Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d6dd00077k

Lightweight privacy-preserving human activity recognition from CSI data using a CNN-temporal attention network

Khondakar Ashik Shahriar, Maruf Ahmed and Hafiz Imtiaz *

WiFi Channel State Information (CSI) has emerged as a powerful sensing modality for device-free Human Activity Recognition (HAR), enabling fine-grained motion understanding without requiring wearable sensors or cameras. However, any type of HAR – either CSI signal-based or device-based, inherently encodes sensitive behavioral patterns, raising significant privacy concerns. In this work, we propose an end-to-end privacy-preserving CSI-based HAR framework that integrates a Convolutional Neural Network (CNN) with a temporal attention mechanism. We perform extensive evaluations on multiple benchmark datasets consisting of varying distance and height factors, as well as different environmental conditions. Our baseline non-privacy-preserving CNN–temporal attention model achieves state-of-the-art performance. Additionally, we incorporate differential privacy (DP) into the training pipeline – enabling rigorous privacy guarantees through controlled noise injection and gradient clipping. We evaluate the proposed framework's privacy–utility trade-off and demonstrate that even strong privacy protection can maintain excellent recognition accuracy. Our framework can progressively approach the non-privacy-preserving performance for some parameter regimes. As such, our experimental results clearly demonstrate that the proposed architecture remains robust under privacy constraints and generalizes effectively across heterogeneous sensing conditions. We argue that our work provides practical insights into deploying secure and privacy-aware WiFi sensing systems for real-world HAR applications.

Received 16th February 2026
Accepted 14th May 2026

DOI: 10.1039/d6dd00077k

rsc.li/digitaldiscovery

1 Introduction

Human Activity Recognition (HAR) automatically identifies and categorizes human actions and plays a vital role in numerous real-world applications, including smart homes, healthcare monitoring, assisted living, surveillance, and security systems.[1,2] As an example in healthcare scenarios, continuous activity monitoring is essential for patients suffering from chronic conditions such as diabetes, cardiovascular diseases, and obesity, where maintaining prescribed physical activity routines is critical.[3–5] Similarly, detecting abnormal behavior in patients with cognitive impairments or mental health disorders can help prevent adverse events and enable timely intervention.[6,7] Recently, incorporating HAR with digital twin systems has translated real-world human activity into virtual replicas, allowing real-time monitoring and predictive decision-making in smart systems.[8] These practical demands have motivated the development of accurate and scalable HAR systems, which are realized through various sensing modalities, including vision-based and wearable sensor-based approaches.[9,10] These methods are limited by environmental sensitivity, continuous user participation, and long-term device attachment.[11,12]

In recent years, WiFi-based HAR has emerged as a promising alternative to vision-based and wearable-sensor-based approaches.[13] WiFi sensing enables device-free activity recognition, overcoming the limitation of requiring users to carry sensors or remain within the field of view of cameras. Compared to existing methods, WiFi-based approaches are robust to lighting conditions and Line-of-Sight (LOS) constraints. WiFi-based HAR commonly relies on two complementary signal representations: Received Signal Strength (RSS) and Channel State Information (CSI). RSS is available on almost all WiFi devices at the MAC layer, but provides coarse packet-level measurements with low sensitivity. In contrast, CSI is obtained at the physical layer on limited devices and offers high sensitivity through fine-grained multipath information with higher temporal resolution. Also, WiFi-based HAR systems still suffer from several factors that degrade performance, including moving objects, human-related variability, and signal interference from other wireless devices. Additionally, noise, hardware heterogeneity, and device placement variations can introduce inconsistencies in CSI measurements. Temporal variations, such as changes in occupancy or environmental conditions over time, further reduce model robustness and generalization

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh. E-mail: hafizimtiaz@eee.buet.ac.bd

capability.[10] As a result, recent research has predominantly focused on incorporating CSI for HAR for its better generalization in complex real-world scenarios.[14,15]

Alongside the rapid advancements, privacy has become a pivotal factor for the deployment of HAR technology.[16,17] Vision-based HAR systems inherently expose sensitive visual information of the user in homes and healthcare setups. Recent deepfake generation technologies can readily alter facial data, significantly amplifying the privacy risk of vision-based systems.[18] Audio-based sensors pose privacy risks by continually capturing speech and ambient sounds, which may reveal sensitive personal information.[19] Inertial Measurement Unit (IMU)-based approaches can reveal sensitive personal attributes (activities and demographics) and behaviors, and remain vulnerable to inference attacks and leakage during data transmission and processing.[20] Although CSI-based HAR avoids direct visual and on-body sensing like traditional approaches, adversaries can still exploit CSI measurements to infer sensitive activities, occupancy patterns, and behavioral routines.[21] Deployed deep learning–based HAR systems are vulnerable to attacks such as model inversion,[22] membership inference,[23] and adversarial manipulation,[24] which allow adversaries to infer private user information from model outputs, parameters or perturbed inputs, compromising data confidentiality.[25] Consequently, ensuring strong privacy protection across all HAR modalities, particularly in a CSI-based framework, has become essential for trustworthy deployment.

Motivated by these challenges, we propose a robust and privacy-preserving CSI-based HAR framework that provides formal and quantifiable privacy guarantees through differential privacy (DP). We design a lightweight convolutional neural network (CNN) augmented with a temporal attention mechanism to effectively capture discriminative temporal patterns from CSI sequences without relying on conventional recurrent architectures (RNN, LSTM, and GRU) to avoid computational bottlenecks. To protect sensitive training data, we integrate DP training using the Gaussian mechanism,[26] ensuring that the learning process satisfies (ε, δ)-differential privacy. We conducted extensive experiments on multiple CSI HAR benchmark datasets – namely the CSI-HAR,[27] WiAR,[28] and CSLOS datasets.[29] The datasets consist of variations in distance, antenna height, and environmental conditions. Experimental results demonstrate that our proposed model achieves high recognition accuracy while maintaining strong privacy guarantees. Notably, across several configurations, DP training incurs only marginal performance degradation (7–8%) and, in some cases, improves generalization due to its regularization effect.[30] These findings indicate that accurate, efficient, and privacy-preserving CSI-based HAR is achievable through careful algorithm and model design. The main contributions of our work are summarized as follows:

- We propose an end-to-end CNN–temporal attention architecture for CSI-based HAR that avoids recurrent networks while effectively modeling temporal dynamics.
- We design a unified preprocessing and training pipeline that supports distance-, height-, and environment-variant CSI data.
- We integrate differential privacy into the training process using the Gaussian mechanism and analyze the privacy–utility trade-off through an extensive parameter sweep.
- We demonstrate our proposed model's robust and privacy-preserving performance across multiple publicly available datasets and heterogeneous sensing conditions, highlighting the practicality of the proposed framework for real-world deployment.

1.1 Notations

Scalars are denoted by lowercase letters (*e.g.*, t and k), vectors by bold lowercase letters (*e.g.*, $\mathbf{x}$), and matrices by bold uppercase letters (*e.g.*, $\mathbf{X}$). For a matrix $\mathbf{X}\in\mathbb{R}^{T\times S}$, T and S denote the number of time frames and CSI subcarriers, respectively, and $\mathbf{x}_t$ represents the CSI vector at time index t. The Gaussian distribution with mean μ and variance σ^2 is represented by $\mathcal{N}(\mu,\sigma^2)$.. In the context of differential privacy, ε and δ denote the privacy parameters, and $\mathcal{M}$ represents a randomized mechanism. Probabilities are denoted by $\Pr(\cdot)$, and expectation by $\mathbb{E}[\cdot]$.

2 Related studies

HAR has been widely explored using diverse sensing modalities, including vision-based, audio-based, wearable sensors, and wireless signal-based approaches. Vision-based HAR approaches have evolved from probabilistic and machine learning methods (*e.g.*, HMM, SVM, and KNN) to state-of-the-art deep learning models leveraging transfer learning such as CNNs, RNNs, and autoencoders, enabling automatic feature learning and robust activity recognition.[31,32] For example, Lin *et al.*[33] proposed Spike-HAR and Spike-HAR++, end-to-end spiking neural network architectures with spiking attention and transformer blocks, to efficiently perform event-based human activity recognition. Liu *et al.* proposed an Aligned Compressed Event (ACE) tensor representation and a Branched Event Network (BEN) to address noise, sparsity, and temporal misalignment in event-based vision and achieved the fastest inference speed of 3.28 ms and 304 fps.[34] Integrating audio with visual information enhances HAR by providing complementary temporal and contextual insights and improving robustness of the system.[35] Kim *et al.*[36] introduced an audio event recognition model for elderly people in household events. Shaikh *et al.* proposed MAiVAR, a multimodal audio–vision fusion framework that jointly learns both audio and visual features.[37] However, audio-visual HAR systems are prone to high computational cost, data-hungry behavior, and prolonged training times, leading to significant privacy concerns, limit scalability, and deployment in practical settings.[38]

To address the aforementioned challenges, wearable sensor-based HAR has been widely adopted in existing literature for its low-cost and privacy-aware sensing of human movements and real-time performance.[39,40] Xu *et al.* introduced a multi-frequency channel attention framework for HAR that replaces global average pooling with DCT-based compression, enabling more informative yet lightweight channel modeling and achieving state-of-the-art performance on multiple benchmark

datasets.[41] Recently, Arafa *et al.* proposed a hybrid CNN–Bi-LSTM framework with oversampling techniques to address class imbalance and spatiotemporal modeling challenges in wearable sensor–based HAR.[42] Bigelli *et al.*[43] introduced an energy-efficient privacy-preserving HAR approach using an autoencoder to obfuscate sensitive attributes while retaining high recognition accuracy on low-power devices. Despite these advantages over vision-based HAR, sensor-based HAR approaches require continuous wearing of devices and reliable connectivity, which can reduce user comfort, increase maintenance overhead, and limit long-term deployment.[44]

CSI-based HAR offers a promising device-free alternative using WiFi signals to recognize human activity without continuous user participation.[45] Chen *et al.*[46] introduced an attention-based BiLSTM model that leverages temporal dependencies in raw WiFi CSI sequences, improving the accuracy of activity recognition through adaptive feature weighting. Ding *et al.* proposed HARNN, a CSI-based framework that combines environmental detection, noise suppression, handcrafted statistical features, and an RNN to robustly model activity–channel relationships in indoor settings.[47] Wang *et al.* developed MCBAR, a multimodal CSI-based HAR system that employs GAN-driven semi-supervised learning to mitigate performance degradation caused by environmental dynamics and data imbalance.[48] Addressing robustness and privacy-aware deployment, Yadav *et al.* presented CSITime, a lightweight attention-enhanced CNN architecture with data augmentation strategies that achieves state-of-the-art performance and strong generalization without relying on intrusive sensing modalities.[49]

To address the abovementioned limitations in current HAR research practice, our work focuses on jointly addressing model efficiency, privacy preservation, and robustness. The proposed framework integrates temporal attention within a lightweight CNN while avoiding computationally expensive recurrent architectures, enabling effective temporal modeling with reduced overhead. Furthermore, the incorporation of DP provides formal privacy guarantees, mitigating inference risks overlooked in prior CSI-based approaches.

