of Tx (Transmitter antenna) and the jth antenna of Rx (Receiver antenna). Fig. 1 shows the CSI signals of a volunteer when performing different actions in Wiar. Only one subcarrier is plotted in Fig. 1. Clearly, different actions result in different CSI patterns.

3.2. Principle of inferring private-attributes from WiFi sensing data

Since the CSI data reflects the information of the WiFi signal on each subcarrier during transmission, interference caused by user activity during WiFi signal transmission can result in variations in the obtained CSI patterns. For example, people of different heights affect the channel state on different subcarriers. The same principle also applies to user weight and gender, and thus we can infer users' privacy attributes from CSI data. We can use deep learning techniques to infer the private attributes from the WiFi sensing data to infer the relationship between CSI data and private attributes. Fig. 2 shows the amplitude of the CSI signal when different volunteers performing the same action (arm wave). Different people display distinguishable patterns when performing the same action, laying the foundation for private-attribute leakage in WiFi sensing.

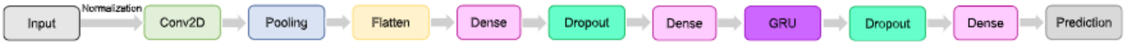

Fig. 3. Simplified structure of our model.

Table 2
User statistics of the two datasets. Left: Widar dataset; Right: Wiar dataset.

User ID	Gender	Height (cm)	Weight (kg)
1	Male	178	70
2	Female	161	62
3	Male	170	74
4	Female	160	57
5	Male	180	75
6	Male	172	69
7	Male	168	58
8	Male	175	85
9	Male	165	54
10	Male	170	72
11	Male	175	70
12	Male	176	66
13	Female	155	56
14	Female	158	55
15	Male	175	80
16	Male	186	88

User ID	Gender	Height (cm)	Weight (kg)
1	Male	173	85
2	Female	180	75
3	Male	165	65
4	Female	160	60
5	Male	162	53
6	Male	170	60
7	Male	165	50
8	Male	155	65
9	Male	180	85
10	Male	175	70

4. The measurement methodology

This section explains our measurement methodology, including our approach, the datasets, and the experimental settings.

4.1. Approach overview

To verify the private attribute leakage in WiFi sensing, we use a DL model and apply it to the CSI data. Before training, we normalize the labels (height and weight) using the Min–Max Normalization method, calculated as follows:

$$x' = \frac{x - \min x}{\max x - \min x} \quad (3)$$

where x' represents the normalized label and x represents the original label.

Fig. 3 shows our DL model, which is similar to [25] in nature. We feed the CSI data to a convolution layer using a 3 × 3 kernel and a pooling layer to extract the data features and then reshape the features to a one-dimensional vector. Then, the one-dimensional vector is sequentially passed through a two-layer fully connected neural network with a dropout probability of 0.5, a GRU layer, and a fully connected layer. For the gender classification task, a softmax layer is added after the fully connected layer, which has an output dimension of 2, representing the probability of the CSI data belonging to a male and a female. Binary cross entropy loss and RMSprop optimizer are used in the gender prediction. For height and weight, the output of the fully connected layer is a number representing the prediction value of height or weight (after normalization). MSE loss and Adam optimizer are used in the weight and height prediction. We set the learning rate to 0.001, batch size to 32, and dropout rate to 0.5.

4.2. Datasets

We use two datasets to demonstrate private attribute leakage in WiFi sensing. The first is Widar [3], a dataset aiming at gesture recognition based on WiFi CSI data. The database includes 9 gesture movements and 16 volunteers with various heights and weights. The other dataset is Wiar [4], which collects CSI data for human activity recognition, which includes 10 volunteers and performs 16 activities. Table 2 gives the user statistics, including the gender, height, and weight attributes.

4.3. Experimental settings

First, we perform operations such as normalization, dimension adjustment, and labeling of the CSI data. Afterward, we split data into training and test sets. Last, the processed data is fed into the model, and appropriate classification or regression models are applied. For example, the height and weight prediction tasks require numerical estimation, while gender only needs to be classified as male or female. Our experiments use Python 3.6, Tensorflow 2.0, and Keras 2.3. For more details, please refer to our released code.

