Physical Communication 75 (2026) 103008

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

A hybrid GAN and attention-based sparse autoencoder framework for robust end-to-end wireless communication

Safalata S. Sindal *, Y. N. Trivedi

Department of Electronics & Communication Engineering, Institute of Technology, Nirma University city, 382481, Gujrat, India

ARTICLE INFO

Keywords:
Generative adversarial networks
End-to-end communication
Autoencoder
Attention mechanism
Channel state information

ABSTRACT

This paper presents a hybrid deep learning framework that integrates a Generative Adversarial Network (GAN) with an Attention-based Sparse Autoencoder (GAN-AAE) for end-to-end wireless communication over Rayleigh fading channels with imperfect channel state information at the receiver (CSIR). Traditional autoencoder models lack the ability to learn underlying signal distributions or correct distortions caused by fading and noise. The proposed GAN-AAE addresses these limitations by using a generator as a learnable channel surrogate to refine encoded signals and an attention mechanism to dynamically prioritize relevant features for improved decoding. The imperfection in the CSI is quantified by a correlation coefficient ρ, where $0 \leq \rho \leq 1$. Perfect channel knowledge is denoted by $\rho = 1$, and decreasing values of ρ correspond to increasingly inaccurate CSIR. The model is jointly trained using adversarial and reconstruction losses to enhance its adaptability. Simulation results show that the GAN-AAE framework significantly outperforms conventional Maximum Likelihood Detection and baseline deep and convolutional neural network-based models in terms of bit error rate (BER). The model is evaluated over M-ary phase shift keying (M-PSK) and M-ary Quadrature Amplitude Modulation (M-QAM) with Rayleigh fading channel. At $\rho = 0.9$ and a signal-to-noise ratio (SNR) of 10 dB, the conventional baseline model achieves a BER of 0.072, whereas the proposed GAN-AAE achieves a lower BER of 0.02404 for Binary phase shift keying (BPSK). A complexity analysis indicates that although the GAN-AAE model introduces some additional computational overhead, the performance gains in reconstruction justify the trade-off. Overall, the GAN-AAE offers a resilient and adaptive solution for end-to-end communication under realistic wireless impairments.

1. Introduction

Modern communication systems are typically built using a block-wise structure, where each function, such as source coding, channel encoding, modulation, and decoding, is designed independently [1]. While this traditional approach is systematic and facilitates modular development and analysis, it may limit the overall system performance due to the lack of joint adaptation among components. In several studies, deep-learning-based communication models have demonstrated superior performance compared to conventional theoretical detection schemes by leveraging diverse training approaches and adaptive learning strategies under challenging channel conditions [2–6], highlighting the strong potential of deep-learning models to achieve high performance in wireless environments. In such scenarios, deep models learn an effective decision rule matched to the actual observed distribution, while analytical detectors remain tied to idealized mathematical assumptions. With the advancement of deep learning techniques, end-to-end communication system design using neural networks has emerged as a promising alternative [7]. In these systems, the transmitter and receiver are modeled as trainable neural networks and are trained together to learn an effective communication strategy directly from data, without relying on the conventional modular design.

However, while end-to-end learning-based systems offer a unified framework, their practical deployment faces challenges. In real-world scenarios, channel conditions are far from ideal, and perfect knowledge of channel state information (CSI) at the receiver is often unavailable. Although many early end-to-end learning frameworks assume perfect CSI to simplify system modeling and learning [8–11], this assumption is impractical in wireless environments. Obtaining accurate CSI in time-varying channels, especially under Rayleigh fading, requires significant overhead and is prone to estimation errors [12]. Therefore, incorporating imperfect CSI into the system model is crucial to developing robust and deployable end-to-end learning-based communication solutions.

Within this paradigm, autoencoders (AEs) [13] have been widely explored as a neural network architecture for modeling end-to-end

* Corresponding author.
E-mail addresses: 22ftphde68@nirmauni.ac.in (S.S. Sindal), yogesh.trivedi@nirmauni.ac.in (Y.N. Trivedi).

https://doi.org/10.1016/j.phycom.2026.103008
Received 20 November 2025; Received in revised form 31 December 2025; Accepted 13 January 2026
Available online 22 January 2026

Table 1
Summary of GAN applications in data-driven communication modeling.

Ref	Type of Neural Network	Contribution	Key Findings	Limitations
[19]	Basic GAN	Proposed a GAN framework to replicate AWGN channel characteristics without theoretical modeling.	Demonstrated GAN's potential for flexible, generalized channel modeling.	Lacks comparison with classical models; limited to AWGN scenarios.
[20]	SEGAN (Self-Attention GAN)	Designed a SEGAN model integrating self-attention for improved temporal feature extraction.	Improved speech enhancement with low memory overhead at selective layers.	Increased training complexity when attention spans multiple layers.
[21]	GAN for Data Augmentation	Employed GANs for data augmentation to improve signal classification under limited data.	Enhanced accuracy under data-scarce settings using synthetic samples.	Model performance relies heavily on initial dataset quality.
[22]	Conditional GAN (cGAN)	Developed a GAN model for joint frequency-domain channel statistics at mmWave/THz.	Captured complex cross-frequency channel behavior.	Limited generalization due to reliance on ray tracing simulations.
[23]	Image-to-Image GAN	Used GAN for reconstructing images from semantic maps to reduce bandwidth.	Maintained quality under distortion with reduced transmission load.	GAN at receiver adds heavy computation; not suited for constrained devices.
[24]	Restoration GAN	Proposed GANs for restoring wireless data loss without retransmissions.	Achieved over 90% restoration in various loss scenarios.	Sensitive to bias and distribution of training data.
[25]	Multiple GAN types	Surveyed GAN applications in wireless, including channel modeling, data augmentation, and security.	Emphasized adaptability of GAN to various wireless tasks.	Unstable training and complex hyperparameter tuning remain major barriers.

communication systems due to their ability to learn compressed representations of data and reconstruct them at the receiver. While effective, standard AEs often suffer from overfitting [14] and poor generalization. To mitigate this, the use of sparse AEs [15] has gained traction. By enforcing sparsity in the latent space, sparse AEs promote the learning of more compact and informative features. However, sparsity alone may not be sufficient to capture the variability introduced by uncertain or rapidly changing wireless channels. To address this limitation, attention mechanisms have emerged as a promising approach to further refine the feature extraction process [16]. These mechanisms enable the model to focus on the most relevant parts of the input signal during encoding and decoding, effectively prioritizing the transmission of critical information. However, several limitations hinder the effectiveness of the attention mechanism with an autoencoder in real-world fading scenarios. First, attention-based autoencoders primarily enhance feature selection but do not learn the underlying distribution of the transmitted signal, limiting their ability to handle unseen or highly perturbed inputs. Second, current architectures lack an internal correction or feedback mechanism that can compensate for inconsistencies in the received signal caused by noise, fading, or imperfect CSI. This leads to limited reconstruction capability of the decoder under rapidly varying or unpredictable channel conditions. Additionally, these models fail to dynamically adapt to fluctuations in the signal due to imperfect CSI, as they do not explicitly model uncertainty in the received data.

These limitations highlight the need for an enhanced mechanism that can model complex signal distributions and refine distorted received signals. This study focuses on addressing the challenges of signal degradation under fading environments with imperfect channel knowledge. To this end, we explore the integration of Generative Adversarial Networks (GANs) [17] into an attention-augmented sparse autoencoder framework for end-to-end communication. GANs are particularly effective in learning complex data distributions and have demonstrated success in refining outputs in various domains, making them well-suited for enhancing signal reconstruction in uncertain wireless channels. Recent studies have also explored GAN and ResNet-based architectures for wireless end-to-end communication design [18], demonstrating the growing applicability of adversarial techniques in this domain. By leveraging the generative capability of GANs alongside the selective feature extraction of attention mechanisms and the compact representation of sparse autoencoders, this work aims to develop a more resilient and adaptive communication model capable of operating effectively under practical, imperfect CSI conditions.