3 Background and problem formulation

In this section, we briefly review the fundamental concepts underlying CSI-based HAR and DP, which form the theoretical basis of this work.

3.1 Multipath propagation

In a typical wireless communication system, the transmitted signal propagates from the transmitter to the receiver through multiple paths due to reflection, diffraction, and scattering caused by surrounding objects. This phenomenon is known as multipath propagation, where signal components are attenuated in power, delayed in time, and shifted in phase or frequency.[50] Based on small-scale fading models, multipath propagation can be broadly categorized into line-of-sight (LOS) and non-line-of-sight (NLOS) propagation. LOS channels are modeled using Rician fading due to the presence of a dominant direct path, whereas NLOS channels are characterized by Rayleigh fading models to describe the statistical time-varying nature of the wireless signal.[51] To this end, human motion can dynamically perturb these propagation paths by introducing additional reflections and obstructions, thereby inducing measurable variations in the received signal characteristics that can be exploited for activity recognition. Mathematically, the received baseband signal $r(t)$ can be expressed as[52]

$$r(t) = \sum_{i=1}^{N} a_i(t)s(t - \tau_i(t))e^{j\phi_i(t)}, \quad (1)$$

where $s(t)$ denotes the transmitted signal, and $a_i(t)$, $\tau_i(t)$, and $\phi_i(t)$ represent the time-varying amplitude attenuation, propagation delay, and phase shift of the i-th propagation path, respectively. In LOS environments, a dominant propagation component exists with significantly larger $a_i(t)$, leading to a Rician fading distribution. In contrast, NLOS environments lack a dominant path, and a Rayleigh distribution explains the fading amplitudes well.

3.2 Orthogonal frequency division multiplexing

Orthogonal Frequency Division Multiplexing (OFDM) is a multicarrier transmission technique widely used in modern WiFi systems. Instead of transmitting data over a single wideband channel, OFDM divides the available bandwidth into multiple orthogonal narrowband subcarriers for accommodating multiple users.[53] This design improves spectral efficiency and mitigates inter-symbol interference (ISI) by ensuring orthogonality among subcarriers.[54] In IEEE 802.11 systems, each subcarrier experiences different channel conditions, enabling fine-grained channel measurements that can be exploited for HAR applications.[55]

3.3 Channel state information

Channel State Information (CSI) describes the frequency-domain response of a wireless channel and captures the effects of multipath propagation. For an OFDM system, the CSI of the k-th subcarrier can be expressed as[56]

$$H(k) = |H(k)|e^{j\angle H(k)}, \quad (2)$$

where $|H(k)|$ and $\angle H(k)$ denote the amplitude and phase, respectively. Phase measurements are often corrupted by carrier frequency offset and sampling frequency offset.[57] Consequently, CSI amplitude is widely adopted in HAR due to its stability. Human activities perturb the multipath structure of the wireless channel, producing distinctive temporal patterns in CSI measurements. These patterns can be used to recognize activities without the use of devices.[40,58] As such, we illustrate a typical CSI-based human activity recognition pipeline in Fig. 1, where raw WiFi CSI signals are captured, processed, and fed into a deep learning model for activity classification.

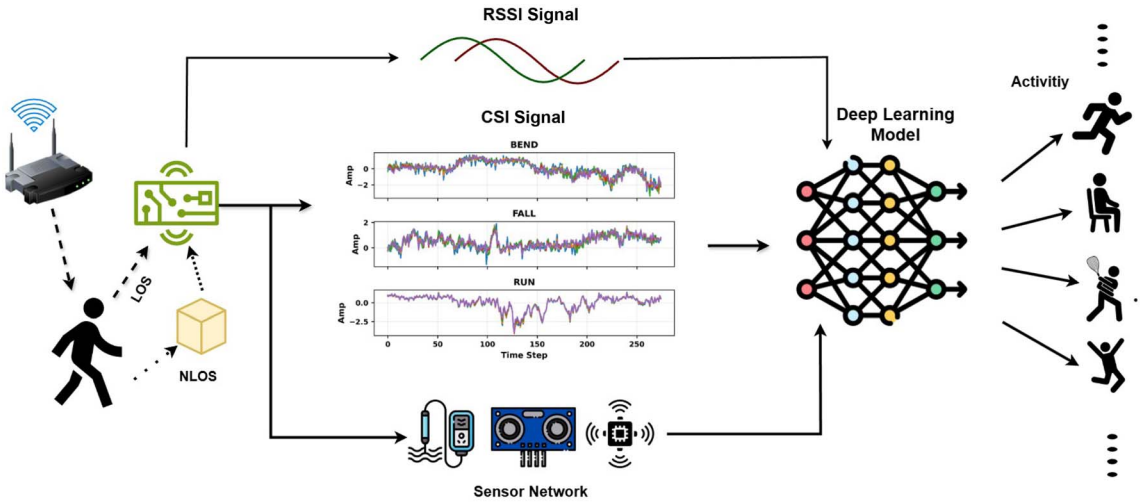

Fig. 1 A typical human activity recognition pipeline.

3.4 Differential privacy

Signal processing and ML algorithms often operate on sensitive data and may leak private information about individuals through their learned representations or outputs.[59] Differential privacy[26,60] provides strong protection by ensuring that the outcome of a randomized algorithm is statistically indistinguishable regardless of whether any single individual's data are included in the input dataset. This property offers robustness against a wide range of privacy attacks, including those leveraging auxiliary or background information.

Definition 3.1 (Differential privacy[26]). A randomized algorithm $\mathscr{M}: \mathscr{D} \to \mathscr{T}$ satisfies (ε, δ)-differential privacy if for all measurable subsets $S \subseteq \mathscr{T}$ and for all datasets $D, D' \in \mathscr{D}$ that differ in at most one entry, it holds that

$$\Pr(\mathscr{M}(D) \in S) \leq e^{\varepsilon} \Pr(\mathscr{M}(D') \in S) + \delta. \quad (3)$$

Here, the privacy parameter ε controls the privacy budget, with smaller values indicating stronger privacy guarantees, while δ can be interpreted as the probability that the privacy guarantee may be violated. In practice, δ is chosen to be negligibly small, typically on the order of the squared inverse of the dataset size.[26] Several mechanisms exist for achieving DP, including input perturbation, output perturbation, and objective perturbation.[61] In this work, we adopt the Gaussian mechanism, which ensures differential privacy by injecting calibrated Gaussian noise into the learning process.[62] During training, per-sample gradients are first clipped to a fixed norm to bound their sensitivity, after which Gaussian noise proportional to the clipping threshold is added.[63] This approach guarantees (ε, δ)-differential privacy while allowing effective optimization of DNN.

4 Proposed framework

4.1 Exploratory data analysis (EDA)

First, we conduct an EDA of CSI data to gain insights into its intrinsic characteristics and sensitivity, focusing on the walking activity across varying distances, antenna heights, and environments. The analysis is based on time–subcarrier CSI amplitude representations.

4.1.1 Effect of distance. As can be seen in the spectrograms in Fig. 2, CSI patterns exhibit clear distance-dependent behavior. At a short distance of 1 m, the CSI amplitudes demonstrate moderate fluctuations with localized temporal variations, indicating strong LOS and near-field multipath components. At 3 m, the CSI responses become more dynamic, with increased temporal variability and deeper amplitude fades, reflecting enhanced sensitivity to human motion. In contrast, at 6 m, the CSI signals appear more uniform across subcarriers with reduced temporal contrast due to signal attenuation and spatial averaging. These observations suggest that mid-range distances provide richer discriminative motion cues for activity recognition.[64]

4.1.2 Impact of antenna height. The spectrograms in Fig. 3 illustrate the influence of antenna height on CSI signal characteristics. At 60 cm, the CSI amplitudes remain relatively stable with limited temporal modulation, indicating weaker interaction between walking motion and dominant signal paths. At 90 cm, pronounced temporal transitions and subcarrier-specific streaks emerge, corresponding to strong Doppler and shadowing effects caused by limb movements intersecting the Fresnel zone. At 120 cm, the CSI patterns become smoother while maintaining consistent temporal variations, suggesting a balance between motion sensitivity and signal stability. This confirms antenna height as a critical factor in capturing discriminative CSI features.[65]

4.1.3 Environmental variability. CSI signal responses across different environments, shown in the spectrograms of Fig. 4, reveal noticeable distributional shifts. Although the temporal structure of walking activity remains consistent, each environment introduces unique subcarrier-level amplitude variations due to differences in layout. Certain subcarriers consistently exhibit higher sensitivity to motion, whereas others remain relatively invariant. This environment-dependency highlights the domain shift challenge in CSI-based HAR.[66,67]

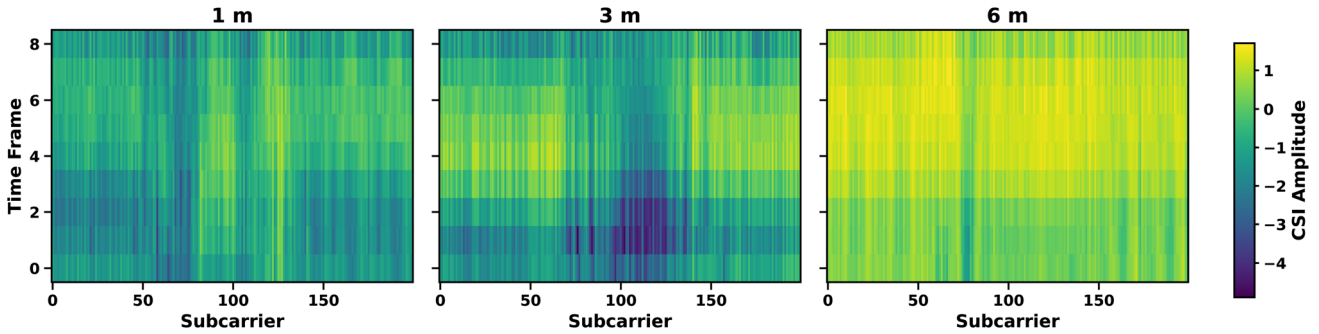

Fig. 2 Spectrograms of CSI data for walking activity across different transmitter–receiver distances (1 m, 3 m, and 6 m) of the WiAR dataset.