We compare our model's performance with the baseline. The baseline is the statistical guess of the user profiles in Table 2. In other words, the baseline always predicts the gender to male, which results in 0.75 (12/16) accuracy for the Widar dataset and 0.80 (8/10) accuracy for the Wiar dataset. Similarly, the baseline always outputs each dataset's mean height and mean weight. Hence, the baseline obtained an average of 7.21 cm and 7.41 cm height prediction errors and 6.73 kg and 9.64 kg weight prediction errors for Widar and Wiar, respectively. If our model performs better than the baseline, it convinces that WiFi sensing leaks private information.

5. Experiment results

We conduct comprehensive experiments and result analysis in this section. Specifically, we provide the overall accuracy analysis in Section 5.1 and the ablation study in Section 5.2.

5.1. Overall accuracy

We train our DL model on the whole dataset including a variety of actions. Although extracting private attributes from the CSI data is challenging due to the confusion from different actions, the experiment results show that our model achieves significantly higher accuracy than the baseline.

Table 3 tabulates the results for the height prediction using the baseline and our DL model. It is clear that ours achieves better accuracy in inferring users' height attributes, with less than a 4 cm error for both datasets. Compared to the baseline, our model reduces the prediction error by more than half. Therefore, it is evident that WiFi sensing data leaks the height information.

Similarly, Table 4 tabulates the results for the weight prediction using the baseline method and our DL model. Our model

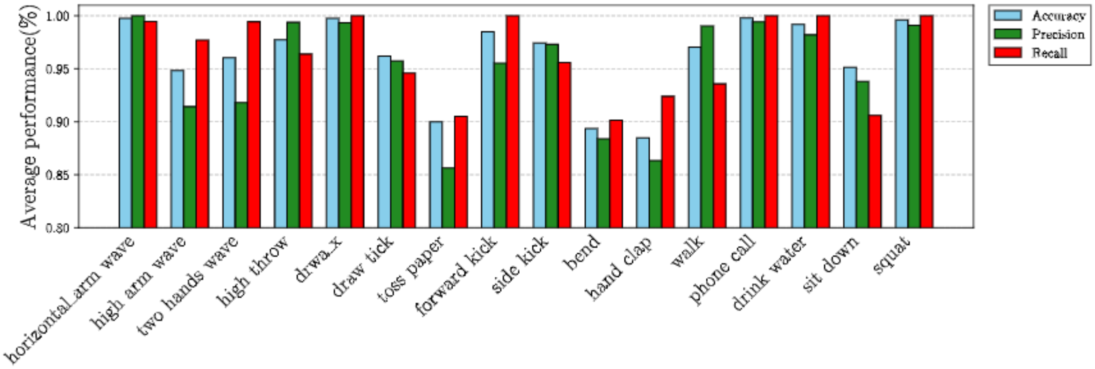

Fig. 4. Gender prediction of our model versus different actions when using the Wiar dataset. The higher, the better.

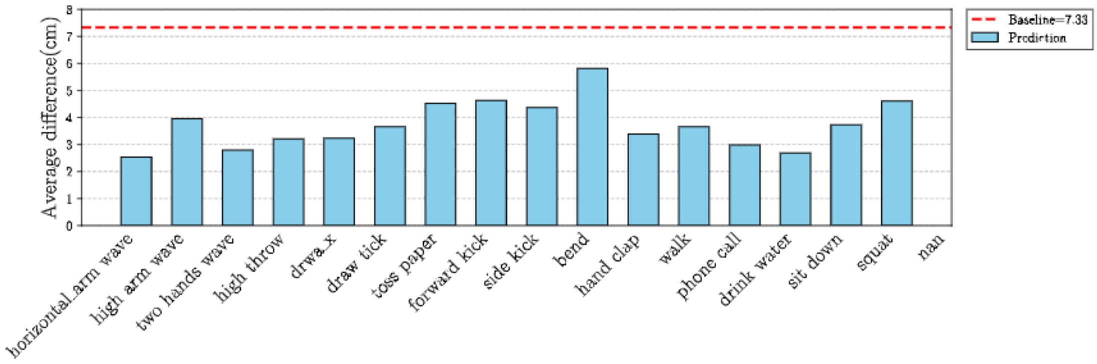

Fig. 5. Height prediction of our model versus different actions when using the Wiar dataset. The lower, the better.