1.1. Related works

The application of GANs in wireless communication has attracted considerable attention due to their capacity to model complex data distributions and learn robust generative mappings without requiring explicit analytical formulations. Early attempts, such as in [19], employed a GAN-based framework for wireless channel modeling, showing that data-driven generative approaches could approximate statistical properties of traditional channel models like additive white Gaussian noise (AWGN) without complex theoretical derivations. However, these works were mostly confined to simplified channel conditions and lacked quantitative benchmarking against classical models. Subsequent efforts extended GAN to tasks beyond channel modeling. In Phan et al. [20], a Self-Attention GAN was introduced for speech enhancement, demonstrating the importance of integrating attention mechanisms to capture long-term dependencies in sequential data. While effective, this method incurred additional training overhead. Other studies, like Tang et al. [21], applied GAN for data augmentation in modulation classification, achieving notable gains under limited training data. However, the dependence on the quantity and diversity of original data posed generalization challenges. In the context of high-frequency communications, [22] proposed a GAN-based framework for modeling multi-frequency THz channels. Their model captured joint frequency-domain characteristics more accurately than traditional approaches. However, it relied heavily on ray tracing data, which might not generalize well to real-world conditions. Similarly, [23] used a pre-trained GAN for image reconstruction in semantic communication systems, illustrating GAN's potential to reduce bandwidth usage. Still, the computational demands at the receiver limited its applicability in edge devices. GANs have also been applied to wireless time-series restoration, as in Han and Na [24], showing high recovery rates for lost data. Nonetheless, the results were sensitive to training data quality. A broader review in [25] summarized these applications, reinforcing the versatility of GANs for wireless systems while cautioning about their training instability and high sensitivity to hyperparameter tuning. An organized summary of these research studies is shown in Table 1. In addition to these developments, recent studies have explored GAN architectures for various physical-layer wireless tasks. For example, GANs have been used for adaptive modulation and coding in next-generation 5G communication systems [26], and for enhancing received signal waveforms in fading environments [27]. GANs have also been applied for wireless channel data generation and modeling to improve training efficiency when channel measurements are limited [28].

Learning from the diverse applications of GANs across different neural network architectures in communication systems, we observed key advantages, including the effectiveness of using pre-trained GANs for enhanced signal reconstruction, the ability of attention mechanisms to selectively emphasize critical features, and the integration of GANs to compensate for data loss or channel distortion. These studies highlight the potential of GANs not only as standalone generative models but also as complementary modules within structured neural architectures. However, their practical deployment is often challenged by unstable training dynamics, sensitivity to hyperparameters, and dependence on large, high-quality datasets. These limitations can significantly impact the performance and generalization capability of communication models under real-world conditions, particularly in scenarios involving Rayleigh fading and imperfect CSI. To mitigate such instability, our design incorporates several stabilizing choices, including a controlled 1:1 generator-discriminator update schedule, small-step Adam optimization, and the use of a pre-trained generator structure that suppresses early-stage adversarial imbalance and reduces the risk of mode collapse. Motivated by these insights, our proposed model is designed to combine the generative strength of GANs with the sparsity and feature prioritization of attention-augmented autoencoders, while accounting for training stability, data efficiency, and robust signal recovery under imperfect CSI environments.

1.2. Main contributions

This study contributes to the advancement of deep learning-based communication systems by proposing a hybrid framework. The key contributions are as follows:

- We propose a hybrid GAN-AAE model, which integrates GAN within an attention-based sparse autoencoder for end-to-end wireless communication under Rayleigh fading with imperfect CSI.
- The training process of the proposed GAN-AAE is presented through an algorithm that enables coordinated interaction between the encoder-decoder path and the GAN components, allowing the model to learn effectively under channel uncertainty while incorporating a stability-aware GAN training strategy to ensure reliable adversarial behaviour within the hybrid architecture.
- To evaluate the effectiveness of the proposed framework, its performance is assessed using Bit Error Rate (BER) versus average signal-to-noise ratio (SNR) on M-PSK and M-QAM modulated Rayleigh fading channel under imperfect CSIR.
- We compare the performance of the proposed model against three benchmarks: a conventional Maximum Likelihood Detection (MLD) system, a baseline end-to-end AE model with Deep neural network (DNN)-based-channel module [29] and a baseline convolutional neural network (CNN)-based end-to-end AE model [30] under similar imperfect CSI conditions.
- We also compare the performance of the proposed model with all baseline models under perfect CSIR ($\rho = 1$) to quantify the gap between ideal channel knowledge and practical uncertainty.
- Through simulation results, we demonstrate improved robustness and signal reconstruction quality, showing that the proposed GAN-AAE model achieves lower error rates and better generalization in unpredictable channel scenarios. We also provide a complexity analysis to quantify the resulting performance-complexity trade-off across models.

The remainder of the paper is organized as follows: Section 2 presents the system model, including the detailed working of the proposed GAN-AAE architecture. Section 3 discusses the performance evaluation of the proposed system through simulation results. Finally, Section 4 concludes the article with key observations and potential future directions.

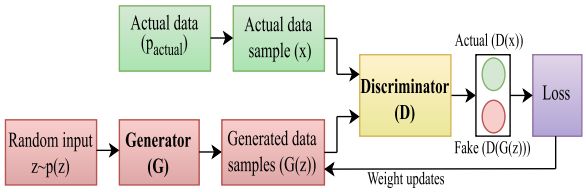

Fig. 1. Basic working principle of a generative adversarial network (GAN).

2. System model

This section presents a GAN-Attention-based end-to-end sparse autoencoder communication system model (GAN-AAE) for a Rayleigh fading channel with imperfect channel state information at the receiver (CSIR).

From a set of M possible messages, the transmitter selects a message $d \in \mathcal{M} = \{1, 2, \ldots, M\}$. The message is transmitted over n channel uses, with $n = k$. Each message d is expressed as a k-bit binary vector with $k = \log_2(M)$. This k-bit message is mapped to a modulated channel input vector $\mathbf{s} = [s_1, \ldots, s_n]^{\mathsf{T}}$, where each s_i is drawn from the set of symbols in an M-ary constellation. The received signal is modeled as $\mathbf{y} = h\mathbf{s} + \mathbf{w}$, where $h \sim \mathcal{CN}(0, 1)$ denotes the complex Gaussian channel coefficient, considered constant across the n channel uses, under the flat fading time-invariant channel model with a single-input single-output link. The additive white Gaussian noise (AWGN) is represented as $\mathbf{w} = [w_1, \ldots, w_n]^{\mathsf{T}}$ where $w_i \sim \mathcal{CN}(0, \sigma^2)$, and the variance is given by $\sigma^2 = (2E_b/N_0)^{-1}$. Here, E_b/N_0 denotes the ratio of energy per bit to noise power spectral density. The system operates under practical conditions where perfect CSI is not available at the receiver. Instead, the receiver accesses CSIR $\hat{h}$, modeled as:

$$\hat{h} = \rho h + \sqrt{1 - \rho^2}\, \delta, \tag{1}$$

where ρ denotes the correlation coefficient ranging between 0 and 1, and $\delta \sim \mathcal{CN}(0, 1)$ is independent complex Gaussian noise. The decoder, based on attention-based sparse autoencoder structure, jointly processes the GAN-refined signal $\boldsymbol{y}_{\text{gen}}$ and the imperfect channel information $\hat{h}$ to learn a compact latent representation and reconstruct the transmitted message $\hat{d}$. The system's performance is measured using bit error rate (BER), representing the fraction of incorrectly decoded bits. BER is computed by converting d and $\hat{d}$ to their k-bit binary forms and measuring the bit-wise error rate.