4.2 Data preprocessing

We applied a consistent preprocessing and augmentation strategy to all the datasets under investigation to ensure robustness and fair comparison.[68] Since CSI data recordings corresponding to different activities and trials exhibit varying temporal lengths, the minimum number of time frames among all recordings is first identified. Each sample is then truncated to this minimum length to ensure uniformity across samples. Although truncation to a fixed length may introduce partial information loss, the use of overlapping sliding windows enables effective recovery of discriminative temporal patterns from CSI sequences, enabling consistent windowing and batch processing without introducing zero padding or interpolation. Then, each truncated sample is normalized independently using z-score normalization to mitigate subcarrier-wise amplitude bias and environment-induced scale variations. Specifically, for a CSI matrix $\mathbf{X}\in\mathbb{R}^{T\times S}$, where T and S denote the number of time frames and subcarriers, respectively, the normalization is performed as

$$\mathbf{X}_{\text{norm}}{}^{(i)} = \frac{\mathbf{X}^{(i)} - \mu^{(i)}}{\sigma^{(i)} + \eta}, \tag{4}$$

where $\mathbf{X}^{(i)}$ denotes the i-th CSI sample, and $\mu^{(i)}$ and $\sigma^{(i)}$ represent the mean and standard deviation computed from the i-th sample, respectively. The constant η is added for numerical stability to prevent division by zero, ensuring that each sample is normalized to have approximately zero mean and unit variance, which improves convergence during training. Each normalized sample is segmented using a sliding window approach. Fixed-length windows of 128 time frames are extracted with a stride of 32 frames, producing overlapping segments that preserve temporal continuity, while increasing the number of training samples. Each windowed segment retains the original subcarrier dimension, forming a two-dimensional time–frequency representation. Gaussian noise is injected into each windowed CSI segment as a form of data augmentation to enhance model robustness and reduce overfitting. Specifically, for each original windowed data $\mathbf{W}$, an augmented version is generated as

$$\mathbf{W}' = \mathbf{W} + \mathcal{N}(0, \sigma_{\text{n}}{}^2), \tag{5}$$

where $\mathcal{N}(0, \sigma_n{}^2)$ denotes zero-mean Gaussian noise with standard deviation σ_{n}. We note that in our experiments, we used $\sigma_{\text{n}} = 0.03$. All windowed CSI segments are stacked to form the final input tensor $\mathbf{X}\in\mathbb{R}^{N\times 128\times S}$, where N denotes the total number of segments and S represents the number of subcarriers. Corresponding activity labels are encoded into a one-dimensional label vector $\mathbf{y}\in\mathbb{R}^{N}$. This preprocessing pipeline ensures consistent temporal structure, normalized feature distributions, and enhanced data diversity for reliable CSI-based HAR.[69,70]

4.3 Model architecture

In the following, we describe the proposed lightweight CNN model with a temporal attention mechanism, designed to efficiently capture discriminative spatio-temporal patterns from CSI data without relying on recurrent layers. The architecture processes CSI inputs of shape $128 \times S$. Our proposed network begins with two 1D convolutional layers that operate along the

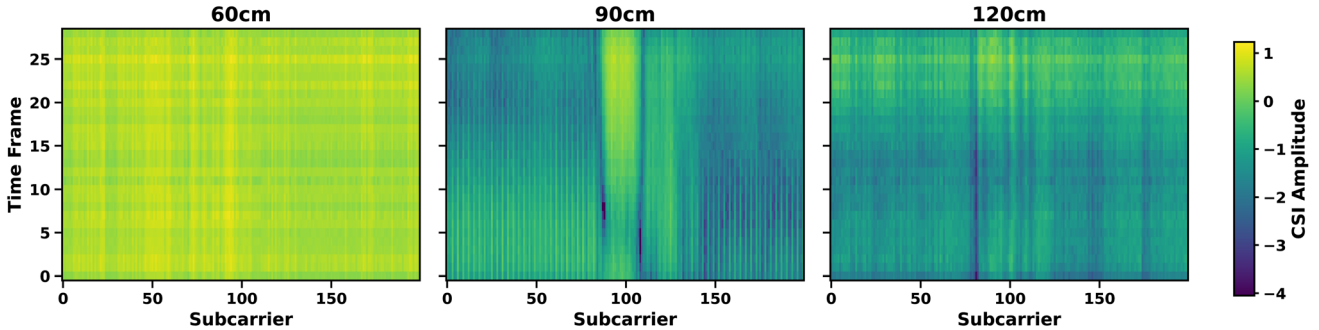

Fig. 3 Spectrograms of CSI data for walking activity captured at different antenna heights (60 cm, 90 cm, and 120 cm) of the WiAR dataset.

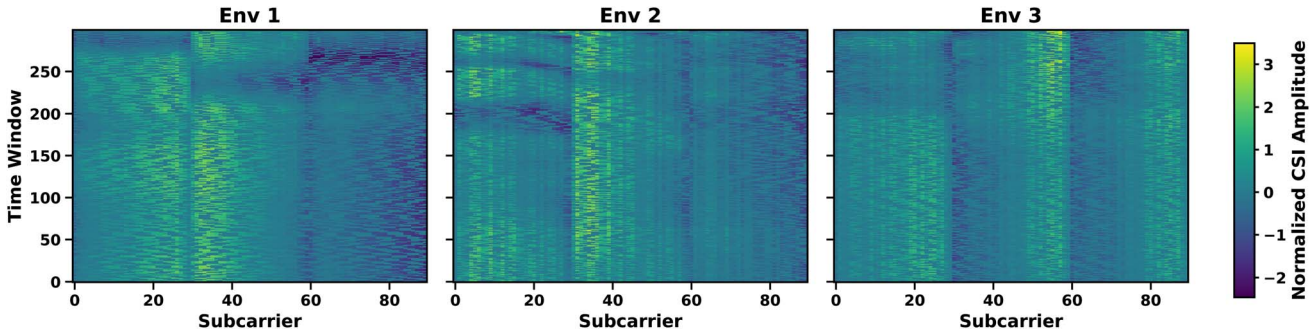

Fig. 4 Normalized spectrograms of CSI data for walking activity across three different environments of the CSLOS dataset.

temporal dimension to extract local temporal dependencies across subcarriers. Each of the convolutional layers is followed by a rectified linear unit (ReLU) activation and group normalization to stabilize training. A max-pooling layer reduces the temporal resolution by a factor of two, enabling hierarchical feature abstraction while reducing computational complexity. Dropout is applied to mitigate overfitting.[71] Next, a temporal attention module is applied to model long-range temporal dependencies.[72] More specifically, the attention mechanism assigns adaptive importance weights to individual time steps, allowing the network to emphasize temporally salient CSI segments associated with human activities. The weighted features are aggregated into a fixed-length representation. Finally, the aggregated feature vector is passed through a fully connected (FC) layer with ReLU activation to an output layer for classification. The details of the network (as shown in Fig. 5) are provided in Table 1. Overall, the model contains only 58 824 trainable parameters, requiring approximately 5.29 million multiply–add operations per inference. We note that the reported parameter counts correspond to the 7-class activity recognition setting. Only the final classification layer parameters change for experiments with 12 and 16 activity classes. Additionally, the total model size is ∼0.22 MB – making it suitable for deployment in resource-constrained edge-devices and privacy-sensitive environments.[73] While the model is lightweight for inference, we note that DP-based training may introduce additional computational overhead due to per-sample gradient operations, which are typically handled in offline or high-resource environments.

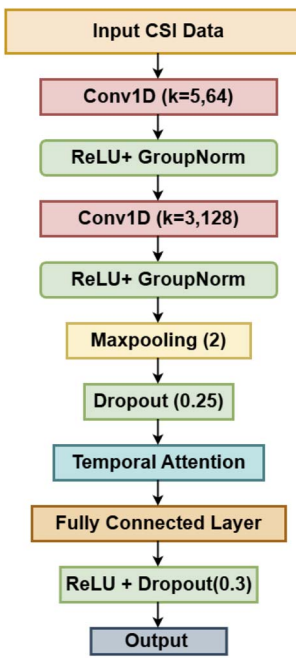

Fig. 5 Proposed CNN–temporal attention architecture for CSI data. The model comprises stacked 1D convolutional layers with group normalization and ReLU activations, followed by temporal attention and fully connected layers for the final prediction.

4.4 Training and evaluation

In this section, we describe the training configuration, evaluation protocol, and performance assessment procedures used to validate the proposed CSI-based HAR framework. Identical data splits, preprocessing pipelines, and model architectures are used across all experiments.