Table 3
Comparison of the height prediction error between the baseline and our model. The smaller the better.

Dataset	Baseline (cm)	Ours (cm)
Widar	7.21	3.57
Wiar	7.41	3.81

Table 4
Comparison of the weight prediction error between the baseline and our model. The smaller, the better.

Dataset	Baseline (kg)	Ours (kg)
Widar	6.73	4.87
Wiar	9.64	3.17

significantly reduces the weight prediction error compared to the baseline. For example, the weight prediction error is decreased from 9.64 kg in the baseline to only 3.17 kg in our model for the Wiar dataset. Thus, we can also conclude that WiFi sensing data leaks the weight information.

Table 5 shows the confusion matrices of our gender prediction. The accuracy is higher than 0.9 for both datasets. In comparison, the accuracy of the baseline is only 0.75 (12/16) for the Widar dataset and 0.8 (8/10) for the Wiar dataset. Therefore, WiFi sensing is very accurate in inferring the gender information of the user, which consequences can raise security concerns.

5.2. Ablation study

In addition to the overall performance of the whole dataset, we explore the performance when a single action is used for gender, weight, and height prediction. We remove some actions in the dataset because they do not have enough data samples for training. In the end, the Wiar dataset is kept with 16 actions (e.g., arm wave and drink water), and the Widar dataset is kept with 4 actions (e.g., clap and sweep).

Fig. 4 shows the gender prediction performance versus different actions in the Wiar dataset. Besides the accuracy metric, we also show the precision and recall metrics. As we can see, many actions have nearly 100% performance in recognizing gender. Figs. 5 and 6 show the precision of our model to predict the height and weight for the Wiar dataset, respectively. We can see a consistent pattern: WiFi sensing from all actions can leak private attributes.

We conducted similar experiments with the Widar dataset. Figs. 7, 8, and 9 show our model's performance for predicting the gender, height, and weight, respectively. As with the Wiar dataset, the same conclusion can be drawn: WiFi sensing leaks the private attributes no matter user actions.

6. Conclusion

Our investigation results confirm that personal information such as height, weight, and gender can be inferred from the WiFi sensing data. We hope this work could raise awareness about privacy leakage in WiFi sensing, which will be a crucial component of the metaverse. Addressing such privacy issues can help ensure a more secure interaction. Our approach can be easily extended to other WiFi sensing applications because it takes general CSI data as the input and a lightweight AI model for the prediction.

This paper does not propose a privacy protection mechanism, but interested readers can refer to ObscureNet [23] to gain insights about protecting inertial sensing data. We leave it as future work to design a privacy protection mechanism specifically designed for WiFi sensing.

Table 5
Gender prediction of our model in the two datasets. Left: Widar dataset; Right: Wiar dataset.

		Prediction	
		Male	Female
Actual	Male	0.99	0.01
	Female	0.03	0.97
Overall accuracy:0.982			

		Prediction	
		Male	Female
Actual	Male	0.9	0.1
	Female	0.09	0.91
Overall accuracy:0.907			

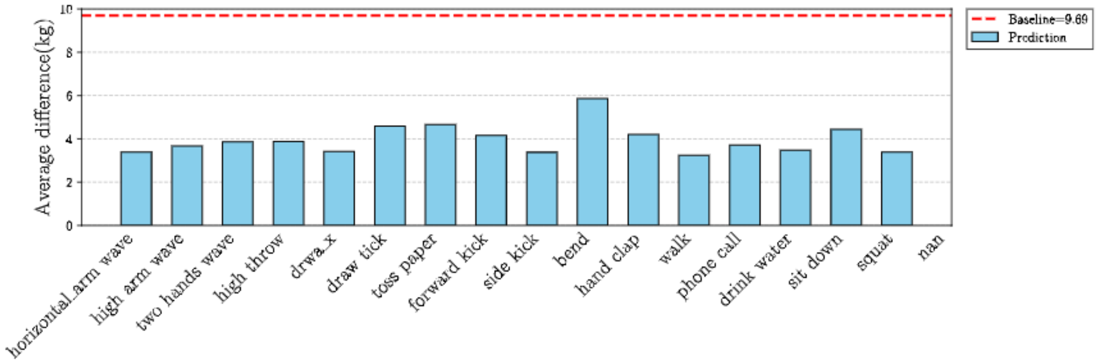

Fig. 6. Weight prediction of our model versus different actions when using the Wiar dataset. The lower, the better.