2.1. GAN overview

To improve the overall signal reconstruction under channel uncertainty, GAN is integrated into an autoencoder-based framework as a learnable surrogate for the wireless channel. Fig. 1 illustrates the working principle of a conventional GAN, agnostic to the communication system model. It consists of two neural networks: a generator G and a discriminator D. The generator learns to produce samples that mimic the actual distribution of the received signals, while the discriminator attempts to distinguish between the actual data samples and those generated by G. Here, the term actual refers to the known or reference data used for training. During adversarial training, G is optimized to generate outputs that can fool the discriminator into classifying them as real, while D is trained to correctly identify actual versus generated (synthetic) signals. This adversarial learning mechanism is expressed through a minimax objective function [13] with the following value function:

$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{actual}}}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))], \tag{2}$$

where $\boldsymbol{x}$ corresponds to the actual data samples, which serve as the input to the discriminator. The term p_{actual} denotes the empirical probability

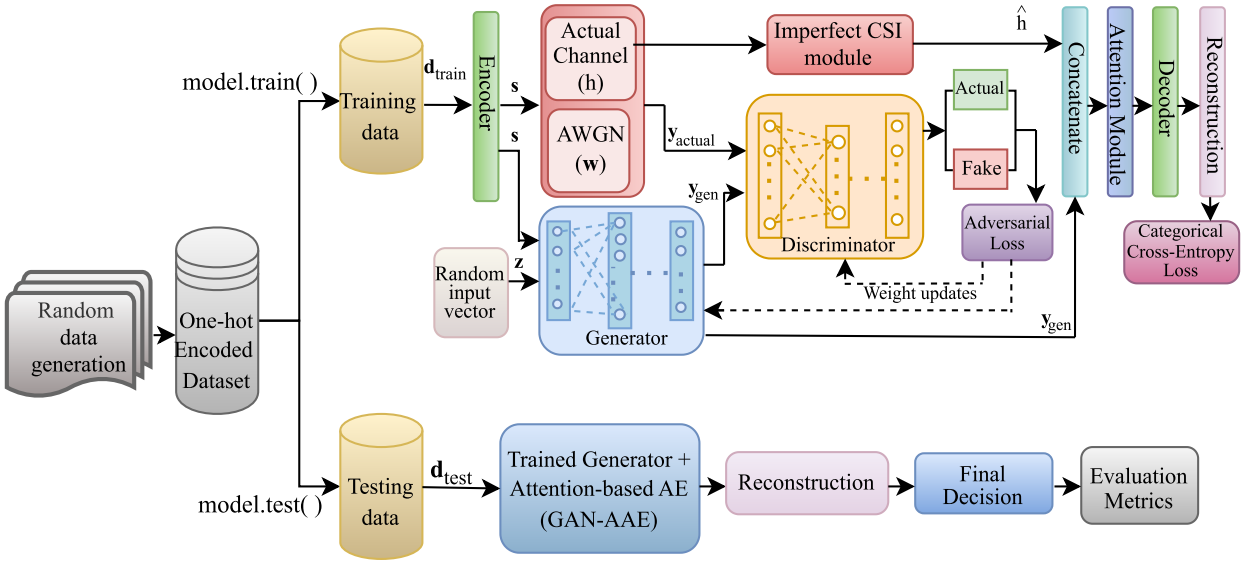

Fig. 2. Workflow of the proposed GAN-AAE model.

distribution of the actual samples. It describes how these samples are distributed within the input data space presented to the discriminator. The term $\boldsymbol{x} \sim p_{\text{actual}}$ denotes that the actual samples $\boldsymbol{x}$ are drawn from the distribution p_{actual}. Additionally, $\boldsymbol{z}$ is a random input vector that serves as the initial input to the generator. It is sampled from a predefined probability distribution $p(\boldsymbol{z})$, typically a standard normal or a uniform distribution. This distribution provides a structured but random input space from which the generator learns to produce realistic samples. The output of the generator is denoted by $G(\boldsymbol{z})$, which is a synthetic sample produced to resemble the actual data. The term $D(\boldsymbol{x})$ represents the output of the discriminator when it receives an actual sample $\boldsymbol{x}$. On the other hand, $D(G(\boldsymbol{z}))$ represents the output of the discriminator when it is presented with a generated sample $G(\boldsymbol{z})$. The term $\mathbb{E}_{\boldsymbol{x}\sim p_{\text{actual}}}[\log D(\boldsymbol{x})]$ represents the average log-probability that the discriminator correctly identifies actual data samples. Meanwhile $\mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$ represents the average log-probability that the discriminator correctly identifies generated samples. The training process follows a two-player minimax game defined by the objective function in Eq. (2). In this adversarial setup, the discriminator D aims to maximize the value of the objective function and the generator G attempts to minimize it. The discriminator is trained to maximize its ability to distinguish actual signals from generated ones using the following loss function:

$$\begin{aligned}\mathcal{L}_D = \quad & -\mathbb{E}_{\boldsymbol{x}\sim p_{\text{actual}}}[\log D(\boldsymbol{x})] \\ & -\mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))],\end{aligned} \tag{3}$$

where the first term encourages the discriminator to assign high probabilities to actual data samples, reinforcing its ability to correctly recognize actual inputs. The second term drives the discriminator to assign low probabilities to synthetic samples generated by the generator, thereby enhancing its capability to detect generated data. Conversely, the generator is trained using the loss:

$$\mathcal{L}_G = -\mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}[\log D(G(\boldsymbol{z}))], \tag{4}$$

which drives the generator to produce samples that the discriminator is more likely to classify as actual samples. Training continues in an alternating manner, first updating D, then G, until convergence is reached. At equilibrium, the generator produces signal representations so close to the actual distribution that the discriminator cannot reliably differentiate between actual and generated data.

In our proposed GAN-AAE framework, the conventional GAN architecture is adapted to align with the structure of an end-to-end wireless communication system. The discriminator D is provided with the actual received signal $\boldsymbol{y}_{\text{actual}}$ received over the wireless channel as input. This signal serves as the empirical reference data, analogous to the role of $\boldsymbol{x} \sim p_{\text{actual}}$ in the conventional GAN setup. The distribution of $\boldsymbol{y}_{\text{actual}}$ across training samples thus forms the counterpart of the empirical distribution p_{actual}. It refers to the empirical probability distribution of the received signal samples generated under the given channel conditions. It is not defined by a closed-form expression but is instead estimated from a large set of simulated training samples. The generator G is designed to take two inputs: the encoded signal $\boldsymbol{s}$, produced by the encoder network of the autoencoder framework, and a Gaussian random vector $\boldsymbol{z} \sim \mathcal{CN}(0, I)$. The encoded signal $\boldsymbol{s}$ represents a learned representation of the transmitted data, reflecting modulation or encoding features of the source message. The additional random input $\boldsymbol{z}$ introduces controlled stochasticity into the generation process, enhancing the diversity of generated samples. In this way, the pair $(\boldsymbol{s}, \boldsymbol{z})$ serves as the input to the generator, extending the conventional formulation where the generator input is solely a random vector. While $\boldsymbol{s}$ captures task-specific structure from the autoencoder, $\boldsymbol{z}$ contributes randomness similar to the role of $p(\boldsymbol{z})$ in traditional GANs, enabling the generator to model the variability inherent in wireless channels.

Fig. 2 shows the workflow of the proposed GAN-AAE model, which consists of two distinct phases: training and testing. During the training phase, the output of the generator is used for adversarial learning against the discriminator and for task-oriented learning through the decoder with cross-entropy loss. During testing, the discriminator is discarded, and only the trained generator is retained as a surrogate channel. The output of the trained generator is combined with the imperfect CSI estimate, processed through the attention module layer and decoder, and finally reconstructed for evaluation using metrics such as BER.

2.2. GAN-attention-based end-to-end sparse autoencoder (GAN-AAE) communication system model

As shown in Fig. 3, the proposed GAN-AAE communication system comprises an encoder (transmitter), a GAN-based channel modeling block, and an attention-augmented decoder (receiver), designed to operate under Rayleigh fading with imperfect CSIR. At the transmitter, the input message d is mapped to a latent space representation $\boldsymbol{s}$ through a series of fully connected (FC) layers, forming the encoded signal that is forwarded to both the channel and the GAN module. To strengthen the encoder representation, sparsity is introduced in the latent space so that only a small number of latent units become active for each

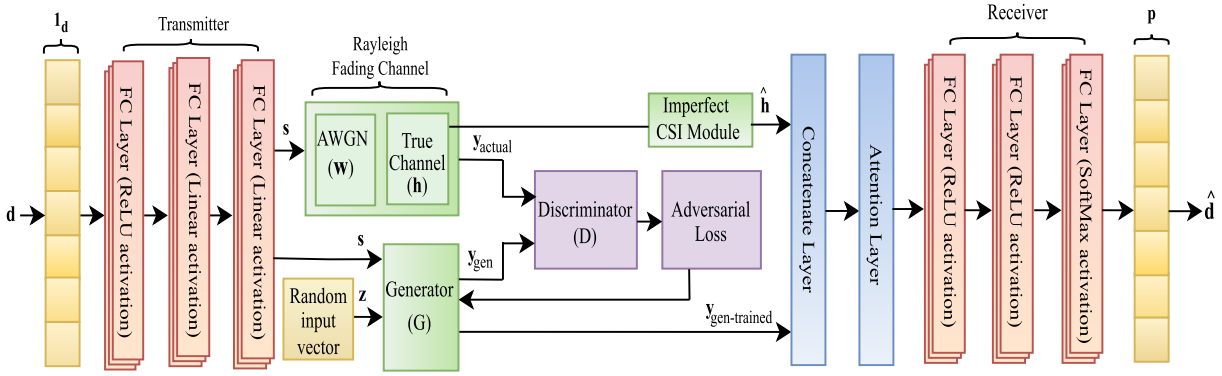

Fig. 3. Architecture of the proposed GAN-attention-based end-to-end sparse autoencoder communication system with imperfect CSIR.