4.4.1 Non-privacy-preserving (baseline) model training. The baseline experiments are conducted using the proposed CNN–temporal attention model trained without privacy constraints. For all datasets, CSI samples are divided into training and test sets using an 80 : 20 stratified split to preserve class distribution. The model is optimized using the AdamW optimizer with an initial learning rate of 10^{-3} and a weight

Table 1 Architecture summary of the proposed CNN–temporal attention network

Layer	Input shape	Output shape	Parameters
Conv1D ($k = 5$, 64)	52 × 128	64 × 128	16 704
ReLU + GroupNorm	64 × 128	64 × 128	128
Conv1D ($k = 3$, 128)	64 × 128	128 × 128	24 704
ReLU + GroupNorm	128 × 128	128 × 128	256
Max pooling (2)	128 × 128	128 × 64	—
Dropout (0.25)	128 × 64	128 × 64	—
Temporal attention	64 × 128	128	8321
Fully connected	128	64	8256
ReLU + dropout	64	64	—
Output layer	64	7	455
Total	–	–	**58 824**

decay of 10^{-4}. Cross-entropy loss with label smoothing is employed to mitigate overconfidence in class predictions. A cosine annealing learning rate scheduler with warm restarts is used to facilitate stable convergence.[74] We employed early stopping based on validation accuracy across the datasets – if the validation performance does not improve for 15 consecutive epochs, training is terminated to prevent overfitting. Model checkpoints corresponding to the best validation accuracy are retained for final evaluation.

4.4.2 Differentially-private model training. DP is incorporated into the training process using the Opacus framework.[75] The same model architecture and data splits used in the baseline experiments are retained to enable a fair comparison between private and non-private training regimes. During DP training, per-sample gradients are computed and clipped to a fixed ℓ_2 norm to bound the sensitivity.[63] Gaussian noise is then added to the clipped gradients, ensuring that the training procedure satisfies (ε, δ)-differential privacy.[76] The privacy parameter δ is fixed to 10^{-5} for all experiments, while the privacy budget ε is varied to analyze the privacy–utility trade-off. The DP model is trained for 50–100 epochs using the AdamW optimizer with a learning rate of 10^{-3}. The effect of DP on recognition performance is analyzed by comparing test accuracy across varying privacy budgets with the non-private baseline. Lower ε values correspond to stronger privacy guarantees but introduce higher levels of noise during optimization, which can reduce classification accuracy.[26] Conversely, larger ε values relax privacy constraints and allow the model to approach non-private performance.

4.5 Computational complexity and efficiency metrics

We analyzed several computational and system-level metrics, including FLOPs, MACs, latency, throughput, memory usage, and energy-related indicators to assess the practical deployability of the proposed model. These metrics are profiled in a Kaggle execution environment using PyTorch-based profiling tools under a GPU/CPU runtime, ensuring reproducibility and consistency.

4.5.1 FLOPs and MACs. The number of floating-point operations (FLOPs) and multiply–accumulate operations (MACs) quantify the computational cost of the model. For a convolutional layer, the MACs can be estimated as:

$$\text{MACs} = K \times C_{\text{in}} \times C_{\text{out}} \times L_{\text{out}}$$

where K is the kernel size, C_{in} and C_{out} denote the number of input and output channels, and L_{out} is the output length. FLOPs are typically approximated as:

$$\text{FLOPs} \approx 2 \times \text{MACs}$$

4.5.2 Latency and throughput. Latency measures the time required to process a single input sample:

$$\text{Latency} = \frac{\text{Total inference time}}{\text{Number of samples}}$$

Throughput represents the number of samples processed per second:

$$\text{Throughput} = \frac{1}{\text{Latency}}$$

4.5.3 Memory usage. Memory consumption includes both model parameters and intermediate activations during inference:

$$\text{Memory} = \text{model parameters} + \text{activation memory}$$

4.5.4 Energy proxy. The energy proxy provides an approximate indication of energy consumption and is typically proportional to computational cost:

$$\text{Energy proxy} \propto \text{MACs} \times \text{memory access cost}$$

4.5.5 Real-time factor (RTF). RTF measures the ratio between processing time and input duration:

$$\text{RTF} = \frac{\text{Processing time}}{\text{Input duration}}$$

4.5.6 Parameter density. Parameter density reflects the compactness of the model relative to its representational capacity:

$$\text{Parameter density} = \frac{\text{Number of parameters}}{\text{Model size(MB)}}$$

5 Experimental results

5.1 Evaluation on the WiAR dataset

5.1.1 Performance variation with distance. As discussed in Section 4.1, the WiAR dataset[28] contains CSI data variations with different distances. First, we analyze the performance of the proposed model architecture (both without and with DP constraints) with varying distances from the receiver. In Fig. 6, we show the classification accuracy of the WiAR system across different transmitter–receiver distances with and without DP. Additionally, in Table 2, we summarize the classification performance with recall and F1-score under the distance settings for the WiAR dataset. At a distance of 1 m, the proposed CNN–temporal attention model achieved near-perfect non-DP performance, with an overall accuracy of 95% and recall, and F1-scores of 0.96 and 0.95, respectively. This indicates that, under close-range conditions with no privacy constraints, the CSI measurements provide highly discriminative patterns for activity recognition. As the distance increased to 3 m, the accuracy slightly decreased to 94%, accompanied by recall and F1-scores of 0.93 and 0.94, respectively. This performance degradation can be attributed to increased multipath complexity and reduced signal sensitivity to fine-grained human motion at moderate distances. At a distance of 6 m,

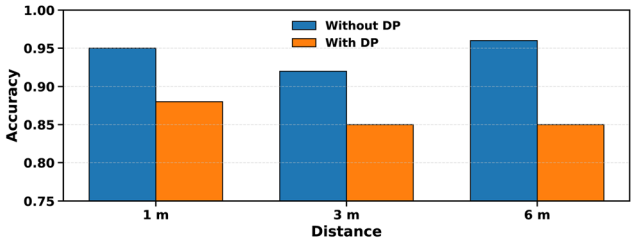

Fig. 6 Classification accuracy across different transmitter–receiver distances on the WiAR dataset without and with DP ($\varepsilon = 12$, $\delta = 10^{-5}$).

Table 2 Performance comparison of the proposed model on the WiAR dataset across different transmitter–receiver distances. DP results are reported at $\varepsilon = 12$ and $\delta = 10^{-5}$. Recall and F1-score are reported for the non-private (Non-DP) setting

Distance	DP Acc	Non-DP Acc	Recall	F1-score
1 m	0.88	0.95	0.96	0.95
3 m	0.85	0.94	0.93	0.94
6 m	0.85	0.96	0.95	0.96

the model achieved an accuracy of 96%, demonstrating a recovery in performance compared to the 3 m setting. This suggests that the proposed architecture effectively captures robust temporal patterns in CSI data even at extended sensing distances, potentially benefiting from more stable multipath structures in the far-field region.

In Fig. 7, we illustrate the privacy–utility trade-off, that is, the variation of performance with the overall privacy budget ε for different distance conditions. The figure reaffirms that lower ε values (corresponding to stronger privacy guarantees) introduce higher levels of noise and thereby reduce classification accuracy. On the other hand, larger ε values relax privacy constraints and allow the model to approach non-private performance, irrespective of distance settings.

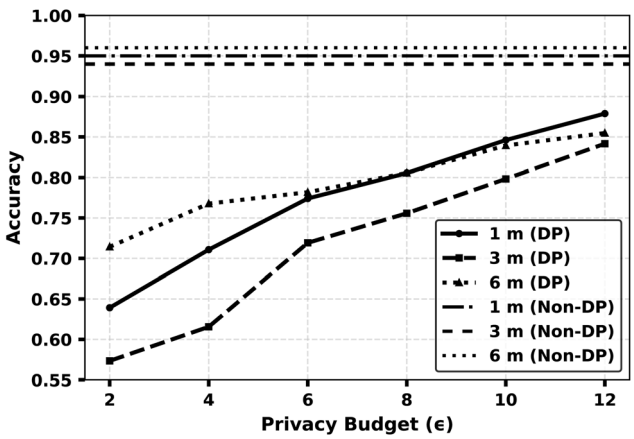

Fig. 7 Privacy–utility trade-off on the WiAR dataset under different distance settings: variation of classification accuracy with the privacy budget ε.

5.1.2 Performance variation with height. Next, we investigate the variation in performance of our proposed model on the WiAR dataset using three receiver height configurations to observe the impact of antenna height: 60 cm, 90 cm, and 120 cm. In Fig. 8, we show the classification accuracy of the WiAR system across different antenna heights with and without differential DP. Additionally, in Table 3, we tabulate the classification performance of the proposed CNN–temporal attention model under both non-private and DP training settings. In the non-private setting, the model achieved consistently high accuracy across different heights, with non-DP accuracies of 99%, 97%, and 99% at 60 cm, 90 cm, and 120 cm, respectively. The corresponding recall and F1-scores indicate balanced classification performance across activity classes, demonstrating the robustness of the proposed architecture to variations in antenna height. When privacy constraint is incorporated into the training process, only a slight reduction in accuracy was observed. Specifically, the model achieved accuracies of 92%, 91%, and 93% for heights of 60 cm, 90 cm, and 120 cm, respectively.

In Fig. 9, we illustrate the variation in performance with the overall privacy budget ε for different height configurations. As in the experiments under different distance conditions, the figure clearly demonstrates that lower ε values reduce classification accuracy. On the other hand, larger ε values allow the model to approach non-private performance, irrespective of height configurations.

5.2 Evaluation on the CSLOS dataset

We conduct experiments on the CSLOS dataset[29] across three distinct indoor environments to assess the robustness of the proposed CNN–temporal attention model. These environments exhibit varying multipath characteristics, spatial layouts, and

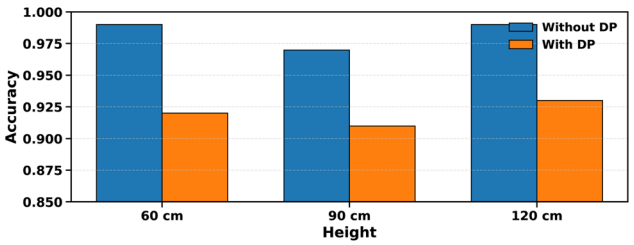

Fig. 8 Classification accuracy across different antenna heights on the WiAR dataset without and with DP ($\varepsilon = 12$; $\delta = 10^{-5}$).

