Perspective

Check for updates

Advancing RF sensing with generative AI: from synthetic data generation to pose completion and beyond

Ziqi Wang, Shiwen Mao

Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA.

Correspondence to: Dr. Shiwen Mao, Department of Electrical and Computer Engineering, Auburn University, 200 Broun Hall, Auburn, AL 36849-5201, USA. E-mail: smao@ieee.org; ORCID: https://orcid.org/0000-0002-7052-0007

Abstract

Radio Frequency (RF) sensing has emerged as a pivotal technology for non-intrusive human perception in various applications. However, the challenge of collecting extensive labeled RF data hampers the scalability and effectiveness of machine learning models in this domain. Our prior work introduced innovative generative AI frameworks - RF-Artificial Intelligence Generated Content using conditional generative adversarial networks and RF-Activity Class Conditional Latent Diffusion Model employing latent diffusion models - to synthesize high-quality RF sensing data across multiple platforms. Building upon this foundation, we explore future directions that leverage generative AI for enhanced 3D human pose estimation and beyond. Specifically, we discuss our recent advances in pose completion using latent diffusion transformers and propose additional research avenues: cross-modal generative models for RF sensing, real-time adaptive generative AI incorporating evolutionary learning for dynamic environments, and addressing security and privacy concerns in intelligent cyber-physical systems. These directions aim to further exploit the capabilities of generative AI to overcome challenges in RF sensing, paving the way for more robust, scalable, and secure applications.

Keywords: Generative AI, 3D human pose estimation and completion, latent diffusion models, multi-modal conditioning, RF sensing

1. INTRODUCTION

The proliferation of wireless communication technologies has ushered in an era where Radio Frequency (RF) signals are not only mediums for data transmission but also carry important information for sensing and perception. RF sensing enables device-free, privacy-preserving monitoring of human activities, offering significant advantages over traditional vision-based or wearable systems. Recently, Deep learning has significantly advanced RF sensing by enabling complex pattern recognition and interpretation of RF signals for applications such as vital sign monitoring, human activity recognition (HAR) and 3D human pose estimation (HPE)[1]. Despite its high potential, RF sensing faces a significant hurdle: the collection of large-scale, high-quality labeled datasets is both time-consuming and costly. RF data is highly sensitive to environmental changes, device configurations, and temporal dynamics, making it challenging to generalize models across different settings. Moreover, the randomness and complexity inherent in RF signals complicate the data acquisition process.

Generative AI offers a promising solution to this data scarcity problem[2,3]. By synthesizing realistic RF data, we can augment limited datasets, enhance model robustness, and reduce the dependency on extensive data collection efforts. In our previous work, we harnessed the capabilities of generative adversarial networks (GANs)[4] and diffusion models[5] to address these challenges. The diagram of generative AI for RF sensing is illustrated in Figure 1. In this perspective paper, we summarize our contributions in this emerging field and propose future research directions that build upon our foundational work. Specifically, we discuss our latest findings in 3D human pose completion using latent diffusion transformers (LDTs) and outline two promising avenues: cross-modal generative AI for enhanced RF sensing and real-time adaptive generative AI for dynamic environments. These directions aim to further exploit the potential of generative AI in addressing the challenges of RF sensing, ultimately enhancing the performance and practicality of RF-based applications.

2. ADVANCEMENTS IN GENERATIVE AI FOR RF SENSING

2.1. Related works

Traditional data augmentation techniques, such as geometric transformations and time-series interpolation, have been applied to RF data[6], but these methods often fail to generalize across unseen environments. GANs, Variational Autoencoders (VAEs), and Diffusion Models have also been explored to generate synthetic RF data for wireless sensing. For example, a prior study has applied GANs to RF sensing, primarily focusing on WiFi Channel State Information (CSI) augmentation for improving HAR performance[7]. Such generated synthetic data lacks the complexity and variability required for practical RF sensing applications. Ultimately, existing methods fall short in generalization, limiting their ability to synthesize diverse and adaptable RF data that accurately represents a wide range of human activities. Diffusion models have been increasingly adopted due to their ability to generate high-fidelity, diverse synthetic signals. As time-series signals with multidimensional features, RF signals present abundant opportunities for sensing applications. Ref.[8] provides a comprehensive survey that discusses the utilization of generative AI to augment wireless sensing.

