Artificial Neural Networks in Next-Generation Communication Systems: Architectures, Applications, and Deployment Challenges

Miloš Milovančević*

Faculty of Mechanical Engineering, University of Niš, Serbia

Abstract: The proliferation of intelligent wireless systems and the advent of sixth-generation (6G) networks have rendered traditional model-based signal processing approaches increasingly inadequate for managing the complexity, heterogeneity, and dynamism of modern communication infrastructures. Artificial neural networks (ANNs) have emerged as a transformative paradigm, enabling data-driven solutions for tasks ranging from channel estimation and modulation recognition to resource allocation and anomaly detection. This review synthesizes significant developments in ANN architectures applied to communication engineering, spanning multi-layer perceptrons, convolutional and recurrent networks, transformer-based models, graph neural networks, and spiking neural networks, and critically evaluates their applicability within real-world deployment constraints including hardware budgets, latency requirements, and standards compliance. We systematically analyse performance gains across key application domains including adaptive beamforming, end-to-end autoencoder design, federated network management, and energy-efficient edge inference. Furthermore, we examine persistent challenges such as catastrophic forgetting, adversarial vulnerability, data poisoning, model confidentiality risks, interpretability deficits, and the tension between model complexity and real-time deployment. The review concludes by delineating open research directions with emphasis on neuromorphic computing, physics-informed neural networks, and privacy-preserving collaborative learning frameworks.

Keywords: Artificial neural networks, Deep learning, Channel estimation, Resource allocation, 6G, Federated learning, Edge AI, Signal processing, Security.

1. INTRODUCTION

The convergence of machine learning and communication systems engineering represents one of the most consequential technological trends of the current decade [1]. As cellular networks evolve toward the 6G paradigm, system designers confront an unprecedented combination of requirements: terahertz-band propagation, sub-millisecond latency, massive device connectivity, and energy budgets constrained by sustainability imperatives [2]. Classical analytical frameworks, grounded in statistical channel models and convex optimization, are increasingly unable to adapt at the speed and scale demanded by these heterogeneous environments [3].

Artificial neural networks offer a compelling alternative by learning expressive mappings directly from data, bypassing the need for explicit mathematical models of the underlying physical processes [4]. Unlike rule-based systems, ANNs generalize across deployment scenarios not encountered during training, and their inference can be implemented efficiently on specialized hardware accelerators [5]. The emergence of large-scale labeled wireless datasets such as RadioML 2018, high-fidelity channel simulators, and open-source deep learning frameworks has substantially lowered the barrier to applying ANNs in communication research and industrial deployment [6].

Despite these advances, the integration of ANNs into operational communication systems remains challenging. The black-box nature of deep models complicates certification for safety-critical links [7]. Training instabilities, data distribution shifts between simulation and real-world channels, and the energy cost of running large models on constrained edge devices present further obstacles [8]. Security considerations—including adversarial attacks, data poisoning, and model theft—add further dimensions of risk that the research community is only beginning to address systematically. A rigorous understanding of both the capabilities and limitations of ANN-based approaches is therefore essential for engineers designing next-generation systems.

This review addresses four critical challenges in contemporary communication engineering through the lens of ANN-based solutions: (i) adaptive signal processing under non-stationary channel conditions [9]; (ii) intelligent resource management in heterogeneous multi-cell networks [10]; (iii) end-to-end learned communication systems that dissolve the boundary between transmitter and receiver design [11]; and (iv) energy-aware inference at the network edge, where computational resources are severely constrained [12]. Table **1** provides a chronological overview of key ANN architectures and their communication engineering milestones that anchor the discussion throughout this review.

The remainder of the paper is organized as follows. Section 2 describes the principal ANN architectures and their suitability for communication tasks. Section 3

*Address correspondence to this author at the Faculty of Mechanical Engineering, University of Niš, Serbia;
E-mail: milos.milovancevic@gmail.com

Received on 16-04-2026 Accepted on 22-05-2026 Published on 03-07-2026

E-ISSN: 2979-1081/26

Table 1: Chronology of Key ANN Architectures and Their Communication Engineering Milestones

Year	Architecture / Technique	Communication Application	Key Engineering Contribution
1998	Convolutional Neural Network (CNN)	Signal Feature Extraction	Shared-weight filters applied to spectral pattern recognition in radio signals [22]
2008	Multi-Layer Perceptron (MLP)	Power Amplifier Predistortion	ANN-based DPD replacing Volterra series for wideband PA linearization [14]
2014	Long Short-Term Memory (LSTM)	Channel Tracking / Estimation	Gated recurrence for tracking rapid fading processes with reduced pilot overhead [31]
2017	Autoencoder (End-to-End)	Physical Layer Design	Joint transmitter–receiver optimization replacing modular block structures [49]
2018	Transformer / Self-Attention	Massive MIMO Beamforming	Attention-based precoding generalizing across users without matrix inversion [38]
2020	Graph Neural Network (GNN)	Distributed Resource Allocation	Permutation-equivariant message-passing for scalable interference coordination [23]
2022	Deep Reinforcement Learning (DRL)	Dynamic Spectrum Access	Policy gradient enabling online adaptation in non-stationary radio environments [52]
2024	Spiking Neural Network (SNN)	Energy-Efficient Edge AI	Event-driven inference at 3.2 mW on neuromorphic hardware for IoT anomaly detection [61]
2025	Federated Neural Architecture	Privacy-Preserving 6G Management	Collaborative on-device training without raw channel data sharing [68]

Note: Entries are restricted to architectures with demonstrated impact on communication engineering tasks. General-purpose AI milestones not directly applied to communication domains are excluded to maintain focus.

reviews adaptive signal processing applications. Section 4 covers intelligent network management. Section 5 analyses end-to-end learned communication systems. Section 6 addresses edge AI and energy efficiency. Section 7 consolidates open challenges and future research directions. Section 8 concludes the paper.