Table 3 Performance comparison of the proposed model on the WiAR dataset under different antenna height configurations. DP accuracies correspond to $\varepsilon = 12$ and $\delta = 10^{-5}$, while recall and F1-scores are reported for the non-private (Non-DP) setting

Height	DP Acc	Non-DP Acc	Recall	F1-score
60 cm	0.92	0.99	0.98	0.99
90 cm	0.91	0.97	0.96	0.97
120 cm	0.93	0.99	0.99	0.99

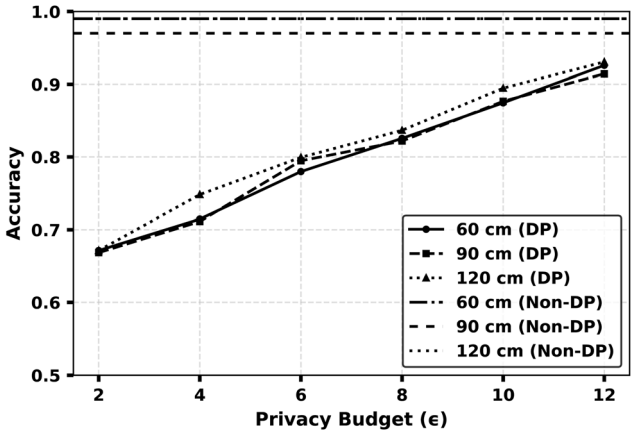

Fig. 9 Privacy–utility trade-off on the WiAR dataset under different height settings: variation of classification accuracy with the privacy budget ε.

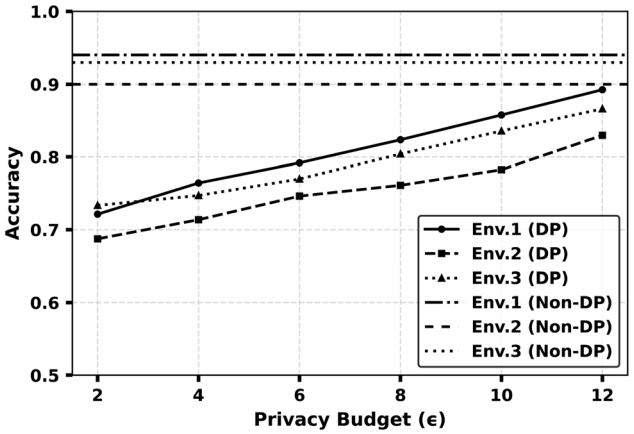

Fig. 11 Privacy–utility trade-off on the CSLOS dataset across different environments, showing classification accuracy as a function of the privacy budget ε.

Table 4 Performance comparison of the proposed model on the CSLOS dataset across different environments. DP results are obtained with $\varepsilon = 12$ and $\delta = 10^{-5}$. Recall and F1-scores are reported for the non-private (Non-DP) setting

Environment	DP Acc	Non-DP Acc	Recall	F1-score
Env.1	0.89	0.94	0.93	0.94
Env.2	0.82	0.90	0.90	0.90
Env.3	0.87	0.93	0.93	0.94

levels of environmental clutter. In Table 4 and Fig. 10, we summarize the classification performance under both non-DP and DP training settings on the CSLOS dataset across different environments. In the non-private setting, the proposed model achieved accuracies of 94%, 90%, and 93% for Environments 1, 2, and 3, respectively. The performance variation across environments reflects differences in signal propagation conditions and multipath complexity inherent to each environment. When DP was incorporated into the training process, the model demonstrated slight performance degradation across the environments. Specifically, accuracies decreased to 89% in Environment 1, 82% in Environment 2, and 87% in Environment 3. The corresponding macro-averaged and weighted F1-scores remained consistent, indicating balanced recognition performance across activity classes.

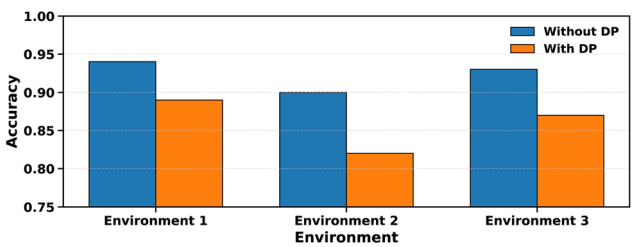

Fig. 10 Classification accuracy across different environments on the CSLOS dataset without and with DP ($\varepsilon = 12$; $\delta = 10^{-5}$).

In Fig. 11, we depict the privacy–utility trade-off by analyzing how classification accuracy changes with the privacy budget ε under different environmental conditions. As in the experiments on the WiAR dataset, the figure demonstrates that lower ε values (stronger privacy guarantees) reduce classification accuracy. On the other hand, larger ε values allow the model to approach non-private performance, irrespective of height configurations, at the cost of weaker privacy.

5.3 Evaluation on the CSI-HAR dataset

Finally, we investigate the performance of the proposed framework on the CSI-HAR dataset[27] to assess its effectiveness under standard indoor sensing conditions. In Table 5, we summarize the classification performance achieved with and without DP constraints. In the non-private setting, the proposed model achieved an overall classification accuracy of 98%, demonstrating strong baseline performance. When DP was incorporated into the training process, the model achieved an accuracy of 91%. In Fig. 12, we present the privacy–utility trade-off of the proposed CSI-HAR framework by showing how classification accuracy varies with the privacy budget ε. We observed a similar trend for this dataset as in the previous cases.

5.4 Performance variation with training sample size

We investigated the effect of training sample size on classification performance and privacy guarantees under DP on the CSI HAR dataset[27] and WiAR dataset.[28] For the WiAR dataset's distance variation, we took the 1 m subset, and for the height

Table 5 Performance comparison of the proposed model on the CSI-HAR dataset with and without differential privacy. DP results are reported at $\varepsilon = 12$ and $\delta = 10^{-5}$

Setting	Accuracy	Recall	F1-score
DP training	0.91	0.90	0.91
Non-DP training	0.98	0.98	0.98

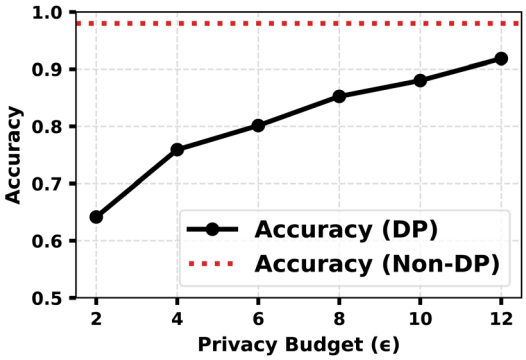

Fig. 12 Privacy–utility trade-off for the proposed CSI-HAR framework, illustrating the impact of the privacy budget ε on classification accuracy.

dataset, we took the 60 cm subset. More specifically, we trained the proposed model with different numbers of training samples using DP-SGD,[63] implemented *via* the Opacus framework,[75] and then tested it on the test set. For each configuration, the resulting privacy budget ε was computed while keeping the noise multiplier and norm clipping parameters fixed to enable fair comparison across different training sample sizes. Specifically, the noise multiplier was set to $\sigma = 1.0$, and the clipping norm was fixed to $C = 1.0$.

As shown in Fig. 13 and 14, increasing the number of training samples consistently improves classification accuracy while maintaining a lower privacy cost. Specifically, Fig. 13 illustrates the performance variation with the number of training samples for the CSI HAR dataset,[27] whereas Fig. 14 shows the corresponding performance trends for the WiAR dataset. For Fig. 14(a), the 60 cm height subset, and for Fig. 14(b), the 1 m distance subset was considered. When the model is trained with fewer samples, the accuracy remains relatively low due to limited data diversity and the regularizing effect of DP noise. As the training dataset grows, the model learns more robust and discriminative features, resulting in

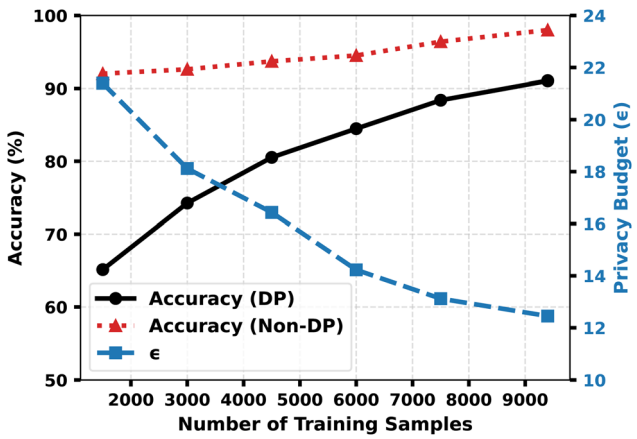

Fig. 13 Effect of training sample size on the classification accuracy (left axis) and the overall privacy budget ε (right axis) under differential privacy for the CSI-HAR dataset.

accuracy improvements exceeding 90% for the largest sample size of the CSI HAR dataset[27] and near 90% for the distance and height variations of the WiAR dataset[28]. Simultaneously, the privacy budget ε decreases monotonically as the number of training samples increases. This reduction is attributed to privacy amplification through subsampling in DP-SGD, where each sample contributes proportionally less to the overall privacy loss as the dataset size grows. Consequently, stronger privacy guarantees are achieved when more training data are available.[77] These results demonstrate that increasing the training sample size yields a dual benefit: enhanced recognition performance and improved privacy protection. This finding is particularly relevant for privacy-sensitive CSI-based HAR, indicating that scalable data can effectively mitigate the privacy–utility trade-off imposed by DP.