2.2. RF-AIGC with conditional GANs

In our first study, we introduced an Artificial Intelligence Generated Content (AIGC) framework termed RF-AIGC, utilizing a conditional Recurrent Generative Adversarial Network (RF-CRGAN). This model used an autoencoder-based GAN network as the generator to synthesize labeled RF data from specific human poses across multiple wireless sensing platforms, including WiFi, radio-frequency identification (RFID), and Frequency-Modulated Continuous Wave (FMCW) radar.

To enhance the realism and diversity of synthesized RF data, we employ a two-stage generator fine-tuning and adversarial learning process. After an initial pretraining phase, where the generator learns from real RF-pose data, we fine-tune it using adversarial training with weakly supervised learning to improve generalization across different RF modalities and activity types. Unlike conventional GANs, which often suffer from
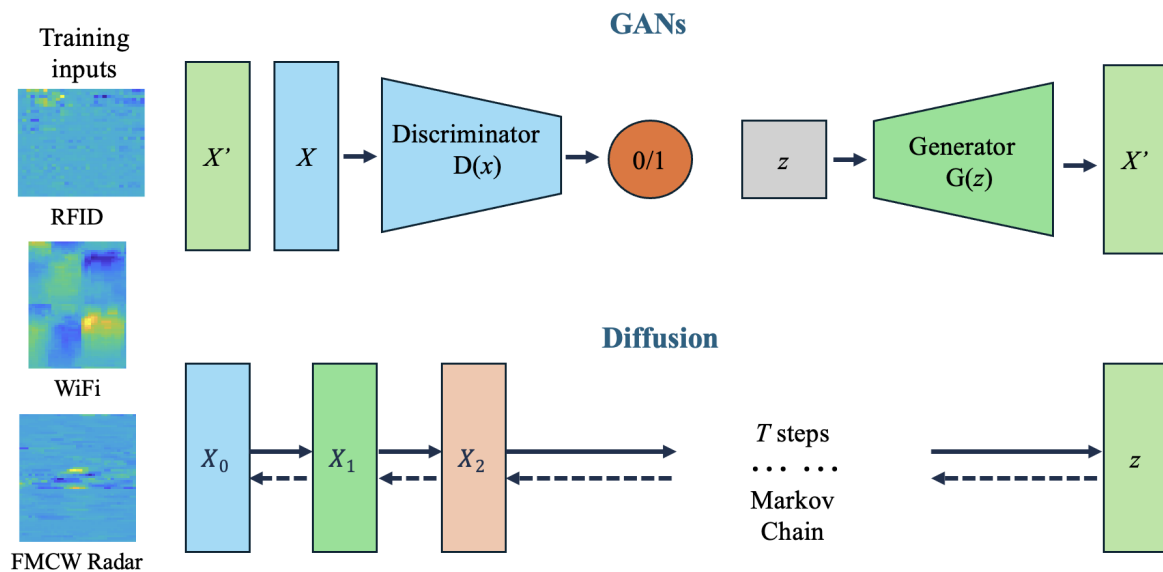
Figure 1. The diagram of our proposed "Generative AI for RF Sensing" framework: After appropriate signal processing for different RF sensory data, GANs or diffusion models can start training and learning the diverse data distributions to generate high-quality synthetic data. The goal is to automatically and interactively (allowing users to provide prompts such as class labels and texts) generate, complete, enhance, edit or repair RF sensing data with fidelity and diversity.

mode collapse, our approach ensures the generator learns to produce diverse and temporally coherent RF signals. This is achieved through RF feature perturbations, where the generator synthesizes data conditioned on pose skeletons and global motion variations, helping it learn fine-grained motion dynamics. By augmenting limited training datasets with synthesized RF data, we demonstrated that models could achieve comparable performance to those trained on extensive real-world data. The ability to synthesize RF data from vision data (more easily acquired and modified) for specified activities allows for targeted augmentation, addressing class imbalances, and improving model robustness.

2.3. RF-ACCLDM with latent diffusion models

Building upon the success of RF-CRGAN, our second study proposed the RF-ACCLDM (Activity Class Conditional Latent Diffusion Model). This framework leverages latent diffusion models to directly generate high-fidelity synthetic RF data based on the user prompt of a body shape and an activity label (e.g., drinking, boxing, and walking). Operating in latent domains, RF-ACCLDM supports various RF technologies and modalities, including RFID, WiFi CSI, and FMCW radar. The use of latent spaces reduces computational complexity and enables the model to capture essential features across different RF modalities. By compressing high-dimensional RF data into latent representations, the diffusion model operates more efficiently, allowing for the synthesis of high-quality data at a fraction of the computational cost of its counterparts. The conditional diffusion process ensures that the generated data aligns with specified activity classes, providing precise control over the data generation process.