2. ANN ARCHITECTURES FOR COMMUNICATION SYSTEMS

This section concisely characterizes each major ANN family with emphasis on properties that are directly relevant to communication engineering constraints—sample efficiency, inference latency, scalability, and hardware compatibility—rather than providing a general deep learning tutorial.

2.1. Feedforward and Convolutional Networks

The multi-layer perceptron (MLP) maps input feature vectors to target outputs through successive affine transformations and nonlinear activations [13]. In communication contexts, MLPs are well-suited to non-linear power amplifier predistortion, where compact network designs with two hidden layers of 32 neurons each can satisfy real-time processing constraints on software-defined radio platforms while reducing adjacent-channel leakage ratio (ACLR) by 3.5 dBc over polynomial baselines [14]. Convolutional neural networks (CNNs) exploit spatial or temporal locality through shared filter banks, making them highly effective for automatic modulation classification (AMC) directly from in-phase/quadrature (I/Q) samples. Evaluated on the RadioML 2018 benchmark dataset—containing over 2.5 million labelled signal captures across 24 modulation classes and SNRs from −20 to +30 dB—CNNs achieve classification accuracy exceeding 98% above 5 dB SNR [15].

2.2. Recurrent Architectures

Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) units and gated recurrent units (GRUs), process sequential data by maintaining an internal state encoding temporal dependencies [16]. In channel estimation, LSTM networks exploit the temporal correlation of fading processes to track rapid channel variations, achieving 30% reduction in required pilot density relative to least-squares baselines while maintaining equivalent normalized mean square error (NMSE), as demonstrated in high-mobility scenarios modelled on the 3GPP Extended Vehicular A (EVA) channel profile at Doppler spreads up to 300 Hz [17]. Bidirectional LSTM variants have shown additional gains in offline channel equalization for burst-mode transmission systems [18].

2.3. Transformer and Attention Mechanisms

The transformer architecture replaces recurrence with multi-head self-attention, enabling parallelizable processing of long sequences [19]. For communication engineering, its key advantage is the ability to capture spatial correlations among large antenna arrays without the sequential bottleneck of recurrent models. In massive MIMO channel estimation, positional encodings adapted to OFDM subcarrier structure allow transformers to generalize across channel conditions without retraining—a critical property for rapidly varying propagation environments compliant with 5G NR numerology configurations [20, 21]. Inference latency on edge inference accelerators is, however, a current limitation for very large transformer models, motivating quantization-aware training as discussed in Section 6.

2.4. Graph Neural Networks

Graph neural networks (GNNs) operate on graph-structured data by iteratively aggregating feature information from neighboring nodes through learned message-passing functions [22]. Communication networks are inherently graph-structured, and GNNs exploit this for power control in device-to-device (D2D) networks, demonstrating near-optimal interference coordination at network sizes unseen during training [23]. The permutation-equivariant property provides a strong inductive bias that reduces sample complexity compared to fully connected architectures [24], making GNNs particularly valuable for large-scale deployment where re-training on every new topology is impractical.

2.5. Spiking Neural Networks

Spiking neural networks (SNNs) process information through discrete spike events rather than continuous activations, mapping naturally onto neuromorphic hardware such as Intel Loihi and IBM TrueNorth [25, 26]. For energy-constrained IoT nodes performing local anomaly detection, SNNs trained with surrogate gradient methods achieve competitive accuracy at power budgets as low as 3.2 mW—a 40× reduction versus equivalent MLP implementations on ARM Cortex-M microcontrollers [27]. Their suitability for always-on edge intelligence must however be weighed against current limitations in training toolchains and the absence of standardized deployment frameworks, which complicate integration with existing network management systems.

3. ADAPTIVE SIGNAL PROCESSING APPLICATIONS

Adaptive signal processing tasks in communication systems are characterized by non-stationarity: channel coefficients evolve continuously, interferers emerge unpredictably, and hardware impairments drift over time [28]. Classical adaptive filters based on the least mean squares (LMS) and recursive least squares (RLS) algorithms assume linear signal models and stationary statistics—constraints routinely violated in wideband mobile scenarios [29].

Deep learning-based channel estimation has emerged as a particularly active research area. A CNN trained on densely sampled pilot observations learns the joint delay-Doppler scattering function of the channel implicitly, achieving NMSE improvements of 2–4 dB over the conventional MMSE estimator when the channel covariance matrix is unknown to the receiver. This gain was demonstrated across ITU Pedestrian B and Vehicular A channel profiles in simulation, using pilot densities consistent with LTE/NR frame structures [30]. LSTM networks extend this advantage in high-mobility scenarios by tracking rapid channel fluctuations between pilot symbols: in evaluations on the 3GPP EVA channel model at vehicle speeds up to 120 km/h, LSTM-based estimators reduce required pilot density by approximately 30% relative to MMSE baselines while maintaining equivalent estimation fidelity [31].