5.5 Ablation analysis

Although our proposed model is quite light-weight, we conduct an ablation study on the CSI-HAR dataset[27] in the non-DP setting to quantify the contribution of the different architectural components of the proposed CNN–temporal attention model. In Table 6, we summarize the classification performance of different model variants. The full model, as shown in Fig. 5, combines two convolutional layers with temporal attention, and achieves the highest accuracy of 97.84%, demonstrating the effectiveness of the complete architecture. When the network was simplified to a single convolutional layer, the accuracy decreased to 94.38%, indicating that deeper convolutional feature extraction plays a critical role in capturing discriminative spatial–temporal patterns from CSI signals. Further performance degradation was observed when the temporal attention mechanism was removed, with accuracy dropping to 93.65%. This result highlights the importance of adaptive temporal weighting in emphasizing salient CSI segments associated with human activities. Without attention, the model treats all time steps equally, limiting its ability to focus on informative temporal regions. Overall, the ablation results confirm that both multi-layer convolutional feature extraction and temporal attention are essential components of the proposed architecture, jointly contributing to improved recognition accuracy and robust performance in CSI-based HAR.

5.6 Edge deployment and computational efficiency analysis

To evaluate the suitability of the proposed model for resource-constrained environments, we analyzed the computational efficiency, memory footprint, and inference performance of the proposed model. The detailed results are summarized in Table 7.

As shown in Table 7, the proposed model is highly efficient in terms of both computation and memory. With only 58 824 parameters and a compact model size of approximately 0.22 MB, the architecture is significantly smaller than typical deep HAR models. The computational cost is also low, requiring only 5.29 MFLOPs per inference, which enables fast execution. The model achieves a latency of 2.071 ms and a throughput of 482.81 samples per second, indicating its capability for real-

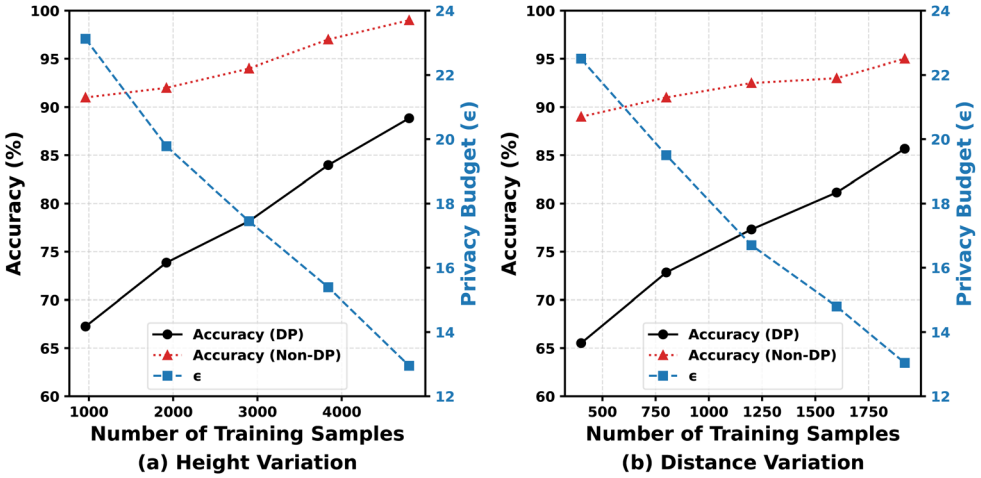

Fig. 14 Effect of training sample size on the classification accuracy (left axis) and the overall privacy budget ε (right axis) under differential privacy for the WiAR dataset: (a) for 60 cm height and (b) for 1 m distance.

Table 6 Ablation study evaluating the contribution of architectural components on the CSI-HAR dataset

Model variant	Accuracy (%)
Full model (CNN + attention)	97.84
One convolution layer	94.38
Without temporal attention	93.65

Table 7 Edge suitability analysis of the proposed model

Metric	Value
Total parameters	58 824
Model size	0.2244 MB
FLOPs	5.29 MFLOPs
MACs	5.29 MMACs
Latency	2.071 ms
Throughput	482.81 samples/sec
Memory usage	28.72 MB
Peak memory	28.72 MB
Energy proxy	10 961.85
Real-time factor	0.001618
Parameter density	0.011115

time processing. The real-time factor (0.0016) further confirms that the system can operate well within real-time constraints. Additionally, the memory footprint remains modest at 28.72 MB, making it suitable for deployment on edge devices with limited resources. Overall, these results demonstrate that the proposed framework offers an excellent balance between performance and efficiency, making it well-suited for real-world, resource-constrained, and privacy-sensitive deployment scenarios.

5.7 Comparison with existing studies

5.7.1 Comparison in the non-privacy-preserving (baseline) setting. In Table 8, we compare the proposed light-weight network with recent CSI-based HAR approaches. ABLSTM,[72] GraphHAR,[78] and TSHNN[79] report strong performance for 16 activities of the WiAR dataset, with TSHNN achieving the highest F1-score among prior studies. PWCNN[80] also demonstrates competitive precision. For reduced activity sets, ABiLSTM,[81] DenseLSTM,[82] and 2DCNN[27] exhibit relatively lower accuracy and incomplete metric reporting. Recent state-of-the-art models such as GrapHAR[78] and TSHNN[79] achieve very high recognition accuracy by employing more complex architectures. Specifically, GrapHAR models CSI sub-carrier correlations using graph attention networks combined with temporal causal convolutions, achieving approximately 98–99% accuracy with around 4.95 M parameters. Similarly, TSHNN adopts a computationally intensive design by converting CSI signals into video-like representations and applying 3D CNNs followed by GRU layers, achieving over 99% accuracy but with more than 1 M parameters and significantly higher computational cost. In contrast, our approach focuses on a computationally light-weight and privacy-preserving design, using only 58 824 parameters (~0.22 MB) while maintaining competitive performance (approximately 94–99%) across 7, 12, and 16 activity

Table 8 Comparison of HAR methods under non-DP conditions in terms of accuracy, precision, recall and F1-score

Method	Activities	Precision (%)	Recall (%)	F1-score (%)
ABLSTM[72]	16	97.75	95.44	95.59
GraphHAR[78]	16	97.88	97.76	97.82
TSHNN[79]	16	98.93	98.90	98.92
PWCNN[80]	16	97.03	—	97.00
ABiLSTM[81]	12	92.86	—	—
DenseLSTM[82]	7	94.70	—	—
2DCNN[27]	7	95.50	—	—
This work	7	**98.00**	**98.00**	**98.00**
	12	**93.50**	**93.00**	**94.00**
	16	**98.30**	**97.60**	**97.30**

Table 9 Comparison of HAR methods under DP conditions in terms of accuracy degradation – accuracy drop indicates performance loss relative to the corresponding non-DP baselines

Method	Activities	Accuracy drop (%)	(ε, δ)
Debaditya *et al.*[83]	18	9–11	$(8, 10^{-5})$–$(33, 10^{-5})$
Bigelli *et al.*[43]	6	≤ 5	—
Nadira *et al.*[85]	12, 16	10–12	$(15, 10^{-5})$
This work	7, 12, 16	**7–9**	$(12, 10^{-5})$

scenarios, thereby offering a more practical solution for resource-constrained and real-world deployments.

5.7.2 Comparison in the privacy-preserving setting. Next, we compare the performance of our proposed light-weight network in a privacy-constrained setting against existing studies. In Table 9, we present a comparative analysis of existing DP-based HAR methods in terms of the accuracy degradation induced by privacy preservation. Prior studies report a noticeable performance drop when DP mechanisms are incorporated into CSI-based or sensor-based HAR pipelines. Specifically, Debaditya *et al.*[83] introduced TEMPDIFF, and observed an accuracy reduction of approximately 9–11% for an 18-activity classification task. Bigelli *et al.*[43] achieved a relatively lower degradation ($\leq$5%) but considered only six activities for low-power IoT devices. In contrast, methods by Piran *et al.*[84] and Nadira *et al.*[85] reported higher performance losses ranging from 10–15% and 10–12%, respectively, as the number of activities increases. Compared to these approaches, our proposed method consistently limits the accuracy drop to 7–9% across varying activity sets of 7, 12, and 16 classes and across different transmitter–receiver distances, antenna heights, and LOS-NLOS environments. This shows that our proposed framework achieves a more favorable privacy–utility trade-off, maintaining robust recognition performance even under stringent differential privacy constraints.

We further compared our proposed framework against the baseline models mentioned in Table 9 on the CSLOS Env. 1 dataset. Table 10 presents a comparison of DP-based HAR methods in terms of accuracy degradation. The accuracy drop indicates the relative performance loss compared to the corresponding non-DP baseline models. Our proposed method achieves an accuracy drop of 5.3% at $(\varepsilon = 12, \delta = 10^{-5})$, which is lower than the degradation in prior studies, such as 9% in ref. 83 6.5% in ref. 43 and 10% in ref. 85. This suggests that the proposed framework is capable of maintaining better utility under privacy constraints.

Table 10 Comparison of HAR methods under DP conditions in terms of accuracy degradation on the CSLOS Env. 1 dataset – accuracy drop indicates performance loss relative to the corresponding non-DP baselines

Method	Accuracy drop (%)	(ε, δ)
Debaditya *et al.*[83]	9	$(21, 10^{-5})$
Bigelli *et al.*[43]	6.5	$(17, 10^{-5})$
Nadira *et al.*[85]	10	$(15, 10^{-5})$
This work	**5.3**	$(12, 10^{-5})$

6 Discussion

In this work, we propose a novel light-weight privacy-preserving CSI-based HAR framework by jointly addressing classification accuracy, and robustness across heterogeneous sensing conditions. Across all three real datasets and different experimental configurations, incorporating differential privacy introduces a predictable yet controlled reduction in classification accuracy. This can be considered as the cost of incorporating privacy into the model training pipeline. The observed performance degradation, typically ranging between 7 and 9%, aligns with the theoretical expectations of DP-SGD due to gradient clipping and noise injection.[63] Despite this, the proposed CNN–temporal attention model maintains excellent accuracy and F1-scores even under strong privacy (*e.g.*, $\varepsilon \leq 12$), indicating that CSI representations retain sufficient discriminative structure under DP-induced noise. The proposed framework demonstrates strong robustness across diverse sensing conditions, including variations in transmitter–receiver distance, antenna height, and indoor environments. Distance-based evaluations on the WiAR dataset reveal that mid- and long-range sensing scenarios preserve discriminative motion signatures, even under DP constraints. This observation aligns with multipath propagation characteristics, where stable far-field components continue to encode meaningful activity-induced perturbations.