Our experiments showed that RF-ACCLDM outperforms traditional diffusion models on raw RF data in terms of quality, computational efficiency, and scalability. Quantitatively, we use Fréchet Inception Distance (FID) scores to measure the realism of the generated RFID data using our various generative models. GAN-based RF-CRGAN had the highest FID (48.89), indicating a relatively large domain gap in realism compared to real RF data. Diffusion-based RF-ACCDM improved fidelity by 47.5% (FID = 25.64), showing better feature alignment with real RF data. Latent Diffusion-based RF-ACCLDM achieved the lowest FID (10.45), a 78.6% improvement over GANs, nearly approaching real RF data quality (FID = 6.22). Latent diffusion produces the most realistic RF data, demonstrating higher fidelity and better structural coherence compared to other

methods, while taking significantly less time to train compared to standard diffusion models (with an over 40% improvement).

The quality improvements are also evident in downstream tasks such as HAR classification. RF-CRGAN reached 91.2% F1-score only when augmented with an adequate amount of real data, but struggled with domain shifts when used as is. However, diffusion-based RF-ACCDM improves performance to 92.1% (+ 1.0%) without the need of mixing real data. Latent Diffusion RF-ACCLDM achieved the highest F1-score at 93.0% (+2.0%), demonstrating superior activity classification, and offering better cross-domain adaptability. For complicated regression tasks such as 3D HPE, synthetic models trained with only RF-CRGAN-generated data achieved a median error of 6.08 cm, substantially larger than 4.89 cm obtained by standard diffusion models, and 4.23 cm obtained by latent diffusion models. This is further supported by the fact that RF-CRGAN estimates unnatural and discontinuous poses. Standard diffusion models perform better, but struggle with outliers and noise. Latent diffusion models achieve natural and temporal poses with smoothness and rare outliers.

2.4. Pose completion with latent diffusion transformers

While our previous models addressed data scarcity, another challenge in RF-based 3D HPE is the incomplete capture of skeletal joints due to sensing constraints. In our current research, we introduce a novel framework that leverages LDTs with cross-attention conditioning to infer missing joints in skeletal poses[9]. By generating high-quality, diverse RFID sensing data and training a transformer-based kinematics predictor termed RF-Former, we can estimate 3D poses with temporal smoothness from RFID data. Our model then completes full 25-joint configurations from these partial 12-joint inputs, marking the first method to detect over 20 distinct skeletal joints using generative AI technologies in wireless sensing-based continuous 3D HPE tasks. This advancement is particularly significant for RFID-based systems, which typically capture limited joint information. Our approach extends the applicability of wireless-based pose estimation to scenarios where collecting extensive paired datasets is impractical, such as pedestrian and health monitoring in occluded environments. The architecture of the LDT system is shown in Figure 2.

To assess the realism and utility of 3D human poses estimated from LDT-generated RFID data in depth, we evaluate the results using key structural and temporal metrics, as summarized in Table 1. These evaluations are conducted across diverse subjects and multiple rounds of synthetic RFID data, ensuring robust analysis. The estimated poses demonstrate strong alignment with anatomical and temporal expectations. Specifically, the average joint position error is 8.99 cm, which, while slightly higher than real data, reflects realistic limb placements given the generative setting. The joint angle error of 6.91° further confirms the structural plausibility of the generated poses. Importantly, the temporal smoothness score of 1.51 cm/frame closely approximates that of real motion data (1.40 cm/frame), demonstrating that our generated sequences preserve continuous and fluid motion trajectories. In terms of generative quality, the FID score of 1.42 reflects high similarity to real data (ground truth FID = 0.73), and our diversity score of 10.98 slightly exceeds the real data baseline (10.35), indicating that the generated samples span a wide range of pose configurations without sacrificing coherence. These outcomes indicate that LDT-generated RFID sequences not only resemble real-world dynamics but also generalize well to different skeletons, as our model was tested on subject configurations not seen during RF-Former training. Altogether, the results support the conclusion that our LDT-based system is capable of producing high-quality synthetic RF data that can be reliably transformed into 3D poses. These poses maintain anatomical realism and temporal fluidity, enabling their effective use in data augmentation, human motion analysis, and AIGC-empowered wireless sensing applications.