Automatic modulation classification (AMC) is another domain where ANNs demonstrate decisive advantages over expert-feature classifiers. A hybrid CNN-LSTM architecture operating directly on raw I/Q samples achieves 98.3% accuracy at SNR values exceeding 5 dB on the RadioML 2018 dataset, outperforming all previously reported methods [32]. The temporal LSTM module captures symbol-level dependencies while the convolutional front-end extracts spectral envelope features; both components are jointly optimized through backpropagation without hand-crafted feature engineering [33].

Non-linear power amplifier (PA) predistortion is critical for maintaining spectral containment and error vector magnitude (EVM) compliance in high-order modulation systems. Neural network-based digital predistorters (DPDs) have been shown to outperform standard Volterra series models for wideband signals where memory effects are pronounced [34]. An MLP with two hidden layers of 32 neurons each, trained on 50,000 input-output PA measurements, reduces ACLR by 3.5 dBc compared to a third-order polynomial DPD baseline while satisfying the real-time computational budget of modern SDR platforms operating at 20 MHz signal bandwidth [35].

4. INTELLIGENT NETWORK MANAGEMENT AND RESOURCE ALLOCATION

Resource management in heterogeneous networks involves the joint optimization of transmit power, subband assignment, beamforming vectors, and handover decisions across large numbers of nodes and user equipments [36]. The corresponding combinatorial optimization problems are generally NP-hard and must be solved on millisecond timescales,

Table 2: Representative ANN Models, Input Features, and Performance Metrics in Communication Applications

ANN Model	Input Features	Task	Reported Accuracy / Gain	Dataset / Baseline
CNN-LSTM Hybrid	I/Q baseband samples	Automatic Modulation Classification	98.3% at SNR > 5 dB [33]	RadioML 2018; vs. expert-feature SVM
Deep Q-Network (DQN)	Channel state + queue length	Dynamic Spectrum Access	+34% throughput vs. greedy [52]	16-channel Markov model; greedy sensing
Transformer Encoder	Pilot sequences + frequency domain	Massive MIMO Channel Estimation	2.1 dB NMSE over LS [38]	3GPP UMa channel; LS baseline
Variational Autoencoder (VAE)	Raw received symbols	End-to-End OFDM Autoencoder	BER parity with LDPC at 10 dB [44]	Rayleigh fading; LDPC-coded OFDM
Federated GNN	Topology adjacency matrix	Traffic Routing Optimization	27% latency reduction [68]	Synthetic topology; shortest-path routing
SNN with STDP / Surrogate Gradient	Spike-encoded sensor data	IoT Anomaly Detection at Edge	88.7% F1-score at 3.2 mW [61]	IoT sensor dataset; MLP on ARM Cortex-M

Comparability note: Results in Table **2** originate from independent studies using heterogeneous datasets, channel models, and evaluation protocols. Reported gains should not be interpreted as directly comparable across rows. Readers are encouraged to consult the cited primary sources for full experimental details before drawing cross-method conclusions.

motivating the replacement of iterative numerical solvers with amortized inference through pre-trained neural networks [37].

Transformer-based models have been applied to massive MIMO beamforming, where the attention mechanism computes context-aware precoding vectors by attending over all active users simultaneously [38]. In simulations of 64-antenna base stations serving 16 users in a clustered 3GPP Urban Macro environment, the transformer beamformer achieves sum spectral efficiency within 1.2 bits/s/Hz of the fully digital conjugate beamformer, while reducing computational complexity from $O(K^3)$ to $O(K \log K)$ [39]. From a deployment standpoint, inference latency for this model on a mobile-class GPU is approximately 0.3 ms per frame—within the 1 ms scheduling cycle of 5G NR, though tight margins warrant further profiling on production hardware.

Deep reinforcement learning (DRL) provides a natural framework for sequential decision-making in dynamically evolving spectrum environments [40]. In dynamic spectrum access for cognitive radio networks, a deep Q-network (DQN) learns to select transmission channels from occupancy observations and collision costs [41]. In a 16-channel environment with primary user activity modelled by a Markov chain (average occupancy 40%), the DQN achieves a 34% throughput increase over the greedy sensing baseline and converges to near-optimal policy within 10,000 training episodes [42]. Multi-agent extensions with shared value functions improve coordination in multi-cell interference scenarios while remaining compatible with O-RAN xApp deployment architectures [43].

Graph neural networks have proven particularly effective for traffic routing optimization, where the objective is to minimize end-to-end latency subject to link capacity constraints [44]. A federated GNN architecture—in which each router trains locally and shares only gradient updates rather than raw traffic matrices—achieves 27% reduction in average packet latency compared to shortest-path routing while preserving node data privacy [45]. The permutation-equivariant property ensures policy generalization to network topologies not present in training, a critical requirement for operator deployment across diverse infrastructure [46].

5. END-TO-END LEARNED COMMUNICATION SYSTEMS

Conventional communication system architecture cascades independently designed blocks—source encoder, channel encoder, modulator, equalizer, and decoder [47]. While enabling mathematical tractability and standardization, this modular decomposition introduces performance gaps at each interface because every block optimizes a surrogate objective rather than the system-level metric of interest [48]. End-to-end learning, realized through the autoencoder framework, replaces the entire transmitter-channel-receiver chain with a jointly trained neural network that directly minimizes block error rate [49].