transmitted message. This reduces redundant feature activations and results in a more compact and structured latent encoding, which improves generalization under varying channel conditions. The actual channel, composed of a Rayleigh fading path and AWGN, transforms s into the actual received signal $\boldsymbol{y}_{\text{actual}} = h\boldsymbol{s} + \boldsymbol{w}$. The generator G in the GAN takes the encoded signal $\boldsymbol{s} \in \mathbb{C}^n$ and an additional random vector $\boldsymbol{z} \sim \mathcal{CN}(0, I)$ of length $2n$ as its two inputs and learns to produce a synthetic received signal $\boldsymbol{y}_{\text{gen}} = G(\boldsymbol{s}, \boldsymbol{z})$, which closely mimics the distribution of the actual received signal $\boldsymbol{y}_{\text{actual}}$. The discriminator D receives both the actual signal $\boldsymbol{y}_{\text{actual}}$ and the generated signal $\boldsymbol{y}_{\text{gen}}$, and is trained to distinguish between them. Adversarial training encourages the generator to improve the quality of its output by minimizing the adversarial loss, enabling it to learn the underlying channel characteristics of the transmitted signal. Once training reaches the Nash equilibrium i.e., the point at which the discriminator can no longer reliably distinguish between actual and generated signals, the generator is considered fully trained and produces $\boldsymbol{y}_{\text{gen-trained}}$, a signal that closely approximates the actual channel output. This refined signal is then forwarded to the concatenation layer, where it is combined with the output of a separate imperfect CSI module $\hat{h}$. Thus, the system benefits from the ability of GAN to learn and replicate complex fading and noise characteristics in a data-driven manner. To enhance the decoder's ability to process distorted and uncertain inputs, an attention layer is applied to the concatenated vector of $\boldsymbol{y}_{\text{gen-trained}}$ and $\hat{h}$. The proposed model employs a dot-product self-attention mechanism operating in the latent feature domain. This attention mechanism computes relevance scores between latent features by taking the dot product of the concatenated feature vector with its transpose, followed by softmax normalization to obtain attention weights. Unlike spatial attention used in vision tasks, this self-attention operates on abstract latent representations and enables adaptive feature re-weighting rather than spatial localization. The attention mechanism dynamically selects the most relevant features for reconstruction, allowing the network to emphasize signal components that are more informative and reliable under fading and imperfect CSI conditions. The selected features pass through a sequence of fully connected layers within the decoder, ultimately producing the reconstructed output $\hat{d}$. This structure not only improves symbol reconstruction accuracy but also increases robustness against channel impairments by leveraging adversarial learning and latent-domain self-attention-based feature prioritization.

By integrating sparse AE, GAN, and attention mechanism within a single learning pipeline, the proposed framework enables the generator-refined signal, the imperfect CSI, and the attention-weighted latent representation to jointly influence the decoding stage. During training, the adversarial refinement and feature prioritization occur in the same forward path, meaning that the improved channel-like samples produced by the generator directly affect the representation that the attention mechanism selects for decoding. This leads to a closed and mutually reinforced interaction where the generator continuously adapts to produce more realistic channel outputs, the attention layer dynamically adjusts the contribution of the refined features based on CSIR reliability, and the decoder maps these selectively emphasized features to the symbol decision space. Thus, in the proposed hybrid framework, the sparse latent representation provides a more stable input distribution for the generator to refine and for the attention mechanism to selectively re-weight. The interaction between autoencoder-based representations and adversarial learning has been noted to enhance robustness and latent-space consistency in wireless communication frameworks [25]. This unified interaction allows the system to form channel-aware symbol representations that better reflect the actual channel behavior under imperfect CSI, thereby improving reconstruction performance in a jointly learned manner. Algorithm 1 presents the end-to-end training procedure for the proposed GAN-AAE communication model under Rayleigh fading with imperfect CSIR. It outlines the coordinated update of four key neural network components: an encoder denoted by f with parameters $\boldsymbol{\theta}_f$, a decoder denoted by g with parameters $\boldsymbol{\theta}_g$, a generator G with parameters $\boldsymbol{\theta}_G$, and a discriminator D with parameters $\boldsymbol{\theta}_D$. Here $\boldsymbol{\theta}$ represents the set of trainable parameters (weights and biases). During each epoch, the dataset $\boldsymbol{\mathcal{X}}$ is shuffled and split into batches $\boldsymbol{\mathcal{X}}_{\text{batch}}$. Each batch is encoded by f to produce latent representations $\boldsymbol{q}_{\text{batch}}$. These representations are passed through a simulated Rayleigh fading channel to obtain the actual received signal $\boldsymbol{y}_{\text{actual}}$, and simultaneously passed to the generator, along with a batch of random complex vectors $\boldsymbol{z}_{\text{batch}}$, to produce $\boldsymbol{y}_{\text{gen}} = G(\boldsymbol{q}_{\text{batch}}, \boldsymbol{z}_{\text{batch}})$, a synthetic version of the received signal. The discriminator D is trained to distinguish $\boldsymbol{y}_{\text{actual}}$ from $\boldsymbol{y}_{\text{gen}}$ by minimizing the loss L_D, while the generator G is trained adversarially to fool D by producing indistinguishable signals from the actual channel output. Both components are updated using gradient descent [13] on their respective loss functions. Once adversarial feedback is incorporated, the channel estimate $\hat{h}$ is expanded to an n-length vector to match its dimension with the received signal $\mathbf{y}$ before concatenation. Then the refined output $\boldsymbol{y}_{\text{gen}}$ is concatenated with $\hat{\boldsymbol{h}}$ to form the combined representation $\mathbf{c}$ and passed through an attention mechanism. This mechanism computes attention score matrix $\mathbf{A}$ by taking the dot product between the concatenated vector and its transpose, followed by applying the softmax function to normalize these scores. The attention scores dynamically weight the importance of different features within $\mathbf{c}$, effectively enabling the model to prioritize more informative components of the input. Subsequently, the context vector $\mathbf{v}$ is calculated as a weighted sum of $\mathbf{c}$ using the attention matrix $\mathbf{A}$, providing a refined representation that captures the most relevant information for decoding. The context vector $\boldsymbol{v}$ resulting from this process is decoded by g to produce the reconstructed batch $\hat{\boldsymbol{\mathcal{X}}}_{\text{batch}}$. The decoder's performance is measured by the reconstruction loss L_{AE}. Simultaneously, $L1$ regularization is applied to the encoder weights as $L_{\text{reg}} = \lambda \sum_i |\theta_f^i|$, and the total loss for encoder

Algorithm 1 GAN-AAE model training under imperfect CSIR.