Furthermore, to fairly evaluate the effectiveness of our pose completion framework, we generate the same number of samples as the ground truth data and report average scores under two scenarios: seen (partial poses seen during training) and unseen (new partial poses never observed during training). The results are presented in Table 2.
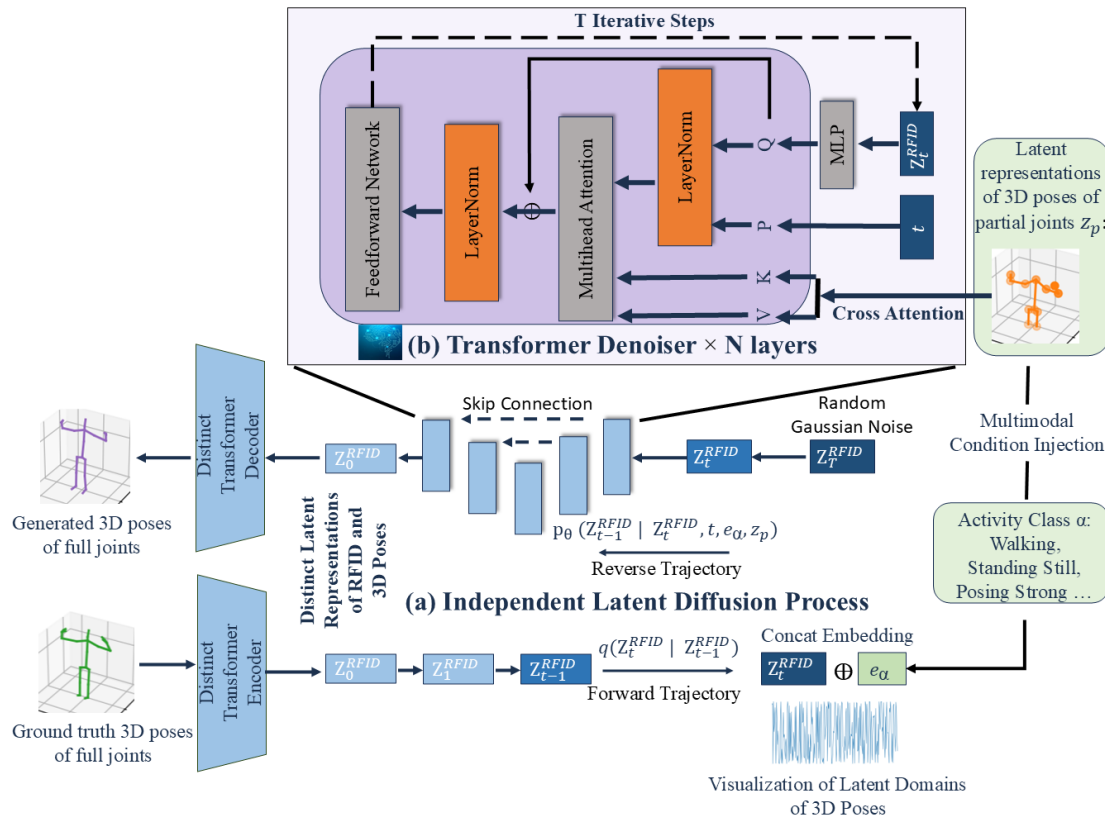
Figure 2. The architecture of the RFID-based 3D human pose completion system empowered by latent diffusion transformers.

Table 1. Evaluation metrics for estimated 3D human poses using LDT generated RFID data

Metrics	LDT	Metrics	LDT
Average joint error (cm)	8.99	FID	1.42
Joint angle error (°)	6.91	Diversity	10.98
Smoothness (cm/frame)	1.51	GT Diversity	10.35
GT Smoothness (cm/frame)	1.40		

In the seen scenario, our model achieves strong structural and temporal performance: an average joint error of 11.74 cm, bone length consistency of 1.77 cm, and a low joint angle error of 6.65°, indicating anatomically plausible completions. The temporal smoothness score of 2.46 cm/frame closely aligns with that of real motion sequences, preserving natural transitions across frames. Furthermore, the generated poses show high fidelity and coverage with a FID of 0.87 and diversity score of 26.59, comparable to real data baselines (FID = 0.15, diversity = 27.12), confirming both realism and motion variety. In the more challenging unseen scenario, the model generalizes reasonably well despite no exposure to the partial pose patterns during training. The joint error increases to 19.23 cm and the angle error to 11.13°, which remains acceptable given the added difficulty of inferring 13 missing joints out of 25. Notably, the smoothness score drops slightly to 1.90 cm/frame, suggesting that certain joints exhibit reduced motion. This is expected, as some generated motions may become overly conservative without matching examples. Given that limb lengths in human skeletons average around 35 cm, a 10 cm deviation corresponds to roughly one-third of an arm or one-quarter of a leg. The reported errors are thus reasonable - especially considering that our model completes full-body poses (25 joints) from partial inputs (12 joints) in motion, not static form.