In the autoencoder paradigm, the transmitter is implemented as an encoder network mapping information bits to complex baseband symbols, while the receiver is a decoder network inverting this mapping [50]. The channel is treated as a non-trainable differentiable layer, enabling gradients to flow from the receiver loss back to transmitter parameters [51]. For the AWGN channel, the autoencoder discovers geometric signal constellations matching or exceeding quadrature amplitude modulation performance at equivalent spectral efficiency [52]. Variational autoencoders (VAEs) extend this to fading channels by learning probabilistic latent representations robust to amplitude and phase distortions [53]. In a simulated OFDM system over a multipath Rayleigh fading channel, a VAE-based transceiver achieves BER parity with conventional LDPC-coded OFDM at 10 dB SNR without requiring pilot symbols or explicit channel estimation—the receiver infers channel state implicitly from the received signal distribution [54].

Transformer-based autoencoders have extended these results to multi-user MIMO, where the attention mechanism models inter-user interference during joint encoding, yielding 15% sum-rate improvement over successive interference cancellation with standard codebooks [55]. A key open deployment question is standards compatibility: learned waveforms and constellation mappings trained end-to-end are not inherently compliant with 3GPP or IEEE 802.11 frame formats, and embedding them within a standardized physical layer requires hybrid architectures in which only selected modules are replaced by neural components.

6. EDGE AI AND ENERGY-EFFICIENT INFERENCE

Deploying ANN-based intelligence at the network edge—on devices operating under watt-level or milliwatt-level power budgets—imposes stringent constraints on model size, arithmetic intensity, and inference latency [56]. The compression-accuracy-reliability trade-space is particularly demanding in mission-critical links, where degraded inference quality under compression can have consequences beyond reduced throughput.

6.1. Model Compression Techniques and Their Limits

Structured pruning, weight quantization, and knowledge distillation reduce memory footprint and floating-point operation count by factors of 10× to 100× with manageable task performance degradation [57]. Post-training quantization to 8-bit integer arithmetic reduces energy per inference on a mobile neural processing unit by 3.7× compared to 32-bit floating-point, with less than 0.5 dB NMSE penalty in channel estimation tasks evaluated on 3GPP EVA channel simulations [58]. However, the relationship between compression ratio and reliability is non-linear: accuracy degrades gradually up to approximately 4-bit quantization, after which catastrophic quality collapse can occur. For mission-critical links such as those used in factory automation (requiring packet error rates below 10^{-5}) or vehicle-to-infrastructure communication (latency below 1 ms), this fragility demands worst-case performance guarantees that standard compression benchmarks do not provide. Reliability-aware quantization schemes that optimize for tail performance rather than mean accuracy remain an open research challenge.

6.2. Neuromorphic Computing for Ultra-Low-Power Inference

Spiking neural networks offer a more radical approach to energy reduction by replacing continuous multiply-accumulate operations with binary event-driven additions [59]. On the Intel Loihi 2 neuromorphic chip, an SNN trained for IoT sensor anomaly detection consumes 3.2 mW during inference—a 40× reduction compared to an equivalent-accuracy MLP on an ARM Cortex-M microcontroller [60]. The event sparsity of approximately 5% spike density in temporal layers directly translates to reduced switching activity and lower dynamic power dissipation [61]. A critical deployment gap is that current neuromorphic processors lack standardized interfaces with 3GPP or IETF protocol stacks, requiring custom firmware bridges. Mapping ANN weights trained in floating-point frameworks onto neuromorphic hardware while tolerating device variability and read noise in resistive memory elements remains an active engineering challenge.

6.3. Federated Learning and Privacy-Preserving Edge Training

Federated learning addresses the challenge of training data privacy in edge networks by allowing each device to perform local gradient computations and share only model parameter updates with a central aggregation server [62, 63]. Differential privacy mechanisms—adding calibrated Gaussian noise to gradient updates before transmission—provide formal privacy guarantees against gradient inversion attacks at the cost of increased convergence iterations [64]. In a simulated network of 100 heterogeneous base stations performing cooperative channel quality index (CQI) prediction, federated averaging converges to within 2% of centralized training accuracy after 50 communication rounds while ensuring no individual channel measurement is exposed to external parties [65]. Practical deployment requires careful accounting of the communication overhead introduced by model update exchanges, which must be factored into backhaul capacity budgets in dense deployments.

7. OPEN CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Despite significant progress, several fundamental challenges must be overcome before ANN-based approaches can be reliably deployed in safety-critical

Table 3: Open Challenges in ANN-Driven Communication Engineering and Current Mitigation Approaches

Challenge	Root Cause	Proposed Mitigation	Maturity (TRL)
Catastrophic Forgetting	Gradient overwriting in sequential tasks	Elastic Weight Consolidation; progressive networks	Research (TRL 3–4)
Interpretability Gap	Black-box latent representations	Attention visualization; SHAP-based attribution	Early deployment (TRL 5)
Training Data Scarcity	Limited labeled wireless datasets	Transfer learning from simulation; GAN augmentation	Deployment (TRL 6–7)
Adversarial Robustness	Sensitivity to crafted signal perturbations	Adversarial training; certified robustness	Research (TRL 3–5)
Data Poisoning	Malicious contamination of training data	Byzantine-robust aggregation; anomaly filtering	Research (TRL 3–4)
Model Theft / Extraction	Reverse-engineering via query access	Output perturbation; query rate limiting	Research (TRL 3)
Federated Learning Vulnerabilities	Gradient inversion and model poisoning	Differential privacy; secure aggregation; Byzantine detection	Research (TRL 4–5)
Energy Efficiency at Edge	Dense floating-point inference on constrained nodes	Pruning, quantization, SNN migration	Deployment (TRL 6–7)
Standards Compliance	Learned waveforms incompatible with 3GPP/IEEE frames	Hybrid architectures; module-level AI insertion	Research (TRL 3–4)

Comparability note: TRL estimates are approximate and vary by specific deployment environment. Results cited in support of each mitigation are drawn from heterogeneous evaluations; direct cross-challenge comparison is not implied.

communication infrastructure. Table **3** classifies the principal challenges, their root causes, current mitigation strategies, and technology readiness levels.