Similarly, antenna height experiments confirm that the proposed model effectively captures temporal motion patterns across different Fresnel zones. Although DP training introduces slightly higher performance degradation in cluttered or dynamically changing environments (*e.g.*, CSLOS Environment 2), the overall accuracy drop remains reasonable. These results indicate that the proposed architecture generalizes well under realistic deployment variability – a critical requirement for practical WiFi sensing systems.

Our ablation study highlights the critical role of temporal attention in modeling CSI dynamics. Removing the attention mechanism leads to a notable decline in accuracy, confirming the importance of adaptive temporal weighting in emphasizing motion-salient CSI segments.[72] Unlike recurrent architectures, such as LSTM or GRU that impose higher computational and memory overhead, the proposed attention module enables effective long-range temporal modeling with minimal complexity. This design choice proves particularly advantageous under DP training, where deeper or recurrent architectures are more sensitive to gradient noise.[86] The results highlight that temporal attention combined with convolutional feature extraction is sufficient for capturing discriminative CSI patterns, supporting efficient and privacy-aware model design.

The training sample size analysis provides quantitative insights into the privacy–utility trade-off inherent in differential privacy. As expected, stronger privacy guarantees (that is, lower ε) lead to reduced recognition accuracy, while increasing the number of training samples simultaneously improves

performance and reduces the effective privacy budget through privacy amplification.[60] This dual benefit suggests that scalable CSI data collection can effectively mitigate DP-induced performance loss. While smaller ε values provide stronger theoretical privacy guarantees, in practice, moderate ε (*e.g.*, $\varepsilon \leq 12$) offers a more realistic balance between privacy and utility for deep learning–based CSI HAR systems.[87,88] In safety-critical healthcare scenarios, the privacy–utility trade-off must be carefully considered, as even modest accuracy degradation (*e.g.*, 7–9%) may impact reliability and necessitate application-specific tuning or complementary validation mechanisms. While differential privacy provides formal guarantees, evaluating robustness against empirical attacks, such as membership inference remains an important direction for future work.[23] From a system design perspective, the proposed framework enables flexible privacy configurations, allowing practitioners to select operating points that balance recognition accuracy and privacy protection according to application requirements. Although the scope of this work does not include a dedicated hardware implementation, the proposed lightweight and infrastructure-free framework is inherently scalable and can be readily deployed across large-scale real-world and industrial environments using existing WiFi systems.[89,90] Additionally, a limitation of the work is that the evaluation was conducted on public datasets, where activities follow predefined scripts. The performance on "in-the-wild" datasets containing natural and unconstrained human activities is yet to be evaluated.

Future work can explore adaptive or layer-wise DP mechanisms, as well as multi-user real-time HAR.[67,91,92] Federated or self-supervised learning and validating the framework in large-scale, real-world deployments with heterogeneous devices can be a promising research direction.[93–95] Emerging approaches such as domain adaptation,[96] multimodal sensing,[97] and a self-supervised learning framework, along with SOTA architectures, can be promising directions for improving robustness and accuracy. While phase information can provide additional spatial cues, its instability and limited availability across datasets motivate our use of amplitude-only features in this work; incorporating calibrated phase information remains an important direction for future work.[2] Although evaluated across multiple environments, explicit cross-environment generalization (training and testing across disjoint environments) remains an important direction for future work to assess robustness under distributional shifts.

7 Conclusion

In this work, we present a lightweight and privacy-preserving CSI-based HAR framework that integrates a CNN–temporal attention architecture with differential privacy–aware training. The proposed architecture effectively models temporal CSI dynamics without relying on computationally expensive recurrent networks, making it suitable for deployment in resource-constrained and privacy-sensitive environments. Extensive evaluations on multiple public datasets and under heterogeneous sensing conditions demonstrate that excellent recognition performance can be achieved, while providing strict (ε, δ)-DP guarantees. Our investigation indicates that DP training incurs only a modest performance degradation (7–8%) while significantly strengthening protection against inference attacks. Furthermore, robustness across distance, antenna height, and environmental variations highlights the practicality of the framework for real-world WiFi sensing applications. Although we employed the standard DP-SGD, designing a CSI-aware privacy mechanism that exploits the structure of OFDM subcarriers can be an important direction for improving the privacy–utility trade-off.[98] Additionally, we report detailed computational efficiency metrics, and evaluating the model on physical edge devices (*e.g.*, Raspberry Pi or Nvidia Jetson platforms) remains an important direction for future work. Overall, our work provides empirical evidence that accurate, efficient, and privacy-aware CSI-based HAR is achievable when privacy is incorporated as a first-class design constraint.

Author contributions

Khondakar Ashik Shahriar – data curation, investigation, methodology, software, validation, visualization, writing – original draft. Maruf Ahmed – resources, supervision, validation, visualization, writing – review and editing. Hafiz Imtiaz – investigation, resources, supervision, validation, visualization, writing – review and editing.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data availability

All three datasets used in this work are publicly accessible. The relevant papers for the datasets can be found in WiAR,[28] CSLOS[29] and CSI-HAR.[27] The implementation code for the proposed CNN–temporal attention network and the scripts used for training and evaluation are publicly available at https://doi.org/10.5281/zenodo.18963950. The download link for datasets can be accessed from the mentioned repository.

Acknowledgements

The authors would like to express their sincere gratitude towards the Department of Electrical and Electronic Engineering (EEE) of Bangladesh University of Engineering and Technology (BUET) for providing support for research.

Notes and references

1 G. Saleem, U. I. Bajwa and R. H. Raza, *Neural Comput. Appl.*, 2023, **35**, 4145–4182.
2 N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi and J. S. Suri, *Artif. Intell. Rev.*, 2022, **55**, 4755–4808.
3 F. Serpush, M. B. Menhaj, B. Masoumi and B. Karasfi, *Comput. Intell. Neurosci.*, 2022, **2022**, 1391906.
4 R. Madhavan, *Macromol. Mater. Eng.*, 2022, **307**, 2200034.
5 R. Madhavan, *Mater. Adv.*, 2022, **3**, 8665–8676.