Finally, we conduct detailed comparisons between our framework and two self-supervised learning (SSL) meth-

Table 2. Evaluation metrics for 3D pose completion with ground truth and unseen partial pose conditioning

Metrics	Ground truth	Unseen
Avg joint error (cm)	11.74	19.23
Bone consistency (cm)	1.77	2.12
Joint angle error (°)	6.65	11.13
Smoothness (cm/frame)	2.46	1.90
FID (-)	0.87	4.67
Diversity (-)	26.59	13.71

ods for pose completion. The autoencoder-based model, despite its temporal-aware architecture, suffers from severe overfitting, leading to poor generalization beyond the training dataset. While it achieves reasonable joint position accuracy, it struggles with high trajectory errors and structural inconsistencies, particularly in unseen cases. It produces unnatural joint twisting, with an average joint angle error of 16.0° and low temporal smoothness (0.3 cm/frame), indicating it prioritizes minimizing reconstruction loss over learning meaningful pose structures. The KNN-based approach, while offering better trajectory consistency, lacks flexibility in pose completion, relying on direct interpolation and stitching of existing training samples rather than generating novel poses. This limitation results in poor adaptability in unseen scenarios.

In contrast, the LDT model achieves state-of-the-art pose completion performance, significantly outperforming both Autoencoder and KNN-based approaches. In seen scenarios (cases where the model is tested on pose configurations that were present during training), LDT achieves an MPJPE of 11.7 cm, a bone length consistency error of 1.77 cm, and a joint angle error of just 6.6°, ensuring that its generated poses are both structurally sound and kinematically accurate. Unlike the autoencoder, which produces rigid and overfitted outputs, LDT maintains fluid motion transitions, reflected in its superior temporal smoothness of 1.51 cm/frame. Even in unseen scenarios, where generalization is critical, LDT maintains strong performance as LDT learns an underlying manifold of human motion, enabling it to generate diverse, anatomically realistic, and temporally coherent pose completions. Its superior generalization ability ensures smooth, accurate motion reconstruction, making it most effective for real-world tasks.

3. FUTURE DIRECTIONS

3.1. Cross-modal generative models for RF sensing

An exciting future direction is the development of cross-modal generative models that bridge RF sensing with other data modalities, such as different wireless technologies, along with audio and vision data. By conditioning RF data generation on these modalities, we can enhance the richness and context of synthetic RF datasets. For instance, using synchronized audio cues or small amounts of visual data, a generative model could produce RF signals that reflect complex human activities or environmental interactions. This cross-modal approach can be particularly beneficial in situations where visual data is limited or privacy concerns restrict the use of cameras. It also opens avenues for multi-sensor fusion, improving the robustness and accuracy of HAR systems.

A practical implementation of cross-modal generative AI could involve aligning WiFi CSI-based human activity data with video-based pose estimation models, where a joint embedding space maps RF signals to visual motion representations. For example, a diffusion-based generative model could condition RF data synthesis on concurrent depth images to improve the quality of RF-generated poses. A key challenge in this integration is data synchronization and alignment across heterogeneous modalities, especially when dealing with different sampling rates and noise characteristics. Another challenge is ensuring privacy compliance when fusing RF with vision-based sensing, where sensitive data must be securely processed while retaining performance benefits.

3.2. Real-time generative AI for dynamic environment adaptation

Another critical research avenue is the application of generative AI for real-time data augmentation in dynamic environments. RF sensing systems often suffer performance degradation due to changes in the environment, such as moving objects, variable crowd densities, body shapes of test subjects, or alterations in room layouts. By developing generative models capable of synthesizing RF data on-the-fly, we can continuously adapt and augment the training data to reflect the current environment. Techniques like online learning, adaptive diffusion models, and evolutionary algorithms could be explored to achieve this goal.