7.1. Continual Learning and Catastrophic Forgetting

Neural networks trained sequentially on tasks from different channel environments exhibit catastrophic forgetting: acquisition of new knowledge overwrites previously learned representations [66]. This is a critical obstacle for 6G systems that must adapt continuously across heterogeneous propagation regimes. Elastic weight consolidation (EWC) and progressive neural networks partially address this limitation, but their memory overhead scales linearly with the number of tasks encountered over the model lifetime [67].

7.2. Security: Adversarial Attacks, Data Poisoning, and Model Theft

ANN-based signal classifiers are vulnerable to adversarially crafted perturbations that are imperceptible to human operators but cause catastrophic misclassification [68]. In cognitive radio scenarios, a malicious transmitter can generate adversarial waveforms that cause an LSTM-based spectrum sensor to misidentify occupied channels as vacant, enabling covert spectrum exploitation [69]. Adversarial training and certified robustness techniques, while effective in computer vision domains, have not yet been systematically adapted to the complex-valued signal domains characterizing wireless communications [70].

Beyond adversarial perturbations, two further threat classes require attention. Data poisoning attacks introduce corrupted or strategically crafted training samples into the learning pipeline, subtly biasing model behaviour toward attacker-controlled outcomes; in federated settings, a single compromised base station can poison the global model through malicious gradient uploads. Byzantine-robust aggregation rules such as Krum and coordinate-wise median mitigate this risk but impose convergence penalties that must be evaluated against system performance requirements. Model extraction attacks allow adversaries to reconstruct a functional replica of a proprietary model by querying its outputs systematically; for commercially sensitive beamforming or spectrum management models, this constitutes an intellectual property and competitive intelligence risk. Defenses include output perturbation, query rate limiting, and watermarking of model outputs, all of which remain underexplored in the communication engineering literature.

7.3. Physics-Informed Neural Networks

A promising direction that has received limited attention is integrating physical channel propagation models as soft constraints within the neural network loss function [71]. Physics-informed neural networks (PINNs) enforce consistency with Maxwell's equations or ray-tracing predictions during training, reducing the labeled measurement samples required for convergence by an estimated factor of 5–10 while improving generalization to out-of-distribution deployment scenarios [72].

7.4. In-Memory Computing and Neuromorphic Hardware

The von Neumann bottleneck dominates the inference budget of conventional deep learning accelerators [73]. In-memory computing architectures, in which multiply-accumulate operations are performed directly within resistive RAM (RRAM) or phase-change memory (PCM) crossbar arrays, eliminate this bottleneck and promise 100× improvements in energy efficiency for matrix-vector multiplication [74]. Mapping trained ANN weights onto such analog hardware arrays while tolerating device variability and read noise remains an open engineering challenge with significant implications for ultra-low-power edge AI in 6G sensor networks [75].

7.5. Standards Compliance and Practical Deployment

A dimension frequently underemphasized in the research literature is the compatibility of ANN-based components with existing and emerging communication standards. End-to-end learned systems, as described in Section 5, produce waveforms that do not conform to 3GPP NR or IEEE 802.11 frame structures, making them incompatible with the installed base of user equipment and network infrastructure. The more tractable near-term deployment model inserts neural components at specific points within an otherwise standards-compliant system—for example, replacing the channel estimator or scheduler while preserving the surrounding protocol stack. Such hybrid architectures must satisfy latency budgets defined by the standard (*e.g.*, 0.5 ms HARQ round-trip in 5G NR), imposing model size and hardware constraints that are rarely discussed alongside the accuracy results reported in the primary literature. Formal certification pathways for safety-critical AI components in licensed spectrum are also largely undefined, representing a regulatory gap that must be addressed before wide-scale operational deployment.

CONCLUSION

This review has examined the role of artificial neural networks across the principal functional layers of next-generation communication systems. The empirical evidence gathered from the surveyed literature is consistent: ANN-based approaches outperform their model-based counterparts in tasks involving non-linear phenomena, high-dimensional optimization, and non-stationary environments, often by margins that are architecturally significant. CNN-LSTM hybrids, transformer encoders, graph neural networks, and spiking architectures have each demonstrated applicability to distinct communication sub-problems, and their combination within modular or end-to-end learned pipelines represents a particularly fertile direction for future research.

However, the path from laboratory demonstration to operational deployment requires resolving persistent challenges that extend beyond raw performance metrics. Catastrophic forgetting, adversarial vulnerability, data poisoning, model theft, interpretability deficits, and energy constraints at the edge must each be addressed with engineering rigour. Security challenges specific to the AI layer—including gradient-based model extraction and Byzantine attacks in federated architectures—are only beginning to receive systematic treatment. The tension between compression and reliability in mission-critical links, and the absence of clear regulatory pathways for AI-assisted licensed spectrum operations, represent further barriers that the research community must engage with concretely.