Input: Learning rate α, epochs N_{epoch}, batch size N_{batch}, $L1$ regularization weight λ, training dataset $\mathcal{X}$.
Initialize: Randomly initialize encoder θ_f, decoder θ_g, generator θ_G, discriminator θ_D. Set epoch = 0.

for epoch = 1 to N_{epoch} **do**
  Shuffle training dataset $\mathcal{X}$
  for each batch $\mathcal{X}_{\text{batch}} \in \mathcal{X}$ **do**
    Autoencoder Forward Pass:
    Encode input: $\boldsymbol{q}_{\text{batch}} = f(\mathcal{X}_{\text{batch}}; \theta_f)$
    Sample $\boldsymbol{z}_{\text{batch}} \sim \mathcal{CN}(0, \boldsymbol{I})$ for current batch
    Generate channel-like output:
      $\boldsymbol{y}_{\text{gen}} = G(\boldsymbol{q}_{\text{batch}}, \boldsymbol{z}_{\text{batch}}; \theta_G)$
    Obtain true channel output:
      $\boldsymbol{y}_{\text{actual}} = h \cdot \boldsymbol{q}_{\text{batch}} + \boldsymbol{w}$
    Discriminator Update:
    • Compute loss:
      $\mathcal{L}_D = -\mathbb{E}[\log D(\boldsymbol{y}_{\text{actual}})] - \mathbb{E}[\log(1 - D(\boldsymbol{y}_{\text{gen}}))]$
    • Update discriminator parameters:
      $\theta_D \leftarrow \theta_D - \alpha \nabla_{\theta_D} \mathcal{L}_D$
    Generator Update:
    • Compute adversarial loss:
      $\mathcal{L}_G = -\mathbb{E}[\log D(\boldsymbol{y}_{\text{gen}})]$
    • Update generator parameters:
      $\theta_G \leftarrow \theta_G - \alpha \nabla_{\theta_G} \mathcal{L}_G$
    Concatenate: $\boldsymbol{c} = \text{Concat}(\boldsymbol{y}_{\text{gen}}, \hat{\boldsymbol{h}})$
    Compute attention:
      $\boldsymbol{A} = \text{Softmax}(\boldsymbol{c} \cdot \boldsymbol{c}^T)$
    Compute context vector: $\boldsymbol{v} = \boldsymbol{A}\boldsymbol{c}$
    Decode output: $\hat{\mathcal{X}}_{\text{batch}} = g(\boldsymbol{v}; \theta_g)$
    Compute AE loss:
      $\mathcal{L}_{\text{AE}} = L(\mathcal{X}_{\text{batch}}, \hat{\mathcal{X}}_{\text{batch}})$
    Apply $L1$ regularization:
      $\mathcal{L}_{\text{reg}} = \lambda \sum_i |\theta_f^i|$
    Total AE loss: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{AE}} + \mathcal{L}_{\text{reg}}$
    Update encoder and decoder:
    • $\theta_f \leftarrow \theta_f - \alpha \nabla_{\theta_f} \mathcal{L}_{\text{total}}$
    • $\theta_g \leftarrow \theta_g - \alpha \nabla_{\theta_g} \mathcal{L}_{\text{AE}}$
  end for
end for
Output: Trained parameters $\theta_f, \theta_g, \theta_G, \theta_D$

update becomes $L_{\text{total}} = L_{\text{AE}} + L_{\text{reg}}$. Once all epochs are completed, the trained parameters θ_f and θ_g are updated accordingly. Although the algorithm operates using the intermediate generator output $\boldsymbol{y}_{\text{gen}}$ during training, this signal progressively improves as the adversarial process continues. Once training converges, the generator is considered fully trained, and its output is denoted $\boldsymbol{y}_{\text{gen-trained}}$. This refined signal is used as the final input to the decoder during inference or deployment. The overall training framework enables the model to effectively learn and compensate for channel distortions, thereby improving reconstruction performance in the presence of imperfect CSI.

3. Performance evaluation

This section describes the simulation setup and performance evaluation of the proposed GAN-AAE end-to-end communication system model.

To assess the effectiveness of the proposed model, a dataset comprising 20,000 independently and identically distributed (i.i.d.) source symbols is generated. These symbols are uniformly drawn from the defined message set according to the modulation configuration adopted in the system. To ensure fair and statistically valid assessment, separate datasets are used for training, validation, and testing, each subjected to independent noise and channel realizations. Specifically, 20% of the training dataset is held out as a validation set to continuously monitor the model's generalization ability and prevent overfitting during training. The BER is employed as the performance metric for system evaluation. Simulations are conducted across a range of average SNR values to evaluate the proposed model's robustness and decoding performance under varying channel conditions.

Table 2
Layer-wise structure of the proposed GAN-AAE model.

Layer	Output Dimensions
Input	M
Encoder Dense + ReLU	M
Encoder Dense + Linear	M
Encoder Dense + Linear	n
Generator Dense + ReLU	n
Generator Dense + Linear	n
Refined Signal y_{gen}	n
Imperfect CSI Module $\hat{h}$	n
Concatenation Layer	$2n$
Attention Layer	$2n$
Decoder Dense + ReLU	M
Decoder Dense + ReLU	M
Decoder Dense + Softmax	M

Table 3
Hyperparameters used for training the GAN-AAE model.

Hyperparameter	Value
Learning rate	0.001
Batch size	32
Number of epochs	150
L1 regularization strength	0.01
Optimizer	Adam
Loss function (AE)	Categorical crossentropy
Adversarial loss function	Binary crossentropy
Gradient update mode	Alternating G and D

Table 2 presents the layer-wise structure of the proposed GAN-AAE model, highlighting the dimensions at each stage of the end-to-end communication pipeline. The generator, although shown as a single module in Fig. 2, is implemented as a lightweight feedforward neural network with two dense layers. It first applies a nonlinear transformation using a ReLU-activated dense layer, followed by a linear projection layer that outputs the refined signal y_{gen}. This structure enables the generator to learn nonlinear mappings from encoded symbols to realistic channel-distorted signals. The discriminator used during adversarial training is also implemented as a lightweight two-layer feedforward network (dense-ReLU followed by dense-sigmoid). Since the discriminator is active only during the training stage to guide the adversarial learning process and plays no role in inference, it is not included in Table 2 and does not affect the final signal reconstruction pipeline. Here, M denotes the number of possible messages, which depends on the modulation scheme used. For a BPSK system, $M = 2$. Each message corresponds to a k-bit binary vector. The parameter n represents the number of discrete channel uses per message, where $n = k = \log_2(M)$. Additionally, while the encoded and received signals are inherently complex-valued, the neural network layers operate on real-valued tensors. Therefore, complex vectors are internally represented as real-valued vectors of twice the original dimension by concatenating their real and imaginary parts.

Table 3 summarizes the hyperparameters used for training the proposed GAN-AAE model. A learning rate of 0.001 is used across all neural network components to ensure stable convergence, especially important in adversarial training scenarios. The training is performed over 150 epochs with a mini-batch size of 32, using the Adam optimizer [13] due to its adaptive learning capability and widespread effectiveness in

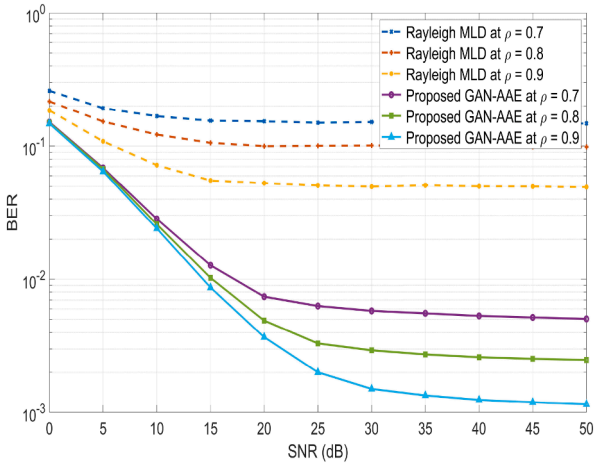

Fig. 4. BER vs average SNR performance comparison of the proposed GAN-AAE and theoretical MLD model over a BPSK Rayleigh fading channel for different values of ρ.

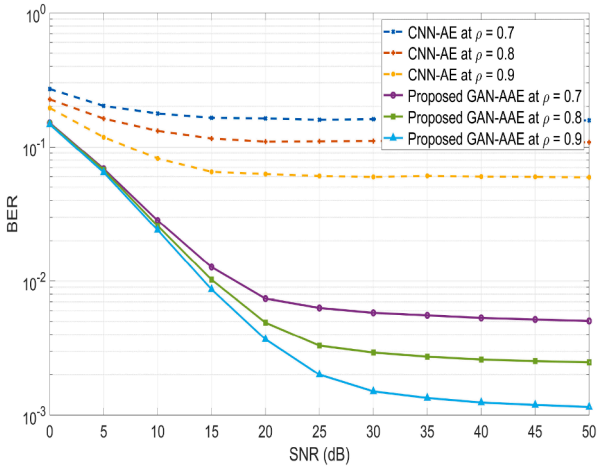

Fig. 5. BER versus average SNR performance comparison of the proposed GAN-AAE and CNN-AE model over a BPSK Rayleigh fading channel for different values of ρ.

training deep models. All the layers use the default Xavier/He initialization [13]. The autoencoder component is trained using categorical cross-entropy [13] loss, which is suitable for multi-class classification of source symbols. $L1$ regularization is applied to the encoder weights to promote sparsity in the latent space, improving generalization under varying channel conditions and preventing overfitting. For the adversarial component, binary cross-entropy is employed as the loss function for both the generator and discriminator. This loss formulation aligns with standard GAN training. During each training iteration, one update step is performed for both the generator and the discriminator in an alternating manner. These hyperparameter choices strike a balance between training stability and model performance in the presence of Rayleigh fading and imperfect CSI.