A practical approach could involve integrating evolutionary learning-based generative AI, where a model monitors variations in RF data streams and adjusts synthetic data generation accordingly. For example, a real-time latent diffusion model could refine its internal noise schedule based on environmental drift, ensuring that generated RF signals remain relevant even under shifting room layouts or varying subject appearances. However, real-time adaptation poses several implementation challenges. First, computational constraints must be addressed to ensure that generative AI operates efficiently on edge devices. Second, real-time models must be robust to transient anomalies, such as sudden occlusions or temporary RF interference, which could mislead adaptation mechanisms.

3.3. Ethical Considerations in Synthetic RF Data Generation

While generative AI significantly enhances RF sensing by addressing data scarcity, it introduces concerns regarding bias and privacy risks. If training data is skewed toward specific body types or demographics, synthetic RF signals may fail to generalize across diverse populations. For instance, HAR models trained predominantly on young adults may struggle to recognize movement patterns of elderly individuals, leading to disparities in real-world applications. Additionally, synthetic data generation must ensure that sensitive information is not inadvertently leaked or reconstructed, as models may inadvertently encode latent representations of original training data. For instance, a diffusion model trained on RF gait signals from a hospital setting could regenerate identifiable gait patterns extracted from detected activities, posing deanonymization risks. Moreover, adversaries could exploit synthesized RF signals to mimic an individual's movement profile, potentially compromising RF-based authentication. To mitigate these risks, ensuring diverse and representative training datasets is essential for reducing bias, while differential privacy techniques can be applied to prevent models from memorizing and reconstructing specific movement signatures. Furthermore, adversarial detection mechanisms capable of distinguishing real from synthetic RF signals can be integrated to prevent unauthorized misuse. Secure multi-party computation and federated learning can enable collaborative model training without sharing raw data, enhancing privacy in distributed systems[10].

4. CONCLUSIONS

Generative AI holds immense potential for overcoming the challenges associated with RF sensing data scarcity and variability. Our work demonstrated that by employing advanced generative models such as conditional GANs and LDTs, we can synthesize high-quality RF data, enhance model performance, and expand the capabilities of wireless sensing systems. Future research in cross-modal generative models and real-time data augmentation promises to further advance the field, enabling more robust, versatile, and adaptive RF sensing applications. As we continue to explore these directions, significant contributions are expected to areas such as HAR, healthcare monitoring, autonomous driving, and beyond.

DECLARATIONS

Authors' contributions

Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Wang, Z.; Mao, S.

REFERENCES

1. Li, C.; Cao, Z.; Liu, Y. Deep AI enabled ubiquitous wireless sensing: a survey. *ACM. Comput. Surv.* **2021**, *54*, 1-35. DOI
2. Xu, M.; Du, H.; Niyato, D.; et al. Unleashing the power of edge-cloud generative AI in mobile networks: a survey of AIGC services. *IEEE. Commun. Surveys. Tuts.* **2024**, *26*, 1127-70. DOI
3. Du, H.; Niyato, D.; Kang, J.; et al. The age of generative AI and AI-generated everything. *IEEE. Netw.* **2024**, *38*, 501-12. DOI
4. Yoon, J.; Jarrett, D.; van der Schaar, M. Time-series generative adversarial networks. 2019. Available from: https://proceedings.neurips.cc/paper_files/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf [Last accessed on 8 Apr 2025]
5. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Mach. Learn.* **2020**, *33*, 6840-51. DOI
6. Zhang, J.; Wu, F.; Wei, B.; et al. Data augmentation and dense-LSTM for human activity recognition using WiFi signal. *IEEE. Internet. Things. J.* **2021**, *8*, 4628-41. DOI
7. Wang, D.; Yang, J.; Cui, W.; Xie, L.; Sun, S. Multimodal CSI-based human activity recognition using GANs. *IEEE. Internet. Things. J.* **2021**, *8*, 17345-55. DOI
8. Wang, J. Du, H.; Niyato, D.; et al. Generative AI for integrated sensing and communication: insights from the physical layer perspective. *IEEE. Wirel. Commun.* **2024**, *31*, 246-55. DOI
9. Wang, Z.; Mao, S. Generative AI-empowered RFID sensing for 3D human pose augmentation and completion. *IEEE. Open. J. Commun. Soc.* **2025**. DOI
10. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: concept and applications. *ACM. Trans. Intell. Syst. Technol.* **2019**, *10*, 1-19. DOI