6 A. Lentzas and D. Vrakas, *Artif. Intell. Rev.*, 2020, **53**, 1975–2021.
7 R. Madhavan, *New J. Chem.*, 2022, **46**, 17596–17609.
8 F. Yu, C. Yu, Z. Tian, X. Liu, J. Cao, L. Liu, C. Du and M. Jiang, *IEEE Internet Things J.*, 2024, **11**, 26314–26328.
9 S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang and Z. Li, *J. Healthc. Eng.*, 2017, **2017**, 3090343.
10 K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu and Y. Liu, *ACM Comput. Surv.*, 2021, **54**, 1–40.
11 L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee and H. Moon, *Pattern Recognit.*, 2020, **108**, 107561.
12 R. Madhavan, *Innov. Discov.*, 2025, **2**, 4.
13 I. Ahmad, A. Ullah and W. Choi, *IEEE Open J. Commun. Soc.*, 2024, **5**, 3595–3623.
14 Y. Zhang, F. He, Y. Wang, D. Wu and G. Yu, *Neural Comput. Appl.*, 2023, **35**, 12415–12432.
15 Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang and J. Liu, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022, **45**, 3200–3225.
16 Y. Yang, P. Hu, J. Shen, H. Cheng, Z. An and X. Liu, *High-Confid. Comput.*, 2024, **4**, 100204.
17 M. Karim, S. Khalid, A. Aleryani, J. Khan, I. Ullah and Z. Ali, *IEEE Access*, 2024, **12**, 36372–36390.
18 A. Golda, K. Mekonen, A. Pandey, A. Singh, V. Hassija, V. Chamola and B. Sikdar, *IEEE Access*, 2024, **12**, 48126–48144.
19 H. Zhang, C. Song, A. Wang, C. Xu, D. Li and W. Xu, *The 25th annual international conference on mobile computing and networking*, 2019, pp. 1–16.
20 C. Sivakumar, V. Mone and R. Abdumukhtor, *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.*, 2024, **14**, e1535.
21 Y. Wang, T. Sun, S. Li, X. Yuan, W. Ni, E. Hossain and H. V. Poor, *IEEE Commun. Surv. Tutor.*, 2023, **25**, 2245–2298.
22 W. Yang, S. Wang, D. Wu, T. Cai, Y. Zhu, S. Wei, Y. Zhang, X. Yang, Z. Tang and Y. Li, *Artif. Intell. Rev.*, 2025, **58**, 242.
23 H. Hu, Z. Salcic, L. Sun, G. Dobbie, P. S. Yu and X. Zhang, *ACM Comput. Surv.*, 2022, **54**, 1–37.
24 H. Wei, H. Tang, X. Jia, Z. Wang, H. Yu, Z. Li, S. Satoh, L. Van Gool and Z. Wang, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024, **46**, 9797–9817.
25 J. C. Costa, T. Roxo, H. Proença and P. R. M. Inacio, *IEEE Access*, 2024, **12**, 61113–61136.
26 C. Dwork and A. Roth, *Found. Trends Theor. Comput. Sci.*, 2014, **9**, 211–487.
27 P. F. Moshiri, R. Shahbazian, M. Nabati and S. A. Ghorashi, *Sensors*, 2021, **21**, 7225.
28 L. Guo, L. Wang, C. Lin, J. Liu, B. Lu, J. Fang, Z. Liu, Z. Shan, J. Yang and S. Guo, *IEEE Access*, 2019, **7**, 154935–154945.
29 A. Baha'A, M. M. Almazari, R. Alazrai and M. I. Daoud, *Data Brief*, 2020, **33**, 106534.
30 Y. Wang, Q. Wang, L. Zhao and C. Wang, *Future Gener. Comput. Syst.*, 2023, **148**, 408–424.
31 A. Ray, M. H. Kolekar, R. Balasubramanian and A. Hafiane, *Int. J. Inf. Manag. Data Insights*, 2023, **3**, 100142.
32 F. Shafizadegan, A. R. Naghsh-Nilchi and E. Shabaninia, *Artif. Intell. Rev.*, 2024, **57**, 178.
33 X. Lin, M. Liu and H. Chen, *Front. Comput. Neurosci.*, 2024, **18**, 1508297.
34 C. Liu, X. Qi, E. Y. Lam and N. Wong, *IEEE Access*, 2022, **10**, 55638–55649.
35 S. Latif, H. Cuayáhuitl, F. Pervez, F. Shamshad, H. S. Ali and E. Cambria, *Artif. Intell. Rev.*, 2023, **56**, 2193–2240.
36 J. Kim, K. Min, M. Jung and S. Chi, *Build. Environ.*, 2020, **181**, 107092.
37 M. B. Shaikh, D. Chai, S. M. S. Islam and N. Akhtar, *Neural Comput. Appl.*, 2024, **36**, 5499–5513.
38 T. Jesus, J. Duarte, D. Ferreira, D. Durães, F. Marcondes, F. Santos, M. Gomes, P. Novais, F. Gonçalves and J. Fonseca *et al.*, *International conference on intelligent data engineering and automated learning*, 2020, pp. 549–560.
39 S. Herath, M. Harandi and F. Porikli, *Image Vis. Comput.*, 2017, **60**, 4–21.
40 J. Wang, Y. Chen, S. Hao, X. Peng and L. Hu, *Pattern Recognit. Lett.*, 2019, **119**, 3–11.
41 S. Xu, L. Zhang, Y. Tang, C. Han, H. Wu and A. Song, *IEEE Trans. Knowl. Data Eng*, 2023, **35**, 12497–12512.
42 A. Arafa, H. Harfoush, N. El-Fishawy and M. Radad, *Pervasive and Mobile Computing*, 2026, p. 102161.
43 L. Bigelli, C. Contoli, V. Freschi and E. Lattanzi, *IoT*, 2024, **26**, 101189.
44 H. Haresamudram, C. I. Tang, S. Suh, P. Lukowicz and T. Ploetz, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2025, **9**, 1–44.
45 N. Damodaran, E. Haruni, M. Kokhkharova and J. Schäfer, *CCF Trans. Pervasive Comput. Inter*, 2020, **2**, 1–17.
46 X. Chen, Y. Zou, C. Li and W. Xiao, *IEEE Trans. Hum.-Mach. Syst.*, 2024, **54**, 68–78.
47 J. Ding and Y. Wang, *IEEE Access*, 2019, **7**, 174257–174269.
48 D. Wang, J. Yang, W. Cui, L. Xie and S. Sun, *IEEE Internet Things J.*, 2021, **8**, 17345–17355.
49 S. K. Yadav, S. Sai, A. Gundewar, H. Rathore, K. Tiwari, H. M. Pandey and M. Mathur, *Neural Netw.*, 2022, **146**, 11–21.
50 F. Fuschini, H. El-Sallabi, V. Degli-Esposti, L. Vuokko, D. Guiducci and P. Vainikainen, *IEEE Trans. Antennas Propag.*, 2008, **56**, 848–857.
51 C. Xiao, Y. R. Zheng and N. C. Beaulieu, *IEEE International Conference on Communications, 2003*. ICC'03., 2003, pp. 3524–3529.
52 G. D. Durgin and T. S. Rappaport, *IEEE Trans. Antennas Propag.*, 2000, **48**, 682–693.
53 Y. G. Li and G. L. Stuber, *Orthogonal frequency division multiplexing for wireless communications*, Springer Science & Business Media, 2006.
54 F. B. Frederiksen and R. Prasad, *Proceedings RAWCON 2002. 2002 IEEE radio and wireless conference (Cat. No. 02EX573)*, 2002, pp. 19–22.
55 Y. Zhong, J. Wang, S. Wu, T. Jiang, Y. Huang and Q. Wu, *IEEE Internet Things J.*, 2020, **8**, 15148–15159.
56 G. Caire and S. Shamai, *IEEE Trans. Inf. Theory*, 2002, **45**, 2007–2019.
57 V. V. Ratnam, H. Chen, H.-H. Chang, A. Sehgal and J. Zhang, *IEEE Trans. Wirel. Commun.*, 2024, **23**, 10820–10833.

58 Y. Ma, G. Zhou and S. Wang, *ACM Comput. Surv.*, 2019, **52**, 1–36.
59 J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten and V. Shmatikov, *2011 IEEE symposium on security and privacy*, 2011, pp. 231–246.
60 C. Dwork, *International conference on theory and applications of models of computation*, 2008, pp. 1–19.
61 Z. Ji, Z. C. Lipton and C. Elkan, *arXiv preprint arXiv:1412.7584*, 2014.
62 B. Balle and Y.-X. Wang, *International conference on machine learning*, 2018, pp. 394–403.
63 M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar and L. Zhang, *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.
64 L. C. Bezerra, N. Kouzayha, H. Elsawy, A. Bader and T. Y. Al-Naffouri, *IEEE Open J. Commun. Soc.*, 2023, **5**, 97–111.
65 S. Palipana, D. Rojas, P. Agrawal and D. Pesch, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2018, **1**, 1–25.
66 T. Fan, S. Qiu, W. Gong and Y. Fang, *IEEE Trans. Mob. Comput.*, 2025, **24**(10), 11034–111045.
67 Z. Shi, Q. Cheng, J. A. Zhang and R. Y. Da Xu, *IEEE Internet Things J.*, 2022, **9**, 24643–24654.
68 B. K. Iwana and S. Uchida, *PLoS One*, 2021, **16**, e0254841.
69 L. Danay, R. Ramon-Gonen, M. Gorodetski and D. G. Schwartz, *Int. J. Med. Inf.*, 2024, **191**, 105565.
70 M. Taib and G. G. Messier, *IEEE Access*, 2024, **12**, 158647–158656.
71 N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *J. Mach. Learn. Res.*, 2014, **15**, 1929–1958.
72 Z. Chen, L. Zhang, C. Jiang, Z. Cao and W. Cui, *IEEE Trans. Mob. Comput.*, 2018, **18**, 2714–2724.
73 V. Casola, A. De Benedictis, S. Di Martino, N. Mazzocca and L. L. L. Starace, *IEEE Internet Things J.*, 2020, **8**, 12724–12733.
74 I. Loshchilov and F. Hutter, *arXiv preprint arXiv:1608.03983*, 2016.
75 A. Yousefpour, I. Shilov, A. Sablayrolles, D. Testuggine, K. Prasad, M. Malek, J. Nguyen, S. Ghosh, A. Bharadwaj and J. Zhao *et al.*, *arXiv preprint arXiv:2109.12298*, 2021.
76 A. Koloskova, R. McKenna, Z. Charles, J. Rush and H. B. McMahan, *NeurIPS*, 2023, **36**, 35761–35773.
77 A. Kurakin, S. Song, S. Chien, R. Geambasu, A. Terzis and A. Thakurta, *arXiv preprint arXiv:2201.12328*, 2022.
78 W. Meng, Z. Liu, B. Li, W. Cui, J. T. Zhou and L. Zhang, *IEEE Trans. Wirel. Commun.*, 2023, **23**, 2755–2770.
79 H. Huang, L. Lin, L. Zhao, H. Huang and S. Ding, *IEEE Trans. Cogn. Commun. Netw.*, 2024, **10**, 2088–2101.
80 I. A. Showmik, T. F. Sanam and H. Imtiaz, *Digit. Signal Process.*, 2023, **138**, 104056.
81 A. Elkelany, R. Ross and S. Mckeever, *Irish Conference on Artificial Intelligence and Cognitive Science*, 2022, pp. 121–133.
82 J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah and J. Cheng, *IEEE Internet Things J.*, 2020, **8**, 4628–4641.
83 D. Roy and Š. Girdzijauskas, *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)*, 2023, pp. 1–10.
84 F. J. Piran, Z. Chen, Y. Zhang, Q. Zhou, J. Tang and F. Imani, *arXiv preprint arXiv:2509.10691*, 2025.
85 N. Pervin, T. F. Sanam and H. Imtiaz, *Signal Image Video Process.*, 2024, **18**, 9141–9155.
86 T. Ha, T. K. Dang, T. T. Dang, T. A. Truong and M. T. Nguyen, *2019 International Conference on Advanced Computing and Applications (ACOMP)*, 2019, pp. 97–102.
87 I. Mironov, *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 2017, pp. 263–275.
88 Differential Privacy Team, *Apple Machine Learning Research*, 2017.
89 R. Madhavan, *New J. Chem.*, 2025, **49**, 1700–1717.
90 R. Madhavan, *J. Mod. Nanotechnol.*, 2024, **4**, 5.
91 D. Winograd-Cort, A. Haeberlen, A. Roth and B. C. Pierce, *Proc. ACM Program. Lang.*, 2017, **1**, 1–29.
92 F. Abuhoureyah, K. S. Sim and Y. C. Wong, *IEEE Access*, 2024, **12**, 112008–112024.
93 S. Zhang, H. Jia, T. Jiang, S. Wu, X. Ding and Y. Zhong, *Measurement*, 2025, 117821.
94 O. Aouedi, A. Sacco, L. U. Khan, D. C. Nguyen and M. Guizani, *IEEE Open J. Commun. Soc.*, 2024, **5**, 7341–7367.
95 K. Xu, J. Wang, H. Zhu and D. Zheng, *arXiv preprint arXiv:2308.02412*, 2023.
96 A. Pandey, M. Zeeshan, J. Torres-Sospedra, A. Kumar and S. Kumar, *IEEE J. Radio Freq. Identif.*, 2026, **10**, 35–46.
97 I. K. Ihianle, A. O. Nwajana, S. H. Ebenuwa, R. I. Otuka, K. Owa and M. O. Orisatoki, *IEEE Access*, 2020, **8**, 179028–179038.
98 Z. He, Y. Yin, M. Bouazizi, T. Ohtsuki, D. Niyato and G. Gui, *IEEE Trans. Cogn. Commun. Netw.*, 2026, **12**, 7186–7200.