Fig. 4 presents the BER performance of the proposed GAN-AAE model under Rayleigh fading with imperfect CSIR. The curves correspond to three different correlation values: 0.7, 0.8, and 0.9, representing increasing levels of CSI accuracy. For $\rho = 0.9$, the GAN-AAE achieves a BER close to 10^{-3}, demonstrating its robustness to mild channel uncertainty. The results are compared with the BER performance derived from a maximum likelihood detection (MLD) scheme under the same imperfect CSI assumption. At 10 dB SNR and $\rho = 0.9$, the MLD model attains a BER of 0.072 as compared to 0.02404 BER of the proposed model. As the channel becomes less correlated (lower ρ), the BER increases, but the proposed model still maintains a clear advantage over the MLD, indicating its ability to learn and adapt to signal distortions caused by imperfect CSI. The superior performance of the proposed GAN-AAE model over the conventional MLD can be attributed to its ability to learn a detection strategy that is directly matched to the joint channel-CSI statistics encountered during training. Unlike MLD, which relies strictly on the analytical detection rule under imperfect CSI, the GAN-AAE is trained on a broad distribution of channel realizations generated by the GAN, enabling it to learn the joint statistical behaviour of the fading channel and its CSI distortions more comprehensively. Furthermore, the sparse autoencoder structure promotes compact and noise-resilient latent features, while the attention module selectively emphasizes the most informative components induced by the channel. These combined features allow the GAN-AAE model to learn more robust decision regions under imperfect CSI.

At higher SNRs, both models exhibit a characteristic flattening of the BER curves, indicating the presence of an error floor. The BER curves of the MLD model exhibit a noticeable error floor beyond an SNR of around 20 dB. In contrast, the proposed GAN-AAE system significantly outperforms the MLD baseline across all ρ values by exhibiting the error floor beyond 40 dB SNR. In the conventional Rayleigh fading system with imperfect CSIR, the high-SNR error floor is primarily caused by the non-vanishing residual channel estimation error. Since the CSI mismatch remains even when the noise level becomes negligible, this residual distortion dominates the detection process at high SNR, preventing the BER from approaching zero. In the deep learning-based proposed system, several factors jointly contribute to the high-SNR error floor. One contributing factor is that the decoder's learned decision boundaries remain imperfect for channel conditions that lie outside the dominant patterns captured during training. Another factor is the latent-space mismatch caused by imperfect CSI, which remains even when the noise level becomes negligible. These influences can lead to a small but non-zero BER floor, even though the proposed model achieves improved BER performance over a wide SNR range. Despite the presence of a residual error floor, the proposed GAN-AAE model manages to shift this floor to higher SNR levels. This behaviour can be attributed to complementary factors such as the end-to-end learning process, which enables coordinated shaping of the encoder, latent representation, and decoder to form more resilient decision boundaries for the dominant channel statistics. The GAN-based refinement helps the model better approximate typical channel-induced distortions, reducing the effect of moderate CSI mismatch. The attention mechanism further strengthens robustness by selectively emphasizing the most reliable latent features under imperfect CSI. Together, these factors allow the proposed model to maintain improved BER performance over a wider SNR range before the error floor is reached.

Fig. 5 illustrates the BER performance comparison between the proposed GAN-AAE model and baseline CNN-AE based end-to-end communication system [30], under the same channel conditions. The CNN-based approach exhibits limited adaptability to channel imperfections. Across all ρ values, the BER of the proposed GAN-AAE saturates at higher SNR compared to the CNN-AE model. While CNNs are effective at local feature extraction, their performance degrades when dealing with varying and uncertain channel conditions without dedicated mechanisms to capture such variations. In contrast, the proposed GAN-AAE system significantly outperforms the CNN baseline across all correlation levels. At 10 dB SNR and $\rho = 0.9$, the CNN model attains a BER of 0.092 as compared to 0.02404 BER of the proposed model. Even for challenging scenarios like $\rho = 0.7$, the GAN-AAE maintains a substantially lower BER than its CNN counterpart, demonstrating its ability to learn more accurate channel-compensated representations.

Fig. 6 compares the proposed GAN-AAE with the baseline DNN-Channel AE model [29] under the same Rayleigh fading and imperfect CSIR conditions for $\rho = 0.7, 0.8, 0.9$. The trainable channel surrogate of the DNN-Channel AE model shows robustness to CSI imperfections as

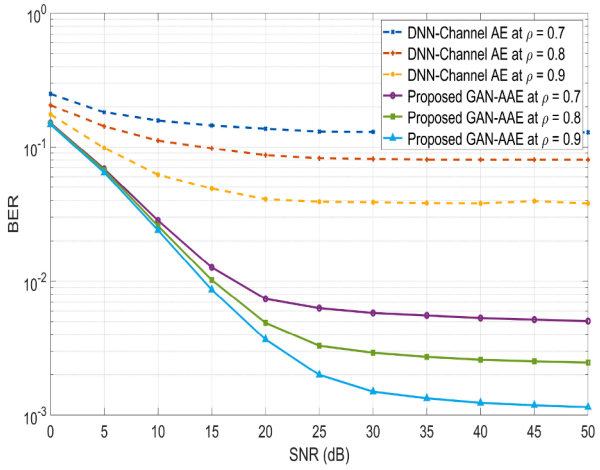

Fig. 6. BER versus average SNR performance comparison of the proposed GAN-AAE and DNN-channel AE model over a BPSK Rayleigh fading channel for different values of ρ.

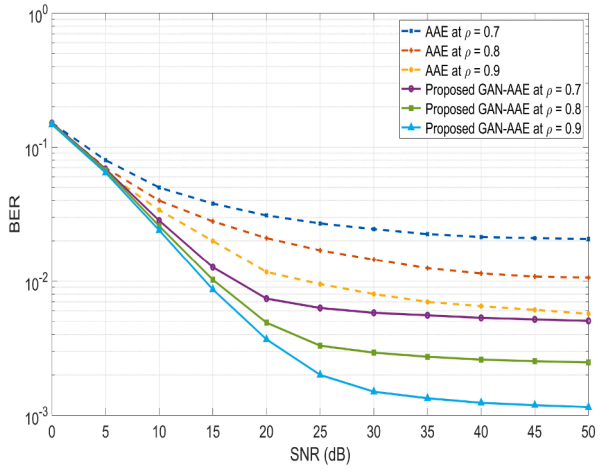

Fig. 7. BER vs average SNR performance comparison of the AAE model and proposed GAN-AAE model over a BPSK rayleigh fading channel for different values of ρ.

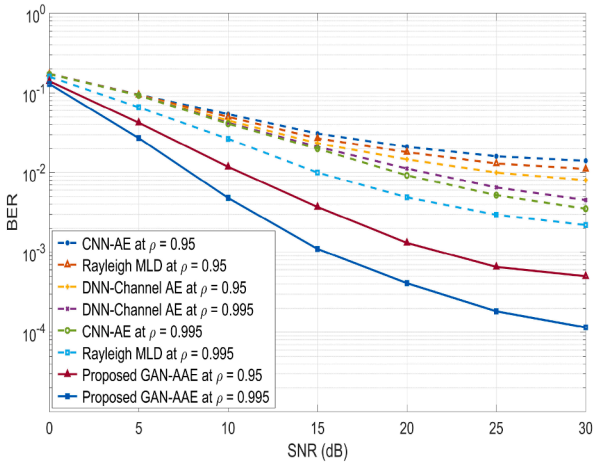

Fig. 8. BER vs average SNR performance comparison of different BPSK rayleigh fading models for $\rho = 0.95$ and 0.995.