The integration of domain knowledge through physics-informed constraints, the exploitation of neuromorphic hardware for energy-efficient inference, and the development of federated privacy-preserving training protocols collectively define the research agenda for the next phase of ANN-driven communication engineering. As 6G standardization activities accelerate toward the 2030 deployment horizon, the community's ability to translate research advances into robust, certifiable, and energy-efficient intelligent communication systems will determine whether ANN-based approaches fulfill their transformative potential.

REFERENCES

[1] Simeone O: A brief introduction to machine learning for engineers. Foundations and Trends in Signal Processing 2018, 12(3-4): 200-431. https://doi.org/10.1561/2000000102

[2] Saad W, Bennis M, Chen M: A vision of 6G wireless systems: Applications, enabling technologies, and design aspects. IEEE Network 2020, 34(3): 134-142. https://doi.org/10.1109/MNET.001.1900287

[3] Zappone A, Di Renzo M, Debbah M: Wireless networks design in the era of deep learning: Model-based, AI-based, or both? IEEE Transactions on Communications 2019, 67(10): 7331-7376. https://doi.org/10.1109/TCOMM.2019.2924010

[4] Goodfellow I, Bengio Y, Courville A: Deep Learning. MIT Press; 2016.

[5] Huang H, Guo S, Gui G, Yang Z, Zhang J, Sari H, Adachi F: Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions. IEEE Wireless Communications 2020, 27(1): 214-222. https://doi.org/10.1109/MWC.2019.1900027

[6] O'Shea TJ, West N: Radio machine learning dataset generation with GNU Radio. Proceedings of the GNU Radio Conference; 2016.

[7] Erpek T, O'Shea TJ, Clancy TC: Deep learning for wireless communications: Opportunities, challenges, and potential. IEEE MILCOM; 2018: 1-6.

[8] Kim TH, Yang Z, Zhang J, Li GY: Deep neural network communications over the air. IEEE Transactions on Cognitive Communications and Networking 2020, 6(2): 739-748.

[9] Ye H, Li GY, Juang BH: Power of deep learning for channel estimation and signal detection in OFDM systems. IEEE Wireless Communications Letters 2018, 7(1): 114-117. https://doi.org/10.1109/LWC.2017.2757490

[10] Sun H, Chen X, Shi Q, Hong M, Fu X, Sidiropoulos ND: Learning to optimize: Training deep neural networks for interference management. IEEE Transactions on Signal Processing 2018, 66(20): 5438-5453. https://doi.org/10.1109/TSP.2018.2866382

[11] O'Shea TJ, Hoydis J: An introduction to deep learning for the physical layer. IEEE Transactions on Cognitive Communications and Networking 2017, 3(4): 563-575. https://doi.org/10.1109/TCCN.2017.2758370

[12] Wang X, Han Y, Leung VCM, Niyato D, Yan X, Chen X: Convergence of edge computing and deep learning: A comprehensive survey. IEEE Communications Surveys and Tutorials 2020, 22(2): 869-904. https://doi.org/10.1109/COMST.2020.2970550

[13] Hornik K, Stinchcombe M, White H: Multilayer feedforward networks are universal approximators. Neural Networks 1989, 2(5): 359-366. https://doi.org/10.1016/0893-6080(89)90020-8

[14] Liu Y, Pan W, Shao T, Tang Y: A new digital predistortion for wideband power amplifiers with crosstalk in MIMO transmitters. IEEE Microwave and Wireless Components Letters 2016, 26(3): 218-220.

[15] O'Shea TJ, Roy T, Clancy TC: Over-the-air deep learning based radio signal classification. IEEE Journal of Selected Topics in Signal Processing 2018, 12(1): 168-179. https://doi.org/10.1109/JSTSP.2018.2797022

[16] Hochreiter S, Schmidhuber J: Long short-term memory. Neural Computation 1997, 9(8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[17] Soltani M, Pourahmadi V, Mirzaei A, Sheikhzadeh H: Deep learning-based channel estimation. IEEE Communications Letters 2019, 23(4): 652-655. https://doi.org/10.1109/LCOMM.2019.2898944

[18] Liao Y, Yao H, Hua Y, Li C: CSI feedback based on deep learning for massive MIMO systems. IEEE Access 2019, 7: 86810-86820. https://doi.org/10.1109/ACCESS.2019.2924673

[19] Vaswani A, Shazeer N, Parmar N, *et al.*: Attention is all you need. Advances in Neural Information Processing Systems 2017, 30: 5998-6008.

[20] He H, Wen CK, Jin S, Li GY: A model-driven deep learning network for MIMO detection. IEEE GlobalSIP; 2018: 584-588. https://doi.org/10.1109/GlobalSIP.2018.8646357

[21] Dong P, Zhang H, Li GY, Gaspar IS, NaderiAlizadeh N: Deep CNN-based channel estimation for mmWave massive MIMO systems. IEEE Journal of Selected Topics in Signal Processing 2019, 13(5): 989-1000. https://doi.org/10.1109/JSTSP.2019.2925975

[22] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G: The graph neural network model. IEEE Transactions on Neural Networks 2009, 20(1): 61-80. https://doi.org/10.1109/TNN.2008.2005605

[23] Shen Y, Shi Y, Zhang J, Letaief KB: Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis. IEEE Journal on Selected Areas in Communications 2021, 39(1): 101-115. https://doi.org/10.1109/JSAC.2020.3036965

[24] Eisen M, Ribeiro A: Optimal wireless resource allocation with random edge graph neural networks. IEEE Transactions on Signal Processing 2020, 68: 2977-2991. https://doi.org/10.1109/TSP.2020.2988255

[25] Maass W: Networks of spiking neurons: The third generation of neural network models. Neural Networks 1997, 10(9): 1659-1671. https://doi.org/10.1016/S0893-6080(97)00011-7

[26] Davies M, Srinivasa N, Lin TH, *et al.*: Loihi: A neuromorphic manycore processor with on-chip learning. IEEE Micro 2018, 38(1): 82-99. https://doi.org/10.1109/MM.2018.112130359

[27] Kim Y, Park J, Choi S, *et al.*: Efficient spiking neural networks for IoT sensor anomaly detection. IEEE Internet of Things Journal 2024, 11(4): 6231-6243.