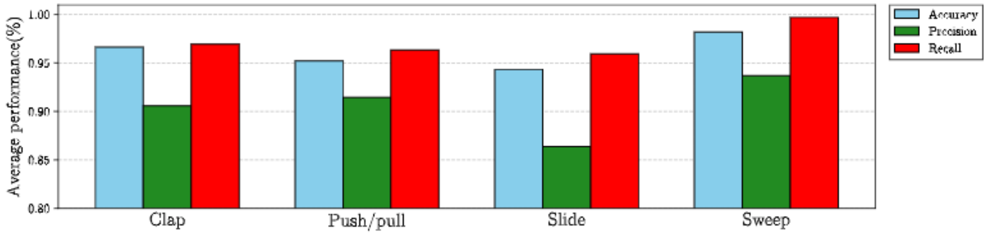

Fig. 7. Gender prediction of our model versus different actions when using the Widar dataset. The higher, the better.

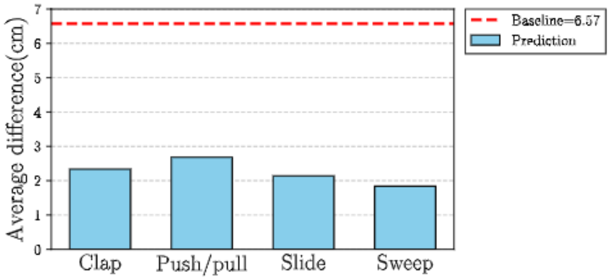

Fig. 8. Height prediction of our model versus different actions when using the Widar dataset. The lower, the better.

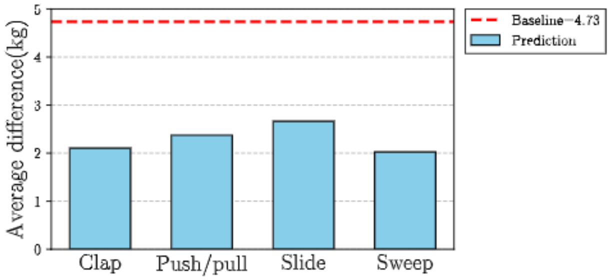

Fig. 9. Weight prediction of our model versus different actions when using the Widar dataset. The lower, the better.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank anonymous reviewers for their insightful comments. This work was supported by the National Natural Science Foundation of China (Grant No. 62302265, U23A20332) and Shandong Provincial Natural Science Foundation (Grant No. ZR2023QF172).

References

[1] H. Jiang, C. Cai, X. Ma, Y. Yang, J. Liu, Smart home based on WiFi sensing: A survey, IEEE Access 6 (2018) 13317–13325.

[2] Y. Ma, G. Zhou, S. Wang, WiFi sensing with channel state information: A survey, ACM Comput. Surv. 52 (3) (2019) 1–36.

[3] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, Z. Yang, Widar3. 0: Zero-effort cross-domain gesture recognition with Wi-Fi, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2021) 8671–8688.

[4] L. Guo, L. Wang, C. Lin, J. Liu, B. Lu, J. Fang, Z. Liu, Z. Shan, J. Yang, S. Guo, Wiar: A public dataset for wifi-based activity recognition, IEEE Access 7 (2019) 154935–154945.

[5] A. Khalili, A.-H. Soliman, M. Asaduzzaman, A. Griffiths, Wi-fi sensing: applications and challenges, J. Eng. 2020 (3) (2020) 87–97.

[6] W. He, K. Wu, Y. Zou, Z. Ming, WiG: WiFi-based gesture recognition system, in: 2015 24th International Conference on Computer Communication and Networks, ICCCN, IEEE, 2015, pp. 1–7.

[7] Y. He, Y. Chen, Y. Hu, B. Zeng, WiFi vision: Sensing, recognition, and detection with commodity MIMO-OFDM WiFi, IEEE Internet Things J. 7 (9) (2020) 8296–8317.