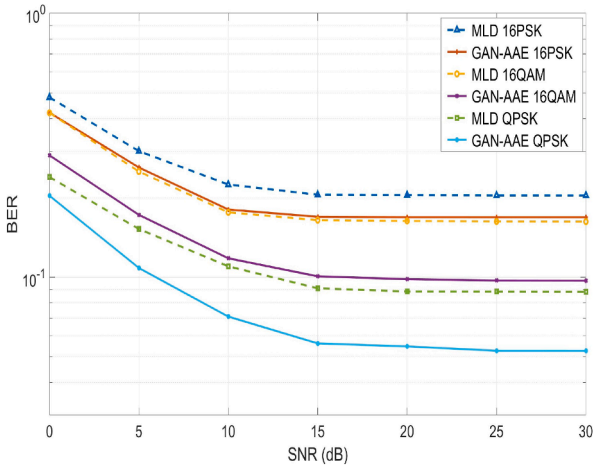

Fig. 9. BER versus average SNR performance comparison of the proposed GAN-AAE and MLD model over Rayleigh fading with imperfect CSIR ($\rho = 0.9$) for QPSK, 16-PSK, and 16-QAM.

compared to the MLD and CNN-AE models, but it is still outperformed by the proposed GAN-AAE. The DNN-Channel AE exhibits its error floor at a higher SNR compared to the CNN-AE. However, this behaviour still occurs substantially earlier than in the proposed GAN-AAE. The proposed model consistently achieves lower BER than the DNN-Channel AE model over the entire SNR range and for all ρ. The gains are most evident at higher SNRs, where the DNN-Channel AE saturates while the GAN-AAE continues to improve, reflecting the benefits of adversarial channel refinement together with attention-based feature selection. At 10 dB SNR and $\rho = 0.9$, the DNN-Channel AE attains a BER of 0.06223, whereas the proposed model achieves 0.02404 BER.

Fig. 7 compares the BER performance of the proposed GAN-AAE model against a baseline autoencoder system, referred to as AAE, under Rayleigh fading with imperfect CSIR. In the AAE configuration, the GAN module is removed from the GAN-AAE architecture, while all other architectural and training components remain identical, including the attention mechanism, encoder-decoder structure, and imperfect CSI modeling. This allows for a controlled comparison to isolate the impact of the GAN component on system performance. As shown in the figure, for all values of ρ, the GAN-AAE consistently outperforms the AAE. The adversarial training process of the GAN provides a more robust latent representation that helps the decoder better separate signal components from channel impairments under imperfect CSI, enabling improved signal reconstruction. The GAN-AAE achieves lower BER across the entire SNR range, with performance gaps widening at higher SNRs. At 10 dB SNR and $\rho = 0.9$, the AAE model attains a BER of 0.034, whereas the proposed model achieves 0.02404 BER. The AAE model exhibits an error floor beginning around 40 dB, whereas the GAN-AAE continues to improve and reaches the error floor beyond 50 dB. These observations highlight the ability of GAN to learn and replicate channel distortions more effectively, thereby enhancing signal reconstruction quality in the presence of imperfect CSI.

Fig. 8 illustrates the relative BER trends of the benchmark CNN-AE, DNN-Channel AE, and Rayleigh MLD models against the proposed GAN-AAE under higher correlation levels of 0.95 and 0.995. These ρ values reflect realistic, high-quality CSI scenarios typically achieved in practical systems. For both ρ values, the proposed GAN-AAE maintains a clear BER advantage across the entire SNR range. The performance behaviour indicates that the benchmark schemes show limited improvement even as SNR increases, whereas the proposed model continues to benefit from higher SNR, giving lower BER for both correlation conditions.

Fig. 9 compares the BER of the proposed GAN-AAE model with the conventional MLD over a Rayleigh channel under imperfect CSIR ($\rho = 0.9$) for QPSK, 16-PSK, and 16-QAM. Across the SNR range, the GAN-AAE achieves lower BER than MLD for all the modulation schemes.

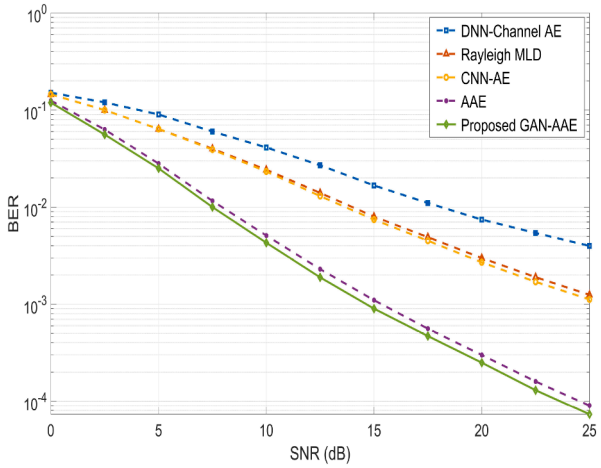

Fig. 10. BER versus average SNR comparison of different BPSK rayleigh fading models under perfect CSI ($\rho = 1$) condition.

The relative gain, however, decreases with increasing constellation order, since denser constellations have smaller decision regions and are more sensitive to errors. Hence, the achievable BER improvement is smaller in higher order modulations.

Fig. 10 compares the proposed GAN-AAE with all the baseline models under perfect CSIR i.e. $\rho = 1$. The DNN-Channel AE uses an extra approximation through the surrogate in the deep neural network-based channel. Consequently, it fails to exploit the exact channel mapping in the perfect CSI environment and exhibits the highest BER across the SNR range compared with the other models. The two attention-based models: AAE and the proposed GAN-AAE, achieve the lowest BERs. The proposed model shows marginally better performance than AAE, reflecting that when CSI is perfect, the GAN's channel-refinement brings limited additional benefit beyond the attention-augmented autoencoder.

To assess the computational efficiency of the proposed GAN-AAE model for BPSK Rayleigh fading channel with imperfect CSI, Table 4 shows a forward-pass complexity analysis in terms of floating point operations (FLOPs) [31]. For fully connected and convolutional layers, the FLOPs are estimated using FLOPs $= 2 \times I \times O$, where I and O are the input and output dimensions of the layer, respectively. For the baseline CNN-AE model, the total FLOPs are 1280 FLOPs per input sample, computed assuming consistent system parameters with the GAN-AAE model. For the DNN-Channel AE model, the total FLOPs are 1542 FLOPs per input sample. Although both the baseline models use convolutional layers, the DNN-Channel block employs much larger channel widths, resulting in a higher operation cost. In contrast, the proposed GAN-AAE model includes an attention-based sparse autoencoder and an additional generator component. The attention-based autoencoder (AAE), excluding the GAN component, contributes 6092 FLOPs while the proposed GAN-AAE model yields a total of 12,248 FLOPs per input sample. These FLOP values correspond to the processing cost of a single input sample. During mini-batch training or inference, this cost scales linearly with the batch size. While the proposed GAN-AAE model introduces a higher computational complexity compared to baseline systems, this increase is accompanied by substantial improvements in system performance, particularly under conditions of channel uncertainty and imperfect CSI. The additional complexity stems from the integration of the attention mechanism and the generative component, both of which contribute to enhanced feature extraction and signal reconstruction capabilities. For the AAE model, the total FLOPs reduce to 6,092, isolating the overhead introduced by the GAN component. This confirms that the majority of the computational cost resides in the encoder-decoder and attention modules. Although the AAE is marginally more efficient, it exhibits significantly lower BER performance and an earlier error floor under high-SNR conditions. This highlights that the modest increase in complexity introduced by the GAN is justified by its substantial contribution to improved robustness against channel uncertainty.

Table 4
Forward-pass complexity comparison.

Model	Total FLOPs
Baseline CNN-AE	1280
Baseline DNN-CHannel AE	1542
AAE (without GAN)	6092
Proposed GAN-AAE	12,248

This trade-off between complexity and performance highlights a common design consideration in modern communication systems, where increased computational cost can be justified by the resulting gains in robustness, adaptability, and decoding accuracy. For applications demanding high reliability in challenging wireless environments, the performance benefits of the GAN-AAE model make it a compelling alternative despite its elevated complexity.