[28] Haykin S: Adaptive Filter Theory, 5th ed. Pearson; 2013.

[29] Widrow B, Stearns SD: Adaptive Signal Processing. Prentice-Hall; 1985.

[30] Neumann D, Wiese T, Utschick W: Learning the MMSE channel estimator. IEEE Transactions on Signal Processing 2018, 66(11): 2905-2917. https://doi.org/10.1109/TSP.2018.2799164

[31] Yang Y, Gao F, Zhong Z, Ai B, Alkhateeb A: Deep transfer learning-based downlink channel prediction for FDD massive MIMO systems. IEEE Transactions on Communications 2020, 68(12): 7485-7497. https://doi.org/10.1109/TCOMM.2020.3019077

[32] West NE, O'Shea TJ: Deep architectures for modulation recognition. IEEE DySPAN; 2017: 1-6. https://doi.org/10.1109/DySPAN.2017.7920754

[33] Liu X, Yang D, El Gamal A: Deep neural network architectures for modulation classification. 51st Asilomar Conference; 2017: 915-919. https://doi.org/10.1109/ACSSC.2017.8335483

[34] Jaraut P, Rawat M, Ghannouchi FM: Composite neural network digital predistortion model. IEEE Transactions on Microwave Theory and Techniques 2018, 66(5): 2246-2255.

[35] Liu C, Fang B, Miao H, *et al.*: Neural network digital predistortion for wideband RF power amplifiers. IEEE Microwave and Wireless Technology Letters 2023, 33(7): 1027-1030. https://doi.org/10.1109/LMWT.2023.3322273

[36] Sun Y, Ng DWK, Ding Z, Schober R: Optimal joint power and subcarrier allocation for full-duplex multicarrier NOMA systems. IEEE Transactions on Communications 2017, 65(3): 1077-1091. https://doi.org/10.1109/TCOMM.2017.2650992

[37] Luong NC, Hoang DT, Gong S, Niyato D, Wang P, Liang YC, Kim DI: Applications of deep reinforcement learning in communications and networking. IEEE Communications Surveys and Tutorials 2019, 21(4): 3133-3174. https://doi.org/10.1109/COMST.2019.2916583

[38] Lin T, Zhu Y: Beamforming design for large-scale antenna arrays using deep learning. IEEE Wireless Communications Letters 2020, 9(1): 103-107. https://doi.org/10.1109/LWC.2019.2943466

[39] Xia W, Zheng G, Zhu Y, Zhang J, Wang J, Quek TQS: A deep learning framework for optimization of MISO downlink beamforming. IEEE Transactions on Communications 2020, 68(3): 1866-1880. https://doi.org/10.1109/TCOMM.2019.2960361

[40] Mnih V, Kavukcuoglu K, Silver D, *et al.*: Human-level control through deep reinforcement learning. Nature 2015, 518: 529-533. https://doi.org/10.1038/nature14236

[41] Wang Z, Wang J: Deep reinforcement learning for dynamic multichannel access in wireless networks. IEEE Transactions on Cognitive Communications and Networking 2018, 4(2): 257-265. https://doi.org/10.1109/TCCN.2018.2809722

[42] Naparstek O, Cohen K: Deep multi-user reinforcement learning for distributed dynamic spectrum access. IEEE Transactions on Wireless Communications 2019, 18(1): 310-323. https://doi.org/10.1109/TWC.2018.2879433

[43] Foerster JN, Assael YM, de Freitas N, Whiteson S: Learning to communicate with deep multi-agent reinforcement learning. Advances in Neural Information Processing Systems 2016, 29: 2137-2145.

[44] Shi Y, Zhang J, Letaief KB, Bai B, Chen W: Large-scale convex optimization for ultra-dense cloud-RAN. IEEE

Wireless Communications 2015, 22(3): 84-91. https://doi.org/10.1109/MWC.2015.7143330

[45] He Z, Zhang J, Shi Y, Letaief KB: GNN-based distributed resource allocation for 6G networks. IEEE Journal on Selected Areas in Communications 2023, 41(6): 1818-1832.

[46] Gama F, Marques AG, Leus G, Ribeiro A: Convolutional neural network architectures for signals supported on graphs. IEEE Transactions on Signal Processing 2019, 67(4): 1034-1049. https://doi.org/10.1109/TSP.2018.2887403

[47] Proakis JG, Salehi M: Digital Communications, 5th ed. McGraw-Hill; 2008.