[8] H.-H. Chen, C.-L. Lin, C.-H. Chang, WiFi-based detection of human subtle motion for health applications, Bioengineering 10 (2) (2023) 228.

[9] Y. Ge, A. Taha, S.A. Shah, K. Dashtipour, S. Zhu, J. Cooper, Q.H. Abbasi, M.A. Imran, Contactless WiFi sensing and monitoring for future healthcare - Emerging trends, challenges, and opportunities, IEEE Rev. Biomed. Eng. 16 (2023) 171–191, http://dx.doi.org/10.1109/RBME.2022.3156810.

[10] B. Tan, Q. Chen, K. Chetty, K. Woodbridge, W. Li, R. Piechocki, Exploiting WiFi channel state information for residential healthcare informatics, IEEE Commun. Mag. 56 (5) (2018) 130–137, http://dx.doi.org/10.1109/MCOM.2018.1700064.

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

[12] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, S. Li, RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices, IEEE Trans. Mob. Comput. 16 (2) (2016) 511–526.

[13] L. Li, H. Xiaoguang, C. Ke, H. Ketai, The applications of WiFi-based wireless sensor network in internet of things and smart grid, in: 2011 6th IEEE Conference on Industrial Electronics and Applications, 2011, pp. 789–793, http://dx.doi.org/10.1109/ICIEA.2011.5975693.

[14] Y. Ren, S. Tan, L. Zhang, Z. Wang, Z. Wang, J. Yang, Liquid level sensing using commodity wifi in a smart home environment, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4 (1) (2020) 1–30.

[15] J. Yang, H. Zou, H. Jiang, L. Xie, Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes, IEEE Internet Things J. 5 (5) (2018) 3991–4002, http://dx.doi.org/10.1109/JIOT.2018.2849655.

[16] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, Y. Liu, Smokey: Ubiquitous smoking detection with commercial WiFi infrastructures, in: IEEE International Conference on Computer Communications, IEEE, 2016, pp. 1–9.

[17] K. Lone, S.A. Sofi, A review on offloading in fog-based Internet of Things: Architecture, machine learning approaches, and open issues, High-Conf. Comput. 3 (2) (2023) 1–10.

[18] Y. Chen, H. Chen, Y. Zhang, M. Han, M. Siddula, Z. Cai, A survey on blockchain systems: Attacks, defenses, and privacy preservation, High-Conf. Comput. 2 (2) (2022) 1–20.

[19] G. Bansal, K. Rajgopal, V. Chamola, Z. Xiong, D. Niyato, Healthcare in metaverse: A survey on current metaverse applications in healthcare, IEEE Access 10 (2022) 119914–119946, http://dx.doi.org/10.1109/ACCESS.2022.3219845.

[20] Q. Yang, Y. Zhao, H. Huang, Z. Xiong, J. Kang, Z. Zheng, Fusing blockchain and AI with metaverse: A survey, IEEE Open J. Comput. Soc. 3 (2022) 122–136, http://dx.doi.org/10.1109/OJCS.2022.3188249.

[21] S. Tayal, K. Rajagopal, V. Mahajan, Virtual reality based metaverse of gamification, in: 2022 6th International Conference on Computing Methodologies and Communication, ICCMC, 2022, pp. 1597–1604, http://dx.doi.org/10.1109/ICCMC53470.2022.9753727.

[22] M.A.I. Mozumder, M.M. Sheeraz, A. Athar, S. Aich, H.-C. Kim, Overview: Technology roadmap of the future trend of metaverse based on IoT, blockchain, AI technique, and medical domain metaverse activity, in: 2022 24th International Conference on Advanced Communication Technology, ICACT, 2022, pp. 256–261, http://dx.doi.org/10.23919/ICACT53585.2022.9728808.

[23] O. Hajihassnai, O. Ardakanian, H. Khazaei, ObscureNet: Learning attribute-invariant latent representation for anonymizing sensor data, in: International Conference on Internet-of-Things Design and Implementation, IoTDI, 2021, pp. 40–52.

[24] D. Halperin, W. Hu, A. Sheth, D. Wetherall, Tool release: Gathering 802.11 n traces with channel state information, ACM SIGCOMM Comput. Commun. Rev. 41 (1) (2011) 53.

[25] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, 2014, arXiv preprint arXiv:1406.1078.