4. Conclusion

This work presented a GAN-integrated Attention-based Sparse Autoencoder (GAN-AAE) framework for end-to-end wireless communication over Rayleigh fading channels with imperfect CSI. The sparse latent representation provides a stable feature space, the generator refines channel-like distortions based on learned statistics, and the attention module selectively emphasizes reliable latent features according to CSI quality. This joint refinement process enables the decoder to operate on channel-aware and noise-resilient representations, improving symbol reconstruction under CSI mismatch. The simulation results show that the GAN-AAE outperforms the conventional MLD model by learning decision regions that are adapted to the joint channel-CSI statistics encountered during training. The proposed model also surpasses baseline DNN and CNN-based systems for M-PSK and M-QAM modulation schemes across different ρ values, demonstrating robustness under varying levels of CSI uncertainty. Furthermore, the proposed model effectively shifts the error floor to a higher SNR region compared to the baseline models. This improvement is driven by the effects of GAN-based channel refinement, end-to-end learning, and attention-guided feature selection. Additionally, a detailed complexity analysis revealed that although the GAN-AAE introduces higher computational overhead, its improved BER performance justifies the trade-off. These results validate the effectiveness of integrating generative modeling and attention mechanisms into an end-to-end autoencoder for robust deep learning-based communication system design under imperfect CSI.

Future research can explore extending the GAN-AAE framework to MIMO systems and more complex channel conditions such as time-varying or frequency-selective fading. Incorporating memory-aware mechanisms can enhance the model's ability to capture temporal dependencies and improve robustness. Additionally, incorporating adaptive or multi-head attention could allow the decoder to selectively focus on relevant features based on CSI reliability or channel variability. These extensions would broaden the applicability of the GAN-AAE model to more complex and heterogeneous wireless environments.

CRediT authorship contribution statement

Safalata S. Sindal: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Data curation; **Y. N. Trivedi:** Writing – review & editing, Validation, Supervision, Project administration.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Declaration of competing interest

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

References

[1] J.G. Proakis, M. Salehi, Digital Commun. 4 (2001) 593–620.
[2] T.J. O'shea, K. Karra, T.C. Clancy, Learning to communicate: channel auto-encoders, domain specific regularizers, and attention, in: IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2016, pp. 223–228.
[3] J.M. Kang, C.J. Chun, I.M. Kim, Deep learning based channel estimation for MIMO systems with received SNR feedback, IEEE Access 8 (2020) 121162–121181.
[4] H. He, M. Zhang, S. Jin, C.K. Wen, G.Y. Li, Model-driven deep learning for massive MU-MIMO with finite-alphabet precoding, IEEE Commun. Lett. 24 (10) (2020) 2216–2220.
[5] L.V. Nguyen, N.T. Nguyen, N.H. Tran, M. Juntti, A.L. Swindlehurst, D.H. Nguyen, Leveraging deep neural networks for massive MIMO data detection, 30, IEEE Wireless Communications, 2022.
[6] J. Zhao, Y. Wu, Q. Zhang, J. Liao, Two-stage channel estimation for mmWave massive MIMO systems based on ResNet-UNet, IEEE Syst. J. 17 (3) (2023) 4291–4300.
[7] T. O'shea, J. Hoydis, An introduction to deep learning for the physical layer, IEEE Trans. Cognit. Commun. Netw 3 (4) (2017) 563–575.
[8] S. Dörner, S. Cammerer, J. Hoydis, Brink S Ten, Deep learning based communication over the air, IEEE J. Sel. Top. Signal Process 12 (1) (2017) 132–143.
[9] M. Lu, B. Zhou, Z. Bu, Attention-empowered residual autoencoder for end-to-end communication systems, IEEE Commun. Lett. 27 (4) (2023) 1140–1144.
[10] Y. Liang, C.T. Lam, B.K. Ng, Compression algorithm for end-to-end communication using CNN, in: th International Conference on Computer and Communications (ICCC), 2021, pp. 318–323.
[11] S. Cammerer, Aoudia F. Ait, S. Dörner, M. Stark, J. Hoydis, Brink S. Ten, Trainable communication systems: concepts and prototype, IEEE Trans. Commun. 68 (9) (2020) 5489–5503.
[12] A. Goldsmith, Wireless Communications, Cambridge university press, 2005.
[13] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep Learning, 1 of *Cambridge*, MIT press, 2016.
[14] H. Zhang, L. Zhang, Y. Jiang, Overfitting and underfitting analysis for deep learning based end-to-end communication systems, 11th international conference on wireless communications and signal processing (WCSP), 2019.
[15] C.S.N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, R. Wang, Development and application of a deep learning-based sparse autoencoder framework for structural damage identification, Struct. Health Monit. 18 (1) (2019) 103–122.
[16] S.S. Sindal, Y.N. Trivedi, Enhancing performance of end-to-end communication system using attention mechanism-based sparse autoencoder over rayleigh fading channel, physical communication, 67, 2024.
[17] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A.A. Bharath, Generative adversarial networks: an overview, IEEE Signal Process. Mag. 35 (1) (2018) 53–65.
[18] J. Zhao, H. Mu, Q. Zhang, H. Zhang, Resnet-wgan-based end-to-end learning for iov communication with unknown channels, IEEE Internet Things J. 10 (19) (2023) 17184–17192.
[19] Y. Yang, Y. Li, W. Zhang, F. Qin, P. Zhu, C.X. Wang, Generative-adversarial-network-based wireless channel modeling: challenges and opportunities, IEEE Commun. Mag. 57 (3) (2019) 22–27.
[20] H. Phan, Nguyen H. Le, O.Y. Chén, P. Koch, N.Q. Duong, I. Mcloughlin, A. Mertins, Self-attention generative adversarial network for speech enhancement, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2021, pp. 7103–7107.
[21] Z. Tang, M. Tao, J. Su, Y. Gong, Y. Fan, T. Li, Data augmentation for signal modulation classification using generative adverse network, 4th international conference on electronic information and communication technology (ICEICT), 2021.
[22] Y. Hu, M. Yin, W. Xia, S. Rangan, M. Mezzavilla, Multi-frequency channel modeling for millimeter wave and THz wireless communication via generative adversarial networks, in: 56th Asilomar Conference on Signals, Systems, and Computers, 2022, pp. 670–676.
[23] M.U. Lokumarambage, V.S.S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, A. Fernando, Wireless end-to-end image transmission system using semantic communications, IEEE Access 11 (2023) 37149–37163.
[24] D. Han, W. Na, Enhancing wireless data transmission: a GAN-based approach for time series data restoration, in: International Conference on Information Networking (ICOIN), 2024, pp. 429–433.
[25] C. Zou, F. Yang, J. Song, Z. Han, Generative adversarial network for wireless communication: principle, application, and trends, IEEE Commun. Mag. 62 (5) (2023) 58–64.
[26] A. Manikandan, R.T. Babu, S.J. Ganesh, T. Sanjay, Generative adversarial networks based adaptive modulation and coding for next-generation 5G communication systems, Discover Appl. Sci. 7 (2) (2025) 91.
[27] S. He, L. Zhu, C. Yao, L. Wang, Z. Qin, A novel approach based on generative adversarial network for signal enhancement in wireless communications, Wireless Commun. Mobile Comput. 1 (2022) 8008460.
[28] W. Xie, M. Xiong, Z. Yang, W. Liu, L. Fan, J. Zou, Real and fake channel: GAN-based wireless channel modeling and generating, physical communication, 61, 2023.
[29] Y. An, S. Wang, L. Zhao, Z. Ji, I. Ganchev, A learning-based end-to-end wireless communication system utilizing a deep neural network channel module, IEEE Access 11 (2023) 17441–17453.
[30] N. Wu, X. Wang, B. Lin, K. Zhang, A CNN-based end-to-end learning framework toward intelligent communication systems, 2019, pp. 110197–110204.
[31] E. Mizutani, S.E. Dreyfus, On complexity analysis of supervised MLP-learning for algorithmic comparisons, in: International Joint Conference on Neural Networks, Proceedings, 2001, pp. 347–352. Cat. No 01CH37222.

Safalata S. Sindal is currently pursuing a Ph.D. from the Institute of Technology, Nirma University, Ahmedabad, Gujarat, India. Her area of specialization is wireless communication networks, deep learning, end-to-end communication system models, and signal processing for communications.

Y.N. Trivedi is a professor in the Electronics and Communication Engineering Department at Nirma University, Ahmedabad, Gujarat, India. He received his BE degree in Electronics from Sardar Patel University, Vallabh Vidya Nagar, in 1993 and his ME degree in Industrial Electronics from M.S. University of Baroda in 1995. He received a Ph.D. degree in Electrical Engineering from IIT Kanpur in 2011. His research interests broadly include performance analysis and simulation of wireless communication systems, cognitive radio, signal processing, and remote sensing.