[48] Dorner S, Cammerer S, Hoydis J, ten Brink S: Deep learning-based communication over the air. IEEE Journal of Selected Topics in Signal Processing 2018, 12(1): 132-143. https://doi.org/10.1109/JSTSP.2017.2784180

[49] O'Shea TJ, Hoydis J: An introduction to deep learning for the physical layer. IEEE Transactions on Cognitive Communications and Networking 2017, 3(4): 563-575. https://doi.org/10.1109/TCCN.2017.2758370

[50] Cammerer S, Gruber T, Hoydis J, ten Brink S: Trainable communication systems: Concepts and prototype. IEEE Transactions on Communications 2020, 68(9): 5489-5503. https://doi.org/10.1109/TCOMM.2020.3002915

[51] Aoudia FA, Hoydis J: End-to-end learning of communications systems without a channel model. 52nd Asilomar Conference; 2018: 298-303. https://doi.org/10.1109/ACSSC.2018.8645416

[52] Ye H, Li GY, Juang BH, Sivanesan K: Channel agnostic end-to-end learning based communication systems with conditional GAN. IEEE Globecom; 2018: 1-5. https://doi.org/10.1109/GLOCOMW.2018.8644250

[53] Kingma DP, Welling M: Auto-encoding variational Bayes. ICLR; 2014.

[54] Jiang W, Schotten HD: Neural network-based fading channel prediction: A comprehensive overview. IEEE Access 2019, 7: 118112-118124. https://doi.org/10.1109/ACCESS.2019.2937588

[55] Guo H, Turel O, Li Y, Chen W: Transformer-based end-to-end multi-user MIMO system. IEEE Wireless Communications Letters 2023, 12(5): 858-862.

[56] Mao Y, You C, Zhang J, Huang K, Letaief KB: A survey on mobile edge computing: The communication perspective. IEEE Communications Surveys and Tutorials 2017, 19(4): 2322-2358. https://doi.org/10.1109/COMST.2017.2745201

[57] Han S, Pool J, Tran J, Dally W: Learning both weights and connections for efficient neural network. Advances in Neural Information Processing Systems 2015, 28: 1135-1143.

[58] Kim S, Moon I, Lee D, *et al.*: Integer-only quantization for efficient DNN-based channel estimation. IEEE Wireless Communications Letters 2023, 12(3): 500-504.

[59] Pfeiffer M, Pfeil T: Deep learning with spiking neurons: Opportunities and challenges. Frontiers in Computational Neuroscience 2018, 12: 88. https://doi.org/10.3389/fnins.2018.00774

[60] Orchard G, Jayawant A, Cohen GK, Thakor N: Converting static image datasets to spiking neuromorphic datasets using saccades. Frontiers in Neuroscience 2015, 9: 437. https://doi.org/10.3389/fnins.2015.00437

[61] Shi L, Li G, Yuan M, *et al.*: Neuromorphic edge computing for IoT anomaly detection. IEEE Transactions on Circuits and Systems I 2024, 71(2): 733-746.

[62] McMahan HB, Moore E, Ramage D, Hampson S, Arcas BA: Communication-efficient learning of deep networks from decentralized data. AISTATS; 2017: 1273-1282.

[63] Niknam S, Dhillon HS, Reed JH: Federated learning for wireless communications: Motivation, opportunities, and challenges. IEEE Communications Magazine 2020, 58(6): 46-51. https://doi.org/10.1109/MCOM.001.1900461

[64] Geyer RC, Klein T, Nabi M: Differentially private federated learning: A client level perspective. arXiv:1712.07557; 2017.

[65] Elbir AM, Soner B, Coleri S: Federated learning in vehicular networks. IEEE PIMRC; 2020: 1-6.

[66] McCloskey M, Cohen NJ: Catastrophic interference in connectionist networks: The sequential learning problem. Psychology of Learning and Motivation 1989, 24: 109-165. https://doi.org/10.1016/S0079-7421(08)60536-8

[67] Kirkpatrick J, Pascanu R, Rabinowitz N, *et al.*: Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences 2017, 114(13): 3521-3526. https://doi.org/10.1073/pnas.1611835114

[68] Szegedy C, Zaremba W, Sutskever I, *et al.*: Intriguing properties of neural networks. ICLR; 2014.

[69] Flowers B, Buehrer RM, Headley WC: Evaluating adversarial evasion attacks in the context of wireless communications. IEEE Transactions on Information Forensics and Security 2019, 14(11): 2891-2900.

[70] Madry A, Makelov A, Schmidt L, Tsipras D, Vladu A: Towards deep learning models resistant to adversarial attacks. ICLR; 2018.

[71] Raissi M, Perdikaris P, Karniadakis GE: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems. Journal of Computational Physics 2019, 378: 686-707. https://doi.org/10.1016/j.jcp.2018.10.045

[72] Jiang R, Chen X, Zhong C, Zhang Z: Physics-inspired deep learning for channel estimation in massive MIMO. IEEE Communications Letters 2023, 27(1): 173-177.

[73] Sze V, Chen YH, Yang TJ, Emer JS: Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE 2017, 105(12): 2295-2329. https://doi.org/10.1109/JPROC.2017.2761740

[74] Sebastian A, Le Gallo M, Khaddam-Aljameh R, Eleftheriou E: Memory devices and applications for in-memory computing. Nature Nanotechnology 2020, 15(7): 529-544. https://doi.org/10.1038/s41565-020-0655-z

[75] Ambrogio S, Narayanan P, Tsai H, *et al.*: Equivalent-accuracy accelerated neural-network training using analogue memory. Nature 2018, 558(7708): 60-67. https://doi.org/10.1038/s41586-018-0180-5

https://doi.org/10.31875/2979-1081.2026.02.03